Master of Science Thesis in Electrical Engineering
Department of Electrical Engineering, Linköping University, 2020

Verification of Powertrain Simulation Models Using Machine Learning Methods

Khalid Pirgul and Jonathan Svensson

LINKÖPING
UNIVERSITY

Master of Science Thesis in Electrical Engineering

Verification of Powertrain Simulation Models Using Machine Learning Methods:

Khalid Pirgul and Jonathan Svensson

LiTH-ISY-EX--20/5299--SE

Supervisor: **Viktor Leek**
ISY, Linköping University
Jianning Zhao
CEVT AB

Examiner: **Lars Eriksson**
ISY, Linköping University

Division of Vehicular Systems
Department of Electrical Engineering
Linköping University
SE-581 83 Linköping, Sweden

Abstract

This thesis is providing an insight into the verification of a quasi-static simulation model based on the estimation of fuel consumption using machine learning methods. Traditional verification using real test data is not always available. Therefore, a methodology consisting of verification analysis based on estimation methods was developed together with an improving process of a quasi-static simulation model.

The modelling of the simulation model mainly consists of designing and implementing a gear selection strategy together with the gearbox itself for a dual clutch transmission dedicated to hybrid application. The purpose of the simulation model is to replicate the fuel consumption behaviour of vehicle data provided from performed tests. To verify the simulation results, a so-called ranking model is developed. The ranking model estimates a fuel consumption reference for each time step of the WLTC homologation drive cycle using multiple linear regression. The results of the simulation model are verified, and a scoring system is used to indicate the performance of the simulation model, based on the correlation between estimated- and simulated data of the fuel consumption.

The results show that multiple linear regression can be an appropriate approach to use as verification of simulation models. The normalised cross-correlation power is also examined and turns out to be a useful measure for correlation between signals including a lag. The developed ranking model is a fast first step of evaluating a new vehicle configuration concept.

Acknowledgements

First of all, we would like to thank CEVT AB for giving us the opportunity to write this thesis. A special thanks to our supervisors, Jianning Zhao and Simon Klacar at CEVT AB for providing us with guidance and helpful discussions throughout this period. The whole Powertrain Strategy Team has been very welcoming and it has been enriching to work in the team within this field. Additionally, we would like to express our gratitude to our examiner Lars Eriksson and our supervisor Viktor Leek at Linköping University for all of the support.

Gothenburg, June 2020
Khalid Pirgul and Jonathan Svensson

Contents

List of Figures

List of Tables

Notation

Notation	Definition
v	Velocity
a	Acceleration
t	Time
P	Power
T	Torque
ω	Angular velocity
F	Force
Q	Electric charge
I	Current
U	Voltage
U_{oc}	Ideal open-circuit voltage
η_e	Thermal efficiency
H_l	Fuel lower heating value
r	Fuel rate
γ	Gear ratio
c_r	Rolling resistance coefficient
m_v	Vehicle mass
ρ_a	Air density
A_f	Vehicle frontal area
c_d	Drag coefficient
g	Gravitational constant
r_w	Wheel radius
α_{max}	Maximum road incline
σ	Standard deviation
R	Cross-correlation
ρ	Pearson's correlation coefficient
β_i	Regression coefficient
λ	Tuning parameter

Abbreviations	Meaning
M&S	Modelling and simulation
FT	Fuel tank
EM	Electric motor
ICE	Internal combustion engine
HEV	Hybrid electric vehicle
SoC	State of charge
AC	Alternating current
DC	Direct current
QS	Quasi-static inverse simulation
EV	Electric mode
ODE	Ordinary differential equations
PDE	Partial differential equations
BSFC	Brake-specific fuel consumption
DCT	Dual clutch transmission
TCU	Transmission control unit
OOP	Optimal operation point
OEM	Original equipment manufacturer
ECU	Engine control unit
NCCP	Normalised cross-correlation power
MLR	Multiple linear regression
Li-ion	Lithium-ion
LASSO	Least absolute shrinkage and selection operator

1 Introduction

A personal vehicle enables a flexible living and is nowadays taken for granted for many people. A challenge due to the increasing amount of vehicles is the energy demand and supply together with more strict legislated restriction of fuel economy and emission pollution.

By adding an electric motor to the powertrain as a second prime mover, the usage of the internal combustion engine can be reduced and optimised. As the electrical motor is propelling the vehicle, the local emission pollution is none. If the electricity comes from a sustainable source, the environmental footprint is reduced compared to a conventional vehicle.

A time- and cost-efficient way of ensuring that the powertrain configuration fulfils all the requirements is to perform simulations of the powertrain. The purpose of a simulation model is to reflect the actual behaviour of the real world. Therefore, it is essential to build the model based on reliable test data of the vehicle components. Verification of the simulation model is also crucial for the interpretation of the results.

1.1 Motivation

Today, modelling and simulation M&S are broadly practised in the engineering industry. It is faster to create a model and simulate it in an appropriate environment as well as the simulations can be conducted faster than real-time testing. Therefore, Research & Development departments in companies virtually simulate the products, which are more cost-efficient than real-world testing.

As George Box once said "*All models are wrong, but some are useful*" is a quote explaining the importance of having a reliable and valid model. A model is developed for a specific purpose and needs to be verified and validated concerning its specific purpose. There are a lot of different approaches to verify and validate a model, and none of them is perfect for every scenario. A valid model is "accurate enough" for its purpose. It is a balance on time spent of developing the model and its accuracy of reflecting the real world. Accurate enough is the amount of accuracy required for the model's intended purpose.

One way to improve and evaluate a model is to run the simulation and compare it with actual test data. By comparing the simulation- and the test data, it is possible to improve the model, which implies a higher correlation. The correlation can, therefore, be used as an indicator of how reliable the model is, and a ranking model can be developed in order of scoring the reliability.

A ranking model can provide a good indicator of understanding how well the simulation data can be trusted. A model usually has a lot of different parameters and, therefore, acts differently in various simulation environments, and a model cannot be trusted the same if any parameter or input is changed. It can, however, based on the correlation between simulation- and test data, indicate how well the simulation data can be trusted compared to a predicted value using machine learning methods.

Since M&S is widely adopted throughout the industry for the development of new systems for vehicle models, engineers develop lots of versions of simulation models. In the vehicular industry, simulation models are developed to investigate performance in early design steps. A fast-running, quasi-static based simulation tool has partly been developed at CEVT AB, and it would be beneficial to have indicators of how well the simulations reflect the real-world fuel consumption. Test data from the drive cycle has been collected and is about to be correlated to the simulation data to be able to develop a robust ranking model and to improve the accuracy of the simulation model.

1.2 Objective

The objective of this thesis is to develop a ranking model which evaluates the simulated fuel consumption without access to corresponding test data using machine learning methods. To develop this model, knowledge of correlation, verification and validation will be a central part together with knowledge of statistics and probability which, therefore, will be introduced in this thesis. The simulation model which will be evaluated does not contain the correct transmission model and needs to be developed. The transmission model will be developed to replicate together with the rest of the quasi-static model the fuel consumption behaviour of the vehicle model and provided test data. It will also contribute to a better understanding of how to model and simulate hybrid electric vehicles and their components.

1.3 Problem Definition

Based on the objective of this thesis, there exist several tasks which have to be solved to achieve a ranking model able to evaluate the results of a simulation model using machine learning methods. A list of four tasks states the most relevant problems, which has to be solved to answer the question: How can the reliability of a simulation model be evaluated based on collected data of vehicle components?

- Develop a dual clutch transmission for the current simulation model to recreate the behaviour of measured fuel consumption from the test vehicle.
- Determine the model error between the simulation model and test vehicle data based on correlation in fuel consumption.
- Implement a gear selection strategy for the dual clutch transmission corresponding to the existing strategy in the test vehicle.
- Design and develop a model, ranking the performance of the simulation model based on predictions of fuel consumption.

1.4 Outline

A short description of the content of the thesis is described here.

Chapter 1: Introduction
In this chapter motivation, objective and problem definition are described.

Chapter 2: Theory
In the theory chapter, all relevant theory for understanding this thesis is presented.

Chapter 3: Related Research
This chapter present relevant related research used as inspiration for the base of the work.

Chapter 4: DCT Model Development
This chapter contains the development of the DCT simulation model.

Chapter 5: Ranking Model Development
The development of the ranking model is presented in this chapter.

Chapter 6: Results
This chapter contains the results for the development of the DCT model as well as the results for the development of the ranking model.

Chapter 7: Discussion
This chapter contains the analysis of the results and the method presented in chapter 4, 5 and 6.

Chapter 8: Conclusion

Conclusions are presented together with the future work.

2 Theory

This chapter presents the theory needed to have an adequate understanding of the report.

2.1 Hybrid Electric Vehicle

The automotive industry is getting more restricted when it comes to emitted emission rates. The legislated emission rates are decreasing every year, making the development of efficient powertrains more crucial. The hybrid electric vehicle (HEV) allows not only fossil fuel as the energy carrier but also electrical energy by including an electric motor (EM) into the drivetrain. The advantage of using a combination of an electric motor and an internal combustion engine (ICE) is the possibility of optimising the use of each engine. By letting the ICE operate in more efficient areas on the brake specific fuel consumption map, the fuel economy increases. An EM has the advantage of zero local emission, no demand of idling at stand-still and the possibility to charge the battery using the EM as a generator, called recuperation. The main advantage for the ICE is the energy capacity in the fuel tank and the power density[7].

Currently, there is a couple of different configurations of the HEV. Series, parallel, mild and a combination of those, are some of them. The model of interest is a parallel HEV, and this thesis will, therefore, only consider the parallel HEV, which is explained in the following section.

2.1.1 Parallel Hybrid Electric Vehicle

The parallel HEV is the configuration where at least two energy sources mounted in parallel are used for the propelling of the vehicle. The parallel configuration enables optimisation of the power split between the ICE and the EM to increase the fuel economy and reduce emissions. It is possible to provide a traction force either by using the ICE, using the EM or both simultaneously. Benefits of the parallel hybrid, compared to a conventional vehicle is the possibility of downsizing the ICE and still fulfil the requirements of the maximum power due to the possibility of using the ICE at its maximum torque together with power from the EM. The configuration also enables the ICE to turn off at idling[7].

A configuration of the parallel hybrid powertrain can be observed in figure 2.1 below. The ICE and the EM are often separated by a planetary gear engaging different or multiple of the planet gear members resulting in the possibility of using the ICE and the EM separately or simultaneously. The ICE-side is working in the same way as for a conventional vehicle using the energy from the fuel tank. The EM-side consists of a battery, a power converter and the electric motor. The battery is the source of energy to the EM. To be able to feed the EM, it has to be converted to AC power which is done by the inverter included in the power converter, converting the DC power to AC power. It also includes a converter, changing the voltage to the correct voltage for feeding the EM.

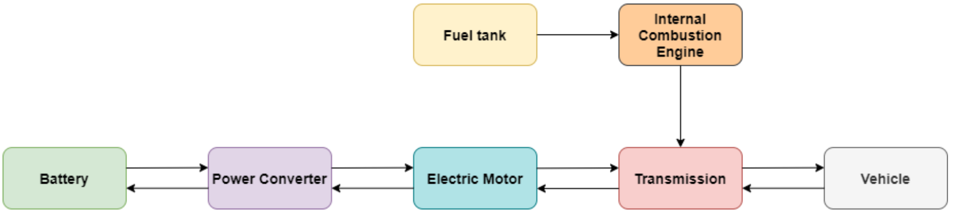

Figure 2.1: *Parallel hybrid configuration, arrows correspond to the energy flow through the powertrain.*

2.2 Powertrain Simulation

Powertrain simulation is one of the main steps of testing and specifying concepts during vehicular development. Using a drive cycle as an input, developers can benchmark fuel consumption in a particular model. Simulations can be performed in various environments using different approaches. The most common approach for powertrain simulations is the forward dynamic simulation, but the powertrain simulation in this thesis is using the quasi-static (QS) inverse simulation. Both are, therefore, explained in the following subsections.

2.2.1 Quasi-Static Simulation

The main purpose of the QS method is to simplify the modelling part and to decrease the simulation time, which is performed by inverting the physical causality chain [10]. This method uses vehicle speed, acceleration and grade angle as inputs that goes through the powertrain and provides the required torque for the propelling of the vehicle. When the wheel torque is calculated, the fuel flow is derived[1]. When the drive cycle is provided for the simulation, the speed, the acceleration and the slope are constant for period time, h. This is the main essence of the QS simulation, the test cycle is divided into time intervals h. These intervals are chosen to be small enough to satisfy the assumption of the velocity and the acceleration being constant as shown in equation (2.1) and equation (2.2).

$$v(t) = \overline{v}_i = \frac{v_i + v_{i-1}}{2}, \quad \forall \quad t \in [t_{i-1}, t_i] \tag{2.1}$$

$$a(t) = \overline{a}_i = \frac{v_i - v_{i-1}}{h}, \quad \forall \quad t \in [t_{i-1}, t_i] \tag{2.2}$$

The QS method has drawbacks as well. Since it is a backward simulation, the physical causality is not considered, and the inputs have to be known a priory. Therefore, QS is unable to handle problems with feedback control [7].

2.2.2 Forward Dynamic Simulation

The forward dynamic simulation approach is, unlike the quasi-static approach a forward simulation, simulating the time-varying behaviour of the system. The motions of the system can be described by ordinary differential equations (ODE) or partial differential equations (PDE). These equations are often nonlinear since most of the real-world events are not increasing- or decreasing linearly. Numerical methods are needed to solve these very time-consuming equations. For the fuel consumption calculation, the forward dynamic simulation results in a more accurate calculation but with loss in calculation time [7].

2.2.3 Quasi-Static Modelling

Modelling a complete hybrid electrical vehicle (HEV) model can swiftly become complex. A common practice is to split the model into separated sub-models representing the components of the vehicle. This method enables the possibility to work with a modular model with autonomous sub-models. Each sub-model has a clear input-output, and the link between is determined by the power flow. With this approach, components in the powertrain can easily be replaced, and different vehicle configurations can be simulated. Every sub-model except the transmission is provided for this thesis work and is, therefore, only briefly explained in the following subsections. Figure 2.2 represent the input- and output parameters of the quasi-static approach for a parallel hybrid. $\boldsymbol{P}$ is the power, ω is the rotational speed, $\boldsymbol{T}$ is the torque, $\boldsymbol{v}$ is the velocity and the $\boldsymbol{F}$ is the force [7].

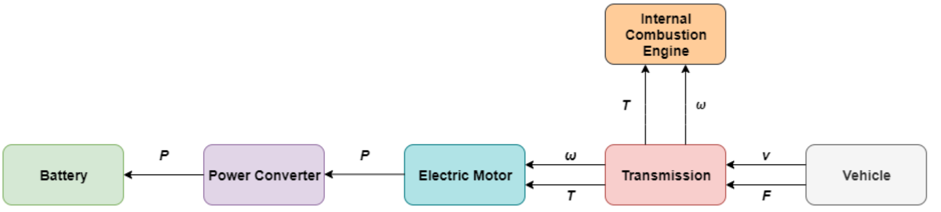

Figure 2.2: Illustration of quasi-static approach for a parallel hybrid. Power flow factors are represented as arrows.

Battery/ Power Converter

One reason why the HEV- and the EV configuration are becoming more common is the progress in battery technology. Initially, the battery was built with lead-acid, which is the oldest rechargeable battery system. It is economically priced but unfortunately also very toxic which complicates the disposal of the batteries. Today, the Lithium-ion (Li-ion) batteries are primly used in many applications that were previously served by other battery technologies. Li-ion batteries are well known for its high specific energy and low maintenance. This technology is more expensive, but due to high cycle count, the cost per cycle is advantageous[7].

When modelling batteries in the HEV, some design specifications are more vital than others. A critical measure is the battery capacity during operation, which has to be accurately determined, also called state of charge (SoC). SoC is a dimensionless parameter that insinuates the percentage of the remaining charge of the nominal capacity [7].

The input for the batteries when using quasi-static approach is the terminal power $P_2(t)$, and the output is the battery charge $Q(t)$, the illustration if the model is shown in figure 2.3.

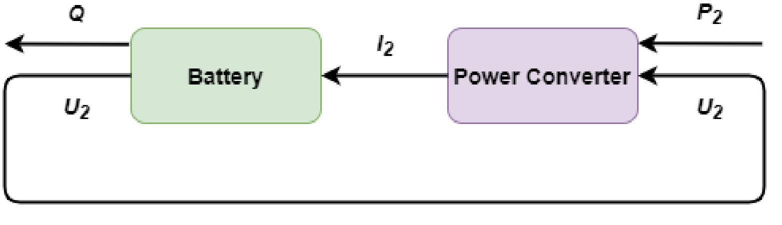

Figure 2.3: Sketch of quasi-static modelling for battery with the power converter.

State of charge q, which is time-dependent, is defined as the ratio between the electric charge Q, and the battery's nominal capacity Q_0. Since direct measurement of Q is not common in the automotive industry, it can be approximated by integrating the discharge current I_2. Equations (2.3) - (2.5) describes the necessary calculations for determining the output variable for the battery model.

$$I_2(t) = \frac{P_2(t)}{U_2(t)} \tag{2.3}$$

$$\dot{Q}(t) = -I_2(t) \tag{2.4}$$

$$I_2 = -Q_0(SOC_{final} - SOC_{start}) \tag{2.5}$$

To determine the terminal voltage U_2 a basic physical model of the battery can be derived by considering an equivalent circuit of the system. Kirchoff's voltage law, equation 2.6, can be used to determine the terminal voltage U_2, where U_{oc} is the ideal open-circuit voltage source represented by the battery.

$$U_2(t) = U_{oc}(t) - R_i(t) \cdot I_2(t) \tag{2.6}$$

Electric Motor

Initially, the use of an electric motor in conventional vehicles was as a starter to boost the engine to its idle speed. Today, the EM is used in electric- and hybrid vehicles as a propulsion component, propelling the vehicle. A crucial advantage of the electric motor is the ability to operate in different modes. First, in propulsion mode by converting electrical power from the battery to mechanical power. Secondly, as a generator converting mechanical power from the ICE to electrical power, recharging the battery. Third, also as a generator, by recuperating mechanical power using the kinematic energy to slow down or stop the vehicle, also called regenerative braking[7].

