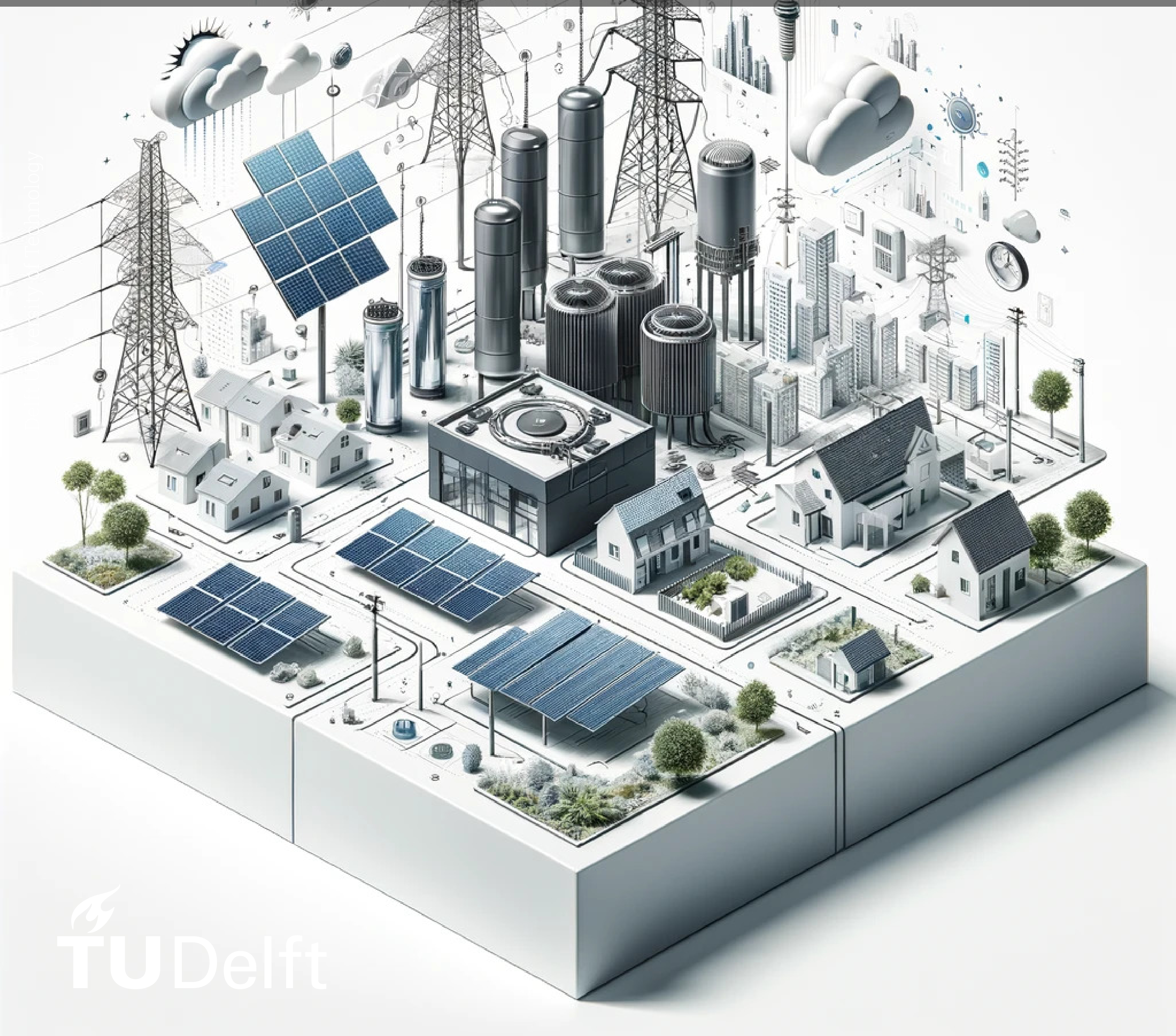
Estimating Flexibility of Distribution Systems by Identifying Assets Based on Aggregated Transformer Measurements

Flexibility to Solve Congestion in Grid Infrastructure

I. Cassabois



TU Delft

Estimating Flexibility of Distribution Systems by Identifying Assets Based on Aggregated Transformer Measurements

Flexibility to Solve Congestion in Grid Infrastructure

by

I. Cassabois

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Thursday, August 29, 2024 at 13:00.

Student number:	5837545	
Project duration:	December 01, 2023 – August 29, 2024	
Thesis committee:	Dr. M. Cvetkovic	TU Delft
	Dr. P. P. Vergara Barrios	TU Delft
	Dr. L. Elizondo Ramirez	TU Delft
	Ir. W. Zomerdijk	TU Delft

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

Cover: Image generated by DALL-E

TU Delft

Preface

I. Cassabois
Delft, August 2024

This master's thesis marks the end of my studies at Delft University of Technology. Throughout this period, I have experienced a memorable journey. I have met people who will probably remain in my life for a long time. I have also challenged myself with the various paradoxes of renewable energy and its technologies. These learnings have given me even more determination to tackle the upcoming challenges our society will face in the future. I am looking forward to the future and how I will be able to apply what I have learned, thanks to the Sustainable Energy Technology master's degree.

First of all, I would like to thank my supervisors for the support they provided in helping me conduct this research work. Without their advice and the weekly meetings that were organized, I would never have been able to complete this thesis. I would like to express my gratitude to my daily supervisor, Wouter, for your assistance during all the moments when I doubted whether the model would function as we intended. You were always available to address the endless questions and remarks I had. This work would not have been possible without you. I would also like to thank Dr. P.P. Vergara Barrios. You consistently raised important issues and provided encouragement throughout this research. I have learned a great deal from your guidance, especially in how one should critically question their own work.

Looking back at the environment that surrounded me during this work, I would like to thank several people. First of all, thank you, Elena, for being there for me during all the periods of doubt I experienced, and for helping me tackle the questions that kept running through my mind. This work would not have been possible without your support and encouragement. Thank you for all your love, positivity, and feedback. I would also like to thank Jeanne for always being by my side this year as we both worked on our respective master's theses. Whether it was sharing a coffee or providing feedback on how to frame my work, I am very grateful. Additionally, I thank my parents for always being available to address the little questions I had and for all the energy they provided throughout this thesis. Thank you, Milou, for the endless refills of coffee that helped me get through the writing and thinking process! Thank you, Patrick, Daan, Mathis, Guiomar, Glo, and all the friends who have surrounded me over these two years.

Abstract

In the current grid infrastructure, the power flow reaching the limits of the grid capacity induces several problems. A way to circumvent this issue is the introduction of flexibility in order to alleviate the impact on the energy network. However, the estimation of the required flexibility remains a problem in the current state of the art. In this thesis, an approach is proposed to estimate the flexibility of distribution systems by identifying assets based on aggregated transformer measurements. Flexibility can be seen as the ability of the system to adapt to changes coming from demand and generation. Different assets can be available such as solar panels, and heat pumps and have their specificities. The problem thus translates into the identification process in aggregated data. Several steps are important in the process, such as the consumption baseline, the study of the different characteristics of the profiles, and the number of iterations. To allow the accuracy to reach high levels and to match the standards of distribution system operators (DSO), the process requires the minimization of squared error between the data and the chosen profiles. Different cases are investigated and validated prior to the study of the results. For the identification of heat pumps and solar panels, individual approaches have been investigated in the literature. However, the combined identification of assets has not yet been studied. A mixed identification process is proposed in this work, and the method has been tested, reaching a high level of accuracy. Within this new framework aiming at separating the assets, options are proposed to enhance the accuracy of the process.

Contents

List of Figures

List of Tables

1 Introduction

This chapter provides the reader with a brief overview of the work covered in this thesis. In Sec. 1.1 the section provides information on the problem as well as the context in which it is taking place. Sec. 1.2 will introduce the problem that is addressed throughout the research covered in this document. Sec. 1.3 discusses the research objectives, and presents the research gaps in the next section Sec. 1.4. Consequently, Sec. 1.5 highlights the research question, as well as its sub-questions. The research approach is then detailed in Sec. 1.6. Finally, Sec. 1.7 addresses the outline of the content of this document.

1.1. Background

The escalating concerns surrounding climate change and the imperative for decarbonization have started a global movement toward the transition to a system based on renewable energy. Such a tendency aims to reduce the emission of greenhouse gas(GHG) and to keep the system reliable as the number of assets connected to the grid will increase in the power distribution facilities. The number of renewable energy sources has consequently multiplied over the last decade. This growth raises the problem of the limitation of capacity that can be connected to the grid. In addition to this issue comes the integration of renewable energy into the power grid, the design of the infrastructure is therefore questioned as it has only been built to act as a single-direction system. The introduction of variable renewable energy resources is then provoking an influx into the power grid this adds up another direction of power flow to the infrastructures. Challenges arise to balance the power flow in the grid as the variability of these sources is high. These problems are therefore becoming an important issue as global electricity demand rises due to electrification of different sectors such as heating and transportation. This electrification will grant different advantages in the future, such as an increase in its reliability. Achieving a high level of electrification will lead to a decrease in the dependency of energy on other countries. Hence, the price of electricity would be more stable as the different sources would be mainly located in the country. Such a process would also guarantee utilities and politics a more stable economic climate [1] as well as emitting less GHG. However, the growth of both variable renewable energy sources (VRES) and the energy demand creates a conflict in the grid, the limit of the power that can be transported in the grid has been reached. Such results create a phenomenon called congestion :

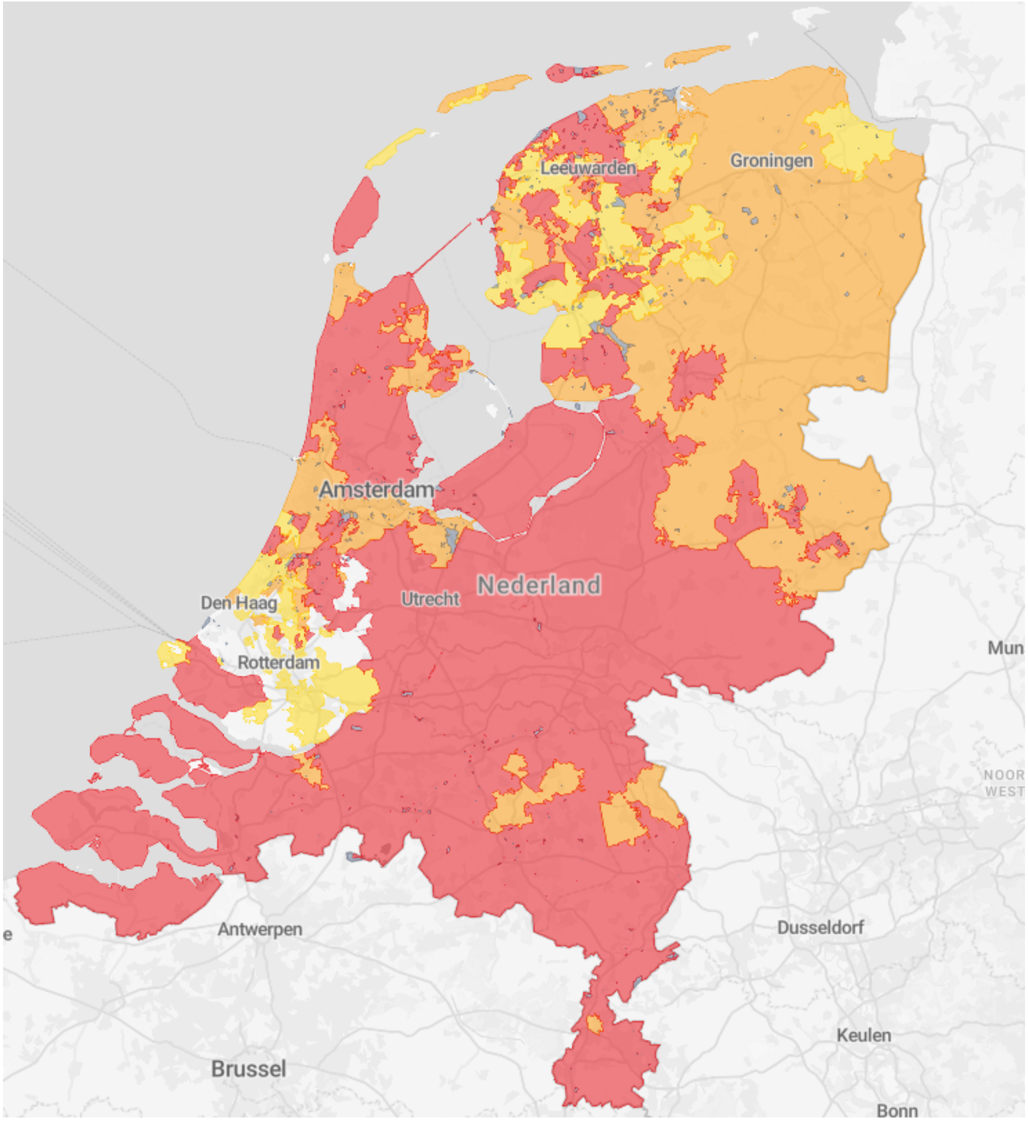


Figure 1.1: Congestion in the Netherlands, Source [2]

Congestion is a phenomenon that happens when transmission and distribution infrastructures are unable to accommodate all desired transactions according to [3], it can also happen due to the damages that the infrastructures may have had to go through due to previous cases of congestion. As one may see in Fig. 1.1 the Netherlands is facing congestion in almost the entire country. The risk of congestion rises with renewable energy integration, urging timely solutions to prevent global energy grid blackouts. The usage of newly penetrating loads is not distributed evenly across the day which creates an overload in the infrastructure [4]. Hence, the imperative to reduce congestion becomes urgent, ensuring the integration of renewable energy as well as improving the reliability of power grids to secure energy distribution. As policies around the globe arise in favour of the development of renewable energy and contribute to the increase of the electrification [5, 6, 7, 8], the congestion is appearing as a bottleneck of the renewable transition. The consequences are far-reaching, leading to voltage deviations and, in the most extreme cases, triggering grid blackouts. Such events not only pose immediate threats to grid stability but also result in substantial additional costs for solving the problem. These current grid problems, amplified by outages and increasing demand [9], necessitate fossil fuel backup generators [10]. Therefore not aligning with the Paris Agreement as the resultant GHG emissions underscore the environmental impact. In addition approximately 1% of global carbon dioxide (CO2) emissions in Germany are attributed to congestion-related problems, posing a challenge to the Paris Agreement's emission reduction goals. Considering the case of Germany, congestion solutions cost 2.25 billion € in 2019 [11].

When looking at a global scale, costs related to congestion could reach 1.3 trillion, particularly affecting developing countries [9], as their electricity demand will increase tremendously. Congestion will then take place in society as a whole, the effect of it will not be discriminatory between companies, households, and utilities. When looking at the grid operator perspective, congestion can happen on every scale and then create a ripple effect for an entire country or even continent. Such as in 2003 when 50 million people suffered the consequences of the outage for up to 4 days [12].

Therefore congestion events should be avoided at any cost, different ways already exist to tackle them, and future expansions of the grid infrastructures could take place to increase the capacity of the grid. Such measures would then not only require a tremendous amount of money to be invested to overcome the physical boundaries but also a consequent period to be dedicated to upscale the power grid. This solution is the most efficient, however it can only be pursued on the long term. Other mechanisms exist to address immediately the problem of congestion such as the use of flexibility. The IEA [5] underlines that decentralized generation combined with flexible loads and storage systems can unlock opportunities for reducing peak demand and help to manage congestion. Numerous approaches have been designed to manage and alleviate congestion within power grids [13]. Among these, the focus of this thesis centres on flexibility as a prominent and effective solution.

