NTNU
Norwegian University of
Science and Technology

Online Modelling of Fuel Efficiency Curves in Generating Machines

Elene Marie Espeland

Master of Science in Cybernetics and Robotics
Submission date: June 2018
Supervisor: Tor Arne Johansen, ITK
Co-supervisor: David Anisi, ABB

Norwegian University of Science and Technology
Department of Engineering Cybernetics

Problem formulation

Given a total power demand, P_d, current practice of equal share in the process industry is to distribute the load among power generating units (e.g., diesel/gas/wind turbines) in accordance with

$$P_i = (\frac{P_{maxi}}{\sum_j P_{maxj}})P_d$$

where P_{maxi} denotes the maximum (active) power for the i^{th} machine. However, the efficiency of generators, as measured by, *e.g.*, Brake Specific Fuel Consumption (BSFC) curves for diesel/gas turbines, are highly individual, even in the case of units from same brand and model. Thus, by considering and utilizing these individual differences, it is possible to share the load in a more fuel/cost/energy optimal manner among the generators. ABB case studies in Paparella et al. (2013), show improvement potential in the range of 4,5-6% which is highly interesting from an operation point of view.

To continue this work the student should perform online modelling of performance for individual machines, that is, real-time estimation of the BSFC curves from noisy and poor quality measurements. The expected output from the master thesis is tailored algorithms for BSFC modelling where the algorithms are simulated with real data from site using MATLAB to study the system response of these algorithms when run in realistic industrial environments. The final algorithm should also be integrable with a load sharing algorithm developed by a fellow M.Sc. student.

Abstract

Gas turbines and diesel engines are important for generating energy in the process industry. Any optimization of the performance in such machines would contribute to reduce the emissions of greenhouse gases in the atmosphere as well as costs regarding fuel consumption. The current load distribution practice in the industry is to use the principle of *equal share*. Some research has been conducted on the alternative approach with optimal load sharing, however, these methods are often based on general fuel efficiency curves, *i.e.*, static curves from the manufacturer, leading to a sub-optimal solution.

In this thesis, four algorithms have been developed to estimate the individual efficiency curves in generating machines, that is the Brake Specific Fuel Consumption curves (BSFC). Classification of operational state with noise filtering, outlier detection and estimation of new curve parameters using least squares regression in real-time are included. The BSFC curves are modelled using a power series in Algorithm A and B; however, the logic of choosing which measurements to include in the estimation is different between the algorithms. Algorithm A is based on a First-In-First-Out logic, while Algorithm B includes weighting based on load, time and ambient temperature measurements. Algorithm C uses instead a second order polynomial to model the BSFC curves and only includes measurements above 40%. Algorithm C is further based on a FIFO logic as in Algorithm A. In addition, an offline algorithm was developed, Algorithm D, which estimate the initial BSFC curve using two models; power series and second order polynomial.

The proposed algorithms are simulated with data from site to study the system response using real operating conditions. The curves are updated every 30 minutes, and the estimated parameters are passed on to a load sharing algorithm developed in Jung (2018), calculating optimal set points regarding efficiency for the k number of generating machines. Finally, the results are discussed and compared to reveal strengths and weaknesses in the systems.

The optimal load sharing algorithm is currently only simulated using fixed BSFC curves, which resulted in cost savings of 2.48%. This result will be more reliable and possibly slightly different when the online estimation of the BSFC curves are applied, this is because of variations in the BSFC curves during the simulation interval due to ambient conditions.

Simulations of the online parameter estimation algorithms developed in this thesis indicate that there are small changes in the BSFC curves through the simulation interval. However, these small changes can have a considerable effect on the fuel consumption in these engines leading to cost savings and less emissions. Hence, online modelling of the BSFC curves should be applied in combination with a load sharing algorithm in the process industry.

Sammendrag

Gassturbiner og dieselmotorer er viktige for å generere energi i prosessindustrien. Enhver optimalisering av ytelsen i slike maskiner vil bidra til å senke utslipp av drivhusgasser i atmosfæren og i tillegg redusere brensel- og vedlikeholdskostnader. I dag, er prinsippet "equal share" benyttet for å fordele last mellom maskiner i industrien. Noen studier er utført på en alternativ metode ved å fordele lasten ved hjelp av å formulere et optimaliseringsproblem, men disse metodene baserer seg ofte på en generell brenseleffektivitetskurve, dvs. statiske kurver fra produsenter som fører til en sub-optimal løsning.

I denne masteroppgaven er fire algoritmer utviklet for å estimere de individuelle effektiviteskurvene i gassturbiner, nemlig "Brake Specific Fuel Consumption" kurver (BSFC). Klassifisering av operasjonell tilstand, redusering av støy og feil målinger, samt estimering av nye kurve parametere ved bruk av minste kvadraters regresjon i sanntid er inkludert. BSFC kurvene er modellert med en potensrekke i Algoritme A og B, men logikken for å bestemme hvilke målinger som estimering skal baseres på er forskjellig mellom de to algoritmene. Algortime A bruker "Først-Inn-Frst-Ut" logikk (FIFU), mens Algoritme B vekter målingene basert på last, tid og lufttemperatur. Algoritme C bruker et andre grads polynom til å modeller BSFC kurven men inkluderer kun målinger med $last > 40\%$. Algoritme C er videre basert på en FIFU logikk, som i Algoritme A. Til slutt ble en algoritme som ikke kjører i sanntid utviklet, Algoritme D, hvor den initielle BSFC kurven be modellert med to modeller; potensrekker og andre grads polynom.

De foreslåtte algoritmene er simulert med data fra et prosesseringsanlegg for å undersøke systemresponsen under virkelige operasjonelle forhold. Kurvene er oppdatert hvert 30 minutt, og de estimerte parameterne er videresendt til lastfordelingsalgoritmen utviklet i Jung (2018), hvor nye sett punkter for de k gassturbinene er beregnet. Videre er resultatene diskutert og sammenlignet for å avsløre styrker og svakheter i systemene.

Den optimale lastfordelingsalgoritmen er på nåværende tidspunkt kun simulert med statiske BSFC kurver som resulterte i en kostnadsbesparelse på 2.48%. Dette resultatet vil bli mer troverdig og trolig oppdatert når algoritmene for å estimere BSFC kurver utviklet i denne oppgaven blir benyttet. Dette er grunnet variasjoner i BSFC kurvene over tid på grunn av ytre påvirkninger.

Resultater fra simulering av algoritmene utviklet i denne oppgaven indikerer at det er små endringer i BSFC kurvene gjennom simulerings intervallet. Disse små endringene vil allikevel kunne ha stor effekt på drivstofforbruket i gassturbinene som kan føre til kostnadsbesparelser og mindre utslipp. Modellering av BSFC kurver med oppdatering i sanntid bør altså benyttes i kombinasjon med en lastfordelingsalgoritme i prosess industrien.

Preface

This master thesis has been executed at the Norwegian University of Science and Technology in the Department of Engineering Cybernetics and in collaboration with ABB. The work was conducted during the spring 2018 semester and is a continuation of a M.Sc. pre-project performed fall 2017 (Espeland, 2017). Hence, background and motivation as well as theory of the basic principles in generating machines are partly reused from the M.Sc. pre-project in this thesis.

The project work has been supervised by Professor Tor Arne Johansen in cooperation with Dr. David Anisi, Principal R&D Engineer ABB. Professor Johansen has acted as the supervisor in charge of quality and structure in the final report and Dr. Anisi acted as the main supervisor throughout the semester. The work in this master thesis has been performed with a high degree of independency with few and efficient advising meetings (Skype) with ABB. Fellow M.Sc. student Byung Kyu Jung has worked in parallel within the related project regarding optimal load sharing.

The system was developed with a data driven method that required data from a customer of ABB. Unfortunately, the work was delayed for approximately one month due to access problems to the ABB database, causing a loss of valuable time which could rather been used to improve the system behavior. In addition, the computer provided by NTNU had limited processing capacity leading to slow simulations, which again reduced the available time to further improve the quality of the system behavior.

I would like to thank Professor Tor Arne Johansen for excellent guidance and quick response on my questions. I am also grateful to Dr. David Anisi, for the time he could use for Skype meetings. I am also thankful for the opportunity given by ABB to study this interesting topic in a combined project and master thesis. Finally, I would also like to express my thanks to Byung Kyu Jung for excellent collaboration and discussions during the project period.

Trondheim, 2018-06-18

Table of Contents

List of Tables

List of Figures

Symbols

ϕ	Model basis function
$\boldsymbol{\theta}^*$	Unknown parameters to be estimated
$\hat{\boldsymbol{\theta}}$	Estimate of $\boldsymbol{\theta}^*$ in a linear system
$\hat{\boldsymbol{\theta}}_{\boldsymbol{ls}}$	Estimate of linear transformed system
$\hat{\boldsymbol{\theta}}_{\boldsymbol{nls}}$	Estimate of original parameters $\boldsymbol{\theta}^*$ in a nonlinear system.
F	Fuel consumption
P	Active power
P_{max}	Maximum power
s	Number of samples
s_r	Sampling rate
T	Time stamp of measurements
w	Window size

Abbreviations

BSFC	Brake Specific Fuel Consumption
ESS	Energy Storage System
FIFO	First In First Out
GT	Generating Turbine
LLSR	Linear Least Squares Regression
MINLP	Mixed Integer Non Linear Programming
MWLS	Moving Window Least Squares
NaN	Not a Number
NLLSR	Non Linear Least Square Regression
RLS	Recursive Least Squares
WLS	Weighted Least Squares

Chapter 1

Introduction

1.1 Background and motivation

Fossil fuels have been widely used to generate power since the industrial revolution in the 18^{th} century. First coal was used to fuel steam locomotives and in the mid 19^{th} century oil was discovered and the consumption of fossil fuels increased drastically (Revelle and Suess, 1957). In this period, the ambient temperature on earth also increased, and it is believed that these events are correlated and the burning of fossil fuels has a negative impact on the environment. It is therefore important to reduce the emission of greenhouse gases from burning fossil fuels.

A contributor to these emissions are fueling of generating machines in the process industry. Generating machines can be turbines and engines fueled with mainly gas or diesel. Today, the current practice is to use the principle of "equal share" between the generating machines, that is, the machines take equal share of their maximum power because it is assumed that similar generating units have identical characteristics. This approach is not optimal because the generating units differ in efficiency with regards to operational hours, external factors like the ambient temperature, quality of the fuel and dirt (Wärtsilä, 2015). Hence the underlying idea is to exploit these differences and share the load optimally. This method can give several benefits like reduction of environmental harmful emissions and maintenance costs.

To measure efficiency of generating machines, the Brake Specific Fuel Consumption (BSFC) can be used. BSFC is the amount of fuel consumed divided by the energy being produced, see Section 2.1.3 for further explanation. Earlier research has used the principle of optimal load sharing between generating machines, but the approach has been to use a BSFC curve provided by a performance test or given from the manufacturer, and not using actual machine efficiency as described in the feasibility study performed by ABB in Paparella et al. (2013). This approach was applied in Skjong et al. (2017), where economic energy management was studied in marine vessels. The data was obtained from a ferry, platform supply and seismic survey vessel and an optimal Energy Storage system is proposed. The Specific Fuel Oil Consumption (SFOC, consumption of fuel oil per energy

unit) used in this study is obtained from supplier and is based on acceptance and certification tests. However, the SFOC/BSFC can differ significantly over time between identical machines due to factors mentioned above, hence the BSFC curves need to be updated continuously to ensure an optimal solution of the load sharing problem.

An alternative approach has been investigated in Lundh et al. (2016), the approach considers online estimation and optimization of load sharing in diesel electric marine vessels. The results obtained are interesting where two online estimation algorithms based on operational data of diesel engines in a cruise ship are considered. The first method uses recursive least squares to find the parameters of a second order polynomial to describe the SFOC, (however this is not complex enough to capture the shape of the measurements in the generating machines used in this project). The second method propose that the SFOC can be approximated by a set of values, however, this is not well suited for the continuous optimization approach which the result in this thesis is applied in.

In Jehlik and Rask (2010) Response Surface Methodology (RSM) was used to develop BSFC maps of a test vehicle, the study was conducted to investigate the effect of ambient temperature on the efficiency of the engine in the vehicle. Least squares regression was applied to estimate the fueling rate and a 6^{th} order polynomial resulted in being the best model describing this. The research performed in Jehlik and Rask (2010) was based on data from cold starts of the engine in a 2005 Chevrolet Equinox, that is the transient operation in the vehicle. By contrast, transient operation of generating units is not the focus in this thesis, rather estimation of a model describing the performance in steady-state operation using a power series instead of a 6^{th} order polynomial.

This thesis is motivated by the lack of research in the field of online estimation of BSFC curves that are updated according to new measurements and ambient conditions. These curves will further be utilized in optimal load sharing between generating machines leading to a more energy efficient operation.

1.2 Goal and method

The goal of this master thesis is to develop and compare tailored algorithms that estimates the BSFC curves online from noisy measurements in all generating turbines (GT). Few similar methods exist; hence the main focus will be on developing a complete system including pre-processing of the data measurements to remove noise and outliers, measurement handling, curve estimation and adaptive update of the curve. The pre-processing step, outlier detection and offline modelling of the nominal BSFC curve for all turbines is already covered in the M.Sc. pre-project in Espeland (2017), hence the remaining parts of the system will be on updating the BSFC curve online and through simulations find the optimal parameters in the different algorithms in addition to connecting all system parts together.

The output of the system is going to be the estimated parameters, these parameters are updated every 30 minutes and act as input in an optimal load sharing problem to increase the overall efficiency in the GTs. The optimal load sharing is performed by Jung (2018), with the optimal set points for each GT as output.

Linear and non-linear least squares regression are used to estimate the BSFC curves, and logic ensuring the curves to represent current ambient conditions are compared. The

system including both online modelling of BSFC curves and optimal load sharing is simulated using measurements obtained from site in approximately a period of one year in addition to temperature data obtained from the Norwegian Meteorological Institute (NMI). The sampling rate of the measurements are $s_r = 10$ seconds, and the 8 first months of the measurements are used in modelling the initial BSFC curve offline (as well as predicting future power demand in Jung (2018)).

Compared to the current practice in the industry, estimating the BSFC curves independently in each GT and accounting for individual differences and further apply optimal load sharing can lead to a more energy efficient operation. This is shown in previous case studies performed by ABB in Paparella et al. (2013), which resulted in 4-6.5 % cost savings and thus decrease in emissions of greenhouse gases.

1.3 Outline of the report

The first part in this master thesis is theory and background covered by Chapter 2. The first section, Section 2.1, presents theory regarding generating machines including basic principles, factors affecting the performance and the definition of BSFC and the optimal load sharing problem. Section 2.2 includes a short summary of the results obtained in the pre project of this master thesis, while Section 2.3 covers theory regarding the data fitting problem. Here the concepts of power series and least squares are presented. The last section in the theory chapter, Section 2.4, includes theory of two adaptive regression methods; Moving window least squares and recursive least squares.

The developed algorithms are described in Chapter 3, with curve estimation using FIFO logic and load intervals is presented first in Section 3.1, followed by a method using weighted data points to estimate the BSFC curve in Section 3.2. Each section includes a detailed description of the process in each algorithm, as well as a flow chart describing the steps. The last section in this chapter, Section 3.3, includes additional scenarios that was implemented including estimating the BSFC curve by using only measurements with $load > 40\%$. The final algorithm is an offline approach where the BSFC curves are estimated using two models; power series and polynomial. However, these methods are not of major interest in this thesis and should be further studied to develop a comprehensive system interacting with the optimal load sharing algorithm.

Chapter 4 describes the results obtained with the proposed algorithms with a discussion of the most important findings. A more general discussion of these results and the applied methods are covered in Chapter 5. Chapter 6 provides a summary and concluding remarks of the obtained results as well as proposing future work.

Chapter 2

Theory and background

2.1 Generating turbines

This section presents the basic principles of generating machines, theory of factors affecting the performance and definition of BSFC. Section 2.1.1 is obtained from the M.Sc. pre-project and is intended for the readers that are unfamiliar with the basic workings of generating machines.

2.1.1 Basic principles of generating machines

In this thesis, measurements from gas turbines are used to develop an up to date performance model describing them. Turbines are one of two basic components needed to create electricity, and the other component is generators. They both use rotating parts, but there are some differences between them.

Turbines

A turbine is a machine that captures kinetic energy from a moving fluid, liquid or gas so it can be exploited. The force in the kinetic energy can come from steam heated by oil, gas and coal or directly from moving air or water. An illustrative example of a turbine is a windmill, it captures the kinetic energy in the wind and transforms it into mechanical energy. The main components in a turbine are the blades, a shaft or axle and a machine driven by the axle Woodford (2017). The blades capture the kinetic energy and are connected to a shaft or axle which is rotating as the blades move. The axle or shaft is connected to a machine, the type of machine depends on the application of the energy. A typical machine could be a generator that converts mechanical energy from the shaft to electrical energy.

There are several types of turbines and they differ in what kind of fuel that is used. Some turbines can run on fuel of different quality, where the fuel quality will affect the shape of the BSFC curve and hence the efficiency.

Gas turbines have several applications both in aviation and electricity generation. The generating machines studied in this project convert the energy stored in fossil fuels to shaft power (mechanical energy). The major parts of these generating machines are the compressor, combustion chamber and the turbine. The compressor compresses air, and the air is sent to the combustion chamber where it is mixed with fuel. This mixture leads to ignition and hence a high pressure air flow enters the turbine, the flow expands down the exhaust pressure, which leads to shaft work. The shaft is connected to an electric generator which converts the mechanical energy to electricity as described in Klein and Abeykoon (2015).

Generators

A generator converts mechanical energy into electrical energy. Michael Faraday discovered electromagnetic induction in 1831-1832, and today's generators are based on this principle presented in Mayer and Varaksina (2017). Faraday discovered that voltage could be induced in a coil by changing the magnetic environment around the coil, and a generator is basically a coil with electric conductors, for example, a copper wire, that turns around inside a magnetic field. The rotation of the coil can come from a diesel motor or turbine, as previously described.

However, the focus of this project is the efficiency measurement of converting energy stored in fossil fuel to active power that moves the shaft in the turbines. The efficiency measurement, Brake specific fuel consumption, is explored in Section 2.1.3.

2.1.2 Factors affecting the performance of generating turbines

The efficiency of a gas turbine can be found with field testing, this is necessary to assure the correct behavior considering performance predictions from the manufacturer. The field test is conducted by taking measurements during steady-state operation at site to determine parameters affecting the efficiency, *i.e.*, full load output power, heat rate and exhaust heat rate. Steady-state operation is when all affecting factors are stable over a minimum 10 minutes interval, that is a maximum variation of approximately $\pm 1\%$ in speed and load, and $\pm 1^oC$ in inlet and ambient temperature, see guideline for field testing of gas turbines in Brun and Nored (2006). It is important that this requirement is fulfilled because the performance test is not valid if the gas turbine is in transient operation. Some of the most important parameters considering performance will next be presented and are based on Brooks (2000).

Air temperature

A gas turbine uses air to transform kinetic energy into electrical energy. Thus the ambient temperature has an impact on the performance of the gas turbine because it affects the density and/or mass flow of the air intake. The effect of ambient temperature on turbine output, heat rate, heat consumption and exhaust flow in a single shaft MS7001 turbine was shown in Brooks (2000). It was explained that any deviations from the ISO conditions (air temperature of 15^oC and 60% humidity), affect these parameters, see Figure 2.1 obtained

from Brooks (2000). As can be observed in the figure, an increase in temperature results in a decrease in the turbine output relative to the ISO conditions.

This was studied in De Sa and Al Zubaidy (2011), and it was concluded that the thermal efficiency and power output varies according to ambient temperature. It was shown that gas turbines operating in ambient temperature higher than ISO conditions were less efficient, this is due to a lower mass flow rate generating less power as well as the compressor consuming more energy. This was also studied in gas turbines located in Turkey, Erdem and Sevilgen (2006), where it was shown that the efficiency decreases and the loss rates varied between 1.67% and 7.22% when the temperature was above 15^oC. In the same study, it was shown that the power output increased when the inlet air was cooled to 10^oC, hence preferred operating conditions are lower temperatures.

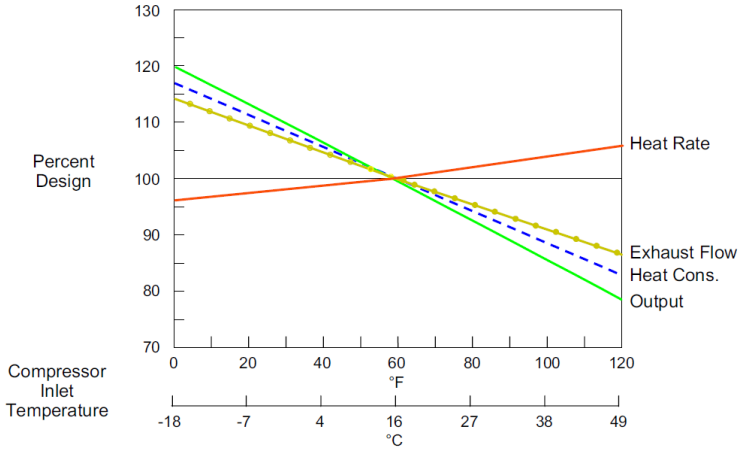

Figure 2.1: Effect of ambient temperature in a single shaft turbine, the output of the turbine decreases when the temperature is higher than ISO conditions. The figure is obtained from Brooks (2000).

Fuel

The amount of work done by generating turbines depend amongst other parameters on the mass flow, that is the sum of compressor airflow and fuel flow. Hence, the power of the turbine is affected by the fuel, and this was also validated in Section 3.1 in the M.Sc. preproject in Espeland (2017). The dataset was divided into eight periods, where every period contained data from approximately one month, and it could clearly be observed different efficiency results in period 5 compared to other periods, see Figure 2.2. It is therefore essential to distinguish between the quality of fuel when modelling the BSFC curve.

Performance degradation

With time all GTs will degrade regarding performance, this can be faults or tear that are recoverable or not recoverable. The recoverable faults are often due to dirt or other small

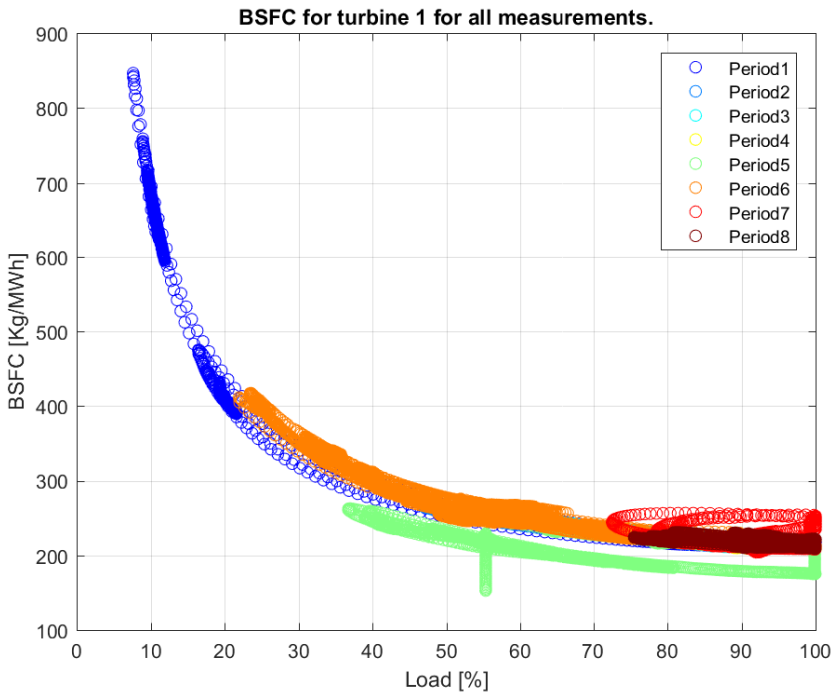

Figure 2.2: BSFC values in GT1 divided into time periods indicated with color to observe the time changes of the values, and each period contains data points from approximately one month. (Figure obtained from the M.Sc. pre-project of this master thesis Espeland (2017)).

particles in the machinery and the problem can often be solved by a thorough wash or external maintenance. The non-recoverable faults, for example worn parts, need to be replaced before the machinery crashes. The performance degradation is also affected by weather conditions, for example will a GT in dry and hot climate degrade slower than the opposite conditions. Because of this degradation, it is important to update the BSFC curve such that the performance of the GT is more exact than the initial performance curve given by the manufacturer.

Other parameters

Several parameters affect the efficiency in a generating turbine in addition to the parameters mentioned above as:

- Barometric pressure
- Power turbine speed
- Gas generator speed

- Inlet and exhaust pressure loss
- Relative humidity of inlet air
- Water/steam injection rate.

However, a requirement from ABB is that the final algorithm developed in this thesis should be data driven and have a low complexity with regards to the number of measurements needed from site. Hence, these parameters in addition to performance degradation are unknown. This is ,of course, a source of error, it is thus essential to keep that in mind when choosing curve update time and number of data points to be included in the update. The fuel quality is also unknown, but the obtained data points are assumed to be measured when using the same fuel quality. The air temperature data can be obtained, but not directly received from site in addition to the other measurements.

2.1.3 Brake specific fuel consumption

This section is partly obtained from the M.Sc. pre-project in Espeland (2017).

Efficiency of a prime mover that burns fuel, for example a turboshaft engine, can be measured by its Brake Specific Fuel Consumption (BSFC). This measurement is defined as the amount of fuel consumed per hour divided by the active power produced. The mathematical formulation is as follows

$$\textbf{BSFC} = \frac{\textbf{F}}{\textbf{P}} \tag{2.1}$$

where **F** is the fuel consumption (Kg/h), and **P** is the power produced (MW). The unit of BSFC is therefor given in Kg/MWh, where a lower value implies higher energy efficiency.

The current practice in the process industry is to share the total power demand, P_d, between the GTs according to the following equation:

$$P_i = \left(\frac{P_{max_i}}{\sum_j P_{max_j}} \right) P_d, \tag{2.2}$$

where P_{max_i} denotes the maximum active power for the i^{th} machine. This will give the wrong base for further calculations because many factors such as running hours, temperature, maintenance, quality of fuel affect the shape of this curve Jehlik and Rask (2010), Lundh et al. (2016). A simple experiment performed in *MATLAB* where BSFC curves were found for two GTs validates this hypothesis. As can be seen in Figure 2.3, GT1 is slightly more efficient than GT3.

It is exactly these differences in efficiency that is going to be exploited in the optimal load sharing problem, which can be formulated as follows

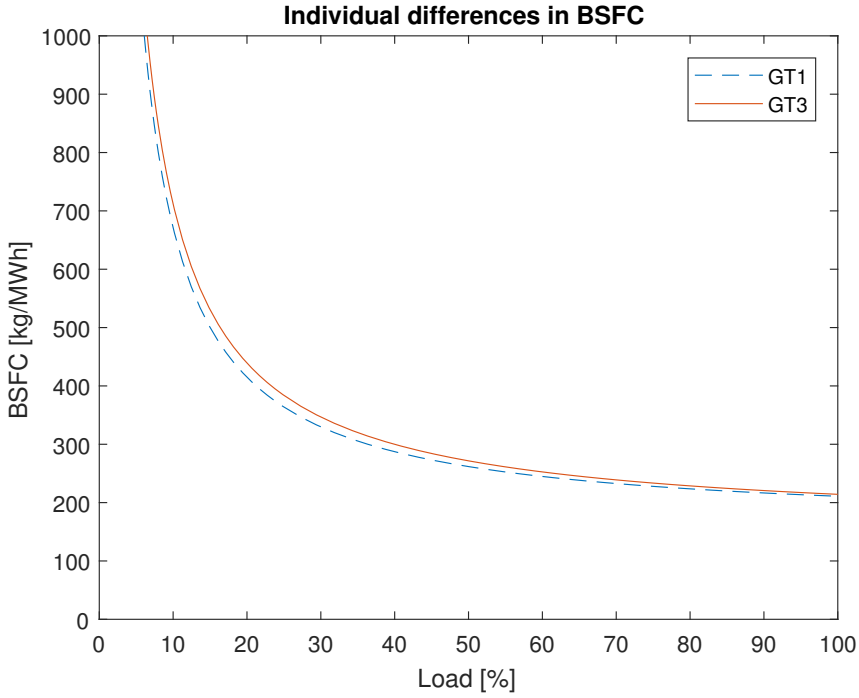

Figure 2.3: Difference in BSFC between turbine 1 and 3

$$\begin{aligned} &\text{Minimize (P)} \\ &\qquad \sum BSFC_i(P_i) \cdot P_i \\ &\text{Subject to} \\ &\qquad \sum P_i = P_d \\ &\qquad 0 \leq P_i \leq P_{max_i} \\ &\qquad \sum (i : P_i > 0) \geq N \end{aligned} \tag{2.3}$$

where BSFC is the specific efficiency curve in each turbine. N is the number of running machines and plays the role as the spinning reserve which is the unused capacity that can be activated if required. The spinning reserve is often used as a redundancy requirement to prevent possible power losses during non-predicted events Skjong et al. (2017).

