Tampere University

Marja Vilkko

ANALYZING LOCAL FLEXIBILITY VERIFYING METHODS AS AN ENABLER OF FUTURE FLEXIBILITY MARKETS

Master's Thesis
Faculty of Information Technology
and Communication Sciences
Sami Repo
Tuomas Rauhala
11/2021

ABSTRACT

Marja Vilkko: Analyzing local flexibility verifying methods as an enabler of future electricity markets.
Master’s thesis
Tampere University
Master’s degree in Electrical engineering
11/2021

The share of variable and weather-dependent electricity production in Finland is growing. Also, the production is scattered around the country. This creates problems with congestion management and remaining the balance between supply and demand. These problems have various types of solutions, and one emerging solution is flexibility. The focus in this thesis is demand-side flexibility. Enabling demand-side flexibility into electricity markets has its problems and problems with settlement and verifying is the subject of this thesis.

When flexible resources are aggregated, especially by an independent aggregator (IA), there is a problem with the settlement process. Every party engaged in the flexibility event should get compensation the right amount and differences in imbalances should be settled. For this reason, a proper verifying and settlement process should take place in every flexibility event.

Verifying is done so that the purchased flexibility is realized at the right time with the correct volume. For the verifying process, there are various methods, and seven of them are examined in this thesis. The verifying methods tested in this thesis are baseline methods, that use the historical data of the examined site. Baseline methods create a baseline for the event day, that represents the load that would have taken place if the flexibility event did not happen. The difference between the baseline and the actual load can be determined by the realized flexibility.

In this thesis, seven verifying methods were tested with data from Tampere University Hervanta campus’ Kampusareena -building. Based on the tests, three methods that used adjustment factors performed better than the three methods that did not use adjustment factors. The seventh method was the Comparable day method, and it performed better than methods with no adjustment but worse than methods with adjustment. The performance was evaluated with quantitative and qualitative methods. Based on those, one method over another could not be determined and the results should be looked as a reference for further work.

More tests with the methods in varying situations should be done to gain more information about them. In this thesis, the Kampusareena was examined alone, but in reality, the aggregator would control the electric vehicle charging or the ventilation and offer them as flexibility. The changes in their load should be determined and the methods tested in this thesis would probably not be able to do that. For further work, the responsible party of the verifying and the level of verifying should be decided.

Keywords: flexibility, demand-side flexibility, verifying, baseline, imbalance settlement, transmission system operator, independent aggregator

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

TIIVISTELMÄ

Marja Vilkko: Paikallisen jouston todentamismetodien analysointi tulevaisuuden sähkömarkkinoiden mahdollistajana
Diplomityö
Tampereen yliopisto
Sähkötekniikan diplomi-insinööri
11/2021

Vaihtelevan ja sääriippuvaisen sähköntuotannon osuus Suomen sähköntuotannosta on kasvussa. Tuotanto on myös hajautettuna ympäri maata. Tämä luo ongelmia siirtojenhallintaan ja kulutuksen ja tuotannon tasapainon ylläpitoon. Näille ongelmille on useita ratkaisuja, joista yksi nouseva ratkaisu on jousto. Tämän työn keskiössä on kysyntäjousto. Kysyntäjouston mahdollistaminen sähkömarkkinoille ei ole ongelmatonta ja ongelmat todennuksen ja selvityksen kanssa ovatkin tämän työn aiheena.

Kun joustokohteita aggregoidaan, erityisesti itsenäisen aggregaattorin kohdalla, tulee ongelmia selvitys prosessin kanssa. Jokainen joustotapahtumaan osallistuva on oikeutettu kompensaatioon sekä jokainen tase johon jousto vaikuttaa pitäisi korjata todenmukaiseksi. Tästä syystä kunnollinen todentaminen ja selvitys prosessi tulisi olla osa jokaista joustotapahtumaa.

Todentamista tehdään, jotta voidaan varmistaa, että ostettu joustotapahtuma tapahtuu oikeaan aikaan ja oikealla suuruudella. Todentamiseen on useita menetelmiä, joista seitsemää on tarkasteltu tässä työssä. Todentamismenetelmät, joita tässä työssä on testattu, ovat baseline -menetelmiä, jotka käyttävät historiadataa joustokohteesta. Baseline -menetelmä luo vertailukäyrän tapahtumapäivälle, joka kuvastaa sitä kuromaa, joka olisi tapahtunut, ellei joustotapahtumaa olisi ollut. Tämän baselinen ja todellisen kuroman välistä saadaan toteutunut jousto.

Tässä työssä seitsemän todennusmenetelmää testattiin Tampereen yliopiston Hervannan kampuksen Kampusareena -rakennuksen datalla. Tehtyjen testien perusteella, kolme menetelmää, joiden laskemiseen käytettiin korjauskerrointa, suoriutuivat paremmin kuin ne kolme menetelmää, joiden laskemiseen ei käytetty korjauskerrointa. Seitsemäs menetelmä, Comparable day -menetelmä, suoriutui paremmin kuin menetelmät ilman korjauskerrointa, mutta huonommin kuin menetelmät korjauskertoimen kanssa. Menetelmien toimivuutta arvioitiin kvantitatiivisilla ja kvalitatiivisilla menetelmillä. Niiden perusteella, ei voida sanoa mikä menetelmä olisi yksinkertaisesti parempi kuin muut ja tuloksia tulisi tarkastella pohjana jatkoselvityksille.

Jotta menetelmistä saataisi enemmän tietoa, testejä niillä erilaisissa kohteissa tulisi vielä suorittaa. Tässä työssä Kampusareenaa tarkasteltiin yhtenä kohteena, mutta todellisuudessa aggregaattori hallitsisi esimerkiksi sähköautolatauspistettä tai ilmanvaihtoa ja tarjoaisi niitä joustona. Muutokset sähköautolatauksen tai ilmanvaihdon kuormassa tulisi pystyä todentamaan ja menetelmät, joita tässä työssä käytettiin, eivät todennäköisesti sitä pystyisi tekemään. Jatkoselvitystarpeena ovat myös kysymykset siitä, kuka on vastuussa todentamisesta ja millä tasolla todentamista tulisi tehdä.

Avainsanat: jousto, kysyntäjousto, todentaminen, baseline, taseselvitys, kantaverkkoyhtiö, itsenäinen aggregaattori

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck –ohjelmalla.

PREFACE

This thesis was written for Fingrid Oyj as a part of the OneNet project, which is part of the multinational European Union's Horizon 2020 project. OneNet researches the possibilities of flexibility and its ways to be part of the electricity markets. This thesis is part of a study about flexibility register. The goal of the thesis was to provide information about different verifying methods and to examine how verifying can be a part of the future electricity and flexibility markets.

I would like to thank my thesis supervisor professor Sami Repo, my supervisor at Fingrid Tuomas Rauhala, and specialist Jukka Rinta-Luoma for the excellent guidance through the thesis process. Also, a huge thank to my steering group, and all the other experts in our Market Innovations team for the wise words and great guidance to the fascinating world of electricity markets.

A huge thanks to my friends at Tampere University for making the years here so enjoyable that leaving is going to be hard. I have received so much more than I could have ever imagined. Special thanks to Jussi, Sähkökilta, and Tampereen Teekkarit.

Lastly, a great thanks my dad, who has been my support through my years in school, and who inspired me to start studying electrical engineering.

Helsinki, 19th November 2021

Marja Vilkko

CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

BRP	Balance Responsible Party
BSP	Balancing Service Provider
CDF	Cumulative Distribution Function
CEP	Clean Energy Package
DA	Day-ahead
DR	Demand Response
DSM	Demand Side Management
DSO	Distribution System Operator
EC	Energy Conservation
EE	Energy Efficiency
EU	European Union
ES	Energy Storage
EV	Electric Vehicle
FCR-D	Frequency Containment Reserve for Disturbances
FCR-N	Frequency Containment Reserve for Normal Operation
FFR	Fast Frequency Reserve
IA	Independent Aggregator
ID	Intraday
ISR	Imbalance Settlement Responsible
LV3	Laatuvahti 3, measuring device used in Kampusareena
MAPE	Mean Absolute Percentage Error
MBL	Maximum Base Load
MBMA	Meter Before – Meter After
MDA	Metered Data Aggregator
MGA	Metering Grid Area
MGO	Metering Generation Output
NEMO	Nominated Electricity Market Operator
PV	Photo Voltaic
RE	Retailer
RES	Renewable Energy Source
RMSE	Root Mean Square Error
SO	System Operator
SP	Service Provider
TSO	Transmission System Operator
VPP	Virtual Power Plant
VRES	Variable Renewable Energy Source

1. INTRODUCTION

In the past, energy was produced in large power plants. Energy system operators could rely on stable and controllable energy production. The electricity grid was planned in a way that electricity flow was from the power plant to the customer. The grid was enforced in places where the consumption and production were, and bottlenecks were solved with a new and stronger grid. The energy was produced with coal, oil, and other cheap fossil fuels. Energy systems relayed on stable energy production by the power plants. Electricity could not be stored in a way it would have been beneficial for the energy system.

Ongoing climate change is forcing the energy industry and power systems to change [1]. Climate change has had many impacts on the decision-making in countries all over the world. One of these is the Clean Energy Package (CEP) [2] from European Union (EU) which is a framework for decision-making in EU countries. Because of climate change, industries and countries are trying to achieve more environment-friendly ways to produce energy. Also aiming to be carbon-neutral or even carbon-free is highly popular [3]. In Finland transmission system operator has a natural monopoly and has pressure to act among the climate change and CEP [4]. Climate change is forcing everyone to evaluate their actions, where the materials come from, and how the energy is produced.

CEP, climate change, and the aims to be carbon-free or -neutral are all thriving to increase the amount of renewable energy production. New renewable energy sources (RES) such as wind and solar power are increasing all over the world. RES is shaping the energy system and the energy production profile. New additions and innovations of the RES are causing energy transitions all over the world [5]. Because of these actions and changes, energy production can locate anywhere, with small or large power plants. And with variable renewable energy sources (VRES) comes batteries and energy storages (ES) to balance the variable energy production coming from VRES.

An increasing amount of RES will have a good impact on climate change [6]. With small steps, the energy production profile is more environmentally friendly. If the number of RES increases enough, the energy system does not have to rely anymore on fossil fuels. With increasing RES the reliability of the energy supply does not increase, rather it will decrease if correct actions to prevent it are not taken. The problem with RES is that it is unpredictable, highly uncontrollable and most of the time depending on the weather and climate [7]. Especially VRES is creating more uncertainty to the production profile. Variable electricity production creates fluctuation of voltage and frequency to the electricity grid and creates a problem with frequency or voltage. [5] [8]

The solution to fluctuation that VRES creates, is flexibility. Flexibility can offer solutions for example to frequency management and congestion management [9]. In CEP flexibility is also one of the topics of the laws that are aiming to secure the integration of RES [2]. Various flexibility solutions will enable for example demand-side load, energy storages, and flexible load into the electricity markets. Flexibility can be a solution to balance the production and consumption, and to congestion manage. But, enabling flexibility into flexibility markets has its problems.

The mechanisms of flexibility are different, but some act with the price of electricity and some with an agreement. The agreements are different for each resource, but they rely on the price and current transmission, frequency, or voltage situation. Problems come along when imbalance settlement isn't enough to cover who offered flexibility and what were the actual flexibility events that happened. Flexibility can be quite local, which cannot be seen directly within the imbalance settlement.

Transmission system operators (TSOs) have a part as a market enabler, to make sure flexible resources can participate in the markets without a particularly large risk of not getting paid for the flexibility offered. Flexibility verifying and settlement will be the answer to this problem. Verifying is a process where the flexibility event is analyzed if it happened, and at what rate. Flexibility can be offered from a source where there is a base-load, normal use, and in the case of flexibility, that normal load changes. Different kinds of baseline methods can be used to determine what would have been the normal use compared to the flexibility event and from that, the amount of flexibility can be calculated.

In the literature, many different baselines and verifying methods are presented. Some of them rely on the historical data from the source, and some need information about weather and climate to determine what could have been the normal use [10]. The verifying method can be implemented to a single flexible resource, or to aggregated sources that contain the same type of resources or different types of resources. The verifying method will be more complicated as if it must be able to present multiple types of resources, and the accuracy of the method might decrease.

This thesis will be examining a case example of a large office building that contains photovoltaic (PV), electric vehicle (EV) charging, air condition, and other flexible sources. The information about this case examination is presented later in this thesis. The goal of the examination of the office building is to test how well the chosen verifying methods work in this specific case. This information and collected information from literature has formed a comparison between the methods. The comparison is done with TSO and the bigger picture in mind. The comparison aims to find verifying methods that would be usable at the TSO level. The main research questions of this thesis are:

- What are the verifying methods that can be implemented in this case examination?
- What are the factors that make verifying methods usable at the TSO level?

In chapter four the case study of this thesis is introduced. The data and site used for the test are presented and examined, and the tests that are done are gone over more precisely. In chapter five is presented the results from the tests. Also, the evaluation process is done to the results and the results are presented. Chapter six is the discussion, and the results are examined with another point of view presented in this thesis, along with the problems that the future will hold. Also, the further studies that should take place are discussed. In chapter seven is the conclusion of the thesis.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 957739.

2. FLEXIBILITY IN ELECTRICITY SYSTEM

In the literature, there are many definitions for flexibility. One definition of flexibility is presented in the article “Comprehensive classification and characterization of power system flexibility resources” [11]:

> *”The ability of power system operation, power system assets, loads, energy storage assets and generators, to change or modify their routine operation for a limited duration, and responding to external service request signals, without inducing unplanned disruptions.”*

This definition defines clearly, which resources can be counted as flexibility, what is the period that it is considered flexibility, and what are the reasons that activate flexibility to it be referred to as flexibility [11]. In this thesis, this definition will be used to clarify what is flexibility in the power system.

Flexibility is a complex and large entity, and this chapter will compose the important notations of flexibility considering this thesis.

2.1 The need of flexibility

The most straightforward need for flexibility comes from the higher penetration of VRES in the electricity system. The decarbonization of the power generation, which is linked to the growing number of VRES, is also the biggest driver for energy transition [5]. The increasing number of renewable electricity resources in the grid is focusing a lot on wind and solar energy production. Numbers to back this is shown in Table 1. In Finland, the connection of wind to the grid will cause the most challenges and changes in the next few years [12].

Table 1** Electricity production in Finland by sources. **[13]

SOURCE	2019	2020
NUCLEAR POWER	34,70 %	33,90 %
HYDROPOWER	18,50 %	23,70 %
BIOMASS	18,90 %	15,30 %
WIND POWER	9,10 %	11,80 %
NATURAL GAS	5,5 %	5,8 %
COAL	7,1 %	4,4 %
PEAT	4,3 %	3,3 %
WASTE FUELS	1,3 %	1,1 %
SOLAR POWER	0,2 %	0,4 %
OIL	0,4 %	0,3 %

Table 1 presents the shares of each production source of electricity production in Finland. The percentages are from the total electricity production of 66 TWh. The total production was the same in 2019 and 2020. As shown, the share of electricity produced with renewable resources is increasing. In 2019 the total share of electricity produced with renewable electricity resources was 47% and carbon-neutral 81 %. The same values in 2020 were 51 % and 85 %. From the total electricity consumption of 81 TWh, 18,5 % was imported electricity. [13]

When a higher amount of renewable and variable electricity resources is connected to the grid, the more electricity grid and system changes from the old electricity production profile. As this means good for the environment, the electricity grid and markets will face problems deploying all the potential of VRES into them. Article “Electricity generation from renewables in the United States: Resource potential, current usage, technical status, challenges, strategies, policies, and future directions” [14] presents various challenges that prevent an integrating large number of VRES into the electricity grid. The most important challenges considering the electricity markets and grid are the distributed nature of RES, maintaining the balance between supply and demand, and grid capacity restrictions.

The challenge of maintaining the balance between supply and demand is linked to the variability of RES. VRES are usually dependent on the weather, whenever it is windy or sunny. As the goal is not to curtail the production from VRES because then the purpose of using them to replace fossil-fuelled power plants would be dismissed. As the production from VRES is weather dependent and not demand dependent, VRES might produce electricity during hours when the demand is low. For this kind of situation, solutions to keep the balance between supply and demand are needed.

Distributed characteristic is natural for VRES, as the smaller unit of wind turbines or solar parks are located where the most value from the wind or sun can be achieved. This creates an electricity system where the production is all over the system, and the topologies of electricity transmission should be considered more carefully. Having production scattered around the system can be helpful during congestion situations. In Finland, the problem is more because most of the wind power is produced in the wester of Finland, and the consumption is in the south of Finland. This can create congestion management issues.

The challenges described above create the need for flexibility. Flexibility solutions are a variable combination of things that can ease the situation with VRES in the electricity system and enable VRES to be connected to the grid. [15] When VRES creates a high amount of electricity in one area, where it is windy, but the consumption in that area might not be enough, flexible demand response or electricity storages can consume that electricity. This way the balance between supply and demand is kept and the electricity system does not have to do additional transmission operations. Another way, if the production from the VRES is lower, but consumption is higher, fast responding generation, discharging electricity storage or demand-side management can be the problem solver.

When dealing with VRES, another solution would be investing in grid hardware. Building more grids and reinforcing the transmission lines, could solve the congestions management issues. But investing that much into the grid would probably not be cost-efficient. Some of the congestion management issues can be and will be solved with a reinforced grid, but some of the highest load profile scenarios are so rare, that the grid's investments would not be cost-effective. For those scenarios flexibility can be used, as an alternative for grid invests and with the grid invests. [12]

2.2 Flexibility solutions

In this thesis, flexibility resources, enablers and actions are altogether called flexibility solutions. This term will hold all the necessary parties needed for flexibility in the power systems.

In subchapter 2.2.1, the classification of flexibility solutions is examined. Classification of these flexibility solutions is important, so that later in this thesis there is a clear understanding of why different kinds of flexibility solutions are investigated more. Subchapter 2.2.2 goes more deeply into the characterization of the flexibility resources. Characteristics of flexibility resources are important because they determine a lot about how much and on which scale the flexibility can be used in a power system examined.

2.2.1 Classification of flexibility solutions

Flexibility solutions can be divided by the party offering the flexibility resources. The crude division could be demand-side, supply-side, network-side, and energy storages [5] [16]. On top of these, some enablers of flexibility should also be considered as a part of flexibility solutions [11]. The classification is presented in Figure 1.

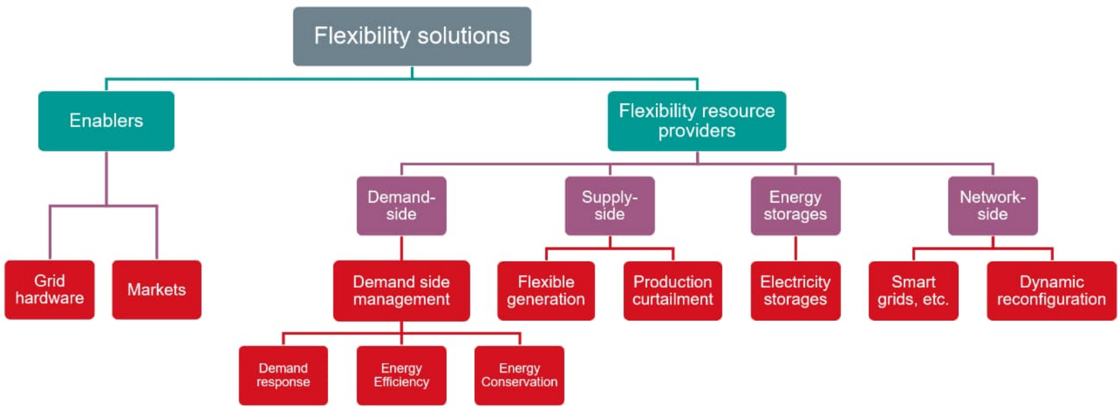

Figure 1 *Classification of the flexibility solutions.* ***[16, 17, 18, 19]***

Demand-side flexibility contains Demand-side management (DSM) which includes demand response (DR), energy efficiency (EE), and energy conservation (EC) [19]. This categorization is done differently in various papers and studies. This division of DSM is used in this thesis because it underlines the many ways demand-side flexibility can be used. Often in literature when flexibility is discussed, only DR and DSM are taken into consideration, this should be noted as there are other flexibility solutions than them. Demand-side flexibility is the focus of this thesis, as other flexibility solutions do not require verifying, in the same way, demand-side flexibility needs. Overall, the growth of interest in demand-side flexibility is natural, as it is the source of flexibility that has not still been utilized in its full potential.