1.2. Problem Definition

Flexibility can be defined as the capability of a power system to cope with the variability and uncertainty that solar and wind energy introduce at different time scales, from the very short to the long term, avoiding curtailment of power from these VRES and reliably supplying all customer energy demand [14]. This solution raises a particular interest due to its capacity to dynamically adapt to the variations of demands as well as power generation [15, 16, 17, 8]. Flexibility can be defined then according to the different utilities that are already existing in the grid. While it holds significant promise in addressing congestion, its widespread implementation is not without challenges. Technical limitations, regulatory complexities, and the need for comprehensive coordination among stakeholders present hurdles that demand careful consideration. An in-depth exploration of these challenges is crucial to developing strategies for their effective mitigation, thereby paving the way for the successful integration of flexibility into power grid operations. The implementation of such solutions depends on the available information in the grid about the different utilities that define the electrical profiles. Without a precise insight into the grid components, the estimation of flexibility can be laborious. Flexibility can require the different characteristics of different units [18] such as active, and reactive power. In case of a deficiency of data, initiatives regarding the estimation of flexibility shall be done to help the distribution systems operators and the transmission system operators (DSO-TSO) handle the congestion events. With the increasing number of distributed energy resources (DERs), it becomes harder and harder to locate the production and to identify the different types of utilities in the grid. As installing smart meters in every neighbourhood to keep track of the production and consumption of data would require massive funds, the need to develop a method to estimate the hidden capacity of flexibility is becoming prominent.

Nevertheless, the lack of data remains and questions arise around the estimation of flexibility. How can flexibility be estimated based on aggregated data? How can it be defined in that case? Shall we refer to different profiles as flexibility assets? All of these questions raise the problems that the energy transition is facing in the electrification of the grid, these problems are emerging as VRES are being incorporated into the grid. How can the estimation of flexibility be computed? What methods exist in this case? The estimation of flexibility is based on different methods concerning the different types of data that are available for the DSO. However with limited information as we look into aggregated data, which data are valuable for the estimation of flexibility? Which utilities can still be identified?

Different energy sources can be the source of flexibility, [19] refers to renewable energy sources as a source of flexibility, wind energy and solar energy are particularly interesting as they have a peak of production and thereby can be identified easier than other type of production. On the other side, the assets of energy consumption can be the consumption of power of heat pumps. As heat pump consumption has peaks it is therefore easily identified.

1.3. Research Objective

The objective of the thesis is to develop a model capable of identifying the flexible assets within the grid. Subsequently, this model will enable the creation of case studies based on collected data, facilitating analysis of the results. Following this process, a sensitivity analysis will be conducted to estimate the amount of flexibility available in the grid based on aggregated data. Leveraging the identified flexible assets, various strategies for enhancing flexibility will be proposed with the aim of minimizing congestion within the neighbourhood. This approach is intended to mitigate the social-economic welfare impacts on neighbourhood scales as much as possible.

1.4. Research Gaps

In current literature, various methods are employed to estimate flexibility within power grids. These methods include optimizing the reactive-active power of assets and delineating flexibility regions. However, existing approaches predominantly focus on identifying individual assets, as emphasized by [20]. Notably, there is a notable gap in research concerning the estimation of flexibility based solely on aggregated power measurements, without detailed asset information. To effectively estimate flexibility, it is essential to understand the characteristics of assets within the grid. As the different profiles are all added up together, the peak and characteristics are all hindering each other. Consequently, flexibility can be assessed accordingly. However, the relationship between the growth of solar panels and heat pumps and their impact on flexibility remains largely unexplored in the literature. To determine the amount of flexibility available in the grid, accurate profiles for both heat pumps and solar panels need to be established. Therefore the need to provide accurate heat pump profiles is crucial. However, within the literature, significant deviations have been observed from the modelled behaviour and the measurements profiles [21].

1.5. Research Questions

This thesis aims to provide an accurate estimation of the flexibility on a distribution level based on aggregated data provided by transformer measurements. Once the flexibility is estimated, the goal will be to provide advice for the implementation of such solutions to profit from the socio-economic welfare of the utilities in the neighbourhood. The research question can be formulated as :

”How can the identification of assets based on aggregated transformers’ measurements provide an estimation of the available flexibility in the distribution system?”

Following the research gaps, different sub-research questions have risen :

1. How can we estimate more accurate heat pump profiles?
2. Which data and steps are needed for disaggregating the profiles with a 90% confidence interval?
3. How can we minimize the uncertainty of the proposed method?
4. What is the correlation between the heat consumption profile and the PV production?

1.6. Research Approach

The research approach will consist of first testing the identification of the PV and HP separately to provide accurate results. Then once the requirements for moving onto the second step have been fulfilled, a model that can combine both of the identification processes will be created. The procedure will be to first identify assets without noise and then implement different levels of noise based on the prevalent asset in the data. For each of the models, the first step will always be to remove the baseline consumption profile from the data and then use the identification algorithms. Optimization of the squared error is the mathematical model that has been chosen for all of the identification. The scenarios will differ in the number of assets implemented and the different steps chosen to optimize. Then the results will be visualized throughout a boxplot graph to assess the accuracy of the methods that have been developed. With respect to the results, an estimation of the reduction of the peak load will be pursued to grand insights into how much could be done by curtailing or using other DR mechanisms.

1.7. Outline

The structure of the thesis is as follows. Chapter 2 provides background knowledge about the use of flexibility in the context of congestion happening across the grid in the Netherlands. A literature review is presented about the different flexibility assets and methods to identify the potential of flexibility in the perspective to establish the framework of the thesis. The research gaps can be found at the end of this review. Then Chapter 3 presents the different steps that have been pursued to deliver a model that can identify the utilities. The different steps and models will be explained to legitimate the final algorithm proposed. At the end of this chapter, the different scenarios and the KPI are detailed. Once the model, scenarios and methods for evaluation are presented, the validation of the model can be found in Chapter 4. Then once the validation is presented, the different scenarios are displayed in Chapter 5. Chapter 6 provides insights into the different important parameters and a sensitivity analysis is detailed. Finally, Chapter 7 concludes on the work that has been covered in this research and discusses the different limitations of the model as well as future work.

2

Literature review

To ensure that this thesis is viewed as a comprehensive work concerning the literature review, this chapter will detail the important aspects of the topic which will give the keys to understanding the current situation. The state of the art of the grid will be the starting point of this work, thereby making congestion management the first focus of this work in Sec. 2.1. Concerning Sec. 2.1 one of the solutions considered to tackle congestion is flexibility, the estimation of it regarding the different flows of power with aggregated data in the grid will be detailed in Sec. 2.2. Sec. 2.3 will give insights into how this flexibility estimation can take place as well as with the methods used to estimate the flexibility. Lastly, the kind of data that will be required and used for the model will be developed in Sec. 2.4.

2.1. Congestion Management

To address congestion issues, various methods can be employed, including both technical and non-technical approaches that involve market-based operations. Methods such as Generation Rescheduling (GR) [22, 23, 24, 25, 26], load shedding [23, 26], Distributed Generation (DG) [27, 28, 29, 30, 19, 31], and Demand Response (DR) [29, 32, 30, 33, 34, 35] are commonly used. Among these methods, dispatching DG in territories to minimize generation costs stands out. However, implementing GR requires several steps, including selecting generators to be rescheduled. Traditional optimization problems focus on minimizing generation and reducing load operational costs, but this approach is slow and ineffective [22]. Load shedding involves curtailing a building's load as quickly as possible in response to an urgent request from the grid operator. However, it can only be applied in emergencies and is not suitable for congestion management as a reserve capacity [29]. Distributed Generation is effective in minimizing power losses and voltage instability but has limitations in congestion management due to implementation challenges. Combining Flexible Alternating Current Transformer Systems (FACTS) with DR has been identified as a reliable and efficient solution for relieving congestion [33]. Different optimization problems have been formulated to minimize power flows in the grid. The Unified Power Flow Controller (UPFC) is widely used for congestion management, optimizing based on location to minimize generation costs and ease congestion [3]. Active power rescheduling, notably improving system flexibility, is highlighted as the most effective way to remove congestion, particularly in the context of the renewable transition [13].

2.1.1. Flexibility an answer to congestion

As the IEA has underlined the lack of flexibility [9] in the current systems, it has a direct incident on the increase of the risk profile for transmission and distribution systems. The need to implement flexibility is becoming rapidly a primordial issue to answer to the tremendous growth of electricity volumes produced by renewable energy sources [9, 8, 32]. Flexibility can be defined according to [36] as "*the ability of a power system to respond to changes in electricity demand and generation*". This ability is therefore dependent on the current state of the grid and the physical infrastructures available for the renewable transition. Flexibility has major advantages, it can be implemented in both demand and supply-side technologies [37]. On-site generations can be considered as one from both sides and can alleviate the congestion from the transmission power grid according to to [29], a significant amount of variable renewable energy sources (VRES) however requires a broader choice of flexibility options [32].

Governments around the world are planning to expand the renewable generation capacity [6], the European Commission plans to raise the share of renewables capacity in Europe to 42.5% by 2030 and as in the Netherlands the share of renewable energy is expected to reach 16% in 2023 whereas in 2020 it was 14% [7]. As VRES are expanding rapidly, the variability of its production increases as well, the investigation of the coupling of VRES and heating systems has been led in [32] to tackle these fluctuations. Demand supply management coupled with VRE can on average reduce by 20% the costs. Heating systems can be considered as flexible assets [38, 39, 15], and the integration of such assets requires the coupling to thermal storage to reveal the potential of these utilities. The electrification of the heating system throughout the excess of production from VRES can open the path to electrify the heating [15, 29]. Thus, the need to couple on-site generations and other assets is becoming urgent to unlock the flexibility that is yet remaining underdeveloped. This combination is not the only one, different options also exist as investigated in [32] where the mix of the vehicle to grid(V2G) and flexible demand is also considered. This enables several assets with considerable advantages such as reserve power, regulation, and emergency load curtailment [15]. The option to convert power to heat throughout district heating systems can also be interesting in the case of the power conversion to maximize the potential of the electricity produced [32, 39].

2.1.2. How can flexibility mechanims reduce congestion ?

According to [13] the rescheduling of generation is one of the most efficient ways to handle congestion and the most used technique for congestion management. This can be translated into the use of a demand response process to shift load when the power is the least needed. The integration of variable renewable energies into power grids would grant more flexibility to the grid and its actors. As mentioned in [29] the correlation between schedulable/flexible loads and controllable local generation can improve the coordination and energy management of power grids. Therefore, this means that certain patterns in the grid could be used to tackle the problem of congestion. Also, the self-consumption as underlined in [29] of the demand side users can contribute to reducing the demand on the power grid, which would then allow more flexibility to the grid. Generation can to some extent cover partially or even fully building loads. Thus, reducing the demand and therefore the stress on energy systems [40]. Flexibility has been introduced as an answer to overloading by using mechanisms such as shaving peaks [41]. According to [4] the role of the prosumer in flexibility mechanisms can become an important factor of the future energy grids in facilitating the integration of VRES.

Flexibility according to the literature seems to be a promising approach to reduce the congestion across the energy grid. Nevertheless, the term flexibility needs to be defined and quantified using different assets. The sources and capacities of the different potential options will detailed in the following section.

2.2. Flexibility Estimation

Flexibility within energy systems is crucial for adapting to dynamic changes in demand and optimizing available resources. According to [29], flexibility can be derived from the supply side. However, further analysis of grid operators' data is necessary to assess which utilities can be used to relieve the transmission infrastructure. On the demand side, certain demand response (DR) processes can highlight the flexibility of loads available on the network, potentially leading to the implementation of flexibility mechanisms. Besides, this requires information on the different utilities in the grid to estimate the available flexibility in the system. Technologies like thermal storage are often coupled with these mechanisms to counter the latency of different profiles, enabling the heat to be used at another moment during the day and saving the consumption of other devices [42].

The combination of end-user assets and associated facilities can lead to flexibility implementation [43]. Various mechanisms, including mid-term bilateral flexibility contracts, curtailment, flexibility bids, and controllable resources connected to the distribution network, can be introduced [44]. However, as highlighted, these mechanisms are information-dependent, which is a weakness of the process. The required information to introduce flexibility assets is not always available to the operator.

Different algorithms, such as the non-intrusive load monitoring(NILM) algorithm [45, 46, 37], can be found in the literature to identify flexibility. These methods provide information regarding different appliances, allowing knowledge of their state. However, they have limitations because households have different behaviours depending on the number of inhabitants, and appliances can function in various ways. Nevertheless, specific behaviours can be identified in the mix of data, and renewable energy and various heat mechanisms can be investigated as flexibility assets. Specific machines, often captured through measurements of the active power they consume [18, 45], can provide flexibility once their behaviours are identified.

Other studies take into account aggregated data to identify production profiles of solar energy [46, 47, 48]. Thus, paving the way for the identification of assets using optimization algorithms. However, these identifications usually focus on a single type of asset and do not mix different technologies. Various machines can be a source of flexibility, with heat pumps as end-user systems coupled with on-site generation being one method to grant more flexibility to the energy infrastructure. Flexibility can be viewed through different prisms. Two sides can be put apart to create two categories of flexibility, either the supply side with options such as VRE generation, energy storage and fast-response power plants, or the demand side with schedulable/flexible load, local and controllable VRE generation.

2.2.1. How to characterize flexibility ?

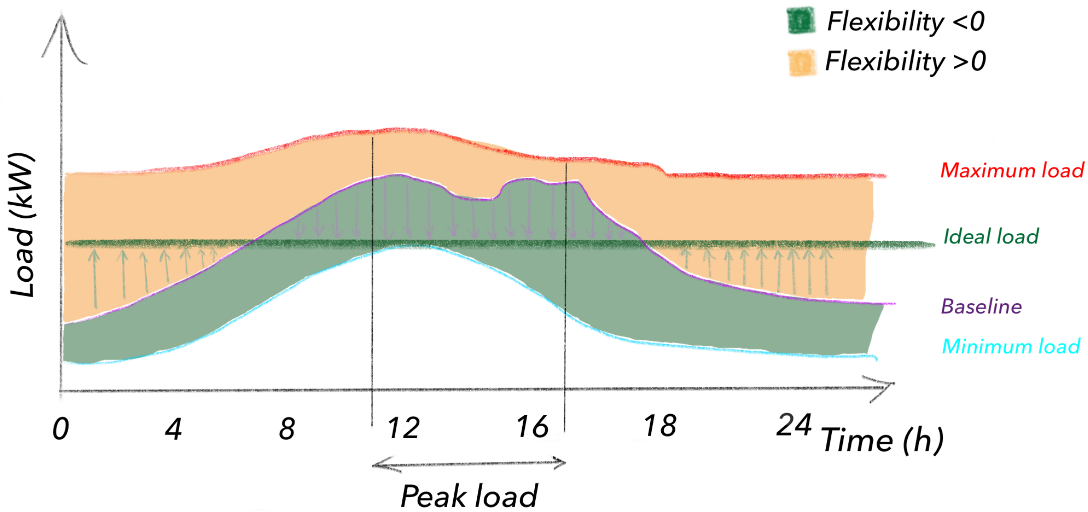


Figure 2.1: The boundaries of the flexibility region can be determined using different loads, the green line represents the load envisioned for the future. The orange region represents the area where flexibility can be developed whereas the green area is the area corresponding to non-available capacity shift.