The subsystem is illustrated in figure 2.4:

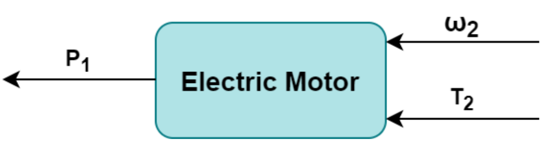

Figure 2.4: *Illustration of quasi-static modelling for electric machine.*

When an electric machine is modelled using the quasi-static approach, the inputs are the required gearbox angular speed ω_2, and the required gearbox torque T_2, at the shaft. The output power P_1, is the power at the DC link. The required power at the shaft is calculated according to:

$$P_2(t) = T_2(t) \cdot \omega_2(t) \tag{2.7}$$

The efficiency of the EM is a function of the input parameters and can be used to determine the power at the DC link, P1. The power flow shows whether the EM is operating as a motor or generator, a positive value of P meaning the machine operating as a motor and negative value of P meaning the machine operating as a generator. The equation (2.8) describes how the EM power is calculated where η_m is a stationary efficiency map as function of the input variables T_2 and ω_2.

$$\begin{cases} P_1(t) = \frac{P_2(t)}{\eta_m(\omega_2(t), T_2(t))} & \text{if } P_2(t) > 0 \\ P_1(t) = P_2(t) \cdot \eta_m(\omega_2(t), -T_2(t)) & \text{if } P_2(t) \leq 0 \end{cases} \tag{2.8}$$

Internal Combustion Engine

The primary mover in the HEV, the internal combustion engine (ICE) has been around for centuries, resulting in well-elaborated engines. Figure 2.5 shows an illustration of the ICE.

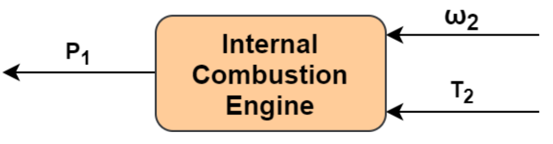

Figure 2.5: *Illustration of quasi-static modelling for ICE.*

The efficiency of the ICE is typically compared using the brake-specific fuel consumption maps (BSFC), which is the rate of fuel consumption r, divided by the produced power P_1:

$$BSFC = \frac{r}{P_1} \tag{2.9}$$

The thermal efficiency of an ICE is remarkably lower than the efficiency of an EM. A map of the ICE efficiency can be observed in figure 2.9 depending on the engine speed and the engine torque. The thermal efficiency is the engine power divided by the produced power.

$$\eta_e = \frac{P_2}{P_1} = \frac{\omega_2 \cdot T_2}{P_1} \tag{2.10}$$

Transmission

Transmission is often referred to the power transmission system that transforms mechanical power with speed ω_1 and torque T_1 to a different speed ω_2 and T_2 utilising the gear ratio γ. Transmissions can also be modelled using a quasi-static approach, illustrated in figure 2.6. The modelling of the transmission is a crucial part of this thesis. Therefore, a more detailed theoretical explanation is presented in the next section.

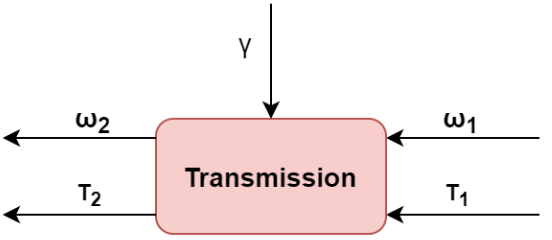

Figure 2.6: Illustration of quasi-static modelling for gearbox.

Vehicle

The input variables for QS is as mentioned before, vehicle speed, acceleration and the grade angle, which are given by the driving cycle. The output of the vehicle is the speed of the vehicle together with force, which can be seen in figure 2.7.

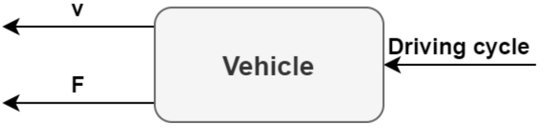

Figure 2.7: Illustration of quasi-static modelling of the vehicle.

The driving cycle is divided into short time periods where the input variables are assumed to be constant. The total force acting on the wheels for the given driving cycle can then be calculated using Newton's second law as:

$$\bar{F}_{t,i} = m_v \cdot \bar{a}_i + F_{r,i} + F_{a,i} + F_{g,i} \tag{2.11}$$

F_r is the rolling resistance:

$$F_r = c_r \cdot m_v \cdot gcos(\alpha) \tag{2.12}$$

where c_r is the rolling resistance coefficient, m_v is the vehicle mass and α is the grade of the road in rad.

F_a is the aerodynamic resistance:

$$F_a = \frac{1}{2} \cdot \rho_a \cdot A_f \cdot c_d \cdot \bar{v}^2 \tag{2.13}$$

where ρ_a is the density of air, A_f is the frontal area of the vehicle and c_d is the drag coefficient.

F_g is the gravitational resistance:

$$F_g = m_v \cdot g \cdot sin(\alpha) \tag{2.14}$$

where g is the gravitational constant.

2.3 Dual-Clutch Transmission

The dual-clutch transmission uses two input shafts, one for odd gears and one for even, connected through two clutches. It can, therefore, be described as two separate manual sub-gearboxes with respective clutch contained in one housing. The benefits of the DCT is that it can use the efficiency of a manual transmission together with the shifting ease of an automatic transmission. The gear ratio can also be selected in a broad range, and the drivability is increased by preselecting the next gear, a smooth and effective shifting can be done by disengage the clutch for the gear in use and engage the upcoming gear. By dividing the gears into an odd and an even shaft, the DCT becomes fully power shiftable, meaning the power flow is never interrupted through the powertrain [14].

Figure 2.8 shows a schematic diagram of a dual-clutch transmission. The sub-gearboxes are not arranged side-by-side in the actual vehicle but is instead nested together to save space.

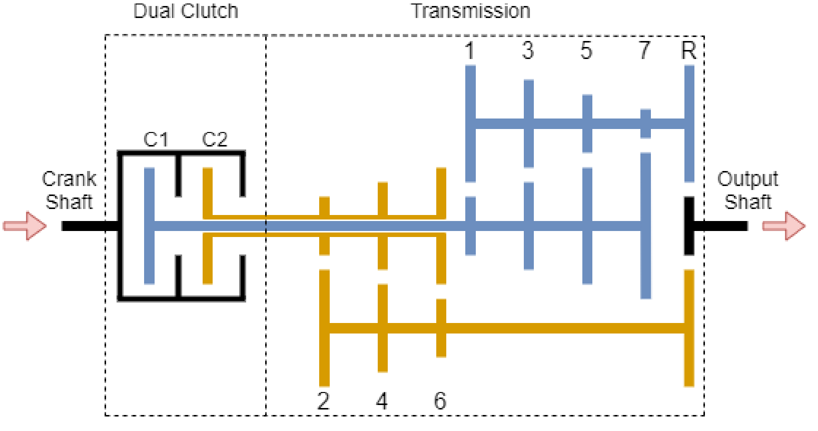

Figure 2.8: *Dual-clutch transmission, where C1 and C2 are clutch number one and two repetitive.*

Assume that a vehicle is driving in second gear which is engaged by clutch C2 and is about to upshift, the gear shifting process is then as follow:

1. The shaft containing the third gear, which is not engaged (clutch C1), preselect the third gear by engaging it. This synchronising process is not noticeable by the driver.

2. The clutch C2, for the shaft of the second gear, is then opened in the same time as the clutch C1, for the third gear, closes. The power flow is, therefore, never interrupted.

3. The shaft of clutch C2 is now free, and another gear available on the shaft can now be preselected for the upcoming gear shift.

4. The upcoming gear is decided by vehicle speed together with throttle position.

5. The process of shifting is the same for both up- and downshifting[14].

The DCT can either be a wet-running system or a dry-running system. The wet-running system is in continual need of a hydraulic supply system for clutch activation and cooling. This result in a higher torque capacity for the wet-running system compared to the dry-running. Therefore, the use of the wet-running system tends toward the market, which requests higher torque and the dry-running system toward the market where less torque is requested. The DCT requires a higher starting torque compared to other transmissions due to its lack of torque increase from the torque converter, and the overall gear ratio is therefore higher. One additional gear is often used to prevent the gear steps of getting too large [14].

2.3.1 Electronic Transmission Control Unit (TCU)

As transmissions turn from manual to automated transmissions, the driver is not supposed to change gear anymore, and new actuators are needed for the clutch and gear-shifting. The new actuators are dependent on signals telling the system to change gear when needed, which is the task for the TCU. The TCU is the "computer centre" of the transmission, it receives signal and sensor information, converts them, evaluates them and provides the actuators with control values. The input- and output parameters for the TCU differ to some extent for different automatic transmissions but are mainly the parameters presented in the table 2.1[14]:

Table 2.1: *Main input- and output parameters for the TCU.*

Input parameters:	**Output parameters:**
Vehicle speed	Shift lock
Wheel speed	Shift solenoids
Throttle position	Pressure control solenoids
Input shaft speed	Torque converter clutch solenoids
Transmission oil temperature	ECU
Kick down switch	Other controllers
Brake light switch	
Cruise control	

For automated manual transmissions, the functions of a transmission falls under three sub-functional groups, vehicle functions, basic functions and hardware-related functions. These functions groups contain different control systems presented in table 2.2 below:

Table 2.2: Function overview automated manual transmission

Vehicle functions	Basic functions	Hardware-related functions
OEM diagnostics	Gearshift	Operating system
OEM communtication	Engine control	Basic diagnostic
Powertrain management	Gearbox brake	BIOS
Brake management	Clutch control	Standard software modules
Special functions	Power take-off control	Basic communication
Central computer interconnection	Safety functions	
Driving strategy	Driving strategy	

Driving strategy is an essential function of an automated transmission which falls under both the vehicle- and basic function. A strategy to optimise the fuel consumption and emission rates but at the same time, meet driver expectations together with engine protection and other operations needs to be implemented in the TCU. It is a difficult task which is further explained in section 2.3.2 referred to as the gear selection strategy. As the calculations are based on so many parameters which in turn require data from many sensors and other components in the vehicle system, a well developed diagnostic and emergency system is required. In fact, the diagnostic and emergency system often makes at least 50% of the total extent of software and functions in the TCU [14].

2.3.2 Gear Selection Strategy HEV

The gear selection strategy decides which gear should be engaged for the even and the odd shaft and also when to shift gear. The strategy differs between different vehicles and configurations, but the main goal is always to run the ICE close to the optimal operation point (OOP) for each gear. The transition from a conventional vehicle to an HEV increases the complexity a lot for the gear shifting, the ICE load is not strictly proportional to the required wheel power anymore and can instead be achieved by several combinations of a load from the ICE and the EM. The EM can be used both as a motor and a generator, moving the operational point in the engine efficiency map both up and down, called load point charging/boosting [6], illustrated in figure 2.9. The efficiency map shows the efficiency of the engine at certain engine speed and engine torque.

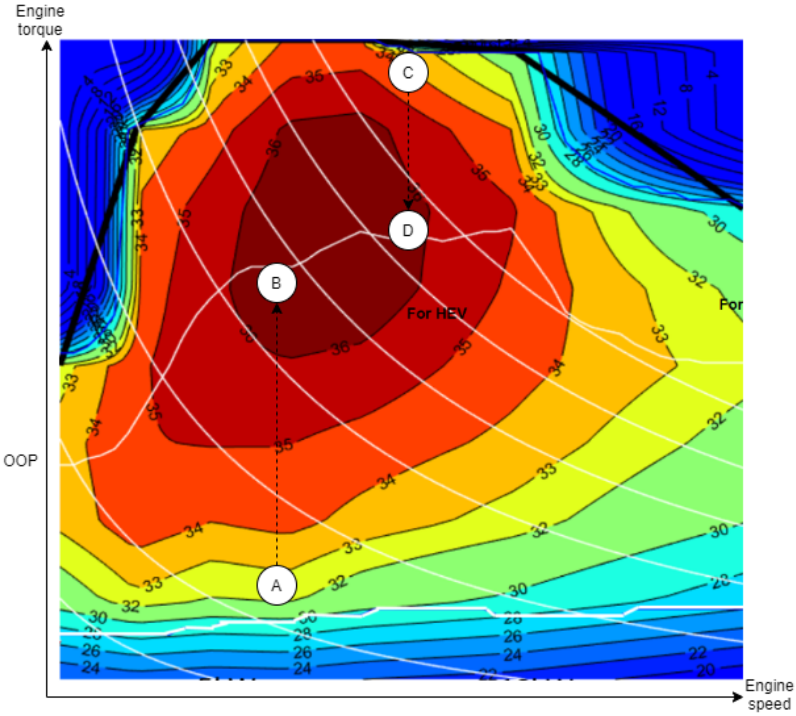

Figure 2.9: *Load point charge (A-B) and load point boost (C-D) illustrated in the efficiency map of an ICE.*

By adding a charge of about 20kW to the EM, the ICE has to increase the work which moves the working point from point A to point B (charging). The working point can also be moved down if necessary by providing power from the EM (boosting), visualised as moving from point C to D in the engine map.

In some situations, a more efficient operating point may be found by downshifting and then charge the battery. This may be the case if the operational point is close to the OOP, but a more efficient area can be entered by downshifting and charge the battery. This is shown in figure 2.10 below. There exist more examples of when it can be beneficial to downshift to let the ICE work in a more efficient area.

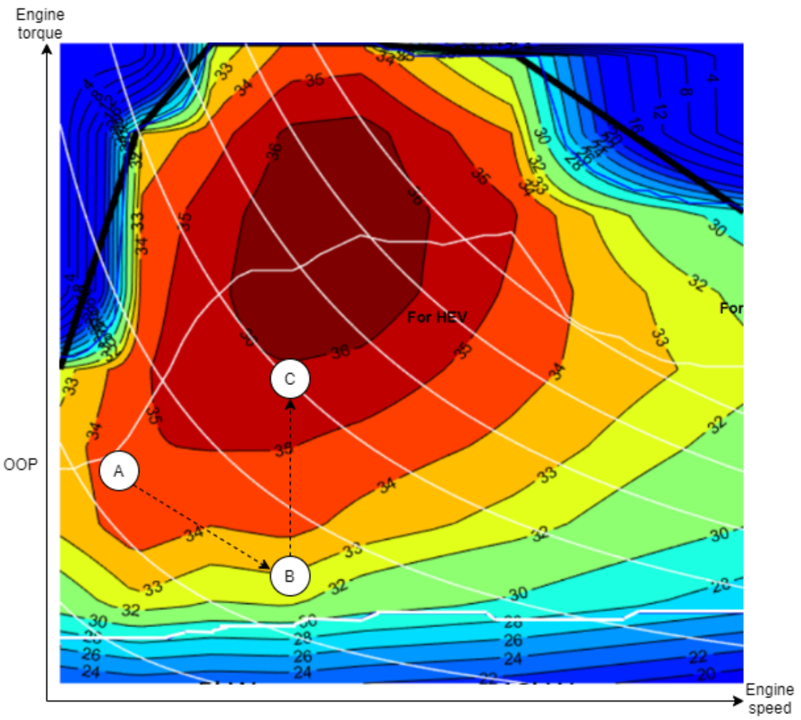

Figure 2.10: Load point shift charge illustrated in the efficiency map of an ICE.

Downshifting from point A result in increased engine speed and decreased torque (point B) and by then increasing the workload for the engine to charge the battery, point C is met which is a more efficient area than point A.

The control algorithms for the gear selection strategy have to be able to find and select the optimal gear based on not only the optimal operation point in the engine map but also constraints caused by different driving states, drive modes and other conditions. Conditions such as standstill, acceleration, deceleration and SoC conditions. For the HEV, there also exist condition of the driving states. The gear selection depends on if the vehicle is in the EV-mode, ICE-mode, regenerative braking or load point charge/boost. The driving states is further explained in chapter 4 about DCT. A solution often used is to construct a gear selection map, which based on vehicle speed and throttle position decides which gear to engage. The usual appearance of the map is that upshifting tend to happen at higher vehicle speeds compared to downshifting at the same gear.

2.3.3 Modelling of the Dual-Clutch Transmission

The most fundamental equation is how the torque and the speed is changed through the gearbox. By neglecting all losses in the transmission, the equations are as follow:

$$\omega_1 = \gamma \cdot \omega_2 \tag{2.15}$$

$$T_2 = \gamma \cdot T_1 \tag{2.16}$$

Where ω_1 and T_1 are the rotational speed, and the torque at the power source which are changed to a new rotational speed ω_2 and torque T_2 depending on the gear ratio γ is the parameter that differs between the gears.

The gearbox efficiency is of course never 100% and, therefore, it has to be taken into account when modelling the transmission. Two separate equations can be used to approximate the losses in the gearbox using an affine dependency between the gearbox input and output power. The losses depend on several factors such as speed, load, temperature, friction and the equations are only approximations of the losses. The equations differ based on torque being positive or negative:

$$\begin{cases} T_2 \cdot \omega_w = e_{gb} \cdot T_1 \cdot \omega_e - P_{0,gb}(\omega_e) & \text{if } T_1 \cdot \omega_e > 0 \\ T_1 \cdot \omega_e = e_{gb} \cdot T_2 \cdot \omega_w - P_{1,gb}(\omega_e) & \text{if } T_1 \cdot \omega_e \leq 0 \end{cases} \tag{2.17}$$

Where $P_{0,gb}$ is the power needed for the gearbox to idle at engine speed ω_e, T is torque and e_{gb} is the efficiency of the gearbox which usually is between 0.95 - 0.97. P_{1_gb} is the losses in the gearbox which affect the fuel cut-off torque (the torque limit that causes the fuel being cut from injection, the second equation)[7].

2.4 Verification and Validation

During model development, verification and validation are vital tools for the developer to ensure the accuracy of the model. The developer uses various techniques to ensure the accuracy of the model and how well the results match the specifications, which is defined as verification.

The next step, validation, usually involves more parts. The developer of the model, together with system experts, jointly evaluates the model. The model is compared to the set specifications and the customers' request. Conclusively all the parts assure that the model has an acceptable level of accuracy.

The difference between verification and validation is thus that verification is the process of checking whether the software meets the specifications or not, and the validation process is the process of checking if the specifications capture the customer's need. The main advantage of verification and validation is to detect bugs or incorrect assumptions early in the developing process to correct the detected errors. Both are required in the developing of a successful model [9].

2.4.1 Data Errors and Data Modelling Errors

The leading cause of modelling errors originates from data errors and data modelling errors. Data errors occur when the produced or provided data are incomplete, it is thus in the input data itself. Data modelling errors are instead referred to as the error occurs when the data is used in an incorrect way for the model.

Data Errors

A couple of usual sources causing data errors:

- The mean values are provided instead of individual values. For example, the mean engine torque is provided for a driving cycle instead of individual values for each time step.
- The cause of deceptive data is not provided. For example, if there exists a static error in the data which could have been reduced or removed if the cause were provided.
- The provided data corresponds to simulations outputs. By using simulation data as input in a developed model, there already exist errors in the data due to that a simulation model never is 100% correct, and the new model is, therefore, build depending on reliable data.