Energy efficiency and energy conservation are similar, and in some articles such as [20] is said that these two mean the same, but there is a difference between them. EE is permanent changes to use energy more efficiently, and EC is more short-term changes. EE can be for example using an automatic thermostat or having good insulation in your home to keep the heat inside the house. EC includes the energy consumers' behaviour such as using the dishwasher when it is full and heating your home only to 19 °C instead of 22 °C. EE and EC are smaller scale flexibility providers and can be affected by changing customers' thoughts about energy efficiency and saving energy. [19] EE and EC are not studied in this thesis further.

Demand response is the actions performed by or with the customer's load. Six programs are used. Peak shaving or peak clipping [21] is reducing high load peaks and lowers the consumption at those times. Valley filling can be used when the cost of electricity is low, and it is more cost-effective for the customers to use electricity during off-peak hours. Load shifting is shaping the load by using electricity during off-peak hours instead of on-peak hours. Typically load shifting requires energy storages, such as water heaters or cold storages. Conservation is similar to EC, and the difference is, that conservation programs are actively trying to make changes so that conservation would occur. Load growth is linked to electrification and the program actively tries to increase it. Load shape is flexible when customers are offered compensation in exchange for shaping their load profile. This includes for example curtailed load or interruptible load. [17, 21]

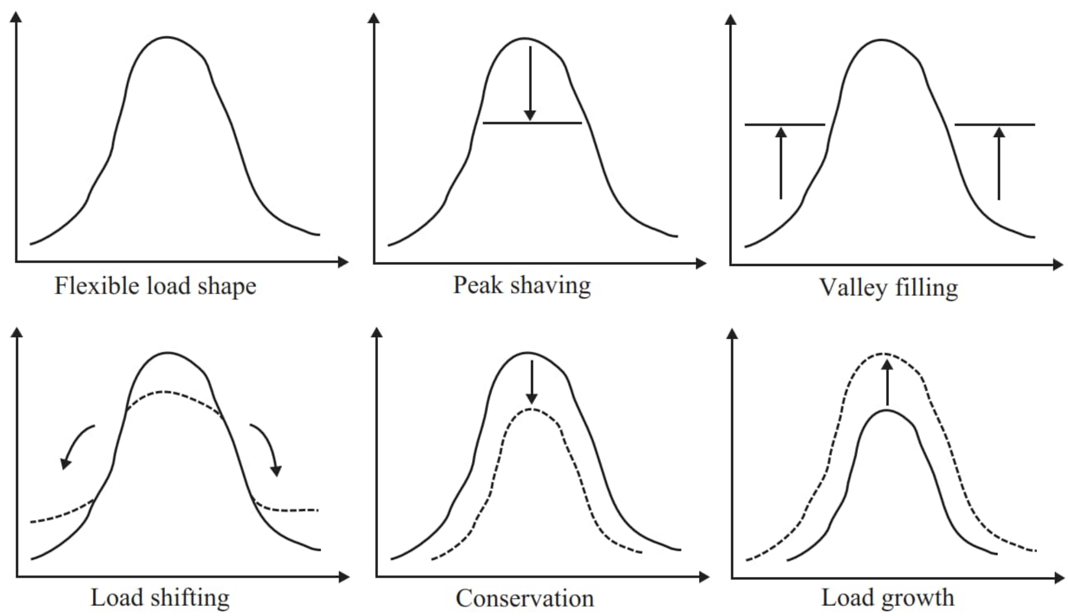

Figure 2** Demand response management programs. **[22], see [17]

The difference between EE, EC, and DR is, that EE and EC are something customers do because they know how to be energy efficient, or they want to reduce their energy consumption. The benefits may be small, as the changes are not always in the times when load reduction is needed. Demand response, on the other hand, can be the same kind of actions done as in EE and EC, but those actions are timed based on which DR program is used and what kind of demand responses are needed.

Another classification used for demand-side flexibility, which is not presented in VIITTAUS, is to define the flexibility by if it is implicit or explicit [23]. Implicit demand-side flexibility means customers reacting to price signals. It is done either with automation or by choice and the goal is for the customers to save money by using electricity when the price is lower. Implicit demand-side flexibility can also be referred to as price-based demand-side flexibility. Explicit demand-side flexibility is usually aggregated and is tradeable in electricity markets. Explicit flexibility is sold beforehand on the electricity markets and the customer is paid specific compensation to change their consumption to the amount that was agreed. Explicit demand-side flexibility is also called incentive-based demand-side flexibility. [23, 24]

Supply-side flexibility is also an important part of the flexibility concept. It can mean the flexible generation of electricity, which will refer to power plants that can easily and quickly change the production [5]. Power plants can be classified by their ability to be flexible, base loads, peaking, or load-following [17]. Baseload power plants such as combined heat and power (CHP) or nuclear plants are run constantly and have only little ability to be flexible. Peaking power plants can be used in extreme conditions when the demand is high or baseload plants are not operating and demand is increasing. Load following plants, which are usually gas turbines or hydropower, can balance the demand and supply at the moment, so they have a higher ability to be flexible. [17] A power plant's ability to be flexible is tied to its ability to have short ramping up or down times [5].

Another way for the supply-side to be flexible is to limit production. Curtailment of the production, especially RES, is a tough subject. As the idea of flexibility in the power system is to allow RES to operate when there is wind or sunny, there is a conflict in curtailing the production from those RES. Sometimes, it is considered appropriate to use strategic curtailment. These situations occur when there is overgeneration or oversupply of RES and the demand-side cannot match it. The electricity markets usually already do the curtailment, as the price is low in day-ahead (DA) markets and the RES may opt to curtail the production. Curtailment can also be used when there is a high use of baseload in the transmission system and the transmission of the electricity need to be controlled. [16]

Network-side flexibility is flexibility in the transmission and distribution systems network. Network-side flexibility can be seen also as an enabler of the flexibility solutions, as presented in [11]. Smart grids, microgrids, dynamic network reconfigurations, interconnections, and different kinds of network expansion planning are all concepts that include multiple solutions to enable to increase the flexibility with the network. Network-side flexibility can be implemented in the different flexibility methods. Network-side flexibility is one of the most important factors in increasing the whole flexibility of the power system. [16]

The last one of the crude divisions of the flexibility resources is energy storages. The mechanics behind these are quite easy. The energy, which is produced during high production or low demand, or during a time when the price is low, is stored in energy storage. The energy is released when production is low otherwise, or demand is high, or when the price is higher. [5] Energy storages can also be classified by the discharge and charge time. Some energy storages can discharge or charge in seconds and some storages in hours or days. Also, the size of the storage can be one way to classify them, as there are available energy storages for houses or energy storages such as water reservoirs or storages for wind power with the wind farms.

As Figure 1 shows one part of the flexibility solutions are enablers. Enablers are different kinds of mechanisms that are enabling the other flexible resources to work. Independently these enablers don’t have flexibility, but when used together with flexible resources, they have a great impact.

Another way to classify flexibility solutions is by time. In [25] the flexibility solutions and impacts are classified chronologically into four categories: long-term, mid-term, short-term, and super short-term. The long-term category presents the solutions that are made in the system planning, which includes grid planning, production, industrial loads, and the electricity markets. The long-term category is usually the enablers presented in Figure 1 but also hydropower in the Nordic countries will need months and years of planning to use as flexibility. Mid-term includes day-ahead and intraday (ID) market actions, which include DSM, peaking load generation, and energy storages. The short-term category is the balancing actions done at the actual hour, so as previously stated, could be load-following generation, DR, or energy storages that can be activated fast as a reserve. The Super short-term category holds the actions done in seconds or even faster. Different kinds of voltage and reactive power control systems are in this category. The market products for this category are Fast Frequency Reserve (FFR) and Frequency Containment Reserve for Disturbances (FCR-D) and Frequency Containment Reserve for Normal Operation (FCR-N) [25]

2.2.2 Characteristics of flexibility resource

In the previous chapter, the classification of the flexibility solutions was presented. In this chapter, one more way for the flexibility resources classification is presented. This classification focuses on the different characteristics of the flexibility resource.

Figure 3 presents the most important characteristics of flexibility resources as a figure of one flexibility event. Flexibility resource providers can either be consumers or producers, or both. Figure 3 illustrates a flexibility event, which can be either production or consumption. Either production or consumption can adjust their load up or down, depending on which is needed.

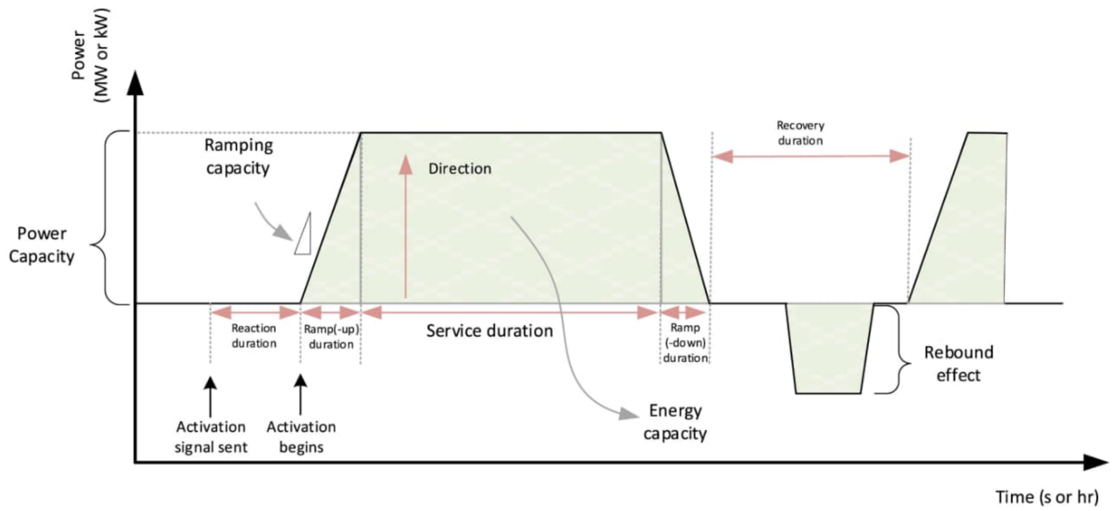

Figure 3 *Important characteristics of flexibility resource.* ***[11]***

It depends on the resource what the activation signal is, it can be a direct or indirect signal, or some resources can receive both signals. As previously stated, demand-side flexibility can be implicit or explicit, and the activation signal in the case of Figure 3 could be either of them. For another kind of flexibility, the same signals are used. Resources can also receive both kinds of signals. Ramping capacity is determined by what is the maximum change of the power output in a time unit. This depends highly on the resource and can determine in which category in the chronological classification the flexibility resource belongs. Service duration, power capacity, and energy capacity should be known from each resource. During a flexibility event the service duration, power capacity, or energy capacity needed might be shorter or smaller than the maximum of the flexibility resource. This should be noted when examining different flexibility events from the same flexibility resource. The last things to note from the events are the recovery time and possible rebound effect that is going to happen after the actual event.

All the characteristics described above are important when the flexibility events and needs are planned and used. When system operators have a clear picture of the characteristics of each flexibility resource, the usage of the flexibility as a part of the power system becomes more accessible.

2.3 Challenges of flexibility

Although flexibility is not a new concept and has always been used in the electricity markets, there are a few challenges that flexibility has regarding its use as a part of normal operation. Especially demand-side flexibility has barriers regarding the participation in electricity markets, as it is usually seen as a newer and complicated concept.

Overall barriers for flexibility are listed in [16], and these barriers should be looked at as challenges. And when these challenges are solved, the flexibility should be easier to utilize to be part of the normal use. These barriers can be categorized as follows: markets and operational environment, different parties' interests, and security. [16] As an addition to this mindset and trust in flexibility can be a barrier for market participants, system operators, and flexibility providers.

Markets and operational environment cover issues with the lack of suitable markets, the regulatory and tariffs, and the business environment. Most of the market barriers concern demand-side flexibility. Demand-side flexibility can participate in the electricity markets, but usually as the flexibility resources are so small, that they need aggregators for participation. For example, an independent aggregator is not still authorized to operate in all marketplaces. More about independent aggregators and challenges with them are described in Chapter 2.5. Market design is one of the flexibility enablers listed in Figure 1. When flexibility can participate as an equal market participant to the electricity markets, the market design will support flexibility. [16] The business environment might be harder to overcome. New flexibility solutions might not have the same value or place on the markets and investments, as older technologies which are proven to be cost-effective [16, 23]. The regulatory and tariffs come along with new market solutions and what comes to overcoming them, is that they should be as transparent as possible, for all the participants [16].

Different parties' interest means that both consumers and investors have an interest in flexibility. Challenges in this are if the solutions and new flexibility opportunities are interesting enough. Some providers might have some conflicts of interest regarding integrating new flexibility solutions to the grid. New flexibility solutions are also expensive, and usually require huge investments towards the technology or mechanism. This could be solved with tariffs or incentives that would promote investing in flexibility and using it as a flexibility provider. Incentives towards customers' actions are also one of the barriers. Demand-side flexibility relies a lot on the behavior of customers, so there is a need to have incentives encouraging the right way of behavior. [16]

The last barrier is the security questions about flexibility. This one is also connected to the customers and DSM. DSM requires a large amount of communication between the system operator and the demand side. This communication can be problematic when taking into consideration the new protocols towards individuals' rights to privacy. With the right permissions, this can be overcome but after that, the challenge is in cyber security. If there is a lot of control over the customer's electricity usage, the risk of the information leak would be high. [16]

Market participants such as system operators can have a hard time trusting flexibility as a congestion management solver. Investing in new grid and reinforcing the old one are physical actions that can be trusted to happen. But when investing in flexibility and making an agreement with the flexibility provider, the SO must rely on the other participant.

Flexibility in the electricity system has a lot of barriers, problems, and challenges ahead of it. The important barrier that should be solved within everything else, is the motivation to change the structure of the system to meet the criteria flexibility solutions create. Also, the flexibility resources should be given interesting enough signals economically to be efficient. [15]

The barriers considering markets and operational environment are especially examined and discussed in this thesis. The problems with independent aggregators and lack of verifying and settlement are part of this problem. The market structure and products do not support independent aggregators to participate in the electricity markets, and the imbalance settlement is not enough to settle the imbalances flexibility events create.

2.4 Imbalance settlement

The idea of the imbalance settlement is to create financial balance in the electricity markets after the operational hour. Imbalance settlement in Finland is done by eSett which is the Imbalance Settlement Responsible (ISR). The imbalance settlement is done on behalf of four TSO's, which are Fingrid (Finland), Energinet (Denmark), Statnett (Norway), and Svenska kraftnät (Sweden). [26] This imbalance settlement is called the Nordic imbalance settlement model. The Nordic imbalance settlement has operated with a single balance model since 1.11.2021 [27]. Before that, the imbalance settlement was operating with separate balances for consumption and production [26].

The roles of the imbalance settlement are Imbalance Settlement Responsible, which is eSett, TSOs, Distribution System Operators (DSOs), Balance Responsible Parties (BRPs), Retailers (REs), Nominated Electricity Market Operators (NEMOs), Service Provider (SPs), and Metered Data Aggregators (MDAs). Figure 4 presents the relations between different market participants and eSett.

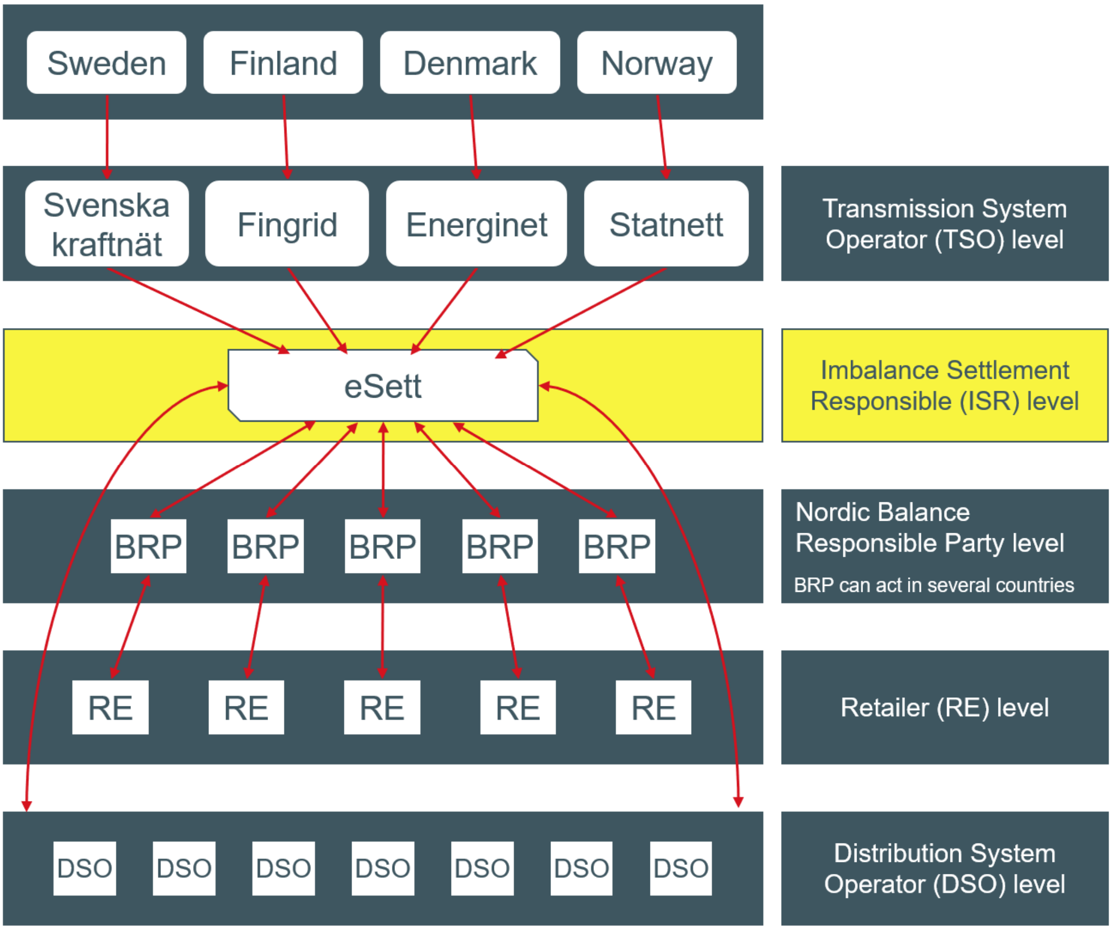

Figure 4** eSett's relations in the Nordic imbalance settlement. Adapted from **[26]

As stated, eSett is responsible for the final imbalance settlement and they also gather all the data and information needed for that settlement. Involved countries' TSOs own eSett and that's why the arrows in Figure 4 go from the countries to the TSOs. The most important task for TSO is to keep the electricity system in physical balance during the delivery day. TSOs are responsible for submitting production plans and activated imbalance adjustments per BRP to eSett. Also, TSOs need to report the information about MBAs and Metering Grid Areas (MGAs) and the relations between them. [27]

DSOs have several duties towards eSett. DSOs in the Nordic Imbalance Settlement include also closed DSOs. The most important responsibility, besides connecting producers and consumers to their grid, is metering the production, consumption, and exchange between other grids, and reporting those measurements to parties involved, including eSett. DSOs in Finland are entitled to register their MGAs to Fingrid. DSOs need to calculate and report load profile shares, calculate the final profiled consumption and correct the imbalance between the DSO and RE after the imbalance settlement reporting is closed. [27]

Retailers are responsible to register with eSett, but the information goes through BRPs. Every RE must have an agreement with a BRP in every MGA they are operating. In Finland, another option is for the RE to have an agreement with another RE which has an agreement with BRP. This arrangement is available to enable the chain of open suppliers. [27]

BRPs need to have an agreement with the TSO they are operating and with eSett. They are responsible for planning the balance schedules, submitting plans of every RO to the TSO, and information about bilateral trades. Bilateral trades need to be verified on the RE level. BRPs are accountable to keep the information about imbalance structure up to date and verifying the data which is reported by eSett. They also need to keep eSett up to date of which Res they are responsible for consumption and production per every MGA. [27]

Earlies mentioned MDA is in Finland’s case soon to start operating Datahub. Datahub will aggregate the information between customers, retailers, and DSOs. Datahub will start operating on 21.2.2022. [28] In Finland, NEMO is for example Nord Pool, which is a power market that operates DA and ID markets [29].