Flexibility can be derived once the boundaries of the consumption of electricity have been established as seen in Fig.2.1. To identify it in the energy grid, the first step is to establish the baseline consumption as seen in Fig.2.2. This variable is highlighted as the inflexible energy consumption that should not be altered [37, 49, 42, 43]. The first step when flexibility needs to be quantified is to determine the minimum and maximum of the operating points of the system to set the boundaries in which the system is operating. This will allow the creation of the minimum and maximum curve set points as proposed in [37] and seen in Fig.2.1.

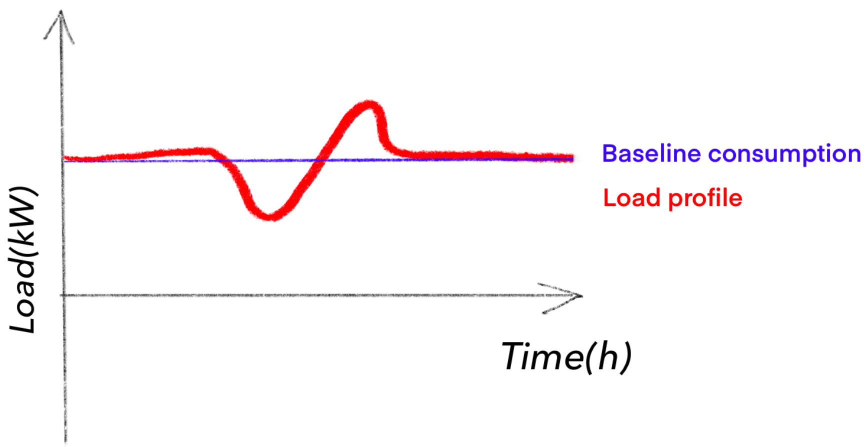


Figure 2.2: The profile represents the load profile for a day, the redline represents the load profile of a neighbourhood, and the blue line the consumption that cannot be changed

Once this is achieved, different requirements have been mentioned in [37] to estimate flexibility, because information on the different flexibility sources needs to be available to the user. Furthermore, if DR is used to approximate the available flexibility, the identification of the assets is dependent on the actions of the users. Moreover, the action needs to take place at an effective time to identify the asset.

The identification of individual assets can be a tedious task due to specific requirements. In the pursuit of approximating flexibility within a system, certain information, such as the temperature set points for heat pumps, can be highly valuable. These set-points are crucial for estimating the pattern of thermal systems, which is essential for assessing the flexibility of heat pumps [49]. Heat pumps operate with a temperature-dependent set-point, influencing their operational modes. As noted in [50], the energy demand varies significantly based on the chosen set-point. Utilizing this information, it becomes possible to determine a feasible region and estimate operating points where flexibility can be implemented. Additionally, renewable energy sources can alleviate stress on the system, potentially enabling more flexibility, especially when coupled with other utilities, as discussed in [42].

Flexibility can be defined using various metrics [19], including ramp magnitude, ramp frequency, and response time, to classify the data flow provided by smart meters and transformers from the grid. However, the choice of metrics and methods for defining flexibility can vary significantly depending on different scenarios. These metrics must be accessible to the system operator to identify various assets, which can be challenging with aggregated data. In [43], metrics are defined based on the type of system they are applied to, highlighting the need to define each system's characteristics to implement flexibility into the grid effectively. Metrics such as peak power reduction, flexibility factor, self-sufficiency, and self-consumption are used to evaluate flexibility in the system.

For heat pumps, relevant metrics include half-hourly average peak electricity demand (GW) and maximum ramp rate (GW/half hour) [51]. Additionally, flexibility can be quantified as the number of consumption hours that can be rescheduled [52]. In [18], flexibility is quantified as a function of the available reactive and active power of a system to adapt. The methods used to identify flexibility will be detailed in Sec. 2.3.

2.2.2. Flexibility savings

The relationship between the supply side and demand side can be investigated to couple solutions for both sides. According to [15], synergies in energy management systems and communication could reduce both investment and fixed costs when the end-uses are combined. Different mechanisms of load management can then lead to the reduction of congestion throughout the use of various types of assets. Self-consumption, as explained in [29], can be considered a type of flexibility through load reduction. Coupled with a solution such as a heat pump system, the load can be shifted, as discussed in [29]. For instance, in [16], load reduction can attain 42% and 61% by managing different types of domestic loads.

Other solutions, such as Vehicle-to-Grid (V2G) or photovoltaic (PV) systems, can be considered to increase the self-consumption of users in the grid [53, 29, 32]. As mentioned in [29], the potential of Power-to-Heat (P2H) for load shifting is interesting, as seen in [38], where 5.5 GW have been effectively shifted from peak loads out of 7.3 GW available for Demand Response (DR). These mechanisms have shown that different assets can introduce flexibility into the grid. However, Variable Renewable Energy Sources (VRES) can also introduce more flexibility, as seen in [32], where a variable defined as the critical excess of energy produced represents the amount of energy needed to overcome problems created by congestion issues.

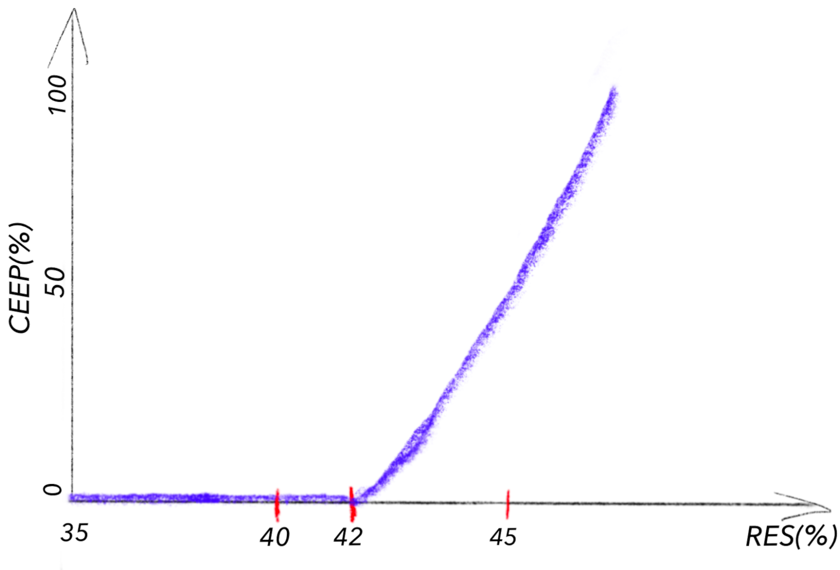


Figure 2.3: CEEP evolution implementing flexible demand, inspired by graph in [32]

Fig.2.3 highlights that with the introduction of flexible demand using RES, the critical excess of energy that needs to be produced can be reduced. About 50% of the demand is flexible, 40% of that flexible demand can be accounted for the daily fluctuations and 30% for the weekly and monthly variations. Such a study also investigates the operating cost of such a system concerning the RES integration and the flexibility index. The flexibility can then be improved by the implementation of VRES into the energy mix in the grid, it may also contribute to the minimization of the operating cost of energy as mentioned in [32].

A decrease of 10% can be seen in the operating costs of the system if the share of RES can reach 100% according to [32]. These different strategies have been the source of the reduction in the power demand but also contributed to the reduction of several costs. On a larger scale, the flexibility can be improved by investigating the relationship between temperature-dependent loads and VRE sources. In [17] 27-141% energy costs have been saved and with additional structures (PV, controllable loads and energy storage) 35% of capacity cost savings. According to [15] the effect of demand-supply management (DSM) with the use of VRE on distribution grids can reduce up to a 20-24% reduction in generator start-up cost. As underlined by Dyson in [50] the residential demand for A/C is highly correlated to the peak system load. It is then interesting to point out the effect that temperature control has on congestion, in the case of global use of heat pumps and a RES penetration level of 36-47%, savings can be done about $33 to $52 per heat pump per year as well as avoiding the congestion problems [54].

Different options have been revealed to be consequent sources of flexibility, the challenge remains in identifying them. Identifying the different assets can be done through the use of different variables such as reactive, and active power. However, cases where the data are aggregated seem to be a difficult situation. To solve these problems the different methods and conditions for finding the solutions will be exposed in the next section.

2.3. Identifying methods being used

When estimating flexibility, different options exist to solve such problems. In the literature, different methods to estimate flexibility exist: the Monte-Carlo method [44, 55, 56], different optimization methods [18, 57, 49, 32, 8, 58, 48, 47, 59] and machine learning [4, 21, 60, 37, 61, 46, 45]. In the following sections, an insight on each the different method will be given on the different pros and cons of each method, as well as to what extent can it be used. An overview of the different methods available in the literature can be found in Fig.2.4.

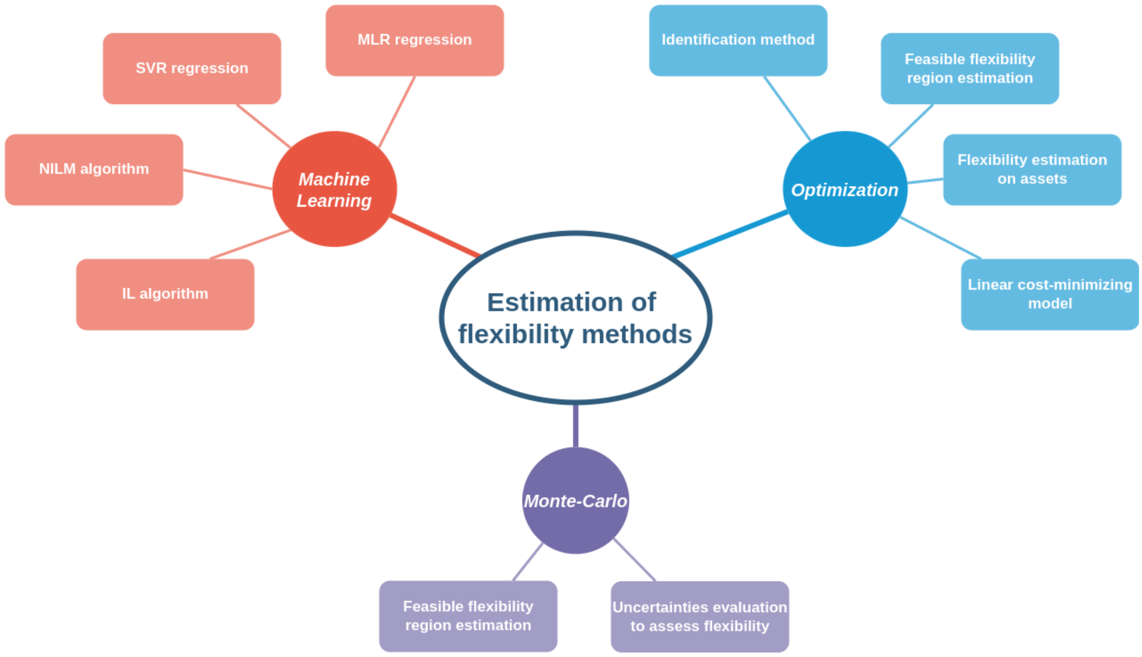


Figure 2.4: Different prism can be viewed for estimating flexibility, optimization, machine learning and Monte-Carlo.

Using the Monte-Carlo method, two ways exist such as the estimation of the feasible flexibility region and the uncertainty evaluation for the assessment of flexibility. As for Machine learning, support vector regression(SVR) and multiple linear regression(MLR) are considered to be good options for approximate flexibility. Other data-intensive algorithms are used such as non-intrusive load monitoring(NILM) and intrusive load monitoring(IL). Optimization problems are also formulated through different methods such as identification, estimation of the feasible flexibility region, estimation of flexibility based on assets and linear cost-minimizing model.

2.3.1. Optimization methods

To identify the different profiles, optimization problems can be formulated to approximate the quantities of such assets [48, 62, 63, 47, 59]. These different articles all identify flexible assets to provide an accurate estimation of the amount of solar panels. The starting point in these articles is that the aggregated power data is formulated as follows:

$$P_{sub}(t) = P_{load}(t) - P_{PV}(t), \forall t \epsilon T \tag{2.1}$$

P_{sub} stands for the power flow from the aggregated data from the transformer measurements, P_{load} is the load from the consumer from the inhabitants and P_{PV} is the power produced from the solar panels. Therefore, identifying the solar profiles can be tedious concerning Eq.2.1. Various techniques have emerged in the literature for disaggregating solar profiles from aggregated data [48, 47, 59]. Each method has specific requirements to obtain the measurements needed for estimating the solar profile.

In [48], two methods for identifying solar profiles are presented, both based on disaggregating solar profiles from aggregated data. The first method uses a linear estimator applied to multiple linear regression to minimize the square errors of the coefficients.

$$P_{PMU}(t) = k_{eff}Q_{PMU}(t) + C_{eff}\phi(t) + R + \epsilon_{PV}(t) + \epsilon_{load}(t) \tag{2.2}$$

Eq.2.2 is the formula to be estimated throughout the ordinary least squares (OLS). $P_{PMU}(t)$ represents the power measurements at the phase measurements sensor units(PMU), $Q_{PMU}(t)$ represents the reactive power measurements from the PMU, $\phi(t)$ is the irradiance, R is a resistive load and the two $\epsilon_{PV}(t)$,$\epsilon_{Load}(t)$ are the noise introduced by respectively solar and the load. The terms C_{eff} and k_{eff} are the terms that are optimized. In contrast, the second method uses contextually supervised generation estimation to find optimal model coefficients that fit the signal closely.

$$\begin{aligned} \min_{Y_i,\Theta_i} & (l_i(Y_i - X_i\Theta) + \eta_i g_i(Y_i) + \gamma_i h_i(\Theta_i)) \\ & \text{s.t.} Y_{agg} = \textstyle\sum_{i=0}^{L} Y_i \end{aligned} \tag{2.3}$$

Eq.2.3 refers to the process of the contextually supervised generation where Y_i is an unknown signal that combines both load and PV generation,l_i refers to a function that penalizes the difference between the reconstructed signal and the linear model, h_i is to counter the overfitting behaviour of the model and g_i represents the penalty function that determines the details of the signal to enhance smoothness. Both methods yield good results but require micro-phasor measurement units, adding constraints for the DSO. These methods can not be applied on a neighbourhood scale as PMUs are not connected to the distribution network. Similarly, in [47], the disaggregation of Photovoltaic (PV) generation occurs at the neighbourhood scale. The method aims to first establish a monthly profile of solar production, then extend it to an hourly format, and finally formulate an optimization problem. The objective is to maximize an hourly-resolution constrained maximum likelihood estimate (MLE) procedure based on weights allocated to each solar exemplar.

$$\begin{aligned} \max_{\omega} \; & (\textstyle\sum_{i=1}^{M} \ln(f(P_{m,n}(i), P'_{m,d}(i), G^E_m(i)|\omega))) - \frac{1}{2}\lambda||\beta||_2^2 \\ \text{s.t.} \; & (\omega^T G^E_h)^T \leq 0, \\ & P_h - (\omega^T G^E_h)^T \geq \beta, \\ & \beta \leq 0, \end{aligned} \tag{2.4}$$

Eq.2.4 corresponds to the optimization problem that is formulated in [47]. $P_{m,n}$ represents the monthly nocturnal power demand, $P'_{m,d}$ is the monthly diurnal power demand, G^E_m corresponds to the hourly PV generation exemplars, ω a weighted vector for the solar exemplars and the term $\frac{1}{2}\lambda||\beta||_2^2$ is a penalty term to maximizing the likelihood function with λ being a tuning parameter and β a non positive vector of elements. However, due to its customer-scale basis and the need for measurement units, this method cannot conduct a neighborhood-wide study. A similar strategy is examined in [59], where the objective is to minimize the sum of squared errors between the calculated and estimated solar generation.

$$\begin{aligned} arg \min_{L_{nt},S_{nt}} & \textstyle\sum_{t=0}^{T} \mu(L_{nt} - \widehat{L}_{nt})^2 + \omega(S_{nt} - \widehat{S}_{nt})^2 \\ \text{s.t.} \; & S_{nt} \geq 0, L_{nt} \geq 0, L_{nt} - S_{nt} = NL_{nt} \end{aligned} \tag{2.5}$$

Eq.2.5 refers to the process mentioned in [59]. L_{nt} represents the load, $\widehat{L}_{nt}$ the estimated load, S_{nt} the solar production,$\widehat{L}_{nt}$ the estimated solar production and ω,μ are some weight associated to the estimation of both cases. A high level of accuracy is achieved, yet the presence of outliers remains an obstacle to be overcome. Although this method appears reliable, it only breaks down one type of asset, in this case, photovoltaic generation.

