Abstract

The future power system faces multiple challenges, such as the integration of renewable energy sources and the rapid increase in energy consumption. Flexibility is proposed as part of the solution to solve these challenges. With the integration of smart systems, the demand side could provide flexibility by reducing the consumption of multiple appliances to meet the supply. Baseline estimation refers to the estimation of the normal operation of the appliances participating in the flexibility process. However, this estimation is difficult due to the measurement of the power consumption usually are at the building level and not appliance level, and baseline estimation do not offer any information regarding the origin of the flexibility. Without the origin of flexibility, the flexibility settlement can be inaccurate. Load disaggregation is the process of acquiring individual appliance information from aggregated consumption measurements. It can provide additional information regarding the individual appliance consumption, such as a power profile for the flexible appliance and the origin of the flexible power. As there are uncertainties in both baseline estimation and load disaggregation, combining the methods can create a more accurate flexibility settlement.

This thesis examines the possibility to combine load disaggregation as an extra validation of flexibility in addition to baseline estimation. The model compares the performance of three different baseline estimation methods: a long short-term memory recurrent network method, an artificial neural network method and an averaging method. Additionally, the possibility for additional verification of demand response by load disaggregation is examined. The combination of baseline estimation and load disaggregation is tested at real costumer locations provided by ENFO AS. The extra validation could optimally provide a power profile for the flexible appliance to help determine the flexibility provided. However, as the system setup only provides active power measurements, the disaggregation method could not be sophisticated enough to provide a power profile. Therefore, the proposed disaggregation examined the possibility to determine if the reduction in power during flexibility events originates from the same appliance. If the flexibility is connected to a single appliance, it is reasonable to assume that flexibility is provided by the expected flexible appliance. The model utilize edge detection to discover changes in power consumption and differentiate the edges of the flexible events from the other changes in power consumption by dynamic time warping.

The three different baseline estimation techniques indicate that models with approximately the same error have significant differences in baseline estimation. Therefore, the choice of baseline technique could create a significantly different outcome of the flexibility settlement. Additional steps to reduce error in baseline estimation are also required to ensure that the accuracy of flexibility settlement is sufficient. In addition, due to the aggregation effect and the expected flexible appliances are multi-state, the edges of the flexible events are not similar enough, and the proposed disaggregation technique is too simple. The flexible events could therefore not be differentiated from other edges in the power consumption. By including more features such as reactive power, voltage and current, in addition to specific appliance signatures, a power profile for the flexible appliance might be estimated to assist baseline estimation and validate the origin of the flexibility.

Sammendrag

Det fremtidige kraftsystemet står overfor flere utfordringer, for eksempel integrering av fornybare energikilder og den raske økningen i energiforbruk. Fleksibilitet foreslås som en del av løsningen for å løse disse utfordringene. Med integrasjonen av smarte systemer kan fleksibilitet tilbys på forbrukersiden av kraftsystemet ved å redusere forbruket for å møte tilbudet. Baseline estimering refererer til estimering av den normale driften av lastene som deltar i fleksibilitetsprosessen. Denne estimeringen er utfordrende på grunn av at måling av strømforbruket vanligvis er på bygningsnivå og ikke lastnivå. Baseline estimering gir ikke informasjon om opprinnelsen til fleksibiliteten og uten opprinnelsen kan fleksibilitetsoppgjøret være unøyaktig. Lastdeling er prosessen med å innhente individuell lastinformasjon fra aggregerte forbruksmålinger. Det kan gi tilleggsinformasjon om det enkelte lastforbruket, for eksempel en effektprofil for den fleksible lasten og opprinnelsen til den fleksible effekten. Siden det er usikkerhet i både grunnlagsestimering og lastdeling, kan kombinering av metodene skape et mer nøyaktig fleksibilitetsoppgjør.

Denne oppgaven undersøker muligheten for å kombinere lastdeling som en ekstra validering av fleksibilitet itillegg til estimering av baseline. Modellen sammenligner ytelsen til tre forskjellige baseline beregningsmetoder: LSTM reccurent network metode, en artificial nerual network metode og en gjennomsnittsmetode. I tillegg undersøkes muligheten for ytterligere verifisering av fleksibilitet ved lastdeling. Kombinasjonen av estimering av baseline og lastdeling blir testet på reelle kundedata levert av ENFO AS. Den ekstra valideringen kan optimalt gi en effektprofil for den fleksible lasten for å bestemme fleksibiliteten. Ettersom systemoppsettet bare gir målinger av aktiv effekt, var lastdelingsmetoden ikke sofistikert nok til å gi en effektprofil. Derfor ble det istedenfor undersøkt muligheten for å koble sammen de fleksible tidspunktene for å verifisere at fleksibiliteten er gitt fra samme last. Hvis fleksibiliteten er koblet til en enkelt last, er det rimelig å anta at fleksibiliteten kommer fra den forventede lasten. Modellen foreslår å bruke kantdeteksjon for å oppdage endringer i effektforbruk og skille kantene på de fleksible hendelsene fra de andre endringene i effektforbruket ved dynamic time warping.

De tre forskjellige estimeringsteknikkene for baseline indikerer at modeller med omtrent samme feil har betydelige forskjeller i estimering av baseline. Derfor kan valget av baselineteknikk skape et vesentlig annet resultat av fleksibilitetsoppgjøret. Ytterligere trinn for å redusere feil i baselineestimering er også nødvendig for å sikre at nøyaktigheten av fleksibilitetsoppgjør er tilstrekkelig. I tillegg, på grunn at de forventede fleksible apparatene har flere innstillinger og for lav oppløsning på dataen, er kantene på de fleksible hendelsene utfordrende å skille fra andre kanter og lastdelingsteknikken for enkel. Ved å inkludere flere funksjoner som reaktiv effekt, spenning og strøm, i tillegg til spesifikke lastsignaturer, kan en effektprofil for den fleksible lasten estimeres for å hjelpe estimering av baseline og validere opprinnelsen til fleksibiliteten.

Preface

The research presented in this master thesis was conducted by Ole Andreas Sloth for the Institute for Electrical Power Engineering at NTNU in collaboration with ENFO. The master thesis is further work on the specialization project written fall 2020. Some parts of the literature review and theory are gathered from the specialization project, especially section 2.2.1-2.2.3 and 3.2-3.4.

"I would like to thank my supervisor, Jayaprakash Rajasekharan, for guidance and feedback during the thesis work. I am also grateful for the additional guidance provided by PhD candidate Surya Venkatesh Pandiyan. In additon, I would like to thank Victoira Fearnley Landmark and Morten Tylden from ENFO for providing advice on project formulation, guidance, and necessary data."

Contents

List of Figures

List of Tables

List of Abbreviations

ANN	Artificial Neural Network
CEER	Council of European Energy Regulators
DEC	Deep embedded clustering
DSM	Demand side management
DTW	Dynamic Time Warping
DTW	Dynamic time warping
FHMM	Fractional hidden Markov models
HMM	Hidden Markov models
HVAC	Heating, Ventilation and Air Conditioning
IEA	International Energy Agency
ILM	Intrusive load monitoring
LSTM	Long Short-term Memory
MLP	Multi Layer Perception
NILM	Non-intrusive load monitoring
RMS	Root Mean Square
SIQCR	segmented integer quadratic constraint programming

1 Introduction

1.1 Background

The future power system faces many challenges with the integration of renewable energy sources and the rapid increase in power demand. As more and more power generation units are renewable, distributed energy generation are increasingly becoming a part of the energy production. These alterations create a more complex power system, and the traditional way of transporting power towards consumers is changed. Power can flow in both directions and can cause reverse power flows, voltage violations, congestion and higher line losses [1]. Renewable energy sources have uncontrollable generation and can create a significant power imbalance in contrast to traditional generation, which is more controllable and predictable. In addition, the increase in power demand grows faster than the expansion of the transmission system. The demand approaches the capacity limit in peak demand hours, while a big part of the system remains idle in low demand hours. Further grid reinforcements will only be needed in peak demand hours, and due to the immense cost, be far from the optimal solution.

As more intelligent and smart systems such as smart-meters are integrated into the power system, flexibility can be a contributor to solve the challenges to the power system. Flexibility is not a unified term, IEA suggested defining flexibility as "All relevant characteristics of a power system that facilitates the reliable and cost-effective management of variability and uncertainty in both supply and demand" in 2018 [2], while CEER suggested "The capacity of the electricity system to respond to changes that may affect the balance of supply and demand at all times" the same year [3]. A generalization of suggested definitions can be "Flexibility relates to the ability of the power system to manage changes" and therefore use the flexibility term as a general term covering various parts and aspects of the power system[4]. Three main categories of flexibility are defined as supply-side flexibility, grid-side flexibility and demand-side flexibility. The different categories of flexibility cover different challenges of grid operation. Supply-side flexibility balances the load and demand in the transmission system and usually consists of altering generation to meet demand. Grid-side flexibility refers to control and level of adaptation in the grid and the grid equipment. Flexibility is provided by adjusting the physical characteristics of the distribution network. Demand-side flexibility is also referred to as demand side management (DSM).

DSM concerns the consumer side of the power system, to manage electrical loads and consumer patterns as efficiently as possible. For example, by moving consumption from peak demand hours to lower demand hours, the consumer pattern is altered more beneficially from the grid perspective. An alteration of power consumption to provide flexibility is defined as demand response [5]. The advantage of DSM is that major grid maintenance and investments can be reduced due to a more consistent load pattern without the highest peaks or lowest valleys. The increase in energy consumption might not create an equal increase in power consumption, as appliance patterns are altered to enhance grid stability. In addition, DSM might be economically beneficial for consumers by implementing compensation schemes for selling flexibility [4]. The flexibility from demand side management often comes from aggregating individual appliances. For enhanced of grid operation, a portfolio optimization of available flexible appliances are crucial [6]. The event where flexibility is delivered is to as demand response Specific appliance information such as how much flexible power

is available, for how long and appliance location are important information. After the flexibility has been offered, ordered and delivered to the market, flexibility settlement validates the flexibility provided. Measurements is compared against an estimated normal operation to verify if the offered flexibility is delivered. The estimated normal operation is referred to as the baseline.

1.2 Motivation

The baseline is an important aspect of the flexible power provision, as the difference between baseline and actual consumption is the amount of flexible power provided. The baseline also validates that flexible power has been provided because there is an unusual difference between estimated normal operation and actual operation. Baseline estimation is a well-studied research topic, and there exists a vast amount of different baseline estimation models. The most common models include regression models and machine learning models, averaging models, and meter-before/meter-after methods [7]. However, baseline estimation still has the challenge of identifying where the change in power originates from directly. It is important to verify that the change in power is determined by the flexible load and not a change in other appliances to increase the correctness of the flexibility settlement. However, intrusive monitoring on the individual flexible load is not common as the cost of monitoring equipment would be high [8]. Smart-meter monitoring of the entire household is the closest measurement to the flexible load. Smart-meter data are usually aggregated over all power consuming appliances connected to the meter. Therefore, decomposition of the measurements into individual appliance consumption is needed to gather the desired individual appliance information and this process is referred to as load disaggregation [9]. This process can locate the flexible appliance and confirm if the flexible appliance has reduced the operation as expected. Load disaggregation can create a power profile for the flexible appliance to compare the baseline estimation against. The disaggregation might also provide evidence that the power reduction is due to the flexible appliance. A combination of baseline estimation and load disaggregation can provide stronger flexibility settlement than baseline can achieve alone. Baseline and load disaggregation are both well-documented in literature, however there exists very little literature that combine both methods for additional verification.

1.3 Scope

The master thesis aims to create an accurate flexibility settlement process for small scale commercial consumers. The consumer data is provided by ENFO AS, and consists of one-minute active power consumption measurements combined with the time of demand response events. A precise estimation of flexibility creates a possibility to compensate consumers correctly, calculate the influence of flexibility on grid operation to measure the reduction in maintenance and investment costs needed. In order to validate demand response accurately, the system model presented consists of a high-resolution baseline estimation model combined with load disaggregation. Baseline estimation is a model approach and estimates the usual operation of the consumer. The high-resolution estimation is needed because of the short demand response events, and low-resolution estimation might not capture important aspects as the correct peak power reduction needed in fast response markets. However, due to high volatility in individual buildings, a near-exact baseline is challenging to achieve. To strengthen the flexibility settlement, load disaggregation is therefore proposed as additional verification. Optimally, load disaggregation would estimate individual appliance power profiles from aggregated power consumption to validate which baseline techniques are most accurate for the given demand response events. However, due to the data provided only consists of active power consumption and no additional appliance signature, it is too challenging to create a power profile for the flexible appliance. The load disaggregation process might instead validate that a change in the power consumption at the start and end of the demand response events originates from the same appliance. If most demand response events can be verified to originate in the same appliance it can be assumed that this appliance is the flexible appliance expected to be reduced.

Three different baseline estimation algorithms are developed and compared, namely an averaged method, an artificial neural network (ANN) algorithm and a long short-term memory recurrent neural network (LSTM) algorithm. The different algorithms are chosen to compare different approaches for high-resolution estimation at demand response events in the consumer data provided by ENFO. ANN have high results detecting connections between dependent and independent variables, such as weather data, and are an established load forecasting method [10]. On the other hand, long short-term memory is better at detecting temporary and long term dependencies between sequential data [11]. The averaging method is a more straightforward approach, easier to implement and is the most widespread baseline estimation method [7]. The load disaggregation technique chosen in the validation process is a combination of edge detection and matching by dynamic time warping. The disaggregation algorithm is simple and chosen because of the absence of appliance signature or multiple features such as reactive power, voltage and current that more sophisticated algorithms depend on. The baseline estimation would for the available data be the main contributor to verify the amount of flexibility and the load disaggregation to verify that the amount of flexibility is reduced by the flexible appliance.

1.4 Contributions

The validation process combines baseline estimation with a load disaggregation technique. The combination is tested as a process to provide more precise information regarding the demand response event. The contributions of this thesis could be listed as follows:

1. A comparison between LSTM, ANN and averaging method for high resolution baseline estimation. None of the models could capture the volatility of the buildings, however the averaging method are more consistent as a baseline estimation method.
2. Edge detection and comparison based on dynamic time warping for active power consumption is examined as a method to connect demand response event to the same appliance. The method was not able to connect the demand response events as more features or appliance signatures is needed.

1.5 Report Outline

To create a correct and relevant system model, chapter 2 of the thesis provides information on relevant literature for both baseline estimation and load disaggregation. Secondly, the theory needed to understand the system model are presented in chapter 3. Parts of both literature review and theory are gathered from the specialization project written in fall 2020. The system setup, as well as construction and description of the system model, are presented in chapter 4. The results of the system model on the acquired building power consumption are presented in chapter 5 and the corresponding discussion of results and future work are presented in chapter 6. Lastly, conclusion of the thesis is presented in chapter 7. The bibliography and appendix for additional system model results are added at the end.

2 Literature Review

To correctly create a model for flexibility validation, it is essential to understand the background and existing literature behind the underlying models utilized in the thesis. Therefore, a review of research papers in baseline estimation and load disaggregation is presented in the following chapters.

2.1 Baseline Estimation

Baseline estimation is the process of estimating normal power consumption in the presence of demand response events. However, as there is no direct measurement on the appliances, the calculation is performed at the lowest measurement point connected to the appliances, usually at the smart-meter level. There are different ways to estimate the baseline in literature, and the most common is the use of regression and machine learning models, averaging methods, also known as XofY days, and meter-before/meter-after methods. The regression and machine learning algorithms tend to be the most accurate [12] and are presented first. Secondly, a review of the research literature for the other methods are presented.

2.1.1 Baseline Estimation Machine Learning Techniques

There is no direct way to supervise a baseline estimation as the actual power load is not the estimation's target. However, load forecasting is the process of predicting the future power demand using currently available information. As baseline is normal power consumption, load forecasting techniques can be used to calculate baseline. The training process of the algorithms is similar, as the training data consists of regular power consumption in the absence of demand response events. The performance of the baseline techniques can be validated where the actual consumption is known to estimate an error of the baseline technique.

Load forecasting can be divided into three different tasks based on the time horizon of the forecasted data [11]. Long-term forecasting assists in infrastructure planning, while mid-term and short-term forecasting can be helpful for system operations. Short-term load forecasting is most relevant to baseline estimation as the short-term forecast are more detailed and accurate. However, most load forecasting algorithms are presented at grid or substation level [11]. Load forecasting on individual building level is more complicated due to the high volatility of the load. Recently more focus on building level forecasting has emerged, below some sophisticated methods are presented and discussed.

Support vector regression (SVR) is tested as a load prediction for baseline estimation [13]. The input variables are ambient temperature up to two hours before the dispatch event and EnergyPlus software to simulate weather parameters influence on power consumption. In addition, working schedules were added as input. The estimation has a time resolution of 1 hour, while the dataset has a resolution of 15 minutes. The decrease in time resolution causes the error to drop significantly due to a more averaged value than lower resolutions. The mean absolute percentage error (MAPE) of each hour is determined to 1.57%, scoring better than methods such as XofY and N days regression

baseline method that uses hourly energy fractions to calculate the fraction of daily load that occurs in a given hour. The use of the EnergyPlus software for building simulations complicates the model for generic use, as the software might not contain standards for buildings at all locations.

Load forecasting by artificial neural networks is a well-established technique [14]. Artificial neural networks have the ability to capture non-linearities and dependencies between independent and dependent variables. Usually, only one hidden layer is needed to capture any continuous function. Different designs for neural networks exist, and single model multivariate forecasting is the most common design for load forecasting neural network algorithms. The design of these models is to have one output neuron for each data point to predict. A drawback of this model is that the hidden layers must be very large to provide decent results. Single model multivariate forecasting is compared against a model with one output neuron, predicting the next datapoint based on the previously estimated datapoint [15]. The multivariate model proved higher results on different experiments. For individual buildings, artificial neural networks with heatning, ventilation and air condition (HVAC) setpoints are explored [16]. The error is reduced, but the method requires additional building information and the eQuest simulation tool to simulate the buildings. The use of simulation tools requires building-specific information that might not be available for general use. Neural networks are also tested on small-scale residential buildings and compared to polynomial regression, and XofY methods [17]. The neural network method has the lowest bias, while the polynomial regression has the lowest absolute error. However, as the method is tested on small residential buildings, user occupancy adds a more prominent aspect of randomness to the power consumption. The randomness is more challenging to capture in machine learning methods than the polynomial regression that compares the consumption before and after the consumption demand response event has been initiated. The neural network also emerges as a superior technique when bias is introduced as an evaluation metric in addition to error calculation [18].

A method using long short-term memory recurrent network (LSTM) load forecasting applied to both individual and aggregated residential buildings are investigated [11]. The method provides good results for aggregated units with a MAPE of 8.9% for 12 step forecasting. However, on individual buildings, the MAPE is 44%. The increase in error is due to the high volatility of the load. The LSTM method still performs better than other sophisticated methods such as backpropagation neural networks and K-nearest neighbours. The input data used are previous power consumption, time of day, day of the week and a holiday indicator. Another variant of LSTM with pinball-loss function can be applied and shows promising results [19]. The pinball-LSTM model has a different evaluation metric, and comparison with other models is problematic.

2.1.2 Other Baseline Estimation Methods

Other baseline estimation strategies are meter-before/meter-after method and averaging methods [7]. Meter-before/meter-after compares the load after a flexibility event with the load directly before to calculate the amount of flexible power. The XofY method utilizes the highest or an average of X days in a period of Y days to calculate the baseline. Deep learning might also be used for baseline estimation [20]. The paper researches baseline estimation based on probabilistic estimation using deeply embedded clustering (DEC), a deep learning clustering method. A daily pool of load profiles is collected without regard to load conditions to provide sufficient load patterns

for the deep learning algorithm to extract representative features to cluster the load patterns. An optimal cluster selection is performed on each day of each customer of the collected load patterns. Quantile regression forest models are then used to generate the estimated baseline demand during dispatch events. Data-driven clustering is tested in [21].

The criteria for an appropriate baseline estimation model is a balance between transparency, accuracy, data needs [12]. Prediction models prove to be more accurate. However, as the transparency is low, it is hard to validate the baseline in case of a dispute. In addition, prediction models need large amounts of data to be calculated. Meter-before/meter-after are more transparent and require less data, but are also less accurate than prediction due to the influence of other appliances. The last XofY lies between these two methods in all three aspects.

2.2 Load Disaggregation

Load disaggregation algorithms are dependent on the input data provided. Different disaggregation algorithms require different input data, and the most common input data varies between active power, reactive power, harmonics, current, voltage, and phase angles. As active power is the most common form of power measurement, many algorithms try to disaggregate with only active power as input data. Usually, additional information is required, either direct information about the appliance signature or by capturing power transients based on event detection [22]. Transient event refers to a switch in signal from one particular steady-state to a new steady-state. The upper limit for load disaggregation with one-hour resolution is researched to 55% even with massive deployment of sensors for appliance sub-metering [23]. Some state of the art algorithms that are based on active power measurement is discussed below.

2.2.1 Fuzzy Logic

Load disaggregation can be performed by fuzzy logic[24]. The need for reactive power is eliminated due to high power inductive appliances have a higher pulse variance than high power resistive appliances. The algorithm is composed of three parts, pulse extraction, pulse clustering and classification, and pulse to appliance association. The model used has a 0.14 Hz signal. The pulse to appliance association is tested on a residential unit and provides a load disaggregation map. However, the user must provide information about the number, type, and labelled power values of the appliances. The collection of that information is not easy to achieve on an aggregated level as many consumers do not have all the information required.

2.2.2 HMM/FHMM

Hidden Markov models are a popular way of handling the non-intrusive load monitoring (NILM) problem. NILM is the process of performing disaggregation without sub-metered measurements. An algorithm utilizing hidden Markov models (HMM) based on duration and differential observations can be used to model appliances [25]. Another way to solve the NILM problem is by modelling a single household as a sparse super-state HMM, and provide the correlation between appliance activation within the unit [26]. To model a household without the need for sub-metering, segmented integer quadratic constraint programming (SIQCR) can be employed to solve the hidden Markov model [27]. The model provided can be developed to work with the current smart-meter structure and utilize an iterative k- means method to fit a hidden Markov model with one typical duty cycle instead of weeks of sub-metered observations. As discussed, these approaches require some form of appliance signature or sub-metering as the methods are non-event based [27].

2.2.3 Neural Networks

Deep neural network architectures can be adapted to perform non-intrusive load monitoring [28]. However, to train, the models need vast amounts of appliance data as these algorithms are trained with up to 150 million parameters. The volume of training data is somewhat solved by using a 50-50 split of synthetic and real aggregated energy data. However, once the training process is completed, the algorithms can run with aggregated power measurements as input, even on unseen residential units with accurate results.

2.3 Disaggregation Based on Multiple Features

Multiple features open the door for some more straightforward methods to disaggregate. Event detection can more easily be defining for the disaggregation algorithms [22]. Event detection algorithms can be further categorized, and three main categories are expert heuristics, probabilistic models and matched filters [29]. Expert heuristics is a simple approach and create a set of rules for each appliance. The method commonly requires the initialization of specific variables to be effective, such as total power demand and power variation. Probabilistic methods calculate a probability of if the event has occurred and require a training process to fix variables and apply statistical models to appliances. Matched filters use waveform signal extraction and match to known patterns. Matched filters do not need previous training and knowledge about the appliance. However, high sampling rates are required for sufficient accuracy. The most common features to use for event detection are active and reactive power, voltage and current. Total harmonic distortion and power factor are additional features to improve disaggregation accuracy. Models using lower resolution need other disaggregation methods, for example, using general appliance models based on apparent power consumption to tune into specific appliance models [30]. However, this method requires sufficient data sets to create a general model for the target appliances. The active and reactive power signal can be split into powerlets [31]. Powerlets are short power sequences that can represent a certain signal to characterize appliances. The powerlets are stored in a dictionary for the specific device. A combination of super-state hidden markov model and a Viterbi aglorithm utilizing apparent power

and current RMS values can be used to disaggregate multi-state loads in real-time [32].