The benefits of having the Nordic Imbalance settlement model are that all four countries can rely on the same rules and calculations and the exchange of energy between countries can be identified easier. The single balance settlement will encourage market participants to stay in the balance and using flexibility to correct the balance should be easier than before. [27]

BRPs are responsible for the imbalance in their area, but when for example independent aggregator acts with BRP’s customers the imbalance is disrupted without the knowledge of BRP. This makes the BRP’s job to keep the imbalance harder, and this is the reason why new ways to verify the happened actions.

2.5 Independent aggregator

Independent aggregator (IA) is according to EU directive 2019/944 “Common rules for the internal market for electricity and amending”: “*a market participant engaged in aggregation who is not affiliated to the customer's supplier*” [30] Independent aggregator is an important part of bringing flexibility available for the electricity markets along with normal aggregating. As IA collects the potential of several flexibility providers, those collections are called Virtual Power Plants (VPP), but in this thesis, the concept of IA is used, as it includes the aspect of electricity markets as well [31].

Based on the EU Directive, IA must be responsible financially for the actions that may cause imbalance into the electricity system, and based on that they must either be balance responsible parties or delegate that to another balancing responsible party [30]. Acting as a BRP or allocating that responsibility to another BRP makes IA financially responsible for its acts to TSOs and eSett.

NordREG’s report “Nordic Regulatory Framework for Independent Aggregation” introduces two different positions for IA: Balance Responsible Parties (BRPs) or Balancing Service Providers (BSPs). BRP has a valid agreement with eSett and is responsible for the imbalances caused between the final position and allocated volumes. Also, BRP must consider the imbalance adjustments [26, 32]. BSP is only responsible for delivering the products it sells to the SO and those actions being deducted from the BRPs imbalances. IAs can act as BRP or BSP but also suppliers can act in these roles. When IA acts as those parties they can be referred to as BRP-IA and BSP-IA and suppliers are BRP-supplier and BSP-supplier. [32] These separate positions are not in practice in the Nordic Imbalance Settlement yet, but the implementation is estimated to 2022 [33].

Every customer has an existing BRP-supplier or BSP-supplier, but BRP-IAs and BSP-IAs should be able to operate in the suppliers’ areas with their customers without consent or separate contract with the existing suppliers. This arrangement will ensure that the flexible resources in suppliers’ areas can be used in the balancing markets more efficiently and more flexible resources overall are brought to the markets. [32]

Figure 5 presents the data flow between customers, suppliers, independent aggregator, and eSett. In this Figure, the IA is acting as BSP-IA, which is still not available position in Finland, but the Figure is fine presentations of the acts that can happen.

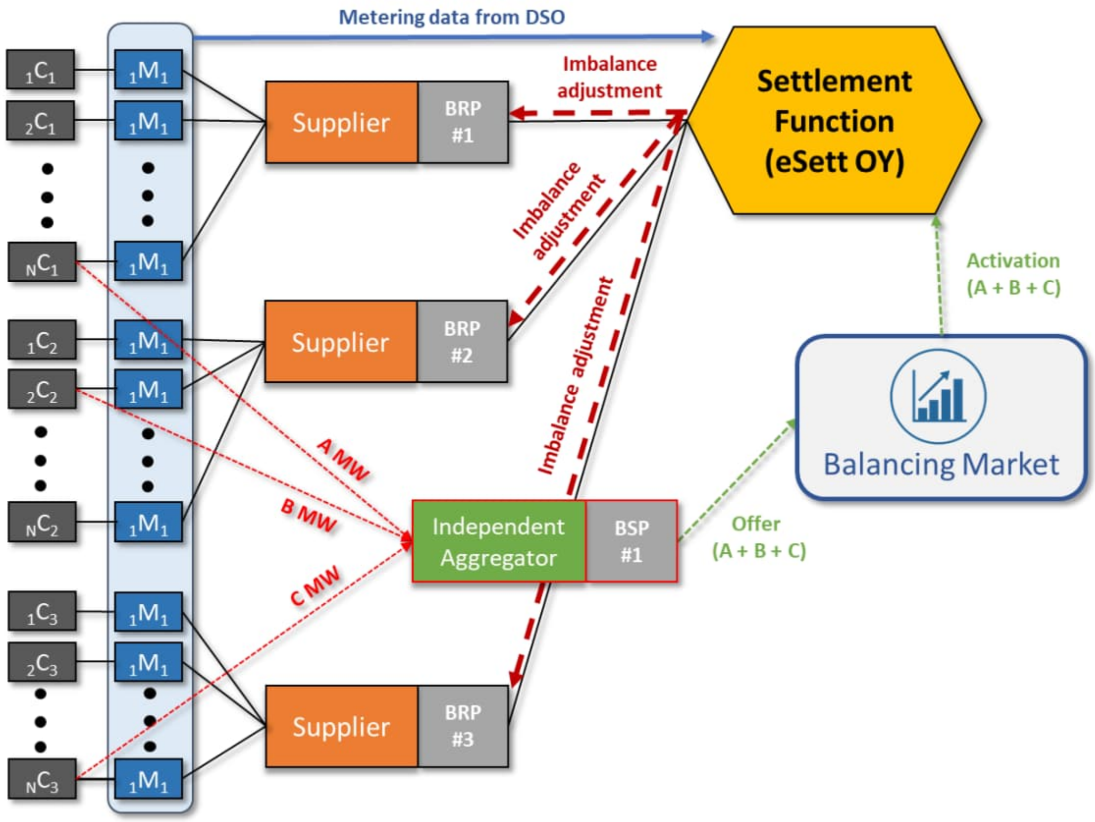

Figure 5** Data flows for independent aggregator and suppliers. **[32]

Figure 5 shows customers marked with dark grey and with blue the metering points that DSOs are responsible for measuring. Each customer has an original supplier and BRP, who are reporting and adjusting the imbalance to eSett. IA is using customers from each BRP to offer to the balancing markets. Here IA is financially responsible to deliver the offered flexibility, but not responsible to report the events to corresponding BRPs.

Problems with IA comes when IA aggregates suppliers' customers, but no information is exchanged between suppliers and IA, as they do not have to. Somehow the actions that BSP-IA makes should be deducted from the responding supplier's imbalance. This is because the acts of IA are impacting the BRPs imbalance. This scenario is presented with an example which is adapted from the Nordic Regulatory Framework for Independent Aggregator -report. [32]

If we have a hypothetical situation, that a customer is expected to consume 5 units of energy, and a BRP-supplier expects to sell that 5 units of energy to them. But an independent aggregator has also a contract with the customer, for 1 unit of flexibility to reduce their consumption. The IA sells the 1 unit of reduction so that the customer consumes only 4 units of energy. This way the IA has received compensation for reducing the consumption and performing a flexibility event, but the BRP-supplier has only sold 4 units of energy and must pay for the excess energy.

The settlement should be done in the case described above. In a report written by Pöyry "Independent aggregator models" [34] is described four different models to solve the problem of IAs causing imbalances to BRPs balances. Model one is that the imbalance is corrected by a neutral party, but no further compensation for the imbalance is paid to BRP. Model two is that no other compensation is done and the BRP is only compensated through imbalance settlement. Model three corrects the balance with the same logic as model one, and the IA pays compensation to the BRP based on pre-determined prices. Model four would be that the customer's non-flexible load would be supplied by the BRP and flexible load by IA. This would make the IA an integrated aggregator. [34]

If IA is aggregating multiple customers from multiple BRPs, the realized adjustment should be verified for each BRP. This creates the need for verifying and baseline, which could be used to settle the imbalances in BRPs imbalance.

3. FLEXIBILITY VERIFYING

The need for flexibility settlement comes from the issues with the imbalance settlement and introducing independent aggregators to the electricity markets. Those issues are presented previously, and in this chapter, the need for accurate verifying, methods for verifying, and the methods that are used in this thesis for tests are examined closely.

In literature verifying methods are referred to as baseline and verifying methods. Baseline methods mean that there is a generated baseline for the flexibility source to represent the actual consumption or production without the flexibility event. The terminology related to the baseline and verifying methods is not unambiguous. From the verifying methods presented in this thesis only *Baseline Type I* and *Type II* are baseline methods, and others are verifying methods.

The categorization of the verification methods is presented in a white paper by EnerNOC [10]:

Table 2 Categorization of verification methods.

Verification methods
Baseline Type I
Baseline Type II
Maximum Base Load (MBL)
Meter Before – Meter After (MBMA)
Metering Generator Output (MGO)

From these *Baseline Type I* and *II* can be also referred to as historical data approaches [35]. Baseline Type I includes the methods *X of Y*, *Regression*, *Comparable day,* and *Rolling average*. And further, these methods have different elements and adjustable parameters that create new verifying methods. In this thesis, the *X of Y* method and *Comparable day* method with some moderations are used, and those methods are also examined more closely than the other methods.

Diverse conditions and different parties usually require varying kind of methods for the verification and each method have their good and bad sides. More detailed studies of baseline methods can be found from articles [36] [37] [38].

3.1 Need of flexibility verifying

From the demand-side flexibility, especially demand response events need verifying. The flexibility that happens because the customer wants to use the electricity during lower price hours, i.e., EE and EC, is not sold or brought event by anyone, and the benefits from this kind of demand response are paying the customer immediately in form of saving money. Other kinds of flexibility resource providers need verifying too, but those are not the focus of this thesis. Supply-side and energy storages do need verifying but verifying is usually easier for them than for the demand-side. The focus of this thesis is demand-side flexibility and verifying methods and for that reason, only the demand-side verifying needs are discussed in this chapter.

The verifying is needed, when there is a customer or third party, such as an aggregator, who responds to flexibility requests by performing a flexibility event in which duration and volume are pre-determined. Different parties involved in the flexibility event have various reasons to confirm, that the pre-determined flexibility event took place the correct way. Some parties such as BRP or TSO want to make sure, that the balance is kept and ensured regarding the event. Parties such as aggregator or independent flexibility providers want to make sure they get the compensation they deserve for the flexibility provided.

As different parties have specific reasons to verify the event, suitable and precise verification is needed. One way to verify flexibility events is with baseline methods. Baseline methods create a baseline for the flexibility provider, to which the actual load of the event is compared. Baseline methods do not work for every flexible resource, and for those different verification methods are used. Verification methods are described more in Chapter 3. In this Chapter, baselines are used to illustrate situations where they are needed and the basic idea behind them. Baseline methods are usually used to verify energy, not power.

Figure 6 presents a simple imaginary flexibility event. With green dots is marked the start and end of the flexibility event. During the event, the load is reduced to a specific level. With red line is presented the baseline. Baseline presents situation if the flexibility event did not have happened. With baseline and the actual load, the flexibility energy can be calculated. This energy is presented with the green lines in Figure 6.

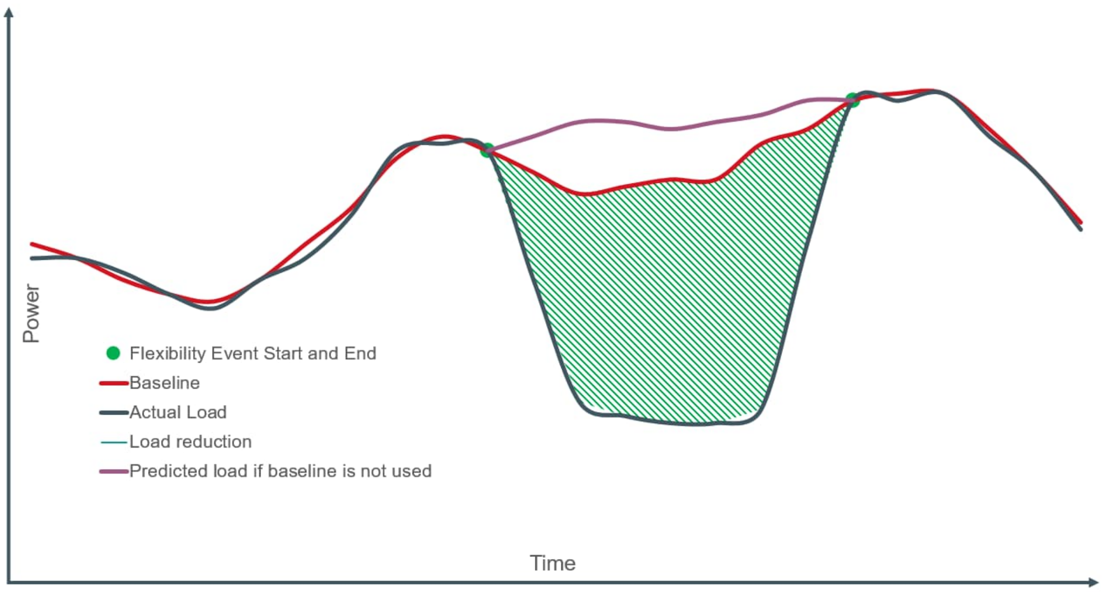

Figure 6 *Example of flexibility event and baseline.*

The example of flexibility event and baseline in Figure 6 shows why there is a need for an accurate baseline during flexibility events. With the purple line is shown how sophisticated guess of the actual load would probably go in this situation. As Figure 6 shows, between the purple and red line is a lot of energy, and in this case, the flexibility provider would benefit from it. It would seem that the flexibility provider has reduced consumption as much as the green-lined and the white areas between purple and red line are together, even when the actual consumption would have been without the event much lower, on the baseline's level.

Figure 7 presents a bit different situation when verifying is needed, in this case, there is presented two ways to act during flexibility event. The dark line presents the actual load, which would have occurred if the flexibility event did not have taken place. The flexibility event starts and ends where the green line starts and ends. The green line also presents the right way to act during flexibility events and the red line presents the wrong way.

If a retailer who bought the event, buys 1 MW of power from 12:00 o'clock to 13:00, means that the retailer is looking to have 1 MWh of energy. Figure 7 shows an exaggerated situation of this event, and this kind of behavior would probably not be seen. In reality, the flexibility event would now follow the baseline, but rather fill the decrease of baseload and cut the increased peak from the end of the event. But ideally, if the committed power is 1 MW and the baseline is known, the flexibility provider would follow the baseline and deliver that extra 1 MW for each moment of the flexibility event.

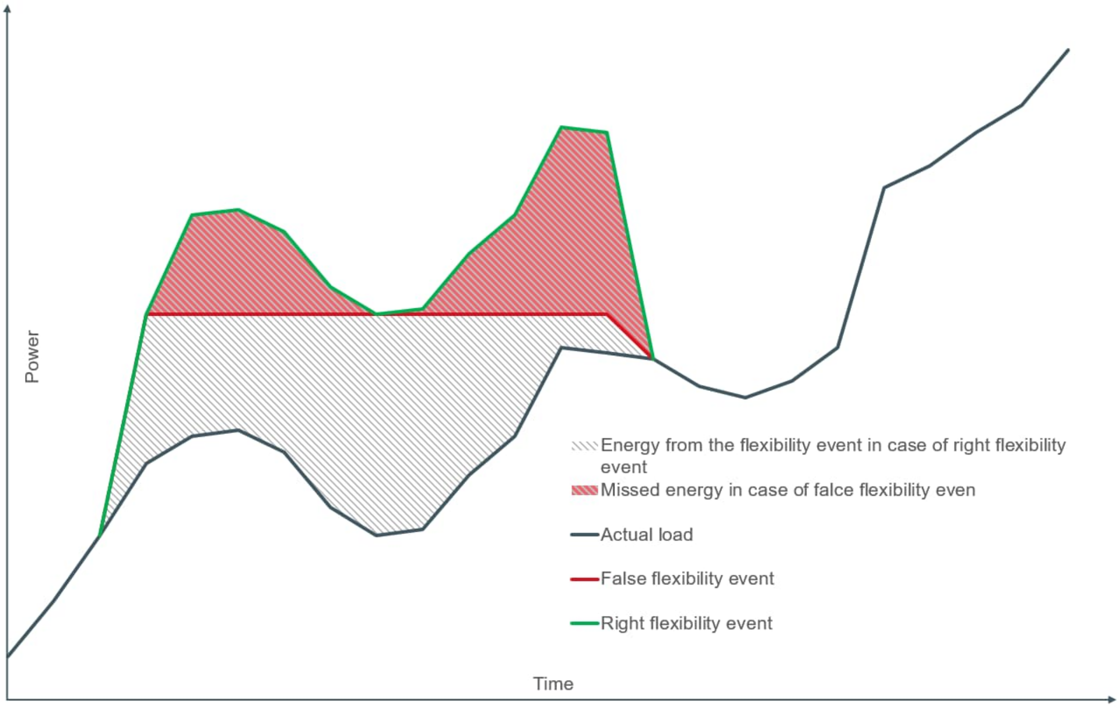

Figure 7 *Example of flexibility event with wrong and right actions.*

Figure 7 also shows the wrong way to act during flexibility events. The red line presents a situation, where the provider does start the event with a 1 MW increase but continues with the same level of power through the event. This way the committed energy is not produced. With the red lines is presented the missed energy that is not consumed or produced if the provider does not consider the baseline during the event.

As seen from Figure 7 with the red lines, there would be a lot of missed energy in case the flexibility provider acts wrong. On top of that, the needed power for each moment is not met. In a real event, the baseline is formed in 15-minute intervals, so changes for 15 minutes might not be seen. In case of faster flexibility events, verifying should be done more precisely than 15 minutes.

Because of Figure 7's situations, the flexibility providers should always know their normal consumption, so that they could perform in a way they are expected during flexibility events. Usually, the baseline is known by the provider beforehand, more about this is described in Chapter 3.3. This example also shows why the verification should be done to the whole event, and not just the beginning, as the beginning power could be right but the rest not. Of course, in the case of energy requests, the start of the event does not tell anything about the performance of the flexibility provider. If a case like Figure 7 happened, and a suitable verification method was used, the wrongful actions would be caught. This would mean, that as the price of the flexibility event was pre-determined, the correct price for the event performed could be calculated. Also, those who are responsible for the electricity system balance can know where the misbalance is from, and which parties should be charged for interrupting the imbalance.

The verifying is more important to smaller units and aggregated flexibility providers, just like in the beginning was told, demand-side operators. Power plants that are in normal use already for reserves do not need verifying in the same way that aggregated EV charging stations, where the load is always changing. The most important reason for accurate verifying comes from loads, that have constantly changing or weather-dependent loads.

3.2 X of Y baselines

The most used and simplest baseline is *X of Y* baselines. *X of Y* baselines are Baseline Type I. The basic idea is to take the average from the previous days and make that average the baseline for event day. The X stands for the days that are considered in the baseline calculations, from the Y days that are before the event. Examples of this method could be 4 of 5 days, 8 of 10 days, or 10 of 10 days. [10]

The Y days number can be also called the look-back window. This is the timeframe for how long back in the historical data the calculation needs to be done. This number can vary from 2 to 60 but usually is something between those. The amount of Y days depends on the verified target, and what kind of load is in question. In the white paper where the look-back window is presented, 60 days was the recommended number of days, and less than that would be too restrictive. [10] However, the most used Y days are around 10 to 20. Many factors can affect this number, and too high number of Y days can include historical data that is irrelevant for the event day.

Also, not all days before the flexibility event are admissible. Y days exclude weekends, holidays, days that include flexibility events, or other days that can be determined as unusual to the electricity consumption or production. On another hand, if the flexibility events happen on weekends the weekdays are excluded and only data from weekends are used. The same logic is used for holidays, usually, weekends or Sundays' data is used for those days.