Other optimization problems allow the estimation of flexibility using different algorithms. In [57], an optimization problem is formulated to minimize congestion by optimizing an Optimal Power Flow (OPF) with boundaries related to congestion. Generator rescheduling is performed to keep individual line flows within reference values at all times. The considered approach is pursued at a transmission level which makes the method inapplicable in the case of the thesis. In [41], a multi-energy flexibility measure is explored to maximize the daily profit of the supplier while considering load classification constraints and energetic boundaries of devices. The study investigates the implementation of Demand-Side Management (DSM) into the grid with load response and load recovery. However, this approach only covers flexibility estimation for one building, making it infeasible for larger-scale implementation. Besides, it requires information on each of the devices in households, making this approach data intensive. Therefore, this approach is not considered.

Similarly, in [8], a linear cost-minimizing model is formulated to minimize the total cost of electricity generation, ensuring that each region meets the demand in each time step. The study explores different DSM techniques to identify available flexibility in the system to relieve congestion, providing relations between different types of assets to unlock potential in the grid. However, this approach is not applicable as it is data intensive.

Another objective of optimization problems is to directly identify flexibility area boundaries using constraints defining the system. Optimization is pursued on the active-reactive power exchange of the grid operator [18, 64]. In [18], optimization is performed without exhaustively selecting angles, achieving earlier convergence than typical optimization problems. The goal is to determine the available flexibility region of Distributed Energy Resources (DERs) in an active distribution with fewer steps. However, this approach requires information about DERs' active and reactive powers, which can only be obtained by having meters on each DER. Conversely, in [32], optimization is conducted differently, aiming to estimate the maximum flexibility available for each option without interfering with others. This approach also requires more information about different types of assets such as reactive and active power therefore making it inapplicable.

2.3.2. Monte-Carlo method

Different cases of estimating flexibility throughout the Monte-Carlo method have arisen [44, 55, 56]. However, it has been revealed that the Monte-Carlo method requires more effort compared to optimization or machine learning. The Monte-Carlo consists of computing a feasible region based on a large sample of data. Such a method allows one to determine a region but the boundaries can differ depending on the set of data. A flexibility region can then be provided to the TSO, [44] states that this information is important to tackle events such as reverse power flows. The simulation runs a power flow scenario and establishes an operating point in the primary substation. One important point underlined by [44] is that the correlation of active and reactive power has a great impact on the p-q flexibility region established. It is already stated that Monte-Carlo has limitations due to the boundaries not being known. In [55] the Monte Carlo method allows to determine the accuracy and variations to estimate flexibility, and an optimization problem is formulated to establish the flexibility available in demand response and demand flexibility programs.

2.3.3. Machine Learning methods

Machine learning has emerged as a powerful tool for solving various problems in the past decade, improving its ability to identify relationships and process large amounts of data based on input data. It has also shown potential in disaggregating power consumption for various utilities [21]. Several machine learning techniques are utilized in the literature for congestion management, including Multiple Linear Regression (MLR), Support Vector Regression (SVR), Non-Intrusive Load Monitoring (NILM), and Interactive Learning (IL). Each of these options has specific requirements for implementation, and preprocessing of data is essential before applying machine learning algorithms [4].

In [4], machine learning is used to identify the energy flexibility potential of residential distributed networks. MLR and SVR models are developed to predict daily power consumption, using three years of data from smart meters. MLR and SVR are highly versatile and offer a high degree of customization, providing greater interpretability. Although MLR and SVR yield similar results, their application can vary depending on the type of profile being processed. Furthermore, in [37], machine learning is used to prepare data for the k-means algorithm, which is employed to estimate the flexibility area for each appliance. The NILM algorithm, coupled with the random forest algorithm, achieves high accuracy in estimating flexibility in the grid. Additionally, the IL algorithm is effective for lower consumption houses, allowing for a reduction in the sample time step to 10 minutes.

The NILM algorithm is utilized in various contexts, such as predicting and estimating power consumption of devices [21]. It disaggregates power consumption by appliances, enabling the identification of specific profiles. Different metrics have been defined for evaluating the algorithm's performance, with the total power change metric being the most precise. The NILM algorithm, particularly when paired with the random forest regression model, outperforms other algorithms [21]. Various uses of the NILM algorithm can be found in the literature, from disaggregating heat pump profiles [21] to solar profiles [45, 46, 48, 65].

While machine learning algorithms like NILM and IL are robust, they require data provided by TSOs or DSOs to solve the problem. Besides, these high-performance methods often require a broad range of data as input, leading to increased investment by TSOs/DSOs to determine different profiles.

	Optimization	**Machine Learning**	**Monte Carlo**
Algorithms	-Minimization of squared errors -MLE -Fit of coefficients	-MLR -SVR -NILM -ILM	-Flexibility region estimation -Uncertainties and variability of variables for flexibility
Pros	-Versatile -Robust -Various functions of OPT	-Versatile -Robust -Various functions of ML	-Versatile -Robust -Handle large dataset
Cons	-Constraint dependent -Data preprocessing	-Broad datasets needed -Number of inputs -ML type choice	-Accuracy, precision problem

Table 2.1: Summary of the methods

From Tab.2.1 it can be noted that optimization is a good solution to estimate the flexibility of the system. However, it will require a sensitivity analysis to have accurate and good results. As for machine learning, it seems to be also a good solution but requires a lot of data to initialize the model and validate it. In the case of aggregated data, it might lack some. Monte Carlo seems not to fit in the case of dealing with aggregated data. It would be a good choice if more data were available to determine a solution.

These methods highlight the different processes available in the literature and show that different methods are preferred over others regarding the case. The optimization problem has risen to be an interesting method to identify profiles based on the minimization process, yet it still needs to improve as the accuracy is dependent on the presence of outliers value [59]. Machine Learning methods can be interesting, but require in most cases a consequent amount of data. Such a requirement is a drag for its implementation. The data needed for most of the cases will be displayed in Sec.2.4

The main issue with estimating the flexibility is to structure the process to achieve a high level of accuracy in the identification. No study has been performed where the identification of different assets takes place. Hence, finding the different steps to be implemented while keeping a high level of accuracy will be a challenge in the design of the protocol. Furthermore, no study has been done that pursue the goal of identifying different assets within aggregated measurements.

2.4. Data-set

Before the methods presented in 2.3, the data used in the literature often goes through different processes and choices before being fed into the algorithms. In the following sections, the process of the data will be detailed as well as the methods to create the profiles required to estimate the flexibility.

2.4.1. Process of the data

To create different profiles, it is necessary to establish the baseline consumption of local consumers and then add various types of profiles to the data. The baseline requires data without any actions from aggregators, and instances of data with errors are replaced using Forward-Backward Autoregression, which outperforms Linear Interpolation [49]. Different groups of houses must be defined to create accurate baselines for various groups of individuals, such as neighbourhoods or boroughs [49]. Consumption can vary based on different energy labels of buildings; for instance, in [66], the group of houses chosen complies with a certain standard. Once the standard has been defined, deviations can occur, such as in [49], where Demand Response (DR) mechanisms introduced by the grid operator can affect consumption patterns. In such cases, a smoothing process can be applied if more accurate data is needed. After establishing and processing the baseline, if necessary, different profiles can be added, such as those of heat pumps and solar panels. The baseline can be estimated by attempting to disaggregate solar production from grid measurements.

Data preparation can also be conducted using machine learning techniques, as shown in [37], where higher complexity algorithms like NILM provide more detailed inputs. In the paper, the NILM algorithm is used to determine the time of use of flexible appliances, based on event detection. When an event exceeds the limits of usual data, it goes through the algorithm, enabling precise energy consumption determination and identification of the operation state of each appliance, as discussed in Sec. 2.3.3. Regarding the data that are analysed, if data are missing from the meter [66] uses linear interpolation to fill the gaps.

2.4.2. PV profile

Solar panels can be modelled with various parameters such as orientations, tilt angles, type of solar panels, nominal power, and module efficiency, all of which contribute to creating solar panel profiles. The PVLIB library is commonly used for this purpose. Additionally, inverter data, as used in [66], can provide insights into the power being fed into the grid.

To improve profiles, optimizing the gap between the modelled profile and the disaggregated one could be beneficial for accuracy. As mentioned in [59], determining a precise baseline of consumption for the base load is crucial for accurately assessing solar production at each instant.

2.4.3. Heat Pumps profile

To create heat pump profiles, data from heat pump meters are retrieved, as discussed in [51, 66]. In [66], measurements provide insight into the electrical household load, including voltage, current, active power, apparent power, and reactive power, at a 10-second resolution.

In the perspective of modelling heat pumps, as described in [51], the assumption that all heat pumps run at the same times of day as conventional heating systems is made. Additionally,[66] provides insight into electric consumption's active power and reactive power for heat pumps and households. The correlation between seasonal changes and higher active power consumption during winter is highlighted in [66]. Heat pumps are predominantly used during the winter months [66].

An analogy between heat pumps and boilers is noted in [51], where consumption peaks hit the same power levels. However, the shapes of the curves for heat pumps and boilers differ significantly during off-peak hours. Therefore, using a gas demand curve to build the heat pump consumption profile is not recommended, or it requires additional steps to fit the heat pump profile. The model gives insight into the usage of heat pumps for heating the domestic hot water when solar radiation is insufficient, most of the consumption of heat pumps comes from heating the space in housing. It is also said that the heat pump is for the moment only used as a boiler and further strategies could be developed to create several options to use heat pump systems more efficiently.

In [67] a machine learning approach using XGBoost to estimate the heat pump profile is pursued, first the required data needs to be selected and prepared. The data that have a significant deviation from the rest of the data set shall be removed using the interquartile range method. Following this step, the data are normalized for the correlation coefficients to be computed using the Spearman coefficient that gives insight into negative and positive correlation. To achieve better efficiency of the algorithm XGBoost is used on linear regression and polynomial regression models, and the least squares approach is applied. The Spearman coefficient allows to filtering of the different kinds of features to be used coupled with the variance inflation factor (VIF). To have precise and accurate results, the VIF needs to be between 1 and 5, therefore eliminating the features that have a high correlation.

In [68] a different approach has been pursued, real data have been used to model the heat demand from countries in Europe as well as the coefficient of performance(COP) of the heat pump. The heating demand contains data for water, space and air heating. They use the wind speed and temperature based on daily reference temperatures and wind speed averages to model the daily demand and then scale it. Following this approach, the building properties and occupant behaviour are utilized to estimate accurate profiles for the heat demand in the buildings on site. As for the COP, the values are computed using ambient air and soil temperatures. A different kind of heat demand is then considered as presented below :

In addition, if further accuracy is needed, [21] indicates that to provide the most precise heat pump profiles, the minimization of the residual sum of squares between the predicted heat pump power and the actual power could be done.

2.4.4. Conclusion

Different methods have been found to tackle congestion and flexibility stands out as one of the most efficient ones. Flexibility has been defined in this chapter as a quantity of energy that can be shifted to relieve the pressure on the power distribution system. The future energy system is required to reduce the variations in its demand. The goal is to minimize the variability in the demand. To accomplish this, the need to identify the different flexible assets is primordial. Different machines can be the source of flexibility, such as VRES, heat pumps and V2G. The problem is to identify the different energy resources in the mix and to find the best method for the process based on the different data available to the problem's owner. As the different assets can mitigate each other, different questions remain unanswered. How does a production one relate to a consumption one? Even if the identification works for different assets, the uncertainty of the method would need to be evaluated. Then, what criteria would be needed to answer the different constraints? What would be the inputs and constraints for the method?

All of these parameters shall be taken into account and kept in mind while creating the methodology. The different steps will have to address these constraints and make sure the frame for the model is correct.

3
Methodology

In this section, the methodology regarding the research work about the estimation of flexibility will be explained. The first section will display the different assumptions that were chosen. As well as the user type that can apply the model that has been developed. Following Sec. 3.1 once the assumptions were selected, the type of data provided to the model will be covered in Sec. 3.2. The different types of data and processes before it can be input will be explained as well. Sec. 3.3 present the different algorithms concerning the different goals of the model. In Chap. 4 the results of the models will be explained and evaluated. Each of the models will be given its own set of variables to assess the accuracy of the results.

The goal of the model that has been built in this research work is to identify flexible assets in aggregated data provided by the power substation based on specific profiles such as solar and heat pumps. Based on the number of assets identified in the aggregated data from the power substation level, an estimation of the flexibility in terms of reduction peak load will be able to be pursued. Therefore, the model should be able to handle noised data and to identify the capacity that can be considered as a flexible quantity. The output of the model would be the capacity of solar and heat pumps that can be utilised to alleviate load from the grid.

3.1. Assumptions

In this section, different types of assumptions have been made in this research to create a frame where the model would be applicable. The assumptions are the following:

1. **Operators**: The model will be used by DSOs, which implies that they have access to the number of people connected to the power substation, the current measurements as well as the voltage from the different power substations at an hourly scale.
2. **PV configuration**: The PV configuration is not studied in this research work, the power produced from solar is considered to be independent of the configuration.
3. **Heat pump consumption**: All heat pumps have the same consumption profile as considered as conventional heating system, no demand response mechanisms can be witnessed.
4. **Consumption**: The trend of the consumption curve in electricity is presumed to be the same as the one provided by the MFFBAS [69] of the AZI profile
5. **Household**: The households from the project are considered to be multi-family households.
6. **Power Factor**: The power factor is considered to be 0.95 [70]
7. **Baseline noise**: There is a noise baseline on top of the consumption and the production of energy in the grid.
8. **House insulation**: The houses have a low level of insulation
9. **Study**: A study has been done for a similar neighbourhood, there are 3089 consumers and a potential solar capacity of 2.05 [GWh], yet 194000 [kWh] has been installed

3.2. Data inputs

Different types of data must be provided to the model to be able to determine the quantity of solar and heat pumps available in the power substation. The different types of information required for the model are detailed in the following subsections. In Sec.3.2.1 the steps to obtain aggregated data are explained. Then, in Sec.3.2.2 the construction of the consumption curve is introduced. Therefore, the profiles needed for the identification are detailed in Sec.3.2.3 and Sec.3.2.4. The weather data are introduced in Sec.3.2.5 and finally, the process for the noise implementation is documented in Sec.3.2.6.