2.3.1 Load classification

After events have been detected or appliances have been modelled, load classification is performed to identify the appliances operating at a given time [22]. The classification process is based on the different types of appliances described in the list in chapter 3.3. For appliance type II, the state of the appliance also needs to be determined. Optimization methods or machine learning algorithms are the most common solvers for the load classification issue. Optimization methods for solving include hybrid programming, genetic algorithms and segmented integer quadratic constrained programming. The machine learning algorithms can be split into supervised and unsupervised techniques. Some common supervised techniques include artificial neural networks [33] and convolutional neural networks [34]. Unsupervised techniques do not require any training and are therefore most desirable for practical use. One unsupervised method is feature clustering and labelling of each cluster, using MLP neural network to classify signature based on current waveform [35]. Another way to perform load identification unsupervised is by DTW [36] Identified edges are compared with all the load signatures in the library using dynamic time warping.

3 Theory

This chapter provides some fundamental theory to understand the process behind the combination of baseline estimation and load disaggregation. Baseline estimation is crucial to verify demand response events, and a more detailed explanation about the relevant aspect of baseline estimation are presented first. As the combination of baseline estimation and load disaggregation is examined in the thesis, load disaggregation is explained in detail. Disaggregation is the process of determining single appliance information, and insight into the characteristics of the relevant appliance types are needed to create a more efficient algorithm for disaggregation. Therefore, an explanation about relevant appliance types is provided, followed by fundamental aspects of load disaggregation. Lastly, the theory of methods utilized in the disaggregation model of the thesis is explained.

3.1 Baseline Estimation

When flexible power is provided, the power pattern will differ from normal operation as one or more appliances are reduced or shut off in a period of usually higher operation. In most cases, there is often no verification of how much flexible power is given to the grid. One way to calculate the amount of flexible power is to estimate how much power the usual operation of the appliances provide and compare the calculation to the actual operation. The difference between the actual operation and the estimated normal operation will be the flexible power. Baseline refers to the usual load operation. An appropriate baseline methodology should offer a balance between accuracy, simplicity, and integrity [7]. Accuracy refers to how accurately the baseline model can estimate the normal load operation without demand response. Simplicity is essential to provide a method that can easily be used and implemented into the power grid. Integrity refers to the ability of the estimation to stop the possibility for participants to game the system by altering the power consumption prior to the demand response event to falsely increase the flexible power provided. The most common models include regression and machine learning models, averaging models, and meter-before/meter-after methods. Meter-before/meter-after is the simplest method, however, the method tends to be more inaccurate due to the influence of other appliances on the consumption data. Regression models tend to be more accurate due to taking other factors that might influence load consumption, such as temperature and weather. The averaging methods are between these models in terms of accuracy and simplicity.

To correctly trade and verify flexibility, estimating and validating the flexible power available as accurate as possible is important. In this thesis, artificial Neural networks (ANN) and long short-term memory (LSTM) recurrent networks are investigated as baseline estimation techniques as these methods are proven to be accurate load forecasting techniques. In addition, the forecasting algorithms are compared with an averaging method. A short paragraph of explanation will be provided along with different algorithms for each of these methods in this chapter.

3.1.1 Averaging method

An averaging method is based on averaged historical data to build the baseline estimation [7]. The method is also referred to as High XofY methods or Middle XofY methods. The method considers the Y most recent days prior to the dispatch event and selects the X most fitting days based on some criteria. The most common criteria include selecting the days with the highest consumption or days with the median consumption. The Y days do not include every day prior to the event, as some days are not eligible to use due to characteristically different load patterns. It is common to exclude holidays and other days with demand response events and separate weekdays and weekends. In addition, some models also exclude days with an average consumption less than a given threshold. If information such as scheduled shutdowns is available, these days can also be considered to exclude. Once the eligible Y days have been selected, the group of days are narrowed down to X days to acquire a better representative of days to average over. Usually, the X days with the highest average consumption during the dispatch event time of day are selected. However, it is also common to use the days with the middle consumption instead of the highest, especially when the DR events do not happen at peak demand hours. The load for the selected X days are averaged to create a baseline estimation.

The baseline is often adjusted to fit the DR day properly, as most days deviate from the average. A timeframe of 2-4 hours is commonly used to adjust the baseline. More than one hour is needed to be a representative deviation from the baseline, and more than four hours are seen as too far away from the event to be representative. The actual load is compared to the baseline load and adjusted either by a scalar or additive approach. The scalar technique is based on the percentage difference, and the baseline is adjusted by the percentage difference between the average value over the timeframe. The additive technique is similar but use the kW value instead of the percentage when adjusting the baseline. The scalar technique can produce a too large adjustment when the consumption is low, and the additive technique is recommended.

3.1.2 Artificial Neural networks

Artificial neural networks can be defined as a connected array of elementary processors defined as neurons [37]. The model is designed to resemble a human brain with many neurons interconnected in a complex, non-linear and massive parallel network. The most common model is the multilayer perceptron (MLP) and consists of an input layer, one or more hidden layers, and an output layer [38]. The MLP is categorized as a supervised learning algorithm due to the need for a target output to learn. Each layer consists of multiple neurons, and each neuron is connected to an adjacent layer with weights. An artificial neuron is a neuron that performs a simple mathematical operation on the inputs. These mathematical operations are referred to as activation functions. The activation function is the enabler to get a neural network to represent more complex and non-linear relationships.

Common activation functions include linear activation function, Sigmoid activation function, tanh function, RELU function and softmax activation function. The linear activation function is proportional to the input and can be defined as:

The Sigmoid activation function can be defined as:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{3.1}$$

The ReLu activation function can be described as:

$$R(x) = max(0, x) \tag{3.2}$$

where R(x) is the activated value. The function is 0 when x is below 0 and equals x when the value is above 0. The tanh function is another common activation function and can be described as:

$$tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{3.3}$$

Multiple hidden layers might cause overtraining, and due to the vanishing gradient problem, the training gets increasingly more complex with more hidden layers [39]. To easily overcome these problems, shallow ANN topologies with one hidden layer are generally used. The advantage of neural networks is the ability to represent linear and non-linear relationships and learn these relationships directly from the data.

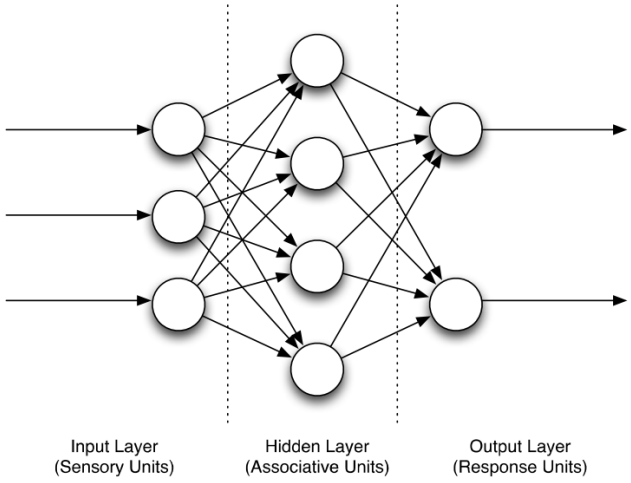


Figure 1: Illustration of a MLP model with one hidden layer [10]

3.1.3 RNN

Recurrent neural networks are networks with a fundamental difference from traditional neural networks. The models are sequence-based which can detect temporary correlations between data close in sequence [40]. Instead of using a fixed number of input vectors as traditional neural networks use, the RNN architecture can use all of the available input information up to the current time frame to predict the value at the current time. RNN have the same activation functions as traditional neural networks, with tanh and sigmoid being common activation functions. A simple RNN using the previous value can be described as [41]:

$$h_t = f(h_{t-1}, x_t) \tag{3.4}$$

where h_t is the single hidden vector in time t and x_t is the input values at time t. An activation function, for example the tanh, is applied:

$$h_t = tanh(W_{hh}h_{t-1} + W_{xh}x_t) \tag{3.5}$$

where W_{hh} is the weight at the previous hidden state and W_{xh} is the weight at the current input state. The output is the given by:

$$y_t = W_{hy}h_t \tag{3.6}$$

where y is the output and w_{hy} is the weight at the output state.

The advantage of RNN compared with traditional ANN is that data is modelled to depend on previous information. The way the data is modelled also acquire some challenges. The vanishing and exploding gradients are common problems while training RNN [42]. The vanishing gradient problem refers to when the norm of the gradient for long-term components decrease exponentially fast to zero, with the result being difficulties with learning long-term dependencies in the data [11]. The gradient exploding problem refers to the opposite where gradients accumulate and might influence the updating of weights in the training period causing oscillating weights and not capturing dependencies correctly.

3.1.4 LSTM

Long short-term memory is an architecture applied to RNN to overcome the problem of vanishing and exploding gradients [43]. The LSTM architecture replaces the regular nodes in the hidden layer with memory cells. Each memory cell contains a node with a self-connected recurrent edge with a fixed weight equal to one, ensuring that the gradient can pass across many time steps without vanishing or exploding. The memory cell consists of simpler nodes in a specific pattern. The simpler nodes are specified as input node, input gate, internal state, forget gate and output gate. The gates are sigmoidal units and pass the information along if the value is one and cuts off the information if the value is 0.

1. Input node: Data for time t is gathered in a standard way from the input layer using the activation function and is combined with the data from the previous time step h_{t-1}. The combined data is then activated commonly using the Tanh activation function.
2. Input gate: If the value is one, the information from the input node flows through. However, if the value is 0, the information is cut off.
3. Internal state: The internal state is the node where the recurrent edge is connected. This edge allows the error to flow across time steps without exploding or vanishing because the node spans adjacent time steps with constant weight.
4. Forget gate: Method to forget the influence of the past time step.
5. Output gate: The internal state is multiplied with an output gate to produce the output value. The internal state is commonly run through an activation function before being multiplied with the output gate.

An overview of the memory cell is presented in figure 2

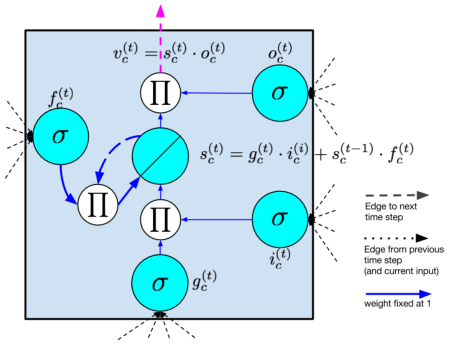


Figure 2: Illustration of a memory cell [43]

3.2 Appliance-specific information

Appliance specific information is needed to analyze the results of load disaggregation more accurately. In the buildings where the system model is applied, the flexible appliances are ventilation, electric water heater and battery. An introduction to the flexible appliances is given below.

3.2.1 HVAC

Buildings account for almost 40% of energy consumption worldwide [44] and approximately half of a buildings energy consumption are used in heating, ventilation, and air-conditioning systems [45]. Therefore, optimal load management of these loads can provide large amounts of flexible power on an aggregated scale. Due to the thermal inertia of buildings, indoor temperature can be controlled within a comfortable zone and create periods where the HVAC loads can be exploited as flexible power. However, HVAC power consumption is multi-state and are more challenging to disaggregate as the flexible power provided greatly depend on which state the appliance is operating in before the demand response and which state the appliance is allowed to be reduced to.

3.2.2 Electric water heater

Electric water heaters are common in all building types and offer a significant thermal storage potential on an aggregated scale. An electric water heater senses the water temperature and a heating element maintains the temperature at the desired setpoint. A deadband of up to 10 degrees celsius is often implemented to avoid too frequent switching. As a precaution of the proliferation of Legionella, the water heater should not be stored at too low temperatures, which is an important consideration when applying water heaters as demand response units. The operation if an EWH can be presented in three steps [46]:

1. If the hot water of the previous time step is less than the minimum temperature $x(t-1) < x_{min}$ then the thermostat in step t is ON $m(t) = 1$.
2. Else, if the water temperature of the previous time step is higher than the maximum temperature, $x(t-1) > x_{max}$ then the thermostat in time t is OFF $m(t) = 0$.
3. Else maintain the thermostat state $m(t) = m(t-1)$.

3.2.3 Battery

Batteries can be divided into two groups, batteries that only can be charged from the grid and batteries that can discharge power back to the grid. The charging time is related to the battery's state of charge (SOC), charging power, charging efficiency and can be expressed as follows [47]:

$$t_f - t = \frac{C_B(1 - SOC)}{P_{charging} * \eta_{charging}} \tag{3.7}$$

where $t_f - t$ is the required charging time, C_B is the capacity of the battery, SOC is the current SOC of the battery, $P_{charging}$ is the charging power, and $\eta_{charging}$ is the charging efficiency.

The different SOC of the battery creates multiple ways to provide flexibility. The battery can have a SOC at 100%, and discharge is the only viable option for use in demand response. If the SOC is greater than 0% and less than 100%, the battery has the option to stop charging to provide additional flexible power. However, as there is less power available at the battery, the SOC might reach 0 % before the demand response event ends, and only delay of charging can be used for flexibility. If the SOC is 0% at the beginning of the demand response event, no dispatch is given, and delaying charging is the only viable option for flexibility. The different use of battery create a complex disaggregation analysis, and a completely accurate model is challenging to achieve without detailed information about the different aspects of the specific battery.

3.3 Load disaggregation

Load disaggregation is the process of obtaining individual load information from an observation point containing multiple loads. There are mainly two classes of load disaggregation, intrusive load monitoring (ILM) and non-intrusive load monitoring (NILM) [48]. ILM requires low-end metering devices applied close to the monitored load, which are a costly investment on an aggregated scale [8]. NILM does not require additional metering devices except smart-meters, which are already deployed in most buildings.

The concept of NILM dates back to the 90s. As more real-life data, methods to handle large amounts of data, and smart-meter infrastructure are developed, NILM can be deployed at a large scale in the near future. The general framework is presented in figure 3. The goal of NILM is to partition the aggregated active power data from a building into appliance level power data without the excessive need for additional metering devices. Electrical loads exhibit a unique power signal pattern referred to as the appliance signature. The methodology of NILM is to use the appliance signature to recognize the operation of the appliance from the disaggregated power data.

Appliance categories are divided into the following categories [49] with typical load profiles.

- Type-I: Appliances with only two states of operation, ON and OFF. Examples are some electric water heaters, fans etc.
- Type-II: Multi-state appliances with finite operating states. Examples are washing machines, dishwashers, dryers etc. These multi-state appliances have repeatable switching states, which makes disaggregation simpler.
- Type-III: Continuously variable appliances are appliances with no fixed number of states. Laptops, TVs are examples belonging to this category.
- Type-IV: Permanent consumer devices are appliances that remain active for days or weeks, and examples include smoke detectors, cable-TV receivers, etc.

The appliance signature category is essential to distinguish the kind of signature disaggregation algorithms are searching for. For example, the knowledge that multi-state appliances have repetitive transition states offers more specific searching criteria for the disaggregation algorithms.

The different steps of the NILM framework are data acquisition, appliance feature extraction, and inference and learning. The data acquisition step is to acquire aggregated data at an adequate resolution to recognize appliance signatures. The data is usually collected from a building smart-meter, and resolution varies from 0.01 Hz to 600Hz [49]. Feature extraction handles the raw data processing, detects events such as appliance state transition and steady-state operation. Then the appliance features are analyzed by identification algorithms in order to identify appliance states. The algorithms are divided into supervised and unsupervised, where supervised algorithms use data from individual appliances to train. As the current smart-meter infrastructure does not measure individual appliance data, training data is a significant obstacle in the large-scale implementation of NILM.

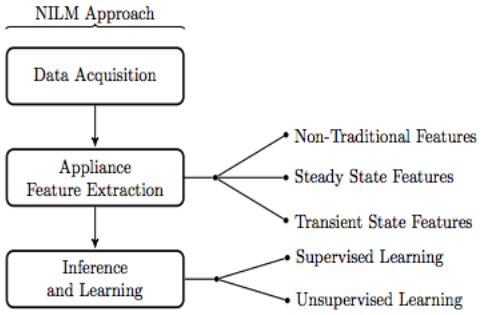


Figure 3: General framework for NILM [49].

3.4 Change point detection methods

In the system model, the load disaggregation algorithm is constructed based on the information from the available buildings. The information only contains active power measurements, the disaggregation provided is restricted to edge detection and a comparison technique. As the edge detection is performed on a single feature, change point detection is an efficient method to detect the edges in the power consumption. An introduction to change point detection is provided, followed by an explanation of the change point detection method preferred in this thesis.

Changepoint detection is the task of finding changes in time series data or signals. In the context of power consumption, changepoints indicate a switch between load states or a new load is turned on or off. Changepoint detection is a well-studied area, and there exist many different methods. Standard change point detection algorithms are binary segmentation method, segment neighbourhood, the optimal partitioning method, and Pruned exact linear-time method [50]. As segment neighborhood and optimal partitioning are exact algorithms, the computational cost is relatively high, with $O(n^2)$.

Changepoint detection assume that there exists an ordered sequence of data $y_{1:n}$. The methods will detect m change points together with positions $\tau_{1:m}$. τ_0 is defined as 0 and τ_{m+1} is defined as n. M change points will divide the data into m+1 segments, where the ith segment contains $y_{(\tau_{i-1}+1):\tau_i}$. Usually the goal is to minimize

$$\sum_{i=1}^{m+1} C(y_{\tau_{i-1}+1:\tau_i}) + \beta * f(m) \tag{3.8}$$

where C is a cost function for a segment and $\beta * f(m)$ is the penalty to guard against over fitting. The negative log likelihood the most commonly used cost function for change point detection and the most commonly used for the penalty function is one which is linear in the number of change points i.e . $\beta * f(m) = \beta * m$.

3.4.1 Binary segmentation search method (BSSM)

Binary segmentation methods are the most established change point detection method and date back to (1974)[51]. The methods extend a single change point method to multiple by repeating the single changepoint methods on subsets. Initially, a single changepoint method is applied to the entire data set. This means that a test to see if a τ exists that satisfies:

$$C(y_{1:\tau}) + C(y_{(\tau+1):n}) + \beta * f(m) < C(y_{1:n}) \tag{3.9}$$

If equation 3.9 is false, no changepoints are detected, and the changepoint search stops. However, if the equation holds, the data is split into two segments, one segment before the identified changepoint and one after. The detection method is applied to the two new segments. The advantage of binary segmentation is that the method is computationally efficient with an upper bound for running time

$O(nlogn)$. However, since the method is approximate, there is no guarantee to find the global minimum of equation 3.8

3.5 Simple disaggregation algorithms

There exist some simple disaggregation algorithms to provide faster results with limited data. These algorithms have lower accuracy than the most advanced HMM or neural network algorithms, but are easier to implement and require less information. Examples of simple algorithms for disaggregation are different matching techniques that use template matching instead of model training [52]. The most common matching algorithm is dynamic time warping (DTW) , and this method is utilized in the system model. A short explanation of the DTW theory is provided below.

3.5.1 Dynamic Time Warping

Dynamic time warping refers to a method to align the signals regardless of length. The result is a minimum accumulated cost or distance between the two signatures [52]. DTW utilitize a recursive updating rule to identify the optimal warping path given by eq 3.11. DTW can be defined as:

$$D(P_a, P_b) := min_{A \in A_{n,m}}[A, \Delta(P_a, P_b)] \tag{3.10}$$

where $A_{n,m}$ the alignment score of n and m. $[A, \Delta(P_a, P_b)]$ is the inner product of the alignment matrix A and the cost matrix $\Delta(P_a, P_b)$.

$$D(n, m) = \delta(p_a^n, p_b^m) + min(D(n-1, m), D(n-1, m-1), D(n, m-1)) \tag{3.11}$$

where p_a^n is the power of P_a in step $n = 1, ..., N$ and p_b^m is the power of P_m in step $m = 1, ..., M$. $\delta(p_a^n, p_b^m)$ describes the distance metric, and can for example be the Eucledian distance, Manhattan distance or the Kullback-Leibler distance. D(N,M) will give the accumulated cost or distance of the comparison. The initial conditions are set as follows: D(0,0) = 0, D(n,0) = ∞ for $n > 0$ and D(0,m) = ∞ for $m > 0$

4 System Model

4.1 System setup

The data set used in this thesis are provided from ENFO AS and consists of main-meter active power data readings from 6 different buildings at different locations from October 10th 2020, to February 12th 2021. In addition, ENFO develops hardware connected to the appliance that can engage some appliances in a controlled dispatch mode for demand response, either reduced operation or a shut down of the appliance. The instances of controlled dispatch modes are referred to as demand response events. A short description of the dispatch appliances used for flexibility is provided. It includes the appliance type, maximum activation time, rest time, maximum flexible power, for some buildings usual flexible power, and the number of demand response events for each site. The main flexible appliance types include ventilation, battery and water heaters. Each building has one primary flexible appliance, and an overview of the flexible appliance set up for each building are given in table 4.1.

Building nr.	Location	Flexible appliance	Max flexible operation [minutes]	Rest time [minutes]	Max flexibility (usual flexibility) [kW]	Number of DR events
1	Grimstad	Ventilation	60	0	300	3
3	Hisøy	Ventilation/heating	240	60	200	41
5	Kristiansand	Battery	60	60	2000 (200)	40
6	Halden	Electric water heater	240	0	900	1
7	Arendal	Ventilation	240	0	150	60
8	Arendal	Ventilation	-	-	-	5

Table 4.1: A summary of buildings with the respective flexible appliances. The building numbers are given from ENFO.

4.1.1 Data Acquisition

The main-meter power readings and the demand response event file are given as json files. The main-meter file consists of measurement timestamps, the sensor ID of the measurement device and the measured active power values at a one-minute resolution. The demand response event value is given as either a 0 or 1, depending on the state of the appliance. A value of 0 indicates normal operation, while 1 indicates dispatch operation. Each value has a corresponding timestamp connected.

4.1.2 Data Pre-processing

Some data processing techniques are applied to get the data in a format possible to use for initial analysis. There are instances where more than an hour of measurement data are missing, and the entire day is removed from the data set. In instances where less than an hour of data are missing,

linear interpolation was applied to fill the gaps. As there is one reading per minute, the timestamps are changed to each whole minute. The demand response event file is cleaned to only contain timestamps where the appliance changes state. In more detail, timestamps where the state has not changed are removed, and instances where the state changes two times in less than a minute are removed. The power usage at each demand response event is studied to interpret the changes in power usage in regard to flexible appliances. The buildings where the flexible appliance not clearly affect the power consumption are not further used in the model because the baseline estimation would mainly depend on the other loads. The error for one-minute resolution estimation will have a more considerable impact than the reduction of the flexible appliance, and the results will be too uncertain to draw useful information. In addition, to eliminate noise, the moving average technique with a window size of ten were applied to the power consumption. However, as edges get stretched, they are harder to detect and the smoothing was not applied to the load disaggregation.