X days are determined based on which method is used. If the baseline is, for example, *High 8 of 10*, this means the 8 days that have the highest energy consumption/production of the 10 admissible days are considered when calculating the baseline. This relation between the X and Y depends on the case the baseline is calculated for. If the flexibility event is demand response on a hot day, the event day consumption is probably higher than usual, so the *X of Y* method such as *High 5 of 10* should be used. Or if the day is otherwise normal, methods such as *Mid 8 of 10* should be used, where the highest and lowest energy consumption days are excluded. [10] [39] It can be hard to determine which of these methods should be used for each day, and if the load is weather dependent, the most accurate results will not be achieved with *X of Y* baselines.

X of Y baselines are simple and widely used, but usually, they need adjustment before they can be used as accurate baselines. Adjustment is usually done with the data from the event day. Some adjustments are directed to the electricity usage of that day and use the prior hours to adjust the event hours baseline and some use regression analysis from the specific customer.

Adjustment factors

X of Y methods are simple, but without any adjustments, quite inaccurate. With different adjustments, *X of Y* methods can achieve high accuracy still being simple to calculate. The need for adjustments comes from the fact, that if the demand response is triggered by emergencies caused by power outages or by peak demand, the load of the day might be quite different from the days before. With the adjustment factors, the baseline can be adjusted to the event day's level referring to the hours before the flexibility event.

Usually, the adjustment is done by the hours 2 to 4 before the event. If more than 4 hours before the event are considered, the data might not be relevant and inaccurate considering the flexibility event. The hour before the event should also be left out of the calculation. If the customer is notified about the event hour before the event, they could use the one-hour window to manipulate the baseline for the event hour. Also, the ramping time should be considered in this matter and should be left out of the calculations to get the most representative adjustment factor for the day. [10]

Figure 8 presents the different timeframes how the adjustment factor can be calculated. Four different models are described in this thesis, from which model D is the one used in the tests later.

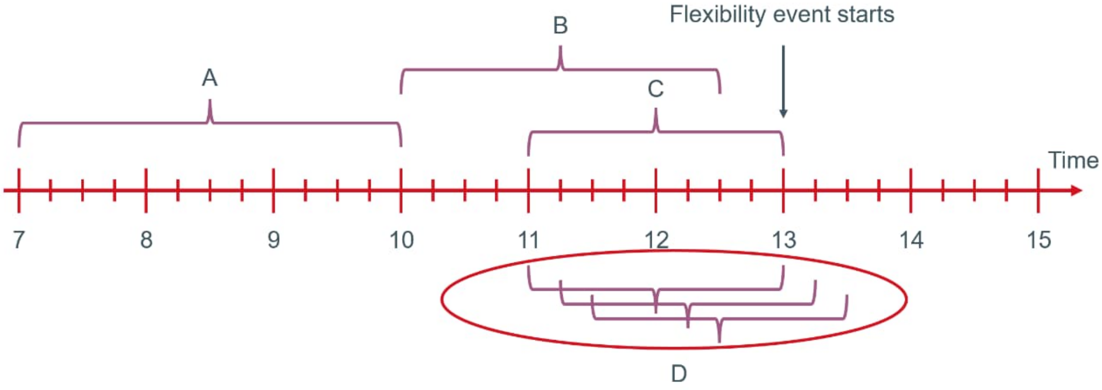

Figure 8 *Timeframes for adjustment factor calculations.*

Model A is a way to calculate the adjustment factor for the whole day based on the morning hours. Usually, the flexibility events occur after 11 am, so the adjustment factor can be calculated earlier than that. This model is quite simple because the adjustment is calculated every day from the same period. This model is also the most inaccurate because the load profile can change a lot during the day and the adjustment is only from the morning hours. Model B is probably the most used and the most accurate model to calculate the adjustment factor. The 30 minutes before the event are left out of the calculation, because during that time the resource might already start preparing for the event and the load might change because of that. Otherwise, the adjustment factor is calculated as normal, 3 hours before the event, leaving 30 minutes before the event out.

Model C is another version of the normal calculation, but the timeframe is directly before the event. This might affect the adjustment factor if the resource starts ramping up or down before the event. Model D is almost like model C, but as with models A, B, and C, the adjustment factor is calculated once for the whole day and the whole event. In model D the adjustment factor is calculated separately for each 15-minute measuring period. This model cannot be used in real life. The adjustment factor is calculated based on the actual load and baseline, and in this model, because there is the actual load for the whole day available without the flexibility event, it is possible to calculate the adjustment factor multiple times.

The most accurate model probably is model D, but as it cannot be used in real life, the second-best model is model B. If the ramping time of the resource is considered, that amount of time can be used for the excluded time before the event, and the adjustment factor can be calculated as close to the start of the event as possible.

The simple way to describe the adjustment factors calculating process is to say if it is scalar or additive. With the scalar adjustment, if the consumption or production is *x* percent above the unadjusted baseline, the baseline for the event would be *100+x* percent. The additive approach is if the production or consumption is *y* kW above the baseline, the event baseline would be the original, but *y* kW added to it. [10]

The adjustment can also be capped or uncapped. Cap could be for example 20 % and would be useful if the customer changed their consumption or production by a large amount just before the flexibility event, for example from 100 kW to 150 kW. If the adjustment was uncapped, the adjustment factor would increase the baseline for the flexibility event and would not probably present the actual usage during the event in a normal situation. With the cap, the adjustment factor would be restricted lower, to be only 10 kW, and the higher consumption or production before the event would not have too much impact on the baseline. [10] Uncapped adjustment is used in this thesis, as the capping could harm the customer.

The third consideration with adjustment factors is that are they used symmetric or asymmetric. Symmetric use is when the adjustment factor can adjust the baseline up and down, and asymmetric is when the baseline is adjusted only up.

In this thesis scalar, uncapped symmetric adjustment factor is presented and used. The adjustment factor is calculated using the averaged baseline and actual load before the event. The adjustment factor can be calculated before the flexibility event. If the event day is d, the time that the event starts is t, the timeframe where the adjustment factor is calculated is from s to e, averaged baseline is $bl(d,t)$ and the actual consumption or production is $al(d,t)$ the adjustment factor $c(d,t)$ can be calculated as follows:

$$c(d,t) = [\sum_{i=s}^{e} al(d,i)]\,/[\sum_{i=s}^{e} bl(d,i)]. \qquad (1)$$

In Equation 1 if the flexibility event starts at noon, the adjustment factor calculation window could be from 9:00-10:59. In this thesis, it would mean using timestamped data from 9:00 to 10:45. With the adjustment factor, the adjusted baseline for the event hour is calculated

$$bl'(d) = c(d) * bl(d) \qquad (2)$$

where all the hours of the event day are multiplied with the adjustment factor to get a new adjusted baseline for the day and the flexibility event. [39]

The adjustment factor can be calculated with hourly or 15-minute data. In this thesis, the 15-minute data is used, and for that reason, the *s* to *e* timeslot should probably be closer to the start of the event, as described previously in this Chapter. In model B’s case, the *s* could be the same, 9:00, but the *e* should be 10:15 or 10:30.

3.3 Comparable day baselines

Comparable day baseline is Baseline Type I and uses historical data to create the baseline. Unlike the *X of Y* baseline which uses multiple days to create an averaging baseline, the *Comparable day* method finds the most similar day from the previous days to use as the baseline. *A Comparable day* does not create new information to form a baseline, it just uses the most similar day’s consumption and production profile as the baseline to compare the flexibility event. [10, 35]

The Comparable day method can consider power and energy, weather such as temperature, and other major factors when comparing the days. It depends on the flexibility source and object which factors are relevant to the consumed or produced energy. For example, if the flexibility event happens at the beginning of the summer, as the first heath waves are coming, the data before that day are not relevant and can be used as the compared day. This method relies on the consideration of when the method is reliable and when more data is needed. [35]

This method also has a few downsides, such as the need for relevant data, but also that this baseline cannot be calculated in advance. This creates problems from the customer’s point of view. Customers should always know their normal load, but in case the reference level is unknown to them, the flexibility might be less than asked. Various forecasting programs do help this situation so that the customer is not unaware of the baseline that is going to be for the event day.

The Comparable day method uses usually more historical data than the *X of Y* method, and one problem with this method is that there might not be relevant historical data for the baseline. This issue is especially with resources where the load profile can be highly variable and does not have a regular pattern to it. For this kind of situation, it would be better to use some other verifying method.

The Comparable day method can also be used with the *X of Y* method, to find from the Y days the most comparable days to use to create an averaging baseline for the event day. This method is used in this thesis.

3.4 Other verifying methods

The methods presented in this chapter are not tested in the case study but are presented so that the reasons for excluding them can be stated. All these methods are important and useful in varying conditions and with different customers. Also, there are verifying methods that are not listed in the first list, which are briefly introduced in this chapter.

3.4.1 Regression

Regression baselines are one modification from Baseline Type I. *Regression* baselines use the load profile, weather, and day-type data to calculate the baseline. A lot of data is needed, and the calculations can be quite complicated compared to regular *X of Y* baselines. There have been studies that regression baselines perform better than *X of Y* baselines, but also studies where for example *High X of Y* baselines performed better than regression baselines. [10]

The Regression baseline is calculated for individual customers and requires a lot of knowledge from the flexibility site. Also, the relations between the load profile, weather, and day-type data must be taught before any baselines can be reformed. [35] *Regression* baselines are not studied more in this thesis, as there are not enough data to use this method or time to teach algorithms to form the baseline. The overall opinion on regression baselines is that they are accurate, and in suitable situations very useful, but they sacrifice too much for the accuracy that the simplicity is compromised. In Finland, load curves are used in normal use already, and they use the same methods as the *Regression* method, and for that reason, the *Regression* method could be used for verification.

3.4.2 Rolling average

In addition to the previously presented Baseline Type I methods there is also a *Rolling average* baseline method. This method also uses a large amount of historical data, and for that reason is not used in this thesis. The baseline is average from the historical data but is weighted so that the days closer to the flexibility event are weighted more than days longer time ago. [10]

This method can be more accurate than *X of Y* methods, as it has more data to form the baseline. The problem with this method comes with flexible sources that have a highly seasonal load profile, such as skiing centers or amusement parks. If the flexibility event happens the first day the center or park is open, the more valued days do not represent the event day and are still weighted more than days the previous year.

The rolling average method is quite similar to the *X of Y* method, but the biggest difference is on the weighted days that consider the seasonal changes of load profile, without adding too many changing parts to the method calculations.

3.4.3 Maximum Base Load

Maximum base load (MBL) is based on the maximum energy usage of the customer. The baseline is determined by the maximum energy usage, and the committed capacity for the flexibility event is deducted from that. The customer knows the baseline in advance and can perform correctly during the event. They must drop their usage to that level. [10]

MBL can also be coincident or non-coincident. Both use the peak hours of the last year, but coincident uses the peak hours by the system lead peaks and non-coincident uses the peak hours based on the individual load profile. [35]

MBL is easy and simple to calculate and is easily accessible by the customer. *MBL* is calculated once a year and will remain the same for the whole year. This creates sometimes situations, where the customer might perform a flexibility event without even doing anything if the usage drops below the baseline determined for that year. *MBL* is also quite inaccurate if compared with *X of Y* methods, but coincident *MBL* can perform better than non-coincident [35].

3.4.4 Meter Before – Meter After

Meter Before – Meter After (MBMA) method is commonly used in faster flexibility events, such as ancillary services, where the event can be only minutes. Other methods presented here are usually used in events that last over an hour.

MBMA method reads the load before and after the flexibility event and sets the baseline between those readings. MBMA method can be easily affected if ramping up or down is done right before the event. This affects the reliability of this method and its accuracy. [35]

3.4.5 Metering Generator Output

The Metering generator output (MGE) method is metering the generation output behind the meter. Most of the verifying methods are used with consumption and demand response, but this method is used with generation to make it easier for the generation to participate in flexibility events [38].

There are three ways to use *MGO*. The first way is to use the difference between the net load and generation load and compare it to normal usage. The second option is to set the baseline to zero and compare the generation metering to it. This is also called zero baseline, and with this method, all the actions with the generator are considered flexibility events. The third way is to combine the first and second methods and create the baseline that way.

3.4.6 Baseline Type II

Baseline Type II is used in sites where sub-metering data is not available. *Baseline Type II* is formed with data from another site, that has the same kind of load behavior. The data could be from an aggregator that has several flexibility providers, but metering data only from a few of them. Also, metering data from aggregated sites can be the base for *Baseline Type II*. [10]

Baseline Type II is usually used with residential customers, as the load must be similar with all the sites that *Baseline Type II* is used. With industrial or commercial customers, the load profile can differ a lot, and using *Baseline Type II* would not be efficient. [10]

Baseline Type II can be quite inaccurate, as it is hard to prove that two customers have the same load profile. It is still used, as it can be rather costly to have the right equipment for metering all the customers individually. This is also one of the reasons why the *Baseline Type II* method is not used with bigger customers, as usually those sites already have reliable metering. [10]

4. CASE STUDY

This chapter introduces the case used in this thesis and tests that are done. First, the data source, and the focus of these tests are introduced. Chapter 4.2 presents the modifications done to the data and more detailed information about the data profile used in this thesis. Lastly, the methods that are tested are examined more precisely and the parameter to evaluate the performance of the methods are gone over.

4.1 Data source

The used data in this thesis is from Tampere university's Kampusareena. It is one of the Hervanta campus' buildings. Kampusareena is an office building, with extensive measurements for electricity consumption. In Kampusareena some companies rent office space and there is restaurants, and a lot of studying places, such as the library and computer rooms.

Kampusareena was chosen for this thesis because it offers metering data that other office buildings cannot provide. This data has been already used in research, was easily accessible to use, and there is no need to make the data anonymous. Another reason to use this data is that there is measured data from different levels of the electricity system. In some cases, the whole building could be one flexible resource, but ventilation or EV charging are flexible resources that can be used individually. Kampusareena provides many ways to test verifying methods with the same data, so the results are comparable.

The electricity connection to Kampusareena is from a 20 kV medium voltage distribution network, and it is transformed to 400 V. The electricity load is divided into two: tenants' electricity and service electricity. Tenants' electricity holds the lightning, appliances, and wall sockets. Service electricity is divided into eight different loads: Ventilation, water cooling 1 and 2, elevator 1 and 2, EV charging station, Solar PV power plant, and others. Figure 9 presents this topology.

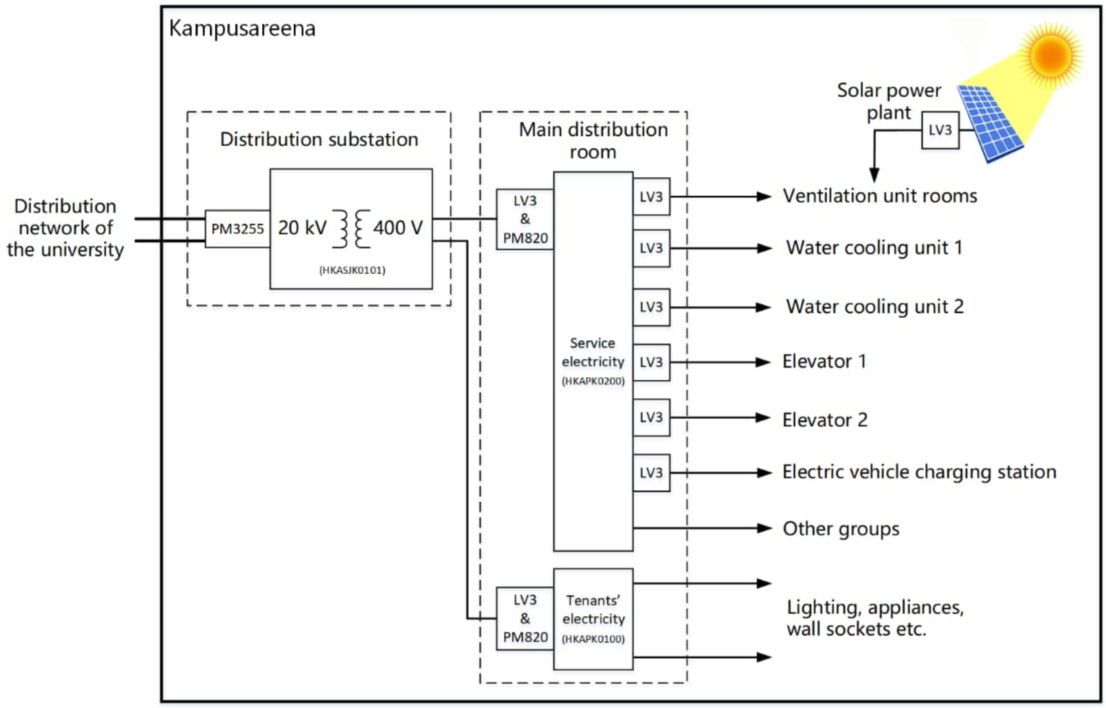

Figure 9** Kampusareena's electricity system's topology. **[40]

As Figure 9 illustrates, the solar power plant supplies the ventilation, but the production from the PV plant is also measured alone. The most interesting measurements are ventilation and EV charging, but also the service electricity is interesting for this thesis.

Measuring devices are Laatuvahti 3 (LV3) [41], Schneider PM3255 [42] and Schneider PM820 [43]. The LV3 measuring device is a power quality measuring device and Schneider PM3255 and PM820 are older consumption measurement devices. All three provide information about the power quality and energy.

From all loads, there was information about frequency, active power, apparent power, reactive power, cumulative energy, harmonics, etc. For this thesis active power, cumulative energy, and tenants' and service electricity frequency were received. More information could have been used, but as verifying methods usually use either the power or energy historical data to form the baseline, these were the needed information. In addition to these, weather data, which is measured with the PV plant, was also available. Information about the temperature inside the building or the CO_2 levels could have been useful, as those indicators would reveal more about the total load.

4.2 Data preparation

When using real-life data, usually some preparation must be done. In this chapter the load profile is introduced, as it has a role in which time slot is chosen for the tests. The cleaning which is done to the data is explained. Lastly, the selected and used data sections are introduced.

4.2.1 Load time series

To get a better understanding of the site this thesis examines, in this chapter the load profile and things that affect it are introduced. As previously stated, this is an office building, which is located on a university campus. The users of this building are university students, employees, teachers, business employees, and visitors. But, because this examination window is 1.6.2020-31.5.2021, the Covid-19 is affecting the use of this building. Also, during summer, there is less usage in this building, as students and workers are on their summer vacation.

During the examination window, the campus was not in complete lockdown, and students and employees were able to use the campus during the daytime. This differs from the normal, as when the campus is open 24 hours a day 7 days a week for those who work or study on the campus. During the examination window, the campus was open for the students and employees from 7:30 to 22:00. The campus would normally be used during those hours the most.

There has been a wide recommendation for everyone in Tampere university to work from home during the examination window, but as Kampusareena also hold business that are not connected to Tampere university and are only renting the space from the building, they have had the opportunity to use the building. Also, during autumn 2020 new students used the campus as normal, and that usage can be seen from the load profile.

Figure 10 and Figure 11 present the load profile in this case. The measurements from PM3255 and PM820 were not suitable to use, as there were a lot of information disconnects, so the total power is gathered from LV3's from service electricity and tenants' electricity.