When estimating the BSFC curves, it is essential that the GTs are in steady-state operation, defined in Section 2.1.2, because varying ambient conditions such as a change in load and temperature will affect the estimate and give a non-exact curve. However, the GTs are seldom in steady-state operation when the load is below 40%, but due to the load sharing algorithm requiring curves describing the entire load interval the curve might be invalid for such low loads.

2.2 Summary of the results obtained in the M.Sc. pre-project

A pre-project investigating offline parameter estimation of BSFC curves in generating turbines was conducted in Espeland (2017). The focus in the project was to compare different methodologies in pre-processing of noisy data and find the family of models that best describe the BSFC. The main results of the algorithms are presented in the following sections.

2.2.1 Pre-processing of noisy measurements

For the curve to be valid and exact, it is necessary that the curve estimation is based on data points received from steady-state operation. The GTs can operate in one of three states

1. Off
2. Transient operation
    (a) Increasing power
    (b) Decreasing power
3. Steady-state operation,

and to ensure this, three methods were compared. The first method included first using a Savitzky-Golay filter to reduce noise as well as data points from transient operation. The method succeeded in reducing noise, but the indication of steady-state operation was not sufficient.

The second method (the classification method) was specifically developed to identify different states in the GT, hence ensuring that the curve estimation is based on data points in steady-state operation. Simulation of the algorithm with all data points in all GTs gave satisfactory results where both slow and fast transients were identified. The slow transients are when the GTs are either being shut off or on, while fast transients are small changes in set points, for example noise. However, the algorithm did not succeed in removing all outliers, but this problem was overcome with an outlier detection algorithm presented in the next section.

The third algorithm was a model created by a neural network to identify the operational states of the GTs. The neural network was trained with operation data from approximately six months and tested with data from three months. The resulting model identified shut off operation well, but the identification of steady-state operation and transient operation was not sufficient, and it was concluded that several parameters were needed in the model rather than just power, maximum power and fuel consumption.

2.2.2 Outlier detection

The classification method performed well in identifying steady-state operation, however, after calculating the BSFC values for each load there existed some outliers in the data. Outliers are measurements deviating from the surrounding measurements caused by for

example noise and incorrect measurements. Consequently, two outlier detection methods were investigated. The first method included using the *MATLAB* function *isoutlier*, however, the outliers were not successfully removed. The second method was an outlier detection algorithm developed specifically for this system, that is finding outliers in BSFC values. The algorithm performed very well, and just about all outliers were removed.

2.2.3 Curve estimation

Three families of functions were investigated to find the curve that best describes the BSFC in the GTs for all loads (0-100%): polynomials, exponentials and power series. Least squares regression was used to estimate curves of different degrees for the three models. Regarding the polynomial, a 7^{th} order was needed to capture the shape of the data points. However, this led to oscillations in data points with a load larger than approximately 60 %. The exponential curve was not steep enough to capture the shape of the data points with low loads, but it was a smooth curve with small errors for larger loads. The best result was when the model was estimated with a power series, the shape of all data points was captured without oscillations, and the error was low.

The results in the pre-project are the basis for this master thesis and will function as the offline modelling of the nominal BSFC curve for the GTs. In addition, the classification algorithm and outlier detection will be applied, and the shape of the curve in the online modelling is chosen based on this research, *i.e.*, the power series.

2.3 The data fitting problem

2.3.1 Power series

After investigating several families of functions in the M.Sc. pre-project (Espeland, 2017), it was concluded that the best model for estimating the BSFC curves was a power series when including BSFC values for loads between 0% and 100%. These results will be further applied in this thesis in the online algorithm when estimating the BSFC curves.

A power series is a sum of terms, often described as a Taylor series of a function (Levi, 1967). The power series in this master thesis will consist of two terms on the following form

$$y(x) = ax^b + c \tag{2.4}$$

where a and c are coefficients and b is the power, y is the BSFC and x is the load. Hence the unknown parameters to be estimated in the online algorithm can be defined as

$$\boldsymbol{\theta}^* = [a, b, c]^T. \tag{2.5}$$

The estimate of $\boldsymbol{\theta}^*$ is denoted $\hat{\boldsymbol{\theta}}$ and will be the result of the online algorithm.

2.3.2 Linear least squares

The data fitting problem includes capturing the overall behavior of data measurements in a model without including noise. A linear parametric model is defined as follows

$$z(\boldsymbol{\phi}, x) = \sum_{j=1}^{n} \phi_j(x)\theta_j^* = \boldsymbol{\Phi}^T \boldsymbol{\theta}^*, \tag{2.6}$$

where $\boldsymbol{\theta^*} = (\theta_1^*, \theta_2^*, ..., \theta_n^*)$ are the unknown n parameters that describes the model and $\phi_j(x)$ are the *model basis functions*, Hansen et al. (2013). The method of linear least squares can be applied to estimate the unknown parameters $\boldsymbol{\theta^*}$ with the parameter $\hat{\boldsymbol{\theta}}$. Linear least squares regression (LLSR) estimates $\boldsymbol{\theta^*}$ with the highest probability of being correct as described in Johnson and Faunt (1992), the method is easy to apply and and quick to use according to Ljung (1976). The estimation of $\boldsymbol{\theta^*}$ is done by minimizing the sum of squares of the residual $\epsilon = y - \boldsymbol{\phi^T}\hat{\boldsymbol{\theta}}$ according to

$$\min_{\theta} \sum_{i=1}^{m} \epsilon_i^2 = \min_{\theta} \sum_{i=1}^{m} (y_i - \phi_i \hat{\theta}))^2, \tag{2.7}$$

where y_i is the model to be estimated. The estimated model ($\hat{y}$) used in this thesis, Equation (2.4), is non-linear, hence non-linear least square regression must be applied.

2.3.3 Non-linear least squares

Non-linear least squares regression (NLLSR) is applied when the model to be estimated is non-linear. How the parameters are estimated with NLLSR are conceptually the same as in LLSR, but the difference is that recursive optimization is needed to solve NLLSR while LLSR can be solved analytically. Also, the starting points in the optimization need to be chosen carefully, because poor starting points increase the risk of not converging, Teunissen (1990).

If the model can be written as a linear parametric model as defined in Equation (2.6), methods like Gauss-Newton, Gradient descent or the Levenberg-Marquardt algorithm can be applied, the theory of these methods are outside the scope of this master thesis, but Chapter 9 in Hansen et al. (2013) can be considered for further explanation. If the model cannot be transformed into a linear parametric model alternative methods are required to solve the NLLS problem, like machine learning or decision trees. However, the estimation model in this master thesis, given in Equation (2.4), can be transformed into a linear parametric model with the following steps:

$$
\begin{aligned}
y &= ax^b + c \\
\ln(y) &= \ln(ax^b) + \ln(c) \\
\ln(y) &= b\ln(ax) + \ln(c) \\
\ln(y) &= b\ln(a) + b\ln(x) + \ln(c) \\
\ln(y) &= \begin{bmatrix} 1 & \ln(x) & 1 \end{bmatrix} \begin{bmatrix} b\ln(a) \\ b \\ \ln(c) \end{bmatrix} \\
\ln(y) &= \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} \\
z &= \boldsymbol{\phi^T \theta}.
\end{aligned}
$$

The resulting parametric model can thus be used to estimate $\boldsymbol{\theta^*}$ with $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ and inserting it in Equation (2.7). The chosen algorithm to solve the minimization problem stops when the estimated model z converges to the real model y. Finally, the estimate of the original parameters is found by transforming the parametric model back to the non-linear system with the following equations:

$$
a = e^{\frac{\theta_1}{\theta_2}}, \quad b = \theta_2 \quad \text{and} \quad c = e^{\theta_3}. \tag{2.8}
$$

2.4 Adaptive regression methods

Least squares described in the previous section is one of the most popular techniques regarding regression Sharma et al. (2016). However, it might be necessary to estimate parameters while the system is in operation using measurements up until current time. Such methods are called adaptive parameter estimation, recursive identification methods or real-time identification. Using the definitions in Ljung (1999), a recursive algorithm is on the form

$$
\begin{aligned}
X(t) &= \hat{\theta}_{t-1} + \gamma_t Q_\theta(X(t), y(t), u(t)) && (2.9) \\
\hat{\theta}_t &= X(t-1) + \mu_t Q_X(X(t-1), y(t), u(t)) && (2.10)
\end{aligned}
$$

where $X(t)$ are states containing known information with fixed dimension, $\hat{\theta}_t$ are the parameters to be estimated at time t. The input $u(t)$ and the measurement $y(t)$ at time t generally have little information compared to the information content from the previous measurements; hence μ and γ are chosen to be small numbers to reflect this information content.

According to Ljung (1999) adaptive parameter estimation methods are often applied when a decision about the system has to be made. For example, tuning of filter parameters or change the model parameters if the accuracy is low. In the following sections, two methods to update the estimated parameter $\hat{\boldsymbol{\theta}}$ are presented.

2.4.1 Moving window least squares

The moving window least squares (MWLS) method estimates the parameters $\hat{\boldsymbol{\theta}}$ when a given number of samples, s, are obtained. That is, the least squares method described previously is repeated every s sample to update the parameters according to current conditions. Here, s is chosen based on the system dynamics, *e.g.*, a small s is chosen if there are rapid changes while a large s is chosen if the changes are slow.

The parameter vector for the MWLS can be found by first defining the error of residuals

$$\begin{aligned}\epsilon^2 &= \epsilon^T\epsilon \\ &= (y-\boldsymbol{\phi_w}\hat{\boldsymbol{\theta}}_{\boldsymbol{w}})^T(y-\boldsymbol{\phi_w}\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}) \\ &= y^Ty - y^T\boldsymbol{\phi_w}\hat{\boldsymbol{\theta}}_{\boldsymbol{w}} - \boldsymbol{\phi_w}^T\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}{}^T y + \boldsymbol{\phi_w}^T\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}{}^T\boldsymbol{\phi_w}\hat{\boldsymbol{\theta}}_{\boldsymbol{w}} \\ &= y^Ty - 2y^T\boldsymbol{\phi_w}\hat{\boldsymbol{\theta}}_{\boldsymbol{w}} + \boldsymbol{\phi_w}^T\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}{}^T\boldsymbol{\phi_w}\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}.\end{aligned}$$

Next, the partial derivative is used to solve Equation (2.7) equal to zero

$$\frac{\partial\epsilon^2}{\partial\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}} = 0. \tag{2.11}$$

Further, ϵ^2 is inserted in Equation (2.11) and solved with respect to $\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}$:

$$\begin{aligned}\frac{\partial\epsilon^2}{\partial\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}} &= 0 \\ \frac{\partial(y^Ty - 2y^T\boldsymbol{\phi_w}\hat{\boldsymbol{\theta}}_{\boldsymbol{w}} + \boldsymbol{\phi_w}^T\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}{}^T\boldsymbol{\phi_w}\hat{\boldsymbol{\theta}}_{\boldsymbol{w}})}{\partial\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}} &= 0 \\ -2\boldsymbol{\phi_w}^Ty + 2\boldsymbol{\phi_w}^T\boldsymbol{\phi_w}\hat{\boldsymbol{\theta}}_{\boldsymbol{w}} &= 0 \\ \boldsymbol{\phi_w}^T\boldsymbol{\phi_w}\hat{\boldsymbol{\theta}}_{\boldsymbol{w}} &= \boldsymbol{\phi_w}^Ty \\ (\boldsymbol{\phi_w}^T\boldsymbol{\phi_w})^{-1}\boldsymbol{\phi_w}^T\boldsymbol{\phi_w}\hat{\boldsymbol{\theta}}_{\boldsymbol{w}} &= (\boldsymbol{\phi_w}^T\boldsymbol{\phi_w})^{-1}\boldsymbol{\phi_w}^Ty.\end{aligned}$$

This gives

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{w}} = (\boldsymbol{\phi_w}^T\boldsymbol{\phi_w})^{-1}\boldsymbol{\phi_w}^Ty \tag{2.12}$$

where $\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}$ are the estimated parameters, $\boldsymbol{\phi_w}$ are the model basis function and y is the real system. Here, w denotes the window size *i.e.*, the latest w samples, hence $\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}$ are the estimated parameters in the current window w (Sharma et al. (2016)).

An alternative approach to update the $\hat{\boldsymbol{\theta}}_{\boldsymbol{w}}$ after receiving s new samples, is to use selective update and update whenever a threshold prediction error is reached, this is implemented in combination with partial least squares in Xu et al. (2014).

The MWLS is ideal to apply in the online algorithm developed in this master thesis. This is because the system is changing based on ambient conditions as explained in Section 2.1.2; hence updating $\hat{\boldsymbol{\theta}}$ is necessary to obtain an exact model. Throughout this thesis the

estimated parameters for the linear transformed system is denoted $\hat{\boldsymbol{\theta}}_{\boldsymbol{ls}} = [\theta_1, \theta_2, \theta_3]^T$, and the parameters transformed back to the non-linear system given in Equation 2.4 is denoted $\hat{\boldsymbol{\theta}}_{\boldsymbol{nls}} = [a, b, c]^T$.

2.4.2 Recursive least squares

The recursive least squares (RLS) estimation of unknown parameters is simple and computationally efficient when the regression vector is a function of time Sharma et al. (2016). There are several variants of this method, where the weighted least squares (WLS) will be considered in this section. The weighted least squares method assigns weights to measurements, that is each measurement is related with a number that reflects the relevance or importance of the measurement in the current time. Weighting the measurements also prevents losing the tracking ability of changes in the regression, that is, if all measurements from previous time instances are included in the regression, new changes might be overseen. The weighted least squares criterion is defined as

$$\hat{\theta}_t = \underset{\theta}{\operatorname{argmin}} \sum_{k=1}^{t} \beta(t,k)[y(k) - \phi^T(k)\theta]^2. \tag{2.13}$$

The equation is identical to Equation (2.7) except β, which is the matrix of weights given to each measurement included in the minimization. The weighted sequence β can be written as a product of λ

$$\beta(t,k) = \prod_{j+1}^{t} \lambda(j) \tag{2.14}$$

this implies that $\bar{R}$ at time t can be written as

$$\bar{R}(t) = \lambda(t)\bar{R}(t-1) + \phi(t)\phi^T \tag{2.15}$$

which leads to the recursive algorithm to estimate $\hat{\theta}_t$

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \bar{R}^{-1}\phi(t)[y(t) - \phi^T(t)\hat{\theta}_{t-1}]. \tag{2.16}$$

For further explanation of the deriving of the recursive algorithm, the reader is referred to Chapter 11.2 in Ljung (1999).

Chapter 3

Algorithm development

Four algorithms were developed in *MATLAB R2017a*, Algorithm A, B, C and D, to estimate the parameters of the BSFC curve in GTs. The shape of the BSFC curves is based on the research performed in the M.Sc. pre-project in Espeland (2017), where it was concluded that a power series representation was best suited. The first three algorithms consist of an offline and online modelling part where the offline modelling uses already obtained data points to estimate the nominal BSFC curve. The online modelling considers updating these parameters to be in line with current conditions, for example, temperature. The fourth algorithm is an offline modelling method and lay the basis for future work.

Least squares regression is used in all algorithms to estimate the BSFC curve. However, there are some differences between the algorithms. Algorithm A uses a First-in-First-out (FIFO) logic to update the current measurements to estimate a new BSFC. Consequently, the new curve describes the BSFC for current conditions, depending on the horizon of the measurement in the estimation. Algorithm B is also taking temperature and current *load interval* into consideration, and weight them accordingly. The measurements with the highest weight in each load interval are hence the basis in the estimation leading to the curve describing BSFC values with similar temperature conditions. Algorithm C is slightly different; here the BSFC curve is modelled using a second-order polynomial and only measurements with $load > 40\%$. The final algorithm, Algorithm D, estimates the BSFC curve using two models, power series and a polynomial to account for the different states in the system. An overview of the difference between the algorithms can be found in Table 3.1.

The primary focus in this thesis is Algorithm A and B because Algorithm C and D, referred to as additional approaches, can not yet be integrated with the optimal load sharing algorithm. This is due to the curves in Algorithm C only describing loads above 40 % and Algorithm D being an offline method in addition to the curve not being smooth. The workflow of Algorithm A, B and C in addition to some of the subparts in the algorithms are presented in flow charts in the respective sections. Where the oval shapes represent input or output, the diamond shapes represent if/else statements and the squares represent operations or functions.

Table 3.1: Overview of the differences between the algorithms developed in this master thesis.

Algorithm	Curve type	Storage logic	Load interval
A	Power series	FIFO, load intervals	0% - 100%
B	Power series	Weighting	0% - 100%
C	Polynomial	FIFO, load intervals	40% - 100%
D	Power series + polynomial	Not implemented	0% - 100%

3.1 Algorithm A - Curve estimation using FIFO logic and load intervals

Algorithm A is based on the same principles developed in Espeland (2017), where the BSFC curve was found for each GT by applying classification, outlier detection and offline parameter estimation with least squares regression. These subparts are connected in an algorithm and altered to function online, including using MWLS and a *working set* to estimate new parameters of the BSFC curve reflecting current conditions. An overview of the main steps is presented in Figure 3.1. A more detailed description of the most important system parts is presented later in this section.

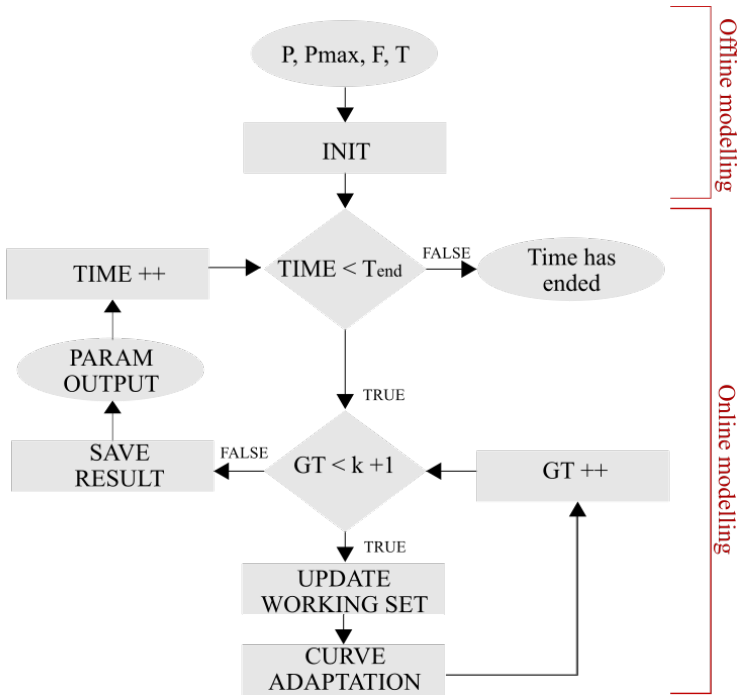

Figure 3.1: Illustration of the system parts in Algorithm A.

3.1.1 Parameters

An overview of the parameters included in Algorithm A is presented in Table 3.2, along with a short description of each. The size of each parameter is indicated in the second column, where v is a number and the remaining sizes are either matrices or vectors. Here, n is the number of measurements, m the number of different types of measurements or load intervals and k the number of GTs. The global storage has a fixed length *size_global_storage* ensuring that old data points are removed, and data points received from k GTs are stored regardless of operational state. New measurements are received every 10 seconds and are stored as a data point containing power, maximum power, fuel consumption and time of when the measurement was taken, that is, $data\_point = [P, P_{max}, F, T]$. The working set is a matrix containing data points used in estimating new parameters describing the BSFC curve, the matrix is updated with new measurements when updating the curve. The logic of the working set is however presented in Section 3.1.3.

Table 3.2: Parameter specifications for Algorithm A.

Name	Size	Description
update_time_curve	v	Update time for the efficiency curve.
size_ws	v	Size of working set.
num_turbines	v	Number of generating machines
size_global_storage	v	Maximum length of the storage containing all measurements.
low_load_threshold	v	Only measurements below this threshold are inserted in the working set when GT is in transient operation.
ws_intervals	m $\times$ 1	Size of the load interval in the working set.
global_storage	n $\times$ m $\times$ k	Measurement storage containing P, P_{max}, F and T.
working_set	n $\times$ m $\times$ k	Parameter estimation set containing P, P_{max}, F and T.
parameter_storage	n $\times$ m $\times$ k	Storage containing estimated parameters for the efficiency curve.
sim_measurements	n $\times$ m $\times$ k	Storage containing ”new” measurements used in simulating the online algorithm.

3.1.2 Initialization

The initialization process includes creating storage and finding the nominal curve, that is, estimating the BSFC curve based on all obtained data points from site using offline modelling. The workflow of the initialization process is illustrated and Figure 3.2 and will be explained further. The first step is to apply the classification method proposed in the M.Sc. pre-project in Espeland (2017), to filter data points measured when a GT is in transient operation or shut off. The classification algorithm also functions as a filter removing fast transients in the data, in contrast to slow transients which is when the GTs change their set

points (loads). The classified data is saved in the global storage ($global\_storage$) and the BSFC values and loads are calculated. Next, the outlier detection algorithm also developed in Espeland (2017) is applied to ensure that all incorrect measurements are removed from the dataset.

As mentioned, the nominal curve is the initial efficiency curve estimated from already received data points, that is all data points from steady-state operation received up until current time. The curve is estimated with the same procedure as in Espeland (2017), by applying least squares regression and has the form of a power series, see Section 2.3. The parameters describing the nominal curve, $\hat{\boldsymbol{\theta}}_{nominal,nls} = [a, b, c]^T$, are calculated from the estimated $\hat{\boldsymbol{\theta}}_{\boldsymbol{ls}}$ and saved in $parameter\_storage$ with the initial time stamp, T_0. This process is repeated for all k GT, and finally, the initial working set of *size_ws* data points from steady-state operation is created, the working set is updated online and is discussed in the next section.

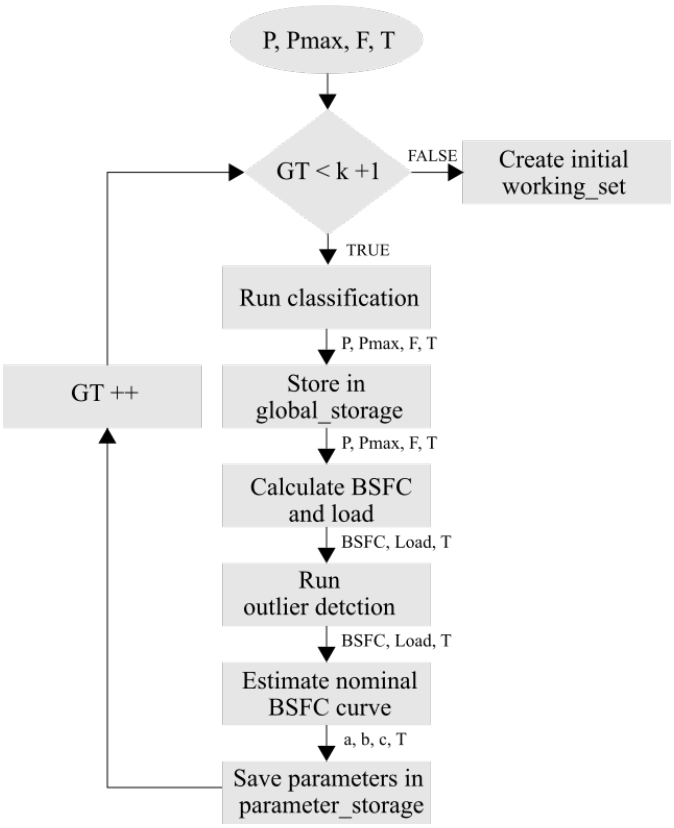

Figure 3.2: Illustration of the work flow in the initialization process of Algorithm A with the matrices P, P_{max} and F, and the vector T as input.

3.1.3 Updating the working set

A working set is created to contain data points to be included in the online update of the BSFC curve. Thus the working set will contain a finite set of data points consisting of P, P_{max}, F and T and are inserted when received. The working set has a fixed length $size\_ws$, and is divided into five intervals. Interval 1 contains high loads, intervals 2-4 contain lower loads decreasing with the interval number and interval 5 contains the lowest loads. This is to ensure that measurements from the entire load interval (0-100%) are represented in the working set, such that the final estimated curve describes the entire interval. Each load interval in the working set is maintained as a FIFO queue. That is, when a new data point is received, the oldest data point is removed, and all the remaining data points are "pushed down" one row. The new data point is then inserted in the first row in the working set in the respective interval. It should also be mentioned that the intervals containing the measurements with higher loads are larger, because these measurements are more relevant when the GTs are in steady-state operation.

When a new data point is received, the operational state is first found by applying the classification method. If the GT is classified as being in steady-state operation based on the current data point, the data point is inserted in the working set in the correct load interval. However, if the GT is in transient operation the data point is only inserted if the load is below $low\_load\_threshold$. This is because the measurements with lower loads are in transient operation, and it is essential not to combine these measurements.

An example of this process is illustrated in Figure 3.3. Assume that the following

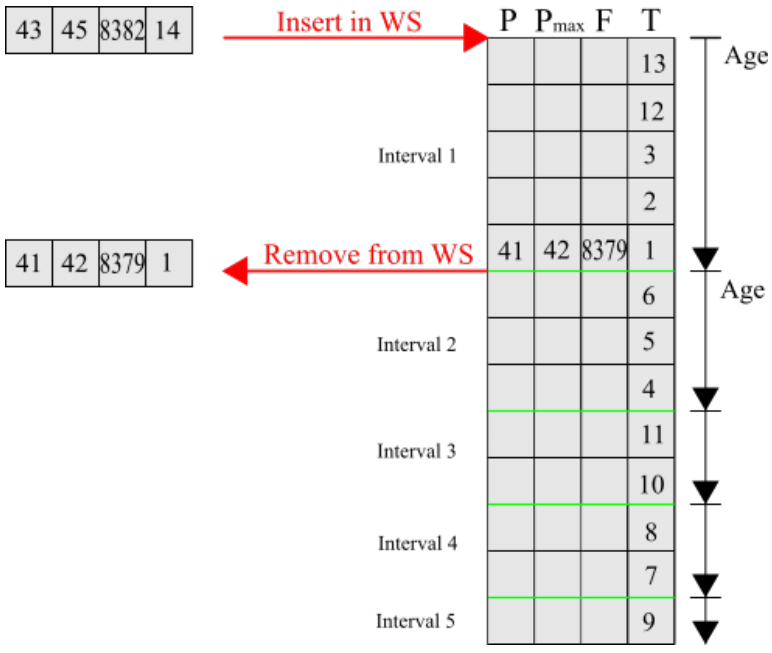

Figure 3.3: Illustration of the structure of the working set and the data point insertion process.

data point is received, $[P, P_{max}, F, T] = [43, 45, 8382, 14]$, where $T = 14$ illustrates the current time. The data point is first classified as being measured when the GT was in steady-state operation by applying the classification algorithm developed in Espeland (2017). The boundary between interval 1 and 2 is in this example set to be 95%; consequently, the new data point should be inserted in interval 1 because the load is calculated to be:

$$load = \frac{P}{P_{max}} 100 = \frac{43}{45} 100 = 96\%. \tag{3.1}$$

Following the FIFO logic, the oldest data point in interval 1,
$[P, P_{max}, F, I] = [41, 42, 8379, 1]$, is removed and the new data point inserted. The rows in each interval in the working set are thus automatically sorted in descending order regarding time. Hence, it is easy to update the working set when new data points are received.

3.1.4 Curve adaptation

After the working set is updated with new data points, it is used in the curve adaptation. The process in addition to system blocks directly influencing the curve adaption is illustrated in Figure 3.4. First, the BSFC values and loads are calculated based on the data points in the working set, followed by outlier detection. This ensures that the curve is estimated based on valid measurements, and reduces the high impact an outlier has on the final curve when estimating with LSR, this is described in Chapter 1.5 in Hansen et al. (2013).