Modelling Errors

A couple of usual sources causing modelling errors:

- The mean value of the data is used when the actual values vary randomly. For example, by using the mean torque instead of the individual values probably result in lower fuel consumption of the vehicle.
- Assume behaviours of the model which is not well strengthened. For example, if a certain behaviour found in the data seems to be the case for all outcomes, but the provided data is not enough to tell.
- Using a normal distribution because it is simple but may in many cases not be valid for the model.

Conclusively, the actual data is better than the statistical summary since the summary can be easily extracted if needed[9].

2.5 Statistical Tools

Statistical analysis of data is an essential step of how to interpret the results and how to handle the data. Different measures tell different things about the data, and a crucial part is to find these measures and correctly compare them. The following sections also present different statistical tool of how predictions of data can be constructed through collected data.

2.5.1 Relation Between Two Variables

When two variables, samples, are compared to examine the connection between them, a statistical relationship can be extracted. This relationship can be defined as the correlation coefficient. The correlation is a measure of the linear relationship between variables. The primary purpose of correlation is not only to use one variable to predict the value of the other but study the cooperative behaviour of two variables and their relationship. In this section, different statistical tools is presented[5].

Covariance

Covariance describes how much two variables change together. Assume that there is two random variable X and Y. If there is a positive relationship between them, i.e. large values of X occurs with large values of Y, that means that the density is affiliated with either positive or negative product of deviation from the mean value $(x-\mu_x)(y-\mu_y)$. For a negative relationship, the product of the deviation from the mean value has the opposite sign. If X and Y cancel one another, then the covariance is approximately 0. Equation (2.18) represent the calculation of covariance between two said variables[5].

$$Cov(X,Y) = E[(X-\mu_x)(Y-\mu_y)]$$

$$= \begin{cases} \sum_x \sum_y (x-\mu_x)(y-\mu_y)p(x,y) & \text{if } X, Y discrete \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-\mu_x)(y-\mu_y)f(x,y)dxdy & \text{if } X, Y continuous \end{cases} \tag{2.18}$$

Where X is the first variable containing a sample of values of x, Y is the second variable containing a sample of values of y, μ is the mean and $p(x,y)$ is the probability mass function which is a function that gives the probability that a discrete random variable is exactly equal to some value. f(x,y) is the probability density function, which is a function of a continuous random variable providing the likelihood of a random value in the sample space of the function whose value at any given sample equal that specific sample.

Standard Deviation

Standard deviation can be described as a representative deviation from the sample mean within a set of values. When the standard deviation has a high value, it is indicating a high spread over a wider range and contrary a low value indicates that the values are close to the mean, i.g. the expected value[5]. Equation (2.19) represent a formula for a standard deviation for a discrete random variable.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{2.19}$$

where N is the number of observation in a sample, x_i is the observed value i in the sample and μ is the mean of the values in the sample.

Standard deviation, σ, can be used to describe the variance σ^2, which gives the dispersion for a set of random values from their average value.

Cross-Correlation

This method is used to compare two time series to determine a degree of similarity. This series can be at different times, and cross-correlation finds how the two series match up with each other. The method is particularly efficient at finding a short known feature in a long signal and reveal periodicity in the compared series [18]. Cross-correlation means that the series is compared with an offset, also called lag. When the lag increases, the possibility of finding a match decreases since the ends of the series do not overlap. When the highest correlation coefficient is found, the lag represents the best fit between the series. Equation (2.20) represents how the cross-correlation is calculated, where x and y is the time-series and τ is the lag. Essentially, y is sliding along to find the highest product between x and y. The purpose is to find how much y must be offset to be similar to x.

$$R_{xy}(\tau) = \lim_{T\to\infty}\frac{1}{T}\int_{0}^{T} x(t)\, y(t-\tau) \quad dt \tag{2.20}$$

If this method is used with one signal, it is called autocorrelation. Autocorrelation is particularly useful for searching repeating patterns, especially if the signal is obscured by noise.

Pearson Correlation Coefficient

Pearson's correlation investigates the statistical evidence of a linear relationship between two variables X and Y. The relationship is represented by the correlation coefficient ρ, which is a parametric measure. This method is based on the method of covariance showed above together with the standard deviation of the variables σ_X and σ_Y. The formula for calculating the Pearson's correlation coefficient is as follow:

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \tag{2.21}$$

Pearson's correlation coefficient value exists in the interval [-1,1]. When ρ is greater than 0, the correlation between the variables is positive. If the coefficient is less than 0, there is an inverted relationship between the variables. The magnitude of ρ is represented in table 2.3[20].

Table 2.3: *The degree of association between two variables represented by the Pearson's correlation coefficient.*

Absolute magnitude	**Degree of association**
$0.8 < \rho \leq 1.0$	extremely strong correlation
$0.6 < \rho \leq 0.8$	strong correlation
$0.4 < \rho \leq 0.6$	moderate correlation
$0.2 < \rho \leq 0.4$	weak correlation
$0.0 < \rho \leq 0.2$	extremely weak or no correlation

Relative Error

The relative error is a measure of the absolute error divided by the actual value. It provides the uncertainty of the measure compared to the size of the measurement, i.e. it puts the error into perspective. The relative error is a well-established measurement easy to interpret. The formula is as follow:

$$Relative\ error = \frac{|Measured\ value - Actual\ value|}{Actual\ value} \tag{2.22}$$

2.5.2 Linear Regression

Linear regression is a commonly used type of predictive analysis. The main idea is to fit a straight line as accurately as possible between data points and then use this line to predict the outcome. Different linear regression methods are further explained in the following sections.

Simple Linear Regression

The primary purpose of the simple linear regression is that a straight line can approximate a relationship between two variables. In this model Y is denoted as the dependent variable (the variable which is to be predicted) and X as an independent variable (the variable that the prediction is based on). The straight line is an estimation that predicts the dependent value, Y, as a function of the independent variable X. The straight-line relationship can be approximated by making a scatter diagram of the two variables. If the dependent values vary in a straight-line fashion, then the simple linear regression model is appropriate to describe the relationship between Y and X. In this model, the outcome variable is dependent on a single predictor, hence the name simple[3].

Linear regression consists of finding the best-fitting straight line through given or measured data points. The most commonly used method of finding the best-fitting line is the least-square method which minimises the sum of the squared errors of the prediction. The errors of each data point are shown in figure 2.11 below.

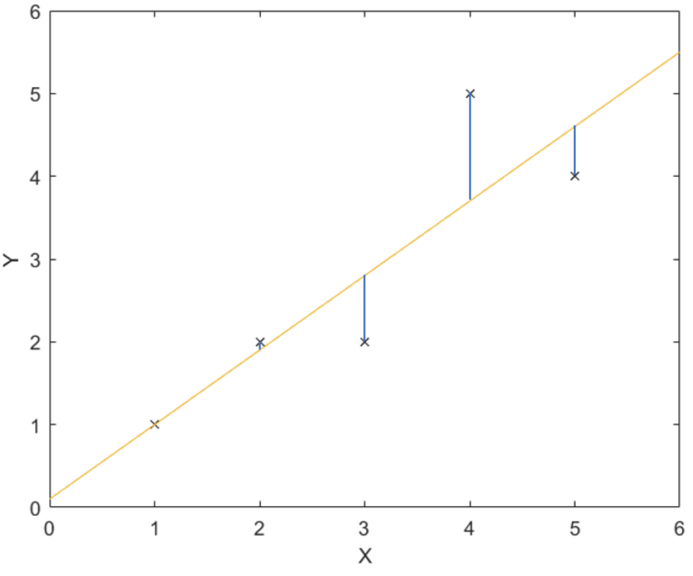

Figure 2.11: *Best-fitting line is the line minimising the sum of the error of each data point, represented by the blue lines in this figure.*

The simple linear regression model has a Y-intercept which is denoted as β_0, and the slope denoted β_1 all of which are represented in equation (2.23). ϵ is an error term that describes the effect on the dependent variable other than the value of the independent variable, for example, measurement error. The slope and the intercept in the equation are called regression coefficients.

$$Y = \beta_0 + \beta_1 X + \epsilon \tag{2.23}$$

Multiple Linear Regression Model

In a situation when a dependent variable is affected by several independent variables, a multiple linear regression model can be used. The model is similar to the simple linear regression model with an extension of several regression parameters[3]. As the number of independent variables increases, the dimension also increases. It is no longer a best-fitting line but a best-fitting plane if there are two independent variables, and the dimensions increase in the same extent as the number of independent variables. Multiple linear regression can be used for several purposes, e.g. to determine the effect of the independent variables on the dependent variable or to predict trends or future values. The model is described in equation (2.24).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + \epsilon \tag{2.24}$$

where Y is the dependent variable, X_i $i = 1, ..., k$ is the independent variables, β_j $j = 0, ..., k$ is the regression coefficients and ϵ is the error term.

To develop an accurate multiple regression model, the following demands have to be fulfilled:

- A linear relationship in the data between the dependent and independent variables.
- A high correlation between independent variables cannot exist.
- The dataset for the dependent variable is chosen independently and random.

Multiple linear regression can be used to estimate, for example, fuel consumption by using it as the dependent variable and other vehicular parameters as independent variables, the example illustrated in figure 2.12.

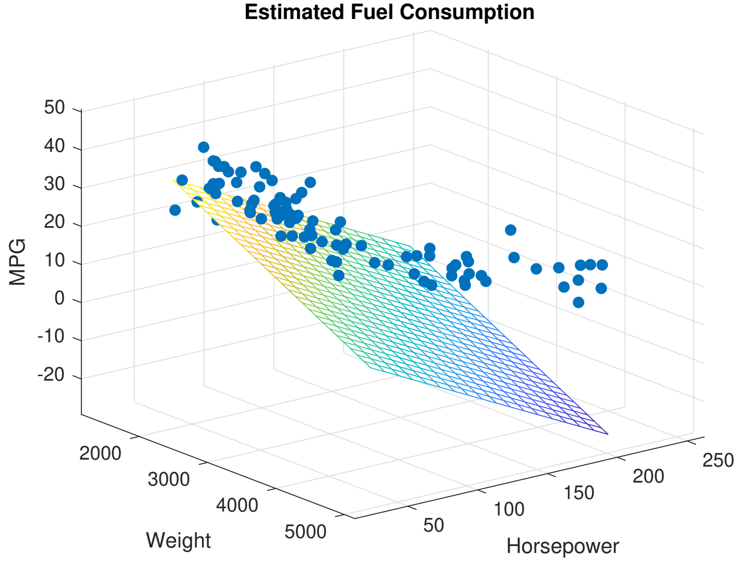

Figure 2.12: *An example of estimated fuel consumption, MPG with two independent variables, vehicle weight and engine horsepower.*

The independent variables can be appended to a matrix **X** and the dependent variable to **Y**. These matrices are described below, where j is the length of each variable, and k is the number of independent variables.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_{j-1} \\ y_j \end{bmatrix}, \quad X = \begin{bmatrix} x_{1,1} & x_{1,2} & . & . & . & x_{1,k-1} & x_{1,k} \\ x_{2,1} & x_{2,2} & . & . & . & x_{2,k-1} & x_{2,k} \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ x_{j-1,1} & x_{j-1,2} & . & . & . & x_{j-1,k-1} & x_{j-1,k} \\ x_{j,1} & x_{j,2} & . & . & . & x_{j,k-1} & x_{j,k} \end{bmatrix} \tag{2.25}$$

The regression coefficient can be calculated by using least point square estimates b_0, b_1, b_2, b_k in the following equation (2.26).

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ . \\ . \\ . \\ b_{k-1} \\ b_k \end{bmatrix} = (X'X)^{-1}X'Y \tag{2.26}$$

With help of this calculation b_k can be called as an unbiased point estimate of β_k which can be used as a regression coefficient for estimating $\hat{Y}$.

LASSO Regression

LASSO is an acronym for the Least Absolute Shrinkage and Selection Operator, which is another method of estimation of linear models. This method uses shrinkage, where data values are shrunk towards the mean. LASSO is used for independent variable selection and regularisation. Since LASSO minimises the residual sum of squares, it can sometimes be unstable, especially if traces of multicollinearity is found (high correlation between independent variables). Therefore, regularisation is particularly useful. LASSO performs *L1* regularisation, which reduces the absolute value of the regression coefficients by adding a penalty factor resulting in that the redundant regression coefficients converge to zero and eliminates. The regularisation significantly reduces the variance of the model, without a substantial increase in its bias. Leading to a simpler model with fewer parameters using only the most crucial variables for estimations of the dependent variable, which can improve the accuracy of the linear regression model. Equation (2.27) describes the LASSO regression where the λ is the tuning parameter describing the amount of shrinkage [17].

$$\hat{\beta}(lasso) = \arg\min_{\beta} \left(\left\| y - \sum_{j=1}^{p} x_j \beta_j \right\|^2 + \lambda \sum_{j=1}^{p} \left| \beta_j \right| \right) \tag{2.27}$$

- $\lambda = 0 \Rightarrow$ the estimates are equal to the ones extracted with linear regression.
- $\lambda \to \infty \Rightarrow$ all regression coefficients are eliminated.

- $\lambda \rightarrow decreases \Rightarrow$ variance increases.
- $\lambda \rightarrow increases \Rightarrow$ bias increases.

The optimal value of lambda can be found by creating several models using e.g. cross validation.

Cross-Validation

Cross-validation is a model evaluation method where a limited dataset is divided into test data and training data. There exist several cross-validation approaches but *k-fold* cross-validation is the one used in this thesis which, therefore, is the one further explained. Cross-validation assesses how different statistical methods perform with an independent dataset, and the central core is to estimate the expected extra-sample error, which is the difference between the predicted value and the test value.

When there is no limitation of provided data, a part of the dataset can be set aside for validation and used for assessment of the prediction model. This case is usually rare, therefore, *k-fold* cross-validation is an effective method where all data is used both as test- and training data by repeating the validation process K times. The data is divided into K subsets of data, and each subset is used once to validate the fitted model which has been trained all data except dataset k, shown in figure 2.13. This process is repeated for $k = 1, 2,..., K$ and all of the estimates of prediction errors are summarised.

A step-by-step approach of performing *k-fold* cross-validation from *The Elements of Statistical Learning* [8] is presented below.

1. Divide the samples into K cross-validation folds at random.
2. For each fold $k = 1, 2,..., K$
   (a) Find a subset of “good” predictors that show fairly strong (univariate) correlation with the class labels, using all of the samples except those in fold k.
   (b) Using just this subset of predictors, build a multivariate classifier, using all of the samples except those in fold k.
   (c) Use the classifier to predict the class labels for the samples in fold k.

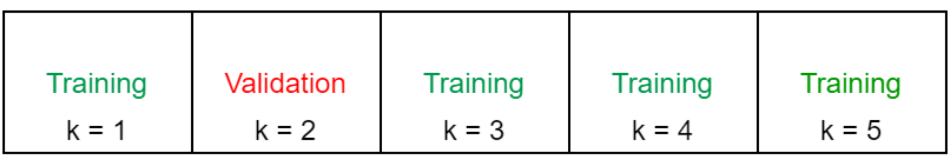

Figure 2.13: *Dataset split into 5 equal-sized sub-datasets, 5-fold. k slides across the data as a validation set.*

3
Related Research

This chapter present relevant related research used as inspiration for the base of the work. The chapter begins by presenting essential papers for the development of the DCT model together with research in the gear selection strategy area, which both together constitutes the modelling part of this thesis. Relevant research is then presented for the intended ranking model area, fuel consumption prediction, correlation and verification analysis.

3.1 Development of a Simulink Powertrain and Hybrid Analysis Tool

In a paper by Lee B., Lee S., Cherry J., Neam A., Sanchez J. and Nam E. [11] presents the development of the vehicular simulation model (ALPHA) and the process of validating the model. Environmental Protection Agency (EPA) has created this simulation model in MATLAB/ Simulink using a forward dynamic simulation approach. The tool can evaluate greenhouse gas emissions (GHG) and fuel consumption for different vehicle configurations and is used to test fuel-saving technologies.

The model contains six subsystems, ambient, driver, electric, engine, transmission and vehicle. The modelling process of the transmission is relevant for this thesis and is, therefore, presented here: The transmission consist of two components, a clutch and a gear, the clutch is modelled as ideal and does not include the slip and efficiency losses when engaging and disengaging the clutch. The gear number and the corresponding gear ratios are defined for each vehicle configuration together with the torque transmitting efficiency, which represents the physical losses in the system.

The model is verified by performing drive cycle tests on a chassis dynamometer which results are compared to the results from the simulation model executing the same drive cycle. The development of the DCT model is going to be verified with test data received from performed tests on the WLTC drive cycle and is, therefore interesting. A second-by-second fuel consumption graph can then be proceeded from the BSFC map and used for comparison with the simulated fuel consumption results. Five different drive cycles are tested and performed three times each to ensure repeatability and reliability. Time trace results of the simulation are first compared with the corresponding second-by-second vehicle data and then the overall fuel consumption and GHG emissions. The following signals are compared for the validation:

- Fuel consumption
- GHG emissions
- Vehicle speed
- Fuel flow
- Engine speed

3.2 Fuel-Optimal Power Split and Gear Selection Strategies for a HEV

In a paper by Ritzmann J., Christon A., Salazar M., and Onder C.[16] presents the development of a fast and efficient optimisation algorithm which determines the optimal state and input trajectories when both discrete and continuous inputs have to be considered. The goal of the optimisation algorithm is to minimise the fuel consumption in a computational time-efficient manner, which, for example, could be used for online optimisation vehicles. The optimisation method is based on Pontryagin's minimum principle (PMP) together with a mixed-integer convex problem and can jointly optimise the torque split, the engine on/off and the gear selection in one optimisation step. The degrees of freedom which have to be set are the clutch position, the selected gear and the torque split between the ICE and the EM.

The driving cycle is known, the optimal torque split can, therefore, be calculated for each gear based on the required torque, the velocity and the equivalence factor. The gears which do violate any constraint, such as rotational speed or maximum torque are excluded. The gear resulting in lowest stage cost is then chosen and applied. An equivalence factor s is introduced to have a charge-sustaining cycle, s relates battery power to fuel power. To be charge-sustaining, the final battery energy has to be equal or greater than the initial battery energy. This constraint is for many values of s, a two-point boundary value problem (TPBVP) is therefore solved to find the smallest value of s which meet the condition. The battery does also have a constraint of the interval from the minimum- to the maximum level of battery energy. By simulating a driving cycle using the optimisation algorithm, and the battery interval constraint is fulfilled along the optimal path, the optimal solution has been found. If instead the constraint is violated anywhere throughout the cycle, the TPBVP is extended to a multi-point boundary value problem (MPBVP). This leads to a value of s, which is changed every time a boundary of any constraint is hit. The method of finding the optimal solution is explained here:

1. Run the driving cycle using the optimisation algorithm if none of the constraints is violated the optimal solution is found.