4.2 Building Selection

Some requirements must be met for buildings to be eligible for demand response validation with the system model. A sufficient amount of demand response events are needed to determine the performance of the baseline estimation over a representative number of events, and the threshold is set to five demand response events in this model. The ratio of flexible power to aggregated power should be notable not to have the demand response events overshadowed by other appliances. Suppose other loads are too power consuming and volatile. In that case, the baseline estimation will greatly depend on the ability to catch this volatility. The baseline error might be more extensive than impact the demand response event altogether. If these requirements are not met, the building is dropped from the system model.

4.3 System Model

This thesis's system model examines the possibility of creating a more robust and more accurate demand response validation by combining baseline estimation and load disaggregation. The baseline estimation model estimates the normal operation to calculate the amount of flexible power and verify that flexible power is consistent during the demand response event. In addition, the possibility to verify that the demand response events are connected to the flexible appliance by load disaggregation is examined. In this way, the validation model can validate the origin of the flexible power to create more transparency in the flexibility settlement process. As there is no signature from the flexible appliance provided, the system model examines the possibility to determine if the demand response edges originate from the same appliance and can be differentiated from the other edges of the data. Suppose most demand response edges can be connected and differentiated from the other edges in the power consumption. In that case, it is reasonable to expect that the demand response events offer a reduction from the same appliance, and it can be assumed that this appliance is the expected flexible appliance. The following chapters describe the model construction.

4.4 Baseline Estimation Model

Three different baseline estimation methods are implemented and compared, two machine learning methods and one averaging method. The machine learning methods are an artificial neural network method and a long short-term memory recurrent neural network method. The averaging method is based on averaging similar days to estimate the baseline. The three methods are described in the next section. After the methods have been created, a baseline for each demand response event is estimated by each method. The baselines are compared with the actual load and visual inspection about performance together with the error estimation by mean absolute percentage error (MAPE), mean absolute error (MAE) and root mean squared error (RMSE), evaluates the performance of the baseline methods.

4.4.1 Machine Learning Methods

The machine learning methods use historical data to estimate the baseline. The process requires more data processing, and the demand response events were removed from the data to create a normal operation load estimation. Day of the week, week number, hourly temperature are added to the dataset. The hourly temperature is collected from Norsk Klimaservicesenter [53], and the input variables are summarized in table 4.2. The machine learning methods require normalized data to better calculate the dependence of the input data, and min-max normalization is therefore performed on the data set. To improve forecasting, the methods also rely on the power consumption directly previous to the estimation period. By including the previous power consumption, the methods are tuned to fit the day of the demand response more accurately.

Input variables
Previous values
Day of week
Week nr
Hourly temperature

Table 4.2: Input Variables for machine learning algorithms.

Different parameters are tuned to improve the accuracy of forecasting. For the LSTM method, the number of layers and neurons do not affect the performance of the method as long as there are multiple levels and a sufficient number of neurons [11]. The number of layers were set to two and the number of neurons to 100. Artificial neural networks work best with one hidden layer and one output layer to prevent the vanishing gradient problem [10]. The number of epochs and neurons are tuned to minimize error. The data set is split up into training and validation with a 90/10 percentage split to validate the methods and estimate error. After training, consumption estimations are applied to the 10% validation set and validated by the error estimation metrics root mean squared error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). An overview of how the construction of the machine learning techniques are illustrated in figure 4. Both methods are implemented in python. The TensorFlow package is utilized for the

method setup, training and estimation, while SciKit-learn and SciPy are used for processing and evaluation.

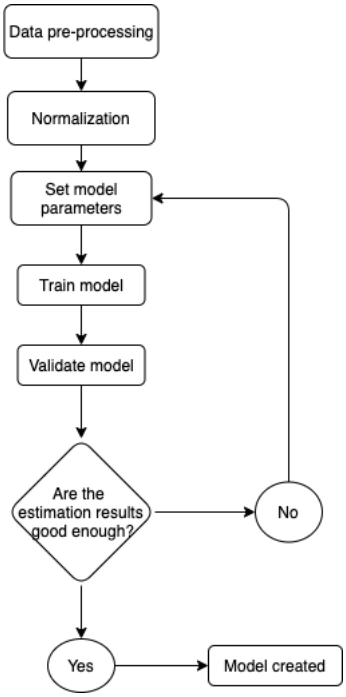


Figure 4: Flowchart of baseline estimation by machine learning approach.

4.4.2 Averaging method

The averaging method is based on the theory in chapter 3.1.1. The possible Y days are created by removing holidays, weekends, and days with demand response from the data set. The exclusion of days is implemented because it is typical for these types of days to have a different load pattern from regular days. When these days are removed, the last Y days remaining before the demand response event are selected. If there are not enough days previous to the event, some days later are added to fill the requirement of Y days due to the small size of the data set. To select the X days, the load in the demand response event period is averaged and compared. As the events usually do not happen at peak demand, the X days with the middle average load in the demand response period are selected. From these, an average load profile is extracted. To adjust the averaged load to the day of the event, the 2 hours before the demand response event are used as an adjustment. The actual power consumption in this time window is averaged and compared to the averaged value of the same period in the estimated load pattern. The difference between the compared values is added to the baseline such that the baseline is adjusted to properly fit the day of the demand response.

4.5 Disaggregation approach

The purpose of disaggregation is to determine if it is possible to verify that the power reduction during demand response events originates from the same appliance. In this case, it is reasonable to assume that the power reduction originates in the expected flexible appliance. The edges from the first selection of demand response events are collected as appliance signatures. Binary segmentation edge detection is applied to the remaining demand response event days, and the edges are stored to compare with the collected edges from the training period. Min-max scaler is used on both appliance signature and the edges from the edge detection. The change in power becomes more apparent after scaling, and the comparison ignores the baseload. Dynamic time warping is applied to the demand response appliance signatures and the located edges from the edge detection algorithm to look for similarities. The advantage of DTW is that the similarity can be detected without equal length on the signals compared. To be able to compare signatures of unequal length is important because the edge detection algorithm detects the largest change between two data points. The detected location can either be the start or the end of the edge or somewhere in between. Therefore the signal from edge detection must include the length of the appliance signal in both directions of the detected edge. Therefore, the extracted signal from the edge detection will be twice as long as the signal from the training period. Suppose the edges can be differentiated from the edges around the event, and the demand response start and end edges are similar for most of the events. In that case, it is confirmed that the same appliance gives the demand response, and it is possible to assume that the power reduction is given by the expected flexible appliance. The disaggregation process is presented in figure 5.

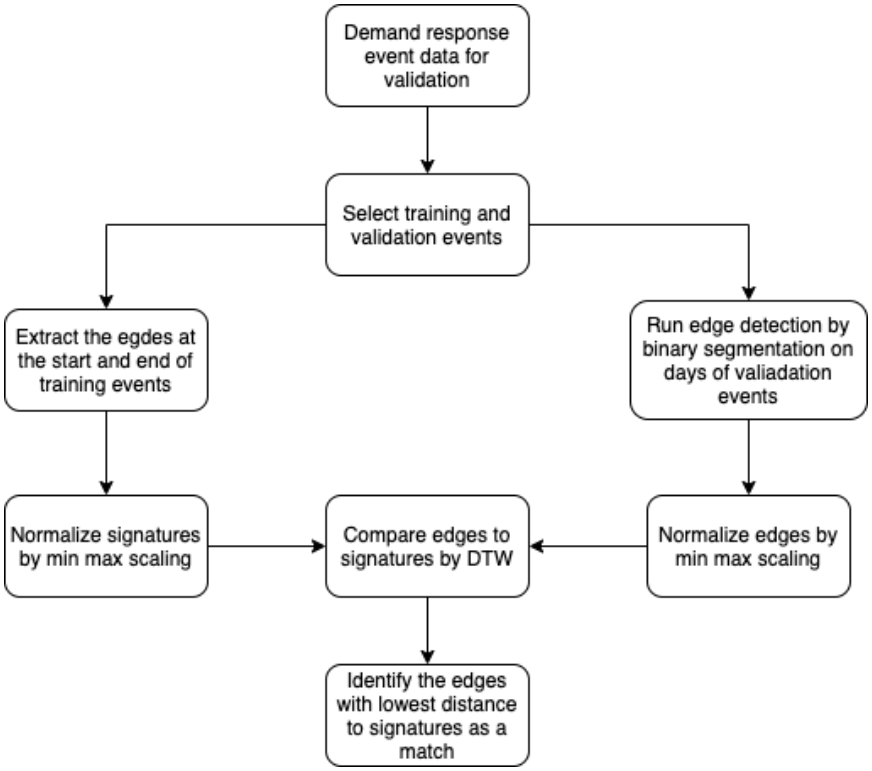


Figure 5: An overview of the disaggregation process

4.6 Summary

The system model is constructed to help create a more accurate flexibility settlement. The system model combines baseline estimation and load disaggregation to have a more robust validation of demand response events. However, due to some limitations in the data, an accurate power profile is too challenging to achieve. Instead, disaggregation is performed to examine the possibility to connect edges of the power consumption to the flexible appliance. This can help ensure that the given flexibility are from the flexible appliance and to prevent gaming of the flexibility settlement. However, as the power profile is too challenging to calculate, the disaggregation cannot help determine which estimation technique to choose for the given demand response events. An overview of the complete system model is presented in figure 6. The following chapter provides the results from the system model. Initially, demand response event analyses for each building are presented to determine which buildings eligible for the demand response validation process. The results from the baseline estimation are presented next, and lastly, the results of the load disaggregation process.

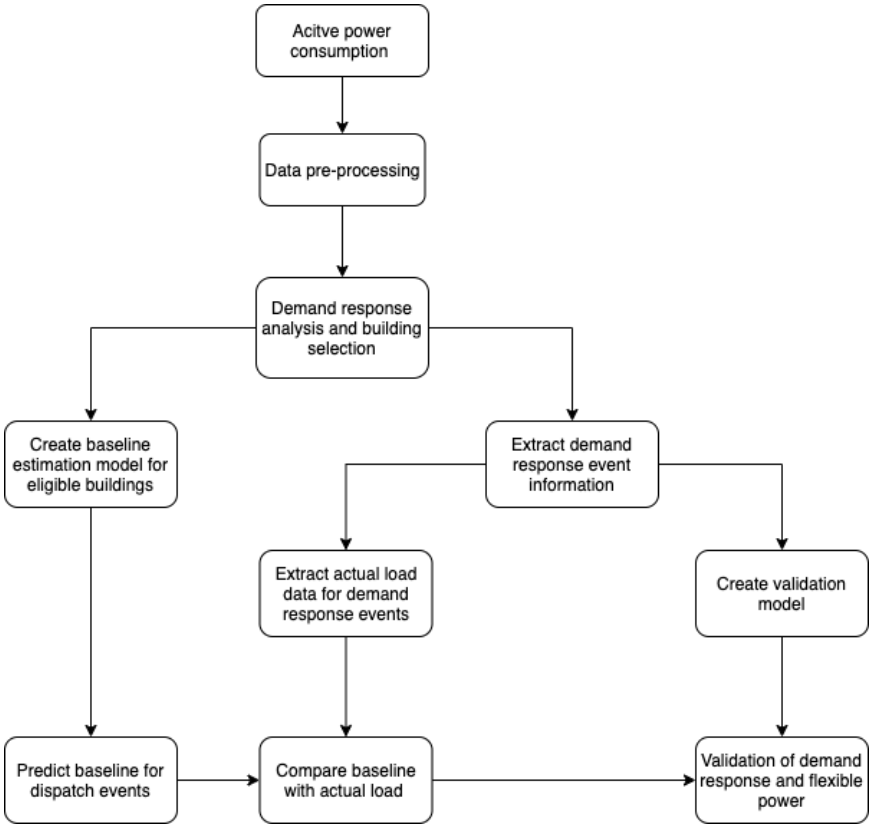


Figure 6: Complete overview of the system model

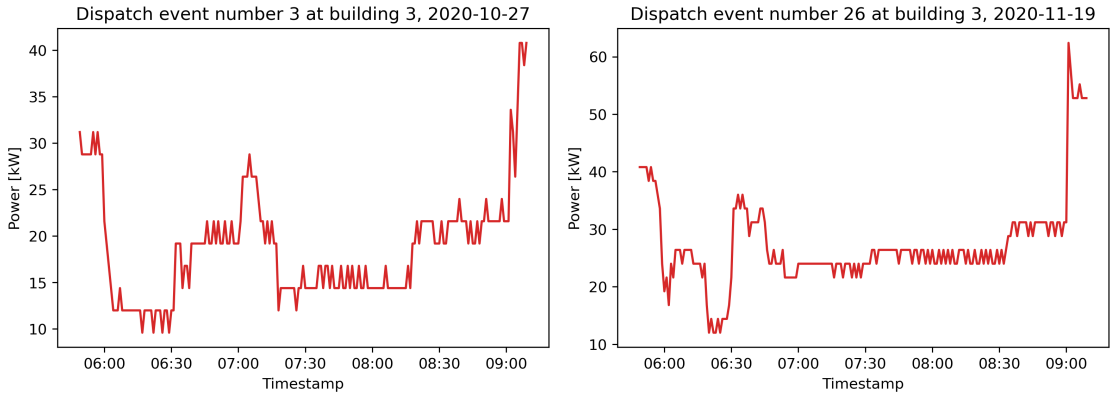
5 Results

The results for the system model are presented in the next chapter. First, an initial analysis of the demand response events at each site is presented. Then, it is determined if the buildings are eligible for use in the demand response validation process, based on the number of demand response events and the impact of demand response at the power consumption. For example, building 1 and building 6 are ineligible due to too few demand response events 4.1. After the building selection, the baseline performance of the three estimation techniques is presented on the eligible buildings. Lastly, the results of the validation process are presented.

5.1 Demand Response Event Analysis

5.1.1 Building 3

Building 3 has a flexible ventilation system, and when dispatch begins, the operation of the ventilation system is reduced. The DR events are located in the power consumption measurements and studied. In figure 7 representative examples for demand response events are illustrated. From the analysis, it is concluded that there is enough impact by the demand response events to perform the baseline validation process. The demand response alters the power consumption significantly, and a power reduction is clearly visible at the beginning of the event and a corresponding increase in power are visible at the end. For building 3, the duration of demand response events is usually three hours. However, some instances of one and two hours exist.



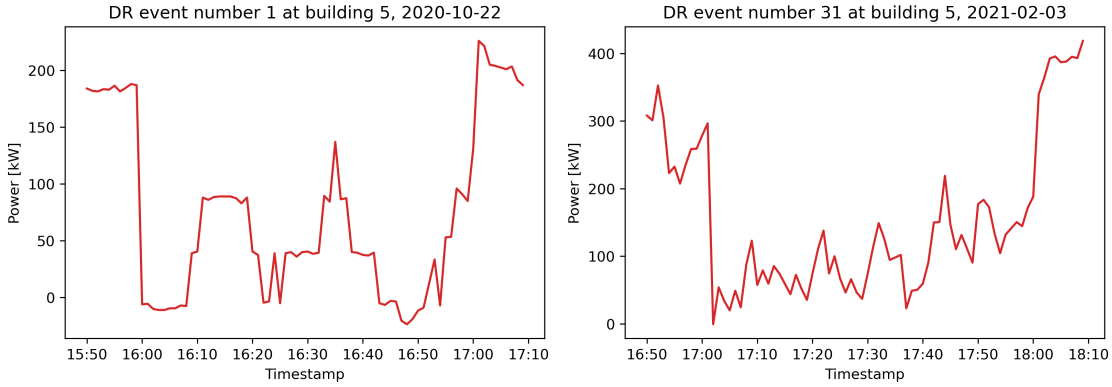
(a) DR event: 06.00-09.00

(b) DR event: 06.00-09.00

Figure 7: An illustration of dispatch events at building 3. The duration of the demand response events are stated below each plot.

5.1.2 Building 5

Building 5 have a battery system as the primary flexible appliance, and when flexibility occurs, the battery discharge power. If the battery were charging when dispatch was initiated, the reduction would include a stop in charging. As the flexible appliance is a battery, the load is curtailable, and the normal discharge from the battery is given as 200kW. All DR events are located in the power consumption measurements and studied. Representative examples of how the power consumption is affected are visualized in figure 8. The impacts of the demand response on power consumption are significant enough to be further analyzed.



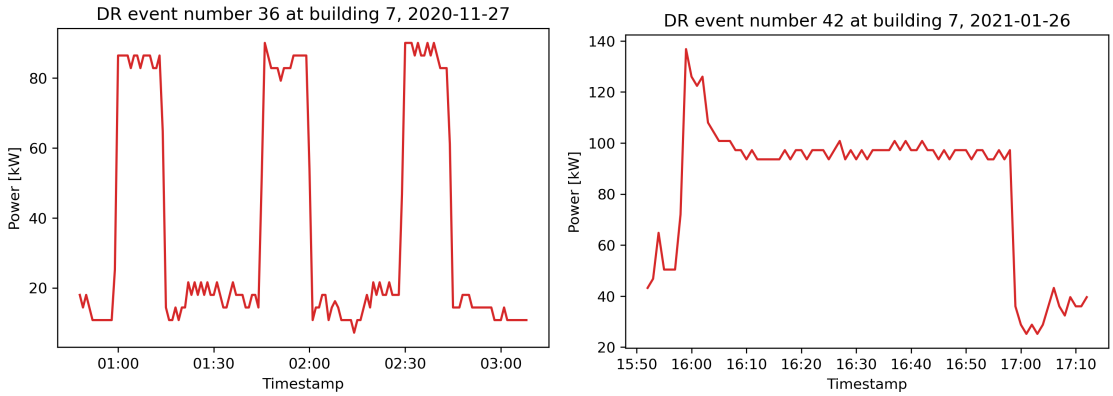
(a) DR event: 16.00-17.00

(b) DR event: 17.00-18.00

Figure 8: An illustration of dispatch events at building 5 with the duration of the dispatch events stated below each plot.

5.1.3 Building 7

Building 7 has a flexible ventilation system and when dispatch is engaged, the power of the ventilation system is reduced. The analysis of the demand response events shows that many other appliances are affecting power consumption. There seems to be one highly volatile appliance that has more significant influence over the power consumption than the ventilation. In figure 9 the power from representative DR events shows that the reduction in ventilation have too little impact on the power consumption to be noticeable, and the increase in another volatile load overshadows the change in the ventilation completely. Therefore, building 7 are ineligible for further analysis.

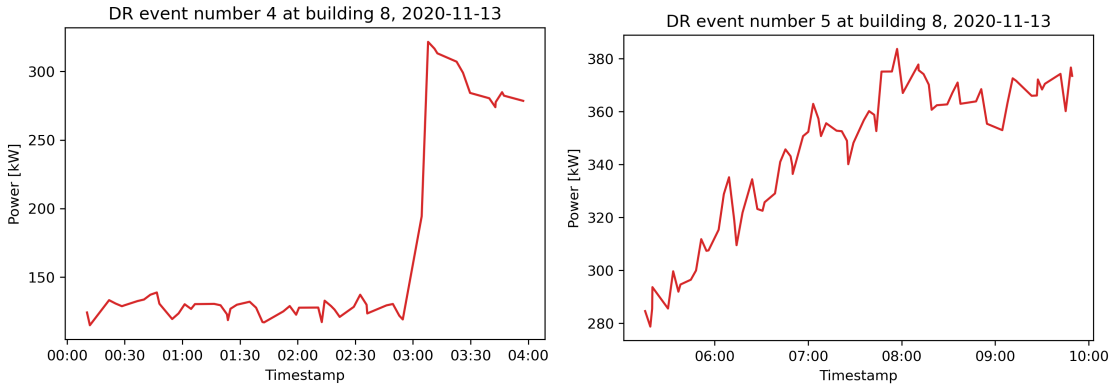


(a) DR event: 01.00-02.00 **(b)** DR event: 16.00-17.00

Figure 9: An illustration of dispatch events at building 7 with the duration of the dispatch events stated below each plot. Increase in another load completely overshadows the reduction in ventilation due to demand response

5.1.4 Building 8

In building 8, the flexible appliance is ventilation. However, for building 8, many consumption measurements points around dispatch events are missing. There are only two complete events, and the resolution of the power consumption is 3 minutes instead of 1 minute. The two events are illustrated in figure 10. Due to there only being two demand response events and the events do not influence the power consumption in an expected way, the building is not further analyzed.



(a) DR event: 01.00-03.00 **(b)** DR event: 06.00-09.00

Figure 10: An illustration of dispatch events at building 8 with the duration of the dispatch events stated below each plot.

5.2 Baseline estimation

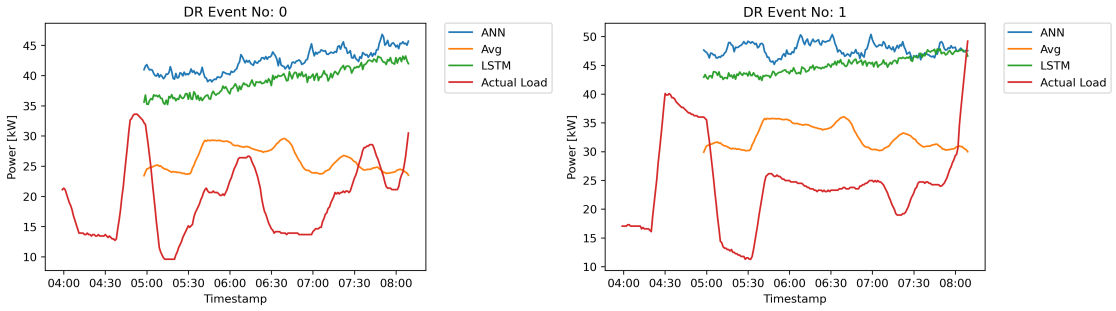
5.2.1 Building 3

As demand response duration is not fixed, the machine learning models must be trained for each duration. The additional training creates a much higher computational cost for both the artificial neural network and the LSTM method. As the construction of the averaging method is not affected by the duration, the computational cost is the same. The evaluation metrics for each method are given in table 5.1. For building 3, the machine learning techniques do not estimate the load in a better way than the simpler averaging method. All evaluation metrics indicate that the averaging method are slightly more accurate than the machine learning algorithms. The parameters used to tune the algorithms are presented in figure 7.1 in the appendix.