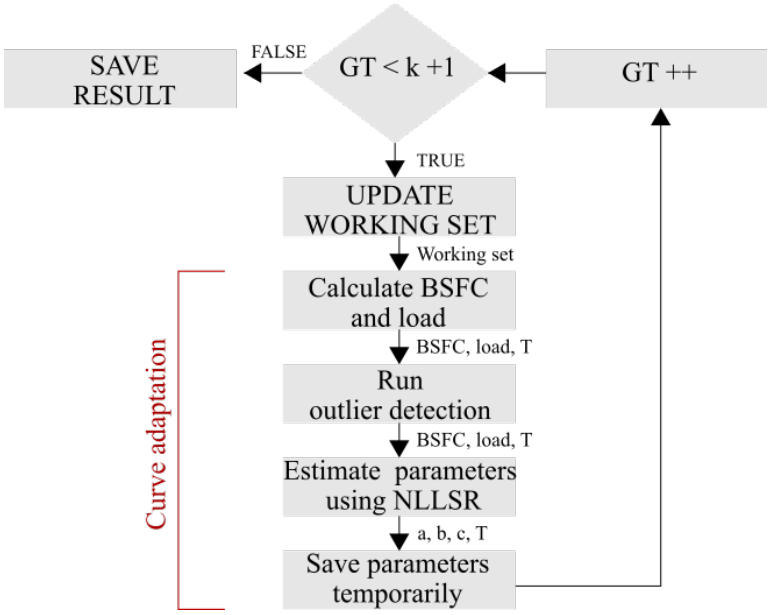

Figure 3.4: Illustration of the work flow in the curve adaptation process in Algorithm A with working set as input and $\hat{\boldsymbol{\theta}}_{\boldsymbol{nls}} = [a, b, c]^T$ and T as output.

Further, the parameters of the new BSFC curve $\hat{\boldsymbol{\theta}}_{\boldsymbol{ls}} = [\theta_1, \theta_2, \theta_3]^T$ are calculated using NLLSR, and transformed back to the non-linear system $\hat{\boldsymbol{\theta}}_{\boldsymbol{nls}} = [a, b, c]^T$. The estimation process is repeated in a time interval for all GTs, using a new window of measurements (working set) in each update. Hence this is a MWLS approach with $s = update\_time$ and $w = size\_ws$. That is, NLLSR is used every s received measurement, including w data points in the working set. The final result of this step is a matrix containing the estimated parameters ($\hat{\boldsymbol{\theta}}_{\boldsymbol{nls}}$) and the update time T.

An alternative method for curve adaptation is using a counting variable, the variable decreases in value when the error from the current curve relative to newly received measurements increases. Thus, the idea is to constantly check this variable and update the BSFC curve if the value is lower than a threshold describing maximum allowed error. For instance this threshold could be chosen to be 0.95, where the closer the value is to one the better the estimation. The error is calculated using R-squared as followed

$$error = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{3.2}$$

where $\hat{y_i}$ is the estimated value of y_i and $\bar{y}$ is the average of the observed data:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i. \tag{3.3}$$

Here, n is the number of observations in both Equation (3.2) and (3.3). The update of the curve is the same as the previous method, that is, using NLLSR, and the resulting parameters $\hat{\boldsymbol{\theta}}_{\boldsymbol{nls}}$ are passed on to the next step. However, because of the fast update time in the previous update method, this method was discarded.

3.1.5 Save result

The matrix with the new parameters is saved in $parameter\_storage$ with the previously estimated parameters. Further, the parameters are passed on to the load sharing algorithm calculating new and optimal set points for the GTs. This process is then repeated for all new measurements.

3.2 Algorithm B - Curve estimation using weighted data points

Algorithm B also incorporates temperature, load and time when estimating the BSFC curve. That is, the current temperature, load and date is important in deciding which data points that should be included in the estimate of the curve. The algorithm consists of offline- and online modelling which can be divided in interdependent subparts and are illustrated in a flow chart in Figure 3.5. The most important systemparts will further be described in the following sections after an overview of the parameters included in the algorithm.

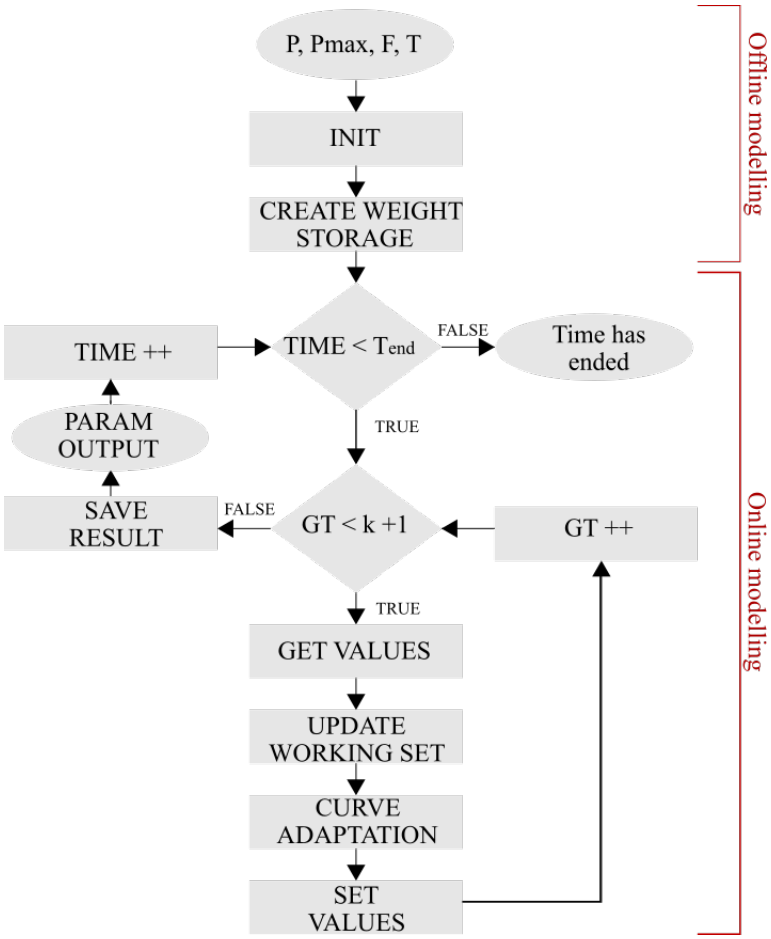

Figure 3.5: Illustration of the steps in Algorithm B.

3.2.1 Parameters

An overview of the parameters included in Algorithm B and a description can be found in Table 3.3. The size of each parameter is also included in the table where v is a number, and the remaining sizes are either matrices or vectors given by the number of measurements n, number of different measurement m and k is the number of turbines or temperature intervals.

Table 3.3: Parameter specification for Algorithm B.

Name	Size	Description
update_time_curve	v	Update time for the efficiency curve.
num_in_es	v	Number of measurements in estimation set used in modelling of new curve.
num_turbines	v	Number of generating machines
max_length_all	v	Maximum length of the storage containing all measurements.
init_measurements	n × m × k	Measurements (P, P_{max}, F, T) already obtained and used in modelling of nominal curve.
classified_init_set	n × m × k	Initial measurements (P, P_{max}, F, T) after classifying steady-state operation.
parameter_storage	n × m × k	Storage containing estimated parameters for the efficiency curve.
working_set	n × m × k	Storage containing all values (load, BSFC, T, index) used in online modelling.
sim_measurements	n × m × k	Storage containing ”new measurements” (P, P_{max}, F, T) used in the online algorithm.
steadystate_counter	k × 1	Counting steady-state measurements for all generating machines until update time is reached.
temp_storage	n × m × k	A storage for each generating machine, containing indexes arranged based on time, temperature and load.
sim_working_set	n × m × k	Storage containing values (load, BSFC, T, index) for simulation.
global_storage	n × m × k	init_measurements + sim_measurements, updated online.
estimation_set	n × m	num_in_buffer measurements obtained from temperature storage and used when updating the efficiency curve.

3.2.2 Initialization

The initialization process is similar to Algorithm A, where a storage containing data points with the initial measurements, *i.e.*, P, P_{max}, F and T is created. Next, the classification algorithm described in Espeland (2017) is used to filter the data points and exclude both slow and fast transients. The classified measurements are stored in a global storage ($global\_storage$) with fixed length, and are updated when new measurements are received using a FIFO logic. Further, the BSFC values and loads for each data point is calculated and outlier detection is applied.

Next, the nominal curve is found based on the pre-processed data points. It should be mentioned that the temperature measurements are not included in the offline modelling. This is because the nominal curve is based on all already obtained data points, not just a set of data points. Hence there are a lot of variation in the temperature in this interval. The parameters describing the nominal curve, $\hat{\boldsymbol{\theta}}_{nominal,nls} = [a, b, c]^T$, are saved with the initial time, T_0, in the parameter storage ($parameter\_storage$).

Finally, the initial working set of the algorithm is created. Each data point in the set contains load, BSFC value, time and index. The index is used to quickly obtain the desired data points in the weight storage described in the next section.

3.2.3 Create weight storage

A weight storage with finite size is created, and it contains data points where the placement indicates the weight. The placement of each data point is found based on the temperature and time when the data point was measured, in addition to the load value. The temperature measurements are obtained from the Norwegian Meteorological Institute and are the mean values for each hour.

The weight storage is arranged in a $n \times m \times k$ - matrix where n is the number of possible data points, m is the number of load intervals and k is the number of temperature intervals, see Figure 3.6 for an illustration. In the illustration, there are ten load intervals in the horizontal direction. The vertical direction represents time with older data points further down; hence, a data point further up in the storage is more relevant for the current condition. The direction into the paper is the temperature intervals, where the intervals extend from the lowest to the highest temperature that is chosen to be included in the system.

The storage is divided into load intervals to make sure that all loads, if existent, are represented in the parameter estimation process. This is to obtain a curve describing the BSFC for the entire load interval (0% - 100%). By also arranging the storage in a third dimension regarding temperature, measurements with similar temperatures and loads can be obtained quickly in the estimation process ensuring similar conditions of the included data points.

An initial weight storage is constructed for all GTs before the online estimation. That is, the storage is filled with data points measured when the GTs are in steady-state operation in addition to measurements with lower loads (below 40%). Each data point is investigated and inserted in the correct place, visualize inserting data points in the illustration in Figure 3.6 and gradually filling the storage. As mentioned the storage has a fixed length and depending on the number of data points obtained offline, the storage might not

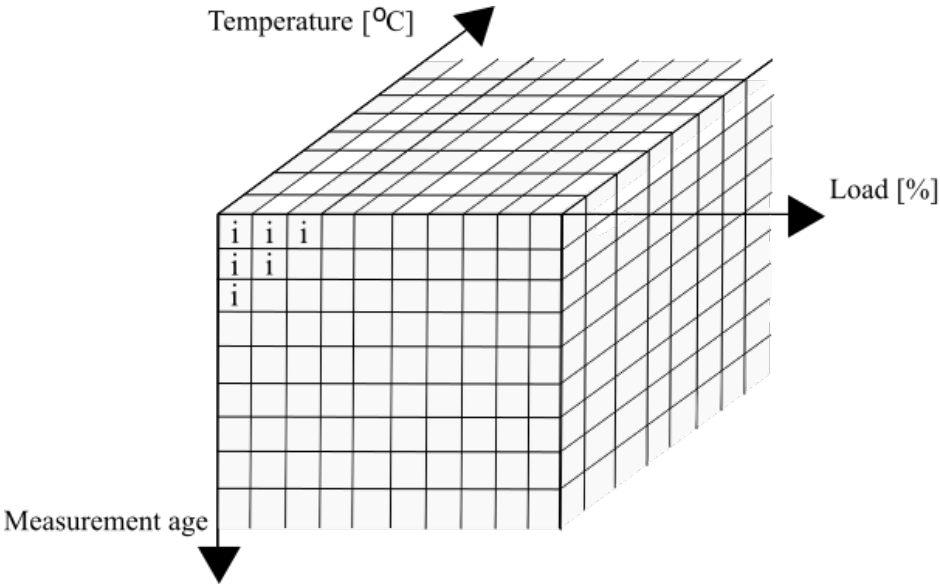

Figure 3.6: Illustration of the weight storage containing the indices for each data point arranged according to load, temperature and age.

be completely filled; hence the unfilled elements contain a -1. The size of the load intervals is an important parameter because the smaller the load intervals are, the more precise is the estimation, however the danger of modelling noise increases.

3.2.4 Get values

The algorithm is iterative, and the processes in the following sections are repeated for all GTs. Hence, this step is developed to retrieve all required datasets for each GT. That is, retrieving the current working set, weight storage and global storage for all GTs. This ensures that data points from different GTs are separated and a higher degree of algorithm control is obtained.

3.2.5 Update working set and weight storage

Following the previous step, the operational state is found by applying the same classification algorithm used during initialization, using the 150 previous obtained measurements. Recall that the three available operational states are steady-state operation, transient operation and shut off, see Section 2.2.1. If the GT is in steady-state operation the load and BSFC value are calculated from the current measurement and inserted in the working set including the time stamp T and index I, and the raw measurements are inserted in the main storage, that is $global\_storage$.

Next, the index of the current measurement is inserted in the weight storage according to load and the specific temperature at the time the data point was measured. As mentioned

in Section 3.2.3, the placement of the data point index reflects the weight of the measurements. Hence, new data points are inserted in the weight storage with a FIFO logic, and this ensures that the most recent data points are placed further to the top in the matrix. That is, the most recent data point is placed according to $matrix\ index = (1, y, z)$ where y is the load interval and z is the temperature. The storage has a fixed length; hence the oldest measurements will be removed from the storage when new measurements are obtained and the remaining elements "pushed down" one row. Finally, a steady-state counter is incremented to ensure steady-state operation when updating the curve, more in Section 3.2.6.

However, if the GT is in transient operation, the insertion process is slightly different. Recall that a GT should be in steady-state operation when the BSFC curve is updated (Section 2.1.3). However, data points from the entire load interval must be represented; hence data points obtained from transient operation with $load < 40\%$ are also inserted in the weight storage. Naturally, the steady-state counter is not incremented, hence there will be no curve updates when the current operational state is transient operation. This ensures that data points above 40% obtained from transient response is not included in the estimation and can thus not interfere with the valid data points.

3.2.6 Curve adaptation

The curve is updated online regularly, for example, every 15 minutes depending on the preset $update\_time$. Since the GTs can operate in one of three states, there exist three scenarios when the curve should be updated. The first scenario is if the GT is in steady-state operation, the steady-state counter is first checked to be above a threshold that indicates that the machine has been in steady-state operation for a specific time. This to follow the steady-state condition described in Section 2.1.2. If this initial check is validated, an *estimation set* is constructed. The estimation set is first filled with data points with the highest weights within the same load interval measured with the same temperature as the current set point. Next, data points with the highest weight from the other load intervals are inserted, but still, with the same temperature.

For example, if the current load is 94% and the temperature is 4^oC, $\frac{3}{4}$ of the estimation set is filled with the newest data points with loads between 90% and 95% that was measured with an ambient temperature of 4^oC. The remaining of the estimation set is filled with measurements from other load intervals, but still with the same temperature. This ensures that some of the affecting factors are approximately the same for the data points in the estimation set.

However, if there are no data points in a given load interval at a specific temperature, the algorithm explores measurements with slightly higher or lower temperatures. How many data points that should be obtained from the weight storage in each load interval is set by the end user of the system

When the estimation set is filled, the new parameters are estimated. The estimated parameters, $\hat{\boldsymbol{\theta}}_{\boldsymbol{ls}}$, are found using NLLSR with the MWLS as in Algorithm A, and transformed back to the non-linear system $\hat{\boldsymbol{\theta}}_{\boldsymbol{nls}} = [a, b, c]^T$. Recall that the efficiency model is a power series:

$$BSFC(load) = a(load)^b + c. \tag{3.4}$$

The estimated $\hat{\boldsymbol{\theta}}_{\boldsymbol{nls}}$ from k GTs are saved in a $1 \times 4 \times k - matrix$ with the update time T. Lastly, the steady-state counter is set to zero.

The second scenario is if the GT is in transient operation or the steady-state counter is below the update threshold (10 minutes of steady-state operation), the parameters are set to be equal to the previous update. This is because the algorithm uses the current data point to create the estimation set, and if this point is obtained from transient operation, an update of the parameters can cause the BSFC curve to be invalid. The third scenario is if the GT is shut off, the parameters are set to NaN (Not a Number) to easily be recognized in the load sharing algorithm.

The estimation of new parameters are relatively fast; hence this step is performed in a time interval. This ensures that the curve always is based on the latest measurements with similar conditions. The flowchart in Figure 3.7 summarize the curve adaptation step with the three updating scenarios due to operational state.

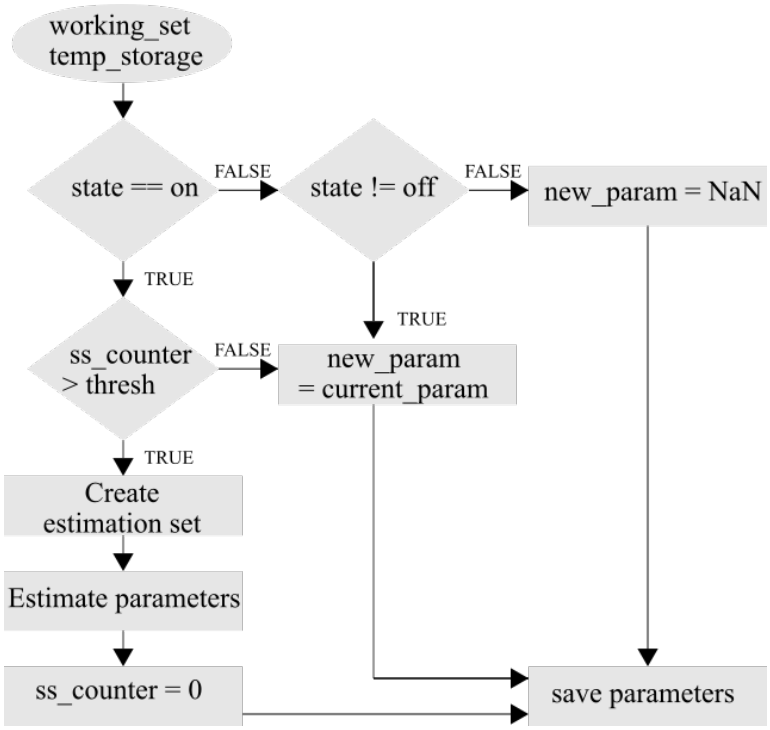

Figure 3.7: Illustration of the curve adaptation block.

3.2.7 Set values

The final step in the online algorithm is to update the storage for each GT, the same storage that was retrieved in Section 3.2.4, that is the updated working set, weight storage and global storage. The update is performed after the new parameters are estimated and before

the update process for the next GT. When all GTs have been through the updating process, the matrix containing the new parameters are passed on to the optimization problem calculating new set points for the GTs.

3.3 Additional algorithms

To investigate the effect of including highly inefficient data points in the curve estimation, an alternative algorithm (Algorithm C) excluding data points with $load < 40\%$ was developed. In addition, an algorithm using offline parameter estimation to find the BSFC curve based on different system responses (steady-state or transient operation) independently. This is referred to as piece-wise curve fitting and is denoted Algorithm D. Having said that, both algorithms require more work to interact with the load sharing algorithm. That is, in Algorithm C the resulting curve only describe BSFC values for loads above 40% and the curve in Algorithm D is not smooth.

3.3.1 Algorithm C - Polynomial estimation for loads $> 40\%$

An algorithm using the same principle as in Algorithm A was implemented. However, the algorithm only includes measurements with $load > 40\%$, to investigate the variations in the estimated parameters when the most inefficient measurements are excluded. A second order polynomial was applied to model the BSFC for these loads and are on the following form:

$$y = p_1 x^2 + p_2 x + p_3, \tag{3.5}$$

where y is BSFC and x is the load. The system can be written as a linear parametric model as in Equation (2.6), where the parameters to be estimated can be written as $\hat{\boldsymbol{\theta}} = [p_1, p_2, p_3]^T$ and no parameter transformation is needed, the regressor can be written as $\boldsymbol{\phi}^{\boldsymbol{T}} = [x^2, x, 1]$. Since $\hat{\boldsymbol{\theta}}$ is linear in the parameters, LLSR can be applied to find the optimal $\hat{\boldsymbol{\theta}}$ that minimizes the error between the estimated and real system, as described in Section 2.3.2.

The parameters included in Algorithm C are similar to the parameters in Algorithm A and are presented in Table 3.4. Here, v is a number describing the one-dimensional size of the parameter, and the remaining parameters are matrices with n being the number of data

Table 3.4: Parameters specification for Algorithm C.

Name	Size	Description
update_time_curve	v	Update time for the efficiency curve.
size_ws	v	Size of working set.
num_turbines	v	Number of generating machines
size_global_storage	v	Maximum length of the storage containing all measurements.
ws_threshold	v	Only measurements below this threshold are inserted in the working set when GT is in transient operation.
global_storage	n × m × k	Measurement storage containing P, P_{max}, F and T.
working_set	n × m × k	Parameter estimation set containing P, P_{max}, F and T.
parameter_storage	n × m × k	Storage containing estimated parameters for the efficiency curve.

points, m the different types of measurements and k is the number of turbines. A short description of the parameters is also included in the table.

The flowchart of the algorithm is the same as in Algorithm A and presented in Figure 3.1, but the content of the system blocks are slightly different.

Initialization

The system is initialized by excluding data points with $load < 40\%$ from the dataset obtained offline. Next, the data points containing P, P_{max}, F and T are filtered with the classification algorithm to remove fast and slow transients in the data. Further, the BSFC values and loads are calculated and outliers found using the developed outlier detection algorithm from the M.Sc. pre-project. When all incorrect data points are removed, the nominal curve is found using LLSR to estimate a second order polynomial, see Equation (3.5), with parameters $\hat{\boldsymbol{\theta}}$. Initial working sets for all GTs with length $size\_ws$ are also created in this step.

Update working set

Unlike algorithm A, the working set is divided into two equal parts. That is, each part of the working set contains the equal number of data points with $ws\_threshold$ dividing the parts, see illustration in Figure 3.8. This is to spread the data points over the load interval between 40% and 100% to avoid only data points with for example $loads > 95\%$.

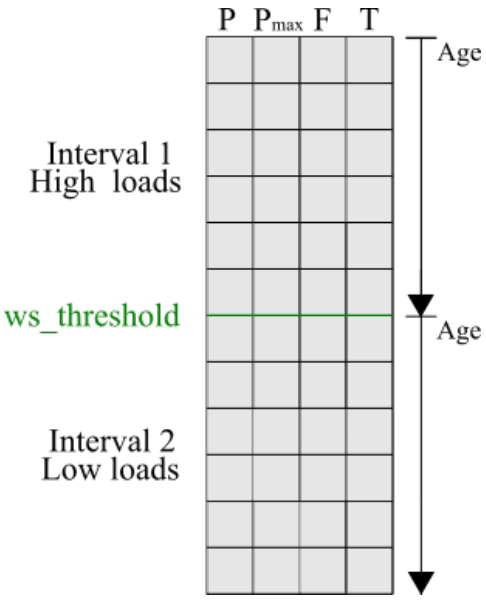

Figure 3.8: Illustration of the working set used in Algorithm C. The working set is divided in two intervals with $ws\_threshold$ being the boundary between the intervals.

When a new data point is received, the classification algorithm is applied to check if the GT is in steady-state operation. If it is, the data point is inserted in the correct interval in the working set given that the load is calculated to be above 40 %. FIFO logic is used to obtain the same length of the working set at all times. That is, the newest obtained data points are placed furthest up in the respective interval. The data points already in the working set are moved down one row, and the data point furthest down in the interval is removed. This is similar to the insertion procedure in Algorithm A, except that the working set in this algorithm only contains two intervals. This process ensures that the current working set for all GTs always contains the newest data points from measurements with $load > 40\%$.

Curve adaptation

The next step in the algorithm is to calculate the BSFC values and load for all data points in the working set for each GT. Further, the outlier detection algorithm is applied to remove incorrect measurements as in the previous algorithms. The curve is updated according to time intervals of size $update\_time$, using LLSR with MWLS to find the new parameters $\hat{\boldsymbol{\theta}} = [p_1, p_2, p_3]^T$. The new parameters are found iteratively for all GTs and passed on to the next step as a $1 \times 4 \times num\_turbines$-matrix containing $\hat{\boldsymbol{\theta}}$ and the current update time.

Saving results

The last step is saving the results in $parameter\_storage$, which is the input of the load sharing algorithm.

3.3.2 Algorithm D - Piece-wise curve estimation

An algorithm using least squares regression with two estimation models was developed. Recall that the GTs can operate in one of three states, steady-state, transient or being shut off. This will affect the efficiency of the GTs and can be observed when plotting the BSFC values. That is, when the GT is in transient response the operation is more inefficient. Hence, the idea is to estimate separate curves describing the different responses and finally connect the curves.

Offline modelling

By studying the active power, fuel consumption and BSFC in Figure 3.9, one can observe two responses in the data. The yellow color indicates more or less steady-state operation, while the blue and orange color indicate transient response. Comparing the BSFC plot with active power and fuel consumption, it can be seen that the data points in interval 1 account for the lower values in P and F. One should notice that when the GT is shutting off, around January 20, the load is still high, that is above 70%. Hence, it is not enough to exclude measurements with lower loads, the classification algorithm must be applied to identify these types measurements.

It can be argued from Figure 3.9 that there are different operational states in the dataset. Thus, the dataset should be divided into separate intervals, such that the estimated curve describes a specific operational state or system behavior. It can be suggested to divide the

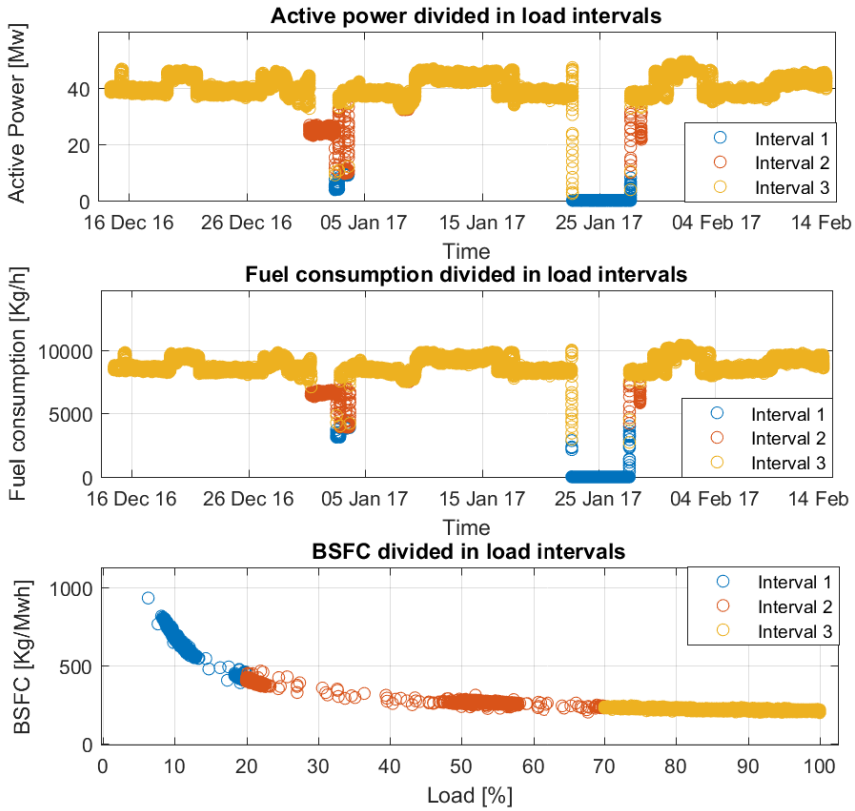

Figure 3.9: Different GT responses in the dataset from January and February indicated with color.

dataset following the intervals in Figure 3.9, then the data points obtained from interval 3 can be estimated with a second order polynomial as in Algorithm C. The measurements from transient operation, interval 1 and 2, are much steeper; hence power series are needed in both cases. The separate models can be written as follows

$$BSFC(x) = a_1 x^{b_1} + c_1, \quad l_1 < x < l_2 \tag{3.6}$$

$$BSFC(x) = a_2 x^{b_2} + c_2, \quad l_2 < x < l_3 \tag{3.7}$$

$$BSFC(x) = p_1 x^2 + p_2 x + p_3, \quad l_3 < x < l_4, \tag{3.8}$$

where l_1, l_2, l_3 and l_4 are the boundaries of the load intervals. The parameters $\hat{\boldsymbol{\theta}}_{\boldsymbol{nls,1}} = [a_1, b_1, c_1]^T$, $\hat{\boldsymbol{\theta}}_{\boldsymbol{nls,2}} = [a_2, b_2, c_2]^T$ and $\hat{\boldsymbol{\theta}}_{\boldsymbol{3}} = [p_1, p_2, p_3]^T$ are to be estimated. When the three separated curves are found, the curves are *merged* together resulting in one curve.