2. If any constraint is violated, the driving cycle is split at the time step where the largest violation is detected. The target battery energy is then set to the path constraint for that point.

3. The optimisation is then run again for the two segments. If any path constraint is still violated, the driving cycle is once again divided at the point where the largest violation is detected.

4. The procedure is repeated until no violations appear.

5. The optimal solution for the algorithm has then been found, which is not guaranteed to be the global optimum.

The method is tested in two different driving cycles, and calculation time, fuel consumption and the difference between the initial and the final SoC is compared to a dynamic programming approach which is commonly used. The result can be observed in figure 3.1 below.

	Method	t_{calc}	$\bar{m}_{f,abs}$	$\xi(t_{fin}) - \xi(0)$
			l/100 km	-
WHVC	DP	4.3 h	6.801	0
	PMP	0.35 s	6.809	0.00017
SWISSelv	DP	6 h	4.703	0.0415
	PMP	0.7 s	4.728	0.0400

Figure 3.1: *Comparison of the Dynamic programming (DP) approach and the developed approach based on PMP. Where WHVC and SWISSelv are two different driving cycles and ξ is the SoC.*

As figure 3.1 shows, the fuel consumption and the difference in SoC are minimal compared to computational time which has been decreased from multiple hours to less than a second.

3.3 Energy Management and Shift Control for Dual-Clutch Transmission

In a paper written by Guoqiang Li and Daniel Görges [12] addresses the integration of shift control and energy management to decrease fuel consumption and emissions for parallel hybrid vehicles with dual-clutch transmission as well as increasing the drivability simultaneously. The paper contains a part of the development of an engine torque smoothing control which is out of scope for this thesis work and is therefore not summarised.

Dynamic programming (DP) is commonly used to solve optimal control problems with continuous and discrete variables. It is therefore often used to optimise the fuel economy based on gear selection and torque split. By only optimising the fuel economy, the result tends to include a lot of gear changes throughout the driving cycle, which decreases the drivability and can cause additional energy use and wear. A novel varying weighting factor (VWF)-based DP algorithm is implemented to increase the drivability and find a balance between fuel consumption and the drivability.

The gear shift strategy uses the current gear position to control the gear position of the next time step $g(k+1)$ according to

$$g(k+1) = g(k) + u_g(k) \tag{3.1}$$

Where k corresponds to the discrete-time step, and u_g belongs to the set $\{-1, 0, 1\}$, where -1 represent downshift, 1 represent upshift and 0 represent sustainment. Jumping gears is, therefore, not allowed for this strategy. $u_p(k)$ is defined as the power distribution coefficient, which is a continuous variable in the interval [0,1]. The optimal control law $\mathbf{u} = [u_p(k), u_g(k)]^T$ is designed to minimise the fuel consumption together with the gear shift frequency during the whole driving cycle which corresponds to minimising the cost function

$$J = \sum_{k=0}^{N-1} \left(\dot{m}_f(k) + \beta u_g(k) \right) \quad \Delta t \tag{3.2}$$

Where N is the length of the driving cycle, $\dot{m}_f$ is the fuel rate which in turn is a function depending on the engine torque and engine speed, Δt is the sample time, and finally, β is the weighting factor for the shift events.

There are a number of constraints listed to ensure normal operation of each component

1. $SOC_{min} \le SOC(k) \le SOC_{max}$
2. $0 \le T_e(k) \le T_{e,max}$
3. $T_{m,min} \le T_m(k) \le T_{m,max}$
4. $\omega_{e,min} \le \omega_e(k) \le \omega_{e,max}$
5. $P_{b,min} \le P_b(k) \le P_{b,max}$
6. $SOC(0) = SOC(N)$

Where constraint 1-5 tell that each parameter should stay within its interval, and the sixth tell that the cycle should be charge sustaining, i.e. the end SOC should be equal to the initial SOC. The state variables are chosen as $\mathbf{x} = [SOC(k), g(k)]^T$.

Different constant values for the weight factor β is first tested and compared. A value of $\beta = 0$ results in 326 gear shift events during the FTP75 driving cycle, and 213 of the gear shift events are within 2s from the previous gear change. The factor β is then increased to $\beta = 0.07$ which decreases the number of shift events to 163 and 38 within 2s, but increases the fuel consumption by 2.1%. By increasing the factor even more, the reduction in gear shift events sacrifice the fuel consumption which is increasing rapidly.

Instead, this VWF is defined as

$$\beta(k) = \begin{cases} \beta_1 & \text{otherwise} \\ \beta_2 & \text{for } k_s \leq k \leq k_s + M \end{cases} \tag{3.3}$$

Where $\beta_1 < \beta_2$. If there is a shift command at time step k_s, the weight factor β is equal to β_2 and remain until minimum gear dwells time M is reached. β_2 is larger than β_1 to prevent gear shifting within the minimum dwell time, which in this study is set to 2s. After time M, β is once again equal to β_1 which focus more on the fuel optimisation.

By using this method, the gear events are decreased to 155 for the FTP75 driving cycle, and none of the gear shifting events is within the time M from each other. The fuel consumption is increased by 3.3% compared to $\beta = 0$, but with more exceptional drivability.

3.4 Benchmarking, Modelling and Validation of a Conventional Midsize Car

In a paper by Newman K., Kargul J., and Barba D. [15] presents the benchmarking and validation process of a 2013 Chevy Malibu 1LS which is a part of a validation project of the Advanced Light-Duty Powertrain and Hybrid Analysis (ALPHA) simulation tool. The simulation tool was explained in the previous section and have since that paper was published been updated to be able to generate more accurate results. This paper contains the first validation of a vehicle in the validation project, which have the ambition to include several vehicles.

The vehicle where first benchmarked and data was collected. The engine and the transmission were then separately tested, and more data were collected. A three-dimensional fuel consumption map with engine speed, torque and fuel consumption on the axles was created. The first step was to loosely fit a surface through the points to be able to estimate the fuel consumption at the corners of the map. The map was extrapolated so that when simulations are performed, the fuel consumption never exceeds the map. The second step was to perform a triangular interpolation to extend the surface and find the surface fitting the test data and the estimated corner data. Other vehicle parameters were also implemented in the model by comparing test- and simulation data. A parametric shift algorithm that dynamically calculates transmission shift points as a function of engine fuel consumption and user-defined operating limits was also implemented.

A fuel meter was installed in the engine compartment to correlate the data from the fuel consumption map and the test data. The result of the correlation became (96.56%) at steady-state operations with an R-squared fit of 0.9999. The fuel meter showed that fuel consumption rapidly increased during high acceleration rates. Therefore, an acceleration penalty factor was calculated and implemented to increase the correlation for the none steady-state test as well. The final results show a good correlation of the average fuel consumption, the fuel consumption rate, the engine speed, the vehicle speed and the transmission output torque.

3.5 Energy Consumption Prediction for Electric Vehicles Based on Real-World Data

Vehicles energy consumption is one of the main constraints that is considered during the development of a vehicle. In the paper by Cedric De Cauwer et al.[4], an energy consumption prediction is modelled and evaluated. The purpose of the paper is to identify the correlation between energy consumption and kinematic parameters in electric vehicle (EV). In this paper, three different estimations models are presented in which aggregation of the input parameters varies.

The first model has aggregated values of the kinematic parameters as inputs which allows a prediction with easily accessible inputs such as travel distance, travel time and temperature. The second model is similar to the first model but with a detailed acceleration data included. The final model uses raw data as input. All models proposed in the paper are statistical and have a direct link with vehicle parameters such as vehicle speed and travelled distance. The first estimated model is presented in table 3.1.

Table 3.1: *Description of the variables used in the estimation model.*

Variable	Decription
β_i	Regression coefficient
Δ E	Energy
v	Average vehicle speed
Δ s	Traveled distance
|20 - T | aux	Scaled temperature value with T
Δ t	time
ΔH_{pos}	Positive elevation
ΔH_{neg}	Negative elevation
ϵ	Error term

Equation (3.4) describes the first model used in this study.

$$\Delta E = \beta_1 \cdot \Delta s + \beta_2 \cdot v^2 + \Delta s \cdot \beta_3 |20 - T| aux \cdot \Delta t + \beta_4 \cdot \Delta H_{pos} + \beta_5 \cdot \Delta H_{neg} + \epsilon \quad (3.4)$$

The generated data that was used in this study was from a logger device installed on a Nissan Leaf for two years. The data was registered at 1 Hz frequency. 10 different drivers drove the vehicle without any restriction and total driven a distance of 24700 km. The results from the multiple linear regression for the first model can be seen in figure 3.2 where R^2 is the coefficient of determination, B_i is regression coefficient, $SE\ B_i$ is standard deviation of the regression coefficient, β_i is standardised beta estimate, P is confidence level and $Avg.$ % contribution is relative average contribution.

	Adjusted R^2	B_i	SE B_i	β_i	P	Avg. % Contribution
Model 1 macro	0.956				<0.001	
Constant		1.19E-1	2.65E-2		<0.001	4%
Rolling resistance		1.32E-1	1.84E-3	0.83	<0.001	78%
Aerodynamic		5.00E-6	4.45E-7	0.11	<0.001	6%
Auxiliaries		1.83E-1	7.35E-3	0.14	<0.001	7%
Pos. elevation		3.08E-3	3.72E-4	0.23	<0.001	26%
Neg. elevation		−2.54E-3	3.68E-4	−0.19	<0.001	−21%

Figure 3.2: *Multiple Linear Regression for the first model.*

This study shows that multiple linear regression with a high coefficient of determination value can be a suitable method for predicting energy consumption. The second model where the acceleration is represented provides a result that shows the importance of acceleration as an independent variable. The inclusions of the acceleration provide a higher level of detail, and with further improvement, the accuracy can be increased. The three models that have been evaluated in this study shows that a model with a higher level of detail can create a more accurate model using multiple linear regression. In conclusion, this study was performed on real-life data, and the models developed to estimated energy consumption was adjusted for the type of data. Overall multiple linear regression is an appropriate approach for this type of scenario.

3.6 Vehicle Fuel Consumption Estimation Using Machine Learning

Argonne National Laboratory has developed a simulation tool, Autonomie, which include over 1 million vehicles with different powertrain types and component technologies. The simulation tool makes good predictions of the fuel consumption of the vehicle but is very time-consuming. The paper by Junlin Yao and Ayman Moawad [19] presents a novel large-scale learning and prediction process (LSLPP) using machine learning to accelerate the prediction of the fuel consumption as well as increasing the possibilities by also be able to predict the fuel consumption from unseen data.

The machine learning and prediction routine follow the steps: preprocessing, training, validation, prediction, and outlier detection if necessary.

The preprocessing step prepares the dataset so that it can be used in the machine learning algorithm. First of all, the machine learning algorithm requires a dataset only consisting of numerical values categories need to be changed to numerical values. The dataset also needs to be grouped, all parameters do not match each other and the data, therefore, need to be separated into subgroups with vehicle parameters matching each other.

The step also includes the feature selection, which refers to selecting relevant variables for the machine learning model. Too many variables do not increase the accuracy of the prediction, it will instead increase the training time as well as increasing the over-fitting of the data. Irrelevant and redundant variables should be removed as well as highly correlated variables which are found by using the Pearson product-moment correlation coefficient. In this study, the number of variables is decreased from more than 30 variables to less than 10. For one powertrain, the selected variables are Vehicle curb weight, Rolling resistance, Drag coefficient, Glider mass, Engine max power, Engine IVA type, Vehicle powertrain and Vehicle class. As written above, the variable cannot be categorical as, for example, vehicle class which therefore need to be changed to a numerical value. It can be achieved by giving one of the vehicle classes number 0 and the other one number 1 if there exist two different vehicle classes. For some machine learning methods, neural networks, it can be efficient to normalise the inputs. It is not used for this thesis work and is therefore not further explained. The final step in preprocessing the data is to split the dataset into training data and test data. Typically, the training data consist of 70% of the dataset, and the test data consist of the remaining 30%. The training data is used to train the machine learning models, and the test data is used to validate the model performance.

The next step in the machine learning routine is the training step where a supervised machine learning algorithm fits a model to the training set. In this study, they utilise multilayer feedforward neural network models, a state-of-the-art machine learning approach for nonlinear regression. There exist lots of different machine learning algorithms which results in different prediction performances.

When the model has been trained, it enters the validation and evaluation step in the machine learning routine where cross-validation techniques are used to find a trade-off between generality and specifics. Over-fitting may occur when training the model leading to unreliable predictions and cross-validation can be used to evaluate the results. The inputs from the test set are implemented in the model resulting in prediction results which then are compared to the ground truth, the outputs from the test set. There exists a couple of measurements which are good indicators of the models' performance. The mean square error (MSE) is computed to measure the prediction residuals as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2 \tag{3.5}$$

where n is the number of data points, $\hat{Y}$ are the predicted values in the machine learning model, and Y is the ground truth obtained by Autonomie.

A model which result in a low MSE value is reliable to use for predictions of unseen data. The correlation factor should be close to 1, which means that there is a high correlation. The validation can be further improved by using the k-fold cross-validation which divides the training- and the test set into k equal-sized subsets. One subset at a time is then used as the test set, and the remaining are used as the training set for k rounds of validation, where the test set is changed each time resulting in that each subset is used once as the test set. Averaging the results eliminates the randomness caused by deciding which set should be used as a test set and training set. In this study, 10-fold cross-validation is used.

Validation results which are compared in this study is the average MSE ($\overline{MSE}$), the standard deviation of MSE (σ_{MSE}), the average square root of MSE ($\overline{RMSE}$) and the coefficient of variation of the $\overline{RMSE}$ ($CV(\overline{RMSE})$). ($CV(\overline{RMSE})$) is measured by dividing the mean of the output $\overline{Y}$ by $\overline{RMSE}$ and is expressed in percentage.

Next step in the machine learning routine is to detect outliers in the training sets. An outlier in the training set is defined as a data point that appears to be inconsistent with the remaining ones in the same dataset. By training the model which includes the outliers results in a worse prediction model compared to a data set excluding the outliers. To detect the outliers, an algorithm named random sample consensus algorithm (RANSAC) is implemented in the machine learning process. RANSAC can exclude the outliers and find the best model by using an iterative process which fits sufficiently many points in the dataset.

The final step in the machine learning routine is to find a confidence interval for predictions of the input variables. To do this, the pairwise bootstrap algorithm is used which estimates the confidence interval as follow:

1. We randomly draw a sample $Z_b^*(n)$ of size n with replacement from the training set $Z(n)$ for $b = 1, ..., B$.

2. A feedforward neural network is trained on bootstrap samples $Z_1(n)^*, ..., Z_B(n)^*$ and we obtain an ensemble of B trained neural networks.

3. v is fed to B neural networks which yield B predicted output variables, $o_1, ..., o_B$.

4. We take the average output $\overline{o} = \sum_{i=1}^{B} o_i$ as the ultimate prediction given the input v.

5. Let $\hat{F}_B$ denote the empirical cumulative distribution function (cdf). Then, the approximate $(1 - 2\alpha)$ central confidence interval of $\overline{o}$ i given by $[\hat{F}_B^{-1}(\alpha), \hat{F}_B^{-1}(1 - \alpha)]$.

where $Z(n) := (z_1, ..., z_n)$ and z_i are a pair of input and output variables (x_i, y_i). x_i is a vector of input variables and y_i is a scalar output variable. v is a vector with input variables and unknown output values for prediction.

3.7 The Relationship of Economic Variables and Final Energy Consumption Using Multiple Linear Regression

Predictions and forecasts are useful tools to get an estimation of a particular future event. This paper by Abdullah L. and Leong W. H. [2] provides a multiple linear regression method of estimating the final energy consumption in Malaysia based on three independent variables, growth domestic product (GDP), population and tourism. The method can also be applied to other predictions, for example, the fuel consumption of a vehicle. The independent variables have been chosen from earlier made studies where these three variables have a great influence on energy consumption. The main goal of the multiple linear regression is to find the best fit between the independent and dependent variables, called the regression equation. The general equation can be written as:

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 x_{i,1} + \hat{b}_2 x_{i,2} + ... + \hat{b}_K x_{i,K} + \epsilon_i \tag{3.6}$$

where b_n is called regression coefficient, $\wedge$ indicates predicted values, ϵ_i is an error term and i = 1,2,...,n.

To develop the regression equation, an analysis is performed according to the following steps:

1. Define the input- and output variable(s), referred to as the independent- and dependent variable(s). For this paper, the dependent variable is the final energy consumption (y) and the independent variables are GDP (x_1), Population (x_2) and Tourism (x_3).
2. Collect and key in the data of independent- and dependent variable(s).
3. Compute the linear relation between the independent- and dependent variable(s). Calculate the multiple correlation coefficient R^2, which is a statistical measure of how close the data is to the fitted regression line. R^2 is in this paper equal to 0.932, which indicates that the independent variables explain 93.2% of the variation in the final fuel consumption. The closer R^2 is to 1 implies that the regression model can make useful estimations.
4. Use the F-test to test the null hypothesis. This step checks if there exist a regression relationship between the independent variables or the dependent variable.
5. Determine the regression coefficients b_n and implement them in the developed multiple linear regression equation, which is the line of the best-fitted data of the independent and dependent variable(s). The final equation for estimating the final energy consumption in this paper then becomes:

$$\hat{Y}_i = -26090248.469 + 0.000003727x_1 + 2.5433x_2 - 0.247x_3 \tag{3.7}$$

3.8 Test Correlation Framework for Hybrid Electric Vehicle System Model

A test correlation framework for HEV was presented in a paper by Meng Y., Jennings M., Tsou P., Brigham D., Bell D. and Soto C. [13]. The main purpose of this paper is to present a Parameter Diagram (P-diagram), which is a model-test correlation framework that focuses on differences between model simulation and vehicle test. In this framework, a vehicle test requirement set, e.g. the battery state of charge (SOC) and state of health has to be properly represented by the simulation model. This framework proposes an organised list of a vehicle subsystem that is used for comparison. Table 3.2 presents quantities from vehicle subsystems that are important for verifying simulation-to-test correlation. The signals that are compared to each other for determining the correlation between simulation and test are shown in table 3.3. These signals are time-varying.