3.2.1. Aggregated data

First of all, the aggregated data from the neighbourhood must be computed to be able to identify any profiles. The measurement units available at the power substation level are the current as well as the voltage level. From these measurements, we can calculate the aggregated power consumption of the neighbourhood using Eq.3.1:

$$P_{sub[kW]} = \sqrt{3} * PF * I_{[A]} * V_{L-L[V]}/1000 \tag{3.1}$$

In this case, the power factor (PF) is assumed to be 0.95 as done in [70] for a neighbourhood level. In Eq.3.1, the voltage level is 10,000 [V], the division by 1000 is for the units to be in [kW] and the current measurement is from the mentioned study in [A]. Thanks to these different data, the aggregated profile can be computed. A day has been chosen over the year for representing the data to be understood, this day is the first one of the year. Some average profiles for PV and HP will be shown in their respective sections. A 24-hour sample is proposed below of the chosen day:

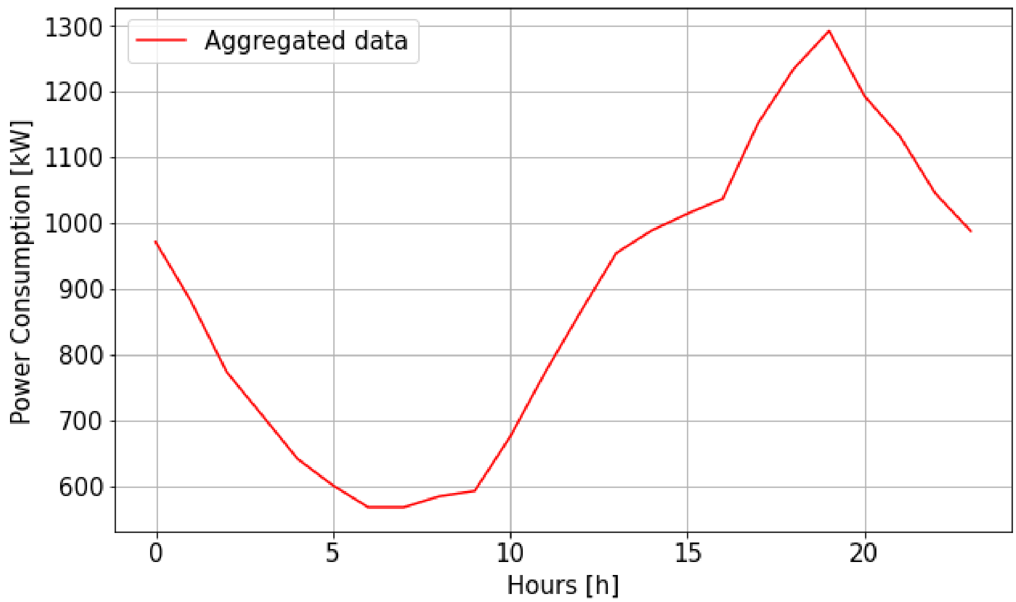


Figure 3.1: Aggregated data from the neighbourhood on 24-hour timelapse

Fig.3.1 highlights the behaviour of consumption, there is a low level of consumption during the night and over the day the peak happens around $\approx$ 7 pm.

3.2.2. Consumption Curve

First, the use of a consumption curve is a crucial point in the model. The starting point is that the load from the neighbourhood from the measurements available at the substation level can be derived from the equation 2.1. Therefore, the term $P_{load}(t)$ can be decomposed as :

$$P_{load}(t) = P_{bsl}(t) + \varepsilon_{nsd}(t), \forall t \epsilon T \tag{3.2}$$

In Eq.3.2, the load is decomposed in two terms,$\varepsilon_{nsd}(t)$ that corresponds to the variations in the power fluctuations and $P_{bsl}(t)$ which represents the baseline of consumption, it is considered as the inflexible part of the consumption, this part of the aggregated data is the first input of the model. To create the baseline of consumption, data from the MFFBAS database [69] are retrieved. The values from the AZI profile are used, they correspond to the consumption values where no inputs are considered. As the sum of the data from the profile is equal to 2, a normalization process is applied and then the baseline is computed :

$$P_{bsl}(t) = n_{consumer} * E_{avg} * AZI(t) \tag{3.3}$$

Where the $n_{consumer}$ is the number of people connected to the power substation, E_{avg} is the average use of energy over a year in [kWh] and $AZI(t)$ is the normalized profile of consumption without any input. This process allows the creation of the consumption curve.

3.2.3. Solar unit profile

To identify the solar profile along the year a solar unit profile is needed. The process for the generation of such data begins by choosing a type of solar panel in a database of the pvlib library and then modelling the output using the different functions available in the literature [71]. Once this has been done, the location must be chosen, in the case of the study, it is Amsterdam. When these conditions have been completed, we size the profile to be a 1 [kW] capacity. Then the values produced by the solar profile need to be checked that they are positive. The unit profile can be used then as input for the model. The code for such a process can be found in the appendix.

3.2.4. Heat pump unit profile

To generate the heat pump unit profile the process differs as no generic consumption data exist for heat pumps. In this research, the work from [66] is used to create the profiles. First, the heat demand curve from the neighbourhood is retrieved. Then the COP produced by [66] for the three types of heat pumps (ASHP, GSHP and WSHP) are used to create an average COP over the year. The heat demand can be divided by the average COP over the year to recreate the heat pump behaviour profile. The profile is normalized in [MW/TWh], as for the model a unit model is needed. To answer this criterion an average consumption for a heat pump is considered to be 4993 [kWh] [66], the profile is then multiplied by this value. To take into account the variations that can take place in the consumption, a noise data baseline can be introduced based on the standard deviation from the heat pump profiles.

3.2.5. Weather data

In the algorithms, the different processes require additional data such as the temperature of the air from the location. These data can be found on the KMNI website [72].. These data correspond to the temperature required to model when the heat pump is not running, or different inputs needed for the solar model.

3.2.6. Noise implementation

The models will have to be validated with noise in the data. A noise baseline will be implemented onto the data based on a Gaussian distribution. The noise value is based on the standard deviation value of the HP profile.

First, the standard deviation(STD) from the data from the aggregated data is computed and then a normal distribution is introduced based on the STD.

$$\rho(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.4}$$

Eq.3.4 is used to implement the noise in the data, where μ is the mean and σ is the value that corresponds to the noise we implement. The different values of σ fluctuate depending on the test. Different cases are studied that can be summed up in the following list :

- Noise case 1 : $\sigma = \frac{\sigma_{std}}{2}$

- Noise case 2 : $\sigma = \frac{\sigma_{std}}{4}$
- Noise case 3 : $\sigma = \frac{\sigma_{std}}{8}$

3.3. Algorithms

In this section, the algorithms that are being used in the different parts of the model are detailed below, Alg.1 provides insight into how the baseline of noise is computed. Alg.2 presents the process initially developed for the single identification model, in this perspective Alg.3 determines the number of customers connected to the power substation. Then, Alg.4 presents the method for the first optimization method concerning the mixed identification model. Finally, Alg.5 is the second method developed for this case. The data are in the hourly format which corresponds to a list of size 8760 for a year and the number of iterations to produce the results is 20 iterations. A list of acronyms used in the algorithms and in the equations is presented below :

1. **BSL** : Baseline
2. **PP** & **NP** : Positive Peak & Negative Peak
3. **AG** : Aggregated data
4. **Nsd** : Noise
5. **Ult** : Ultimate
6. **Irr** : Irradiance

Algorithm 1 : Noise baseline process

Require: $i_{\text{(no PV)}} = i[P_{AG} > 0]$ & $i_{\text{(PV)}} = i[P_{AG} < 0]$
  $BSL \leftarrow P_{AG}(i_{\text{(no PV)}})$
  for i in $i_{\text{(PV)}}$ **do**
    $BSL(i) = \frac{1}{k}\sum_{j=n-k+1}^{n} BSL(j)$
  end for
  for i in BSL **do**
    $BSL_{ult}[k : k+6] = \frac{1}{k}\sum_{i=k}^{n=k+6} BSL(k)$
  end for
  $P_{AG} = P_{AG} - BSL$

Alg.1 needs to have as input the power consumption flow from the considered area. Two lists are needed as inputs, the list where all the indexes that correspond to negative power flow, and a similar one but for the positive power flow. For the indices where the power is positive, the value of the aggregated data is kept for the noise baseline, and for all of the other indices where the power flow is negative, the values are assigned 0. The data where a zero has been assigned are then computed using a smoothing average based on the value of the closest non-zero value. Finally, for having constant value over specific periods a smoothing average process is applied to the data computed for the noise baseline. The process explained in Alg.1 provides the baseline of noise for a whole year. A mean value for 6 hours is computed to have a constant baseline over periods. Otherwise, the noise fluctuations would be too important and then represent an asset.

Algorithm 2 : Single asset identification

Require: P_{AG} & P_{unit} & Irr_{PV}

$$\min_{\alpha,\alpha\in\mathbb{N}}(\sum_{i=1}^{8760}(P_{AG,i}-\alpha P_{unit,i})^2+\eta f_{Nsd})$$

if Asset == PV **then**
  $i_{err} = Irr_{PV} == 0$
  for i in i_{err} **do**
    $P_{PVscaled,i} = 0$
  end for
  $E_{PV} = \sum_{i=1}^{8760} P_{PVscaled,i}$
end if
if Asset == HP **then**
  $P_{HPscaled} = \alpha HP_{unit}$
end if

Alg.2 corresponds to the identification in the case of the single asset model. The inputs needed for the code consist of the aggregated data from the substation, the unit profile of the chosen asset, and the irradiance data from the chosen location. Then the optimization process can take place and the output is the integer α from the minimization of the squared error. If the asset is PV, the irradiance is used to remove the outliers in the profile, and then the annual yield can be derived. Otherwise, for the heat pumps the profile is derived using the integer determined by the optimization problem.

Algorithm 3 : Number of consumer check

Require: AZI & P_{AG} & $i_{noPV} = i[P_{PV} == 0]$
for i in i_{noPV} **do**

$$\min_{\gamma}\sum_{i=i_{noPV}[1]}^{8760}((P_{sub,i}-\gamma AZI_i^2)$$

end for

Alg.3 is used in the case there is only PV in the mix to enhance the accuracy of the model. The indices where the PV production is 0 are retrieved. For the corresponding indices, the number of consumers connected to the power substation is derived using a minimization process of the squared error. Then the correct consumption profile from the neighbourhood is computed and used for the identification.

Algorithm 4 : HP & PV identification algorithm method 1

Require: PP $\leftarrow max(P_{AG})$ & **NP** $\leftarrow abs(min(P_{AG}))$ & P_{AG} & PV_{unit} & HP_{unit} &Irr_{PV}
  if PP > NP **then**
    Do HP identification :
    for i in range(8760) **do**:
      $$\min_{\alpha,\alpha\in\mathbb{N}}(\sum_{i=1}^{8760}(P_{AG,i}-\alpha HP_{unit,i})^2+\eta f_{Nsd})$$
    end for
    $P_{HPscaled}=\alpha HP_{unit}$
    Create $BSL_{Nsd}=P_{AG}-P_{HPscaled}$
    Iteration for determination of baseline noise
    Indices where $HP_{unit}<0.6$ & $PV_{unit}==0$
    $BSL_{ult}[i]=\max(0,BSL_{Nsd}[i])$
    for i in Indices **do**:
      k=6 ▷ It corresponds to the sample size divided by 2
      $BSL_{ult}=\frac{1}{k}\sum_{j=n-k+1}^{n}BSL_{ult}(j)$
    end for
    $P_{AG}=P_{AG}-P_{HPscaled}-BSL_{ult}$
    Do PV identification :
    for i in range(8760) **do**:
      $$\min_{\beta,\beta\in\mathbb{N}}(\sum_{i=1}^{8760}(P_{AG,i}-\beta PV_{unit,i})^2+\eta f_{Nsd})$$
    end for
    $P_{PVscaled}=\beta PV_{unit}$
    $i_{err}=Irr_{PV}==0$
    for i in i_{err} **do**
      $P_{PVscaled,i}=0$
    end for
    $E_{PV}=\sum_{i=1}^{8760}PV_{scaled}(i)$
  else
    Do PV identification :
    for i in range(8760) **do**:
      $$\min_{\beta,\beta\in\mathbb{N}}(\sum_{i=1}^{8760}(P_{AG,i}-\beta PV_{unit,i})^2+\eta f_{Nsd})$$
    end for
    $P_{PVscaled}=\beta PV_{unit}$
    $i_{err}=Irr_{PV}==0$
    for i in i_{err} **do**
      $P_{PVscaled,i}=0$
    end for
    $E_{PV}=\sum_{i=1}^{8760}P_{PVscaled}(i)$
    Create $BSL_{Nsd}=P_{AG}+P_{PVscaled}$
    Iteration for determination of baseline noise
    Indices where $HP_{unit}<0.6$ & $PV_{unit}==0$
    $BSL_{ult}[i]=\max(0,BSL_{Nsd}[i])$
    for i in Indices **do**:
      k=6 ▷ It corresponds to the sample size divided by 2
      $BSL_{ult}=\frac{1}{k}\sum_{j=n-k+1}^{n}BSL_{ult}(j)$
    end for
    $P_{AG}=P_{AG}+P_{PVscaled}-BSL_{ult}$
    Do HP identification :
    for i in range(8760) **do**:
      $$\min_{\alpha,\alpha\in\mathbb{N}}(\sum_{i=1}^{8760}(P_{AG,i}-\alpha HP_{unit,i})^2+\eta f_{Nsd})$$
    end for
    $P_{HPscaled}=\alpha HP_{unit}$
  end if

Alg.4 corresponds to the first optimization method to identify solar energy and heat pumps. The data inputs to this algorithm consist of aggregated data from the substation level, a PV production unit of 1 kW, an HP unit consumption profile and irradiance data from the location. The process relies on determining first the amplitude of the negative peak and the positive peak. This step allows the algorithm to identify the prevalent asset in the mix. Then based on the highest absolute value, either solar or heat pump identification happens first. Following this step, a noise baseline is computed by subtracting the scaled asset profile from the aggregated data. To construct such a profile, the indices where the PV production is 0 and the HP normalized profile is below 0.6 are retrieved. The positive values from the baseline are kept. Otherwise, they are assigned a value of 0. The smoothing process explained previously in Alg.1 is applied to provide a noise baseline for the year. It can be subtracted from the aggregated data as well as the first scaled asset to provide the data for the second asset identification. The identification of the second asset can take place, and the minimization of the squared error between the remaining data and the unit profile can happen. Whether PV is the first or second asset to be identified, the power produced is filtered using the irradiance data from the location. The outputs of the algorithm are the number of HP units and the amount of PV energy produced over the year.

Algorithm 5 : HP & PV identification algorithm method 2

Require: P_{AG} & PV_{unit} & HP_{unit} & Irr_{PV}
Do HP identification :
$y_{noPV} = P_{PV} == 0$
$\min_{\alpha,\alpha\in\mathbb{N}}(\sum_{i=y_{noPV}[1]}^{y_{noPV}} (P_{AG,i} - \alpha HP_{unit,i})^2 + \eta f_{Nsd})$
$P_{HPscaled} = \alpha HP_{unit}$
Create $BSL_{Nsd} = P_{AG} - P_{HPscaled}$
Iteration for determination of baseline noise
Indices where $HP_{unit} < 0.6$ & $PV_{unit} == 0$
$BSL_{ult}[i] = \max(0, BSL_{Nsd}[i])$
for i in Indices **do**:
  k=6 ▷ It corresponds to the sample size divided by 2
  $BSL_{ult} = \frac{1}{k}\sum_{j=n-k+1}^{n} BSL_{ult}(j)$
end for
$P_{AG} = P_{AG} - P_{HPscaled} - BSL_{ult}$
Do PV identification :
for i in range(8760) **do**:
  $\min_{\beta,\beta\in\mathbb{N}}(\sum_{i=1}^{8760}(P_{AG,i} - \beta PV_{unit,i})^2 + \eta f_{Nsd})$
end for
$P_{PVscaled} = \beta PV_{unit}$
$i_{err} = Irr_{PV} == 0$
for i in i_{err} **do**
  $P_{PVscaled,i} = 0$
end for
$E_{PV} = \sum_{i=1}^{8760} PV_{scaled,i}$

Alg.5 has the same inputs as in Alg.4. The algorithm consists of first identifying the heat pump outside of the window of influence from the solar energy by minimizing the squared error between the HP profile and the aggregated data. This translates to when the PV production is 0. This feature is designed to avoid any influence from other assets to hinder the identification. It is this perspective that the minimization of the squared error is taking place. The output of the first optimization process is an integer number of the number of heat pumps identified in the aggregated data. The profile of the computed number of HP units is then subtracted from the mix. Once the disaggregation has been achieved, the baseline of noise is computed with respect to the procedure previously detailed in Alg.4 and deduced from the aggregated data. Finally, the PV identification can take place. The output of the second optimization method is the energy produced by solar throughout the year. The outputs of the algorithm consist of the number of HP identified and the amount of PV energy identified in the mix.