Since this algorithm at the current time is an offline approach, steps like storage handling, online estimation and saving of results are not included.

Chapter 4

Simulation result

The four developed algorithms were simulated using two datasets. One containing measurements from an eight month period obtained offline from site, denoted the *initial dataset*. The other includes new data points and is denoted the *simulation dataset*. Figure 4.1 presents the loads in the GTs throughout the simulation interval to easier discuss the findings in results presented in later sections. It is important to notice that GT4 in average runs on a lower loads compared to the remaining GTs.

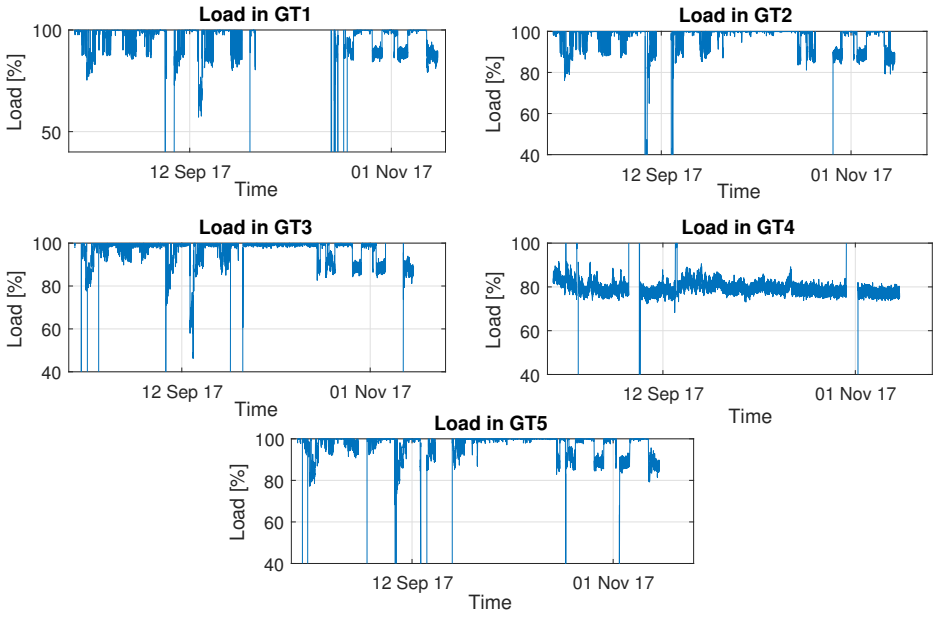

Figure 4.1: Load in all GTs over the simulation interval.

New measurements are received every 10 seconds, and the curve is updated every 30 minutes. The updated parameters are then passed on to the load sharing algorithm developed in Jung (2018), where the new set points in the GTs are calculated.

This chapter presents the simulation results in the same order as the algorithms were presented in the previous chapter, with a short summary of the algorithms at the beginning of each section. Important aspects will be commented, but a more thorough discussion of the results and methods are found in Chapter 5.

4.1 Algorithm A

Algorithm A is a method using NLLSR with a moving window to estimate new parameters describing BSFC curves in GTs. A FIFO logic was applied to always ensure that the working set contained the newest measurements. The nominal curve was found by applying classification and outlier detection and using a power series to estimate the BSFC curve based on the initial dataset. The online algorithm was then simulated with the simulation dataset containing data from approximately three months.

The number of data points in the working set is an essential parameter in this algorithm and can affect the BSFC curve significantly. If too few data points are included in the working set there exist a danger of modelling noise, and the curve is updated on false assumptions. In contrast, to include too many data points can cause the algorithm to not respond to changes. Hence, the system was simulated with three sizes of the working set equal to 300, 1000 and 3000 data points. This corresponds to a time frame of approximately 50 minutes, 2.7 and 8.3 hours, respectively. The chosen parameters specifications are presented in Table 4.1, and can easily be changed for other use cases of this algorithm.

Table 4.1: Parameter choices for Algorithm A.

Parameter	**Size**
update_time_curve	30 min
num_in_ws	Variable
num_turbines	5
max_length_all	1 year
max_length_ws	1 year
ss_threshold	10 min

The working set was divided into five load intervals, see Figure 3.3 for a reminder. The size of each interval depends on the size of the working set and is presented in Table 4.2, with the load intervals indicated. One should notice that interval 1 always contains the highest number of data points. This is because the GTs are most likely in this interval when in steady-state operation.

Table 4.2: Number of data points included in each interval in the working set, for simulation 1, 2 and 3.

Sim	Total ws_size	Interval 1 85%-100%	Interval 2 70%-85%	Interval 3 50%-70%	Interval 4 30%-50%	Interval 5 0%-30%
1	3000	1200	700	700	300	100
2	1000	350	250	250	100	50
3	300	100	70	70	40	20

4.1.1 Generating turbine 4 (GT4)

In this section, the simulation results for GT4 is presented. For the sake of space, equivalent curves for the remaining GTs are presented in Appendix A. However, the following section includes results from all GTs regarding estimation error in the algorithm.

Figure 4.2 presents the time change in the estimated parameters transformed back to the non-linear system, $\hat{\boldsymbol{\theta}}_{\boldsymbol{nls}} = [a, b, c]^T$. The non-continues intervals of the curves are due to the GT being shut off. One should notice that the results of simulation 3 seem to be more stable compared to simulation 1 and 2 which both contain small oscillations in the parameters. However, between September 12 and 20 one can observe a substantial change in the parameters from simulation 3. This is due to a sudden increase in the set points in

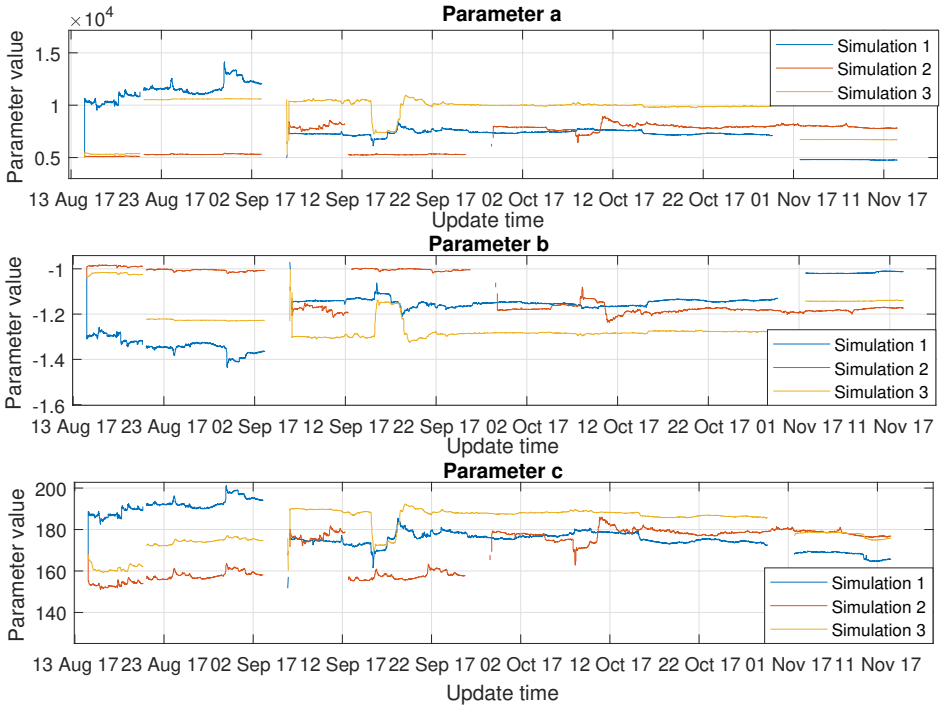

Figure 4.2: Change in the parameters in $BSFC(x) = ax^b + c$ over time for three simulations.

GT4, seen in Figure 4.1, and since the working set contains 3000 data points, it is more likely to include this measurement.

The BSFC curves are found using the estimated parameters and passed on as input to the optimal load sharing algorithm. BSFC curves calculated with parameters from three dates, the initial date, a date in the middle of the simulation interval and the final date is presented in Figure 4.3. The left column of the figure contains the curves for the entire load interval between 0% and 100%. At first glance, these curves look identical for $load > 30\%$; hence the right column includes the curves plotted for loads between 65% and 100% to observe the differences.

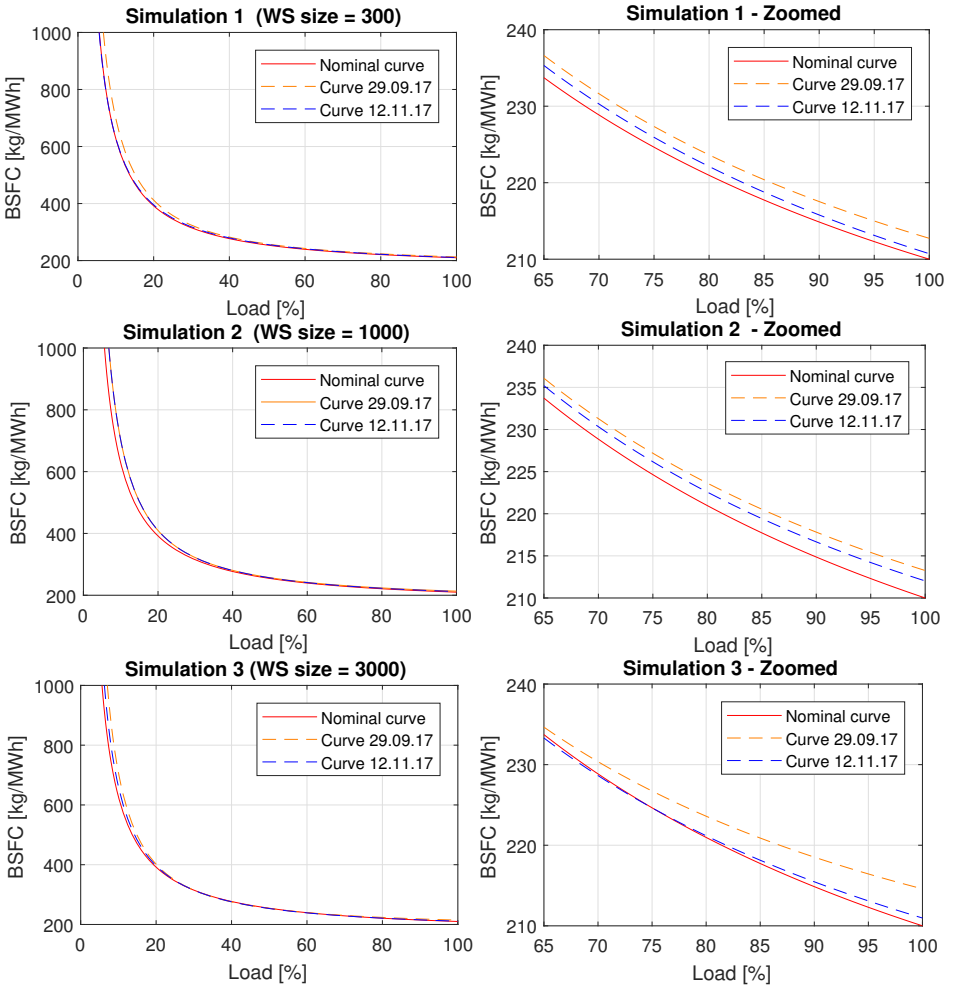

Figure 4.3: The estimated BSFC curve in GT4 from three time stamps for all simulations. To the left the curves are plotted for the entire load interval (0-100%), and to the right between 65% and 100%.

It should be noticed that in Figure 4.3, the curve from September 29 is always furthest up making GT4 being slightly less efficient at that date compared to the other dates. It is also interesting that GT4 is most efficient at the time when the nominal curve was estimated, and the performance of the GT decreases before it slightly increases again at the final date. If the curve from the final date have had been placed furthest up one could have argued that this decrease in performance is due to machine degradation as explained in Section 2.1.2. However, this is not the case. Hence the efficiency must be affected by another factor, for example, temperature.

4.1.2 Estimation error

The error between the measured and the estimated BSFC value was calculated for each curve update and presented in Figure 4.4. In addition, the total percentage error was calculated using Equation (4.1):

$$\%error = 100\% \times \frac{\Sigma_{i=1}^{N} \parallel BSFC_{measured}(l) - BSFC_{estimated}(l) \parallel}{\Sigma_{i=1}^{N} \parallel BSFC_{measured}(l) \parallel}. \tag{4.1}$$

Here, $BSFC_{measured}$ are BSFC values calculated based on the real data points from site, while $BSFC_{estimated}$ are the BSFC values found with the proposed model. Further, N is the number of updates, and l is the load. One should observe that GT5 and GT3 have high percentage errors, especially seen in the results from simulation 2 where GT5 has a total percentage error of 11.69%.

Table 4.3: Total percentage error between measured and estimated BSFC.

Simulation	GT1	GT2	GT3	GT4	GT5
1	1.64 %	1.27 %	0.47 %	0.66 %	9.05 %
2	0.89 %	1.66 %	1.19 %	0.90 %	11.69 %
3	1.00 %	1.52 %	3.92 %	0.63 %	5.54 %

This can be explained by GT5 being turned on and off several times throughout the simulation interval. In addition, the set points are varying much between loads of 80% and 100% compared to, for example, GT4 which is more stable, see Figure 4.1. Hence, it can be argued that many variations in the set points greatly affect the accuracy of Algorithm A. As one can observe in Figure 4.1, GT4 is running on loads around 80% with little variations. Hence, the results are more accurate, and this is validated in Table 4.3 where the percentage error is the overall lowest compared to the other GTs.

It seems like a larger working set is beneficial in this algorithm, it is more stable and not that affected by changes in the GTs. Except from the results from GT3 in simulation 3, where the algorithm performs poorly in the last half of the simulation. A solution to this problem is varying the number of data points included in the estimation. This is, however, more discussed in Chapter 5.

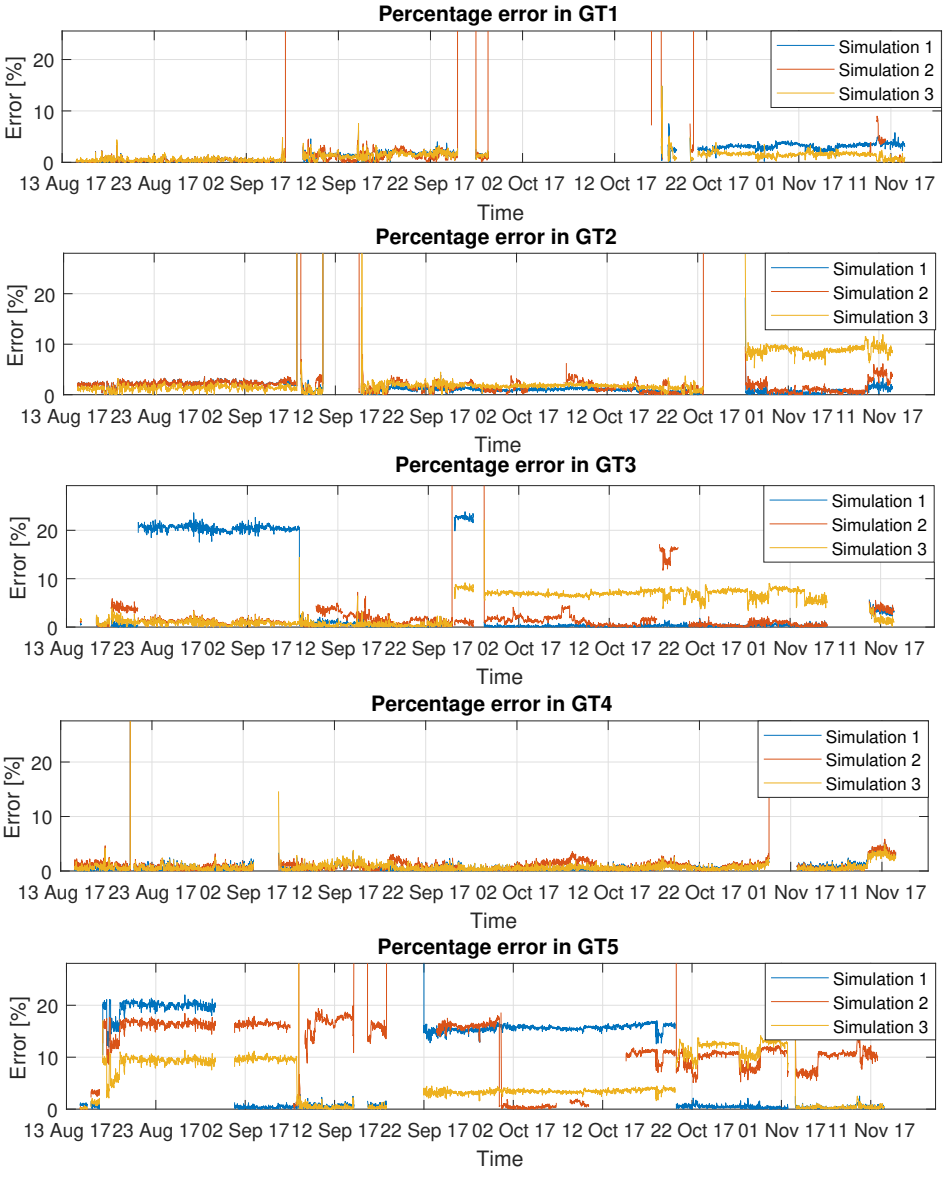

Figure 4.4: Errors between measured and estimated BSFC in the simulations using Algorithm A. The results are plotted for all GTs.

4.2 Algorithm B

The BSFC curves in Algorithm B are based on measurements with similar conditions regarding load, temperature and time. The received measurements was weighted and inserted in the weight storage consisting of 10 load intervals and 50 temperature intervals, that is a $n \times 10 \times 50 - matrix$.

The goal of dividing the load axis in 10 load intervals was to spread the measurements in the estimation set throughout the interval, such that all loads are described in the final BSFC curve. The number of load points (one load point is 1%) included in each load interval, was decided based on the number of measurements in each load point, see Figure 4.5. One can easily observe that the highest number of measurements are when the load is

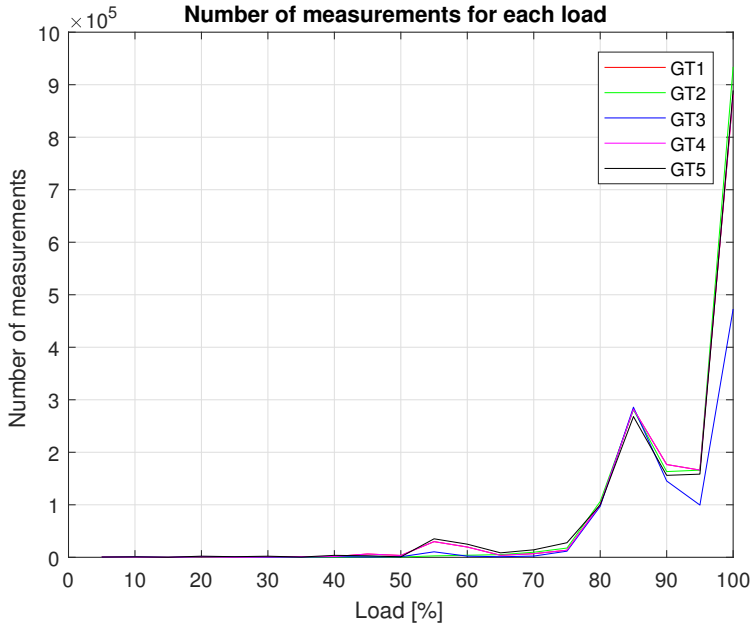

Figure 4.5: Measurement density with different loads, each turbine is indicated with different color.

above 50%. Hence, more load points are required in the low load intervals ($load < 50\%$) to guarantee that these are filled up, for example, could a load interval for lower loads be 0% to 20%. Also, the load intervals containing higher loads must include fewer load points, for example, could one interval contain loads between 95% and 100%. This is because a high accuracy is desired and the GTs are varying a lot more between higher loads. The final choices of the boundaries between the load intervals are illustrated in Figure 4.6.

The estimation set is filled with data points based on the indices in the weight storage, and is naturally also divided in 10 intervals, denoted estimation intervals. The number of measurements included in each estimation interval is varying according to current set point. For instance if the set point is 76%, there are more measurements included in

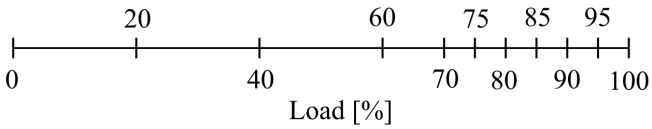

Figure 4.6: Chosen sizes of load intervals in weight storage. The intervals containing lower loads includes more load points compared to the intervals containing higher loads.

this interval (75%-80%) and neighboring intervals (70%-75% and 80%-85%) than the remaining estimation intervals.

The weight storage is further divided into 50 intervals in the third dimension, where each interval is a specific temperature, for example, 8^oC. The number of temperature intervals can easily be chosen and should reflect the possible temperatures at the site where the system is installed. The measurements in this master thesis are obtained from a location with large variations in the temperature during the year. Hence, many temperature intervals were required.

An overview of the remaining parameter choices are presented in Table 4.4, where the size of the working set, $size\_ws$, is changed in every simulation. That is, $size\_ws$ is equal to 300, 1000 and 3000 in simulation 1, 2 and 3, respectively. The remaining parameters presented in Table 3.3 are kept constant during the simulations and are chosen due to available measurements, where n is varying, $m = 4$ and $k = num\_turbines$.

Table 4.4: Parameter choices for Algorithm B.

Parameter	**Size**
update_time_curve	Variable
size_ws	Variable
num_turbines	5
max_length_all	1 year
max_length_ws	1 year
ss_threshold	10 min

4.2.1 Temperature data

Algorithm B is strongly based on choosing data points with the same conditions, *i.e.*, same temperature and load. As explained in Section 2.1.2, there exist a correlation between the ambient temperature and the efficiency of a GT. Several other factors can affect the efficiency, but since this algorithm is developed with a data-driven approach based on as few measurements as possible, such factors will not be considered.

The average temperature each hour was plotted with the average BSFC per day for easier readability, and the values were normalized by applying the following equation

$$A = \frac{A - min(A)}{max(A) - min(A)}, \tag{4.2}$$

where A are the input values. The result for GT2 and GT4 over a period of approximately one month can be seen in Figure 4.7, and one can observe a trend that when the temperature increases the efficiency decreases, that is the efficiency of the GTs is higher for lower temperatures, this is in line with the theory described earlier in Section 2.1.2. However, this is not always true, for example for GT2 around March 18, one can observe that the temperature is increasing, but the BSFC is decreasing. The explanation for this can be that the GT is drastically changing its set points, uncertainty in measurements or other ambient conditions affecting the efficiency.

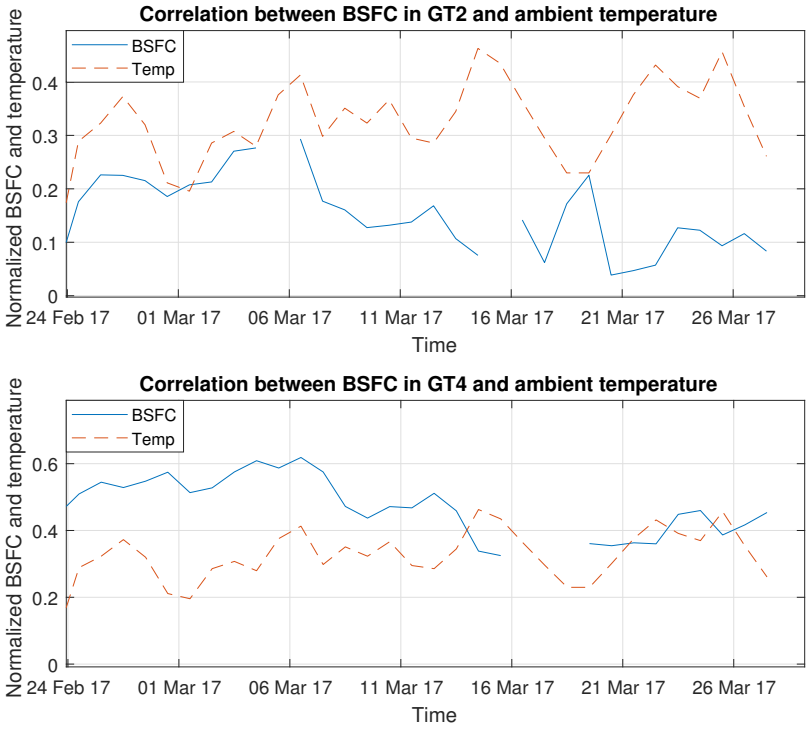

Figure 4.7: Correlation between BSFC and ambient temperature using the normalized BSFC values from GT2 and GT5.

4.2.2 Generating turbine 4 (GT4)

The simulation result of GT4 is presented in this section, the simulation results for the remaining GTs are presented in Appendix B for the sake of space. Figure 4.2 presents the estimated parameters using Algorithm B. One should observe that the parameters are varying in a square shape and change quite often compared to the results in Algorithm A. This is because of the temperature variations, if the temperature increases with one degree, it is a different set of measurements being the basis for the simulation. In addition, if the

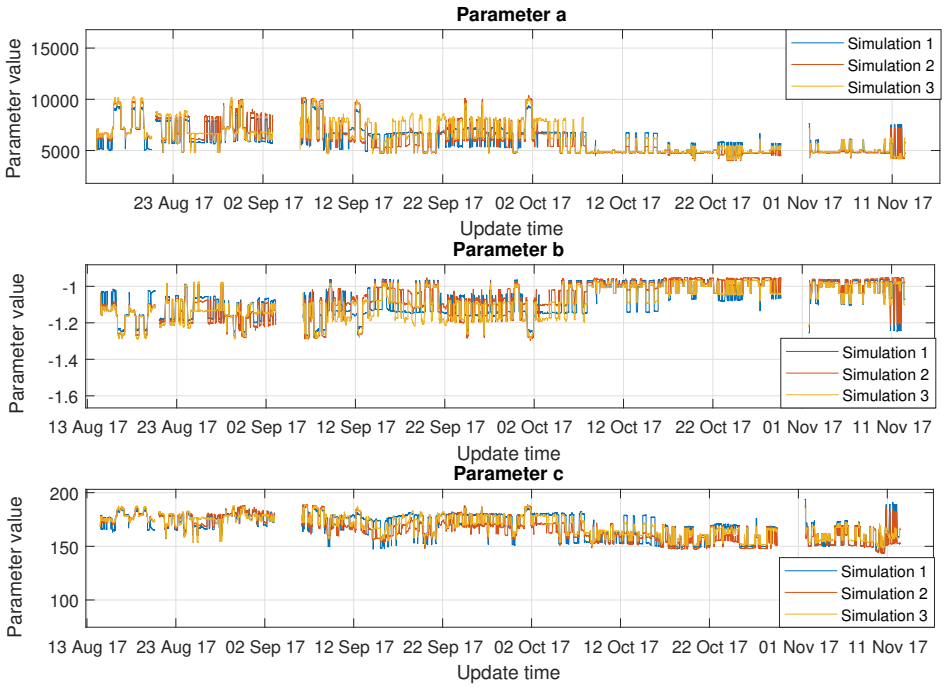

Figure 4.8: Estimated parameters using Algorithm B over time, the three simulations are indicated with color.

temperature is stable, for example, for two hours, many of the same measurements are included in the estimation, hence the flat areas in the parameters.

It should also be commented that the variation in parameters is most significant in the results from simulation 1. That is, the parameters are more stable when simulated with a larger working set. The reason for this is that the working set in simulation 1 includes 300 measurements, which correspond to approximately 50 minutes. Hence, if the temperature is varying with one degree the next hour, it is an entirely different set of measurements included in the update. In contrast, with a working set containing 3000 measurements, corresponding to approximately 8 hours. Consequently, a change in the temperature over one hour will not affect the parameters significantly, since many of the previous measurements are most likely included.

The reader should notice that there are no substantial irregularities in the data, the parameters are oscillating in the same area and form throughout the entire simulation interval. That is, there are no ”surprises”. Compared to Figure 4.2 in Algorithm A, where there were large differences between the simulations and sudden ”jumps” in the parameter values.