Table 3.2: *Cycle averaged and integrated quantities used for vehicle simulation-to-test correlation.*

Subsystem	Quantity	Units
Powerplant	Engine Work	kJ
	Engine on time	s
	Engine operating efficiency	%
	Number of engine starts/stops	-
	Fuel consumption	g
Electrical	HV battery energy	kJ
	HV battery delta SOC over cycle	%
Transmission	Generator mechanical work	kJ
	Motor mechanical work	kJ
	Motor electrical work during engine off periods	kJ
	Motor operating efficiency during engine off periods	%
Chassis	Wheel work	kJ

Table 3.3: *Time varying signals used for vehicle simulation-to-test correlation.*

Subsystem	**Quantity**	**Units**
Powerplant	Engine speed	Rpm
	Engine torque	Nm
Electrical	HV battery current	A
	HV battery voltage	V
	HV battery SOC	%
	Accessory load current	A
	Accessory load voltage	V
Transmission	Generator speed	Rpm
	Generator torque	Nm
	Motor speed	Rpm
	Motor torque	Nm
Chassis	Vehicle speed	kph
	Wheel speed	Rpm
	Wheel torque	Nm

Statistical analysis is used to determine whether the simulation results are within the test data distribution, assume that the test data vary based on several noise factors. The simulation results for cycle averaged, and integrated quantities are compared with the mean of the corresponding quantities from the set of selected test data. The critical metric for correlation is based on the percentage deviation of the simulation result from the test mean. The test standard for each quantity can be calculated and used to compute the number of standard deviations between corresponding simulation data and mean test data. With help from this metric, an explicit indication if the simulation data can be compared with test data can be provided.

Since the test-to-test data is varying depending on noise factors, the framework suggests the following actions, which is further elaborated in [13]. This approach for statistical analysis is shown in figure 3.3.

To calculate the single metric indicating the level correlation between two signal traces x and y, normalised cross-correlation power (NCCP) is applied, which can be seen in equation (3.8). If the correlation value is higher than 0.9, then the correlation between these two signals is high. If the value is less than 0.9, the correlation is poor.

$$NCCP = \frac{max[R_{xy}(\tau)]}{max[R_{xx}(\tau),\ R_{yy}(\tau)]} \tag{3.8}$$

where

$$R_{xy}(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_0^T x(t)\ y(t-\tau) \quad dt \tag{3.9}$$

This framework provides a structured method of using correlation processes

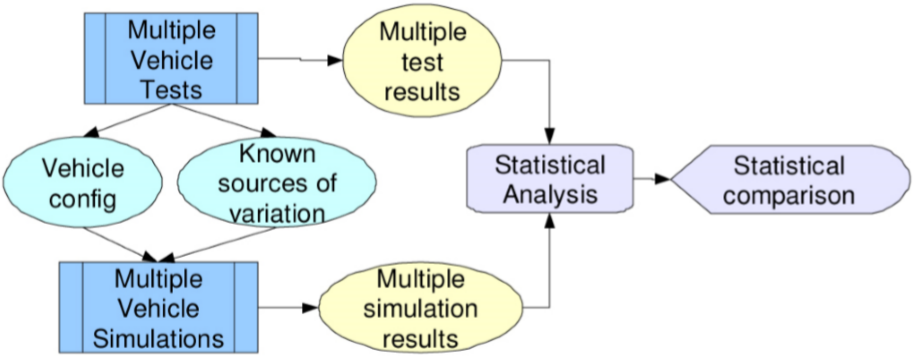

Figure 3.3: *Process flow diagram for statistical analysis approach to test to simulation correlation [13].*

and ideal signal output to an accuracy level on a simulation model. With help thoroughly chosen set of vehicle tests and statistical comparison, a correlation that meets specific criteria is performed.

4

DCT Model Development

This chapter contains the development of the DCT simulation model, both as gearbox process of achieving the correct gear shifting but also the TCU, including the gear selection strategy. The DCT model consists of both the TCU and the gearbox itself. The main goal of the DCT model is to provide simulations of the fuel consumption similar to the test data received from tests performed on the same vehicle configuration. The approach of achieving a gear shifting strategy as similar as possible to the test vehicle data follow the steps stated below:

1. Develop the gearbox model so that the gears are engaged correctly as the clutches position is open/closed.
2. Implement an optimisation algorithm for the selection of the gear.
3. Implement propelling states/modes of the vehicle in the TCU, such as EV-mode, HEV-mode and ICE-mode.
4. Implement constraints based on shifting behaviour found by studying the drive cycle of the test vehicle.
5. Design the gear selection strategy.
6. Examination of differences between simulation- and test result.

4.1 HEV Configuration

An overview of the HEV configuration is shown in figure 4.1.

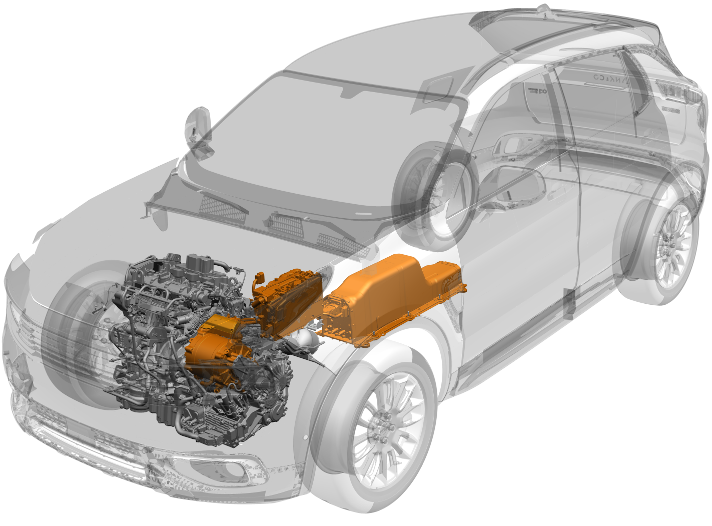

Figure 4.1: *HEV configuration.*

The configuration of the HEV is based on the Lynk & Co 01. A simplification of the HEV configuration is presented in figure 4.2 below.

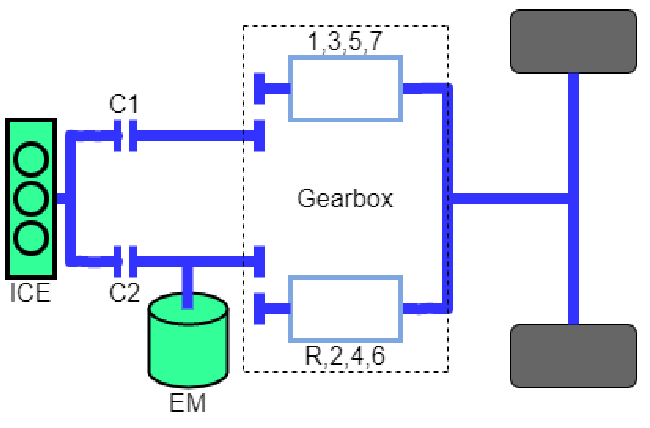

Figure 4.2: *HEV configuration including the transmission, the EM and the ICE.*

The ICE is connected to both gearbox shafts and can, therefore, provide power through either the odd shaft or the even shaft. The EM is connected to the even shaft between the clutch and the gears, a so-called P2.5 hybrid architecture, and can therefore only add additional power through the even shaft. The configuration of placing the EM between the clutch and the gears enables the opportunity to disengage the ICE and run the vehicle fully electric. A step gear is also added to increase the rotational speed of the EM, implying further use at efficient areas in the EM efficiency map. The map in figure 4.3 below shows at which torque and engine speed the most efficient use of the EM is found for the EM implemented in this thesis.

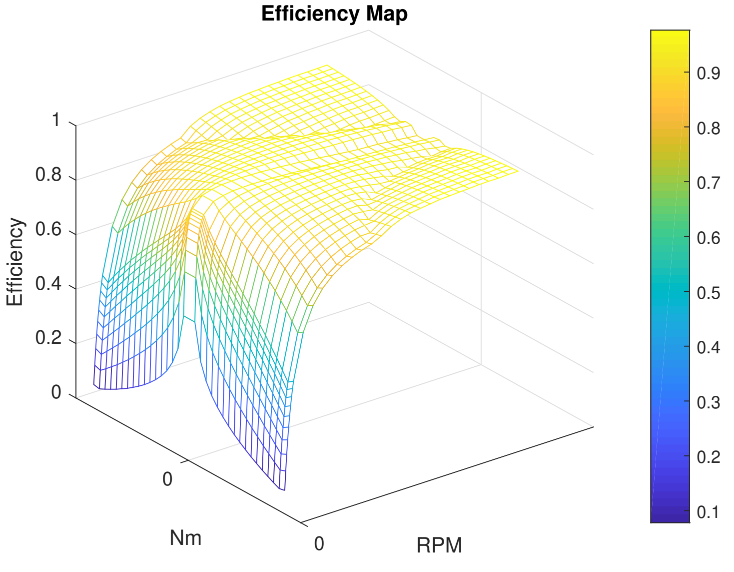

Figure 4.3: *EM efficiency map, increased efficiency as turning to yellow.*

Multiple options of propelling the vehicle exists, figure 4.4 shows the propelling variations of the vehicle. A in the figure corresponds to when both the ICE and the EM are propelling the vehicle simultaneously on the even shaft. It also corresponds to when the ICE is propelling the vehicle along with charging the battery through the even shaft. B corresponds to the same propelling as A, except that the ICE is propelling through the odd shaft and charging the battery through the even shaft. C corresponds to fully electric drive, either by propelling the vehicle fully electric or charging the battery through regenerative braking. D corresponds to the ICE propelling the vehicle through the odd shaft. A, B, C and D are all having their pros and cons depending on the load, vehicle speed, throttle position and SoC.

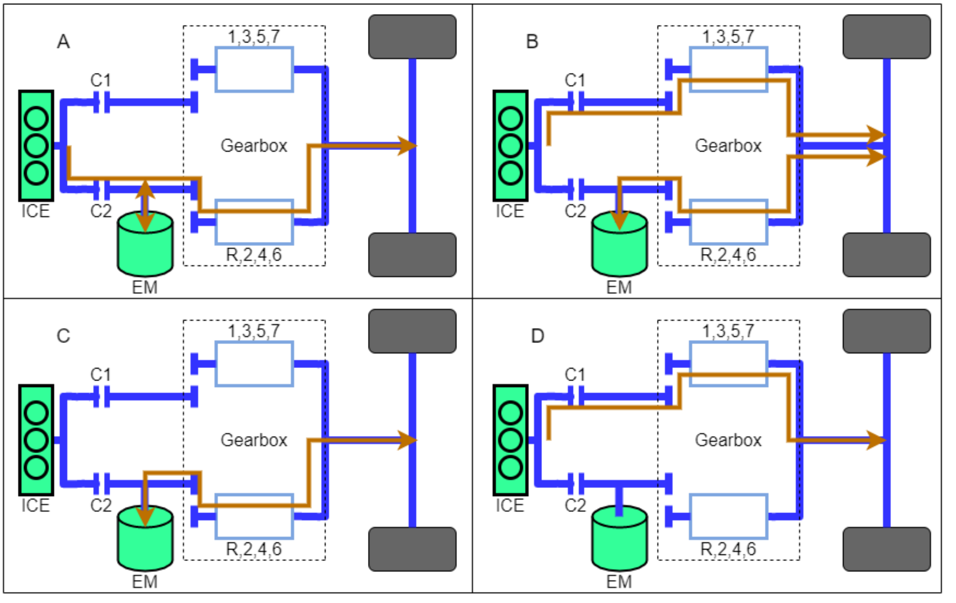

Figure 4.4

Assumptions made to simplify the modelling of the DCT and the gear selection strategy:

- Efficiency of the ICE is represented in the BSFC map where the fuel consumption depending on torque and rotational speed is given. Additional 5% of losses in the ICE is also implemented, representing losses such as friction and heat losses.
- Efficiency of the EM is represented in the motor efficiency map, depending on the torque and rotational speed.
- The losses in the battery are not considered; neither is the temperature of the battery, which has a great impact on battery performance.
- The charge/ discharge rate of the battery is given depending on the SoC.
- The gearbox efficiency is given for each gear, and all other losses are not considered, such as the higher losses when charging the battery through the odd gear shaft.
- The fact that the EM is rotating when the ICE is propelling the vehicle through even gears is not considered.
- The ICE is assumed to be off when regenerative braking.
- Charge sustaining throughout the driving cycle is not considered.

4.2 Gearbox Development

As mentioned in the first paragraph in section 2.3, the DCT can be described as two separate manual gearboxes with a respective clutch. The gearbox is therefore modelled as two manual gearboxes, the first for even gears and the second for odd gears. The even gear shaft is the complicated one due to the possibility of either propelling the vehicle by the ICE, the EM or both simultaneously.

The obvious constraints of the gearbox are:

- The ICE may either select a gear on the even shaft or the odd shaft.
- The EM can only select a gear on the even shaft.
- The even shaft can only engage one gear at a time. If the ICE and the EM are demanding different gears at the same time, one has to be prioritised.

The gear selection strategy explained later in this chapter decides the gear selection of the even and the odd shaft. The task of the gearbox system is never to exceed any conditions of actions valid for the gearbox. An overview of the DCT model is shown in figure 4.5. The model consists of two systems, the TCU which is explained later, and the gearbox, filled as green explained in this section.

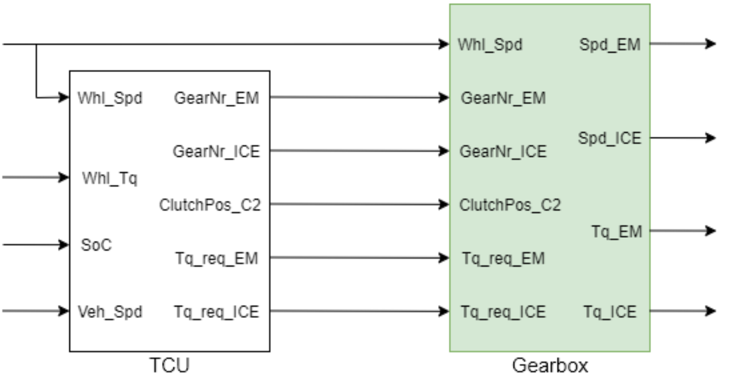

Figure 4.5: *DCT model, gearbox filled as green. Arrows at the left side of the block corresponds to input signals and the right side corresponds to output signals.*

The gearbox system itself contains two subsystems, the even shaft and the odd shaft. Both subsystems have their input- and output parameters presented here:

Inputs of the even shaft:

- Gear number request from the TCU for the ICE and EM, if the gear numbers differ, the requested number from the EM is prioritised.
- Clutch position, if the TCU requests a gear from ICE or EM on the even shaft, the clutch is engaged, otherwise disengaged.

- Torque request for the ICE used to calculate the required torque from the ICE.
- Torque request for the EM used to calculate the required torque from the EM.
- Wheel speed used to calculate the gearbox speed for the ICE and the EM.

Outputs of the even shaft:

- Rotational speed EM.
- Rotational speed ICE.
- Torque EM.
- Torque ICE.

The rotational speed of the EM and the ICE are calculated using equation (2.15) where ω_1 is the rotational speed of the EM, ω_2 is the rotational speed of the wheel and γ is the gear ratio of the selected gear. The torque of the ICE is calculated by rewriting equation (2.17) so it holds for torque instead of power. Power $P_{0,gb}$ and power $P_{0,gb}$ is assumed to equal zero and both sides are divided by the rotational speed ω, the new equation is therefore:

$$\begin{cases} T_2 = e_{gb} \cdot T_1 & \text{if } T_1 > 0 \\ T_1 = e_{gb} \cdot T_2 & \text{if } T_1 \leq 0 \end{cases} \tag{4.1}$$

Where T_1 is the torque of the power source (ICE or EM), T_2 is the wheel torque, and e_{gb} is the gearbox efficiency depending on the gear number.

The same equations are used to calculate the rotational speed of the ICE and the required torque of the ICE on the odd shaft, which also are the outputs of the odd shaft subsystem. The inputs are similar to the even shaft, except the ones regarding the EM.

- Gear number request from the TCU.
- Torque request for the ICE.
- Wheel speed.

The clutch position for the odd shaft is not needed, and instead solved as: If the TCU request an odd gear, the clutch is engaged, otherwise disengaged. Outputs of the gearbox block are shown in figure 4.5, where Tq_ICE is the sum of torque required from the even shaft and the odd shaft. If the ICE does not request gear from a shaft, the torque equal zero for that shaft.

The sample time of the simulation model decides how many times all values are recalculated for the model. For each time step, all values are recalculated, and new gears are selected.

4.3 Optimisation Algorithm of Fuel Consumption

The purpose of the algorithm is to minimise the fuel consumption of the vehicle by selecting the optimal gear of each time step t. As section 2.3.2 describes, the transition from a conventional vehicle to an HEV increases the complexity for the gear shifting, the ICE load is not strictly linked to the required wheel power anymore. It can instead be achieved by several combinations of loads from the ICE and the EM. The algorithm optimises the ICE load, and additional torque is added or subtracted from the EM to deliver the required wheel torque.

Problem

For the DCT, there exist 7 gears which all have to be evaluated to find which gear to select. Each gear has a fixed rotational speed determined by the wheel speed together with the gear ratio of each gear. The load of each gear is, however, not fixed. It can be chosen as long as the difference between required wheel torque and ICE torque does not exceed the maximum torque of the EM, or the ICE torque exceeds the maximum engine torque. The possible load for the ICE is, therefore, in a wide range due to the large EM torque output. The problem is to find the gear and its corresponding torque minimising the specific fuel consumption which does not violate any constraints.

Solution

The BSFC map is structured as a matrix with breakpoints as rotational speed and engine torque shown in figure 4.6 below. The values in the BFSC map is not from the intended ICE, but just an example for better understanding.

BSFC [g/kWh]	Engine Torque [Nm]	15.6	31.2	46.8	62.4	78	93.6	109.2	124.8	140.4	156	171.6
Engine speed [rpm]	500	612	442	370	331	306	313	302	324	288	288	288
	1000	612	403	337	308	284	286	283	284	288	288	288
	1500	612	407	346	306	281	263	255	247	252	288	288
	2000	594	396	335	301	275	259	245	245	248	263	288
	2500	558	382	320	284	266	247	245	245	245	256	288
	3000	522	365	311	279	263	255	248	245	245	259	288
	3500	468	340	288	281	277	268	259	248	248	263	288
	4000	522	356	318	304	283	275	263	254	262	266	288
	4500	540	392	351	317	293	279	272	270	273	275	288
	5000	684	461	378	333	306	284	283	279	283	284	288
	5500	756	511	421	365	328	295	299	295	292	290	288
	6000	756	529	432	374	335	313	299	295	292	288	288

Figure 4.6: *Brake-specific fuel consumption map, which is a measure of the fuel efficiency at given rotational speed and torque.*

The algorithm minimises the specific fuel consumption by finding the torque corresponding to the lowest value of specific fuel consumption in the BSFC map for the given rotational speed of each gear one by one. The gear to be selected is the gear resulting in the lowest specific fuel consumption value that does not violate any constraints. A step-by-step explanation is presented below, explaining the algorithm. The input of the algorithm is the rotational speed of each gear.

1. An input, called RotSpd, consists of the rotational speed of each gear.

2. Check if the rotational speed of any gear violates the minimum- or maximum rotational speed constraint of the engine, if so, remove the gear from the available gears to select.