3.4. Work flow

In this section, the processes behind the identification of the quantity of PV and HP will be explained. The first step of the model is to identify assets separately and to enhance the accuracy in each case. Therefore, providing a solid basis for the identification process. Besides, identifying the important steps needed for enhancing the accuracy of the PV identification and HP identification separately, allows one to have a better insight into the advantages and disadvantages of each asset. Firstly, in Sec.3.4.1 the identification of the solar in the mix of aggregated data will be explained in detail. Then, the heat pump process will be introduced in Sec.3.4.2 and finally, the mixed identification is presented in Sec.3.4.3.

3.4.1. PV identification

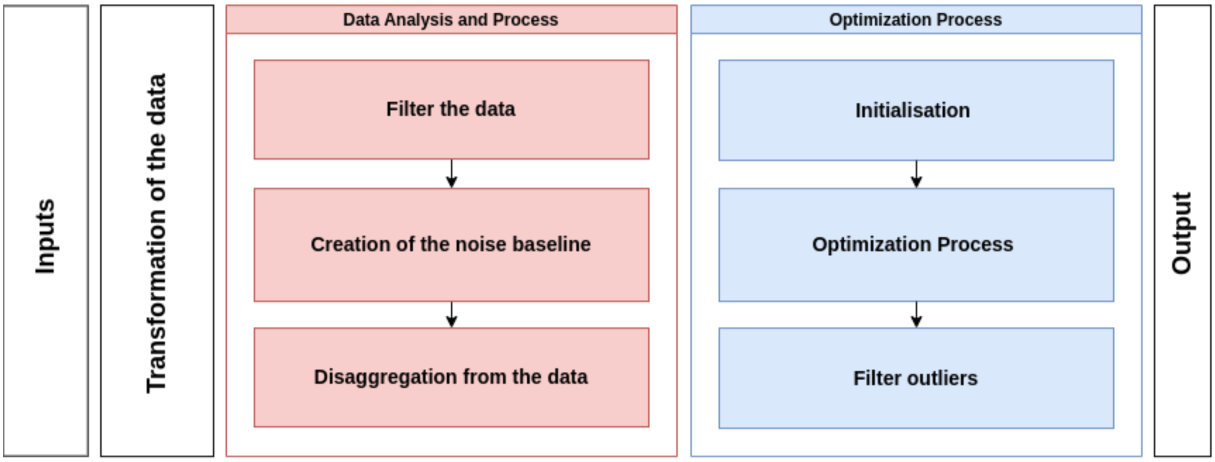


Figure 3.2: PV identification process

The PV identification goes through different parts presented in the flowchart above. As shown in Fig.3.2, the first step consists of providing the inputs to the optimization process. The inputs consist of the following:

1. Consumption data from the power substation level
2. AZI profile
3. PV unit profile of 1 kW
4. Irradiance data

The data for the identification of PV can be seen in the following figure :

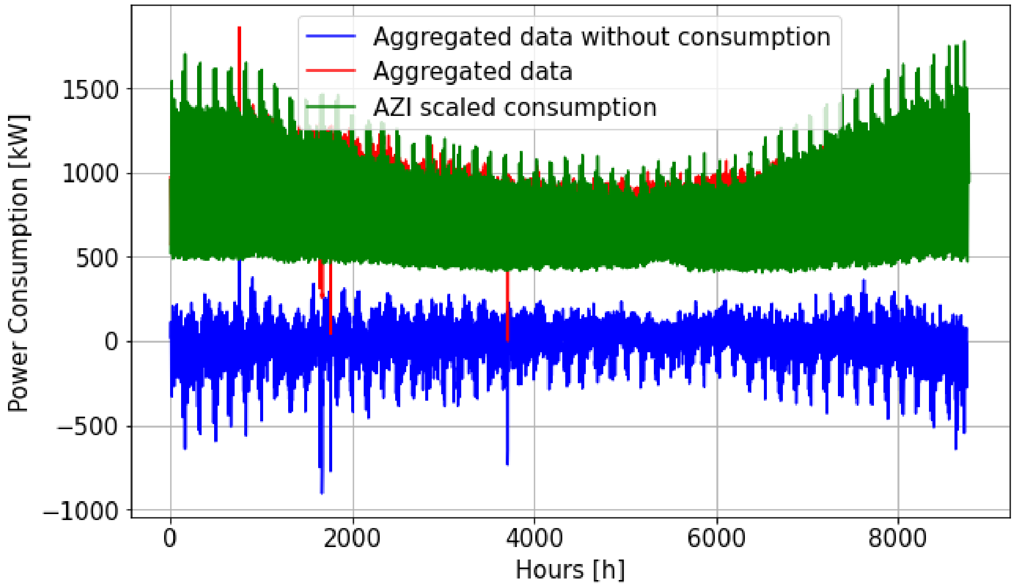


Figure 3.3: Data for the PV identification

Fig.3.3 highlights the different curves that are being used. The red curve represents the consumption data that are retrieved from the power substation and computed using Eq.3.1, and the green curve corresponds to the consumption that has been scaled according to the number of customers throughout Eq.3.5. The blue curve is the remaining data when green is subtracted from red. The negative power flow from the blue curve is accounted as an energy production asset, in this case, it is solar energy. However, one can see that the blue curve is suffering from a lot of variability. Such phenomenon comes from the remaining asset connected to the power substation level and that degrades the PV profile. These data can be accounted for as a noise baseline of consumption during the day induced by different utilities in the neighbourhood. The AZI profile is built by checking the number of customers connected to the power substation through a minimization of the squared difference between the data from the substation and the AZI profile. This process can only take place outside of the PV production hours to avoid any indirect influence as seen in Alg.3.

$$\min_{\theta} \sum_{i=i_{noPV}[1]}^{8760} ((P_{sub,i} - \theta P_{AZI,i})^2) \tag{3.5}$$

Eq.3.5 corresponds to the minimisation process between the aggregated data from the substation level and the power consumption, $P_{sub,i}$ is the power flow from the substation and $P_{AZI,i}$ is the power consumption that is scaled using θ, the goal is to minimize the square error. The process takes place outside of the window of solar influence, which translates to PV production being 0. The algorithm's outcome is the number of customers connected to the power substation, which then allows a more accurate estimation of the consumption. After the check of the number of customers, the consumption profile can be subtracted from the data from the power substation. The outputs of such a process are the inputs to the Data Analysis and Process part. The baseline process can take place, first of all, the data needs to be classified based on the condition presented in Alg.1. Using the smoothing process introduced in Alg.1 the noise baseline can be estimated and subtracted from the data to provide an accurate profile for the solar estimation. Once these steps have been completed the following data can be used for PV identification :

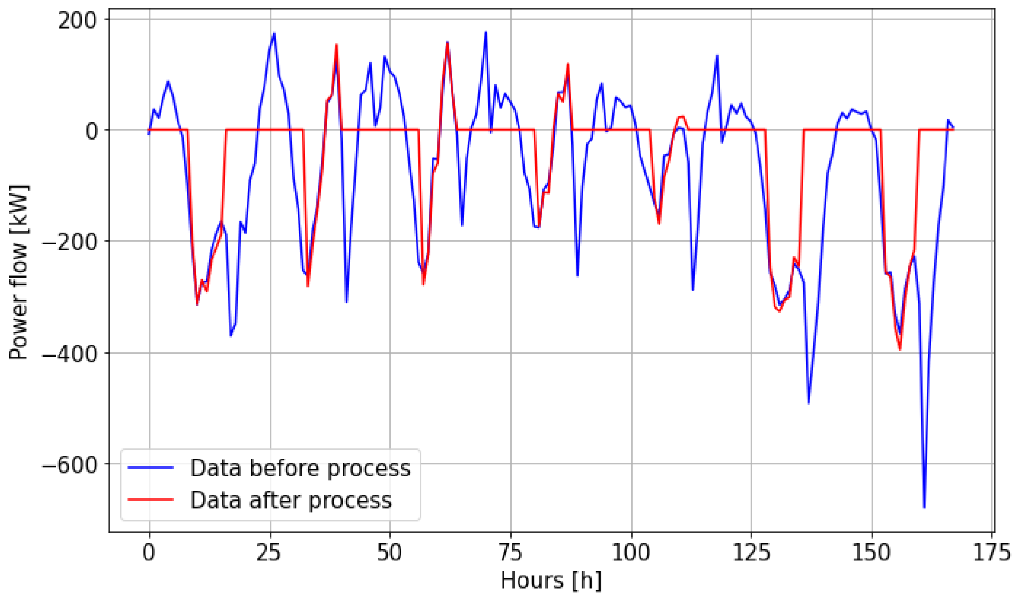


Figure 3.4: The graph highlights the process used in the initialization prior to the optimization to reduce the error in the results

According to Fig.3.4, the data in blue represents the overall data without any filter. The data in red represent the data fed into Eq.3.6. A filter for the irradiance is applied to the data as seen on the red curve. When the irradiance is 0, the data are assigned a 0 value in order to enhance the optimization output as seen in Fig.3.2. The profile on which we can estimate the energy produced by PV can be seen in Fig.3.4.

$$\min_{\beta, \beta \in \mathbb{N}} \left(\sum_{i=1}^{8760} (P_{AG,i} - \beta PV_{unit,i})^2 + \eta f_{Nsd,i} \right) \tag{3.6}$$

Eq.3.6 refers to the method used to identify the solar energy production along the year, the term $\eta f_{nsd,i}$ corresponds to a noise implementation process, $P_{sub,i}$ is the power from the substation level and $PV_{unit,i}$ is the power from the PV unit and β is the integer parameter used for the optimization. Eq.3.6 is minimised over the year to provide a fixed amount of solar panels.

To validate the model, the data from the project are used, the solar energy produced over the year is 194,000kWh. Once the algorithm identifies the number of solar panels present in the aggregated data the computation of the annual yield can be derived in order to check how much energy is produced as seen in Alg.2. The goal is to validate the model by having high accuracy. The validation of the model can be found in Sec.4.2.1.

3.4.2. HP identification

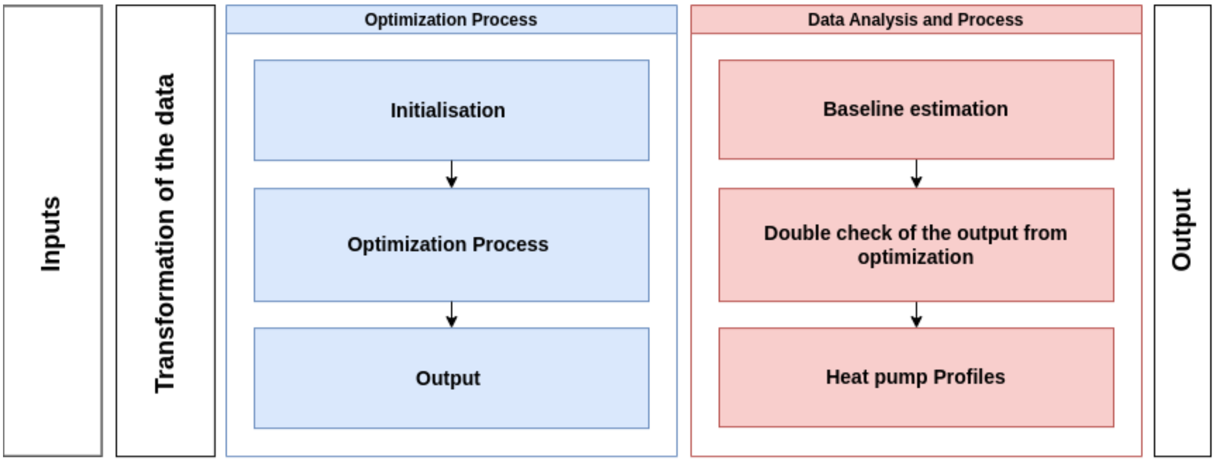


Figure 3.5: Process for heat pump identification

The HP identification goes through different parts presented in the flowchart above. As shown in Fig.3.5, the first step consists of providing the inputs to the optimization process. The inputs consist of the following:

1. Consumption data from the power substation level
2. AZI profile
3. HP unit profile
4. Temperature data

In the Transformation of the data in Fig.3.5, the AZI profile is subtracted from the data from the power substation which corresponds to the blue curve in Fig.3.6, it aims to enhance the accuracy for the heat pump identification. Following this step, the data can be used for the identification process. A sample of the data can be seen in the following graph :

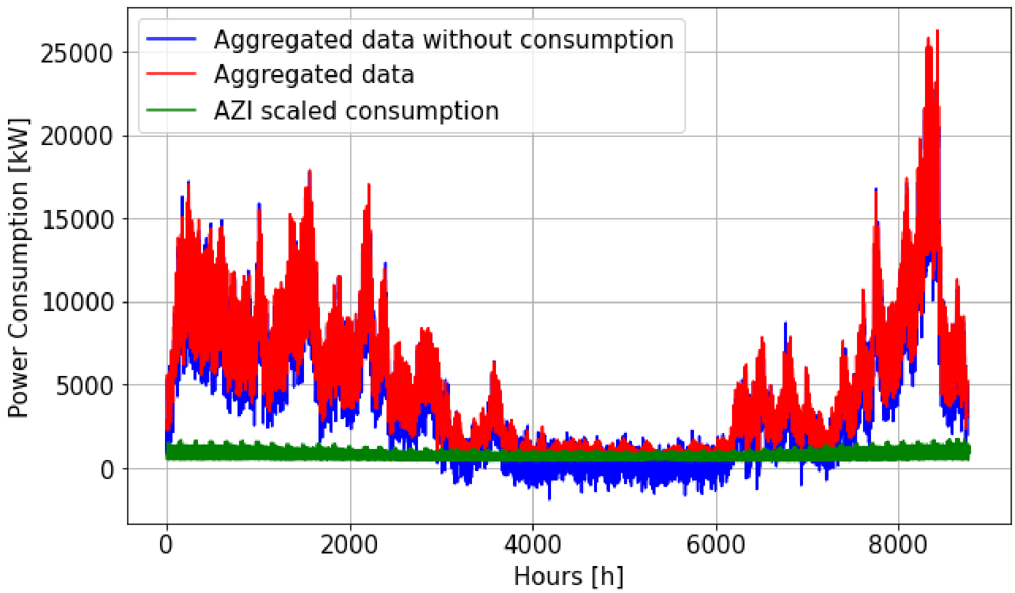


Figure 3.6: Data input for the heat pump identification

Fig.3.6 represents the data before and after consumption is subtracted from the aggregated data. The red curve represents the data from the power substation with heat pumps already implemented in it, and the blue curve represents the remaining data once the consumption has been subtracted. In the case of the HP identification process, the heat pump profile can already be identified by looking at the data. It has a strong influence compared to consumption, and more specifically it has two daily peaks that can reduce the error as one peak can still be identified if another asset mitigates the other one. These data consist of the inputs of the Optimization process that can be seen in Fig.3.5. The data then can go through the process of minimization of the squared error between the two curves as introduced in Eq.3.7 and in Alg.2 :

$$\min_{\alpha,\alpha\in\mathbb{N}}(\sum_{i=1}^{8760}(P_{AG,i}-\alpha HP_{unit,i})^2+\eta f_{Nsd,i}) \tag{3.7}$$

Eq.3.7 refers to the identification process of the heat pumps, $P_{sub,i}$ is the power flow from the substation level, the term $\eta f_{nsd,i}$ corresponds to a noise implementation process, $HP_{unit,i}$ is the power consumption from the heat pump unit and α is the parameter that is used to scale the HP profile. The output of the algorithm corresponds to the number of heat pumps estimated. Once the number of heat pumps is estimated the Data Analysis and process can begin, by using this result a noise baseline can be estimated. It is the result of the different domestic assets that remain in the mix, their fluctuations can be witnessed once the heat pump profile is subtracted. After the noise baseline is estimated, a second check of the estimation of heat pumps can be done to assess the accuracy of the method throughout a retroactive loop. This iterative process is pursued in order to reach high accuracy, finally, the heat pump profiles are computed and the process is over.