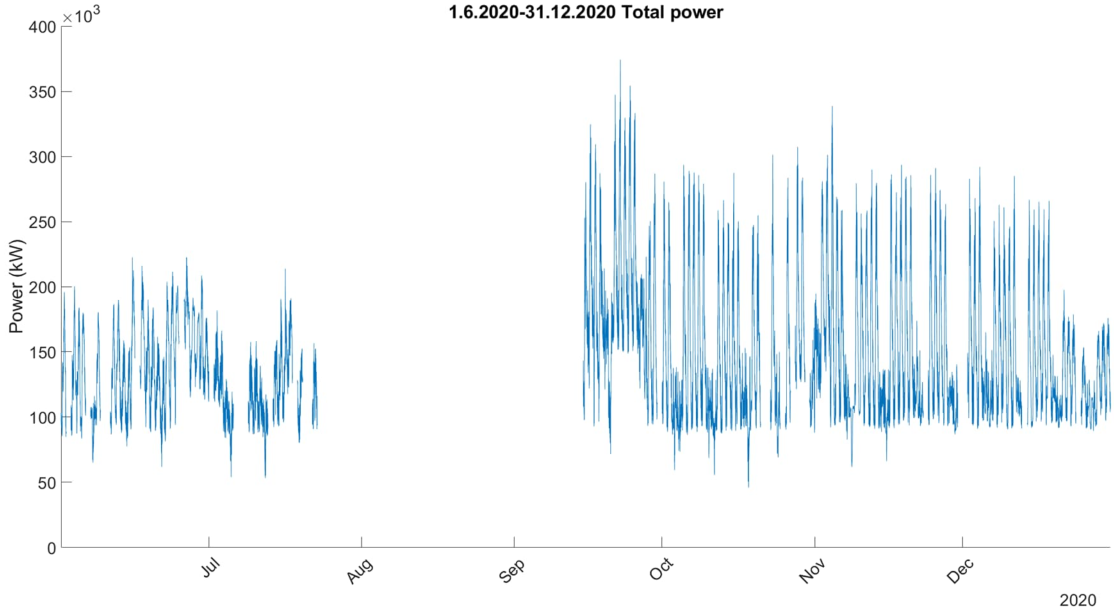

Figure 10 *Total power of the Kampusareena 1.6.-31.12.2020*

In Figure 10 from late July to the middle of September, there was missing data. The gaps between data are missing data. Figure 10 shows that the data from the autumn of 2020 is much more fractured than the data shown in Figure 11. These figures are done with the 15-minute interval data, so this is not the original data which is in a 1-second interval.

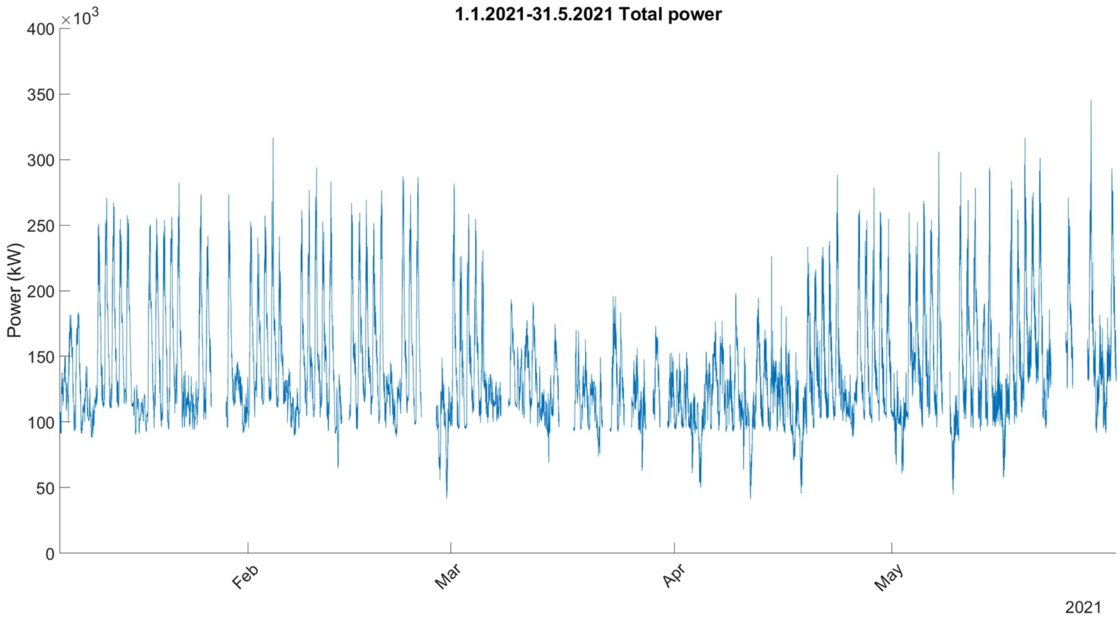

Figure 11 *Total power of the Kampusareena 1.1.-31.5.2021*

These graphs show the total load profile that is examined in this thesis. Weekends can be detected from the curve, and Figure 10 shows that during the summer vacation the consumption is lower than the other months of the year. The lower consumption in March and April can be related to the springtime when there was a bigger lockdown because of Covid-19.

4.2.2 Data selection and intervals

The original data was in 1-second interval measurements, and in this case, it was too precise to use. The amount of data was huge as there is data from each of the different consumptions, as presented in Figure 9. Also, there is no need to have data from each second, as the settlements considering flexibility events are done in an hour interval. One-second data might be relevant when the trade is done closer to the usage hour. But this thesis is focused on the problems with IA and imbalance settlement, so the 15-minute interval is enough.

Imbalance settlement is done with a 1-hour interval, and in the future in a 15-minute interval. As the 15-minute imbalance settlement is not too far in the future, in this thesis all the data is handled within a 15-minute interval. 15-minute data and baselines will give more accurate results than 1-hour baselines. This means, that if the verifying was done by each hour, the changes inside that hour could not be detected, but by doing it by 15-minute intervals, the hour is divided into four sections and the changes between them can be detected. But for some resources even 15-minute intervals are not enough, as with solar power of EV charging the consumption and production can change in minutes.

The conversion from 1-second data to 15-minute data was done with average values. The first 15 minutes is the average value of the second value from 00:00:00 to 00:14:59. The last 15 minutes of the day is 23:45:00-23:59:59. The timestamps in the 15-minute data began from 00:00 to 23:45.

The verifying methods that use historical data, need continuous data from the previous days. A good number of days, with weekends, is around 21 days. 21 days includes enough weekend days to calculate a baseline for a weekend event, and enough days to calculate a baseline for a weekday event.

As Figure 10 and Figure 11 present, there are a lot of interruptions in the measurements and also the load profile itself is not consistent throughout the year. This makes it harder to find which timeslots should be used to test the verifying methods. The best 21 days continuously having all data were in November and in April. November has quite a good presentative of what the load profile in this case is, but April has a lot lower consumption profile than other months. Timeslot 5.-25.4.2021 was still chosen, as it can present the summer months quite well, as there was not good enough data from the actual summer months to use.

The chosen periods are 2.-22.11.2020 and 5.-25.4.2021. These periods do not have any holidays or special days, and all the weekdays and weekends are admissible to use for the verification calculations. Figure 12 and Figure 13 present the total power and temperature on the chosen examination periods. From these Figures, the load profile can be seen even more precisely. The first day of the examination window in both cases is Monday.

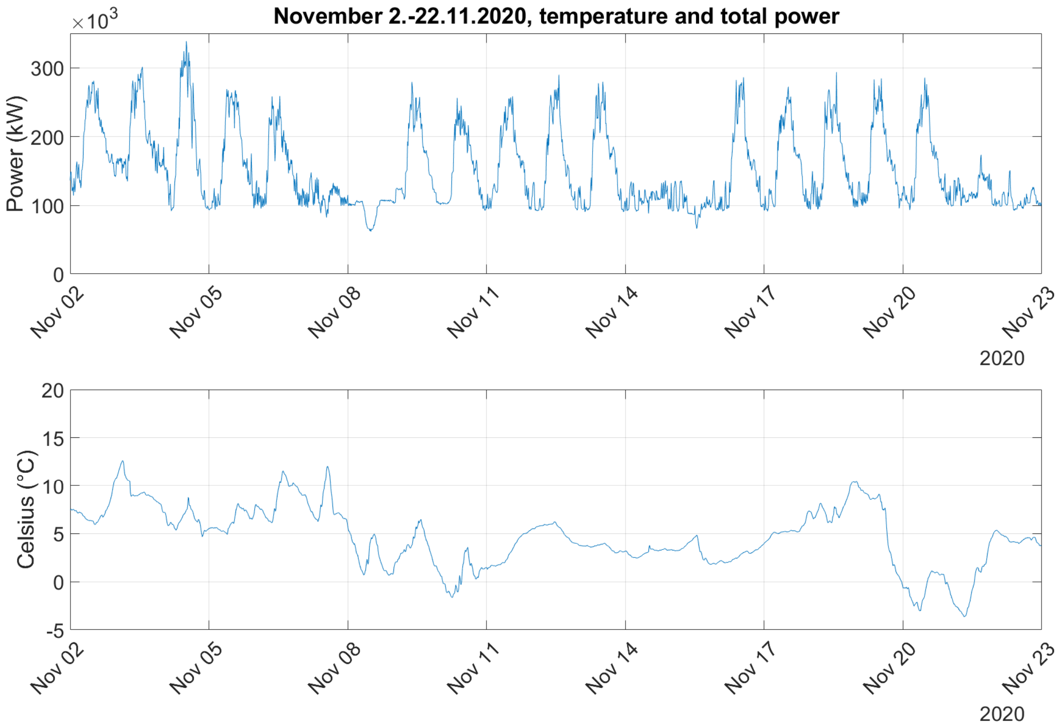

***Figure 12** Chosen examination period, November, total power, and temperature.*

Figure 12 shows that the load profile is quite frequent, the first week is a bit different than the two weeks after it. Weekends can be noticed easily, as the consumption is a lot lower than during the weekdays. The PV production is added to the total load and can cause a small difference between November and April. The baseload is around 100 kW according to Figure 12, the same baseload is shown in Figure 13.

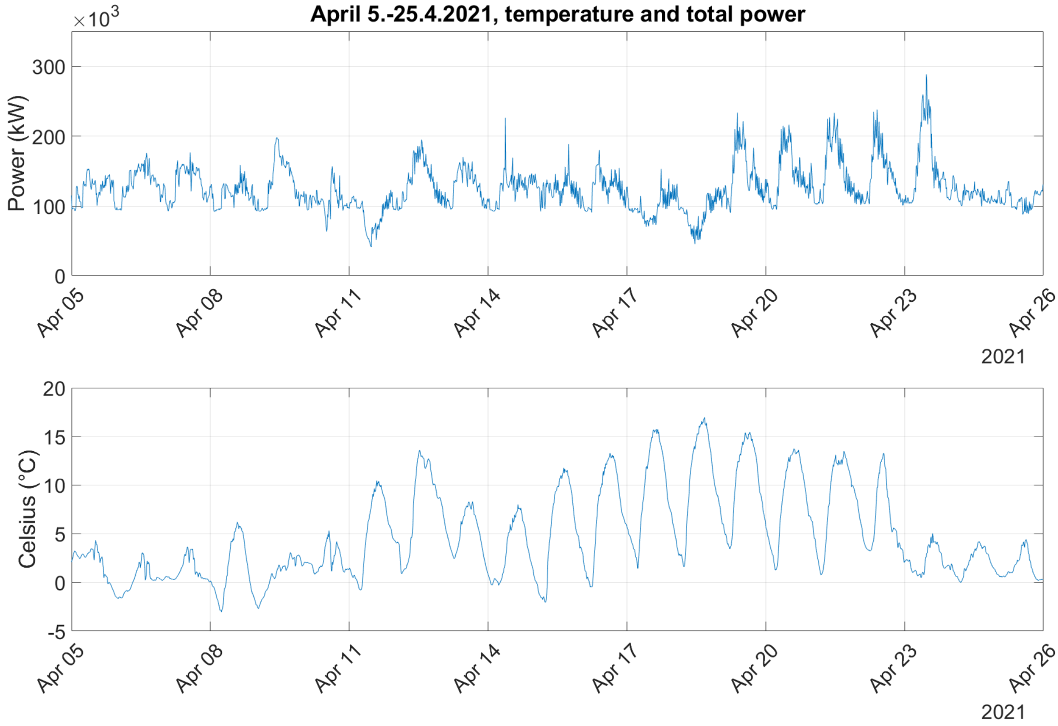

Figure 13 Chosen examination period, April, total power, and temperature.

April had a lower consumption than other months, and it is shown in Figure 13 too. The load profile does not have consistency, as the Novembers load profile. The first week has the lowest load, and this can be partly explained because the Easter holiday ends on the 6th of April that week. The Monday 5th is Easter Monday, which makes this day holiday and not admissible to use for the baseline calculations. Also, the last week, 19.-25.4. has more of the same load profile as November. This will affect the verification results, as the event days are going to be on that week, and the previous two weeks are different load profiles.

Temperature does not seem to have any effect on the power consumption in this case. The temperature is presented in the figures, to give more information about this specific location and case. The temperature is higher in April, and the changes between day and night are bigger, probably caused by the sun. Also, the temperature for November is quite high, especially in the first week. But as said, the temperature changes do not affect the total power, and it won’t be discussed further in this thesis.

4.3 Used methods

As previously stated, in this thesis only Baseline Type I methods are tested. Baseline Type I is the most used method overall, and it has many sub-methods, which are all presented in Chapter 3 from which the used ones are presented in Table 3.

Table 3 Used verification methods in this thesis.

Baseline Type I
X of Y baseline
10 of 10
Mid 8 of 10
High 5 of 10
Comparable day

All the methods presented in Table 3 are described more below. In addition to these, an adjustment factor is applied to all *X of Y* methods, and the method without and with the adjustment factor are compared.

The adjustment factor is an important part of the verification methods, as it considers the event day's characteristics. Methods are all quite similar, but each of them has some adjustment in themselves to be more accurate than the previous, in theory. The simplest method is *10 of 10* method, and all other methods in this thesis are kind of moderations from it.

10 of 10 baseline

10 of 10 baseline calculates the baseline based on the last 10 admissible days, before the event day. In this case study, it means the 10 previous weekdays, as all the week-days are admissible, excluding the Monday 5th of April. *10 of 10* baseline will not be accurate alone, as it will only make the average from the previous days and will not consider any other information.

This baseline method is chosen, because it is the easiest baseline method that there is, and it gives a great basis for comparison for the other methods. This method will not need complicated calculations or a huge amount of data. *10 of 10* method could also be done with a higher number of days. A too high number of days could affect the outcome, as the higher and lower load days can have a bigger impact on the average value. If the event day is normal, and the consumption is on an average level, then more days would probably give more accurate results.

Mid 8 of 10 baseline

Mid 8 of 10 baseline takes the previous 10 days and the highest and lowest load days are excluded. This results that the left 8 days being considered in the average calculation. Mid 8 of 10 baselines have a higher chance to be accurate. Higher and lower loads will have a bigger impact on the average values and passing them will probably give better results. This is only if the event day is considered to be normal, or in the average load profile. Usually, the flexibility events happen on days, when the load profile is a little bit unusual, such as higher consumption. If that is the case, *Mid 8 of 10* might be even more inaccurate than *10 of 10* baseline.

Mid 8 of 10 method is chosen in this thesis because it is usually a better fit for baseline calculations, but exceptions that were described can make it behave poorly. *Mid 8 of 10* method is also easy to calculate, as all of the *X of Y* methods, and for that reason will be useful in this thesis.

High 5 of 10 baseline

High 5 of 10 baseline includes the highest load days of the previous 10 admissible days to the baseline calculations. This method can have better performance than the other two described earlier. Because the flexibility event usually occurs on days, when the consumption is high, or otherwise unusual, taking the days that also have higher consumption into the calculation, will most likely give a better baseline.

This method is included in this thesis, for the same reasons as *10 of 10* and *Mid 8 of 10* methods. The high 5 days is chosen because it is presented in [38] and [39]. Other alternatives for 5 highest days are 4 or 3 highest days [39]. In this case, 5 days is used, because the examination period's load is quite similar, and using more days, the better results should be accomplished.

Comparable day

The comparable day is not the *X of Y* baseline method. And the *Comparable day* method used in this thesis is an alternated version of the one described in 0. *The comparable day* method is usually used when there is a large number of data available and one similar day is used as the comparison for the flexibility event. But because this case is examining only 21 days in each examination window, there will not be a similar day enough to use itself as the baseline. This is solved by looking at all the days before the event day, within the 21 days, to find the three most similar days to the event day. From these three days, an average is calculated and used as the baseline for the event day.

The calculation to find the three most similar days compared to the event day is calculated as shown in the Equation 3.

$$difference = \left|energy_{event\ day} - energy_{compared\ day}\right| \tag{3}$$

The energy of each day is calculated and summed up, and the difference between the energies is compared. The days that had the smallest difference between them were chosen. One day is from 0:00 to 23:59.

The comparable day method is easy to calculate in this case. But it can be calculated more complicated, for example, when using temperature or other weather data to match a similar day. If the comparable day is found only by the load profile, usually the amount of data is reasonable, and calculations are easy.

This method was chosen for this thesis because a different method than the *X of Y* method needed to be tested. *The comparable day* method is easy to calculate and modify for the purpose. *The comparable day* is in this case close to the *X of Y* method, as there is a specific number of days chosen for the average calculation but is still a different method.

4.4 Performance evalutation

To present usable results for choosing the right baseline for a specific situation, specific evaluation parameters are needed. The most usual way to evaluate baselines is with quantitative methods, but in this case, those will not be enough. In addition, qualitative evaluation criteria are needed, and references of which criteria should be used can be found from the literature.

In CoordiNet's report "Markets for DSO and TSO procurement of innovative grid services: specification of the architecture, operation and clearing algorithms" the evaluation is divided into XYZ dimension space. From the axes, the Y axe is the evaluation criteria for choosing the baseline. Accuracy, simplicity, integrity, and efficacy are listed as notable evaluation criteria. [35]

Miriam L. Goldberg, Ken Agnew, and Michael Messenger also list many features and evaluation criteria for baselines in the report "Development of a Standard Baseline Calculation Protocol for Demand Response" [44]. There are similarities, but they also list the ease of use for program participants and administrators and the ease of understanding, lack of bias, costs, and consistency with other market parties [44].

Taking these two reports into account, and keeping the TSO point of view considered, Table 4 presents the evaluation criteria used in this thesis.

Table 4 Evaluation criteria for verifying methods

EVALUATION CRITERIA	QUANTITATIVE	QUALITATIVE
SIMPLICITY		x
EASE		x
ACCURACY	x	
COSTS	x	x
INTEGRITY		x

The simplicity of the verification method is important. Simplicity can mean the simplicity of the verification methods calculations and the availability of data that is needed for the verification calculation. Both are important and have an impact on the results. If the method is easy to calculate and does not require too much memory or complicated algorithms to achieve the baseline, it is better. Simplicity is also affected by the amount of data needed for the calculations. Using historical data is one way, but the amount of it can vary from 10 days to multiple years. In addition to historical data of the load, some methods use weather data, data from other sources, or information related to the examined site. More data to be processed means more complicated method.

Ease can also mean two things, the other is the ease of use and the other is the ease of understanding. Both are linked to simplicity, and in this thesis, the ease of understanding is the more important one. The ease of use is covered in the simplicity evaluation. The ease of understanding means that the parties involved in the verification process understand the mechanisms and reasons for the specific verification method. The goal is to have so simple and easy-to-understand method for the verification.

Accuracy is quite easy to understand as one of the evaluation criteria. Accuracy is measured with quantitative evaluation criteria presented later. Accuracy is one of the most important aspects of verification, to achieve the best results and to make different parties satisfied. Verifying flexibility events accurate is important for the providers and the aggregators, DSOs, and TSOs, so the right amount of compensation for the event is found.

The accuracy errors of the chosen method should also be understood and accepted by all participants. One universal method that would work perfectly for all the situations, cannot be found and this needs to be recognized.

Costs are also a few different things, in this case, costs of doing the verification, costs of the calculation system, and the costs that the flexibility provider might face if they have to install new equipment in order to measure relevant data. There will not be precise calculations of the actual costs that these things would bring in this thesis, but things that might cause higher costs are evaluated. Costs are closely linked with simplicity.

Integrity, or the minimization of gaming, means that the method cannot be tricked into wrong results. This also includes that the method will not encourage flexibility providers into wrongful actions to benefit from the methods. Or at least the most obvious ways should be avoided.

The quantitative, accuracy, performance of load estimation is usually evaluated with root mean square error (RMSE) and mean absolute percentage error (MAPE) [45] [46]:

$$RMSE = \sqrt{\frac{1}{96}\sum_{t=1}^{96}(bl(d,t) - al(d,t))^2} \tag{4}$$

$$MAPE = \frac{100}{96}\sum_{t=1}^{96}\frac{|al(d,t) - bl(d,t)|}{al(d,t)} \tag{5}$$

Where $bl(d,t)$ is the baseline and $al(d,t)$ is the actual power on a day d at the time t. Time intervals are 15-minute, and the diver 96 comes from the number of 15-minute intervals in one day.