3. Find the minimum specific fuel consumption of each gear and the corresponding torque. This is done using the BSFC map, for the given rotational speed, find the column in the map resulting in the lowest value of specific fuel consumption together with its corresponding torque shown in figure 4.7. An interpolation is made if the values are not in the table. Figure 4.7 shows an example, if the rotational speed of the gear is 4000 rpm, the lowest value of the specific fuel consumption is then found at 254 g/kWh, marked with red borders. The corresponding torque can then be found in the same column; in this case, 124.8 Nm and also marked with red borders.

BSFC [g/kWh]	Engine Torque [Nm]	15.6	31.2	46.8	62.4	78	93.6	109.2	124.8	140.4	156	171.6
Engine speed [rpm]	500	612	442	370	331	306	313	302	324	288	288	288
	1000	612	403	337	308	284	286	283	284	288	288	288
	1500	612	407	346	306	281	263	255	247	252	288	288
	2000	594	396	335	301	275	259	245	245	248	263	288
	2500	558	382	320	284	266	247	245	245	245	256	288
	3000	522	365	311	279	263	255	248	245	245	259	288
	3500	468	340	288	281	277	268	259	248	248	263	288
	4000	522	356	318	304	283	275	263	254	262	266	288
	4500	540	392	351	317	293	279	272	270	273	275	288
	5000	684	461	378	333	306	284	283	279	283	284	288
	5500	756	511	421	365	328	295	299	295	292	290	288
	6000	756	529	432	374	335	313	299	295	292	288	288

Figure 4.7: *How the BSFC map is used to find the lowest specific fuel consumption and its corresponding torque.*

4. There are two more constraints that have to be evaluated before selecting gear. The first is to check so that the torque does not exceed the maximum torque of the engine. The second is to evaluate so that the difference between required wheel torque and the ICE torque does not exceed the maximum torque of the EM. If any gear violates any constraint, it is removed from the list of available gears to select.

5. The final step is to select gear. It is done by selecting the gear corresponding to the lowest value of specific fuel consumption.

The output of the algorithm is the gear number and the torque of the ICE.

4.4 Vehicle Modes and States

There are three driving modes for the vehicle, electric vehicle mode (EV-mode), hybrid electric vehicle mode (HEV-mode) and internal combustion engine mode (ICE-mode). The modes are prioritised in the order written above, i.e. the EV-mode has the highest priority, and ICE-mode has the lowest priority. The vehicle enters EV-mode when the electric machine can provide all power needed for propelling the vehicle by itself, which also include if the vehicle is standstill. Due to the assumption that the ICE is turned off when regenerative braking, regenerative braking is also included in the EV-mode. The vehicle is in HEV-mode when the ICE and the EM are propelling the vehicle simultaneously and in ICE-mode when the vehicle is propelled only by the ICE. To further categorise the driving modes, vehicle states are introduced. There are six possible states for the HEV configuration, explained below together with figure 4.8.

1. ICE off - The electric mode is prioritised at low engine speed and low torque due to the low efficiency for the ICE at those conditions.

2. ICE on - As the requested torque or vehicle speed increases, the ICE is turned on. The state ICE on is equivalent to the ICE-mode, i.e. that the EM is not in use.

3. Load point shift charge - The vehicle should enter the load point shift charge state every time it is possible to gain efficiency for the ICE and the SoC level allows it. If the SoC level is low, the vehicle enters this state to charge the battery back to a higher level.

4. Performance boost - If the requested torque exceeds the maximum torque for the ICE, the EM can be used to increase the maximum torque for the vehicle. This state can only be entered if the SoC level is far from its lower boundary.

5. Load point shift boost - The load point shift boost state is entered if the requested torque is within the maximum torque for the ICE, but a more efficient operational area for the ICE can be entered by adding torque from the EM.

6. Regenerative braking - The regenerative braking should always be prioritised when braking as long as the battery is not fully charged. If the battery is full, engine braking and the brakes are used to slow down the vehicle.

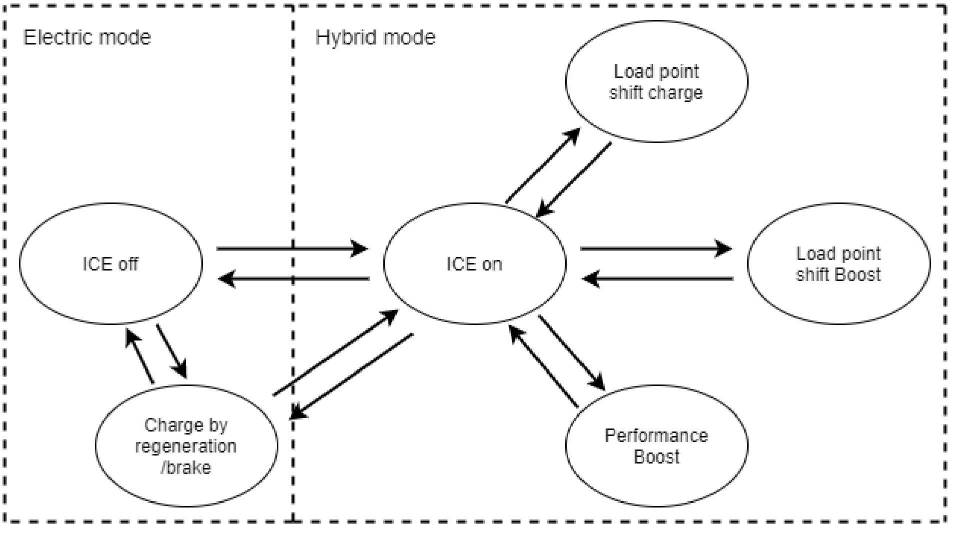

Figure 4.8: *The six possible states for the vehicle configuration represented as circles. Arrows corresponds to the available state transitions from each state. The states are divided into vehicle modes, ICE on also corresponds to ICE mode.*

The driving mode can be categorised using the fuel flow and the current into the EM as shown in table 4.1.

Table 4.1: *Configuration of the fuel flow and current into EM determining driving mode.*

Mode	EM current [A]	Fuel Flow [g/s]
EV	$\neq \; 0$	$= \; 0$
HEV	$\neq \; 0$	$\neq \; 0$
ICE	$= \; 0$	$\neq \; 0$

4.5 Simulation- to Test Data Comparison

Since the vehicle configuration already exists, vehicle tests have been performed for the WLTC drive cycle, see the figure 4.9, enabling the opportunity of extracting driving constraints and threshold values by observations in the test data. Many different comparisons where made, but only the method of the cases where a clear behaviour or threshold was found is presented here. Unfortunately due to confidentiality, some figures will not be presented.

Firstly, the behaviour of the EM gear shifting was examined. A plot of the vehicle speed throughout the WLTC driving cycle and a plot of engaged gears for the even shaft as the EM add power throughout the driving cycle was compared to

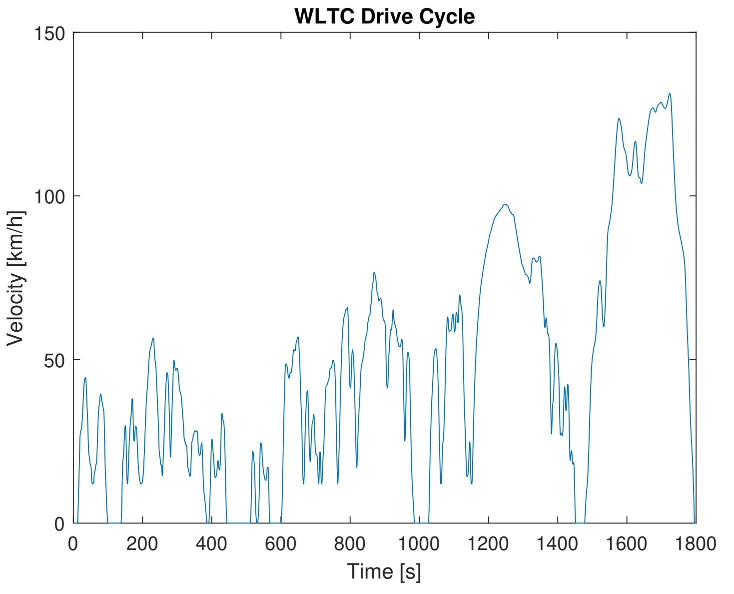

Figure 4.9: *WLTC Drive Cycle used in this thesis.*

find patterns in the gear shifting. By studying this particular plot, thresholds of at which vehicle speed the strategy tend to shift gear can be found. As it is written in section 2.3.2, the upshifting and downshifting events usually appears at different vehicle speeds. With that in mind, upshifting- and downshifting events are studied separately. Another interesting plot to study is the gear number versus the vehicle speed, visualising at which vehicle speed the EM gears are engaged. The last studied plot used for finding the gear shifting thresholds for the even shaft when the EM is propelling the vehicle by itself or simultaneously with the ICE is a plot showing the upshifting and downshifting events versus the vehicle speed. This plot shows at which particular vehicle speed the actual gear shifting event happens.

The transition from EV-mode to ICE-mode appear when the ICE turns on and when the EM turns off. The gear selection strategy is used to optimise the use of battery power, depending on several variables. It would be possible to use the EM until the battery is running out of energy or the load is too high, but this is not a beneficial use of the EM. Instead, the EV-mode is used at low-load, power and velocity. To study the transition behaviour between the modes, vehicle speed-, wheel torque-, wheel power- and SoC vs time are plotted together for each time step indicating if the vehicle is in EV-mode or not. The plots are well studied looking for patterns between the variables and the event switching from EV-mode to HEV-mode and back again.

To see if any particular pattern for the gear shifting can be observed, the vehicle speed vs gear number is plotted and studied. The gear number of the even- and the odd shaft are plotted separately and presented in section 6.1.

4.6 Gear Selection Strategy

The gear selection strategy is implemented in the TCU, filled as green in figure 4.10. The simplified TCU for this work is a composition of several control systems to reduce the number of subsystems.

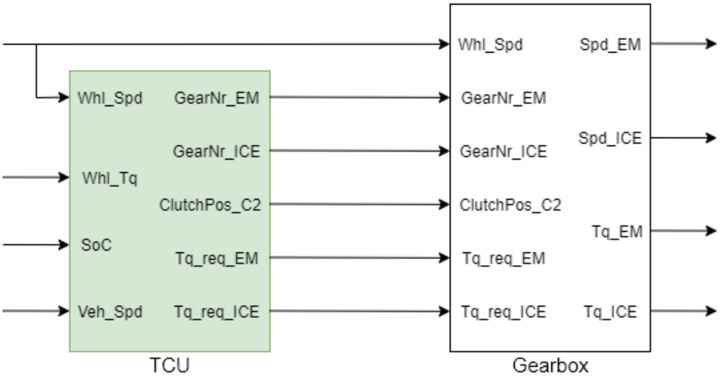

Figure 4.10: *DCT model, gear selection strategy is implemented in the TCU block, filled as green.*

The input parameters of the TCU are:

- Wheel speed which is the rotational speed of the wheels.
- Wheel torque which is the torque at the wheels.
- State of charge which is the level of charge of the battery capacity, which also is the only signal fed back in the model. The SoC signal is coming from the battery.
- Vehicle speed which is the speed of the vehicle.

The output parameters are:

- Gear number EM which is the requested gear for the EM.
- Gear number ICE which is the requested gear for the ICE.
- Clutch position C2 which is the clutch position of the clutch connected to the even shaft.
- Torque requested EM which is the requested torque for the EM.
- Torque requested ICE which is the requested torque for the ICE.

EV-mode has the highest priority, and if any of the EV-constraints are fulfilled, the EM is propelling the vehicle by itself. Two EV-constraints are implemented in the model; the first is if the SoC level is above or equal to a certain percentage and the wheel power is below the EV-mode max power. The second constraint is that if wheel torque is negative (regenerative braking), the EM charge the battery and no torque is requested from the ICE.

If none of these two constraints is fulfilled, the vehicle enters HEV-mode where the gear number and the requested torque from the ICE are calculated using the optimisation algorithm described in section 4.3. The required torque for the EM can then be calculated as

$$T_{EM_{req}} = \gamma \cdot T_{wheel} - T_{ICE} \tag{4.2}$$

The required torque for the EM can be provided by any gear on the even shaft as long as the constraints of the EM are fulfilled. The constraints consist of the maximum torque the EM can provide at a given rotational speed. The rotational speed of the EM is calculated using equation (2.15) for each time step and is used to find the maximum torque of the EM which is found from a datasheet of the EM specifications. The gear selection of the EM is then decided by the available gears depending on the speed of the vehicle. A relation between vehicle speed and gear number of the EM was found in the previous section which is used to decide which gear to engage.

In the case of requests of different gears on the even shaft by the EM and the ICE, an implemented rule downshift the requested gear from the ICE by one, resulting in that the ICE instead operates through the odd shaft. The rule of decreasing the number of the ICE by one comes from the fact that the EM has to change its gear at least two gear numbers, resulting in a higher gear ratio change.

Another boundary found via examining the test data is that the 7th gear only is engaged above a certain vehicle speed which also is implemented in the gear selection strategy. If the 7th gear is suggested for a vehicle speed less than the boundary, 6th gear is instead engaged.

ICE-mode is only entered if T_{wheel} and T_{ICE} assume the same value, resulting in that $T_{EM_{req}}$ is zero in equation (4.2). $T_{EM_{req}}$ equal zero is unusual and the vehicle is almost always operating in HEV-mode if neither of the EV-mode constraints are fulfilled.

To reduce the number of gear shifting events through the drive cycle, a constraint disabling a gear shift within 10 seconds is implemented. This constraint is implemented to obtain a more similar behaviour to the test vehicle. The drivability is important in vehicles, leading to a reduced number of shift events, which this simple constraint replicates.

4.7 Examination of Correlation Between Simulation- and Test Results

To evaluated the performance of the simulation model, correlation is used as a measure. Pearson's correlation coefficient explained in section 2.5 and NCCP, explained in section 3.8 are used as measurements of how well the simulation model performs compared to the results received from the test of the vehicle configuration. The mean relative error, $\overline{RE}$, is also calculated and used as a measure of how well the results of the simulation model follows the test data.

The vehicle speed and travelled distance of the vehicle for each time step received from the test data are used as input for the simulation model. The purpose of the simulation model is to reflect the behaviour of the test data, the simulation model is, therefore, evaluated against the test data. The performance of reflecting the test data differ between the components of the powertrain. The purpose of this thesis is though only to develop a DCT model resulting in that the overall vehicle model estimates the fuel consumption as similar as possible to the test data. The simulation model is, therefore, verified on a model level where the fuel consumption of the test- and simulation results are evaluated based on correlation. The correlation is calculated for each time step throughout the WLTC drive cycle.

5

Ranking Model Development

This chapter contains the development of the ranking model. The purpose of the ranking model is to evaluate the simulation result and score the level of reliability for the user based on fuel consumption. The QS simulation model of the vehicle, including the DCT model, described in section 2.3 is defined as "ground truth" and is always assumed to be correct. A multiple linear regression model is developed and used to provide predictions of fuel consumption of the vehicle throughout the WLTC driving cycle. Results from the simulation model and the MLR model underlying the score provided by the ranking model based on NCCP explained in section 3.8. The approach of the development of the ranking model follow the steps stated below:

1. Establish the QS model as a template simulation model.
2. Determine dependent- and independent variables for the regression models.
3. Generate a database used as test- and training data for the regression model.
4. Perform LASSO regularisation for variables selection.
5. Develop an MLR model.
6. Use the MLR model to estimate fuel consumption for a dataset outside the factor interval of the generated database.
7. Use NCCP to correlate the results from the QS model and the MLR equation and define a scoring system for the ranking model.

5.1 Template Model

To be able to construct a regression model, a large amount of data is required. The regression model intends to provide the fuel consumption rate at each time step. Since the simulation results are time-dependent, collected data of the driving cycle with the same sample time is appropriate. An extensive database consisting of several runs of different combinations of components throughout the driving cycle could not be provided for this thesis for the development of the ranking model.

Instead, to gather a well-structured database with vital signals for the development of the regression model, the QS simulation model can be used as a template model. The simulation result from the template model is defined as ground truth, meaning that the simulated signals are reliable and always assumed to be correct. The template model is needed to have consistency in the ranking model and as a reference in the verification and validating process of the ranking model. To be able to measure the performance of a model, the correct value which the model relates to is required.

5.2 Independent Variables Selection for Regression Model

As section 2.5.2 describes, the independent variables are the variables which the predictions are based on. To be able to make accurate predictions of the fuel consumption (the dependent variable for the regression models), it is important to evaluate the available variables for the model carefully. First of all, the variables implemented in the regression model has to have an impact on fuel consumption. By studying the QS model, all variables having an impact on fuel consumption are stored as possible independent variables, shown below:

- Vehicle curb weight
- Accessory power
- Tire road dynamics
- ICE torque
- BSFC map
- Gear ratio
- Gearbox efficiency
- Battery capacity
- Battery voltage charge/discharge rate
- Battery resistance charge/discharge rate

- EM torque
- EM efficiency map

It is also important to only implement the variables valuable for the prediction. With too many variables implemented, the model tends to overestimate the result (overfitting). With too few variables (underfitting), the model cannot adequately capture the underlying structure of the data, resulting in poor prediction performance. By comparing the possible independent variables with the suggested variables in related research, section 3.8 together with engineering intuition of valuable variables, the number of variables was reduced to:

- Vehicle curb weight
- ICE torque
- EM torque
- Battery capacity

The next step is to investigate the chosen variables and ensure that no multicollinearity exists, also mentioned in section 2.5.2. The independent variables are plotted one by one individually against each other. Multicollinearity means that there is a strong correlation between two or more independent variables resulting in unstable regression coefficients for the regression model.

5.3 Generation of the Database

As the independent variables for the regression models are determined, a database can be generated. The purpose of the generation of the database is to produce a database containing fuel consumption and component data of the WLTC driving cycle for several vehicle component specifications.

The template model uses initial inputs that affect how the vehicular components act during the homologation cycle. The template model inputs are altered using factors of each variable proposed in table 5.1. The database is generated by incrementally increase the factor of each variable one by one. In between the increase of the factor, the template model performs a simulation run and store the fuel consumption rate and component data of each time step. The purpose of these variable factors is to have diversity in the database by replicating the simulation with different inputs. The generation of the database is performed using the following factors and incrementally increase of the variable inputs.

Table 5.1: *The list of variables altered to generate a database.*

Variable Input	Factors	Description
BSFC Map	0.9:0.05:1.2	Brake-specific fuel consumption look-up map. Uses torque and speed as breakpoints.
Vehicle Curb Weight	0.8:0.1:1.2	Vehicle total mass.
ICE Torque Breakpoints	0.8:0.1:1.2	The breakpoints that is used in the BSFC map.
EM Torque Breakpoints	0.8:0.1:1.2	An efficiency look-up map with torque and speed as breakpoints.