The goal is to validate the model by having high accuracy. The validation of the model can be found in Sec.4.2.2.

3.4.3. PV and HP identification

Once the algorithms from the PV identification and the HP identification have reached a high level of accuracy in the results, the following method will be investigated to identify two assets in the mix. Therefore, the next challenge relies upon combining them to see what happens when both of the algorithms are merged. In this case, the inputs consist in :

1. Consumption data from the power substation level
2. AZI profile
3. HP unit profile
4. PV unit profile
5. Temperature data
6. Irradiance data

However, the process behind the identification is different. The data that the optimization process needs to handle can be chaotic and less prone to an identification process. A sample of data can be found below :

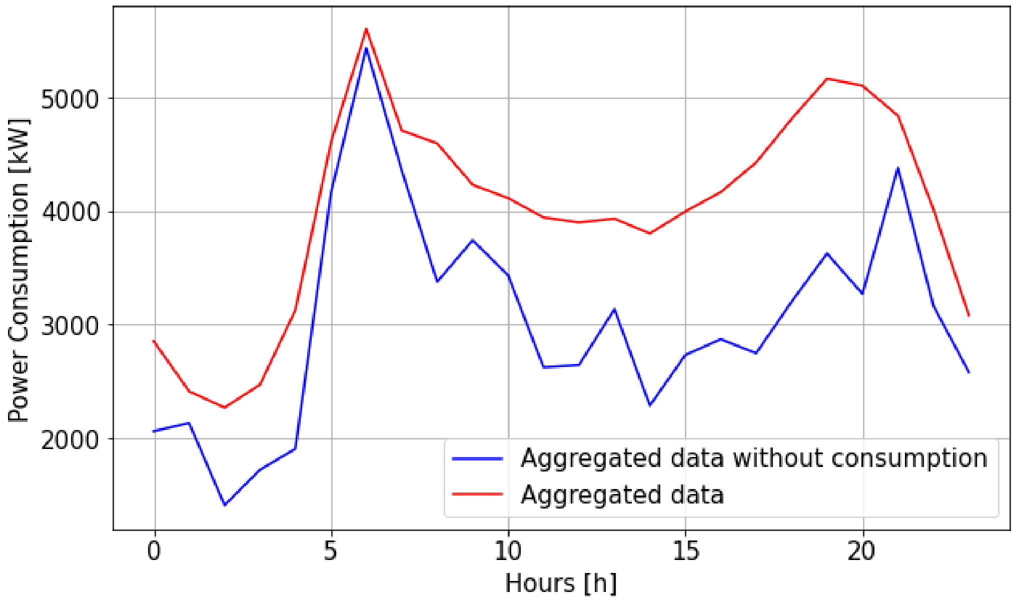


Figure 3.7: Aggregated data for mixed identification

Fig.3.7 provides insight into the data consumption in the neighbourhood. It can be seen that identifiable peaks for heat pumps can already be highlighted. However, in the case of a bigger capacity for solar panels, the first peak of heat pumps can be mitigated by solar production. This is why the two different methods for the algorithm can be found in the Alg.4 and Alg.5.

The process presented in Alg.4 highlights the mixed identification process. The starting point is to first identify which asset is predominant in the mix, therefore identifying the maximum and the minimum over the year. If the maximum value is higher than the minimum value then the heat pump consumption is predominant, and the other way around. Once the prevalent asset is determined, the process of identification of the chosen asset is pursued. Then the disaggregation of the asset is done in the mix, allowing the noise baseline estimation to take place. In this identification process, a problem arises, mixed identification only deals with the estimation of two assets in the power substation data. Therefore, the other types of assets present in the neighbourhood are not taken into account, inducing errors in the estimation of heat pumps and solar panels. Different scenarios will be tested in Ch.**??**, to evaluate the impact of the different inputs to the model. Once the estimation of the noise baseline is achieved, the data for the final identification can be prepared by subtracting the previously estimated noise from the data. Finally, the estimation of the last asset can then take place. The optimization process relies on the same minimization process as seen in Eq.3.6 and Eq.3.7.

A second method can be pursued to counter the mixed estimation, the procedure is similar to the one above. However, in the case of both solar and heat pumps implemented into the mix the profile of consumption from the heat pump can be mitigated by one from solar. Such phenomena can lead eventually to an increase in the error in the estimation of the assets. In this perspective, the following method is developed, the identification of the heat pumps only takes place when the PV production is 0. The heat pump identification happens first regardless of the prevalent asset. Then the noise baseline is estimated and subtracted from the data. Finally, the PV identification is pursued. The goal is to compare the accuracy of the two methods and then decide which one is worth being investigated.

The accuracy of the two methods will be analysed and compared. The goal is to reach a high level of accuracy in both methods. The validation of the two methods can be found in Sec.4.2.3

4

Validation of the model

This chapter will consist of three sections, Sec.4.1 will detail the KPIs used for the evaluation of the results. Then the validation of the model will take place in Sec.4.2. The number of assets for the inputs will be discussed in Sec.4.3. Finally, a conclusion will take place in Sec.4.4.

4.1. Evaluation of results

Different graphs and KPIs will be introduced to assess the results. The accuracy of the algorithm will be evaluated for each of the scenarios by using boxplots. The accuracy of the results can be defined as the proximity measurements accepted to the accepted value. In such perspective, the following equation is used to evaluate the accuracy :

$$\sigma_{relative} = \frac{abs(X_{objective} - X_{result})}{X_{objective}}, \tag{4.1}$$

Eq.4.1 refers to the formula used to determine the relative error of the result. $X_{objective}$ represents the value which need to be obtained and X_{result} the value obtained using the model. The accuracy will be evaluated through the relative error for estimating the energy produced by PV with respect to the value of 194,000 kWh for validating the model. As for the heat pump, the accuracy will be measured through the relative error of estimating heat pump units concerning the value of 3089 units for the validation of the model. The goal of the model is to have the highest level of accuracy for all methods. The variation of the relative error will be evaluated to grant more insights into the DSO's options. Furthermore, it should be able to handle noise in the data. For the validation of the model, the accuracy for HP should be close to 3089 units of HP identified, as for solar it should be 194,000 kWh. The model will be run in 20 iterations to provide accurate results for each scenario.

4.2. Validation of the model

For the validation of the model, three types of validation will have to take place, the PV identification model needs to be verified as well as the HP identification model. Finally, the validation of the mixed identification will take place.

4.2.1. PV identification

For the solar identification process, different versions of the optimization have been investigated over time. The different results of the different versions can be seen in the Tab.4.1. The relative error is with respect to the goal of identifying the 194,000 kWh.

In Tab.4.1, the solar identification model is shown to be accurate up to 5% of relative error using a minimization algorithm to find the solar energy in the aggregated data. Different processes have been documented to show the evolution of the variations of the relative error along the introduction of the different parts of the algorithm. The consumer check algorithm can be found in Alg.3, as for the baseline

Version	Number of Customers	Energy identified [kWh]	Number of solar panels	Relative Error [%]
Initial	3089	92,511.4	80.61	52
Consumer connected check	3066	79,625.9	69.38	58
Noise baseline	3089	209,613	182.6	8.0
Baseline + Consumer check	3066	203,247	177.1	4.7

Table 4.1: Evolution of the results along the different processes

it is in Alg.1. The introduction of the baseline of noise in the algorithm has proven to be efficient for enhancing the accuracy of the estimation of solar. This step must be pursued very carefully if other assets were to be implemented.

Version	Energy to be identified [kWh]	Mean Energy identified from PV [kWh]	Relative Error [%]	Relative error range [%]
Noise case 1	194,500	202,870	4.6	1.36
Noise case 2	194,500	203,241	4.7	0.79
Noise case 3	194,500	203,385	4.8	0.4

Table 4.2: Evolution of the KPIs over the different cases

The accuracy has reached a threshold of 95% and for the variations of the relative error, the interval of fluctuation is smaller than 2% of the objective value. Hence, the algorithm leads to a high level of accuracy and as the model is run 20 iterations, the results stay within a 2% fluctuation of relative error which leads to high levels of accuracy without many outliers. The noise implementation does not seem to have a great impact on the results. The algorithm has led to an estimation accurate up to 95%, therefore allowing the DSO to use this capacity as a flexibility asset.

4.2.2. HP identification

The result of the algorithm for the identification of the heat pumps is promising without several steps to enhance the process. The implementation of heat pumps in the mix is achieved according to the process described in Sec.3.2.4. The noise for the data is implemented using the computation of the standard deviation from the data and then implementing a Gaussian distribution of noise based on the value of the standard deviation. The accuracy is measured through the number of heat pump units identified in the mix and will be compared to the number implemented being 3089. The results for the identification of heat pumps can be found below :

Version	Number of Assets implemented [u.]	Mean number of Assets identified [u.]	Relative Error [%]	Relative error range [%]
Noise case 1	3089	3070	0.6	0.32
Noise case 2	3089	3072	0.55	0.13
Noise case 3	3089	3075	0.45	0.1

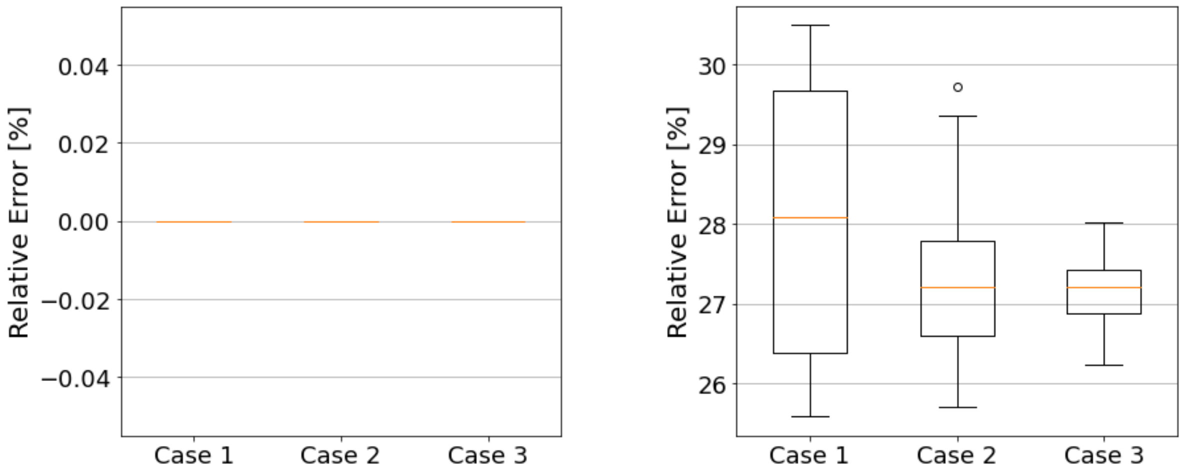
Table 4.3: Results for the case 1

The accuracy of the identification of heat pump units is reaching a level of confidence of 99%. The results are stationary regardless of the noise implemented as it can be seen throughout the standard deviation of the values of the relative error. In that perspective, the algorithm for the estimation of flexibility available to the DSO is providing accurate results.

For the estimation of one asset, the single identification model's accuracy is excellent given that DSO usually operates with 20% of relative error in the estimation of the available flexibility. The accuracy of the individual models has reached both a minimum of 95% and has proven to be stable throughout the iterations, the flexibility quantity can be estimated with less than 5% of relative error. This allows the DSO to identify assets in the mix and to determine how much quantity in the aggregated data can be considered flexible. However, the model can handle only data where one type of asset is present. The model that can identify both assets will be investigated in the next section.

4.2.3. Mixed identification

In this section, the mixed identification model will be validated through 4 different validation scenarios. First of all, the model will have to identify first PV, then a second scenario consists of identifying HP. Consequently, the identification of both PV and HP will be covered through the two methods that have been proposed in Sec.3.



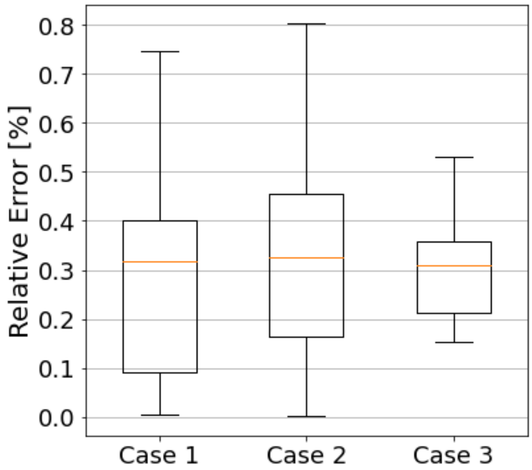
(a) HP identification for scenario 1 **(b)** PV identification for scenario 1

Figure 4.1: Boxplots of the results for Scenario 1(model validation)

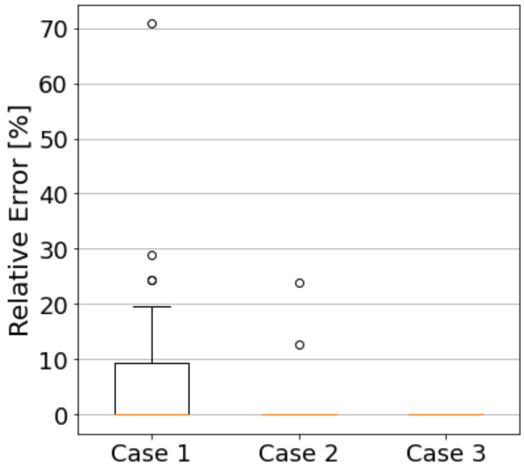
Fig.4.1 represents the result of the mixed model for the output of both PV and HP. For the identification of HP, Fig.4.1a shows that the relative error is 0% which corresponds to no heat pumps introduced in the mix.

Fig.4.1b shows the result for the PV identification, with 27.5% error on average. In case 1, relative error covers a broad range of values [25.4%,30.7%], in case 3 the values are within [26.5%,28%]. A reduction of 20% of the relative error fluctuation interval can be noticed in between each case. The interquartile range between the first quartile and the third is considerably reduced as the noise decreases. The median of the relative error decreases as the noise implemented decreases. In this case, the accuracy of the algorithm is reaching a minimum level of 70% with respect to the estimated goal of 194,500 kWh. Given that the method for a mixed identification process takes into account different noise levels and different processes for the noise baseline, the identification of only one asset solar is hindered by the process of mixed identification. Moreover, as solar energy is produced in small quantities the identification accuracy is less performant than at a higher level of asset implemented. In such a perspective, the estimation of the flexibility is harder at a low level of integration of assets, the accuracy of the model can then be considered to be dependent on the number of assets present in the grid. As for the estimation of flexibility, the relative error is reaching 28% here and is expected to decrease with the increased capacity of solar introduced. The accuracy for the DSO needs to be tackled at low capacities, in this case, the accuracy will be further evaluated at a higher level of PV introduced.