The BSFC curves were calculated using the estimated parameters and Equation (2.4), the result from the beginning, middle and end of the simulation interval are presented for GT4 in Figure 4.9. As can be observed, there are not vast differences in the BSFC curves,

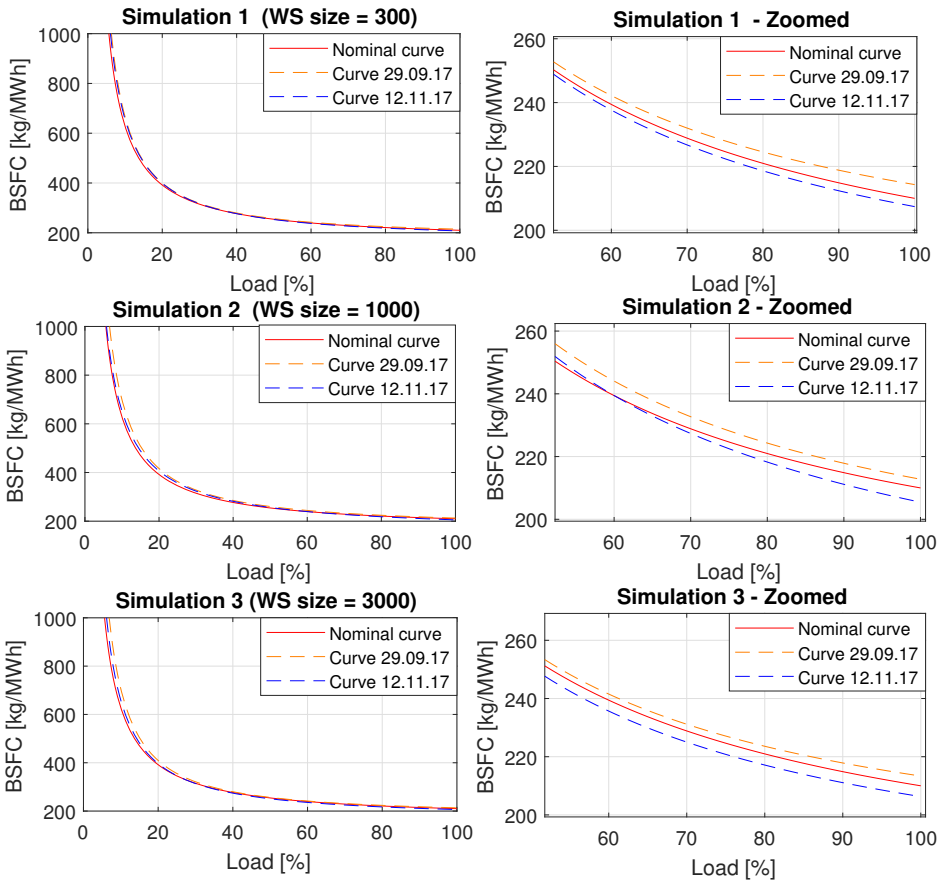

Figure 4.9: The estimated BSFC curve in GT4 using Algorithm B. The result is presented for all three simulations and three dates, that is the beginning, middle and end of the simulation interval.

and this can also be observed in Appendix B where the curves from the remaining GTs are presented. Overall, the algorithm seems to be quite stable and not that affected by noise and measurements from transient operation.

It should be noticed that the curve from November 12 is placed furthest down in all simulations when $load > 70\%$. This indicates that the GT4 is slightly more efficient in November compared to the other dates. This is also the case for the remaining GTs (see Appendix B), which mean that it is an ambient condition causing this, not internal conditions in each GTs. Such an ambient condition might be the air temperature. Referring to Section 2.1.2, it was presented that operation in colder air temperatures can affect the efficiency in a generating unit positively. The temperatures at the dates of the presented BSFC curves are given in Table 4.5. As can be seen in the table, the temperature is lowest in November, so the placement of the BSFC curves is in line with the theory that colder temperatures lead to higher operational efficiency. For example, the average temperature

November 12 is lower than the average temperature September 29; hence the GT should be more efficient in November. It should be mentioned that the results from Algorithm A are contradicting because the curve from November 12 is placed above the nominal curve in Figure 4.3. The nominal curve is the same in both algorithms, hence one of the algorithms is more accurate than the other.

Table 4.5: Temperature and efficiency grading of the presented BSFC curves in Figure 4.9. The grading scale is: *medium - good - best*, hence the curve from November 12 indicate highest operational efficiency.

	Nominal curve (Jan 17 - Aug 17)	**Curve from 29.09.17**	**Curve from 12.11.17**
Temperature	3.1 ^oC	7.4 ^oC	0.1 ^oC
Efficiency	Good	Medium	Best

4.2.3 Estimation error

Figure 4.10 presents the percentage error between the real and estimated BSFC values for all GTs. As can be observed the percentage error in GT3 and GT5 from simulation 1 and partly simulation 2, is much larger compared to the remaining GTs. This was also the case in Algorithm A; however the error was *stable* when using Algorithm A (Figure 4.4), compared to Algorithm B where the error is varying between being very high or very low. Importantly, one should also observe that the errors from simulation 3 are quite low in all GTs. That is, except from a few places around shut on and off.

These errors are also reflected in the total error for GT presented in Table 4.6, which were calculated with Equation (4.1). What is interesting to observe, is that the variations in the set points in GT3 and GT5 do not create large issues for Algorithm B, that is when simulating with $ws\_size = 3000$ (simulation 3). Comparing with simulation 1 and 2 in Figure 4.10 it can be observed large errors, with the maximum percentage error of 3.66% in GT3, simulation 1. However, there are point errors above 20% in some areas not reflected in the total percentage error. Hence, it is clearly seen that a larger working set is beneficial when there are many changes in the set points of the GTs. Also, in a more stable operation simulation 3 performs best, except from in GT4 where the total percentage error from simulation 2 is the minimum error of all results (0.47%).

The reason for the more significant errors in GT3 and GT5 in the results from simulation 1 and 2, is the same as for Algorithm A. Namely, the many changes in GT set points. However, the results from simulation 3 indicate that these changes are captured by

Table 4.6: The total error between the real and estimated BSFC values using Algorithm B.

Simulation	**GT1**	**GT2**	**GT3**	**GT4**	**GT5**
1	1.84 %	1.48 %	3.66 %	0.62 %	1.50 %
2	0.91 %	0.98 %	2.43 %	0.47 %	0.82 %
3	0.86 %	0.81 %	0.68 %	0.57 %	0.60 %

the algorithm, probably because measurements from other dates with similar conditions are included. That is, if the GT switches from running on 80% to 90%, the largest number of measurements included in the working set are measurements with $load = 90\%$. Hence, some new measurements obtained from transient response is included, but the number of measurements from steady-state operation is more significant. Consequently, these changes in set points do not affect the algorithm largely.

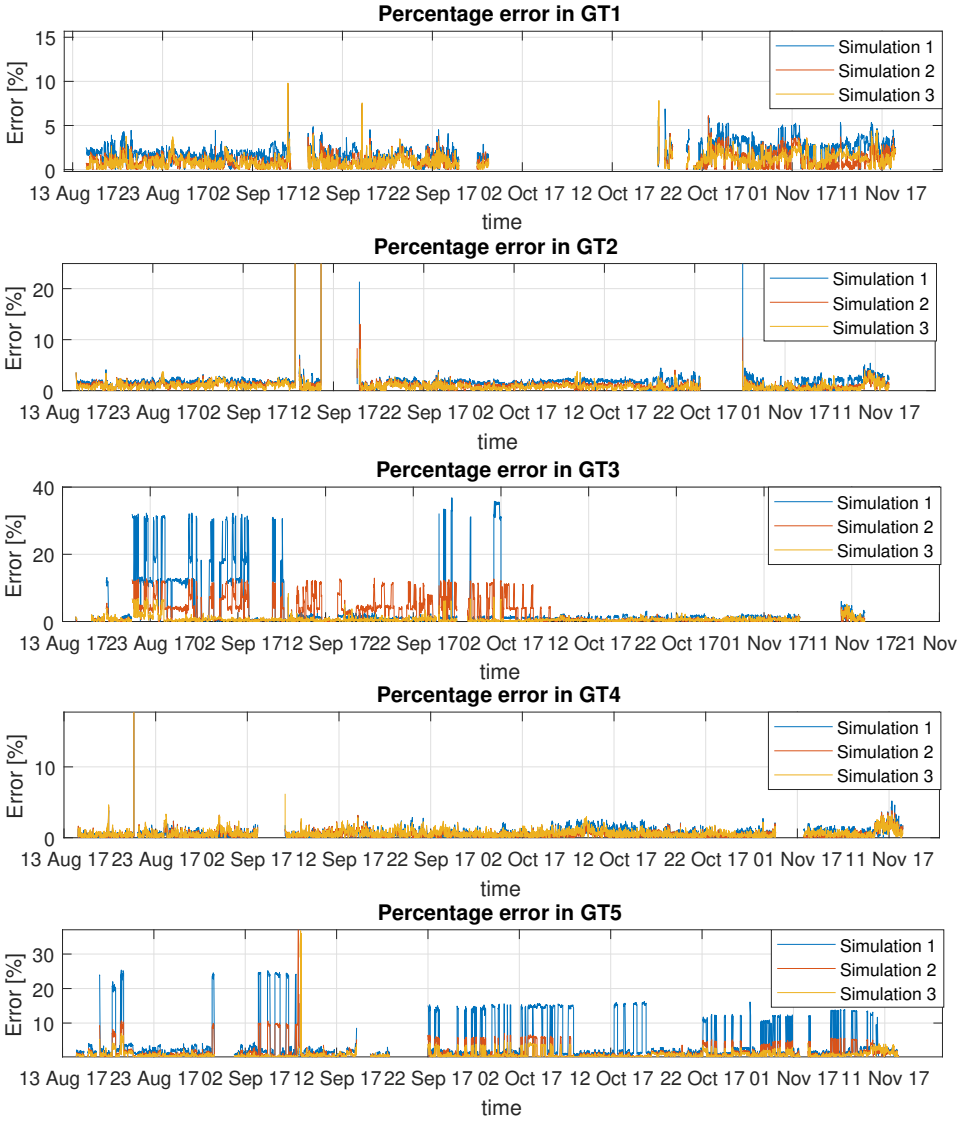

Figure 4.10: The errors between the real and estimated BSFC values for all GTs using Algorithm B.

4.3 Additional scenarios

4.3.1 Algorithm C

Algorithm C estimates BSFC curves with a second order polynomial using LLSR. The estimation is in contrast to the previous algorithms only based on measurements with $load > 40\%$. This is to exclude the ineffective measurements, which can cause more significant errors in the BSFC curve.

The working set was divided into two intervals, where data points with loads above $ws\_threshold$ were inserted in interval 1, and data points with loads below this value were inserted in interval 2. Before the measurements were inserted in the working set, the classification algorithm was applied. Further, the BSFC values and loads were calculated and outliers removed. Finally, the new parameters, $\hat{\boldsymbol{\theta}} = [p_1, p_2, p_3]^T$, were estimated and passed on to the load sharing algorithm.

The algorithm was simulated with the initial dataset to estimate the nominal curve. Further, the simulation dataset with measurements from approximately three months were used in the online simulation, where new measurements were received every 10 seconds. Both datasets included P, P_{max}, F and T. The curve was updated every 30 minutes, and simulated with three sizes of the working set: 300, 1000 and 3000 samples. The parameter specifications applied in the simulations are presented in Table 4.7.

Table 4.7: Parameter choices in Algorithm C.

Parameter	Size
update_time_curve	30 min
num_in_ws	Variable
num_turbines	5
max_length_all	1 year
max_length_ws	1 year
ws_threshold	70 %

Generating turbine 4 (GT4)

The results of the three simulations for GT4 are presented in this section, results from the remaining GTs are presented in Appendix C for the sake of space. Figure 4.11 presents the simulation results for the estimated parameters $\hat{\boldsymbol{\theta}} = [p_1, p_2, p_3]^T$, describing the second order polynomial.

One should observe that the parameters from the different simulations are similar, especially simulation 2 and 3. The parameters have the same fluctuations meaning that the estimation captures the change in the GTs independent of working set size. That said, there are fast oscillations in the parameters, it almost looks like noise. This can indicate that the working set should be chosen to be larger because it would not be ideal to change set points according to these small changes constantly. There are also large peaks in the parameters corresponding with the large deviations in the loads at the same times, see Figure 4.1. Hence, the algorithm is affected by transients even if the most inefficient data

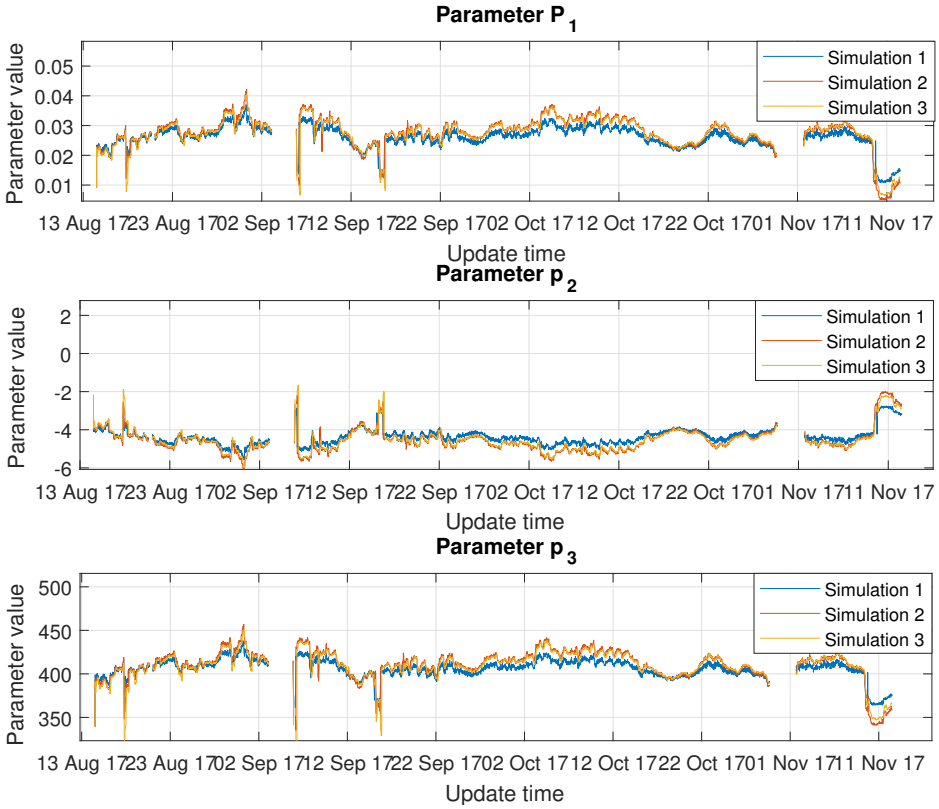

Figure 4.11: Estimated parameters $\hat{\boldsymbol{\theta}} = [p_1, p_2, p_3]^T$ with Algorithm C describing a second order polynomial.The working set contained 300 data points in simulation 1, 1000 in simulation 2 and 3000 in simulation 3.

points are removed.

The BSFC curve at three dates was calculated using the estimated parameters and are presented in Figure 4.12. It should be noticed that the BSFC curve is similar for all simulations, this was however expected since the estimated parameters were very similar for all simulations in Figure 4.11.

A possible issue with Algorithm C can be observed in Figure 4.12. That is the curve from September 29 is curving upwards around 95%. This can be explained by either including too few measurements in the working set or not dividing the working set into enough intervals. Imagine that the GT is running at $load = 80\%$ for a long time. Naturally, interval 1 is filled with measurements with loads approximately equal to 80%. Hence, the working set will not include measurements with higher loads, so when the curve is updated the LSR has no measurements to fit the curve. Consequently, the LSR must keep the natural shape of the second order polynomial in this load area.

As in Algorithm B, the curve from November 12 is placed furthest down indicating that the GT4 has a more efficient operation at this date. The same applies for the remaining

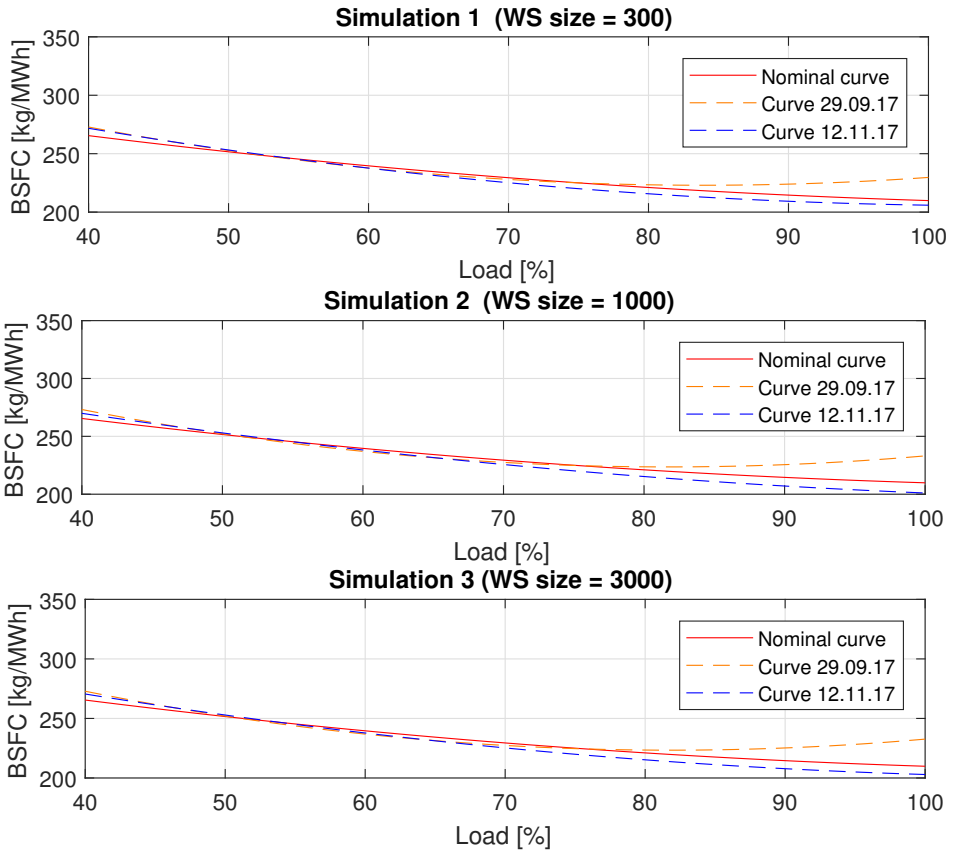

Figure 4.12: The estimated BSFC curve in GT4 from three time stamps using Algorithm C. The result is presented for all simulations with varying size of the working set.

GTs in the results from simulation 3 presented in Appendix C, except from in GT5 where the curve is not accurate for higher loads. The discussion from Section 4.2.2 is not repeated here, but the results from both Algorithm B and C indicate that the air temperature has an affect on the efficiency.

Estimation error

The percentage error between the real and estimated BSFC value is plotted for every parameter update in Figure 4.13, and the total percentage error calculated with Equation (4.1) is presented in Table 4.8. The errors in simulation 3 partly cover the error from simulation 1 and 2, but the error is approximately the same in all simulations. This is validated in Table 4.8, where the total percentage error is largest in the results from simulation 3, except GT3 and GT2 where the error is the same as the other simulations or slightly lower.

Comparing Algorithm A and C, the percentage error from simulation 2 in GT5 is

11.69% using Algorithm A, and 0.29% when using Algorithm C. This is a substantial decrease in the error and may indicate that the errors in Algorithm A are due to the included ineffective data points. In addition, Algorithm A performed best when estimating the BSFC curves for GT4, in contrast, the same simulation using Algorithm C have the

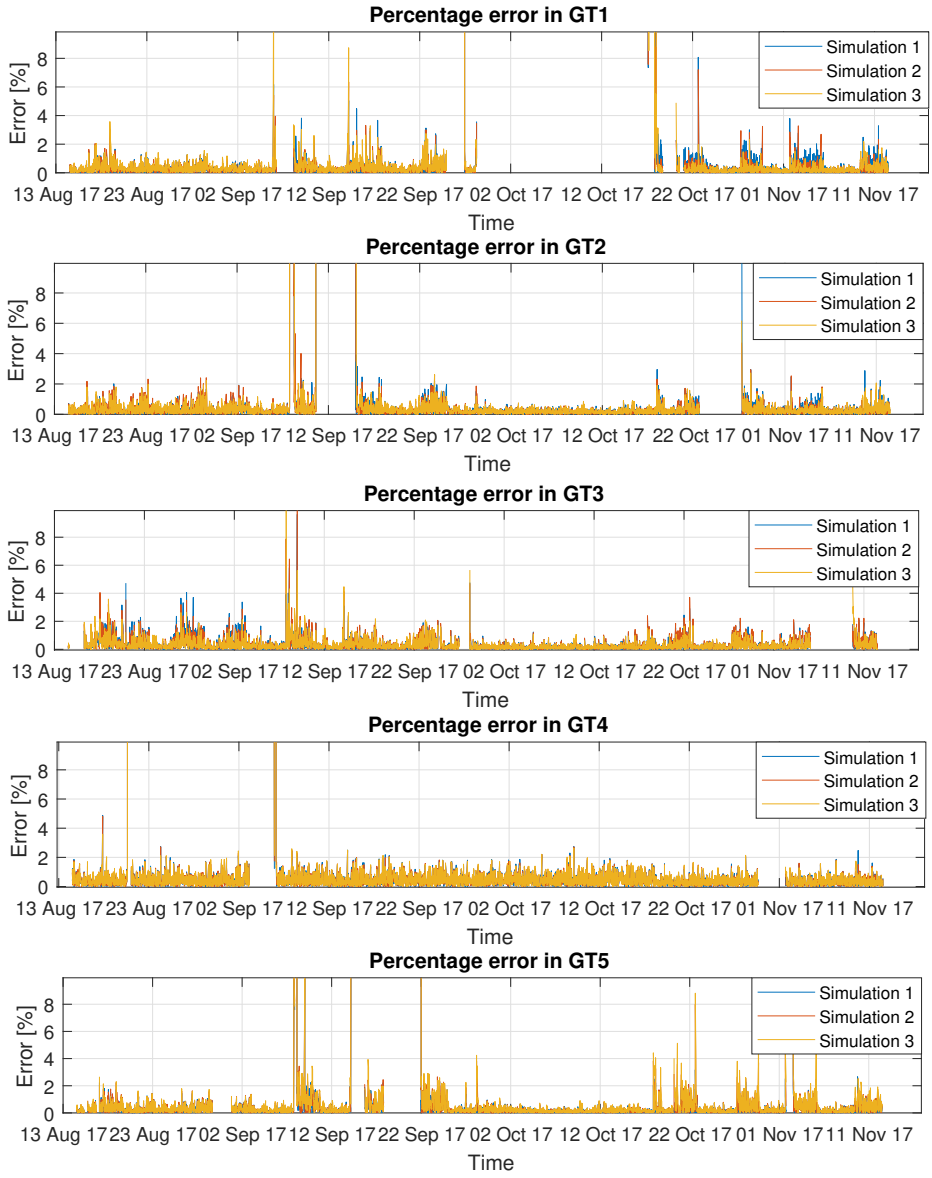

Figure 4.13: The percentage error between the real and estimated BSFC value when using Algorithm C.

Table 4.8: Total percentage error between measured and estimated BSFC in all GTs when using Algorithm C.

Simulation	GT1	GT2	GT3	GT4	GT5
1	0.35 %	0.29 %	0.32 %	0.42 %	0.26 %
2	0.33 %	0.29 %	0.32 %	0.44 %	0.29 %
3	0.36 %	0.29 %	0.29 %	0.45 %	0.33 %

maximum errors. GT4 has only two shut-offs during the simulation interval, and the set points are relatively stable which can explain the low errors using Algorithm A. However, when using Algorithm C the measurements from transient response are removed, hence the measurements creating trouble when using Algorithm A is not present in Algorithm C. Consequently the total percentage error is lower.

4.3.2 Algorithm D

A method using both polynomials and power series to estimate the BSFC curve was developed. The idea was to capture the different trends in the dataset due to different operational states. Three intervals was chosen to separate the dataset with the constraints in Equations (3.6)-(3.8) set to $l_1 = 0$, $l_2 = 20$, $l_3 = 70$ and $l_4 = 100$.

The nominal curve was found using data points from a two month period divided into the three intervals above. Next three separated estimation process was performed, which resulted in three curves presented in Figure 4.14. The upper subfigure contains the entire

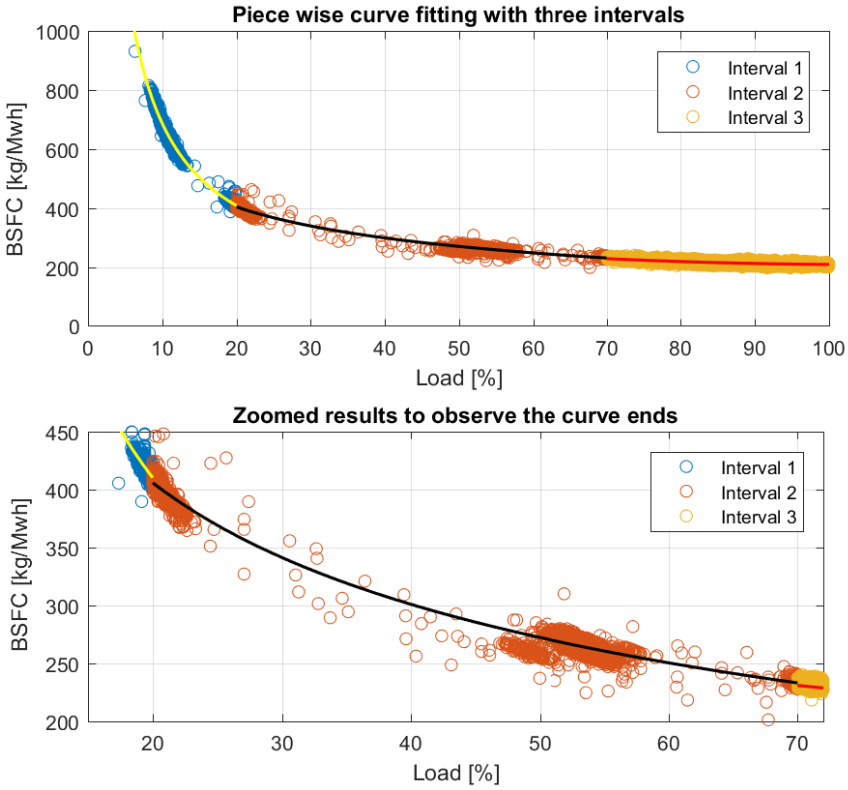

Figure 4.14: Results when estimating BSFC curve using piece wise curve fitting.

window of the data points, and the lower subfigure is zoomed to observe how the ends of the estimated curves lay in relation to each other. The curves were then connected to obtain a smooth curve. The result is compared to the estimate of the nominal curve using Algorithm A, with the same data points included in Figure 4.15.

Several methods to connect the curves were tested, including non-linear least squares, defining an optimization problem and cubic spline. However, the performance of these methods was far from sufficient compared to the results in Algorithm A. Hence; an online simulation was not performed with Algorithm D.

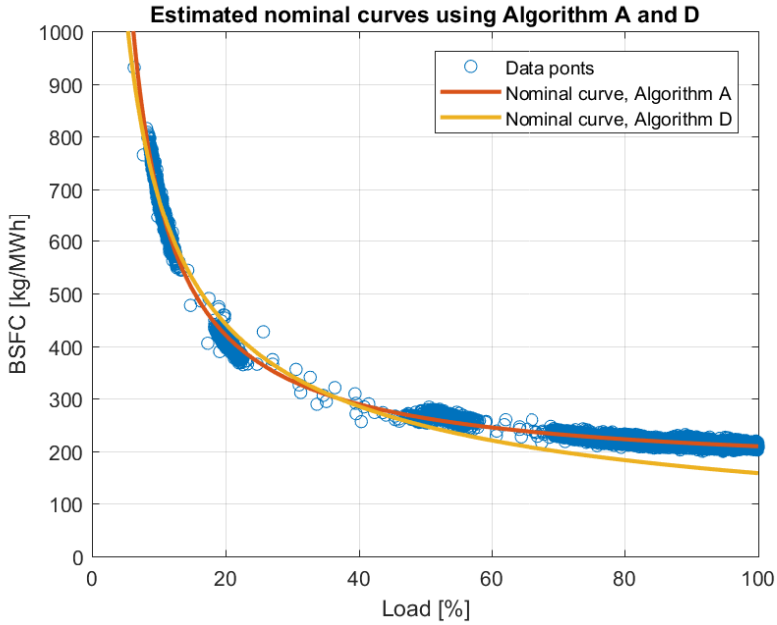

Figure 4.15: Comparing the estimated nominal curves found using Algorithm A and D. The curve estimated with Algorithm A is indicated with red, and yellow for Algorithm D.