The factors are increased incrementally, either by 5 or 10 %. The generation of the database is performed by using for-loops in MATLAB. The generation is divided into four different loops with the factors proposed in table 5.1, one loop for each variable variation. The factor interval is based on engineering intuition of what seems to be reasonable for each variable.

To generate the database, four for-loops are involved. The loops are nested, meaning that for each factor in the outer loop, the inner loop performs the whole factor incremental interval of loops, in a total of four levels. In the inner loop, the simulation template is initiated with altered variables.

The results are stored in the database for each simulation run. Append the results enabling the database size to increase by the number of performed simulations. At the end of each for-loop, the variable value is reset to its initial value so that the next variable may be varied instead.

The purpose of running four loops is to receive the data from every possible combination of variable variations. The regression models require data for training and testing the models and the database generation provide enough foundation for constructing a regression model. The increment size of the factors is presented in table 5.1. The pseudocode for the generation of the database is presented below.

Pseudocode for Generation of the Database

```
Define factor interval for each variable
Set initial values for each variable
for g=1:m % where m the length of the interval for BSFC
  Multiply the factor with initial BSFC

  for h=1:n % where n is length of the interval for curb weight
    Multiply the factor with initial curb weight

    for j=1:o % where o is length of the interval for ICE
      Multiply the factor with initial ICE variable

      for k=1:p % where n is length of the interval for EM
        Multiply the factor with initial EM variable

        Run the template model
        Save and append the variables and the fuel
        consumption in the database

      end
    end
  end
end
Save the results
```

5.4 LASSO Regularisation

LASSO regression is, for this thesis, used as an extension of independent variable selection. Therefore, only the regularisation part of LASSO regression is implemented. LASSO is an effective method performing variable selection yielding sparse models, meaning it is an effective method of reducing the number of redundant independent variables. The reason why LASSO is used for variable selection is due to its powerful way of confirming that the selected independent variables contribute to the result of the regression model. As section 2.5 describes, LASSO performs L1 regularisation to achieve the independent variable reduction by adding a penalty term to the loss function. The LASSO regularisation is performed using a function named `lasso` provided by the Statistical Toolbox in MATLAB.

Cross-validated `lasso` is used to remove redundant independent variables. The dependent- and independent variables are implemented in vector- and matrix form together with the choice of 5-fold cross-validation. The resulting variables are the variables within one standard error of the minimum MSE. By performing the LASSO regularisation, battery capacity is removed from the list of independent variables.

5.5 Development of MLR Regression Model

With an appropriate amount of data and the independent variable selection, the MLR regression model can be developed. Section 2.5 explains the fundamentals of MLR and how it may be used for predictions. The variables that fulfilled the requirements for building a regression model are presented in table 5.2. The regression model are built using the build-in MATLAB function `regress` which provides the intercept β_0 together with the regression coefficients β_i.

Table 5.2: *The variables used for modelling the regression model.*

Variable	**Type**
ICE Torque	*Independent*
EM Torque	*Independent*
Curb Weight	*Independent*
Fuel Consumption	*Dependent*

Since the data from the database is generated in ascending order, the extraction of the regression coefficients is done in the following manner to avoid bias estimation. To determine the regression coefficient a 5-fold cross-validation is executed, see figure 2.13 for better understanding. The data from the generated database is split into five equal-sized data sets. `regress` is then performed five times separately using each data set one time as validation data, i.e. not used for training of the model and regression coefficient extraction. This results in five different values of the regression coefficients for each independent variable. The mean of the regression coefficients is then calculated and used as the final regression coefficients for the MLR model. The MLR equation for fuel consumption estimation is presented in (5.1).

$$\hat{FC} = \beta_0 + \beta_1 Tq_{ice} + \beta_2 Tq_{em} + \beta_3 m_{veh} \tag{5.1}$$

5.6 Estimation of Fuel Consumption

Fuel consumption is estimated using the MLR model from the previous section. Equation (5.1) describes how the estimation of fuel consumption is performed. The values of the independent variables are implemented in the equation for each time step. Therefore, the MLR equation provides the estimated fuel consumption of each time step which then is summed for the total fuel consumption for the WLTC cycle.

5.7 Ranking With NCCP

NCCP is used to evaluate the result provided by the QS model. The vehicle data the user intends to study is implemented in the QS model, and the drive cycle is performed. The fuel consumption of each time step is saved. The values of the variables from the vehicle data are implemented for each time step in the MLR model. The MLR model calculates the fuel consumption of each time step and saves the result. The correlation between the simulated data and the MLR data is then calculated using NCCP. The value received by NCCP is then multiplied by a factor of 10 and rounded to the nearest integer resulting in that the score of the ranking model is on the interval [0,10]. A score of 9 or above infers a reliable simulation performed by the QS model.

The estimation from the previous section can now be used as a signal for determination of the trustworthiness of the simulated fuel. The simulated and estimated are compared to each other using NCCP for each time step, which is described in equation (3.8). If the NCCP score is above 0.9, the simulated fuel consumption is considered reliable.

6

Results

This chapter contains the results for the development of the DCT model as well as the results for the development of the Ranking model.

6.1 DCT Model

The results of the behaviour throughout the drive cycle for different measurements are presented in this section. A scatter plot of the simulation results is plotted against the results from the test data received for the same vehicle configuration for each time step. The actual values throughout the drive cycle are also plotted against each other. The correlation between the simulated data and test data are also presented, both with NCCP and Pearson's correlation coefficient, ρ. The mean relative error, $\overline{RE}$ is also presented.

Figure 6.1 shows the change in the SoC throughout the drive cycle, both for the simulated SoC and the SoC from the test.

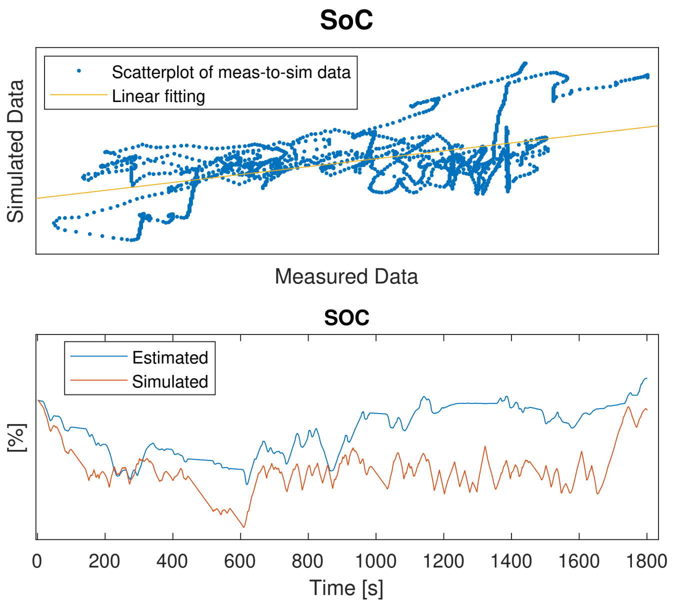

Figure 6.1: *Comparison of the SoC.*

Figure 6.2 shows the fuel consumption of the vehicle throughout the drive cycle for the simulated fuel consumption and the tested fuel consumption.

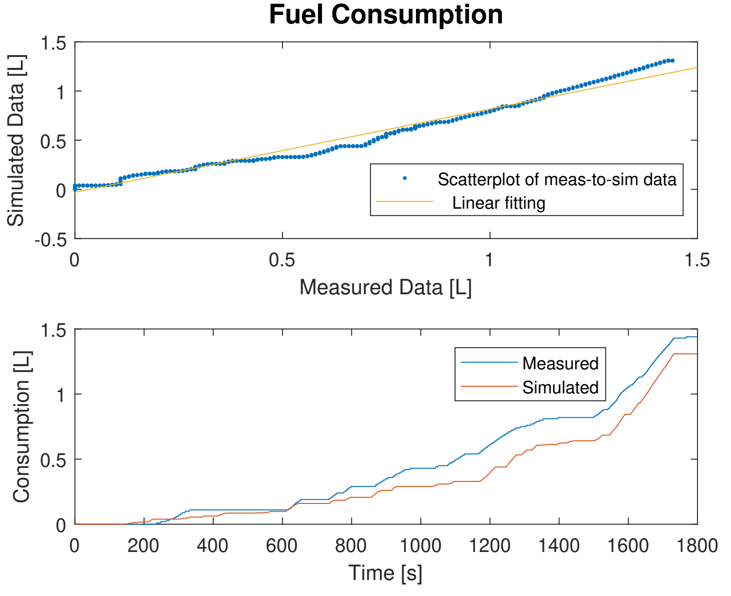

Figure 6.2: *Comparison of the Fuel Consumption.*

Table 6.1 shows the resulting correlation between the simulated- and tested value, both calculated from NCCP and Pearson's correlation coefficient together with the mean relative error, expressed in percentage.

Table 6.1: *NCCP, ρ and $\overline{RE}$ for the fuel consumption and the SoC.*

Subsystem	**Signal**	**NCCP**	ρ	$\overline{RE}$[%]
Powerplant	Fuel Consumption	0.8142	0.9889	25.71
Electrical	SoC	0.8579	0.5535	14.29

Figure 6.3 shows the selected gear of the EM throughout the drive cycle received from the simulation and the test.

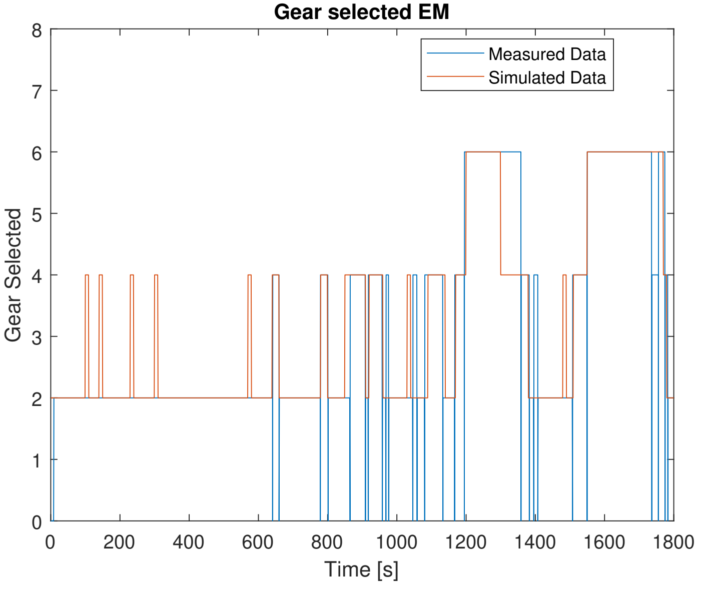

Figure 6.3: *The engaged gear for the EM during the WLTC cycle, measured and simulated.*

Figure 6.4 shows the selected gear of the ICE throughout the drive cycle received from the simulation and the test.

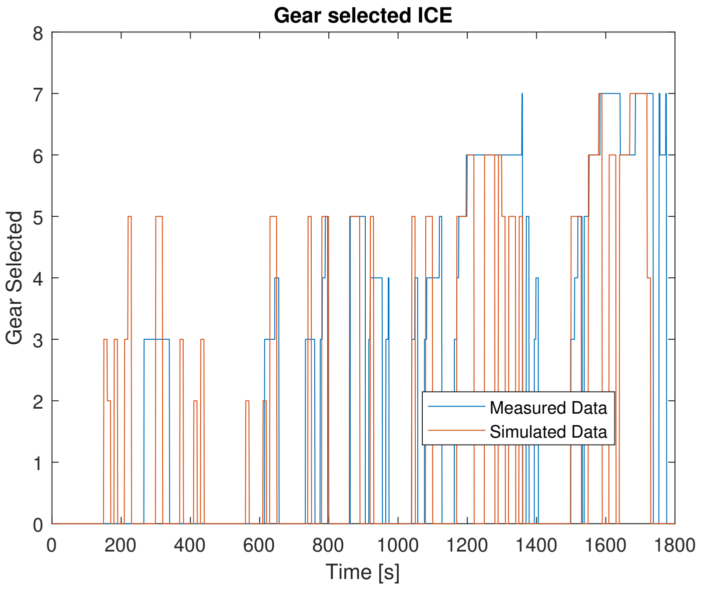

Figure 6.4: *The engaged gear for the ICE during the WLTC cycle, measured and simulated.*

6.2 Ranking Model

The ranking model was primarily developed to evaluate simulation results without access to test results. In this section, results from the developed ranking model together with results of the development process are presented.

LASSO Regularisation

Figure 6.2 shows the result of performing the `lasso` function in MATLAB. The result corresponds to the variables selected as the independent variables for the MLR model.

Table 6.2: *Resulting independent variables for the MLR model from the LASSO regularisation.*

Independent variables:
Vehicle mass
ICE torque
EM torque

To clarify the reason why the battery capacity was removed as an independent variable as the LASSO regularisation was performed, varying battery capacity is plotted against the resulting fuel consumption throughout the driving cycle.

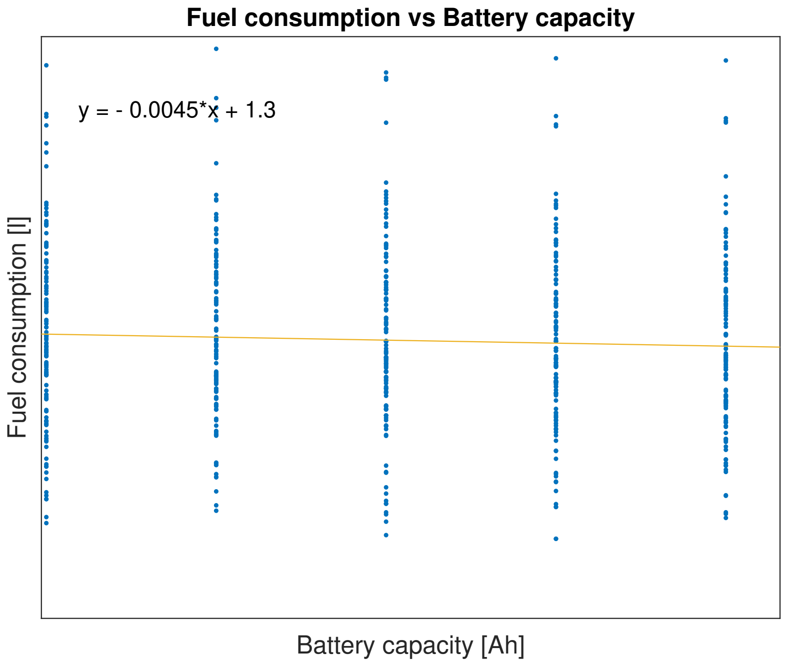

Figure 6.5: The corresponding fuel consumption for different battery capacities, including the equation of the best-fitting line.

MLR Model

To build the MLR model, the MATLAB function `regress` was used, resulting in the regression coefficients presented in table 6.3.

Table 6.3: The regression coefficient extracted using `regress` for the independent variables.

Variable	$[\beta_i]$ MLR
Intercept	-0.0655
ICE Torque	0.016
EM Torque	4.2092e-04
Curb Weight	2.539e-05

Ranking Model

To evaluate the MLR model, three different sets of variable values were tested with the QS model, and the simulated fuel consumption was correlated with the estimated by the MLR model, resulting in a score. Table 6.4 presents the results from the three simulations. The estimated and simulated fuel consumption is represented as well as the value of NCCP and the corresponding score. The value of each independent variable represents the factor multiplied with the initial value of the data implemented in the QS model.

Table 6.4: *Results from the three simulations compared with the estimated values.*

Variable	**Sim. 1**	**Sim. 2**	**Sim. 3**
ICE Torque	1.1	1.5	1.2
EM Torque	1.05	0.6	1.2
Curb Weight	0.86	1.3	1.2
Fuel Consumption Estimated [L]	1.3469	2.4051	1.5188
Fuel Consumption Simulated [L]	1.3992	3.1879	1.7165
Fuel Difference [L]	0.052298	0.78271	0.19774
NCCP	0.92	0.69	0.82
Score	9	7	8

Figure 6.6, 6.7 and 6.8 illustrated the estimated and simulated fuel consumption throughout the WLTC cycle.

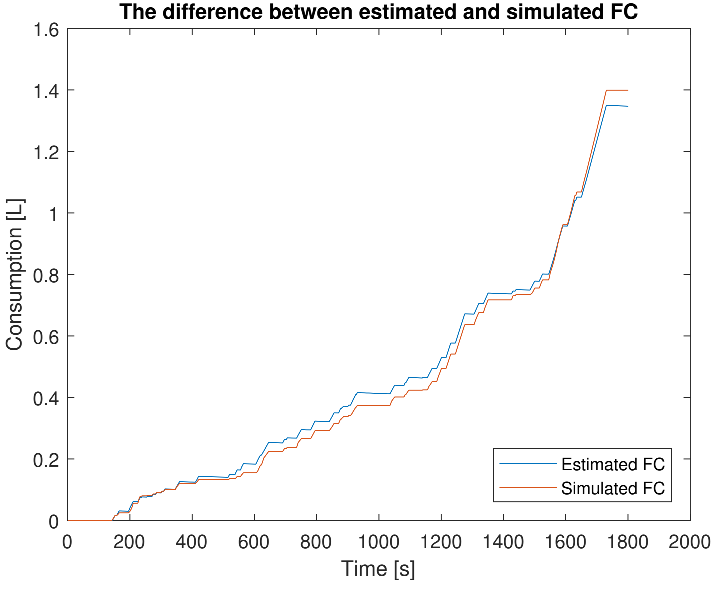

Figure 6.6: *Fuel consumption from simulation 1 with the component properties within the boundaries of the MLR.*

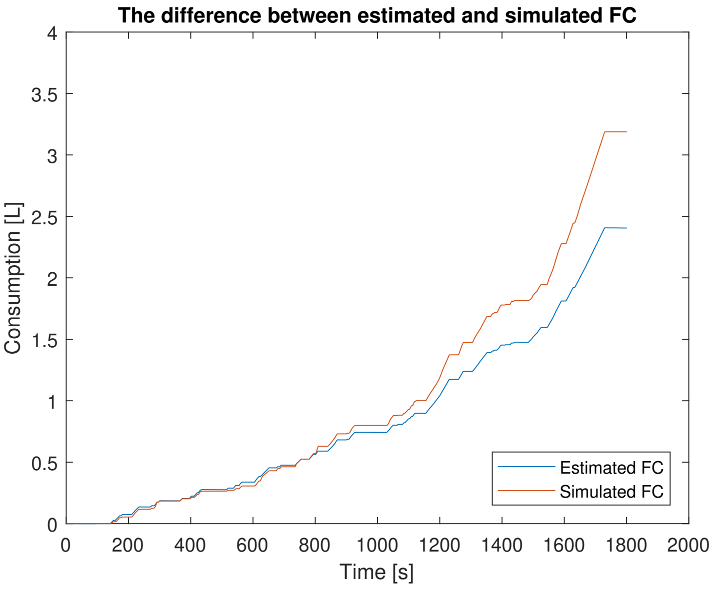

Figure 6.7: *Fuel consumption from simulation 2 with the component properties outside the boundaries of the MLR.*

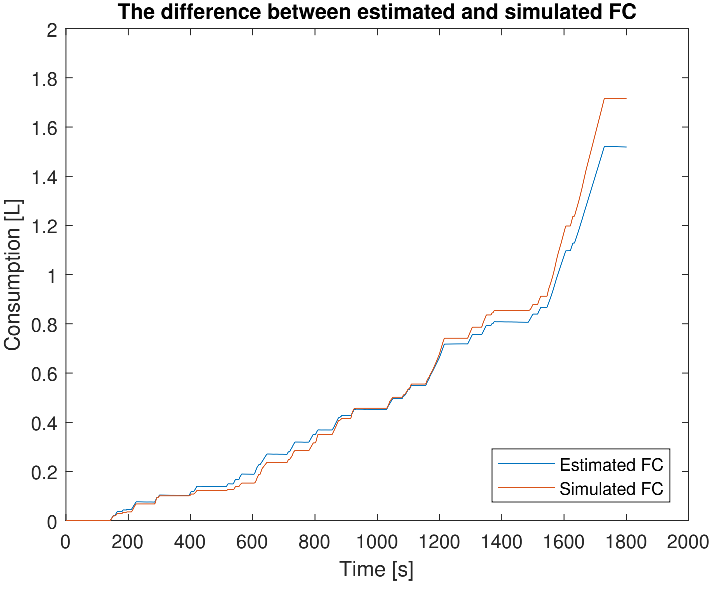

Figure 6.8: *Fuel consumption from simulation 3 with the component properties outside the boundaries of the MLR.*

7 Discussion

This chapter contains the analysis of the results and the method presented in chapter 4, 5 and 6.