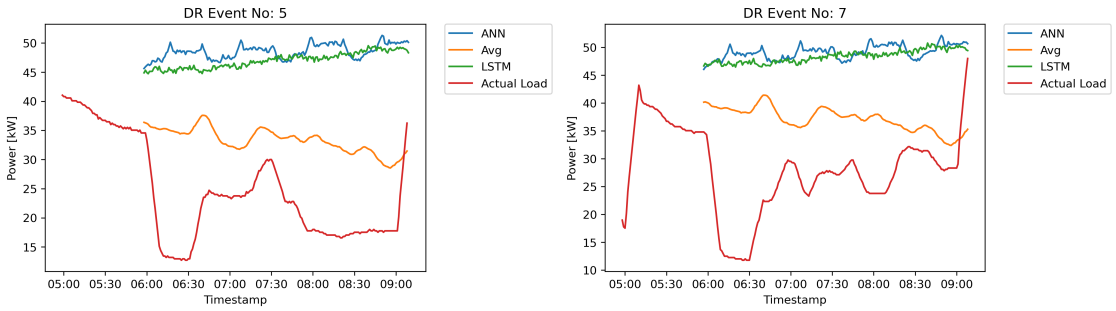
	MAPE	MAE	RMSE
LSTM 1h	11.64	6.77	8
LSTM 2h	11.96	7.03	8.44
LSTM 3h	12.89	7.52	9.12
ANN 1h	12.25	7.08	8.41
ANN 2h	13.37	7.75	9.29
ANN 3h	13.62	7.86	9.52
8of10	11.14	6.41	7.78
8of10 2h	11.00	6.37	7.95
8of10 3h	11.18		6.44
8.18			

Table 5.1: A comparison of the error of baseline estimation methods. The error is tested at the last 10% of the data set where the actual load is known

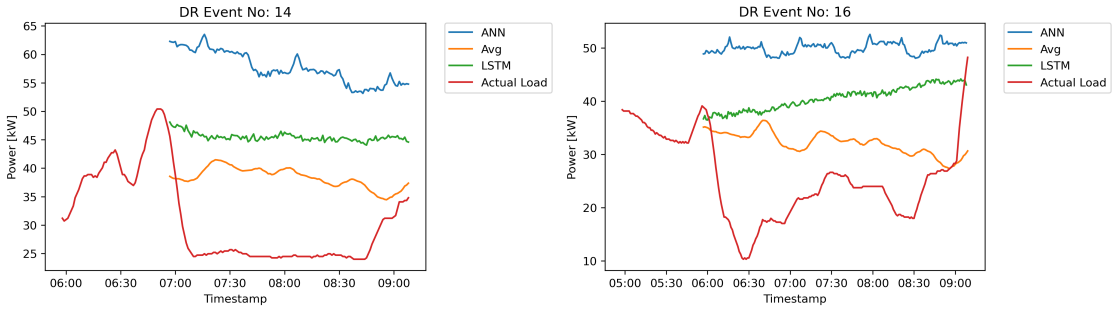
In figure 11, representative demand response events with baseline estimation from the three methods are presented. ANN refers to the artificial neural network, LSTM refers the long short-term memory method and Avg refers to the averaging day method. The rest of the demand response events with baseline are gathered in figure 17-21 in the appendix. At this site, the evaluation metrics seems to be representative of which model performs best. As visualized in figure 11 (a)-(d), the machine learning models clearly overestimate the baseline, ANN at a slightly higher degree than LSTM. The LSTM method performs significantly better than ANN in some situations as illustrated in figure 11 (e)-(h). The averaging method seems, without doubt, better in most situations but are prone to underprediction, as illustrated in figure 11 (a) and (g).



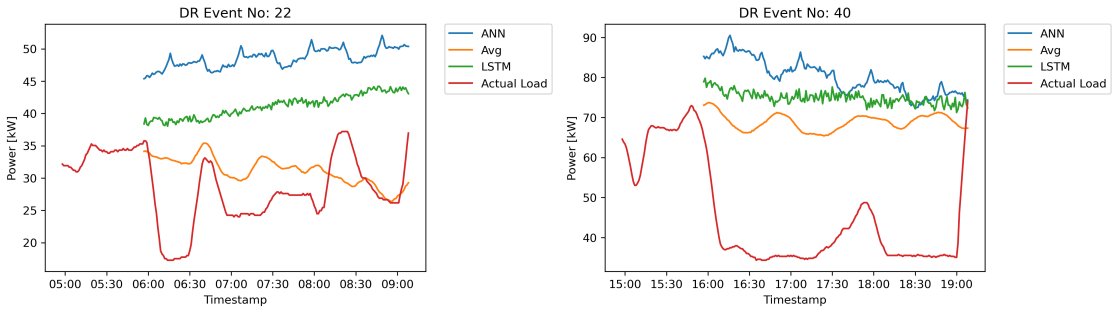
(a) DR event: 05.00-08.00 **(b)** DR event: 05.00-08.00



(c) DR event: 06.00-09.00.00 **(d)** DR event: 06.00-09.00



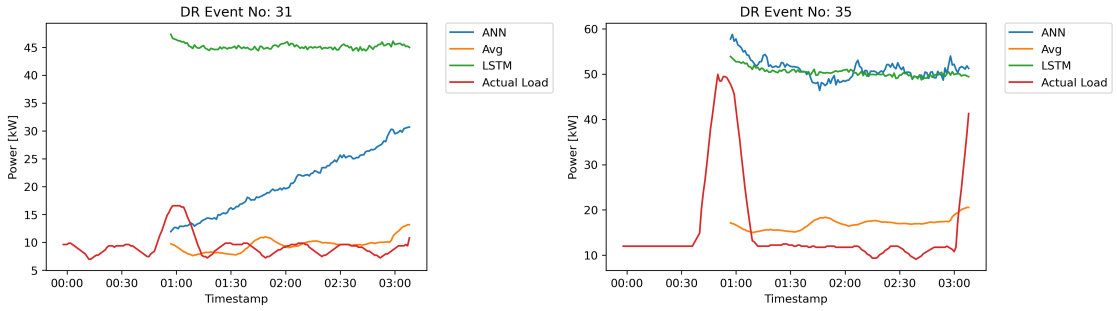
(e) DR event: 07.00-09.00 **(f)** DR event: 06.00-09.00



(g) DR event: 07.00-08.00 **(h)** DR event: 16.00-19.00

Figure 11: An illustration of baseline estimation on demand response events for the three different estimation methods. The timestamps indicate the beginning and end of the demand response event.

It was discovered that events during nighttime between 00.00-04.00 seldom offer any reduction instantly as the demand response event begins. The air quality is most likely sufficient to have the ventilation system already running on a reduced operation. However, the demand response event might prevent the ventilation system from starting up again. The performance of the baseline methods in these situations is illustrated in figure 12. The machine learning algorithms seem to estimate that increase in power is expected in the time period.



(a) DR event: 01.00-03.00 **(b)** DR event: 01.00-03.00

Figure 12: An illustration of baseline estimation on demand response events for the three different estimation methods at night time. The timestamps indicate the beginning and end of the demand response event.

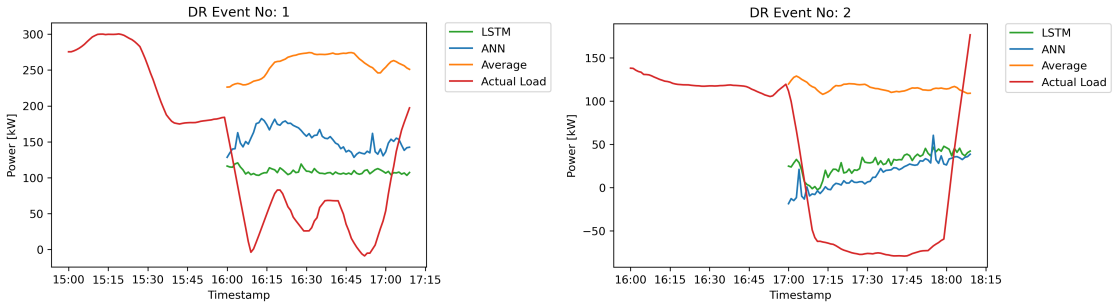
5.2.2 Building 5

The baseline estimation methods were applied to building 5 as there was a sufficient amount of dispatch events, and the flexible load offers a significant change in the main-meter readings. The evaluation metrics of the baseline estimation methods for building 5 are given in table 5.2. The LSTM method has the lowest error in all evaluation metrics, while the averaging method performs better than the ANN method. The parameters used to tune the algorithms are presented in figure 7.2 in the appendix.

	MAPE	MAE	RMSE
LSTM	16.04	36.38	43.60
ANN	19.35	41.44	50.78
8of10	18.64	39.91	46.79

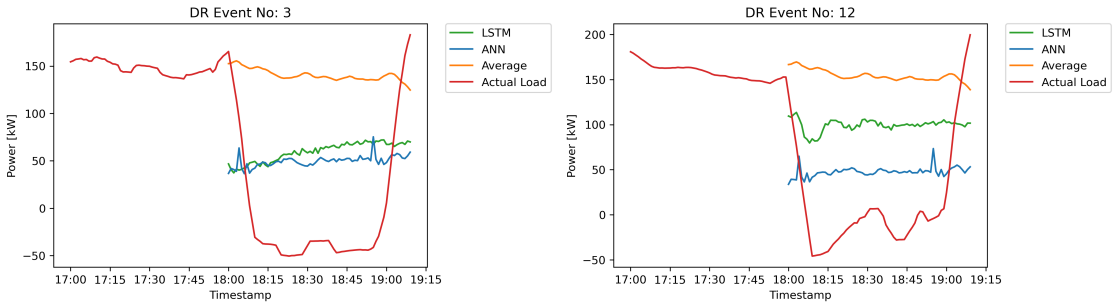
Table 5.2: A comparison of the error of baseline estimation methods. The error is tested at the last 10% of the data set where the actual load is known

The estimation methods estimate quite differently for most of the events and illustrate using only baseline estimation in flexibility settlement. In almost half of the events, the machine learning models clearly underpredicts the baseline as illustrated in figure 13 (a)-(d). The machine learning methods seem to capture the baseline slightly better than the average method on the rest of the event as illustrated in 13 (e)-(g). The artificial neural network method has a slightly higher error than LSTM for building 5, and are more dependent on the power consumption in the minutes before the demand response events as illustrated in figure 13 (h). The averaging method is indeed more averaging and rarely miss entirely. The only event that the averaging method clearly misses is illustrated in figure 13 (a).



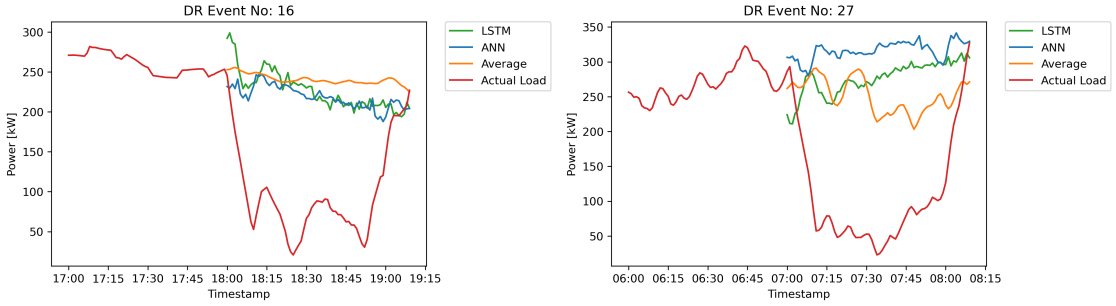
(a) DR event: 16.00-17.00

(b) DR event: 17.00-18.00



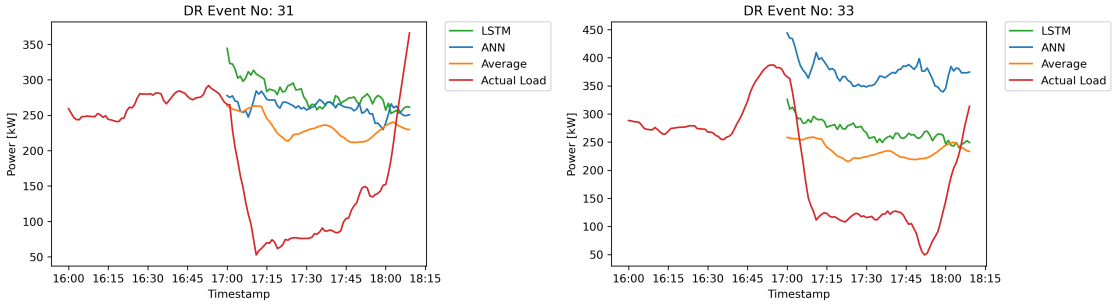
(c) DR event: 18.00-19.00

(d) DR event: 18.00-19.00



(e) DR event: 18.00-19.00

(f) DR event: 07.00-08.00

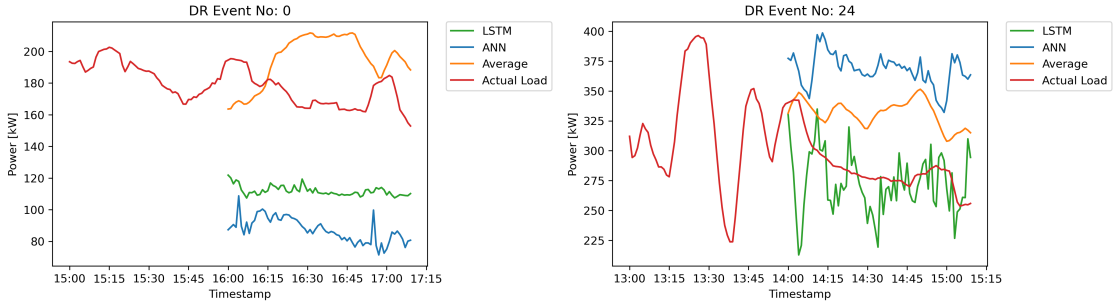


(g) DR event: 17.00-18.00

(h) DR event: 17.00-18.00

Figure 13: An illustration of baseline estimation on demand response events for the three different estimation methods. The timestamps indicate the beginning and end of the demand response event.

Figure 14 illustrates unusual demand response events, where the demand response does not operate as expected. Possible reasons for the demand response event to not operate as expected can be that the battery had no power to dispatch or other appliances increase consumption accordingly. The averaging is the only model that can reasonably estimate the baseline.



(a) DR event: 16.00-17.00

(b) DR event: 14.00-15.00

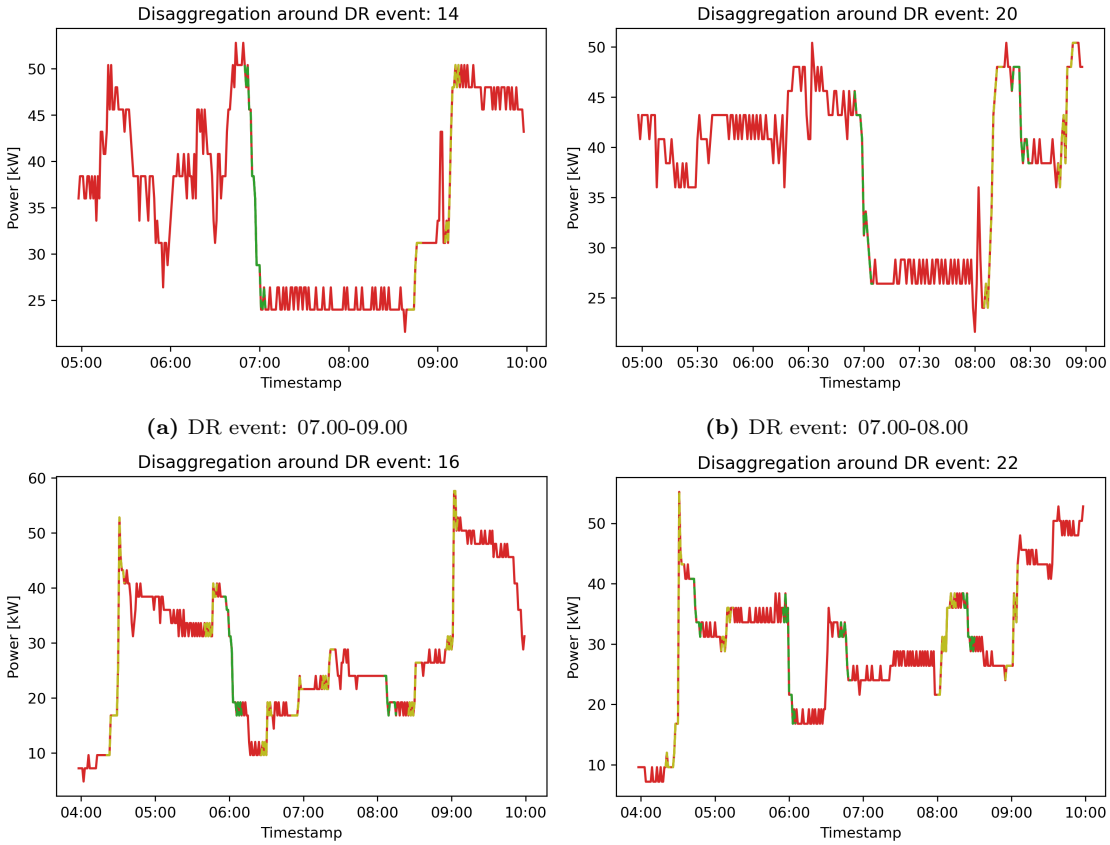
Figure 14: An illustration of baseline estimation on unusual demand response events for the three different estimation methods. The timestamps indicate the beginning and end of the demand response event

5.3 Load disaggregation

Load disaggregation was performed at building 3 and 5. As there was no appliance signature, some of the demand response events were used as training. Therefore, a reference signature is extracted from the beginning and end of the demand response events instead.

5.3.1 Building 3

The disaggregation algorithm was implemented at building 3, using the first ten demand response events as training. Figure 15 illustrates representative events with disaggregation from the remainder of the demand response events while the remaing events are gathered in figures 26-29. The representative signature is compared to the detected edges by DTW, and the comparisons with distances lower than 15% of the largest distance recorded are considered a match. The green line indicates a match connected to a reduction in the flexible appliance, while a yellow indicates an increase in the flexible appliance. From figure 15 (b)-(d), it is illustrated that the DTW cannot differentiate edges of demand response to other edges in the power consumption. Multiple edges are detected as a match, and there is no verification of which matches outside the demand response events are correct. As the ventilation system is a multi-state appliance, the amplitude in state change will probably vary between the different states, and the amplitude of edges are hard to use as verification.



(a) DR event: 07.00-09.00

(b) DR event: 07.00-08.00

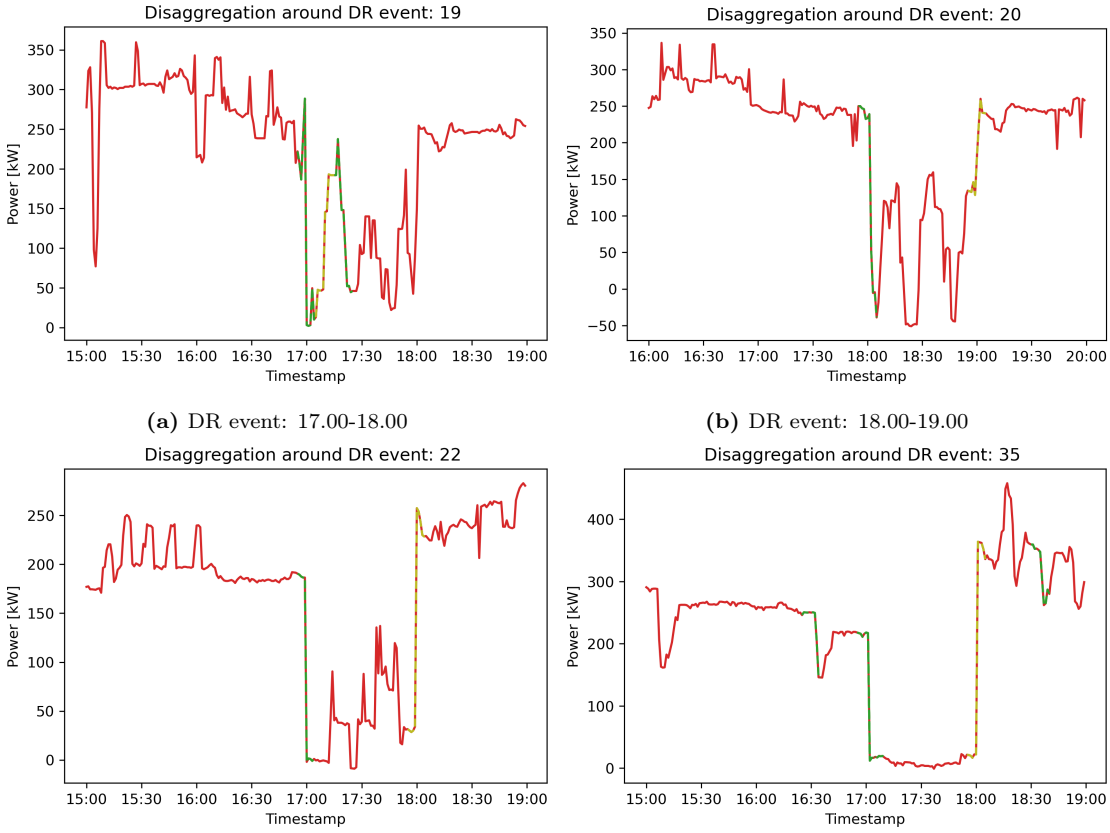
(c) DR event: 06.00-09.00

(d) DR event: 06.00-09.00

Figure 15: An illustration of the load disaggregation method applied at building 3. The green egdes illustrates match in reducing flexible appliance, and the yellow edges illustrates match in increasing flexible appliance

5.3.2 Building 5

The load disaggregation process was performed at building 5. It is assumed that the battery seldom discharges as much power outside of the demand response events. Therefore, the battery operates as a multi-state appliance and excluding edges by amplitude are challenging. Therefore, it is impossible to know if edges previous to demand response events are connected to the flexible appliance or other appliances. The results then greatly depend on the threshold of similarity set in DTW comparison. Figure 16 visualize the performance of the load disaggregation around representative demand response events where the 20% most similar edges are considered a match and the rest of the events are gathered in the figures 30-32. Figure 16 (a) illustrates the main concern, as multiple edges are considered a match. This threshold still captures most of the demand response event edges as figure 16 (b)-(d) illustrate, however not all as illustrated in figure 16 (a).



(a) DR event: 17.00-18.00

(b) DR event: 18.00-19.00

(c) DR event: 17.00-18.00

(d) DR event: 17.00-18.00

Figure 16: An illustration of the load disaggregation method. The green edges illustrates match in reducing flexible appliance, and the yellow edges illustrates match in increasing flexible appliance

6 Discussion

6.1 Model selection

The system model, based on combination of baseline estimation and load disaggregation, was created to meet the specific requirements of the system setup, and different promising baseline estimation techniques were tested. LSTM was chosen because of the ability to capture both long term and temporary connections in sequential data, and as power consumption are dependent on previous power consumption. ANN are a well-established load forecasting technique and have an excellent ability to capture non-linearities and connections between independent and dependent variables. In addition, an averaging method was chosen as this is the most common baseline technique to implement due to simplicity, integrity, and relatively accurate results. Even though support vector regression would be a promising algorithm, the need for hyper-parameter tuning and high computational cost excluded the model from this system. The disaggregation process was chosen based on the information available. As the information did not include appliance signature, multiple features or high frequency measurements, the algorithm chosen became a simple edge detection using binary segmentation change point detection and matching with dynamic time warping as dynamic time warping are able to compare signals of unequal length.

6.2 Site selection

Initial study on the number of demand response events shows that site 1 has three demand response events, and site 6 has only one demand response event as presented in table 4.1 . Therefore, these sites are excluded from the system model as the results will not give enough information on how the model operates on demand response events in general. An initial demand response analysis was performed on the remaining events to determine sites to include in the system model. For building 7, other appliances are too volatile and overshadow the drop in power due to the demand response event as illustrated in figure 9. As the baseline estimation estimates the aggregated power consumption, even a tiny error in estimation could be larger than the influence of the demand response events, and any inspection of the performance of the baseline techniques would therefore be a problem. For building 8, three demand response events have gaps in the data, providing an analysis hard to execute. In addition, the two remaining events do not operate as expected. A possible reason could be that the flexible appliance already works on reduced operation, so there is no or small power drop when the demand response event occurs. This seems to be the case of figure 10 (a). At the start of the event, there is no change. However, in the end, there is a significant increase in power consumption. Another possible reason could be that the flexible appliance is too small compared to the total power in the building, and the change in power does not influence the measurement. For example, figure 10 (b) shows a steady increase in power during the demand response events, which corresponds to the start of the working day. The flexible appliance is probably too small to impact the increase of power due to the start of the working day.