4.4 Results from optimal load sharing

The main results from part two of this project are summarized in this section, the entire results and report can be obtained in Jung (2018). The Mixed Integer Non-Linear Programming (MINLP) model is calculating the optimal set points in GTs, including grid and an Energy Storage System (ESS) in the model. The model uses BSFC curves modelled offline using a curve fitting tool in *MATLAB* from approximately three months, using the same dataset as used in the simulations of the algorithms presented in this thesis.

The simulation of the optimal load sharing algorithm using nominal (but fixed) BSFC curves gave a saving potential of 2.48%. However, one should expect this result to change when the algorithm is simulated with the BSFC curves updated online. That is because the BSFC curve varies due to the factors described in Section 2.1.2 and thus not accounted for in the fixed BSFC.

Chapter 5

Discussion

The main goal of this master thesis was real-time estimation of the BSFC curves from noisy and poor quality measurements. That is, development of tailored algorithms that include pre-processing of measurements, measurement handling and update of the parameters of the BSFC model online. A data-driven method was used and is discussed first in Section 5.1 followed by a discussion of the least squares method. Next, the condition that states that the data points included in the estimation need to be measured in steady-state operation are discussed in Section 5.3. Important aspects in the simulation result has already been discussed in Chapter 4, hence Sections 5.4-5.7 include a more general discussion of the simulation results of the four algorithms.

5.1 The data driven method

A requirement defined by ABB was that the end system should be minimalist and relatively easy for a potential customer to install at their plant. Hence only four measurements were received and intended to be used in the system. From a product salesman point of view, this is a smart strategy because it may be easier to convince a customer to buy the system and test it without giving up to much information about their site and minimizing engineering cost. However, in the development stage of the algorithms, it is not optimal to model a system affected by outside conditions, as explained in Section 2.1.2, without obtaining measurements of these conditions and their impact. For example, if there are significant deviations in the parameter vector $\hat{\boldsymbol{\theta}}_{\boldsymbol{nls}}$, but the data points applied in the parameter estimation are stable, a question arises if the reason for the deviations is ambient conditions or an error in the algorithm. Not including measurements from ambient conditions will also affect the robustness of the algorithm since their effects are unknown.

However, the implementation of the system can be more straightforward regarding not having to derive complex mathematical models. As done in Klein and Abeykoon (2015) who used theoretical modelling with an extensive list of modelling parameters to model the specific fuel consumption. Klein and Abeykoon state that the model developed is as close to the reality as possible. Hence one should notice that ambient conditions such

as altitude, temperature, and pressure are included in the parameter list. The question is therefore if the same can be stated in the case of the model in this master thesis, using only four types of measurements, namely P, P_{max}, F and T. Having said that, it is possible to argue that the obtained parameters capture the essential behavior of the GTs including ambient conditions that will affect these measurements.

5.2 Least squares regression

Both linear and non-linear least squares regression have been applied to estimate the BSFC curves. As described in Section 2.3.2, LLS is easy to apply and quick to use according to Ljung (1976). This has also been experienced during the work with this master thesis. The method is relatively easy to understand and implement and gives a satisfactory result with small errors. However, the LS methods are sensitive to outliers as described in Chapter 1.5 in Hansen et al. (2013). Hence, the outlier detection algorithm developed in the M.Sc pre-project in Espeland (2017) was used to exclude all outliers from the dataset.

Alternatively, an approach for robust data fitting using the generalized maximum likelihood estimation (M-estimation) defined in Huber (2011), could have been applied as argued by Hansen et al. (2013). This method is something in between ordinary least squares and least squares combined with outlier detection, but as argued in Espeland (2017) the outlier detection algorithm performed well when estimating the BSFC curve offline. The algorithm has also performed sufficient in the online estimation of the BSFC curves, removing definitive outliers.

As mentioned in the introduction, recursive least squares (RLS), explained in Section 2.4.2, was applied to estimate the SFOC in diesel engines with a second order polynomial in Lundh et al. (2016). In comparison, MWLS was applied in this thesis where a *window* moves to include the new obtained measurements. An argument to do it this way is because of the form of the model, which in Algorithm A and B lead to $\hat{\boldsymbol{\theta}}$ being non-linear in the parameters in contrast to the linear system developed by Lundh et al. The recursive method is favorable regarding speed because the estimate of $\boldsymbol{\theta}^*$ is corrected based on new information instead of expanding the measurement space and solving the LS problem from the beginning. However, all measurements obtained from the beginning are not included in the estimation. Instead, a finite set of measurements decided from other conditions, for example, temperature as in Algorithm B. Consequently, the speed is not an issue and the simplicity of the system is higher when not using RLS which can be beneficial on a customer's point of view.

5.3 The steady-state condition

As presented in Section 2.1.2 and 2.1.3, it is essential for the GTs to be in steady-state operation when the BSFC curve is estimated. Most likely, the GTs run on loads above 40% when the operation is classified as being in steady-state because this is more efficient. Referring to for example Figure 4.3, where it can be observed lower BSFC values for higher loads resulting in the GTs being more efficient for higher loads. Hence, a steady-state counter is included in the algorithms to ensure that the GTs have been in steady-state

for at least 10 minutes when the curve is updated. This is essential due to the criteria for a performance test, see Section 2.1.2.

However, a problem arises due to the load sharing algorithm requiring a BSFC curve representing the entire load interval, that is a BSFC value for each load between 0% and 100%. The GTs are not restricted to operate in lower loads but are only operated in such inefficient conditions in transient operation. Hence, measurements from the transient operation are added in the estimation set, but only when the load is below 40% to not interfere with the steady-state measurements. Consequently, the optimization problem may be based on a curve that is invalid for lower loads, which can lead to a non-optimal output from the load sharing algorithm.

Despite this, the curves seem to be little affected by including measurements from transient operation, when the GTs are in steady-state operation. This is because the measurement handling process in the algorithms ensures that there are significantly more measurements from steady-state operation than transient operation. This leads to the estimation "favoring" the steady-state measurements. Consequently, as long as this condition is taken into account in the final system with storage handling, and possibly adding an extra constraint in the optimization problem in Jung (2018) avoiding set points below a particular value, the problem is overcome.

5.4 Algorithm A

Algorithm A estimates BSFC curves using NLLSR based on a working set divided into five intervals. The choice of boundaries between the intervals was based on the desire to spread the measurements over the entire load interval (0%-100%). However, many measurements are received in one area of the load interval, due to the GTs running on the same set point for a longer time. Consequently, if the working set had not been divided, the curve would be estimated based on measurements concentrated in one or two areas of the load interval, leading to significant errors. This can be observed in Figure 4.12 when using Algorithm C.

However, it is not guaranteed that the GTs will run on set points in all the load intervals in the working set. For instance, GT4 run on loads between 75% and 90% in the entire simulation interval (except shut off), see Figure 4.1. Consequently, most of the measurements will be inserted in interval 2 (70%-85%) in the working set, and the other intervals will be filled with measurements obtained offline. These measurements are measured under other ambient conditions. Hence, the curve may describe other conditions than the current, leading to calculating non-optimal set points in the load sharing algorithm.

That said, not including new measurements in some of the intervals cannot be fixed by dividing the intervals into even smaller intervals, the measurements will nevertheless end up in the same area of the load axis ($0\% - 100\%$). Also, these intervals will contain old measurements; thus this is only a temporary solution. Another approach can be a weighting of the measurements, which the newer measurements in each interval have the most significant impact on the curve. Alternatively, a curve describing only a part of the load interval as in Algorithm C can be used. That said, the optimal load sharing algorithm is not likely to calculate the new set points to be lower loads. Thus it is not crucial for the curve to be based on the newest measurements in the lower parts of the load axis.

Some comments on the percentage error in Figure 4.4 have already been mentioned in Section 4.1.2 and will be further discussed here. The percentage error is overall relatively low for GT1-4. However, there are some significant deviations. As can be observed in the figure, these deviations often arise before a shut-off or on in the GTs. This behavior is expected because when the GTs are in transient operation, the parameters are not updated. Instead, the parameters from the previous update are kept. Consequently, when the estimated BSFC is calculated in transient response, the model for steady-state operation is used and naturally gives a more substantial error. This argument supports the importance of not updating the curves in transient operation which can cause the entire curve to be based on wrong assumptions.

From the percentage error, it can also be observed that the performance of the algorithm with different simulations varies. That is, in GT3 for example, the error is most substantial in the results from simulation 1 at the beginning of the simulation interval, but from the middle of the simulation interval, the result from simulation 3 has the largest error. To increase the accuracy of the algorithms, a working set with varying size could be implemented. That is when there are small and fast changes in the measurements a smaller working set should be used, and with slow variations, a larger set is preferred. Besides, a functionality checking the estimation error should be developed, rejecting parameter updates if the error is above a threshold.

From the resulting BSFC curves in Figure 4.3, one can argue that the size of the working set has an impact. The differences between the curves from simulation 1 and 2 are not large, but they deviate from the results from simulation 3. As mentioned in Section 4.1.1, the three simulation results correspond in having the curve from September 29 furthest up, that is indicating the least efficient GT operation. It was then discussed that machine degradation does not affect the curve since the curve at a later date indicates a more efficient operation. However, if it was performed maintenance on the GT between the estimation of these curves, it could still be a possibility. After all, it was a shut down of GT4 in the end of October where maintenance could have been performed, see Figure 4.1. Another, aspect to consider is the effect of ambient temperature. As described in Section 2.1.2 and discussed in Section 4.2.2, colder temperatures may have a positive impact on the performance. One could assume that the temperature is lower in mid-November compared to the end of September and this can be the reason for the curve from November 12 indicating more efficient GT operation than the curve from September 29.

However, it is difficult to conclude regarding the placement of the curves due to all other factors affecting the curves that are not included as parameters in the algorithm. That said, the results from Algorithm B and C support the hypothesis, that the curve from November 12 is below the curve from September 29 due to temperature differences, see Figure 4.9 and 4.12. Then again, the curves from November 12 are also below the nominal curve differing it from the results in Algorithm A.

The resulting BSFC curves and estimated parameters from the remaining GTs are for the interested reader presented in Appendix A. Nevertheless, the results are commented here. Studying the percentage error from estimating the BSFC curve in GT2 (Figure 4.4), one can observe that there are significant errors around November 11 for simulation 3. Comparing with the BSFC curves for GT2 in Figure A4, one can observe that the curve from simulation 3 is different from the curve from simulation 1 and 2. By inspecting

the load at this time in Figure 4.1, one can see that this area has variations in set points. Because of the large working set in simulation 3, these variations may be included in the working set leading to an incorrect curve. In contrast, when a smaller working set is used fewer measurements are included; hence the variation within the data points in the working set decreases. Consequently, a smaller working set is preferred when there are large variations in the set points, that is the set points vary from, for example, 80% to 95%.

The error is also significant for GT3 in the interval between approximately August 23 and September 12. This is seen in the estimated parameters for GT3 presented in Figure A5 in Appendix A. The results from simulation 1 deviate significantly compared to the results from simulation 2 and 3. This can be explained by investigating the loads of the GTs presented in Figure 4.1. As observed, the set points are varying in different degree, that is, small and large amplitudes. The working set is maybe to small to capture this change. Consequently, when the BSFC is calculated using the current model, the model might not be updated to describe the current conditions because of the many variations.

5.5 Algorithm B

Algorithm B weights measurements according to load, time and ambient temperature and further estimates the BSFC curves in the GTs using a power series. The weight storage was divided in $n \times 10 \times 50$ indexes, where n is the number of possible measurements in each load interval, 10 load intervals and 50 temperature intervals. Hence, the estimation set is also divided into 10 intervals denoted estimation intervals, of sizes varying with the current set point. In comparison with Algorithm A, where the working set is divided into five intervals, the measurements will be more spread along the entire load axis (0%-100%) when using Algorithm B. It can be argued that this can be the reason for better performance regarding total percentage error presented in Table 4.6. Recall that the total percentage error in GT5, simulation 3 was 5.54% when using Algorithm A, while the corresponding error using Algorithm B was 0.60%. That is a substantial improvement!

Another reason for this improvement can be the varying load interval sizes in the estimation set, as mentioned in Section 4.2.3. The total size of the estimation set is fixed in each simulation; however, the number of included measurements from each load interval in the weight storage varies. Hence, more measurements around the current set point are included leading to the LSR striving to fit the model to the highest density of measurements, making the accuracy higher in these areas.

On the other hand, if the current set point is not representative of the current system behavior, for example, if a wrong measurement is received with $load = 60\%$, and the GT actually runs on 95% the updated curve is based on wrong assumptions. However, the number of measurements included for the higher loads are always large to ensure measurements in the ”important” load interval (70%-100%). In addition, the included measurements are measured with the same ambient temperature, resulting in the ”correct” measurements being included anyway. This is because it is assumed that the temperature is stable; hence the temperature is the same for the wrong and correct measurements.

This method is highly based on ambient temperature measurements, so the question is; how much does actually temperature affect the performance. By comparing the simulation results with Algorithm A, Algorithm B seems to be more robust and not that affected by

changes in set points. With that said, it is not straightforward to argue if this increase in algorithm performance is due to the weighting logic, varying buffer sizes or the fact that the working set is divided into smaller load intervals. Hence, Algorithm A needs to be simulated with smaller and varying load intervals to decide this.

Another aspect to consider is the computational complexity. The simulation of Algorithm B is much slower compared to Algorithm A and C. This is due to the extra weighting logic of the new measurements, and the matrix operation inserting in the weight storage. However, this will not affect the real-time behavior if the measurements are received every 10 seconds as simulated with in this thesis. Then it is more enough time to classify, filter and weight the new measurements before the next measurements are received. Algorithm A is much faster to simulate because the new measurements are directly inserted in the working set after classification and filtering, avoiding the extra weighting time and the time to obtain the relevant measurements from the weight storage. The actual update of the BSFC curve is the same for all online algorithms and is relatively fast compared to the measurement handling performed in advance.

The performance of the algorithm seems to be stable which is also indicated from the results from the other GTs in Appendix B. Studying the estimated parameters in Figures B1, B3, B5 and B7, the same squared shape behaviour as mentioned in Section 4.2.2 is observed and there are few large deviations. However, the algorithm meet some challenges when simulating with measurements from GT3 (see Figure B6 in Appendix B), there are many substantial changes in the parameters, especially in the results from simulation 1 and 2, but compared to the Algorithm A (Figure A6) the result is better for the curve from September 29. Also, comparing Algorithm A and B with regards to the BSFC curves for GT5 (Figures A8 and B8), Algorithm B outperforms Algorithm A and is therefore more robust against changes in the set points.

5.6 Algorithm C

Algorithm C estimates the BSFC curves excluding measurements with loads below 40%. The working set is divided into two intervals, measurements with $load > 70\%$ are inserted in one interval, and the remaining in the other interval. There are expected small changes in the estimated parameters since there are only included measurements above 40% and the GTs does not vary significantly in fuel consumption for higher loads. The results presented in Figure 4.11 supports this assumption. This is also observed in Appendix C (Figures C1 - C4), the estimated parameters in the GTs are not varying significantly over the simulation interval.

However, there are some exceptions where the estimated parameter values deviate significantly, this is especially seen at the beginning of the simulation of GT2 (Figure A3). A possible explanation can be either large changes in the GT set points (transient response), or the GTs being shut off or on. This can indicate that the working sets contains too few measurements or should be divided into several intervals. An alternative solution can be to shorten the load interval. Referring to Figure 3.9, the load interval could have been shortened to be between 70% and 100%, and still have all steady-state measurements included.

As commented in Section 4.3.1, the working set should perhaps be divided into several intervals because of the phenomena seen in Figure 4.12, and this can also be observed

in the results from the remaining GTs in Appendix C. That is, the BSFC curves bend upwards at the end of the load interval ($loads > 85\%$). This has been explained for GT4, but studying the loads in Figure 4.1, the GTs are run on approximately 80% at the end of the simulation interval. Comparing this with the BSFC curves from November 12 (Figures C1-C4 in Appendix C), all curve upwards when the working set is small (300 measurements). This validates the discussion in 4.3.1, and a solution can be to divide the working set into several intervals. Alternatively, the size of the working set must increase to include enough measurements such that all loads are covered in the dataset. The first solution is however probably best because there need to be included a significant number of measurements in the latter approach, that can result in the BSFC curve not capturing changes in the dataset.

Another problem with Algorithm C is the small oscillations in the estimated parameters, and this is best seen for GT4 in Figure 4.11. As discussed in Section 4.3.1, this can be due to the working set containing too few measurements. This combined with dividing the working set into only two intervals makes the algorithm prone to errors and changes because the collection of measurements moves around in the interval affecting the estimated parameters. However, it does not seem that these small oscillations affect the final BSFC curves significantly, but should nevertheless be investigated further.

With all this in mind, using the load interval 40%-100% to estimate the BSFC curve is a good idea due to the steady-state condition. However, this requires more work on the optimal load sharing algorithm allowing curves only describing a subset of loads, not the entire interval from 0% to 100%. Also, the working set should be divided into several intervals spreading the measurements over the load axis and avoiding the natural shape of the second order polynomial. If these measures are met, the algorithm will probably perform better and be more robust to noise. This is because the BSFC values are not varying significantly for higher loads and the shape of the curve will probably not change. Hence it is a problem of moving a fixed graph up and down.

5.7 Algorithm D

Algorithm D was developed to test an alternative approach by recognizing different responses in the BSFC values, and estimate the BSFC curve with separate models. The idea was then to connect the curves to one smooth curve that could be passed on to the load sharing problem.

As observed in Figure 4.14, one can argue that there are three responses in the measurements. That is, interval 1 which is very inefficient operation compared to interval 2 and 3. Interval 3 is the interval the GT should be operated in to have the highest efficiency in the operation, and interval 2 is the efficiency when the GT changes it set points, this is validated in Figure 3.9. The estimated curves fit well with the surrounding measurements. However, as seen in Figure 4.15, where the estimated curves from Algorithm A and D are compared, the merge of the separated curves are not sufficient. The accuracy is worse than in Algorithm A, and Algorithm A also has a more straightforward estimation process.

However, to identify different responses in the BSFC values, and for example, only update the BSFC curves for interval 3 when there are changes in the measurements could give a more robust system. That is, a smooth curve describing the entire load interval,

as well as eliminating the fitting error when measurements from transient response are included. Of course, a method merging the separate curves into a smooth continuous curve must first be developed.

Chapter 6

Concluding remarks and future work

6.1 Concluding remarks

Three tailored algorithms have been developed in this thesis to estimate the BSFC curves in generating machines in real-time using least squares regression. All algorithms contain an estimation of the nominal curve, classification of the operational state of new measurements and outlier detection. The differences in the algorithms are in how the measurements are chosen to be the basis of the estimation as well as the model being used. In addition, a fourth algorithm was explored regarding offline modelling of the nominal curve using separate models to estimate different responses in the system. The first three algorithms were simulated using real data from site and updating of the BSFC curve every 30 minutes.

The first algorithm, Algorithm A, applied a FIFO logic to always include the newest received measurements in the working set used for updating the curve. The simulation results indicate that the working set should be divided into several load intervals to be more robust against changes in the operational conditions in the GTs. Nevertheless, the percentage error between the measured and estimated BSFC values suggest that the algorithm performs well in steady-state operation, but when there are fast changes in the set points, as in GT3 and GT5, the algorithm gives non-satisfactory results.

The second algorithm, Algorithm B, applied a weighting of the measurements based on temperature, load and time when the measurements were received to decide which measurements to include in the estimation. This was done to account for the effect of ambient temperature and estimate new curves based on current conditions. The result of the algorithm indicates that using a working set including a higher number of measurements is beneficial regarding the robustness of the algorithm. The algorithm gives stable results throughout the entire simulation interval for all GTs, even resisting the fast changes in set points in GT3 and GT5 when simulated with a large working set.

Algorithm C was the third algorithm presented. Here, the estimation was only based

upon measurements with higher BSFC values, that is $load > 40\%$, and the curve was modelled using a second-order polynomial. The simulation results suggest that the working set should be divided into several intervals to avoid a collection of measurements in one area at the load axis. In addition, it was suggested to shorten the interval described by the curve, that is only to include measurements above 70%. This will have a great impact on the stability of the estimated parameters.

The final algorithm, Algorithm D, estimates the nominal curve using an offline estimation of separate models to describe different responses in the BSFC values. The dataset was divided into three intervals, two intervals containing measurements from transient operation and one interval containing measurements from steady-state operation. This method was applied to attempt to decrease the estimation error because different system responses should not be estimated using one curve. That is because the BSFC curve is not valid when the GTs are not in steady-state. However, the curve modelled with this algorithm is not sufficient at the current time and needs more work to obtain a smooth curve.

Since this is a data-driven method based only on four measurements (five in Algorithm B), it is hard to account for the ambient conditions affecting the BSFC curve. Hence, there are some uncertainties in the model as well as in the simulations because the cause of the error cannot be guaranteed with 100% certainty. Also, in Algorithm B, the ambient temperature is obtained from the Norwegian Meteorological Institute and is the average temperature every hour. Hence there are errors in the temperature data itself due to measuring equipment, and the actual temperature in every data point. The latter is due to the averaging of the temperature because all measurements from one hour all have the same temperature which may not be the case.

Research performed in this master thesis has validated that the performance characteristics are highly individual in each GT in addition to also changing over time. To exploit these differences and model BSFC curves that are updated in real-time, combined with an optimal load sharing algorithm calculating new set points in the GT have great potential. The proposed solution can reduce the harmful emissions from burning fossil fuels and decrease costs regarding fuel consumption and possibly maintenance. It is a win-win situation for the environment and the process industry!

6.2 Future work

6.2.1 Further development

The accuracy of the algorithms developed in this thesis could benefit from being developed further. That is, adding a functionality keeping track of the updated parameters and rejecting parameters that are very different from the previous estimates. Hence, the large errors could be avoided, and the robustness of the systems would increase.

As suggested in Section 5.4, a varying buffer size should be included in Algorithm A or varying the number of measurements in the working set due to current set point, as performed in Algorithm B. This is because the simulation results from Algorithm A varies according to the conditions in the GTs, that is fast or slow variations in set points. Hence, a smaller working set is better for fast changes in the set points, while a more extensive working set would be better if there is much noise in the measurements or slow

oscillations.

It would also be interesting to analyze the cause of the errors in the different algorithms using several measurements from the site. Maybe humidity in the air or machine degradation have a high impact on the efficiency. However, this would require more data measurements from the customer resulting in ignoring the simplicity requirement of the system. That said, it would be possible to account for the disturbances in the algorithms using feedforward and thus increase the accuracy of the BSFC curves leading to greater soundness in the calculated set points.

Algorithm D should also be further developed, that is, improve the merging of separate curves such that the resulting curve is continuous and smooth. The offline system should next be extended to include an online adaptation of the estimated parameter, concentrated on updating the curve describing interval 3 in the dataset.

As described in Section 2.1.2, the performance and hence the BSFC curve is affected by the quality of fuel in the GTs. The algorithms should therefore be extended to include a functionality to recognize the quality of fuel used. This can be achieved using a neural network or a classification algorithm recognizing different behavior of the BSFC curve. This is important because it is maybe needed a completely different model to estimate the efficiency when using for example diesel. This is seen in Lundh et al. (2016) where a second order polynomial was applied to fit BSFC values in a diesel electric marine vessel.

An alternative use case scenario with online modelling of BSFC curves is that the curves can be used in condition-based monitoring. The curves are a measure of how efficient the GTs consumes fuel, and if this curve changes drastically, it can indicate a fault in the GT. Hence, an unforeseen shut down of the GT can be avoided and instead, a planned shut down can be performed leading to reduce maintenance costs.

6.2.2 Optimal load sharing algorithm

The development of the optimal load sharing algorithm using a MINLP approach developed in Jung (2018), applies fixed BSFC curves estimated based on the entire simulation interval, that is, not updated in real-time. A more reliable result will be obtained by using one of the algorithms developed in this master thesis, where the BSFC curves continuously are updated with regards to new measurements, and in Algorithm B including ambient temperature.

Also, it should be possible to use the optimal load sharing algorithm with BSFC curves describing only a part of the load interval, for example measurements with loads from 40% to 100%. This can increase the robustness of the algorithm, since measurements with high BSFC values are not included.

Bibliography

Brooks, F. J., 2000. GE Gas Turbine Performance Characteristics. GE Power Systems, Schenectady, NY.

Brun, K., Nored, M. G., 2006. Guideline for Field Testing of Gas Turbine and Centrifugal Compressor Performance. Tech. rep., Gas Machinery Research Council, Southwest Research Institute.

De Sa, A., Al Zubaidy, S., 2011. Gas turbine performance at varying ambient temperature. Applied Thermal Engineering 31 (14-15), 2735–2739.

Erdem, H. H., Sevilgen, S. H., 2006. Case study: Effect of ambient temperature on the electricity production and fuel consumption of a simple cycle gas turbine in Turkey. Applied Thermal Engineering 26 (2-3), 320–326.

Espeland, E. M., 2017. Offline Modelling of Fuel Efficiency Curves in Gas Turbines. Tech. rep., Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology.

Hansen, P. C., Pereyra, V., Scherer, G., 2013. Least Squares Data Fitting with Applications. JHU Press.

Huber, P. J., 2011. Robust Statistics. In: International Encyclopedia of Statistical Science. Springer, pp. 1248–1251.

Jehlik, F., Rask, E., 2010. Development of Variable Temperature Brake Specific Fuel Consumption Engine Maps. Tech. rep., SAE Technical Paper.

Johnson, M. L., Faunt, L. M., 1992. Parameter estimation by least-squares methods. Methods in enzymology 210, 1–37.

Jung, B. K., 2018. An MINLP Approach Integrating Predictive Method for Optimal Load Sharing in Generating Machines. Master's thesis, Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology.

Klein, D., Abeykoon, C., 2015. Modelling of a Turbojet Gas Turbine Engine. In: Internet Technologies and Applications (ITA), 2015. IEEE, pp. 200–206.

Levi, H., 1967. Polynomials, Power Series, and Calculus. University series in undergraduate mathematics. Van Nostrand.

Ljung, L., 1976. Consistency of the Least-Squares Identification Method. IEEE Transactions on Automatic Control 21 (5), 779–781.

Ljung, L., 1999. System Identification: Theory for the user. Prentice-Hall PTR.

Lundh, M., Garcia-Gabin, W., Tervo, K., Lindkvist, R., 2016. Estimation and Optimization of Vessel Fuel Consumption. IFAC-PapersOnLine 49 (23), 394–399.

Mayer, V., Varaksina, E., 2017. Electromagnetic Faraday generator and its application. Physics Education 52 (4), 045018.

Paparella, F., Domínguez, L., Cortinovis, A., Mercangoz, M., Pareschi, D., Bittanti, S., 2013. Load Sharing Optimization of Parallel Compressors. In: Control Conference (ECC), 2013 European. IEEE, pp. 4059–4064.

Revelle, R., Suess, H. E., 1957. Carbon Dioxide Exchange Between Atmosphere and Ocean and the Question of an Increase of Atmospheric CO_2 during the Past Decades. Tellus 9 (1), 18–27.

Sharma, S., Khare, S., Huang, B., 2016. Robust online algorithm for adaptive linear regression parameter estimation and prediction. Journal of Chemometrics 30 (6), 308–323.

Skjong, E., Johansen, T. A., Molinas, M., Sørensen, A. J., 2017. Approaches to Economic Energy Management in Diesel-Electric Marine Vessels. IEEE Transactions on Transportation Electrification 3 (1), 22–35.

Teunissen, P., 1990. Nonlinear least squares.

Wärtsilä, 2015. Improving Engine Fuel and Operational Efficiency. Accessed 09.02.18.
URL `https://www.wartsila.com/services/learning-center/business-white-papers`

Woodford, C., 2017. Turbines. Accessed 17.09.17.
URL `http://www.explainthatstuff.com/turbines.html`

Xu, O., Fu, Y., Su, H., Li, L., 2014. A Selective Moving Window Partial Least Squares Method and Its Application in Process Modeling. Chinese Journal of Chemical Engineering 22 (7), 799–804.

Appendix A

Additional simulation results for Algorithm A

The additional simulation results when using Algorithm A is presented in this Appendix. The estimated parameters, $\hat{\boldsymbol{\theta}}_{\boldsymbol{nls}} = [a, b, c]^T$, are presented in Figures A1, A3, A5 and A7 for GT1, 2, 3 and 5, respectively. The resulting BSFC curves calculated with $y = ax^b + c$ is presented in Figures A2, A4, A6 and A8, where the BSFC curves from three dates are presented for all simulations. The results from GT4 are presented in Section 4.1.1.