7.1 DCT Model

The analysis of the results and the method of the development of the DCT model is presented in this section.

7.1.1 Method

For the development of the DCT model, several assumptions and simplifications where made. The optimisation algorithm was developed by studying the papers from related research, section 3.3 and section 3.2 together with the theory in section 2.3 and the supervisor at CEVT AB. The optimisation algorithm resulted in a time-inefficient model and the simulation time increased by a factor of six compared to a model without the algorithm. This further affected the database generation, and fewer variables and factor increments could be selected. Instead of implementing the optimisation algorithm, a gear shift map based on vehicle speed and torque can be implemented. It would reduce the computational time and also better reflect the execution of the gear shifting process of a modern car, due to that a gear shift map is implemented in the majority of modern cars.

The developed simulation model does not consider the constraint of being charge sustaining, meaning that the final SoC is equal to the initial SoC. The constraint is one of six constraints listed in section 3.3 implemented to ensure normal operation of the HEV. The absent of the constraint results in a model which does not consider the final value of the SoC. The reason why the final SoC is close to the initial SoC for the performed simulations is due to the simulation-to test data comparison explained in section 4.5. This disadvantage would cause problems if any other drive cycle were intended to be analysed.

The model lacks other SoC constraint as well. No constraint is implemented for the minimum and maximum SoC. To increase the robustness of the model, minimum and maximum SoC should be implemented but also an SoC target. The SoC target may be a function of vehicle speed, higher SoC target at higher vehicle speed resulting in increased use of the EM at lower speeds. The SoC target is the value of SoC which the gear selection strategy strives to. The SoC level is high/ low compared to the target SoC resulting in charge/ boost from the EM. The lack of SoC constraints results in a model only simulation compatible with the WLTC drive cycle.

There is a delay of the gear shifting to reduce the number of shifting events throughout the driving cycle. This delay is set to 10 seconds due to an extremely fluctuating behaviour of the SoC, and the strategy tend to change gear very often resulting in many gear shifting events through the driving cycle. It would instead be beneficial to implement a cost function with a varying weighting factor to decrease the shifting events, and also take into consideration the increased fuel consumption of the ICE as it is turned on/off.

By only modelling the DCT, there exist several simplifications and assumptions for the other components in the vehicle model, which affects fuel consumption. Therefore, it is more challenging to develop the gear selection strategy due to that it is supposed to compensate for the simplifications and assumption made in other parts of the vehicle model to reflect the vehicle behaviour of the fuel consumption through the drive cycle. Instead, it would be beneficial to develop the DCT model separately so that the input and the output signals of the DCT model could be evaluated compared to the test data separately and not for the overall vehicle model. This would result in a DCT model being able to better replicate the behaviour of the transmission in the test vehicle.

7.1.2 Results

The results of the QS model is compared to the measured test data to verify the performance of the QS model. The results for the model is presented using Pearson's correlation factor ρ, NCCP and the mean relative error. Pearson's factor is an efficient way to describe a correlation between two signals, but it does not consider time delays in signals, that is where NCCP comes in. NCCP is developed explicitly to consider the delays and lag in the signals.

The SoC result is a good indicator of the performance of the model throughout the WLTC cycle. The purpose of the gear selection strategy explained in chapter 4 was to develop a strategy providing a fuel consumption of the simulation model as similar to the fuel consumption in the test data as possible. The change in SoC is equivalent to the use of the EM and the ICE and is, therefore, a vital signal to observe. The SoC is expressed in percentage, and the SoC value of each time step depends on the previous value of the SoC leading to secondary fault for the SoC throughout the drive cycle. Here, NCCP shows its beneficial way of calculating the correlation when it considers the secondary fault and gives a more fair value of 0.8579. Pearson's correlation ρ does only consider the difference for each time step resulting in a value of 0.5535. The relative error of 14.29% is caused by the difference in the second part of the cycle. The SoC is fluctuating around a fixed value due to few implemented constraints on the EV-mode, causing the vehicle to often enter and leave the EV-mode.

Fuel consumption is the most crucial signal for the DCT model development. Fuel consumption does not have a fast response, which results in smooth behaviour throughout the cycle. The scatterplot in figure 6.2 shows that the fuel consumption for the measured data is higher than simulated through almost the whole cycle. This may be caused by the optimisation algorithm developed to minimise the specific fuel consumption, resulting in lower fuel consumption for each time step for the simulated fuel consumption compared to the measured. The NCCP value is 0.8142, and ρ equal 0.9889 together with the relative error of 25.71%. ρ indicates an extremely strong correlation, but NCCP does not.

The gear selection priority order of the EM depends on the vehicle speed and if the vehicle is in regenerative braking mode. The strategy selects the gear highest in the priority order able to deliver the requested torque of the EM. The result of this strategy is shown in figure 6.3. The following of the selected gear is good with a couple of exceptions. The strategy selects gear number four for regenerative braking, causing some differences in the gear selection at the beginning of the cycle.

The gear selection of the ICE depends on the requested load for the ICE. The strategy selects the gear resulting in the lowest specific fuel consumption, which fulfils the implemented constraints. The result is shown in figure 6.4, where it is possible to see that the simulation shift gear more frequently compared to the measured data. A penalty function for gear events could be implemented to reduce the error.

By studying the selected gear for the measured- and simulated signal in figure 6.4. It is possible to note that the simulation model tends to shift gear more frequently and that it jumps a gear, which is not the case for the measured signal from the test vehicle. Both the frequent gear shifting and the jumping of the gears are caused by the lack of implemented constraints. The only constraint affecting the gear shifting frequency is the constraint of not allowing a gear shift within 10 seconds after a shift event. No constraints of not jumping gears exist which clearly can be observed in the figure.

7.2 Ranking Model

The analysis of the results and the method of the development of the ranking model is presented in this section.

7.2.1 Method

The ranking model is based on a database generated from the QS model, causing low diversity in data generation. By defining the QS model as ground truth, the score of the ranking model is based on the MLR model and the QS model, which removes the error between the QS model and the real world, since errors between QS and the real world are not determined. This causes a problem when interpreting the results of the ranking model. The user of the RM is interested in the performance of the QS model compared to the MLR model based on real-world data and not by data generated from the same QS model. The problem may be solved by either improving the QS model, so its results correspond better to the test result, which also leads to higher correlation values between the QS model and the test data. Another solution can be to include an error term corresponding to the error between the QS model and the test data. The error is not constant, so extensive work is needed to develop the error term.

As written in the section above, the database is generated by the QS model. A simulation model resulting in poor correlation implies an inadequate database. It would instead be beneficial to have a database generated from values provided from vehicle tests. This would also result in predictions made by the MLR model corresponding better to the real world.

Many of the potential independent variables were removed before performing LASSO regularisation caused by lack of computational power. For best result, all potential independent variables would have been implemented in the LASSO regularisation function when performing the variable selection. The database generation took approximately 13 [h] needed for the LASSO regularisation. The value of the variables was generated for five different factors, meaning that for an additional variable, the computational time is increased by a factor of five.

The regression model is based on the assumption of a linear relationship, which can cause inaccuracies in the estimation of the signals. A non-linear regression model might estimate fuel consumption with greater accuracy. There also exist several deep learning methods which could be considered for the estimation of fuel consumption.

The normalised cross-correlation power, NCCP, used for determining the score between two signals was published in a paper by Yan Meng at Ford Motor Company. After contacting Yan and further discussion, it was decided that the NCCP is a reasonable approach for comparison of these type of signals.

7.2.2 Results

LASSO regularisation was used to thoroughly evaluate which signal should be included in the regression model, so-called variable selection. Table 6.2 shows the results of the independent variables contributing to the result of the dependent variable fuel consumption. LASSO only removed the battery capacity as an independent variable due to that it barely affects the fuel consumption as it is varied, showed in figure 6.5. The QS model is only simulated for the WLTC drive cycle. Therefore, the capacity of the battery is enough for all tested values. The battery capacity would probably have a greater impact for a longer drive cycle or if the intention was to drive as far as possible fully electric.

To increase the need for the LASSO regularisation, an increased number of variables would have to be implemented in the function. Unfortunately, it requires a larger database which was not possible to generate caused by lack of computational power. It would, however, be beneficial to only use LASSO regularisation as the variable selection process, instead of reducing the number in advance.

The development of the MLR model follows the steps of how to receive a model resulting in well-performed predictions. It would be interesting to develop a non-linear regression model and study the differences, already discussed earlier. The largest factor of misleading predictions for the developed model is that the model is trained och tested on a database generated by the QS model, which does not entirely follow the behaviour of the tested vehicle.

The MATLAB function `regress` is an efficient function providing clear and structured outputs needed for building the MLR model. Three different simulations were performed to evaluate the performance, shown in table 6.4.

The first simulation presents a simulation including a slightly more powerful ICE and EM together with a reduced curb weight. The simulation result is close to the estimated fuel consumption, which is not surprising since all of the component changes are within the interval of the database, presented in table 5.1. The value of NCCP is 0.92 resulting in the score of 9, which corresponds to a reliable simulation.

The second simulation contained a more powerful ICE but on the contrary, a less powerful EM. The curb weight was increased to weight out of the interval of the database. The simulated fuel consumption is larger than the estimated, resulting in a lower value of NCCP, 0.69, equivalent to the score of 7. This result is expected since both the ICE torque and the curb weight is above the interval of the database used for generating the MLR equation. The result indicates for implementing a non-linear regression model. The poor correlation indicates that the simulation is not reliable, and further investigation should be performed.

The third simulation uses variable factors at the edge of the database interval. The torque of the ICE and EM were increased as well as the curb weight. This simulation test was performed to see how well an estimation work with a simulation close to the database interval edge. The value of NCCP was 0.82, leading to a score of 8, which cannot be considered as a reliable simulation.

The first and third estimation occasionally has negative fuel consumption. This is abnormal in the physical world but the MLR is a mathematical model and, therefore, further constraints should be implemented. The main focus was on the final fuel consumption and NCCP is a good method to confirm that the result of the estimation is not at random.

The results from these three simulation tests can be expected because of the method of generating the database. By generating a database based on the simulation model, which then is used to calculate the correlation between the simulated and estimated values, relatively high correlation is expected. A database generated based on test data from different driving cycle tests would probably result in lower values for NCCP and the ranking score.

8 Conclusions

This chapter presents the conclusions and suggested future work based on the results.

8.1 Conclusions

The first step in this thesis was to develop a DCT model for the current simulation model, which was achieved. The implemented DCT is built on two separated gear shafts with the ability to engage the requested gear correctly. Model errors were also determined, evaluated and implemented in the gear selection strategy resulting in a model behaviour more similar to the test vehicle. General gear selection constraints of an HEV was also implemented, increasing the similarities of the strategy to the test vehicle.

A ranking model was also developed. It can predict the fuel consumption of the WLTC drive cycle based on a generated database. The ranking model provides a score based on the evaluation of the simulation result by calculating the correlation between the simulation result and the predicted result. The ranking model is a fast first step of evaluating a new vehicle configuration concept. As mentioned in the background of the thesis, the intention of the ranking model is to perform fast-running simulations of new concepts.

NCCP is a useful measure when the signals tend to lag. The ranking model is a solid base for the evaluation of simulation results. The MLR method is an efficient way of performing simple estimations of the fuel consumption of vehicle concepts where no test data yet exists. The lack of data provided by real-world tests causes inaccuracies in the ranking model, causing an unreliable scoring system. Therefore, real-world test data should be used for the estimation of fuel consumption to provide a reliable scoring system for the ranking model.

8.2 Future Work

This section contains suggestions of necessary work for improving the models.

8.2.1 DCT Model

- Implement a gear shift map based on vehicle speed and wheel torque instead of the optimisation algorithm. A gear shift map is implemented in modern cars and is much easier to modify for different results. Selecting the gear earlier in the model will also simplify the calculation throughout the model. The calculations do not have to be made for all gears in every stage. It will also reduce the computational time for the simulation model. It also allows for testing different shift maps.
- Include a penalty function for the events of the ICE turning on, this will reduce the gear shifting events through the driving cycle resulting in more similar gear shifting behaviour to the test vehicle. Section 3.3 in related research describes a useful method.
- Include an SoC map in the model which contains the lower- and upper bound of SoC together with the SoC target. The SoC target is the level of battery charge the gear selection strategy is aiming for. The SoC target is a function of the vehicle speed, the higher the vehicle speed, a higher SoC target is set to increase the battery level and instead use the EM at lower speeds later on.
- Implement a charge sustaining constraint. This is important when comparing results to test data performed by vehicles having the constraint implemented in the gear selection strategy.

8.2.2 Ranking Model Development

- Include more independent variables for the gear selection using LASSO regularisation. By increasing the number of variables, the probability of finding suitable independent variables for the regression model is increased.
- Use a database based on actual vehicle data for the training and testing of the MLR model instead of a generated database. This will result in a ranking model able to provide a reasonable score of the result received by the simulation model.

Bibliography

[1] L. Nielsen A. Fröberg. Efficient drive cycle simulation. *IEEE Transactions on Vehicular Technology*, 57(3):1442–1453, 2008.

[2] Lazim Abdullah and W Leong. The relationship of economic variables and final energy consumption: multiple linear regression evidence. *MATEC Web of Conferences*, 189:10025, 01 2018. doi: 10.1051/matecconf/201818910025.

[3] O'Connell R. T. Murphree E. S. Bowerman, B. L. *Regression Analysis : Unified Concepts, Practical Applications, Computer Implementation*. Business Expert Press, first edition, 2015.

[4] Cedric De Cauwer, Joeri Van Mierlo, and Thierry Coosemans. Energy consumption prediction for electric vehicles based on real-world data. *Energies*, 8:8573–8593, 08 2015. doi: 10.3390/en8088573.

[5] Jay Devore. *Probability and statistics for engineering and the sciences*. 01 2016. ISBN 9781305251809.

[6] Enrico Fichera. Vehicle cycle simulation: gear and load-point shift for optimum efficiency of hybrid powertrains. *Chalmers University of Technology*, 2018.

[7] Sciarretta A. Guzzella L. *Vehicle Propulsion Systems- Introduction to Modeling and Optimization*. Springer, second edition, 2005.

[8] Trevor Hastie, Rob Tibshirani, and Jerome Friedman. The elements of statistical learning: Springer. *Elements*, 1:241–246, 01 2009.

[9] Carson II and John S. Verification validation: model verification and validation. pages 52–58, 12 2002.

[10] A. Amstutz L. Guzella. Cae tools for quasi-static modeling and optimization of hybrid powertrains. *IEEE Transactions on Vehicular Technology*, 48(6): 1762–1769, 1999.

[11] Byungho Lee, Soduk Lee, Jeff Cherry, Anthony Neam, James Sanchez, and Ed Nam. Development of advanced light-duty powertrain and hybrid analysis tool. 04 2013. doi: 10.4271/2013-01-0808.

[12] Guoqiang Li and Daniel Görges. Optimal integrated energy management and shift control in parallel hybrid electric vehicles with dual-clutch transmission. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, page 095440701985741, 06 2019. doi: 10.1177/0954407019857419.

[13] Yan Meng, Mark Jennings, Poyu Tsou, David Brigham, Douglas Bell, and Ciro Soto. Test correlation framework for hybrid electric vehicle system model. *SAE International Journal of Engines*, 4:1046–1057, 06 2011. doi: 10.4271/2011-01-0881.

[14] Ryborz J. Novak W. Naunheimer H., Bertsche B. *Automotive transmissions*. Springer, second edition, 2011.

[15] Kevin Newman, John Kargul, and Daniel Barba. Benchmarking and modeling of a conventional mid-size car using alpha. 2015:1–12, 04 2015. doi: 10.4271/2015-01-1140.

[16] Johannes Ritzmann, Andreas Christon, Mauro Salazar, and Christopher Onder. Fuel-optimal power split and gear selection strategies for a hybrid electric vehicle. 09 2019. doi: 10.4271/2019-24-0205.

[17] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58:267–288, 01 1996. doi: 10.1111/j.2517-6161.1996.tb02080.x.

[18] US Naval Academy. Cross Correlation. `https://www.usna.edu/Users/oceano/pguth/md_help/html/time0alq.htm`, 2020. Online; accessed 27 April 2020.

[19] Junlin Yao and Ayman Moawad. Vehicle energy consumption estimation using large scale simulations and machine learning methods. *Transportation Research Part C Emerging Technologies*, Volume 101:276–296, 03 2019. doi: 10.1016/j.trc.2019.02.012.

[20] Xia Zhi, Song Yuexin, Ma Jin, Zhou Lujie, and Dong Zijian. Research on the pearson correlation coefficient evaluation method of analog signal in the process of unit peak load regulation. pages 522–527, 10 2017. doi: 10.1109/ICEMI.2017.8265997.