6.3 Baseline estimation

To train forecasting models, the data is processed in more detail. As the training data should forecast normal operation, the days containing demand response events will differ from the normal state operation days. Therefore, they should optimally be removed to increase the accuracy of the model. However, due to the number of demand response events, removing the whole day would cause a significant smaller training set. Therefore, only the hours affected by the demand response events were removed. To not remove the whole day of demand response events probably reduce the accuracy of the model. If the training set had been more extensive, this would definitively be an area to examine to improve the accuracy of the machine learning algorithms.

A few more variables are added to improve the forecasting model. Hourly temperature is added as the temperature has a great influence on load consumption, especially in Norway as heating and cooling are mainly electric [54]. In addition, the day of the week and week number are additional variables, as the load profile are more similar for the same day of the week. Due to the vast differences between weekdays and weekends in office buildings, the weekends are removed from the training data as all demand response events happen during weekdays. It is common for load forecasting models to add additional weather information such as humidity and solar radiation as variables. However, such weather data are not available for the locations of building 3 and 5. The period of the demand response events are usually at the same time of day, which create a smaller training set for those specific periods. This might influence the machine learning algorithms to achieve poorer results during these events than the rest of the validation set. In addition, there are multiple demand response days in succession. The accuracy of the averaging method most likely be reduced as possible similar days are removed from the Y days as days with demand response are not eligible for use.

The performance of the baseline estimation methods is evaluated by different evaluation metrics and visual inspection of the performance in demand response events. The error results are quite similar over all different baseline estimation algorithms. LSTM has the lowest average MAPE of 16.04% at building 5 while the averaging method scores best at building 3 with an average MAPE of 10.89 %. However, the estimation baseline trends differently as visualized in figure 11 and figure 13. The difference in trends probably confirms that the error of the baseline estimation is too high compared to the relative error between the models. The machine learning estimations are more volatile than the averaging method and the error in some situations is much higher than for the averaging method as illustrated in figure 13 (a)-(b) and figure 11 (a)-(g). The averaging method seems to be more stable around the DR events, and the estimation has a higher tendency to estimate the correct beginning of the DR event. As illustrated in figure 13 (h), the ANN method has a higher tendency to be affected by the load directly previous to the estimation. This can cause a more accurate estimation in some situations, however, it means that the estimation method is more prone to gaming than the other methods. None of the estimation techniques correctly captures the spikes during the demand response events and therefore the estimation of flexible power can be more difficult to calculate. In addition, some demand response events due not have the expected power profile where load consumption are clearly reduced, illustrated in the figures 14 and 12. For building 5, the battery might be out of power in the beginning of the events and therefore do not provide any discharge. For site 3, the events during night time are different. Most likely, the air quality is sufficient and the ventilation system is already operating at lower power.

However, the demand response events might delay the start of the ventilation system. The baseline is even more challenging to predict in the situations with unusual power pattern. Even though the machine learning methods overestimate the baseline regularly, it seems like these methods expect an increase in load. The increase can somehow be determined by figure 12 (b) as power consumption are increased before demand response and instantly decrease as the event starts. The average model does not predict the same increase. However, the average model might suffer because the adjustment period previous to the event has very low power consumption. Overall, the baseline estimations do not capture the volatility of the load. In spite of this, all of the estimation techniques provide significantly better results than the similar LSTM method in [11]. However, the models are only tested at the two most eligible buildings.

After the baseline estimation has been implemented, the flexibility calculation can be calculated differently depending on what the goal is for flexibility. For example, if flexibility is needed for fast response markets to balance supply and demand, the power reduction at each time step would be a reasonable calculation method. This is because the power from the flexibility is the most essential aspect. However, if the flexibility is used for energy flexibility, the volume of the flexibility during the entire demand response event would be the most reasonable calculation method. In general, the three estimation methods have significant errors as expected for high-resolution estimation, and it is challenging to determine which estimation technique to rely on. Additional validation to help decide which baseline estimation to choose is needed. Unfortunately, it was not provided enough information to create a load disaggregation model to estimate the power profile of the flexible appliance. If information regarding transient and steady-state signature were given, sophisticated algorithms such as Fractional Hidden Markov Models could be utilized to estimate a load profile. In addition, as only active power was provided, too little information regarding the edges could be extracted.

6.4 Disaggregation

The disaggregation is an additional step to create a more transparent flexibility settlement, and can optimally help to determine if the origin of the power reduction during demand response is from the flexible appliance. However, the edge matching disaggregation process are in need of either multiple features or high resolution to achieve good results. As the system setup did not include either, the structure of the disaggregation became a simple detect and match algorithm with only active power to base on. In addition, as there was no appliance signature to compare against the first demand response events were chosen as representative signatures. A direct comparison would have given poor results, as the baseload is different. Scaling of the signature and edges were performed to get a more accurate comparison. Both ventilation system and battery are multi-state appliances, and the amplitude of state changes are inconsistent. For this reason, the loss of amplitude by scaling does not influence the comparison as much as it would have for single-state appliances. However, multi-state appliances create a more complex analysis as amplitude cannot be a factor to separate the edges. The load disaggregation based on edges extracted from the first portion of demand response events proved to detect most demand response events. However, quite a few edges outside of the demand response events were also characterized as matches. Unfortunately, as the flexible appliances are multi-state, there is no direct way to determine if the matches outside the demand response events are correct or not. Therefore, load disaggregation does not provide the desired information to create a more robust demand response validation given the current information. If multiple features such as reactive power, current or voltage were present, more information would be gathered from edges. The load disaggregation might at least be used to verify that edges are connected to the flexible appliance.

6.5 Further work

The baseline estimation with high resolution proves challenging to keep error at an acceptable level when the volatility is as high as for individual buildings. Improvements to reduce error such as clustering of days, or the possibility to add more weather parameters could be further investigated. Given the process of combining load disaggregation and baseline estimation had been optimal, possible topics to further research include; 24h ahead prediction of available flexibility by power profile of flexible appliances and predicted baseline. Investigative optimal market to trade flexibility, based on appliance characteristics and power profile.

7 Conclusion

This thesis has examined the possibility of adding load disaggregation as an extra demand response event validation in addition to baseline estimation to create more accurate and transparent flexibility settlement. Optimally, a power profile for the flexible appliance could be estimated by disaggregation to help determine the correct baseline estimation method. However, as disaggregation methods to create power profiles require appliance-specific signatures, multiple features or extremely high-frequency consumption data, the power profiles could not be created by the given system setup. Instead, the possibility of utilizing simple load disaggregation to connect the demand response events edges to the same appliance and verify the origin of the flexible power are investigated. The baseline estimation techniques have pretty significant differences in the estimation, and the result of flexibility settlement will greatly depend on which algorithm is chosen. For building 5, the LSTM method provides the lowest error compared to the actual load. However, the machine learning models are prone to miss in some situations. The averaging model is more consistent, but it will seldom be completely accurate. For building 3, the averaging model has the lowest error and is by far more consistent as the machine learning algorithms clearly overestimate the baseline in most events.

Due to low resolution, only active power consumption and the fact that the flexible appliance is multi-state, a large majority of edges are too similar to distinguish the edges connected to the demand response from the other edges. The simple disaggregation technique does not have sufficient information to verify the origin of the flexible power. Given the current system setup, load disaggregation can not validate the demand response, and the flexibility settlement solely rely on the choice of baseline estimation method.

References

[1] X. Jin, Q. Wu, and H. Jia, *Local flexibility markets: Literature review on concepts, models and clearing methods*, Mar. 2020. DOI: 10.1016/j.apenergy.2019.114387.

[2] *Status of Power System Transformation 2019: Power system flexibility – Analysis - IEA*. [Online]. Available: https://www.iea.org/reports/status-of-power-system-transformation-2019.

[3] "Council of European Energy Regulators asbl Cours Saint-CEER Paper on DSO Procedures of Procurement of Flexibility Distribution Systems Working Group," Tech. Rep., 2020. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32009R0714.

[4] E. Hillberg Antony Zegers, B. Herndler, S. Wong, J. Pompee, J.-Y. Bourmaud, and S. Lehnhoff, "Power Transmission & Distribution Systems Flexibility needs in the future power system Discussion paper Disclaimer," Tech. Rep. [Online]. Available: https://www.iea-isgan.org/wp-content/uploads/2019/03/ISGAN_DiscussionPaper_Flexibility_Needs_In_Future_Power_Systems_2019.pdf.

[5] H. J. Jabir, J. Teh, D. Ishak, and H. Abunima, *Impacts of demand-side management on electrical power systems: A review*, 2018. DOI: 10.3390/en11051050.

[6] H. d. Heer, M. van der Laan, and A. Armenteros, "USEF FRAMEWORK," no. May, 2021. [Online]. Available: https://www.usef.energy/app/uploads/2021/05/USEF-The-Framework-Explained-update-2021.pdf.

[7] Enernoc, "The Demand Response Baseline Determination Methods," 2010. [Online]. Available: https://library.cee1.org/sites/default/files/library/10774/CEE_EvalDRBaseline_2011.pdf.

[8] A. Ridi, C. Gisler, and J. Hennebert, "A survey on intrusive load monitoring for appliance recognition," in *Proceedings - International Conference on Pattern Recognition*, Institute of Electrical and Electronics Engineers Inc., Dec. 2014, pp. 3702–3707, ISBN: 9781479952083. DOI: 10.1109/ICPR.2014.636.

[9] A. Moradzadeh, O. Sadeghian, K. Pourhossein, B. Mohammadi-Ivatloo, and A. Anvari-Moghaddam, "Improving residential load disaggregation for sustainable development of energy via principal component analysis," *Sustainability (Switzerland)*, vol. 12, no. 8, p. 3158, 2020, ISSN: 20711050. DOI: 10.3390/SU12083158.

[10] K. Zor, O. Timur, and A. Teke, "A state-of-the-art review of artificial intelligence techniques for short-term electric load forecasting," *2017 6th International Youth Conference on Energy, IYCE 2017*, 2017. DOI: 10.1109/IYCE.2017.8003734.

[11] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851, 2019, ISSN: 19493053. DOI: 10.1109/TSG.2017.2753802.

[12] A. Ramos, "Consumer Access to Electricity Markets: The Demand Response Baseline," in *International Conference on the European Energy Market, EEM*, vol. 2019-Septe, IEEE Computer Society, Sep. 2019, ISBN: 9781728112572. DOI: 10.1109/EEM.2019.8916212.

[13] Y. Chen, P. Xu, Y. Chu, W. Li, Y. Wu, L. Ni, Y. Bao, and K. Wang, "Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings," *Applied Energy*, vol. 195, pp. 659–670, 2017, ISSN: 03062619. DOI: 10.1016/j.apenergy.2017.03.034. [Online]. Available: http://dx.doi.org/10.1016/j.apenergy.2017.03.034.

[14] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Transactions on Power Systems*, vol. 16, no. 1, pp. 44–55, 2001, ISSN: 08858950. DOI: 10.1109/59.910780.

[15] A. Marinescu, C. Harris, I. Dusparic, S. Clarke, and V. Cahill, "Residential electrical demand forecasting in very small scale: An evaluation of forecasting methods," *2013 2nd International Workshop on Software Engineering Challenges for the Smart Grid, SE4SG 2013 - Proceedings*, pp. 25–32, 2013. DOI: 10.1109/SE4SG.2013.6596108.

[16] Z. Jing, M. Cai, M. Pipattanasomporn, S. Rahman, R. Kothandaraman, A. Malekpour, E. A. Paaso, and S. Bahramirad, "Commercial Building Load Forecasts with Artificial Neural Network," *2019 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2019*, no. May 2020, 2019. DOI: 10.1109/ISGT.2019.8791654.

[17] J. Jazaeri, T. Alpcan, R. Gordon, M. Brandao, T. Hoban, and C. Seeling, "Baseline methodologies for small scale residential demand response," *IEEE PES Innovative Smart Grid Technologies Conference Europe*, pp. 747–752, 2016. DOI: 10.1109/ISGT-Asia.2016.7796478.

[18] S. Pati, S. J. Ranade, and O. Lavrova, "Methodologies for customer baseline load estimation and their implications," *2020 IEEE Texas Power and Energy Conference, TPEC 2020*, 2020. DOI: 10.1109/TPEC48276.2020.9042538.

[19] Y. Wang, D. Gan, M. Sun, N. Zhang, Z. Lu, and C. Kang, "Probabilistic individual load forecasting using pinball loss guided LSTM," *Applied Energy*, vol. 235, no. November 2018, pp. 10–20, 2019, ISSN: 03062619. DOI: 10.1016/j.apenergy.2018.10.078. [Online]. Available: https://doi.org/10.1016/j.apenergy.2018.10.078.

[20] M. Sun, Y. Wang, F. Teng, Y. Ye, G. Strbac, and C. Kang, "Clustering-based residential baseline estimation: A probabilistic perspective," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6014–6028, Nov. 2019, ISSN: 19493061. DOI: 10.1109/TSG.2019.2895333.

[21] S. Park, S. Ryu, Y. Choi, J. Kim, and H. Kim, "Data-Driven Baseline Estimation of Residential Buildings for Demand Response †," vol. 8, pp. 638–643, 2015, ISSN: 1996-1073. DOI: 10.3390/en80910239. [Online]. Available: https://www.mdpi.com/1996-1073/8/9/10239.

[22] A. Ruano, A. Hernandez, J. Ureña, M. Ruano, and J. Garcia, "energies NILM Techniques for Intelligent Home Energy Management and Ambient Assisted Living: A Review," 2019. DOI: 10.3390/en12112203. [Online]. Available: https://www.mdpi.com/1996-1073/12/11/2203.

[23] J. Z. Kolter, S. Batra, and A. Y. Ng, "Energy disaggregation via discriminative sparse coding," *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010, NIPS 2010*, pp. 1–9, 2010. [Online]. Available: https://papers.nips.cc/paper/2010/hash/7810ccd41bf26faaa2c4e1f20db70a71-Abstract.html.

[24] G. C. Koutitas and L. Tassiulas, "Low Cost Disaggregation of Smart Meter Sensor Data," *IEEE Sensors Journal*, vol. 16, no. 6, pp. 1665–1673, Mar. 2016, ISSN: 1530437X. DOI: 10.1109/JSEN.2015.2501422.

[25] Z. Guo, Z. J. Wang, and A. Kashani, "Home appliance load modeling from aggregated smart meter data," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 254–262, Jan. 2015, ISSN: 08858950. DOI: 10.1109/TPWRS.2014.2327041.

[26] S. Makonin, F. Popowich, I. V. Bajic, B. Gill, and L. Bartram, "Exploiting HMM Sparsity to Perform Online Real-Time Nonintrusive Load Monitoring," *IEEE Transactions on Smart Grid*, vol. 7, no. 6, pp. 2575–2585, Nov. 2016, ISSN: 19493053. DOI: 10.1109/TSG.2015.2494592.

[27] W. Kong, Z. Y. Dong, J. Ma, D. J. Hill, J. Zhao, and F. Luo, "An Extensible Approach for Non-Intrusive Load Disaggregation with Smart Meter Data," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3362–3372, Jul. 2018, ISSN: 19493053. DOI: 10.1109/TSG.2016.2631238.

[28] J. Kelly and W. Knottenbelt, "Neural NILM: Deep Neural Networks Applied to Energy Disaggregation," DOI: 10.1145/2821650.2821672.

[29] K. D. Anderson, M. E. Berges, A. Ocneanu, D. Benitez, and J. M. Moura, "Event detection for Non Intrusive load monitoring," *IECON Proceedings (Industrial Electronics Conference)*, no. October, pp. 3312–3317, 2012. DOI: 10.1109/IECON.2012.6389367.

[30] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "An unsupervised training method for non-intrusive appliance load monitoring," *Artificial Intelligence*, vol. 217, pp. 1–19, 2014, ISSN: 00043702. DOI: 10.1016/j.artint.2014.07.010. [Online]. Available: http://dx.doi.org/10.1016/j.artint.2014.07.010.

[31] E. Elhamifar and S. Sastry, "Energy disaggregation via learning towerlets' and sparse coding," *Proceedings of the National Conference on Artificial Intelligence*, vol. 1, pp. 629–635, 2015. [Online]. Available: https://ojs.aaai.org/index.php/AAAI/article/view/9249.

[32] S. Makonin, "Real-Time Embedded Low-Frequency Load Disaggregation," no. August, 2014. [Online]. Available: https://www.researchgate.net/publication/264534650_Real-Time_Embedded_Low-Frequency_Load_Disaggregation.

[33] H. H. Chang, "Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses," *Energies*, vol. 5, no. 11, pp. 4569–4589, 2012, ISSN: 19961073. DOI: 10.3390/en5114569.

[34] Q. Wu and F. Wang, "Concatenate convolutional neural networks for non-intrusive load monitoring across complex background," *Energies*, vol. 12, no. 8, 2019, ISSN: 19961073. DOI: 10.3390/en12081572.

[35] D. Srinivasan, W. S. Ng, and A. C. Liew, "Neural-network-based signature recognition for harmonic source identification," *IEEE Transactions on Power Delivery*, vol. 21, no. 1, pp. 398–405, 2006, ISSN: 08858977. DOI: 10.1109/TPWRD.2005.852370.

[36] J. Liao, G. Elafoudi, L. Stankovic, and V. Stankovic, "Power Disaggregation for Low-sampling Rate Data," Tech. Rep. [Online]. Available: http://nilmworkshop.org/2014/proceedings/liao_power.pdf.

[37] A. Baliyan, K. Gaurav, and S. Kumar Mishra, "A review of short term load forecasting using artificial neural network models," *Procedia Computer Science*, vol. 48, no. C, pp. 121–125, 2015, ISSN: 18770509. DOI: 10.1016/j.procs.2015.04.160. [Online]. Available: http://dx.doi.org/10.1016/j.procs.2015.04.160.

[38] M. Hayati and Y. Shirvany, "Artificial Neural Network Approach for Short Term Load Forecasting for Illam Region," *International Journal of Electrical and Computer Engineering*, vol. 1, no. 4, pp. 667–671, 2007, ISSN: 2088-8708. DOI: 10.5281/zenodo.1328642.

[39] S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowlege-Based Systems*, vol. 6, no. 2, pp. 107–116, 1998, ISSN: 02184885. DOI: 10.1142/S0218488598000094.

[40] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997, ISSN: 1053587X. DOI: 10.1109/78.650093.

[41] *Understanding RNN and LSTM. What is Neural Network? | by Aditi Mittal | Medium.* [Online]. Available: https://aditi-mittal.medium.com/understanding-rnn-and-lstm-f7cdf6dfc14e.

[42] J. F. Kolen and S. C. Kremer, "Gradient Flow in Recurrent Nets: The Difficulty of Learning LongTerm Dependencies," *A Field Guide to Dynamical Recurrent Networks*, no. March 2003, 2010. DOI: 10.1109/9780470544037.ch14. [Online]. Available: https://www.researchgate.net/publication/2839938_Gradient_Flow_in_Recurrent_Nets_the_Difficulty_of_Learning_Long-Term_Dependencies.

[43] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A Critical Review of Recurrent Neural Networks for Sequence Learning," Tech. Rep., 2015. [Online]. Available: https://www.researchgate.net/publication/277603865_A_Critical_Review_of_Recurrent_Neural_Networks_for_Sequence_Learning.

[44] L. Pérez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy and Buildings*, vol. 40, no. 3, pp. 394–398, 2008, ISSN: 03787788. DOI: 10.1016/j.enbuild.2007.03.007.

[45] T. Jiang, Z. Li, X. Jin, H. Chen, X. Li, and Y. Mu, "Flexible operation of active distribution network using integrated smart buildings with heating, ventilation and air-conditioning systems," *Applied Energy*, vol. 226, pp. 181–196, Sep. 2018, ISSN: 03062619. DOI: 10.1016/j.apenergy.2018.05.091.

[46] S. Wong, W. Muneer, S. Nazir, and A. Prieur, "Designing, Operating, and Simulating Electric Water Heater Populations for the Smart Grid," no. JANUARY, p. 88, 2013. DOI: 10.13140/RG.2.1.1232.9686.

[47] J. Wang and H. Shen, "A new method of evaluating dispatchable potential for aggregated residential loads," in *Asia-Pacific Power and Energy Engineering Conference, APPEEC*, vol. 2016-December, IEEE Computer Society, Dec. 2016, pp. 1384–1388, ISBN: 9781509054183. DOI: 10.1109/APPEEC.2016.7779774.

[48] L. Jiang, J. Li, S. Luo, J. Jin, and S. West, "Literature review of power disaggregation," in *Proceedings of 2011 International Conference on Modelling, Identification and Control, ICMIC 2011*, 2011, pp. 38–42, ISBN: 9780956715708. DOI: 10.1109/icmic.2011.5973672.

[49] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive Load Monitoring approaches for disaggregated energy sensing: A survey," *Sensors (Switzerland)*, vol. 12, no. 12, pp. 16 838–16 866, 2012, ISSN: 14248220. DOI: 10.3390/s121216838.

[50] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1590–1598, Jan. 2012, ISSN: 01621459. DOI: 10.1080/01621459.2012.737745. [Online]. Available: http://dx.doi.org/10.1080/01621459.2012.737745.

[51] A. J. Scott and M. Knott, "A Cluster Analysis Method for Grouping Means in the Analysis of Variance," *Biometrics*, vol. 30, no. 3, p. 507, Sep. 1974, ISSN: 0006341X. DOI: 10.2307/2529204.

[52] P. A. Schirmer, I. Mporas, and M. Paraskevas, "Energy Disaggregation Using Elastic Matching Algorithms," *Entropy*, vol. 22, no. 1, p. 71, 2020, ISSN: 1099-4300. DOI: 10.3390/e22010071.

[53] Norsk Klimaservicesenter, *Norsk Klimaservicesenter*, 2021. [Online]. Available: https://seklima.met.no/%20https://seklima.met.no/observations/.

[54] Energifakta Norge, *Energy use by sector - Energifakta Norge*. [Online]. Available: https://energifaktanorge.no/en/norsk-energibruk/energibruken-i-ulike-sektorer/.

Appendix

7.1 Tuned parameters for building 3

The model parameters tuned to acquire the lowest error are provided in table 7.1

	Epochs	Neurons	Minutes before demand response	Timewindow of adjustment [min]	XofY
LSTM 1h	100	100	30	-	-
LSTM 2h	100	100	35	-	-
LSTM 3h	100	100	45	-	-
ANN 1h, 2h, 3h	100	10 000	50	-	-
Averaging	-	-	-	120	8of10

Table 7.1: Tuned parameters with the lowest error for site 3. The parameters were discovered through trial and error.