One should notice that the BSFC curves are quite affected when there are large changes in the set points in the GTs. This can especially be seen for GT2 and GT5.

Generating turbine 1 (GT1)

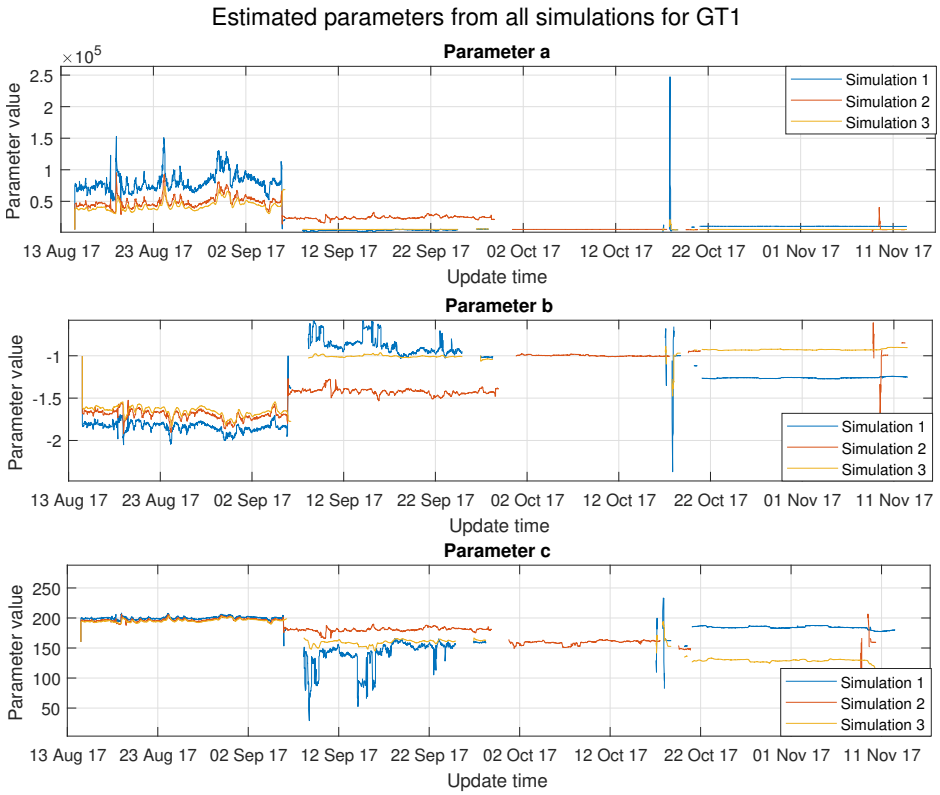

Figure A1: The estimated parameters of the BSFC curve describing the efficiency in GT1 using Algorithm A.

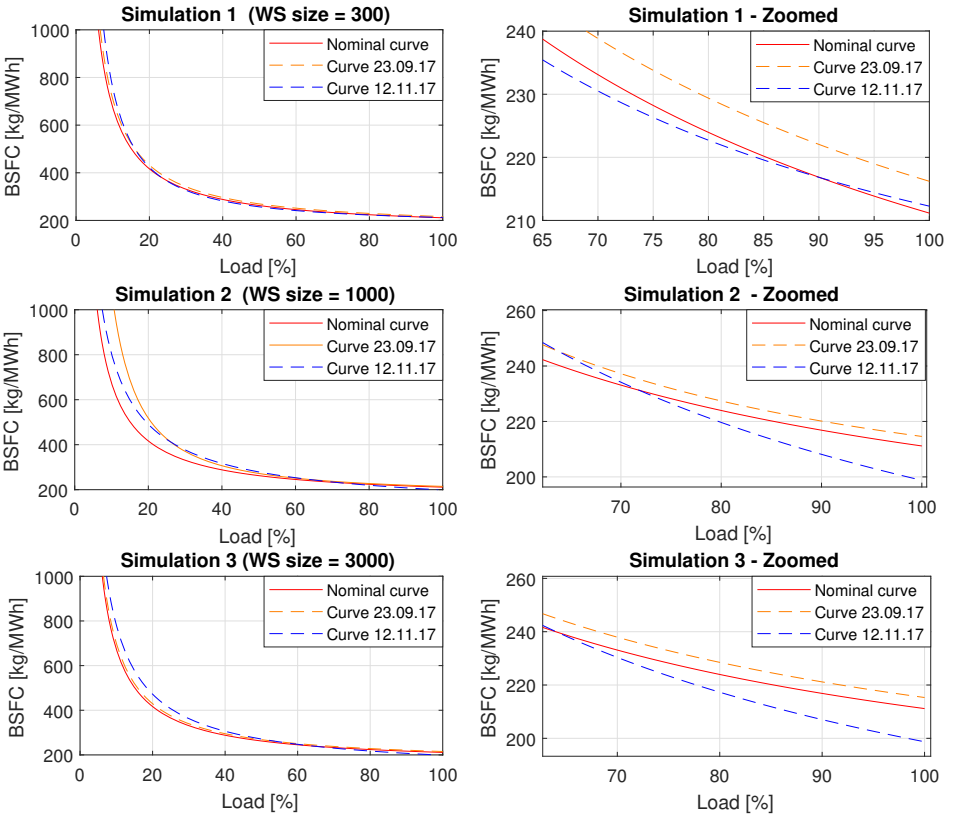

Figure A2: To the left are the resulting BSFC curves for the entire load interval (0%-100%) from three dates, to the right is the same curves but presented in a shorter interval, all for GT1.

Generating turbine 2 (GT2)

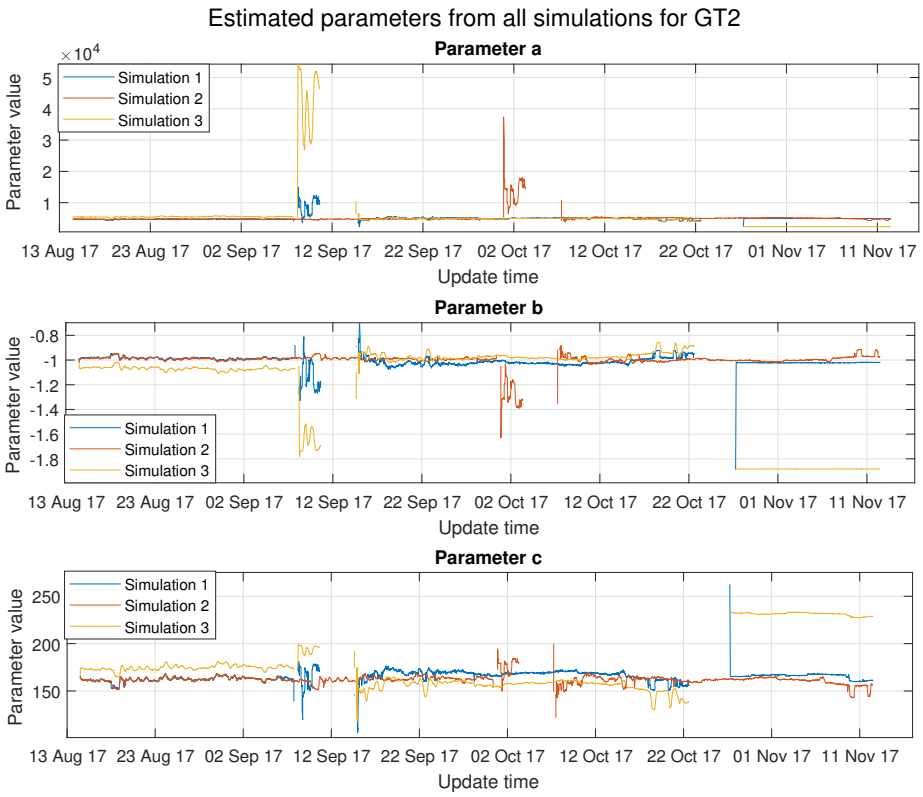

Figure A3: The estimated parameters of the BSFC curve describing the efficiency in GT2 using Algorithm A.

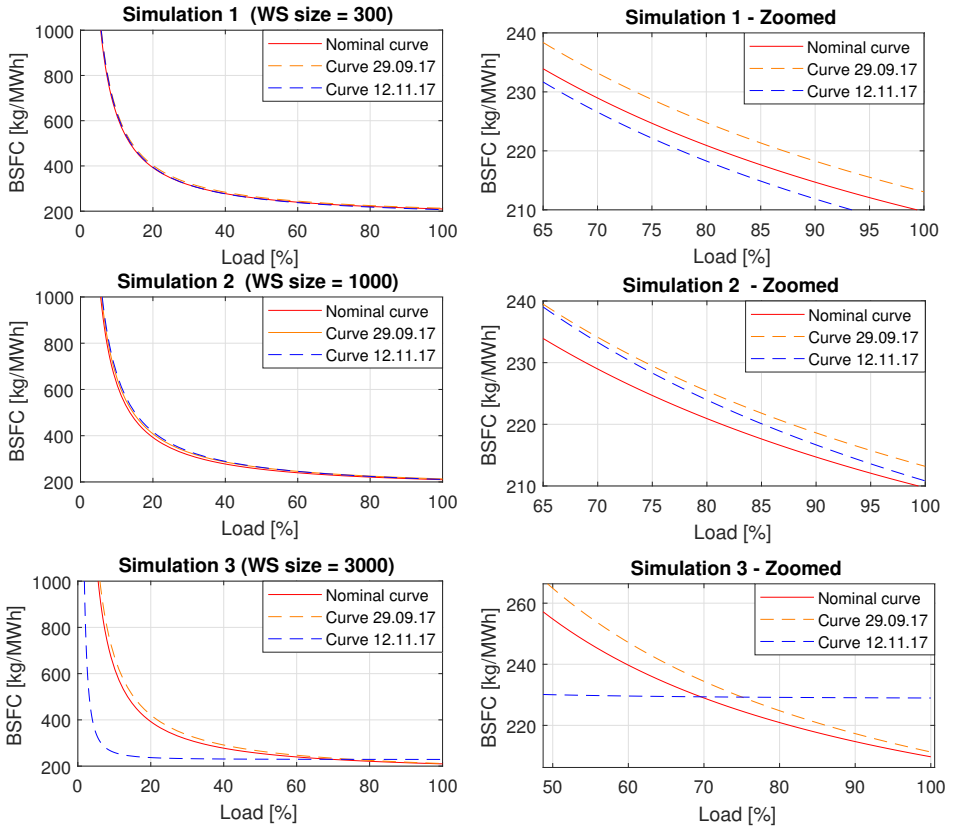

Figure A4: To the left are the resulting BSFC curves for the entire load interval (0%-100%) from three dates, to the right is the same curves but presented in a shorter interval, all for GT2.

Generating turbine 3 (GT3)

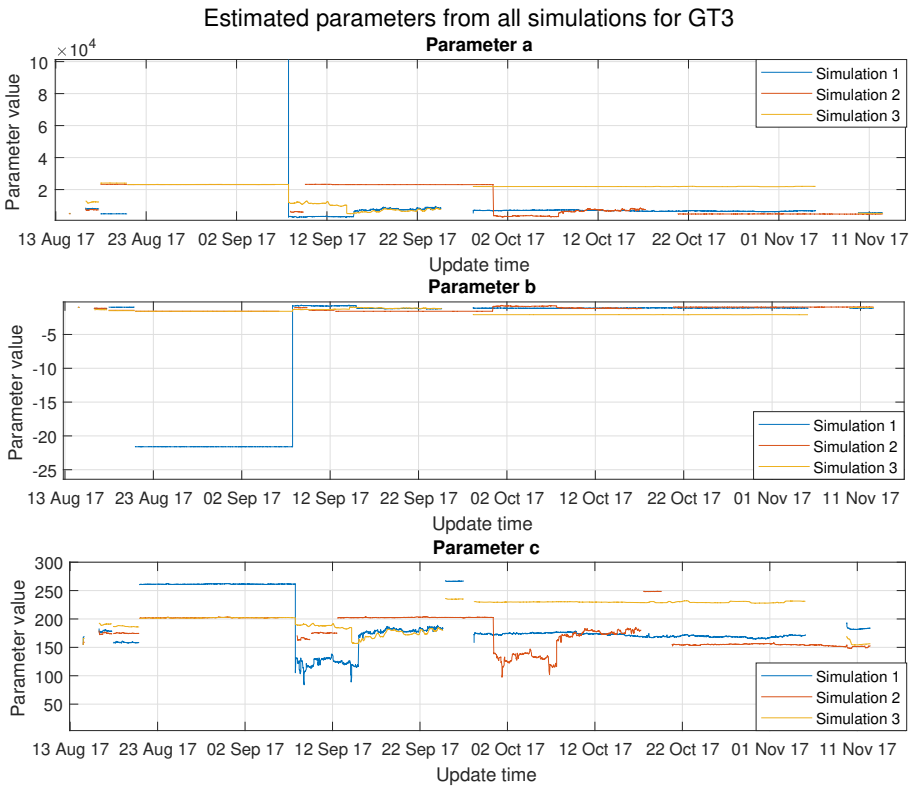

Figure A5: The estimated parameters of the BSFC curve describing the efficiency in GT3 using Algorithm A.

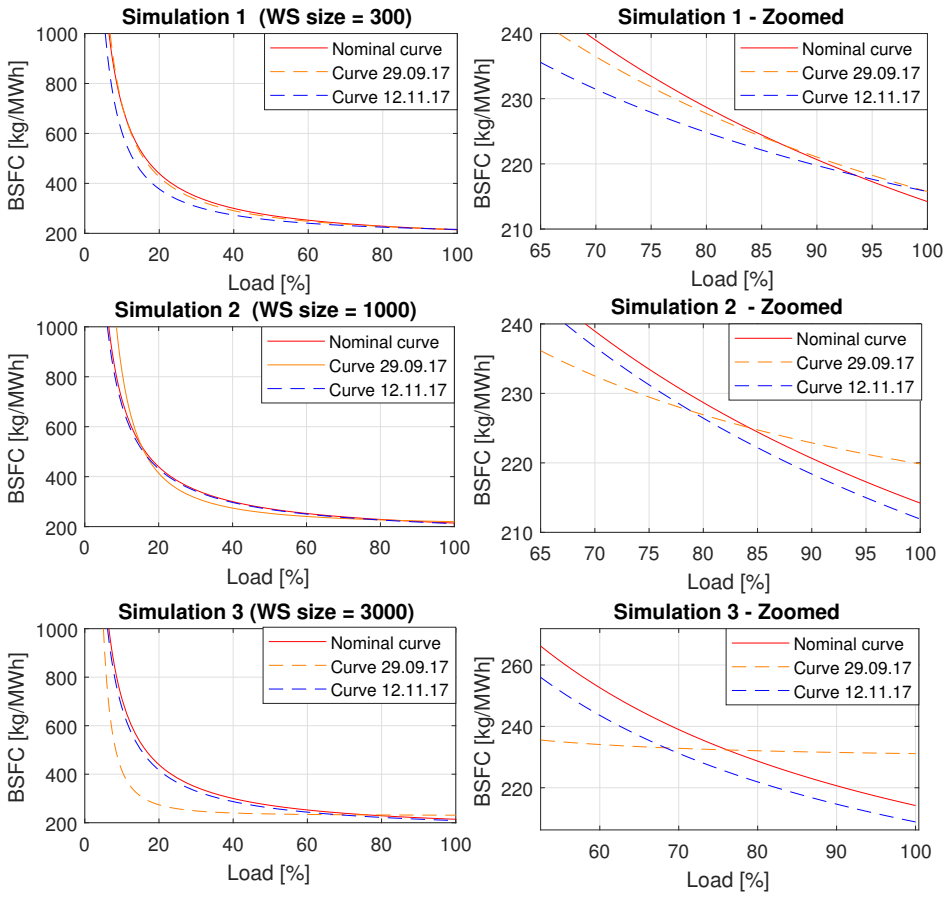

Figure A6: To the left are the resulting BSFC curves for the entire load interval (0%-100%) from three dates, to the right is the same curves but presented in a shorter interval, all for GT3.

Generating turbine 5 (GT5)

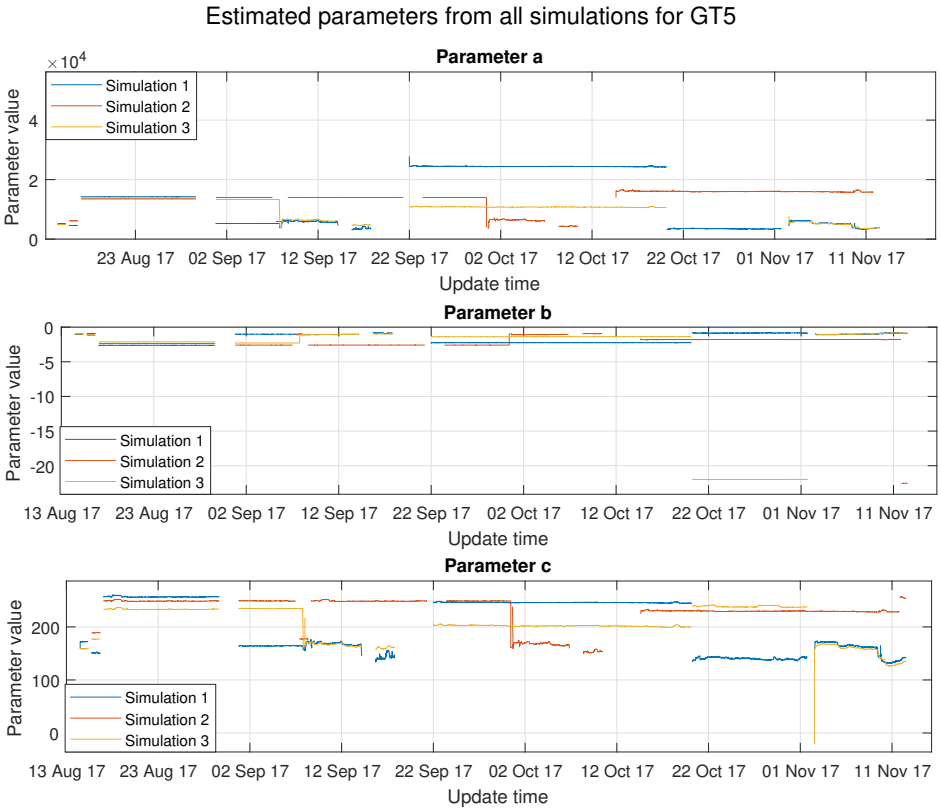

Figure A7: The estimated parameters of the BSFC curve describing the efficiency in GT5 using Algorithm A.

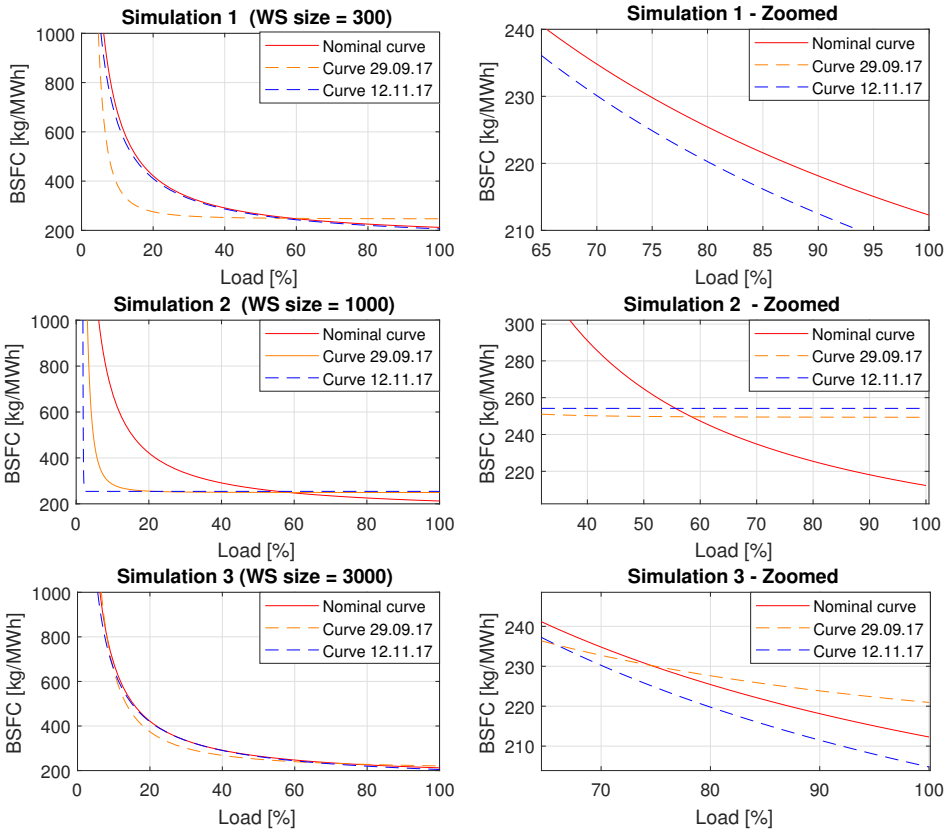

Figure A8: To the left are the resulting BSFC curves for the entire load interval (0%-100%) from three dates, to the right is the same curves but presented in a shorter interval, all for GT5.

Appendix B

Additional simulation results for Algorithm B

The additional simulation results when using Algorithm B is presented in this Appendix. The estimated parameters, $\hat{\boldsymbol{\theta}}_{\boldsymbol{nls}} = [a, b, c]^T$, are presented in Figures B1, B3, B5 and B7 for GT1, 2, 3 and 5, respectively. The resulting BSFC curves calculated with $y = ax^b + c$ is presented in Figures B2, B4, B6 and B8, where the BSFC curves from three dates are presented for all simulations. The results from GT4 are presented in Section 4.2.2.

The results are overall stable, with simulation 3 rejecting many of the changes in the GT set points. However, the reader should notice the results for GT3 and notice the difference in performance in the simulations. For example, studying parameter b, the results from simulation 2 is having major deviations in the values. Simulation 1 is similar to simulation 2, but the result is partly covered in the results from parameter b.

Generating turbine 1 (GT1)

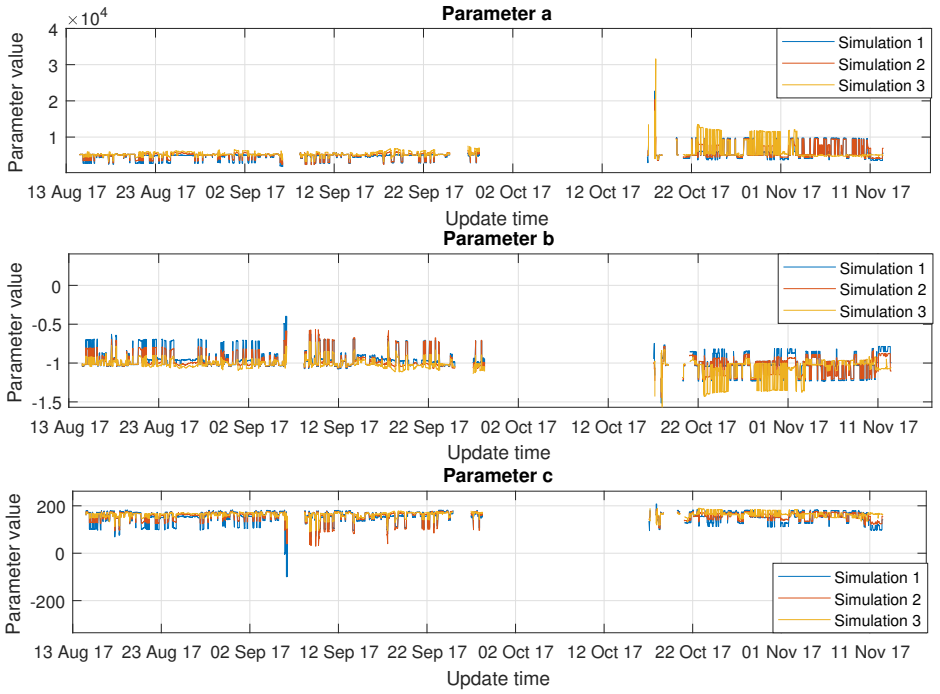

Figure B1: The estimated parameters of the BSFC curve describing the efficiency in GT1 using Algorithm B.

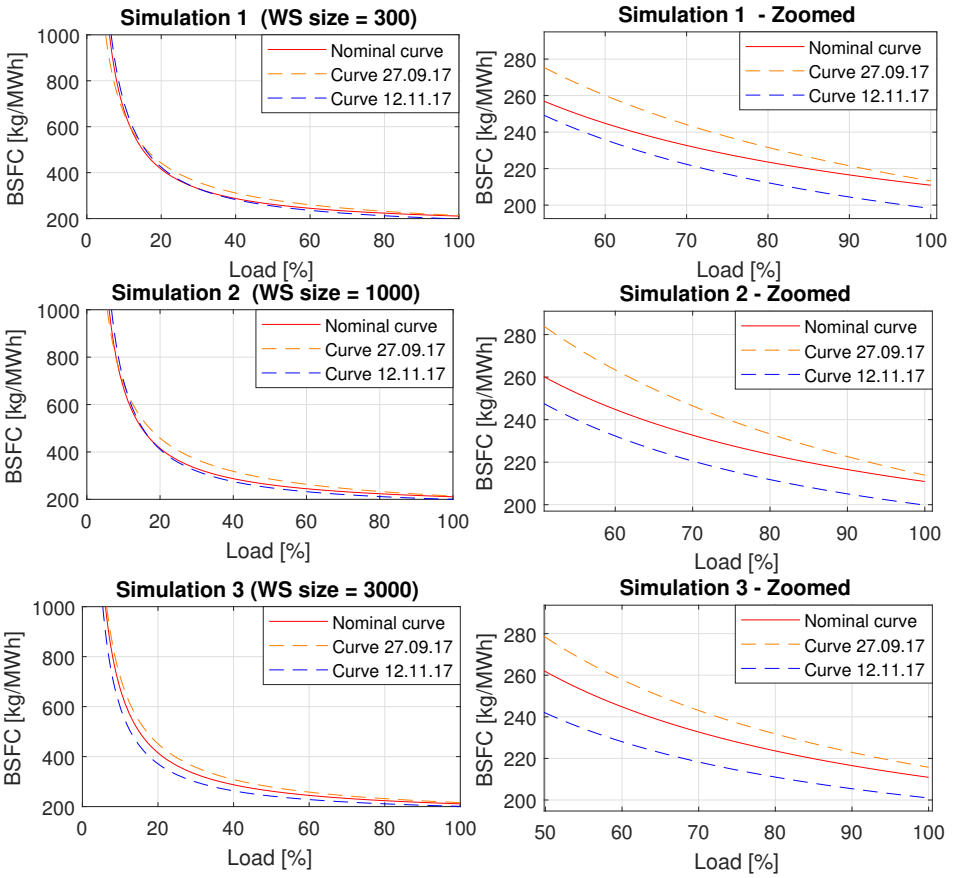

Figure B2: To the left are the resulting BSFC curves for the entire load interval (0%-100%) from three dates, to the right is the same curves but presented in a shorter interval, all for GT1.

Generating turbine 2 (GT2)

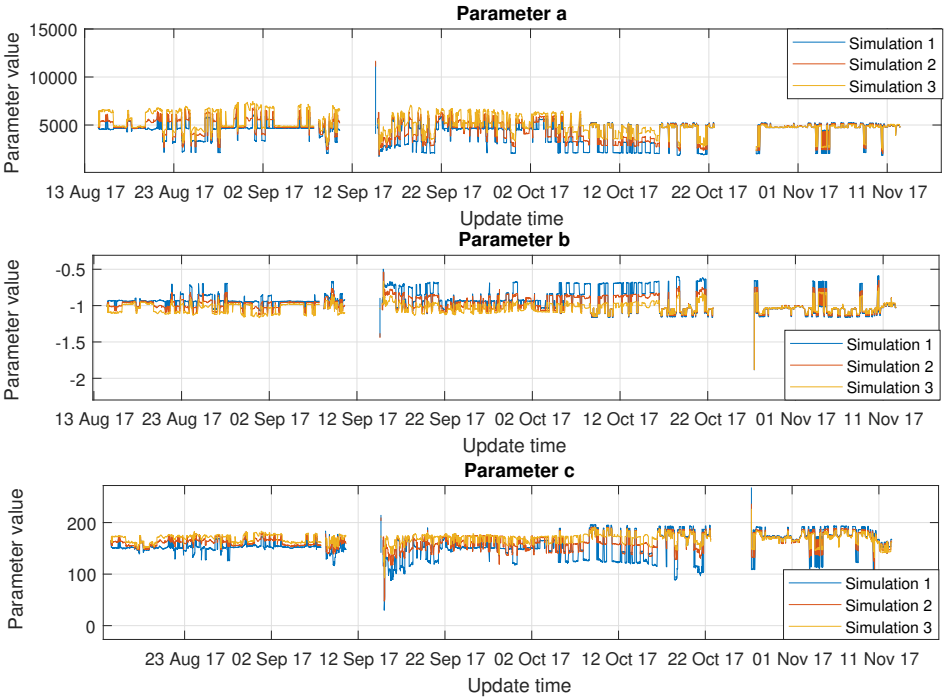

Figure B3: The estimated parameters of the BSFC curve describing the efficiency in GT2 using Algorithm B.