MAPE value is the percentage value of the absolute error between the actual power and baseline and is a good indicator if the baseline works overall. It does not state if the baseline is over or under the actual load, only if there is a difference either way. MAPE can be used as a base indicator for the evaluation, but further evaluations should be done with MAPE. Absolute percentage error can also be calculated for each value,

Root mean square error highlights bigger differences more than small errors between baseline and actual load. RMSE also shows only the error, not that if the baseline is lower or higher than the actual load.

With RMSE and MAPE values it is good to have a Cumulative Distribution Function (CDF) to demonstrate how the errors are distributed. With CDF the errors can be divided into when the baseline is smaller than the actual load and when it is higher than the actual load. For CDF the error between baseline and actual load is calculated as follows:

$$bl(d,t) - al(d,t) = difference(d,t) \tag{6}$$

If the difference is a positive value, the baseline is higher than the actual load and if the difference is negative, the baseline is lower than the actual load. These values are arranged into cumulative order, the closer the values are to 0 on the function, the better.

5. RESULTS

This Chapter presents the results from the test done to the case examination site. As Chapter 5 describes, 4 different baseline methods are tested, and an adjustment factor is used in the *X of Y* baseline tests.

5.1 Baseline method tests

Figure 14 presents the baselines and actual data from November 20th, 2020. Figure 14 shows, that all the tested baselines work quite well.

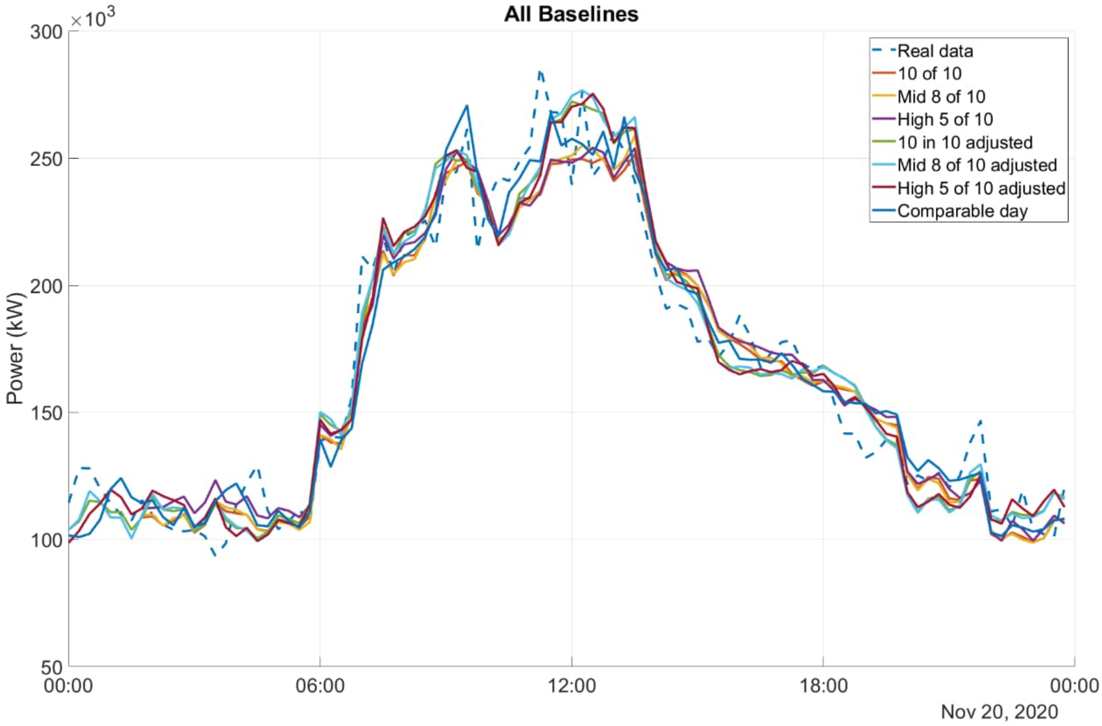

***Figure 14** All tested baselines 20.11.2020.*

In Figure 14 illustrates the baseload, which is around 100 kW, and the load starts to increase after 6 am. Towards the evening, load again starts to decrease, and all of the baselines do present the actual load well. From the figure one method over another cannot be seen, but for example the peak and decrease after it at around 10 am is quite well predicted by the methods.

Figure 15 presents the same baselines tested as in Figure 14, but for April 23rd, 2021. The baselines do not work as well as in Figure 14, but *High 5 of 10* with adjustment, *Mid 8 of 10* with adjustment, and *10 of 10* with adjustment perform the best.

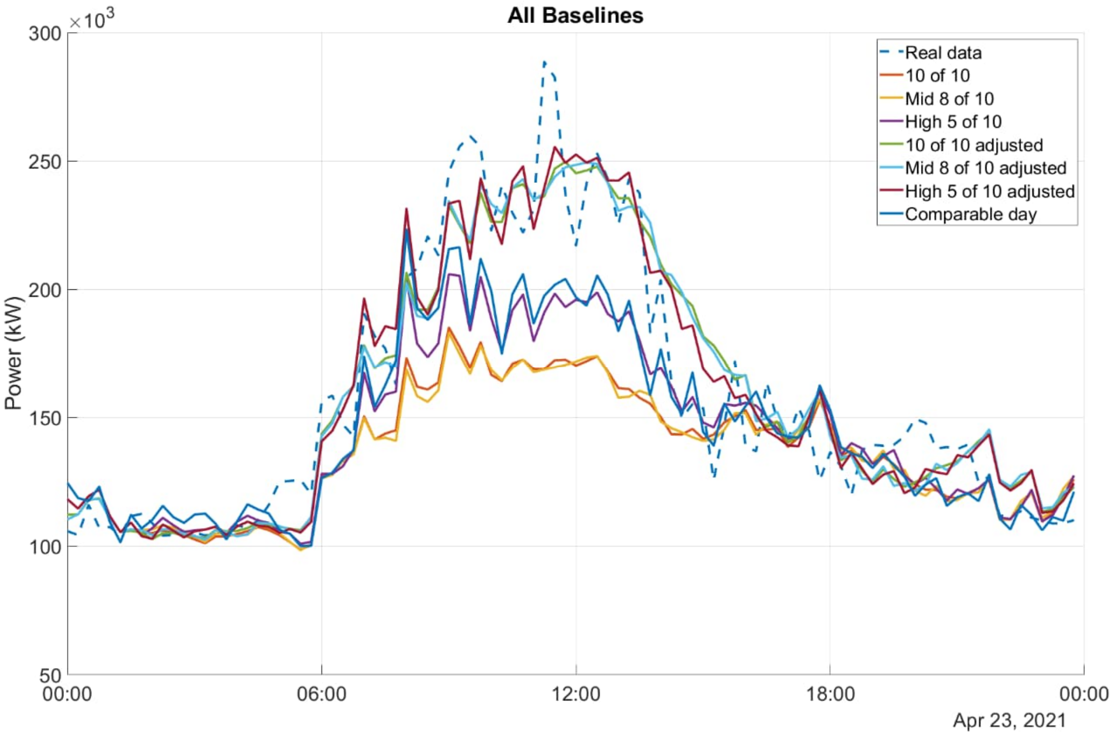

***Figure 15** All tested baselines 23.4.2021.*

Figure 15 shows the reality of the baselines when consistent load pattern from historical data is not available. All of the non-adjusted methods give a baseline that is a lot lower than the actual load. Also the peaks and decreases in the load are not predicted as well as in the November's case.

Figure 16 presents again the same baselines, but for the weekdays 16.-20.11.2020. As All of the days are quite good, and the baseline pattern seems good enough. Some scattering can be seen at the beginning of the week.

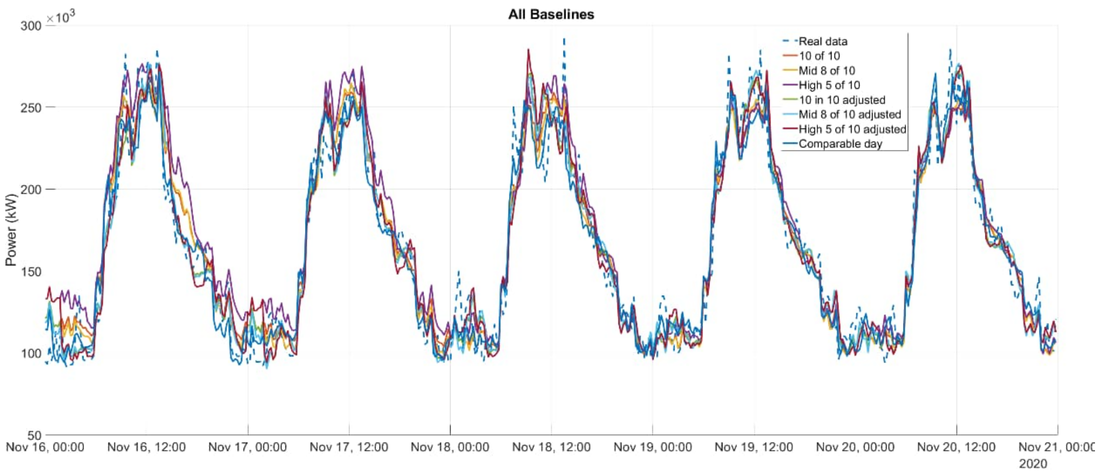

***Figure 16** All tested baselines from time period 16.-20.11.2020.*

Figure 16 illustrates that when consistent and reliable historical data is available, quite accurate baselines can be formed with rather simple methods. The scatters that can be seen in the beginning of the week, can be explained with the load profile which had some irregularities in the first week of the examination period. Those days do not impact the baseline reformed for the Friday 20th and for that reason the baselines show to be more precise for the end of the week.

Figure 17 presents the baselines tested on the period 19.-23.4.2021. Figure 17 shows that Friday 23rd is a bit different load profile day than the previous four, and as stated in Chapter 4.2.1 the whole week had a little bit different load profile than the previous two weeks. This can be seen from the *10 of 10* and *Mid 8 of 10* baselines, which are much lower than the actual load.

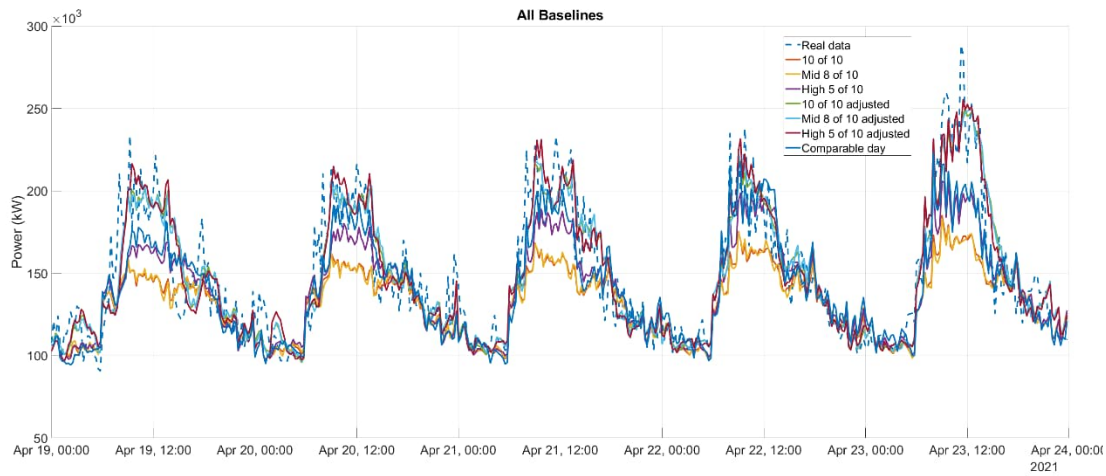

Figure 17 *All tested baselines from time period 19.-23.4.2021.*

Figure 17 shows, that the Friday 23rd has the most inaccurate baseline, but also the load profile that day is much higher than for the beginning of the week. As the Figure 16 showed great results with even simplest methods, Figure 17 shows that the same methods might not always work.

The results from the baseline tests are successful. They show that with only 21 days of data quite accurate baselines can be formed but also that with same amount of data and the same site, but during different season, the result are not well. Small error should be tolerated and with slight moderations to the baselines, in both cases better results should be achieved.

5.2 Quantitative evaluation

The evaluation criteria are presented in Chapter 4.4. The quantitative performance evaluation methods MAPE, RMSE, and CDF results are presented in Figure 18, Figure 19,

Table 5, and Table 6. Figure 18 and Figure 19 presents the Cumulative Distribution Function of the error between baselines and actual load on 20.11.2020 and 23.4.2021. These same baselines are presented in Figure 14 and Figure 15. In literature [45] the CDF is presented as the RMSE values, but as RMSE values exaggerate the higher values, and does not specify negative or positive values. The CDF in this thesis is presented based on the actual errors between the baseline and actual load.

Figure 18 illustrates that the baselines all work quite well. The biggest differences are with *10 of 10, Mid 8 of 10,* and *High 5 of 10* baselines and are almost 50 kW under the actual load. The curve raises rather straight around 0 kW, which means, that most of the errors between the baselines and actual load are small, and around from -10 kW to 10 kW. As Figure 14 shows, the baselines work well on this day and the CDF confirms this.

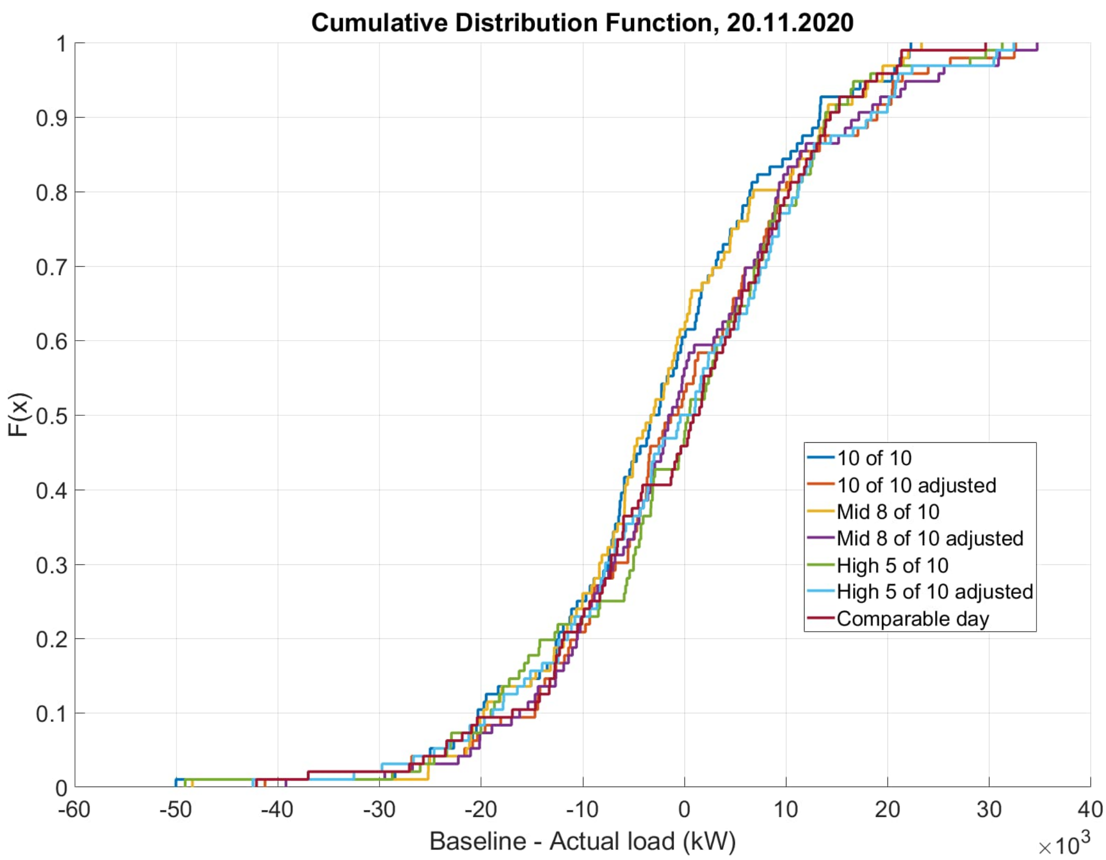

***Figure 18** Cumulative Distribution Function for the error between baselines and actual load 20.11.2020.*

As Figure 15 also shows, the baselines do not perform as well in 23.4.2021. The CDF of the errors shown in Figure 19 also presents the same. Methods *10 of 10* and *Mid 8 of 10* show errors big as -120 kW, which is almost one-third of the whole load. Also, almost all the errors are between -80 kW to 0 kW, which means that most of the baselines underachieve in this case. *10 of 10* with adjustment, *High 5 of 10* with adjustment, and *Mid 8* *of 10* with adjustment methods seem to have much better performance, where the errors fluctuate from -50 kW to 50 kW, almost the same variation as in 20.11.2020.

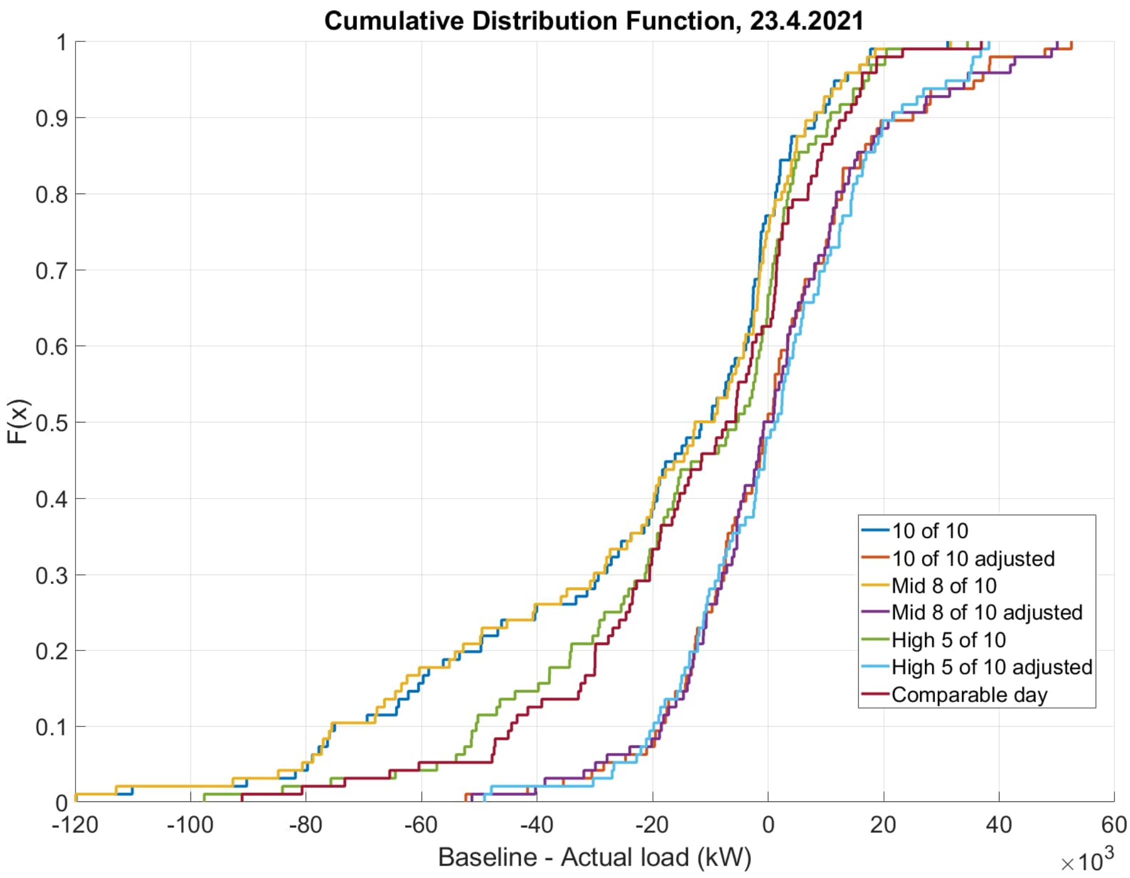

Figure 19 *Cumulative Distribution Function for the error between baselines and actual load 23.4.2021.*

Also, Figure 19 illustrates that *High 5 of 10* and *Comparable day* methods on their own are performing better than *10 of 10* and *Mid 8 of 10*. To note that the *Mid 8 of 10* with adjustment, *10 of 10* with adjustment, and *High 5 of 10* with adjustment seem to work better than methods shown in Figure 18, that the scale of X axle is in Figure 18 is from -60 kW to 40 kW and in Figure 19 from -120 kW to 60 kW.