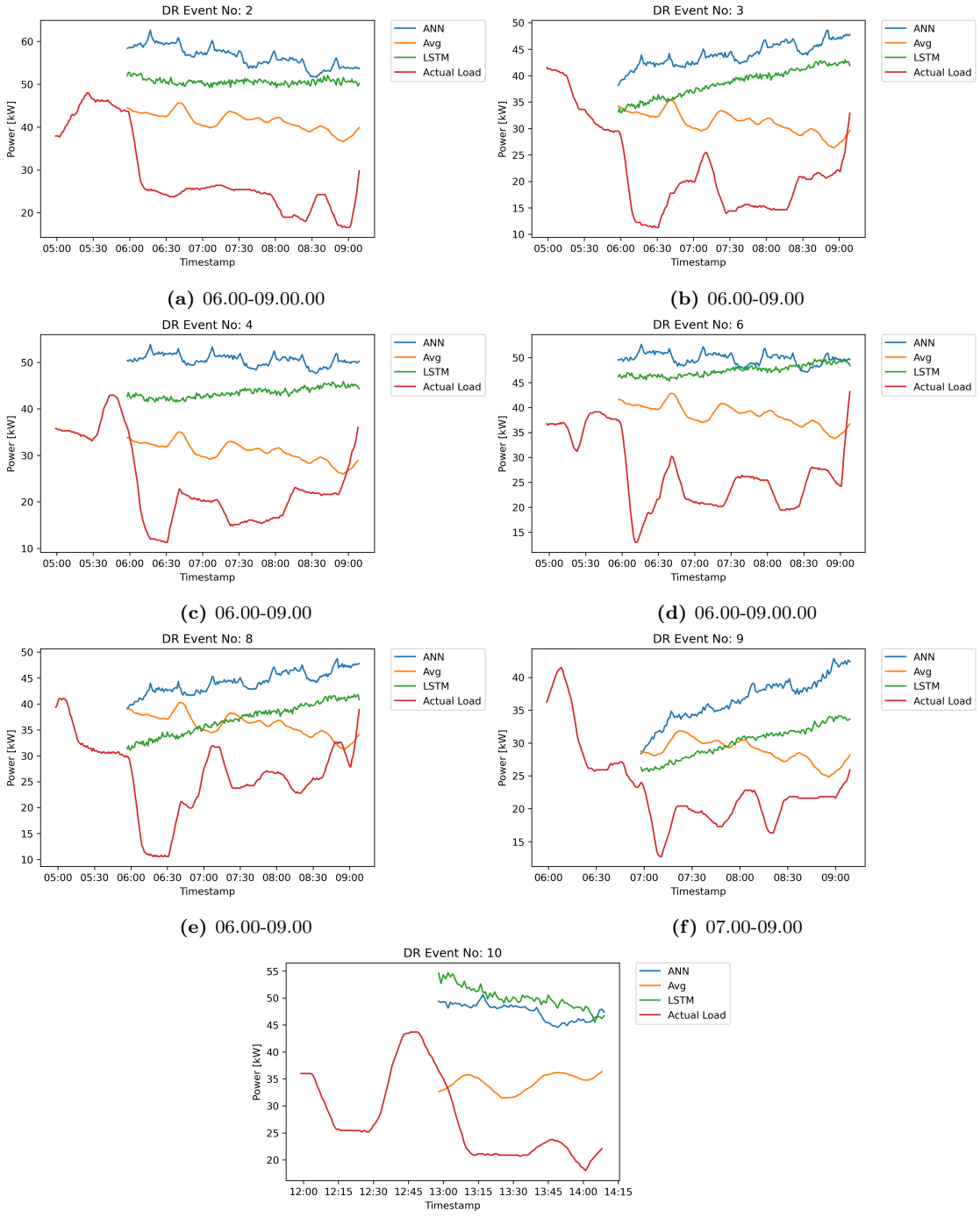
7.2 Tuned parameters for baseline estimation techniques for building 5

The parameters of the different models were tuned by trial and error. The parameters with the lowest corresponding evaluation metrics are presented in table 7.2

	Epochs	Neurons	Minutes before demand response	Timewindow of adjustment [min]	XofY
LSTM	100	100	30	-	-
ANN	100	10 000	60	-	-
Averaging	-	-	-	120	8of10

Table 7.2: Tuned model parameters with lowest for site 5. The parameters were discovered through trial and error.

7.3 Baseline estimation for additional demand response events at building 3



(a) 06.00-09.00.00

(b) 06.00-09.00

(c) 06.00-09.00

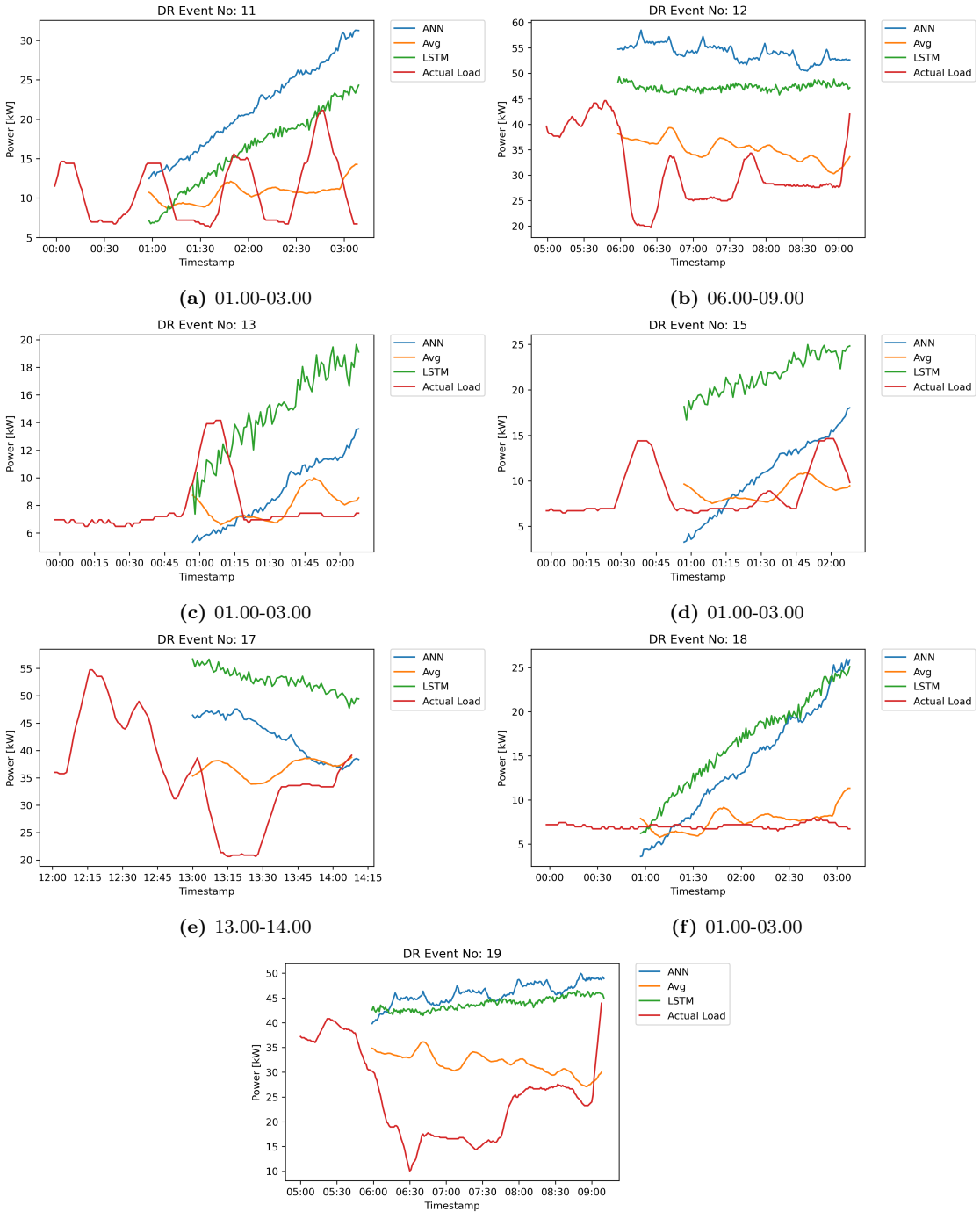
(d) 06.00-09.00.00

(e) 06.00-09.00

(f) 07.00-09.00

(g) 13.00-14.00

Figure 17: An illustration of baseline estimation on demand response events for the three different estimation methods part 1.



(a) 01.00-03.00

(b) 06.00-09.00

(c) 01.00-03.00

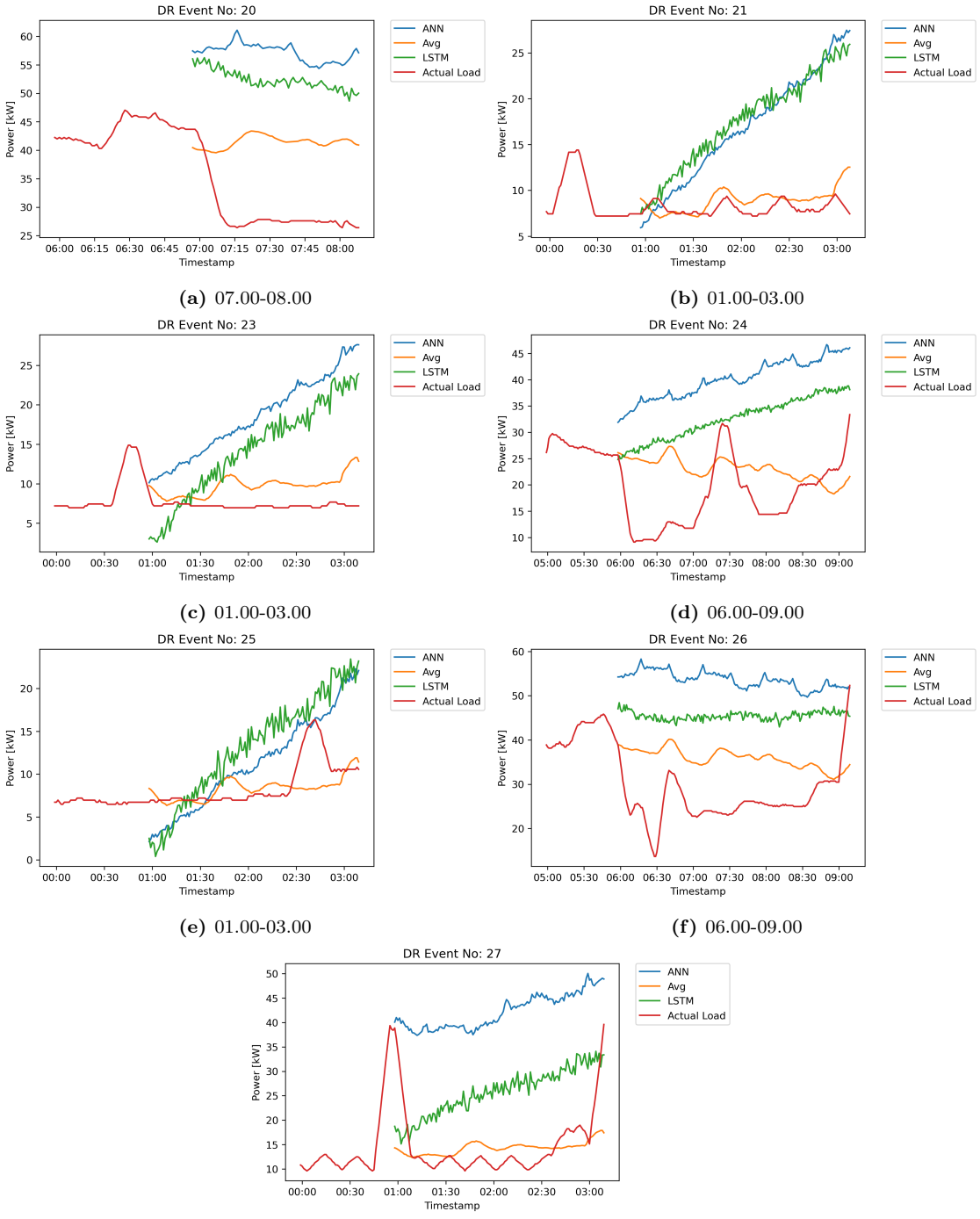
(d) 01.00-03.00

(e) 13.00-14.00

(f) 01.00-03.00

(g) 06.00-09.00

Figure 18: An illustration of baseline estimation on additional demand response events for building 3 part 2.



(a) 07.00-08.00

(b) 01.00-03.00

(c) 01.00-03.00

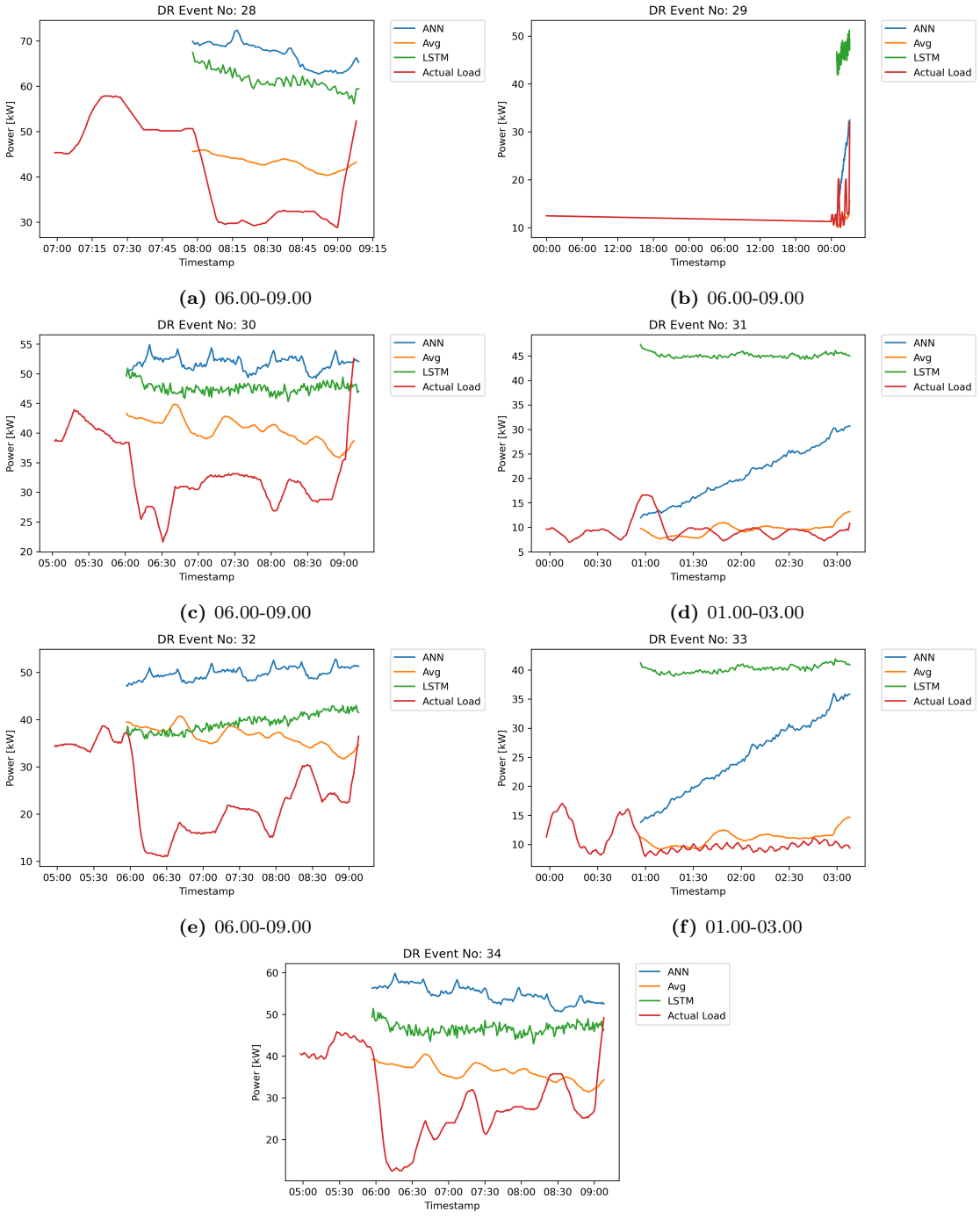
(d) 06.00-09.00

(e) 01.00-03.00

(f) 06.00-09.00

(g) 01.00-03.00

Figure 19: An illustration of baseline estimation on additional demand response events for building 3 part 3.



(a) 06.00-09.00

(b) 06.00-09.00

(c) 06.00-09.00

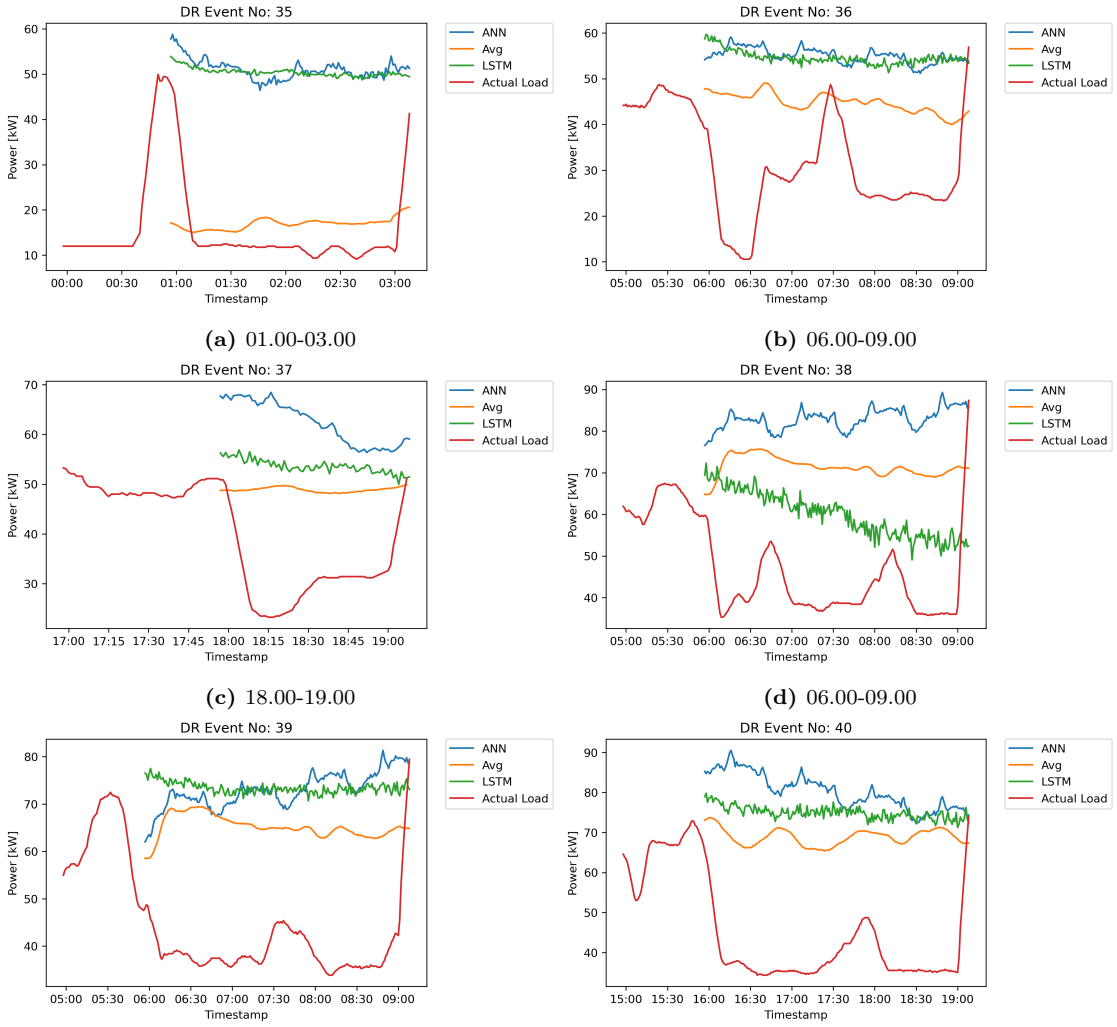
(d) 01.00-03.00

(e) 06.00-09.00

(f) 01.00-03.00

(g) 06.00-09.00

Figure 20: An illustration of baseline estimation on additional demand response events for building 3 part 4.



(a) 01.00-03.00

(b) 06.00-09.00

(c) 18.00-19.00

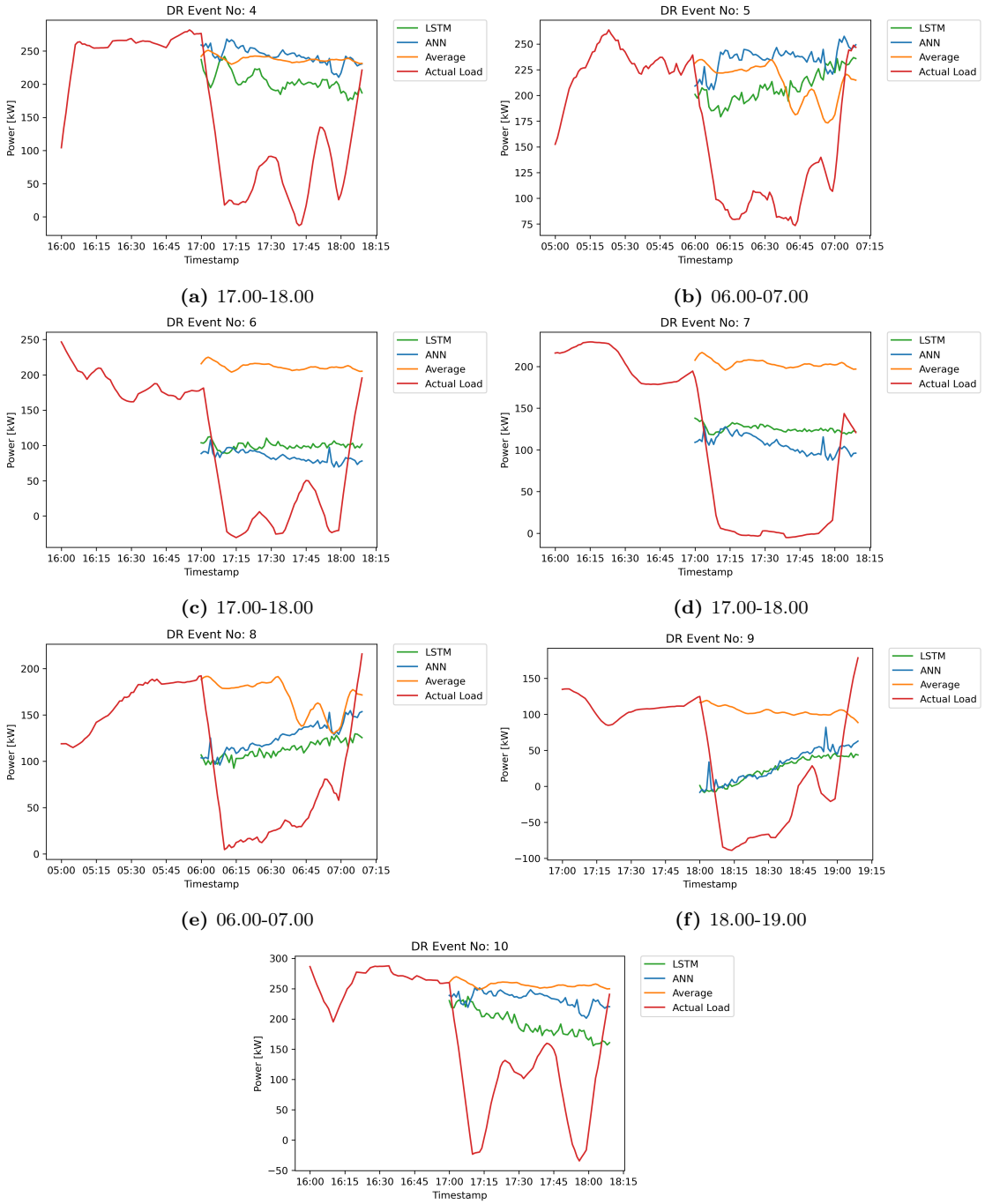
(d) 06.00-09.00

(e) 06.00-09.00

(f) 16.00-19.00

Figure 21: An illustration of baseline estimation on additional demand response events for building 3 part 5.