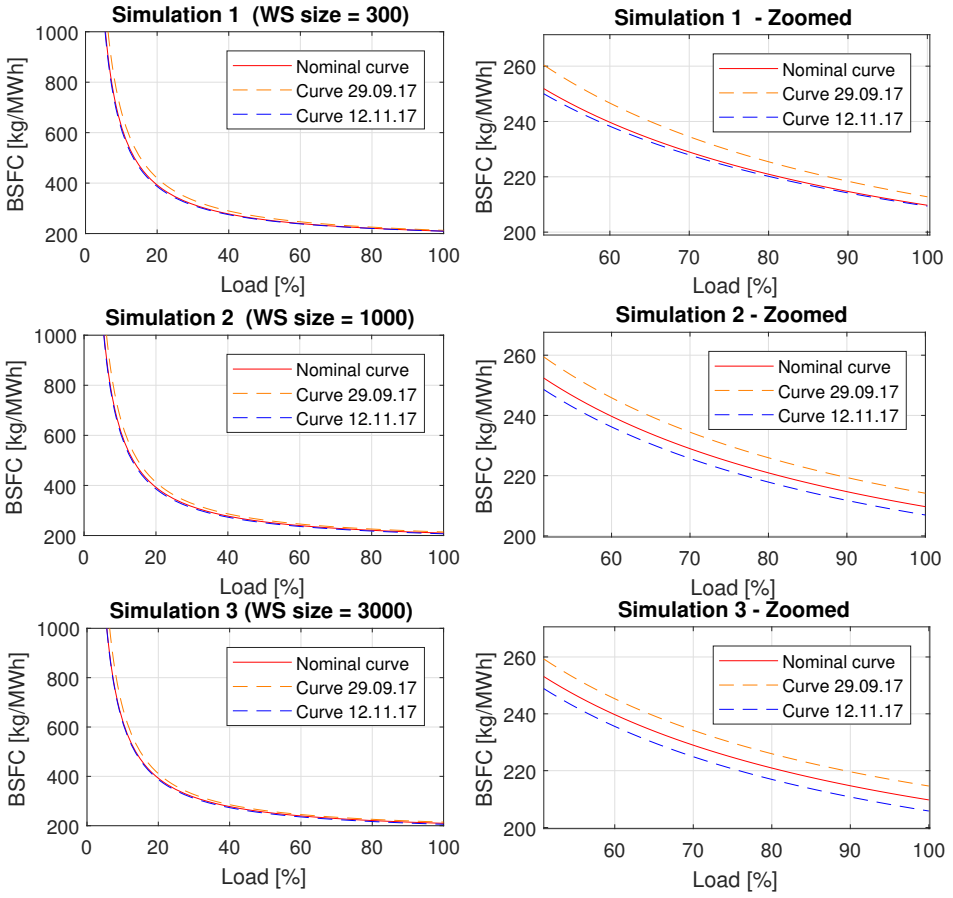

Figure B4: To the left are the resulting BSFC curves for the entire load interval (0%-100%) from three dates, to the right is the same curves but presented in a shorter interval, all for GT2.

Generating turbine 3 (GT3)

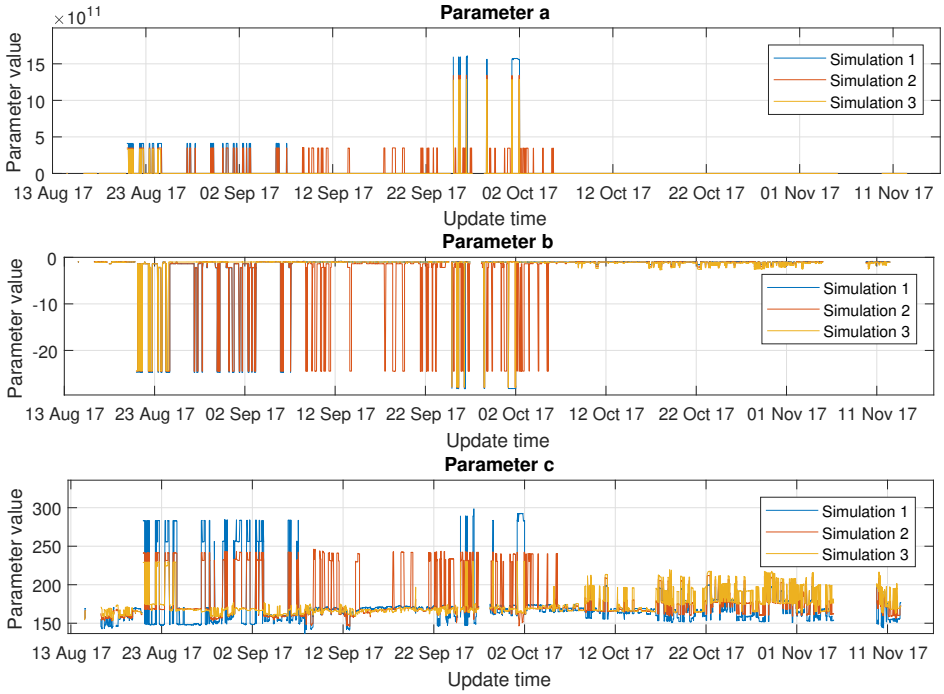

Figure B5: The estimated parameters of the BSFC curve describing the efficiency in GT3 using Algorithm B.

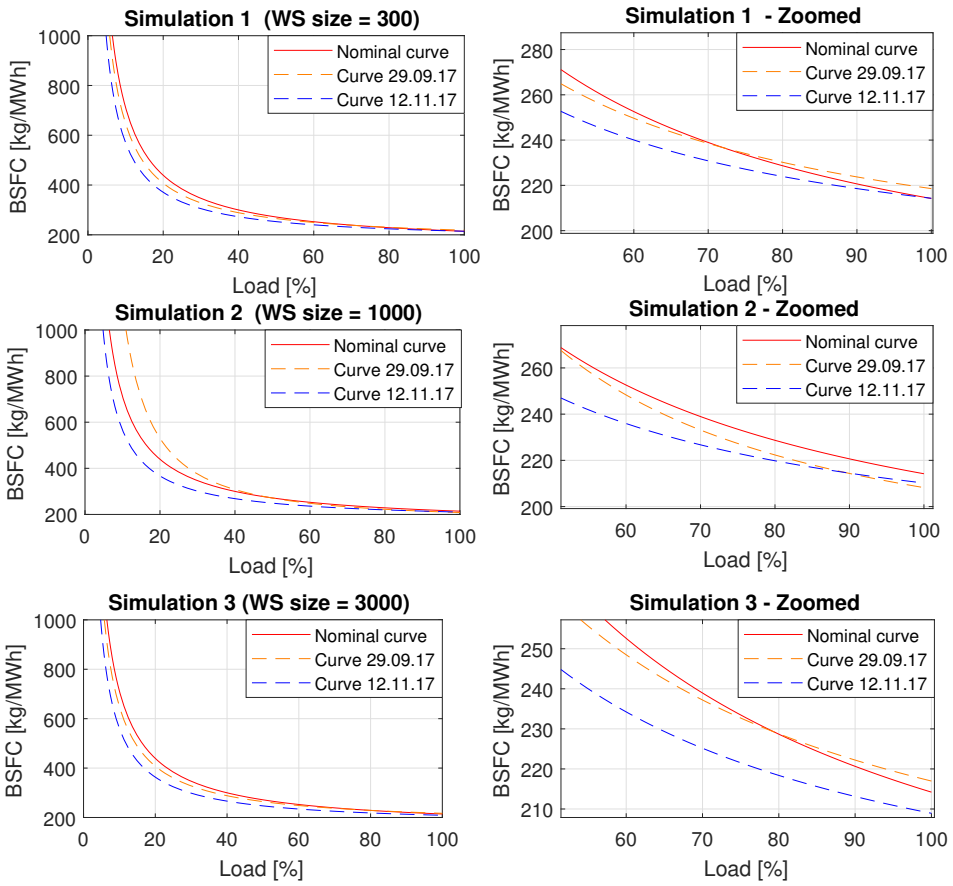

Figure B6: To the left are the resulting BSFC curves for the entire load interval (0%-100%) from three dates, to the right is the same curves but presented in a shorter interval, all for GT3.

Generating turbine 5 (GT5)

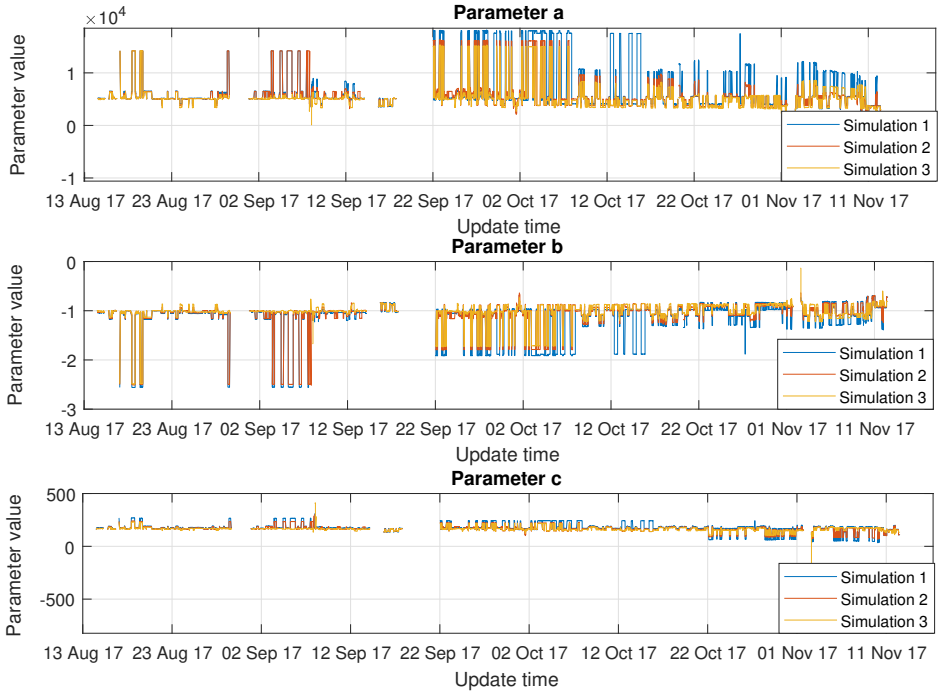

Figure B7: The estimated parameters of the BSFC curve describing the efficiency in GT5 using Algorithm B.

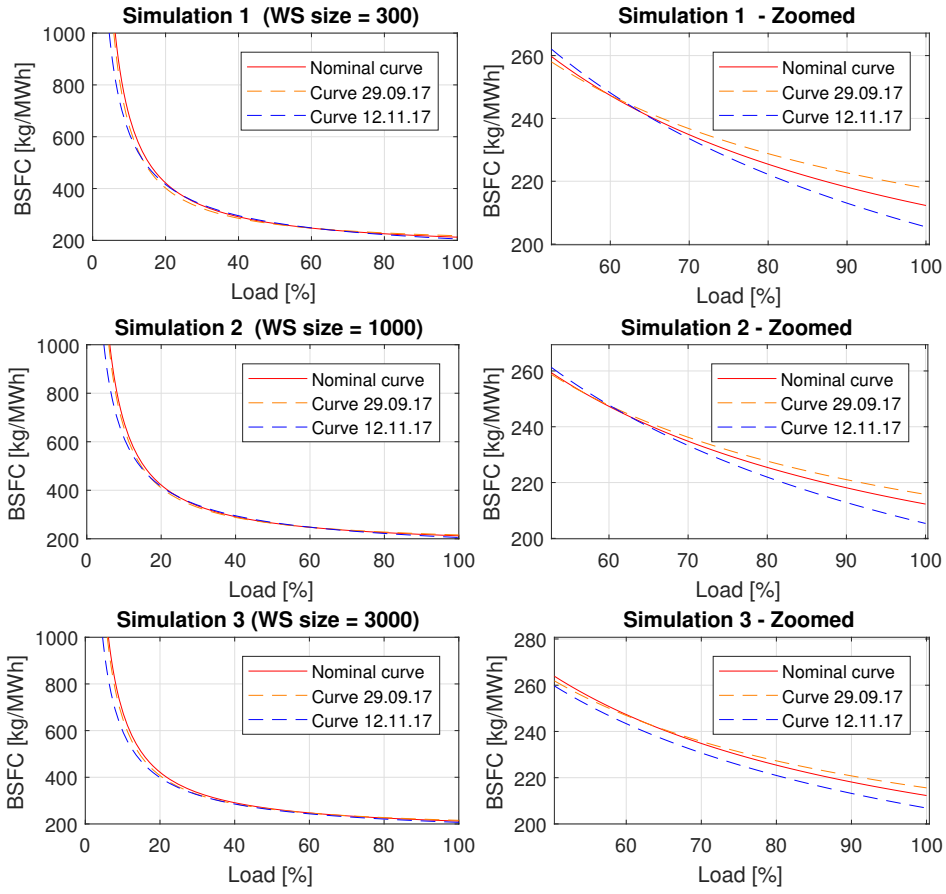

Figure B8: To the left are the resulting BSFC curves for the entire load interval (0%-100%) from three dates, to the right is the same curves but presented in a shorter interval, all for GT5.

Appendix C

Additional results from Algorithm C

The additional results from Algorithm C are presented in this Appendix, the results are presented in Figures C1-C4 for GT1, 2, 3 and 5. Each figure contains the estimated parameters $\hat{\boldsymbol{\theta}} = [p_1, p_2, p_3]^T$ in the three upper subfigures, and the resulting BSFC curves from 40%-100% in the three lower subfigures. The results from GT4 are presented in the simulation result chapter in Section 4.3.1.

One should notice that the BSFC curves are curved slightly upwards for the smaller working sets, that is $ws\_size = 300$, 1000. This is also the case in GT5, but in addition, the curve simulated with the largest working set (3000 data points) is curving upwards. The reason for this is that there are no loads in the interval above 90% in the working set, see Figure 4.1. Hence the natural shape of the second order polynomial is kept.

Generating turbine 1 (GT1)

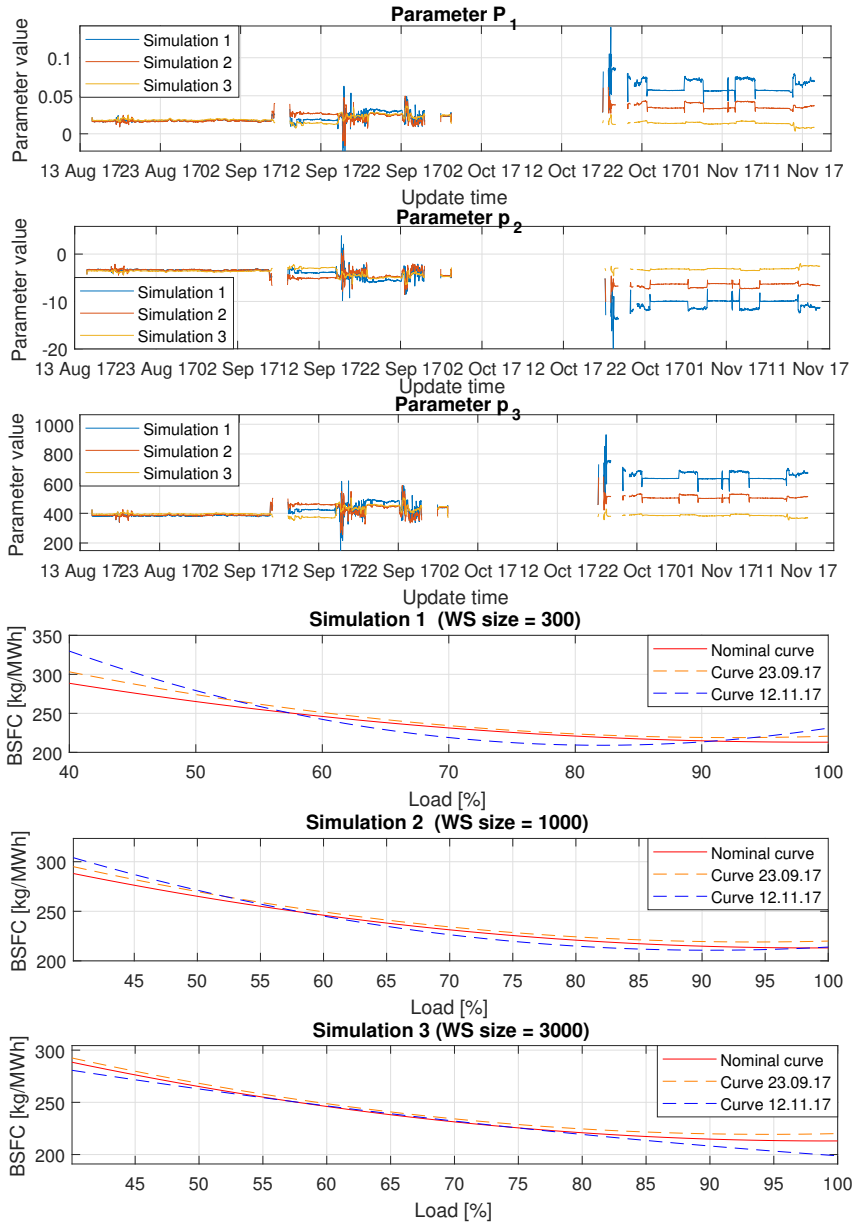

Figure C1: The estimated parameters and resulting BSFC curves from GT1 using Algorithm C.

Generating turbine 2 (GT2)

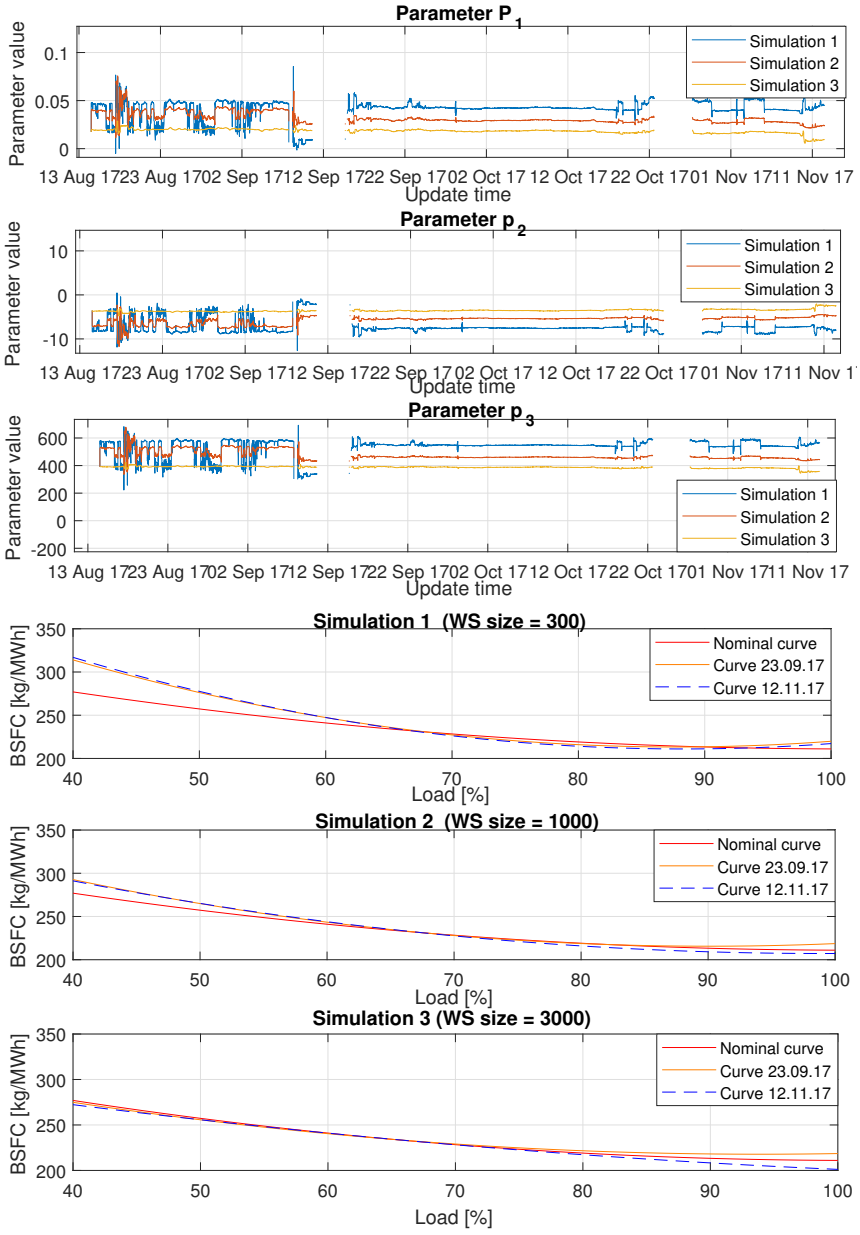

Figure C2: The estimated parameters and resulting BSFC curves from GT2 using Algorithm C.

Generating turbine 3 (GT3)

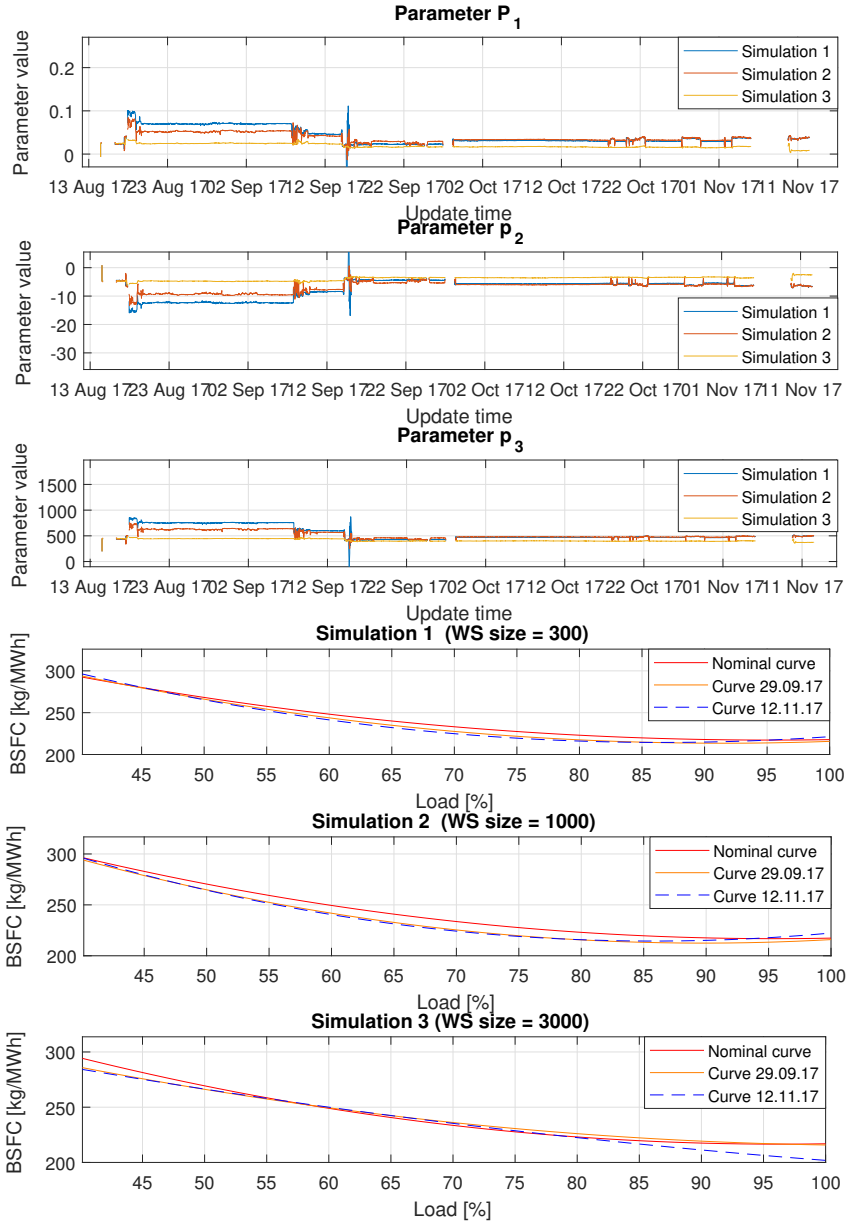

Figure C3: The estimated parameters and resulting BSFC curves from GT3 using Algorithm C.

Generating turbine 5 (GT5)

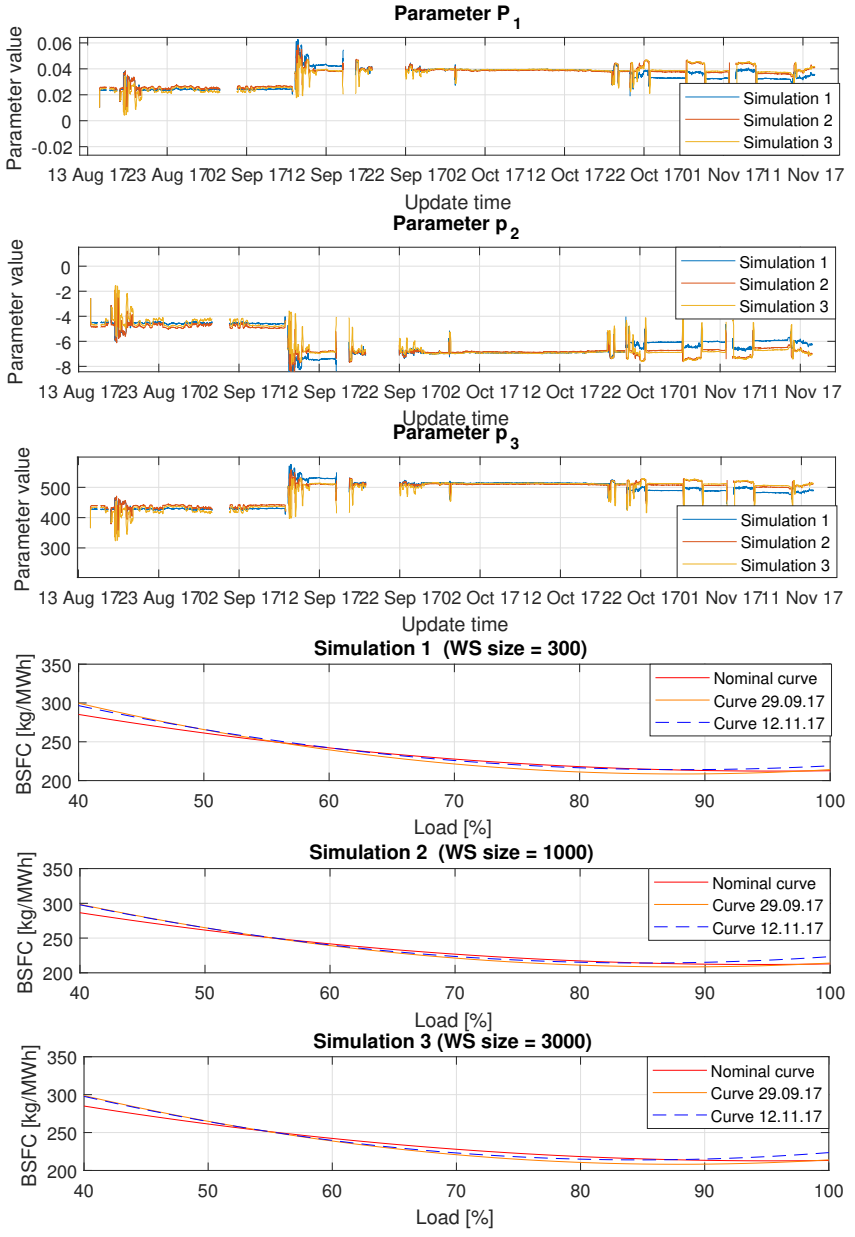

Figure C4: The estimated parameters and resulting BSFC curves from GT5 using Algorithm C.