Table 5 presents the RMSE values for each baseline on specific days that could be tested in this thesis. It shows that the days in November, in general, are performing better than the days in April. The best and worst values are highlighted in the table. The best day measured with RMSE is 17.11.2020 with the *Comparable day* method with only 11,61 kW. The worst day is 23.4.2021 with the *Mid 8 of 10* method RMSE value of 38,59 kW. At the bottom of Table 5 is also presented the average values of the RMSE for each verification method. Considering those values, the adjusted versions of *10 of 10*, *Mid 8 of 10,* and *High 5 of 10* methods perform the best.

Table 5 *Root Mean Square Error*

	10 of 10	10 of 10 adjusted	Mid 8 of 10	Mid 8 of 10 adjusted	High 5 of 10	High 5 of 10 adjusted	Comparable day
16.11.2020	18,66 kW	15,64 kW	17,39 kW	15,21 kW	27,30 kW	18,39 kW	16,16 kW
17.11.2020	12,95 kW	13,07 kW	12,38 kW	13,06 kW	19,63 kW	15,87 kW	11,61 kW
18.11.2020	15,70 kW	16,90 kW	16,28 kW	17,47 kW	17,68 kW	17,97 kW	15,76 kW
19.11.2020	13,13 kW	13,66 kW	12,94 kW	13,49 kW	13,88 kW	14,64 kW	13,98 kW
20.11.2020	12,91 kW	13,36 kW	12,83 kW	13,34 kW	13,64 kW	14,19 kW	13,08 kW
19.04.2021	28,29 kW	18,79 kW	27,96 kW	18,70 kW	23,99 kW	19,82 kW	22,87 kW
20.04.2021	24,96 kW	16,47 kW	25,68 kW	17,50 kW	20,15 kW	17,75 kW	17,67 kW
21.04.2021	29,06 kW	17,33 kW	29,38 kW	18,43 kW	21,27 kW	18,07 kW	19,08 kW
22.04.2021	21,22 kW	12,75 kW	21,68 kW	13,36 kW	14,76 kW	12,56 kW	14,22 kW
23.04.2021	37,96 kW	17,87 kW	38,59 kW	18,03 kW	28,02 kW	17,05 kW	26,18 kW
Average	21,48 kW	15,58 kW	21,51 KW	15,86 kW	20,03 kW	16,63 kW	17,06 kW

As previously stated, the RMSE value exaggerated the large errors, and for that reason, even when some of the verification methods seem to work as good on 23.4. and on 20.11., the bigger errors on both ends of the CDF presented in Figure 18 and Figure 19 makes the RMSE value higher.

Table 6 presents the MAPE values for each day and method. Again, the worst and best values are highlighted. The worst MAPE value is on 16.11.2020 with a *High 5 of 10* method of 17,37 % and the best is on 22.4.2021 with a *High 5 of 10* with an adjustment of 5,85 %. MAPE value presents the overall performance of the baseline compared to the actual load. The average of each method's MAPE values are on the bottom of the table, and according to them, *10 of 10* with adjustment and *Mid 8 of 10* with adjustment perform the best. Also, November days are performing better with MAPE as well.

***Table 6** Mean Absolute Percentage Error.*

	10 of 10	10 of 10 adjusted	Mid 8 of 10	Mid 8 of 10 adjusted	High 5 of 10	High 5 of 10 adjusted	Compara-ble day
16.11.2020	11,26 %	7,54 %	9,99 %	7,29 %	17,37 %	9,27 %	8,13 %
17.11.2020	7,81 %	7,56 %	6,92 %	7,24 %	12,06 %	9,11 %	5,89 %
18.11.2020	7,31 %	7,87 %	7,54 %	7,99 %	8,40 %	8,33 %	7,51 %
19.11.2020	6,20 %	6,76 %	6,12 %	6,56 %	6,65 %	7,36 %	6,84 %
20.11.2020	6,05 %	6,60 %	6,12 %	6,56 %	6,64 %	7,15 %	6,57 %
19.04.2021	12,85 %	9,93 %	12,49 %	9,73 %	12,14 %	10,44 %	11,85 %
20.04.2021	11,15 %	8,08 %	11,57 %	8,41 %	10,44 %	9,29 %	8,98 %
21.04.2021	11,64 %	8,22 %	11,81 %	8,74 %	8,98 %	8,56 %	9,05 %
22.04.2021	8,95 %	5,94 %	9,11 %	6,09 %	7,19 %	5,85 %	7,76 %
23.04.2021	13,29 %	8,46 %	13,45 %	8,55 %	10,47 %	8,50 %	10,44 %
Average	9,65 %	7,70 %	9,51 %	7,72 %	10,04 %	8,38 %	8,3 %

Figure 22 presents the CDF for all errors with each baseline when the data from November and April are combined. The methods seem to be in the same range and performing quite well, but again the scale of X axle should be noted. All the baselines act quite similarly between 0 and 60 kW and there is a lot of distribution between -120 kW and 0 kW. The best performing baselines according to Figure 22 are *10 of 10* with adjustment, *High 5 of 10* with adjustment, and *Mid 8 of 10* with adjustment. All their biggest errors are between -70 kW and -61 kW and 52 kW and 54 kW.

Figure 20 and Figure 21 show one more time the absolute percentage errors for each baseline method compared to the actual load. The time of the day is added, to show to what time of the day the biggest errors take place.

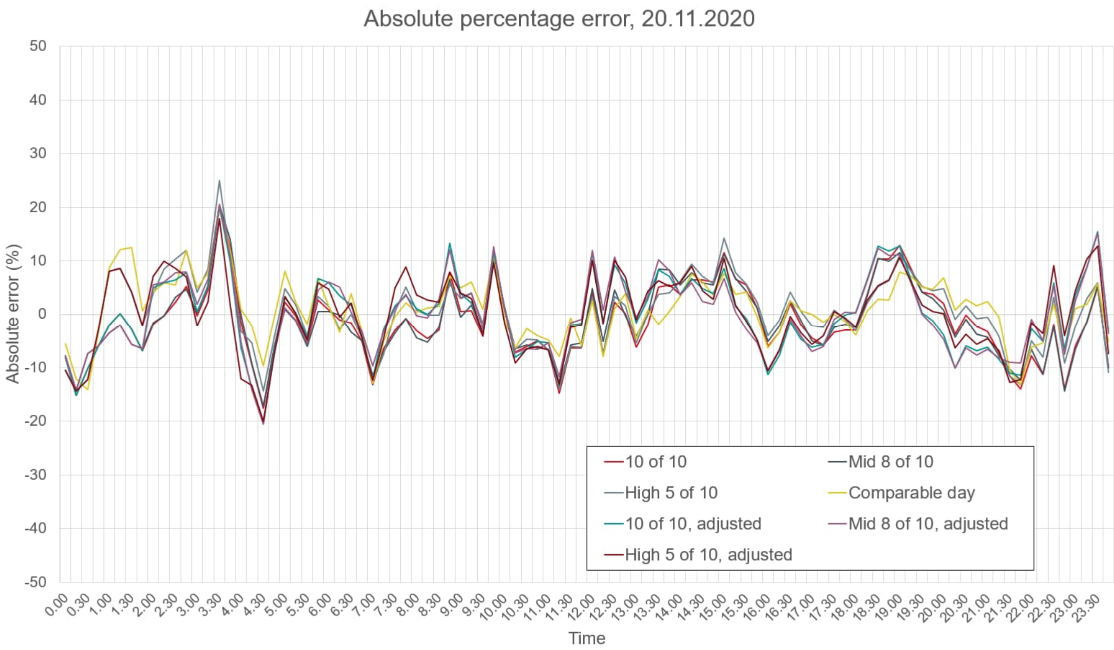

Figure 20 *Absolute percentage error between baselines and actual load 20.11.2020.*

Figure 20 presents the Friday 20th and the absolute percentage error which is calculated subtracting the actual load from the baseline. This means, that when value is negative, the baseline is under the actual load, and when it is positive, the baselines is over the actual load. Biggest errors in this case are during the early morning hours, and the errors during the day are around 10 percent.

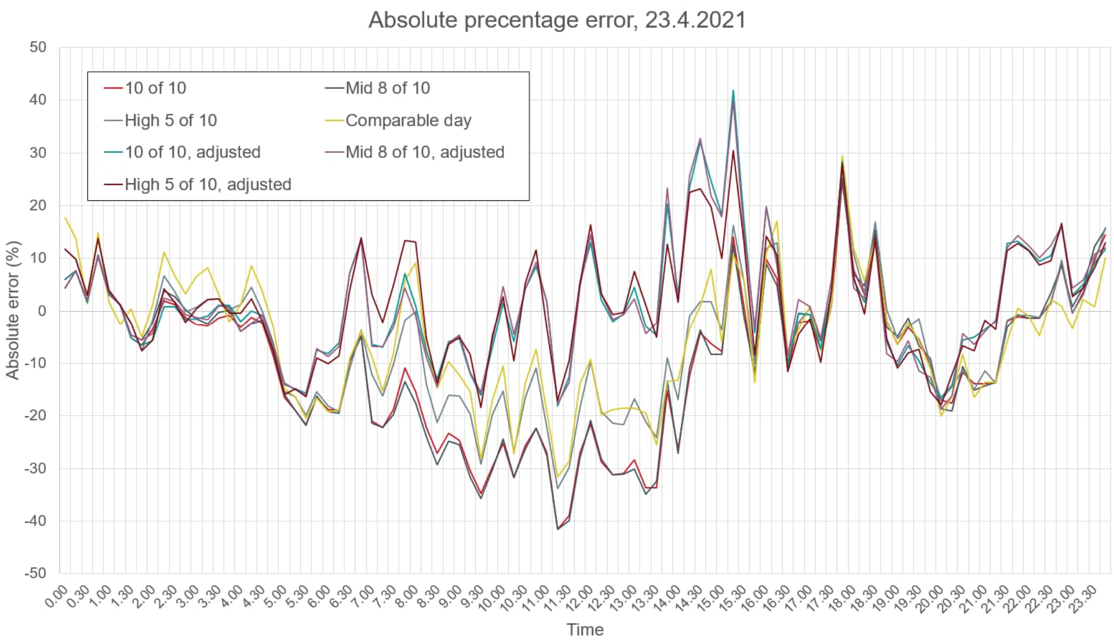

Figure 21 *Absolute error between baselines and actual load 23.4.2021.*

Figure 21 shows the same values as for the November, but for the April 23rd. The errors are much bigger than in the November's case. Also the alarming thing is, that the biggest errors occur during the daytime, and during the night-time the errors are only around 10 percent. The adjusted methods seem to have the smallest errors. Same results are shown in all of the figures presented in this work. But also with them, around the afternoon, the errors are quite huge, considering that during that time most of the flexibility events usually happen.

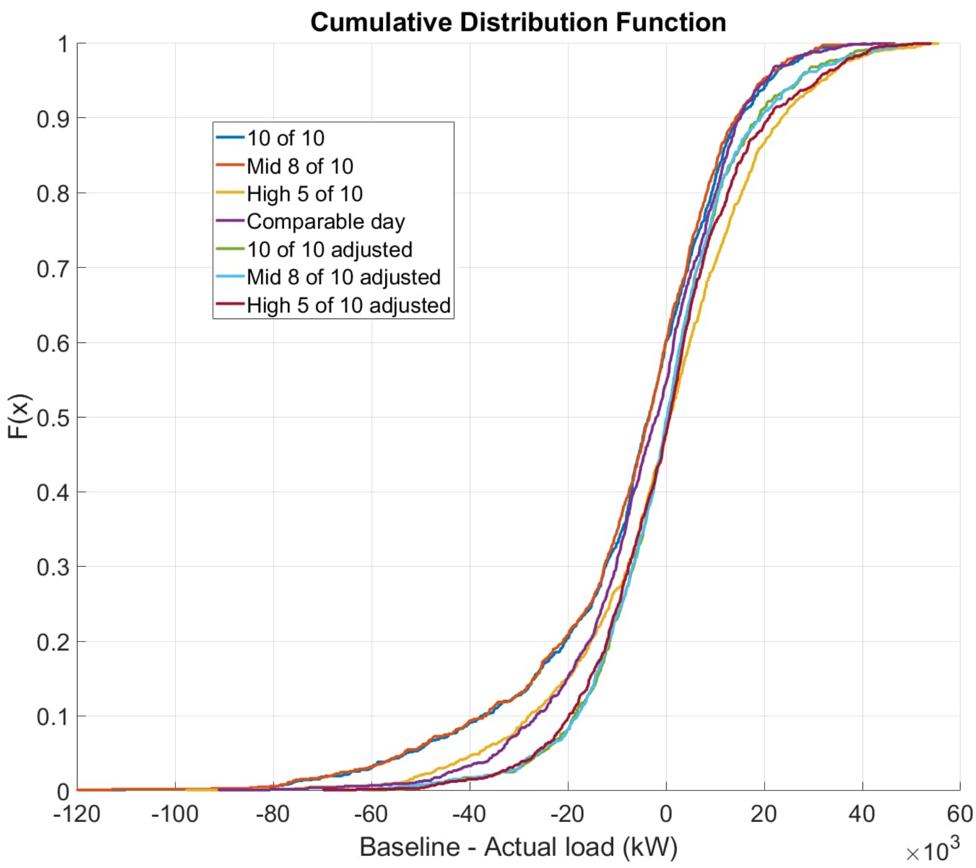

***Figure 22** Cumulative Distribution Function for all baselines, April and November data combined.*

The tests presented here are all done to the whole actual load. These same baselines could be used also with smaller and specific sites such as using the methods individually for solar panels or EV charging. The load EV charging creates is highly variable and the baselines it creates are not the best with the methods used in this thesis.

5.3 Qualitative evaluation

As presented in Table 4, there are also four qualitative evaluation methods used in this thesis. The simplicity, ease, costs, and integrity of the tested methods are estimated in this Chapter. Costs could be a quantitative evaluation method, but in this thesis, the costs are only evaluated with words, because the actual costs of each method are whole different question and research to be done.

Simplicity

All of the used methods are quite simple. *X of Y* method is a bit simpler than *Comparable day* in this thesis' situation at least. *10 of 10* method is the simplest, as no further selection or exclusion of days needs to be done. Choosing the Y days can be complicated if previous event days and holidays had to be found from the data, but in this case, because there were only one, excluding the weekends and one Monday was rather easy. The same applies to *Mid 8 of 10* and *High 5 of 10* methods with the Y days. Finding the excluded days from *Mid 8 of 10* and *High 5 of 10* methods can be done with very simple calculations. Comparing the X of Y methods' simplicity is difficult, but the *10 of 10* method is the simplest way to calculate baseline. It is fairly obvious why it is easy and simple, but also, if no baseline method was used, probably the way to calculate some kind of reference level, would be by calculating the average load from previous days.

The comparable day method is somewhat more difficult to calculate, but still rather easy. In this case, *the Comparable day* method does not include into consideration any other aspects than the load. Some *Comparable day* methods are including information about the weekday, weather, or time of the year when choosing a similar day. In this thesis, the weather seemed to not have any impact on the load, so weather and temperature examinations were left out.

Using adjustment factors with the baselines increases the difficulty of calculating the baseline. Adjustment factor needs to be calculated for each situation and each day and baseline separately, so this means more information and more calculations. Also again, as in this thesis, the weather examinations were left out, the adjustment factor considers only the actual load and baseline. Using regression models to add the information about weather or other relative conditions is more complicated than the adjustment factor used in this thesis. Using regression would increase the complexity of the calculations but would not be considered a hard way.

Ease

As all the methods are relatively simple, the ease point of view does not differ too much. Because the methods are simple, the mechanisms of the methods are easy to understand.

The comparable day itself is not hard to understand as a method, but the use of it brings problems. *Comparable day* baseline can be only calculated after the event day, and one of the parameters of ease of use and ease of understanding is that the user can understand and know the baseline beforehand. When knowing the baseline, the provider knows exactly what the buyer of the flexibility is looking for from them.

X of Y baselines are on themselves quite easy to understand and easy to calculate beforehand. Adding adjustment factors does decrease the ease of understanding. As described before, when using adjustment factors, the baseline is corrected just right before the event, and this makes it for the flexibility provider harder to follow the baseline throughout the event. When adding the weather information to adjustment factors, the factors can be seen as confusing. Adjustment factors used in this thesis are just comparing the actual load and baseline, and adjusting the upcoming baseline based on that. It is quite straightforward, that if the baseline is lower than the actual load, the adjustment factor will be over 1 and correct the baseline upwards and vice versa. When using weather conditions as well, the correction can be smaller or much bigger than expected.

Methods studied in this thesis are all easy to understand and can be calculated with a small effort by the flexibility provider. Or the party doing the verifying can rather easily inform the flexibility provider about the verifying.

Costs

Costs of verifying methods are whole of many specific things, such as the costs of the system calculating the baselines and costs that the method can cause to the flexibility verifier. Also, the flexibility provider can face costs when submetering is required or if the method works against the provider. Costs of wrongful working of the verifying method are evaluated in the efficacy part.

Because the methods used in this thesis are all rather simple and would not need complicated calculations, the cost of each method based on this could all be seen the same with costs in this area.

But because there are also costs from installing submetering and sending the information to the verifier, the most cost-effective method would be *a comparable day* method, as it is the one method that does not need its own submetering and can use another similar site's submetering to use as the baseline. Each of the *X of Y* methods needs submetering, and quite accurate data, and the devices that can do it are expensive.

One possible way to decrease the costs with *X of Y* methods would be if instead of submetering, the baseline used the metering from the whole site where the flexibility resource is located. This could be possible if the flexibility resources' effects on the site could be determined from the whole load. In this case, it would not be possible with EV charging. The flexibility resources share from the whole load must be bigger and the flexibility events large enough to see results when events occur.

Integrity

X of Y methods used in this thesis is somewhat manipulable, as the use of adjustment factor encourages it. When using the adjustment factor, the time window the adjustment factor is calculated should be informed to the flexibility provider. If they are supposed to decrease their consumption for the flexibility event, they could slightly increase their consumption during those hours from where the adjustment factor is calculated. This would make the baseline, which has been ideally following the actual load, to be too low compared to the actual load and the adjustment factor would be something over on and increase the baseline for the event. This way the flexibility provider does not have to increase in total their consumption during the event. Of course, this is a quite complicated way to benefit from the flexibility event, but it is possible.

The comparable day has similar issues, as the flexibility provider can try to mess up the day's load profile, so a higher consumption day would be selected to be the comparable day. With *Comparable day* all of the hours of the day need to be tempered if the flexibility provider wants to benefit from the flexibility event because of the baseline. This can be solved in weather-sensitive sites, by including the weather information in the comparison. *The comparable day* method is not considered to be gameable.

X of Y methods without adjustment factor can be gamed, but with quite large actions, that could hurt the provider more financially than just acting as their baseline and receiving the actual compensation for the right flexibility event. Most gameable baselines are the ones using adjustment factors.

6. DISCUSSION

In this thesis, seven different verifying methods were tested. These methods were *X of Y* methods and *Comparable day* methods. Chapter 5 presents the results from those tests and an evaluation of the performance of those methods. All of the methods worked quite well, taking into consideration the data and environment they were tested. As a result of the tests, this thesis presents a comparison of the methods and reasons why the methods work, and which reasons are making specific methods usable.

In this chapter, the results of this thesis are discussed. This thesis describes a few baseline methods in a specific site, and the further tests and research that should be done are also presented in this chapter.