7.4 Baseline estimation for additional demand response events at building 5



(a) 17.00-18.00

(b) 06.00-07.00

(c) 17.00-18.00

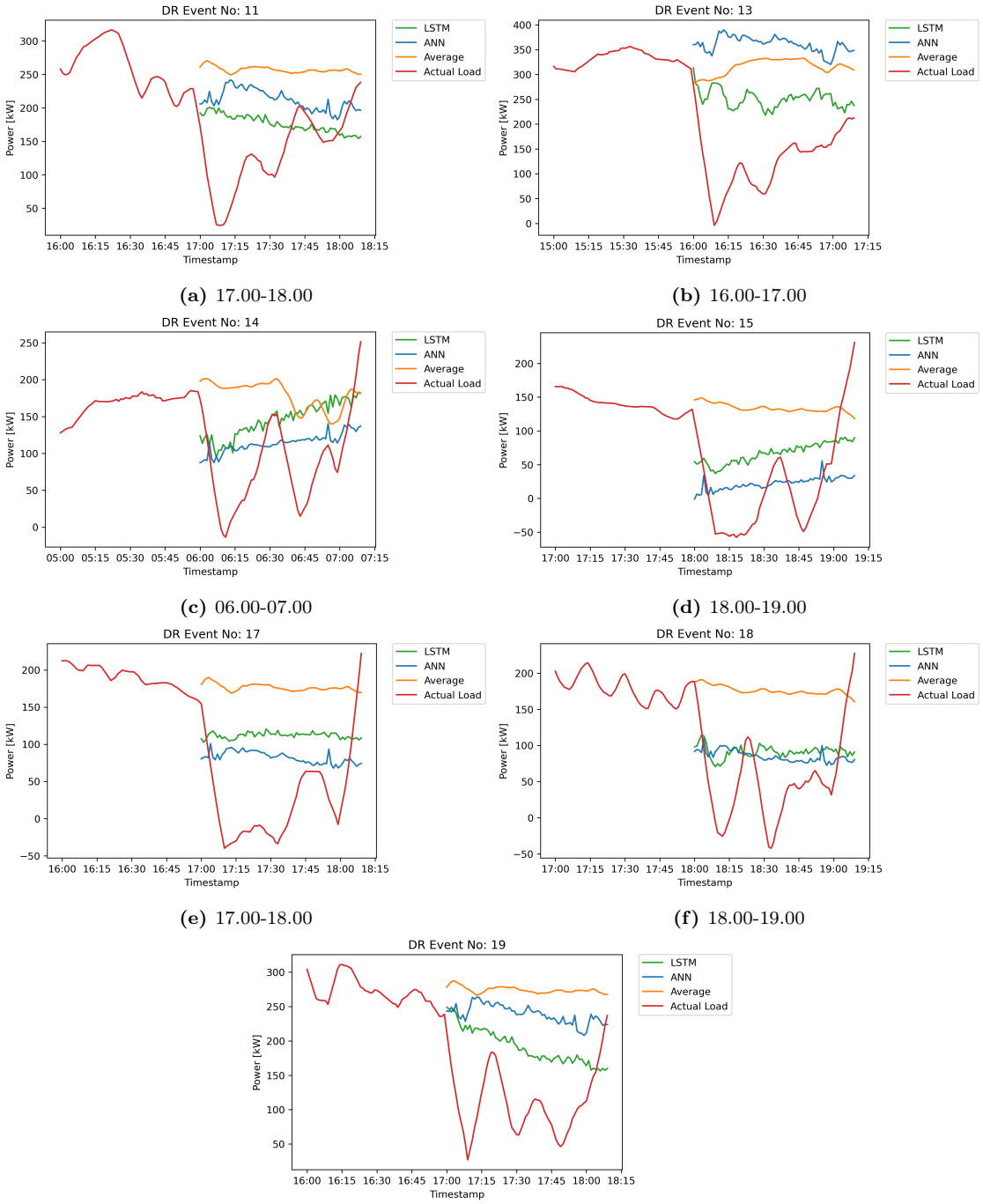
(d) 17.00-18.00

(e) 06.00-07.00

(f) 18.00-19.00

(g) 17.00-18.00

Figure 22: An illustration of baseline estimation on additional demand response events part 1.



(a) 17.00-18.00

(b) 16.00-17.00

(c) 06.00-07.00

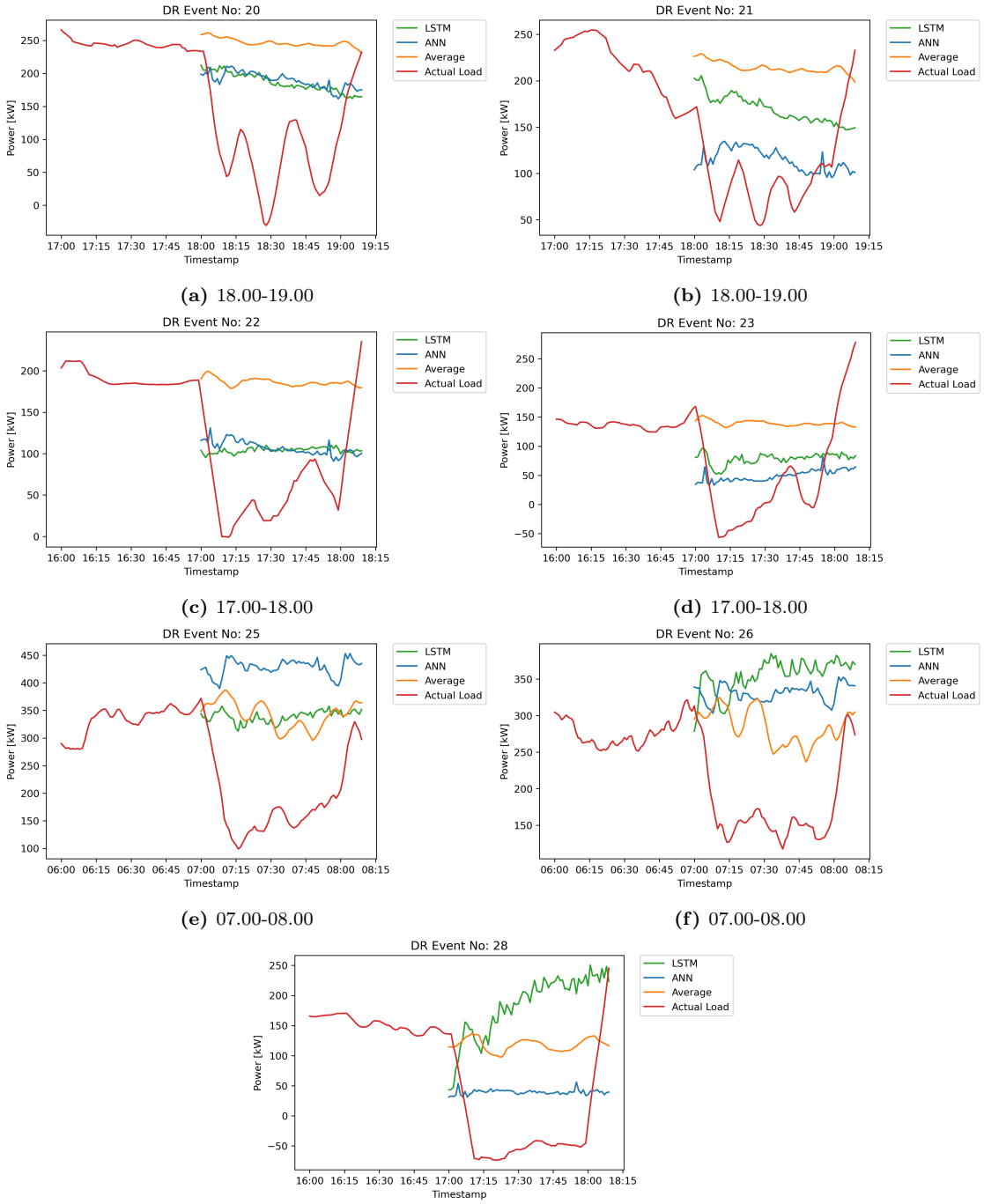
(d) 18.00-19.00

(e) 17.00-18.00

(f) 18.00-19.00

(g) 17.00-18.00

Figure 23: An illustration of baseline estimation on additional demand response events part 2.



(a) 18.00-19.00

(b) 18.00-19.00

(c) 17.00-18.00

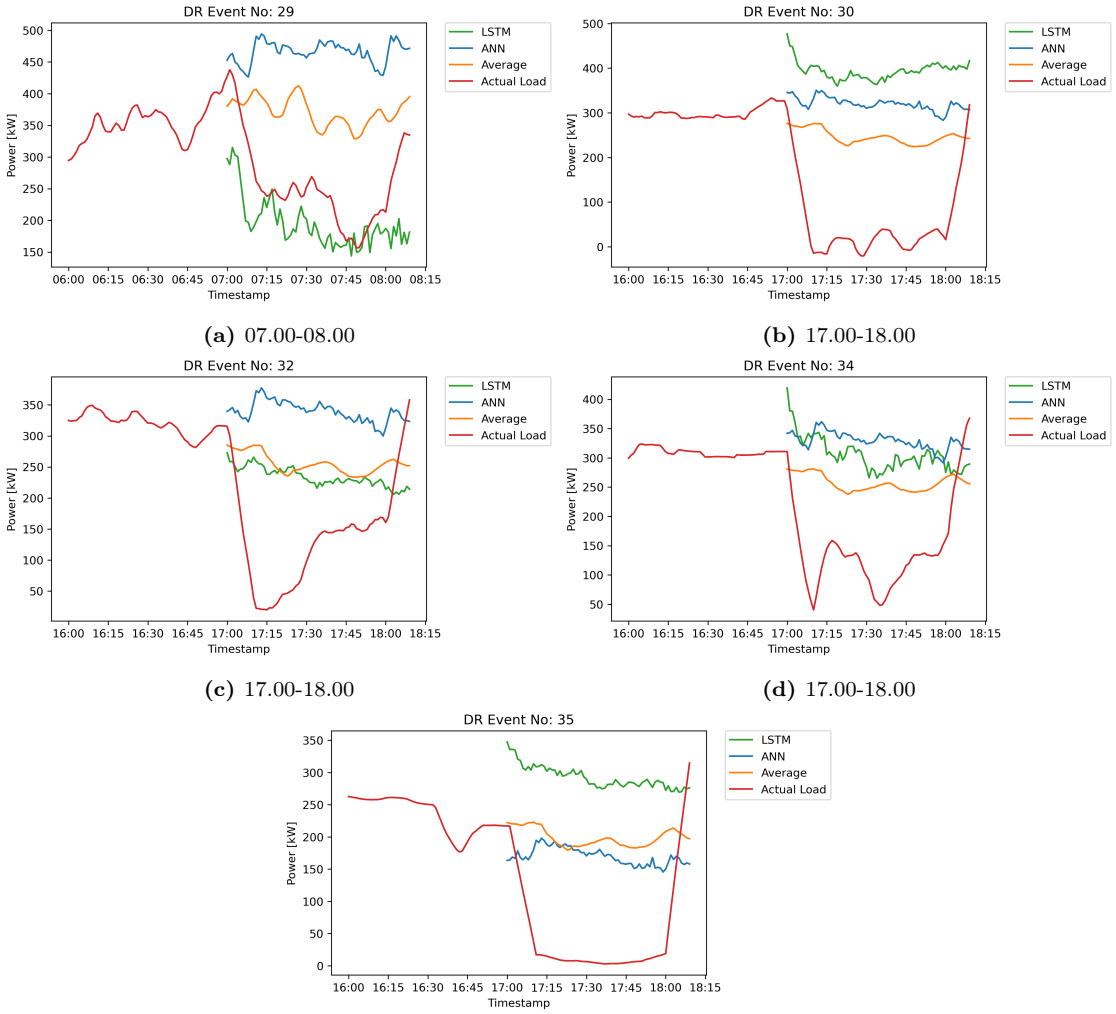
(d) 17.00-18.00

(e) 07.00-08.00

(f) 07.00-08.00

(g) 17.00-18.00

Figure 24: An illustration of baseline estimation on demand response events for the three different estimation methods part 3



(a) 07.00-08.00

(b) 17.00-18.00

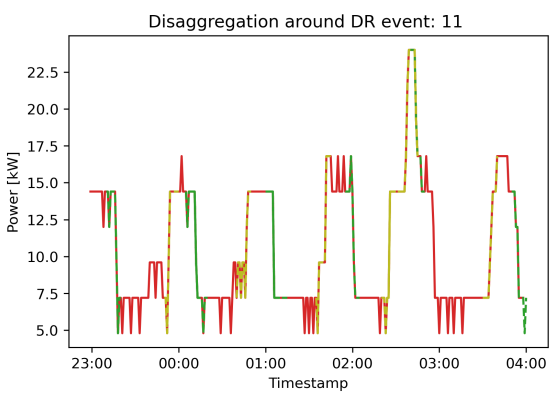
(c) 17.00-18.00

(d) 17.00-18.00

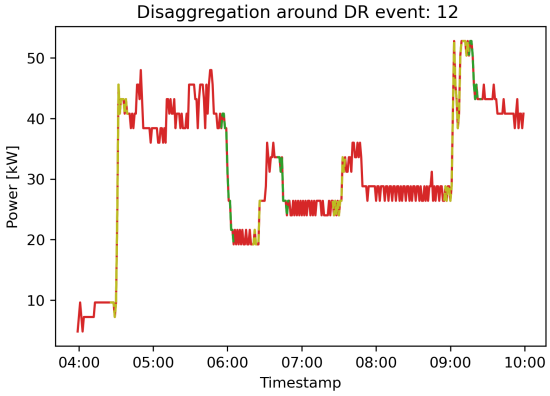
(e) 17.00-18.00

Figure 25: An illustration of baseline estimation on demand response events for the three different estimation methods part 4.

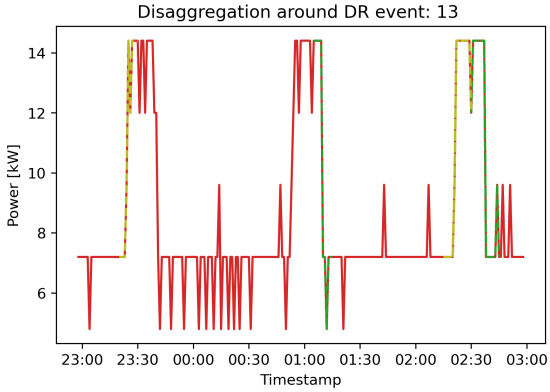
7.5 Load disaggregation for additional demand response events at building 3



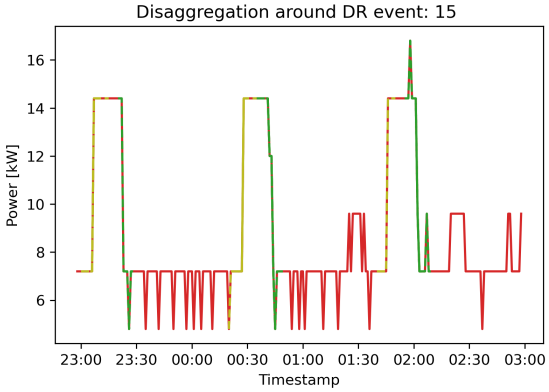
(a) 01.00-03.00



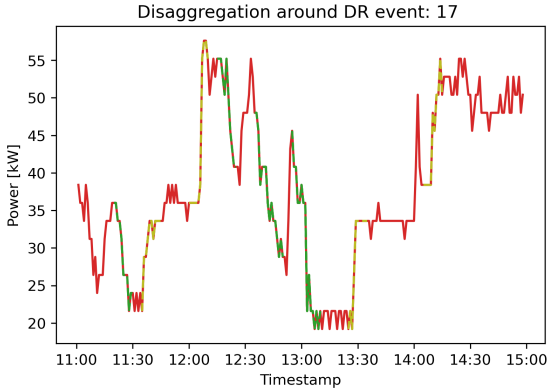
(b) 06.00-09.00



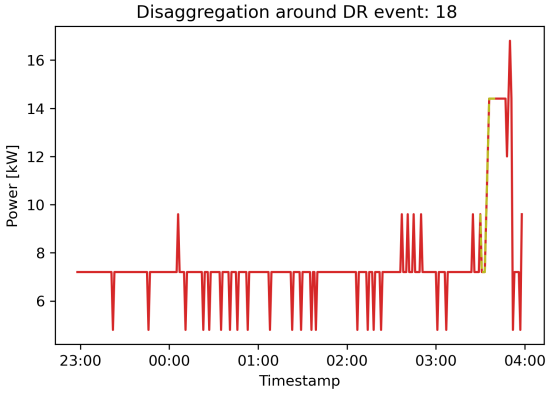
(c) 01.00-03.00



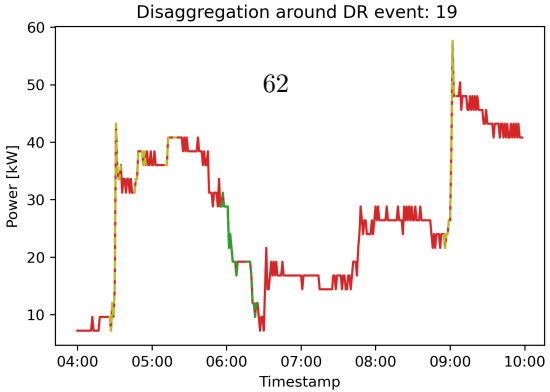
(d) 01.00-03.00

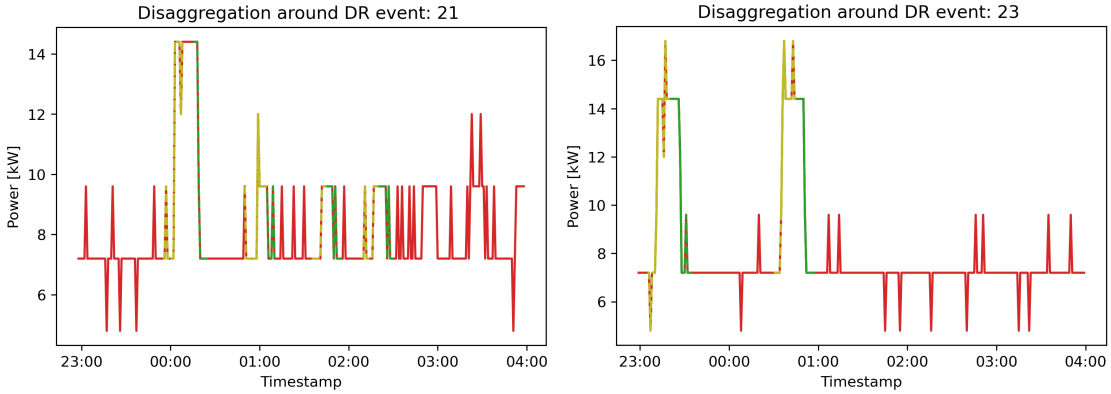


(e) 13.00-14.00



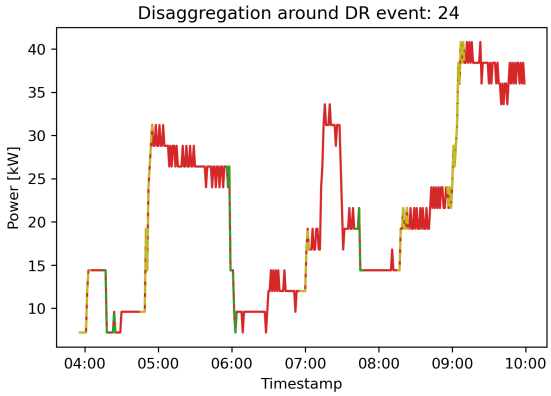
(f) 01.00-03.00



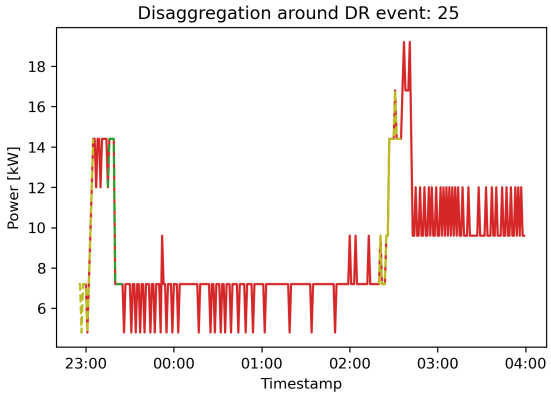


(a) 01.00-03.00

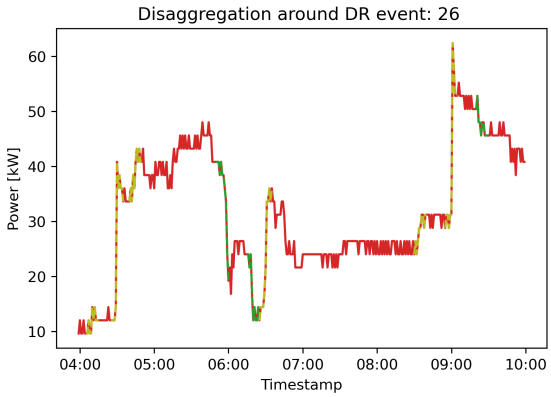
(b) 01.00-03.00



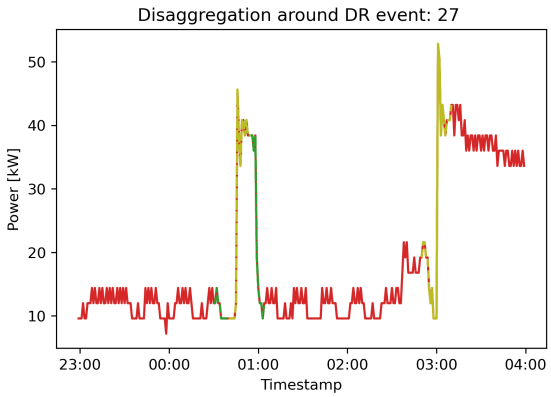
(c) 06.00-09.00



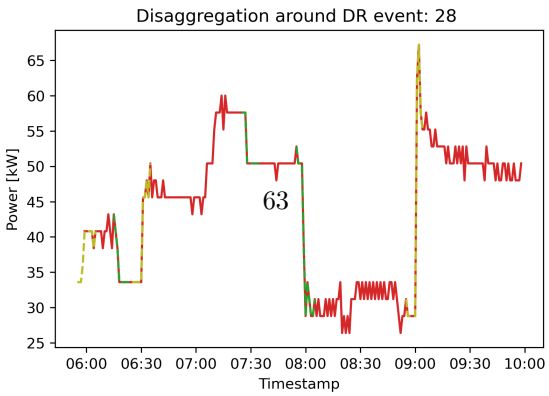
(d) 01.00-03.00



(e) 06.00-09.00

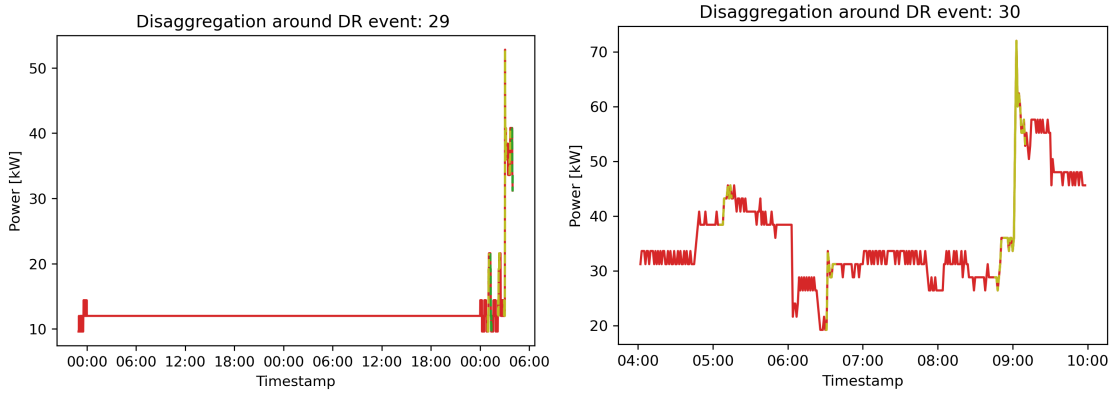


(f) 01.00-03.00



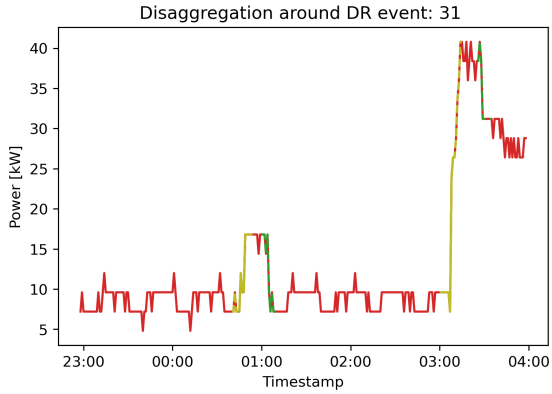
(g) 08.00-09.00

Figure 27: An illustration of the load disaggregation method applied at rest of demand response events