Baselines

When choosing the best baseline method from the tested methods, accuracy should be the most important aspect to consider. Based on the tests it is hard to choose one method over others. The measured data was limited, because of problems with missing data, limited time and resources, and a small testing environment. As the tests that were completed were successful, the results from them are not so comprehensive.

Based on the results, methods with adjustment factors worked seemingly better than without the adjustment factor. In conclusion, if *X of Y* methods are used, some kind of adjustment factor should be used with them. There are a few reasons why the methods with adjustment factors worked as well as they did compared to the non-adjusted ones.

The adjustment in this thesis was done for each 15-minute value separately, but in a real event, the adjustment would be done based on the few hours before the event, for the whole event, which could be hours long. This creates more errors, and the methods could perform on the same level as the worst-performing methods in this thesis. Another reason for seemingly good performance on behalf of all the methods is that the analysis periods were quite short and had similar load profiles. Tests for April's days showed a little bit of what the baselines could look like if the load in previous days is a lot different than on the event day. In real events, it is more possible for events to occur during days when the load profile is unusual or other conditions are different from the normal.

The focus of this thesis is demand-side flexibility, and even though the examined building had a load of around 300 kW, there was only 10-100 kW of the actually flexible load. The most interesting would be the EV charging station, but as there are only a few charging points, the data is irregular, and in real life, those charging stations would be aggregated, and the effect of one station would not be relevant in the big picture. Figure 23 presents the load and baselines of this site's EV charging.

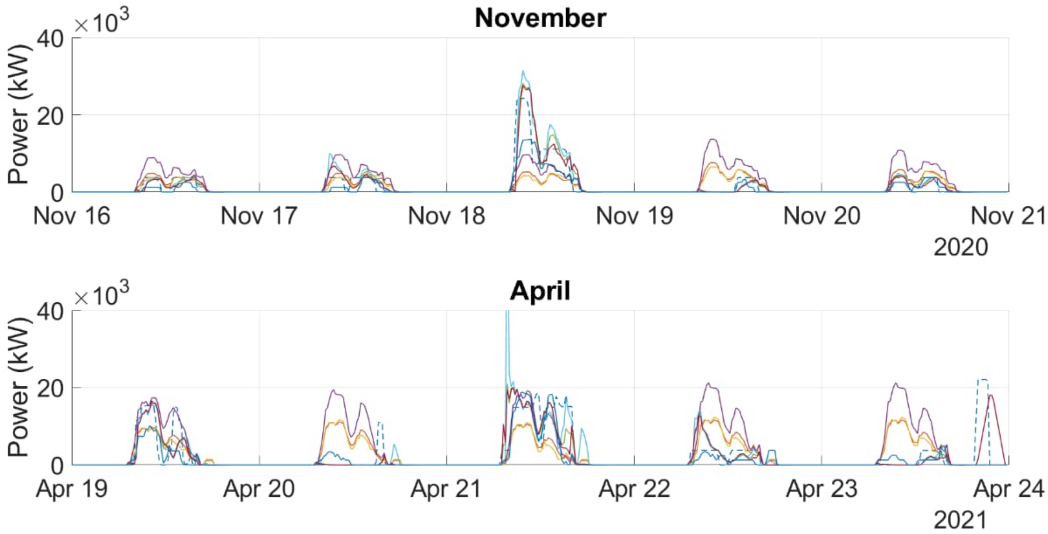

Figure 23 *EV charging station load and baselines.*

The colors of the lines are the same as in the previous figures in this thesis so that the blue dashed line is the actual load. The load is small, around 5 kW to 20 kW and because there are only a few charging stations, the connection of the car can be seen, and the baselines are trying to impose that. But the EV charging could be demonstrated with baseline if more than one station was used, as then there would be a more consistent pattern of load.

Figure 23 presents the other problem with the tests and examination in this thesis. The load profile is quite regular when the whole load is considered. If the load profile would be as regular as in this case, the baseline forming would be easy. The baselines tested in this thesis with this data did perform well, but the results cannot be generalized to all flexible resources and situations.

Verifying methods for other sites

The scalability of the method is an important factor when considering the TSO point of view. From the results of this thesis, the baselines do not work when the load is irregular and small. The methods could work if more EV charging stations were grouped, and the load pattern would have a better consistency. This would mean having sub-metering at each charging station and using an aggregator to control them and selling the potential from them.

CoordiNet's report presents a "decision tree" for the choosing of the verifying method based on if the flexibility event provider is scheduled, are the resources are aggregated, if the resources are the same type of DER and if there is sub-metering available [35]. Based on the decision tree, from the *X of Y* methods, only *High X of Y* is mentioned and used in the case when the resources are not aggregated and are small consumers, heat pumps, or large consumers. *A comparable day* method should be used, based on the tree, only when the resources are aggregated with a VPP concept and no submetering is available. [35]

The problem with the decision tree presented in CoordiNet's report is that almost every method is relying on the sub-metering. Requiring sub-metering from everyone participating in the flexibility markets will limit out participants. As previously stated, sub-metering is expensive, and every flexibility resource provider cannot provide it.

Flexibility register

This thesis is done as part of the EU-project OneNet and the flexibility register, which is developed in the OneNet. A flexibility register is a concept, where the distributed resources are brought together so that they would be easier to implement into the electricity markets. This concept is also presented in the report "An integrated approach to Active System Management" by CEDEC, E.DSO, Eurelectric, GEODE, and ENTSO-E [47].

In the flexibility register, the functionalities of the register are divided into two: controlling the resources' information who are participating in the markets and supporting the verifying and settlement of the flexibility that is provided by distributed resources. This thesis gives information to the verifying part of the register. It is not still decided which is the participant that should be responsible for the verifying.

The flexibility register's main idea is to give support to the market participants to participate in the electricity markets as fairly as possible. The question is still, if the flexibility register is the right party to validate the verification, or should the register just give a framework for the aggregators and other flexibility providers how to calculate the baseline. Providers could calculate their baseline and inform it to flexibility register or flexibility register could have established practices for calculating the baseline for all flexibility providers.

In sthlmflex's pilot, participants are given two options for verifying, they can either use the baseline FlexHub calculates based on the last 5 weekdays averages on one hour or they can suggest their own verifying method which is evaluated by Vattenfall Eldistribution & Ellevion [48]. Choosing this method to let participants choose the verifying method can be effective, as participants with irregular load patterns can still get more specific and accurate verification to their flexible resources. If the participant chooses FlexHub's baseline, and it would be too incorrect, the responsibility for that would be on the participant.

For the flexibility register, further investigation into different verification methods should be done and this thesis should be used as the framework for the subject. The flexibility register would probably be owned by TSO and DSOs, which are discussed earlier as the neutral parties, and good candidates to be responsible for the verifying process. The same questions about who is responsible for the verifying and on what level the verifying should be done are important in the flexibility registers case. Verifying and settlement models will also depend on what are the models and ways the flexibility providers can offer their flexibility to the markets.

Further work

For further work, interviews with flexibility providers, BRPs, independent aggregators, and other parties relevant to the verifying process should be done. With the interviews the opinion of the flexibility event participants could be heard about topics such as what methods seem workable for them, who they think should be responsible for the verifying and what would be the propel level to do the verifying.

TSO cannot alone decide what is the best opinion for other participants, and especially the question of how large errors of verifying can be tolerated should be discussed with the parties in question.

In addition to the interviews, further tests on different sites and cases of flexibility should be done. Tests on different sites that have flexible resources behind them to see how the flexibility events can be detected without sub-metering and tests on a larger amount of same flexibility resources grouped to see if these methods would work for them too. Also, further tests on other methods than the methods tested in this thesis should be done, as some recourses probably will need different methods.

The need of sub-metering can be a tough question to research. The relatively high cost of installing sub-metering to all flexibility resources is discussed, but if there are other reasons to increase sub-metering, would the cost be even that high after all. The further research can go either way, if other reasons to install sub-metering is found, are the problems with verifying solved? Or if the sub-metering is just too high in costs and no other reasons for doing it is found, how the flexibility resources actions can be detected from the whole load of the site?

This thesis is done in the mindset that it is possible to use different methods for verifying in varying situations, with different flexibility resources and market participants. The categorization of the baseline methods does classify methods by on which situation each of them is usually used and this kind of classifying should be reconsidered if laws or regulations forbid the use of different methods.

From the interviews with flexibility providers, BRPs and IAs most important aspects are the thoughts about the method’s performance, ease of use and understanding, simplicity, and costs. With the interviews also the question about who’s responsibility it is to create the baselines and should a neutral party give the framework for the baseline calculations.

7. CONCLUSIONS

In this chapter, the results and studies done in this thesis are summarized. The research questions of this work are presented and answered, and the discussion is summarized.

What are the verifying methods that can be implemented in this case examination?

The case examination in this thesis was done to Kampusareena's data from 1.6.2020 to 31.5.2021. The methods that were able to be tested in this thesis were *X of Y* and *Comparable day* methods. Other methods, such as *Regression*, could not be used, as the load did not show to be weather dependent. Also, the difficulty of calculating the *Regression* method was a barrier when choosing which methods should be tested. *Maximum Base Load*, *Meter Before – Meter After,* and *Metering Generator Output* methods are just not used in this kind of site and were ruled out of the examination based on that. *The Baseline Type II* method was not available for the test, as it would have needed some other sites metering as well.

Methods showed good performance when taking into consideration the data available. Also, considering the scale of the test done, which is small, results should be used as a reference to further work and more extensive examinations should be done to achieve more universal results. Based on the tests done in this thesis, cannot be concluded which methods should be used in specific sites and situations. One method over another could not be determined, but adjusted methods did perform better than non-adjusted ones. Based on this adjusted method is recommended to use if the baseline is wanted to be calculated.

What are the factors that make verifying methods usable at the TSO level?

Qualifications and factors that make a good baseline method are accuracy, simplicity, integrity, costs, and ease. These factors were evaluated in the thesis and in addition to these, the importance of scalability should be noted. The same methods that work for verifying flexibility at the DSO level do work also for the TSO level. More important questions considering TSOs part on the verifying methods is that is the TSO responsible for creating the baselines for the flexibility providers. TSO is neutral party considering the flexibility events, but would it be more efficient for example aggregator to provide accurate baselines. And TSO is not the party responsible for the specific baselines, should TSO give a framework for participants for choosing the baseline?

As a result, from the tests and study in this thesis, the conclusion is that choosing one method for all specific flexibility resources and situations is hard. There should be further examination with other methods, and with real-life flexibility events so that more information about their performance can be collected. The opinions of BRP, aggregators, DSO, and IA should be collected and a settlement model for IA should be chosen before a corresponding verifying method can be chosen.

REFERENCES

[1] P. Seljom, E. Rosenberg, A. Fidje, J. E. Haugen, M. Meir, J. Rekstad and T. Jarlset, "Modelling the effects of climate change on the energy system - A case study of Norway," *Energy policy,* vol. 39, pp. 7310-7321, 2011.

[2] "Clean Energy for all Europeans," European Union, Luxembourg, 2019.

[3] K. Franke, F. Sensfuß, C. Bernath and B. Lux, "Carbon-neutral energy systems and the importance of flexibility options: A case study in China," *Computer & industrial engineering,* vol. 162, p. 107712, 2021.

[4] “Fingrid Introduction,” Fingrid Oyj, [Online]. Available: https://www.fingrid.fi/en/pages/company/in-brief/#4e5e4c6d. [Accessed 11 November 2021].

[5] O. Badatunde, J. Munda and Y. Hamam, "Power system flexibility: A review," Elsevier Ltd, Okinawa, Japan, 2019.

[6] J. Cronin, G. Anandarajah and O. Dessens, "Climate change impacts on the energy system: a review of trends and gaps," *Climatic change,* vol. 151, pp. 79-93, 2018.

[7] L. Shinn, "Renewable Energy: The Clean Facts," Natural Resources Defence Council, 15 June 2018. [Online]. Available: https://www.nrdc.org/stories/renewable-energy-clean-facts#sec-whatis. [Accessed 11 November 2021].

[8] L. J. De Vries and R. A. Verzijlbergh, "How Renewable Energy is Reshaping Europe's Electricity Market Design," *Economics of Energy and Environmental Policy,* vol. 7, pp. 31-49, 2018.

[9] IRENA, “Power System Flexibility for the Energy Transition, Part 1: Overview for policy makers,” International Renewable Energy Agency, Abu Dhabi, 2018.

[10] EnerNOC, "The Demand Response Baseline," EnerNOC, 2011.

[11] M. Z. Degefa, I. B. Sperstad and H. Sæle, "Comprehensive classification and characterizations of power system flexibility resources," Elsevier B.V, Trondheim, Norway, 2021.

[12] Fingrid Oyj, "Fingrid Verkkovisio," January 2021. [Online]. Available: https://www.fingrid.fi/kantaverkko/kehittaminen/verkkovisio/. [Accessed 15 September 2021].

[13] Energiateollisuus, "Energiavuosi 2020, Sähkö," 4 2 2021. [Online]. Available: https://energia.fi/files/4428/Sahkovuosi_2020_netti.pdf. [Accessed 6 5 2021].

[14] A. Osmani, J. Zhang, V. Gonela and I. Awusu, "Electricity generation from renewables in the United States: Resource potential, current usage, technical status, challenges, strategies, policies, and future directions," *Renewable and sustainable energy reviews,* vol. 24, pp. 454-472, 2013.

[15] A. Akrami, M. Doostizadeh and F. Aminifar, "Power system flexibility: an overview of emergence to evolution," *Journal of modern power systems and clean energy,* vol. 7, pp. 987-1007, 2019.

[16] M. R. Cruz, D. Z. Fitiwi, S. F. Santos and J. P. Catalão, "A comprehensive survey of flexibility options for supporting the low-carbon energy future," *Renewable and Sustainable Energy Reviews,* vol. 97, pp. 338-353, 2018.

[17] P. D. Lund, J. Lindgren, J. Mikkola and J. Salpakari, "Review of energy system flexibility measures to enable high levels of variable renewable electricity," *Renewable and sustainable energy reviews,* vol. 45, pp. 785-807, 2015.

[18] “4 Differences between Demand Side Management and Demand Response,” RESPOND, [Online]. Available: http://project-respond.eu/4-differences-between-demand-side-management-demand-response/. [Accessed 3 November 2021].

[19] F. Boshell and O. Veloza, “Review of developed demand side management programs including different concepts and their results,” in *IEEE/PES Transmission and Distribution Conference and Exposition: Latin America*, 2008.

[20] A. Fattahi Meyabadi and M. Deihimi, "A review of demand-side management: Reconsidering theoretical framework," *Renewable and Sustainable Energy Reviews,* vol. 80, pp. 367-379, 2017.

[21] C. W. Gellings, "The Concept of Demand-Side Management for Electric Utilities," *Proceedings of the IEEE,* vol. 73, pp. 1468-1470, 1985.

[22] C. W. Gellings, W. Barron, F. M. Betley, W. A. England, L. L. Preiss and D. E. Jones, “Integrating Demand-Side Management into Utility Planning,” *IEEE Transactions on power systems,* vol. 1, pp. 81-87, 1986.

[23] P. Hermans, L. van Rensburg, M. Lallemand, G. De Jong, A. Pinto-Bello and H. de Heer, “Demand Side Flexibility: Percieved barriers and proposed recommendations,” European Smart Grids Task Force, Expert Group 3, 2019.

[24] M. Albadi and E. El-Saadany, "A summary of demand response in electricity markets," *Electric Power System Research,* vol. 78, pp. 1989-1996, 2008.

[25] M. Alizadeh, M. Parsa Moghaddam, N. Amjady, P. Siano and M. Sheikh-El-Eslami, "Flexibility in future power systems with high renewable penetration: A review," *Renewables and sustainable energy reviews,* vol. 57, pp. 1186-1193, 2016.

[26] eSett, "Nordic Imbalance Settlement Handbook," eSett, 2021.

[27] Nordic Imbalance Settlement Handbook: Introductions and Rules for Market Participants after adapting of Single Balance Model, Helsinki: eSett Oy, 1.10.2021.

[28] "What is Datahub?," Fingrid Oyj, [Online]. Available: https://www.fingrid.fi/en/electricity-market/datahub/. [Accessed 17 November 2021].

[29] "Nord Pool, about us," Nord Pool, [Online]. Available: https://www.nordpoolgroup.com/About-us/. [Accessed 17 November 2021].

[30] "Common Rules for the Internal Market for Electricity and Amending Directive, 2019/944," 14 June 2019. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019L0944.

[31] IRENA, "Innovation landscape brief: Aggregators," International Renewable Energy Agency, Abu Dhabi, 2019.

[32] “Nordic Regulatory Framework for Independent Aggregation,” Nordic Energy Regulators, 2020.

[33] “Balance Service Provider Model,” eSett, [Online]. Available: https://www.esett.com/projects/balance-service-provider-model/. [Accessed 29 October 2021].

[34] Pöyry, “Independent Aggregator Models,” Pöyry Management Consulting, 2018.

[35] N. Stevens, K. Merckx, P. Crucifix, I. Gómez, M. Sanots-Mugica, Á. Díaz, A. Sanjab, K. Kessels, E. Rivero, Y. Mou, J. P. C. Ávila, L. Lind, L. O. Camancho and O. Valarezo, "Markets for DSO and TSO procurement of innovative grid services: Specification of the architecture, operation and clearing algorithms," CoordiNet, 2021.

[36] J. W. Taylor, L. M. de Menezes and P. E. McSharry, "A comparison of univariate methods for forecasting electricity demand up to a day ahead," *International journal of forecasting,* vol. 22, pp. 1-16, 2006.

[37] J.-R. Won, Y.-B. Yoon, J.-H. Choi and K.-K. Yi, "Case study of customer baseline (CBL) load determination in Korea based on statistical method," in *Transmission and Distribution Conference and Exposition: Asia and Pacific*, Seoul, Korea (South), 2009.

[38] DNV KEMA, "Development of Demand Response Mechanism, Baseline Consumption Methodology - Phase 1 Results," Australian Energy Market Operator, 2013.

[39] K. Coughlin, M. A. Piette, C. Goldman and S. Kiliccote, "Estimating Demand response Load Impacts: Evaluation of Baseline Load Models for Non-Residental Buildings in California," eScholarship, University of California, 2008.

[40] A. Hilden, "Power Quality and Power Monitoring in a Modern Office Building Utilizing Diverse Metering," 2018.

[41] "Laatuvahti 3 introduction," [Online]. Available: https://de.cdn-website.com/7369e24fb176447094a81ef49b4a1710/files/uploaded/eQL_Laatuvahti3.pdf. [Accessed 31 October 2021].

[42] "PM3200 series User Manual," [Online]. Available: https://www.se.com/ww/en/download/document/DOCA0006EN/. [Accessed 11 November 2021].

[43] "PowerLogic Series 800 Power Meter PM820, PM850, PM870 Reference Manual," [Online]. Available: https://cdn.chipkin.com/files/2018/Chariot/ModbusPointsLists/PM820_PM850_PM870_Reference_Manual.pdf. [Accessed 11 November 2021].

[44] M. L. Goldberg, K. Agnew and M. Messenger, "Development of a Standard Baseline Calculation Protocol for Demand Response," in *Energy Program Evaluation Conference*, Seattle, 2003.

[45] S. Park, S. Ruy, Y. Choi, J. Kim and H. Kim, "Data-Driven Baseline Estimation of Residential Buildings for Demand Response," *Energies,* vol. 8, pp. 10239-10259, 2015.

[46] G. Li, C.-C. Liu, C. Mattson and J. Lawarrée, "Day-Ahead Electricity Price Forecasting in a Grid Environment," *IEEE Transactions on Power Systems,* vol. 22, pp. 266-274, 2007.

[47] "An integrated approach to Active System Management," CEDEC, E.DSO, Eurelectric, GEODE, ENTSO-E, Berlin, 2019.

[48] sthlmflex, "Onboarding-information för flexleverantörer på sthlmflex," 29 October 2021. [Online]. Available: https://www.svk.se/contentassets/47c91e83919c4ffab23ce89493fa2d83/onboarding-info-sthlmflex-21-10-29.pdf. [Accessed 31 October 2021].