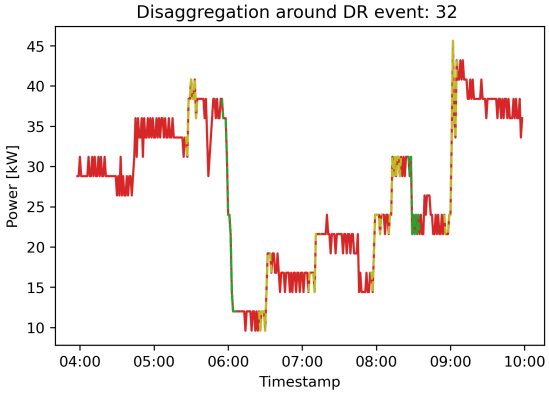


(a) 18.00-19.00

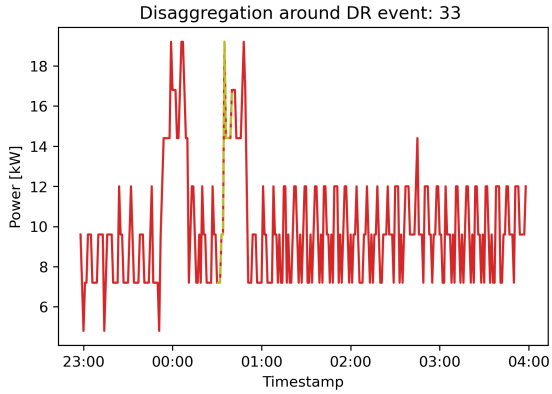
(b) 06.00-09.00



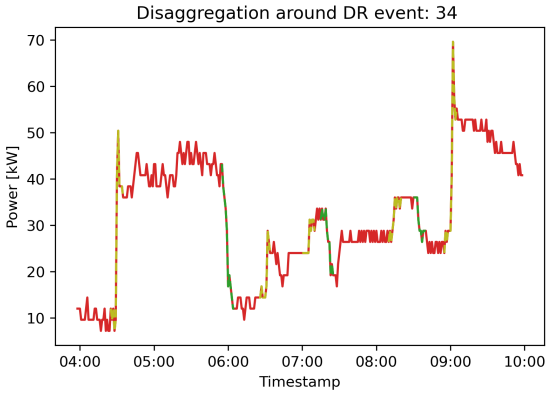
(c) 01.00-03.00



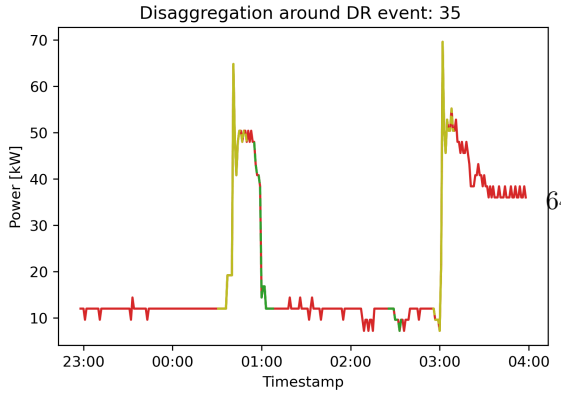
(d) 06.00-09.00



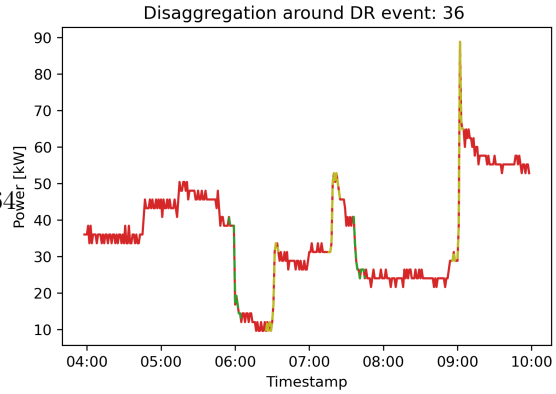
(e) 01.00-03.00



(f) 06.00-09.00

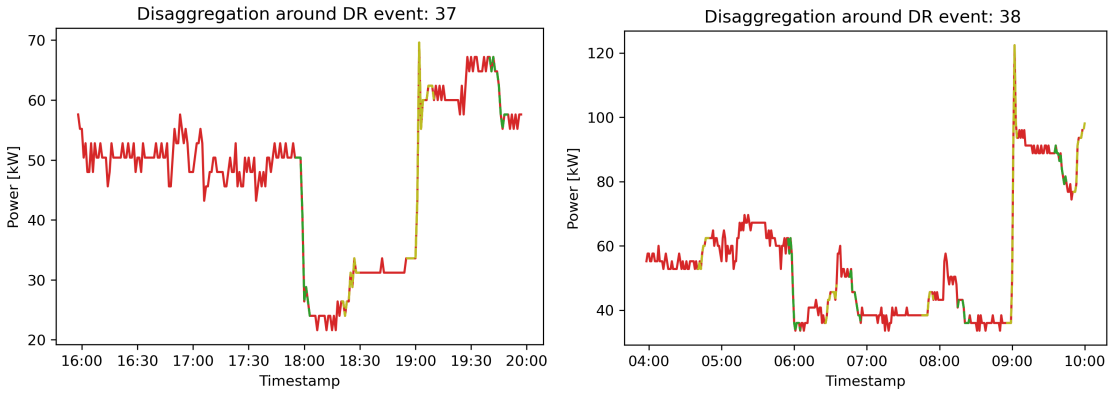


(g) 01.00-03.00



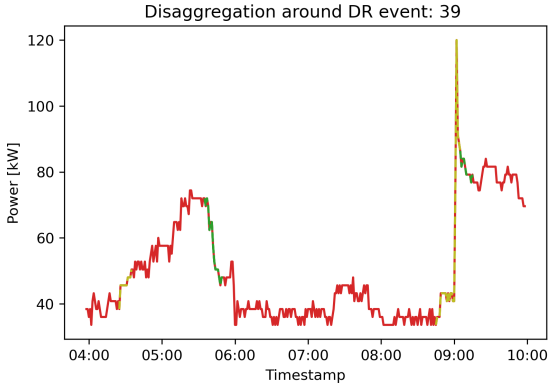
(h) 06.00-09.00

Figure 28: An illustration of the load disaggregation method applied at rest of demand response events building 3 part 3

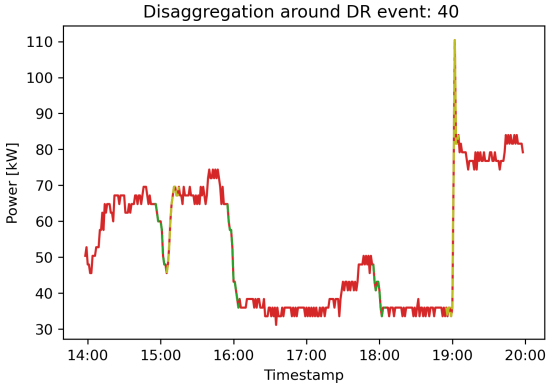


(a) 18.00-19.00

(b) 06.00-09.00



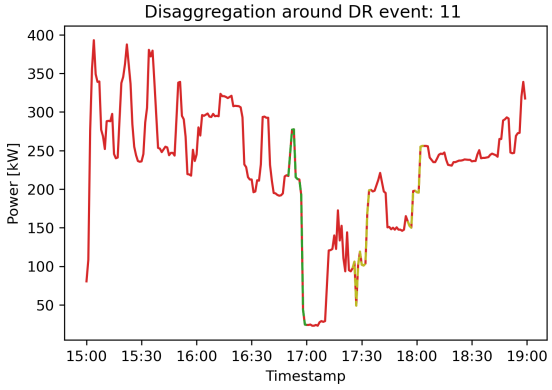
(c) 06.00-09.00



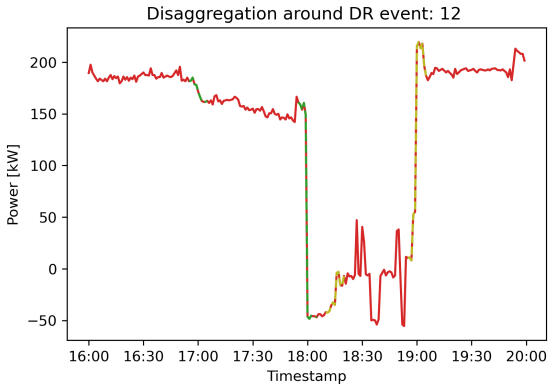
(d) 16.00-19.00

Figure 29: An illustration of the load disaggregation method applied at rest of demand response events building 3 part 4. T

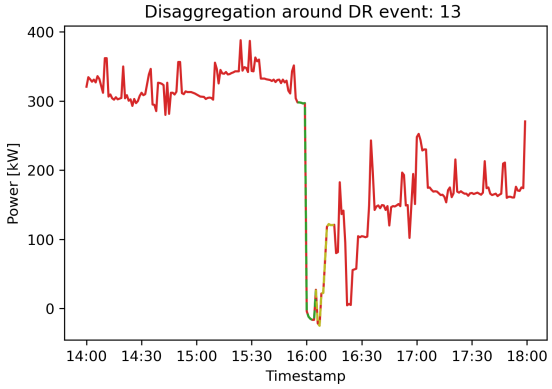
7.6 Load disaggregation for additional demand response events at building 5



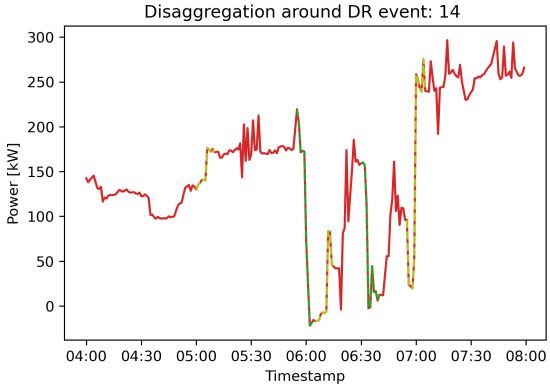
(a) 17.00-18.00



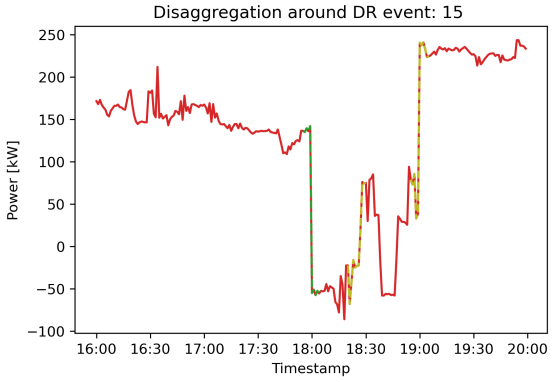
(b) 18.00-19.00



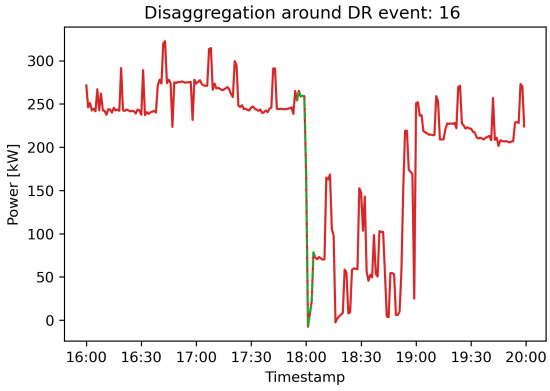
(c) 16.00-17.00



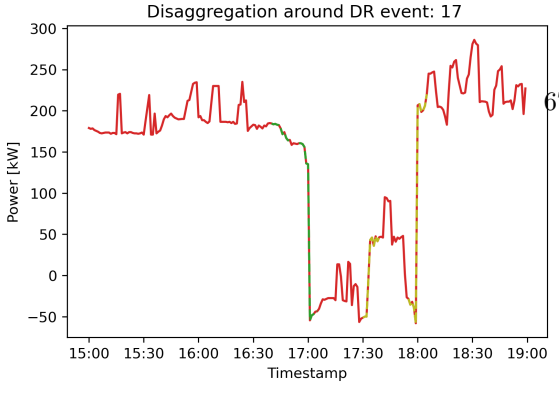
(d) 06.00-07.00

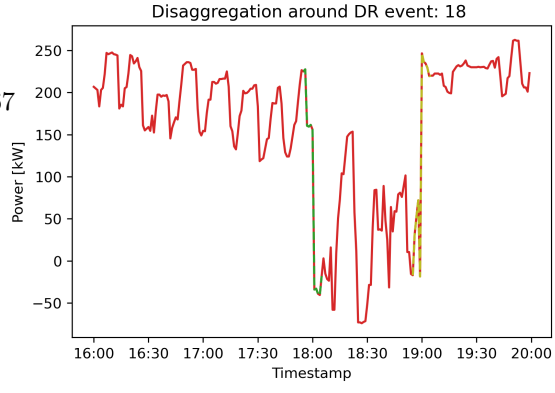


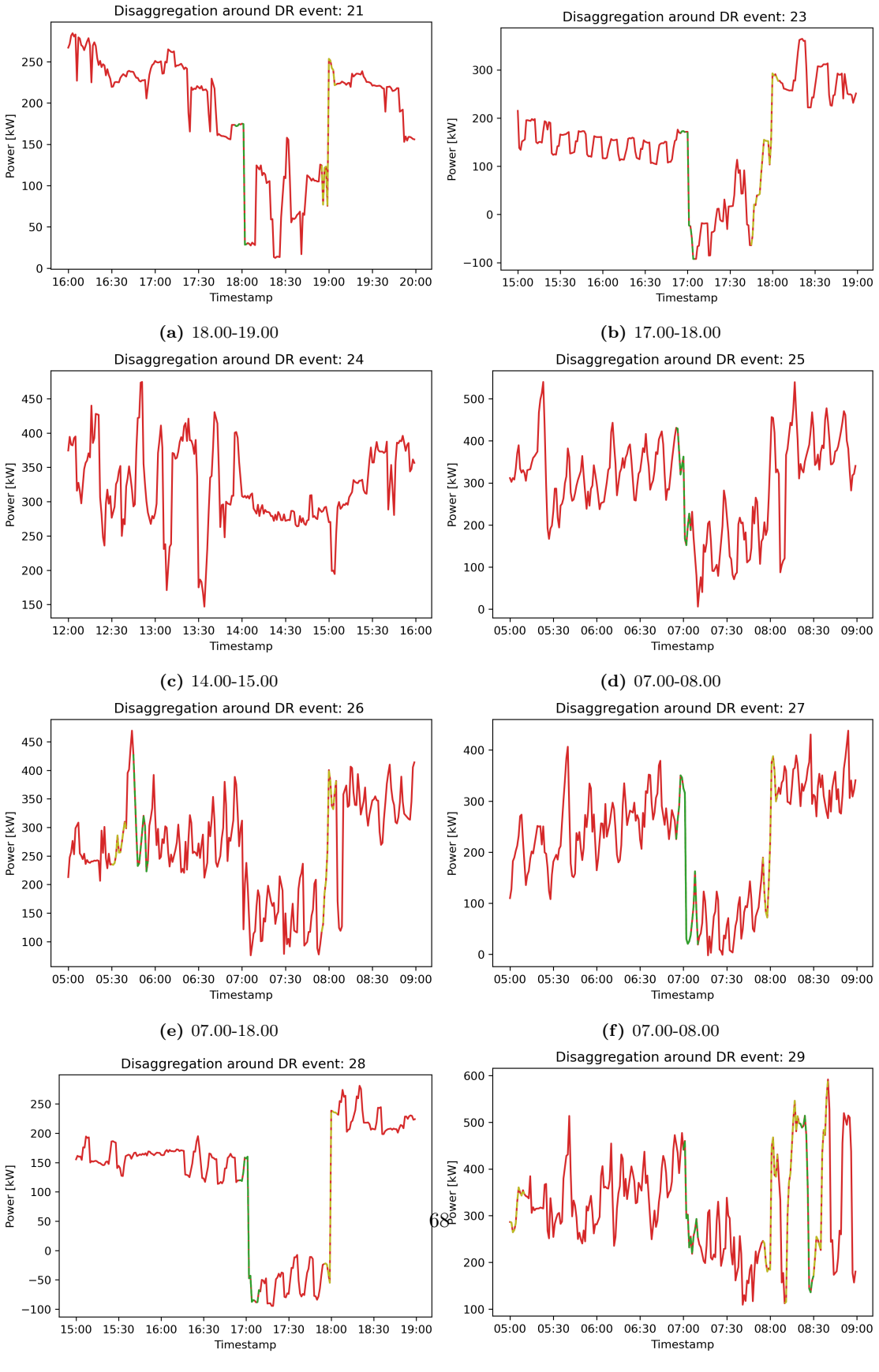
(e) 18.00-19.00



(f) 18.00-19.00







(a) 18.00-19.00

(b) 17.00-18.00

(c) 14.00-15.00

(d) 07.00-08.00

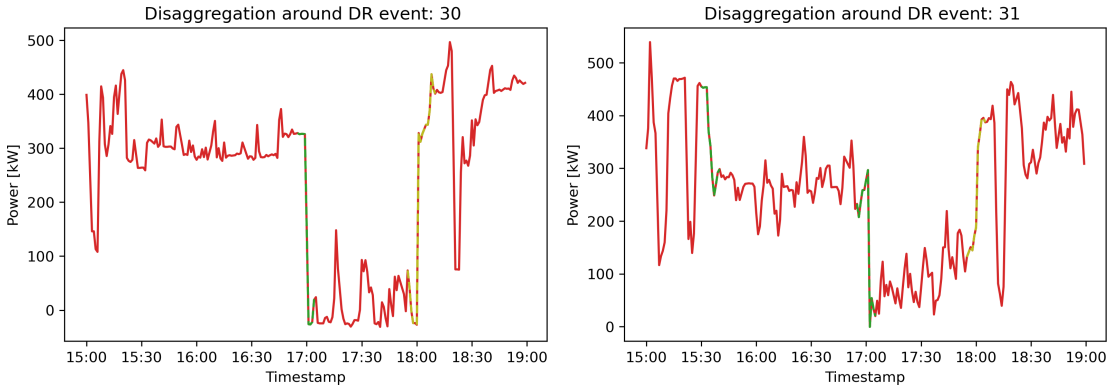
(e) 07.00-18.00

(f) 07.00-08.00

(g) 17.00-18.00

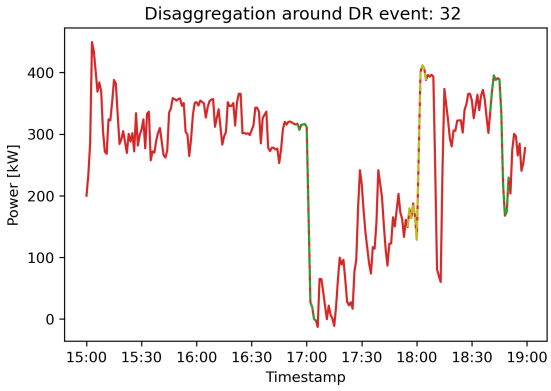
(h) 07.00-08.00

Figure 31: An illustration of the load disaggregation method applied at rest of demand response events building 5 part 2.

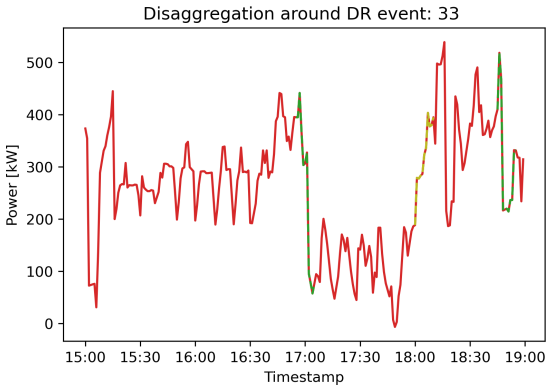


(a) 17.00-18.00

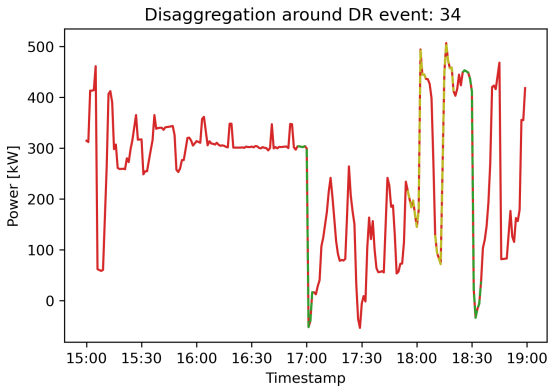
(b) 17.00-18.00



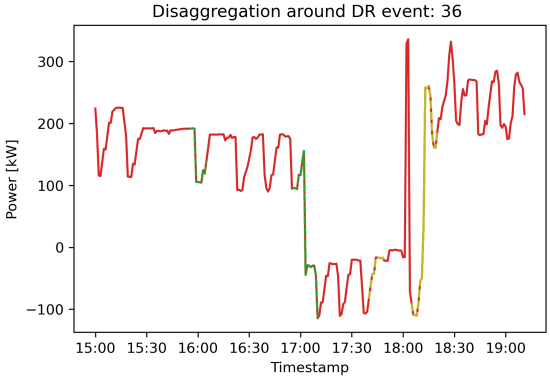
(c) 17.00-18.00



(d) 17.00-18.00



(e) 17.00-18.00



(f) 17.00-18.00

Figure 32: An illustration of the load disaggregation method applied at rest of demand response events building 5 part 3.