The next step in the algorithm is then to identify the number of heat pumps in the mix. The following results correspond to the scenario where only heat pumps are introduced.



(a) HP identification for scenario 2



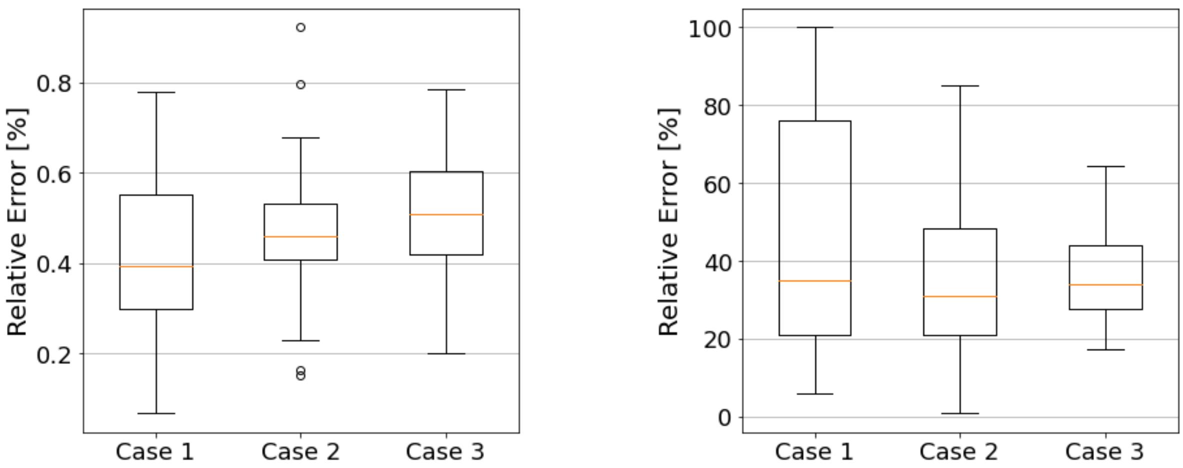
(b) PV identification for scenario 2

Figure 4.2: Boxplots of the results for Scenario 2(model validation)

Fig.4.2 represents the output for the scenario 2. In this case, Fig.4.2a underlines a high-accuracy model for the identification of heat pumps in the mix. The accuracy is high given that the relative error is $\approx$0.3 for the identification of heat pumps. In case 1, the relative error of the estimation of the number of heat pump units covers a range of [0%,0.75%] and in case 3 it is [0.15%,0.53%]. The interquartile range between the first and third quartile is within [0.1%,0.4%] in case 1 and [0.2%,0.35%] in case 3. The fluctuations of the results are small and accuracy is reaching high levels for the HP. The tendency previously noticed in the PV identification case is incorrect for HP identification. As the error in the estimation of HP units is very low, the decrease of the relative error follows a different trend as the accuracy is above 99%. Besides, the algorithm's accuracy respects the DSO standards which correspond to 80%.

However in Fig.4.2b, no PV shall be identified, the introduction of noise in the data shows that PV can be wrongly estimated if a level of noise higher than $\frac{\sigma_{std}}{4}$ is implemented in the data. In case 1, the relative error ranges between [0%,20%] and in case 2 & 3 the relative error is 0%. Conversely to solar identification, relatively speaking the energy consumed by heat pumps over the year is much higher than the solar ones, therefore the identification is easier. The flexibility estimation in this case contains very low error, therefore for the identification of HP, this algorithm is performing greatly. DSO may use the algorithm for the identification in the neighbourhoods of HP units therefore allowing accurate estimation of the amount of flexibility to take place.

As both the PV identification and the HP identification have reached sufficient levels to be tested together the next step consists in testing them with similar conditions as scenarios 3 and 4. The results for the validation of the method 1 can be found below :



(a) HP identification for scenario 3

(b) PV identification for scenario 3

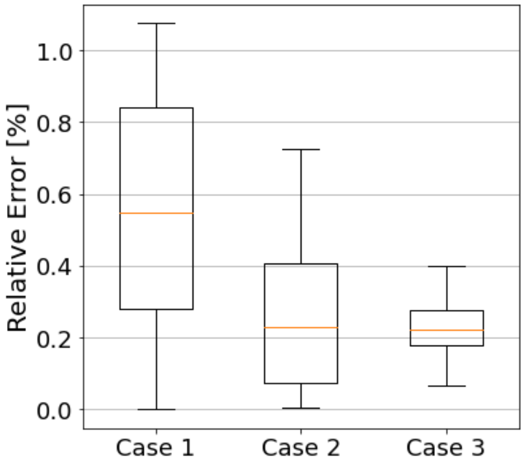
Figure 4.3: Boxplots of the results for Scenario 3(model validation)

For method 1, Fig.4.3 presents the results for both heat pump and pv identification. Fig.4.3a shows that the relative error for the heat pump is as low as [0.05%,0.8%] for the heat pump with little fluctuations in the relative error achieved. In case 1 the relative error is within [0.05%,0.8%] and the interquartile range is [0.3%,0.55%]. For case 3 the relative error ranges in [0.2%,0.7%] and the interquartile interval in [0.4%,0.6%]. The fluctuations here are once more small and the accuracy of the HP identification is above 99%, the estimation of the available flexibility in the mix can be performed with a high level of confidence.

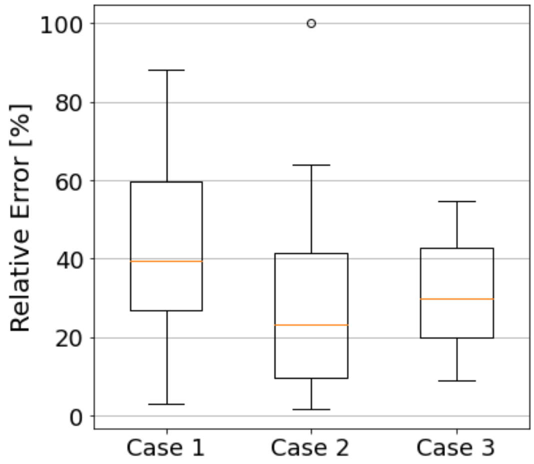
As for the PV identification, Fig.4.3b highlights that the accuracy of the PV identification is highly dependent on the implementation of noise. The relative error interval is first within [10%,100%] in case 1 and then [18%,62%] in case 3. The interquartile range is in case 1 [20%,77.5%] and case 3 [30%,44%]. As for the noise influence when looking at this scenario, the decrease in between each case does not follow the linear relation seen in the case where the model needed to identify one asset. There is first a decrease in the relative error of the PV identification of 10% and then of 50%. Compared to the case where there was no HP in the mix, the noise influence on the results worsened. PV identification is much harder than in Sec.4.2.1 as HP mitigate and introduces more errors in the data. Given that PV is present in small quantities, the identification is tedious. The estimation of the available flexibility for the DSO is difficult at a low level of integration of RES and with another asset to be identified, the noise is becoming an important problem. In such cases, the algorithm is confronted with the limit of the identification process and the interdependency between the quantity of assets to be identified. In the case of a large number of heat pumps and a small amount of PV to be identified, the prevalent asset identification error can lead to a massive loss of accuracy as well as an increase in the spread of the obtained values for the other asset identification.

The PV identification can struggle to identify the solar energy produced over the year. The overall energy consumed by the heat pumps is much larger than the power produced by the solar. Hence, the identification of the units in small quantities remains a problem. The error for the solar identification is increasing as the heat pump identification is also taking place.

The results for the method 2 can be found below :



(a) HP identification for scenario 4



(b) PV identification for scenario 4

Figure 4.4: Boxplots of the results for Scenario 4(model validation)

For method 2, Fig.4.4 presents the HP and PV identification results. Fig.4.4a shows that the accuracy for the heat pump is as high as [99.75%,99.8%] regardless of the noise. In case 1 the relative error is within [0%,1.1%] and the interquartile range is [0.3%,0.82%]. For case 3 the relative error ranges in [0.08%,0.4%] and the interquartile interval in [0.18%,0.28%].

As for the PV identification, Fig.4.4b highlights that the accuracy of the PV identification is highly dependent on the implementation of noise. Besides, in the results, method 2 proves to have better results than method 1 for identifying solar panels. The relative error interval is first within [5%,90%] in case 1 and then [10%,55%] in case 3. The interquartile range is in case 1 [27%,60%] and case 3 [20%,42%]. The median of the relative error in the PV estimation is lower than in method 1. This suggests that method 2 enables the identifying process to be improved. As for the noise, a decrease of 20% of the relative error can be noticed between cases 1 and 2, similar phenomenon is seen between cases 2 and 3. Hence, with this method, the noise relation is linear with the decrease of the relative error of the guess of the PV.

Therefore, the results obtained using method 2 decrease the relative error in the PV identification. However, the accuracy is not reaching as high levels as one would expect, the reason lies in the number of assets considered for the solar identification. If a higher level of assets were present in the data for energy production, the algorithm's accuracy would increase. As for the DSO, the estimation of PV is tedious when a small number of assets are implemented. The estimation of the heat pumps in the case of 3089 units has been successful regardless of the scenarios. The accuracy of the algorithm for the HP identification reached a minimum of 99% regardless of the noise introduced in the data. The results do not show signs of fluctuations in the estimation as the range of fluctuations in the relative error is below 1%. As for PV, the identification has proven to be harder than expected with relative error in the energy produced reaching 100%. However, at lower levels of noise implemented, the median of the relative error seemed to be stable with a value of $\approx$ 30%. The values for the relative error fluctuating over a large range indicates that the identification process at a low level of PV penetration is not stable as the range for the relative error covers a range of $\Delta\%$ = 90%. This underlines the error that can be obtained when using this model with a low number of PV panels in the aggregated data. When looking at the DSO perspective, taking into account that they operate with around 20% of relative error in their models, the HP identification is quite promising. The relative error in the HP identification is below 1%, the one for PV is harder to deal with. The result for the PV identification can be considered accurate with respect to the DSO's norms as at a small number of assets the identification is tedious. More information regarding the limitations of the model is presented in Chap.7.1

4.3. Implementation of assets

The number of assets that are implemented has a great influence on the level of accuracy. As can be seen with the number of heat pumps implemented in the following graph :

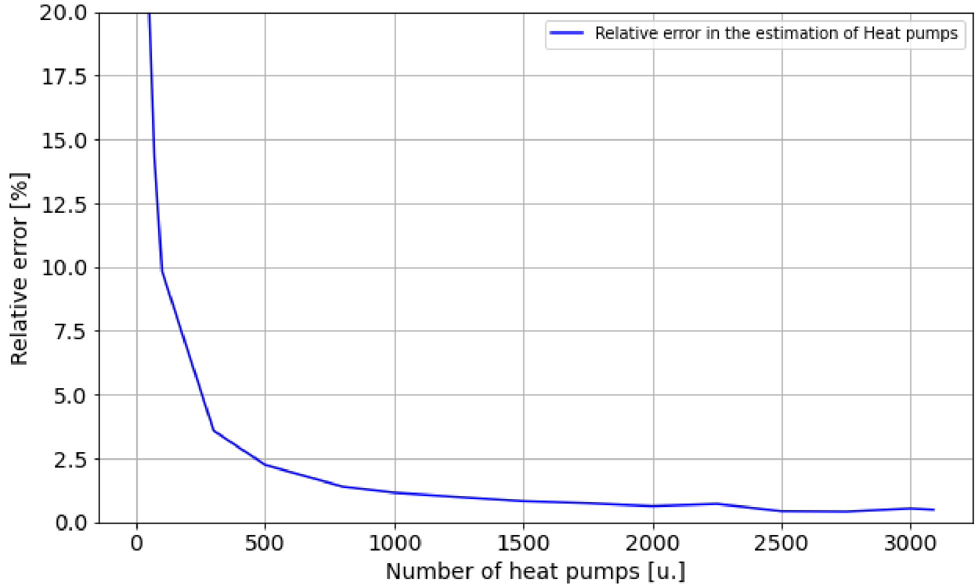


Figure 4.5: Accuracy of the identification based on the number of assets implemented

Fig.4.5 represents the relative error of the identification based on the number of assets implemented. It can be seen that below a number of 300 units, the relative error grows rapidly and reaches the limit of 20 % around 50 units implemented. This behaviour applies to any type of asset considered to be connected to the power substation. Therefore, the highest number implemented in the grid, the highest the accuracy of the model. As a small number of assets allow the range of the relative error to spread, the accuracy of the model is expected to be enhanced using a larger number of assets.

4.4. Conclusion

Overall, the model is performing well and accurately. The relative errors in the case of PV fluctuate tremendously, hence the accuracy is then degraded by the wide range of relative error values at a low PV penetration. However, one problem arises in the case of the model validation. As a small number of solar panels are present in the data, the identification with the mixed process is tedious and cannot reach a level of accuracy as high as for the individual identification. Such a problem can be considered as a usual problem that DSOs encounter when identifying assets in small quantities, the peak and the characteristics are then more prone to be mitigated by variations and other assets that proportionally have a bigger influence than on the heat pump. The identification at a low level of penetration of assets is tedious and requires different approaches. In this perspective, the accuracy of the PV identification can be considered sufficient. One would expect a decrease in the relative error if larger PV penetration were considered.

5

Scenarios analysis & results

This master thesis research work is pursued in the frame of a study on a neighbourhood. Scenarios corresponding to the different options that have been considered in the project will be tested. The value used for the solar corresponds to the maximum capacity that can be implemented in the study. As for the number of heat pumps, it has been decided based on the number of households. Different scenarios will be investigated to assess the model and evaluate the different cases. Depending on the different results, a sensitivity analysis will follow to discuss the different results and evaluate the proposed approaches in Chap.6. The different scenarios are the following:

- Scenario 1: Solar implemented (Annual energy yield of 2.05GWh)
- Scenario 2: Heat pump implemented (3089 units)
- Scenario 3: Solar and Heat pump implemented using method 1 (2.05 GWh and 3089 units)
- Scenario 4: Solar and Heat pump implemented using method 2 (2.05 GWh and 3089 units)
- Scenario 5: Solar and Heat pumps implemented using method 2 for the prevalent periods of the assets (2.05GWh and 3089 units)
- Scenario 6: Solar and Heat pumps implemented using method 2 with different numbers of iterations (2.05GWh and 3089 units)

The flexibility estimation will take place using the output of the model. The flexibility of the heat pumps is quantified in terms of the peak load reduction using curtailing or DR mechanisms. As for PV, the mitigation of the peak load is evaluated under the assumption, that solar energy can be used to alleviate the load at the effective time of the peak load. Therefore, decreasing the power demand when it is most needed.

5.1. Scenario 1

In this scenario, the basic case is developed where the PV generation is considered to be the only asset connected to the grid. Therefore, the model should identify the PV energy generated over the year using the Alg.4. The accuracy is evaluated through the relative error of the amount of yearly energy produced by PV compared to the value of 2,05 GWh. The range of accuracy values will be evaluated to gain more insights into the flexibility estimation. In this scenario, the PV identification cannot be mitigated by other types of assets. The quantity of solar implemented in the data corresponds to 2.05GWh solar energy produced over the year. Solar production does not fluctuate as much as the heat pump, therefore no noise has been implemented into the solar production. The data at the beginning of the algorithm have a peak of production as seen in the following figure :

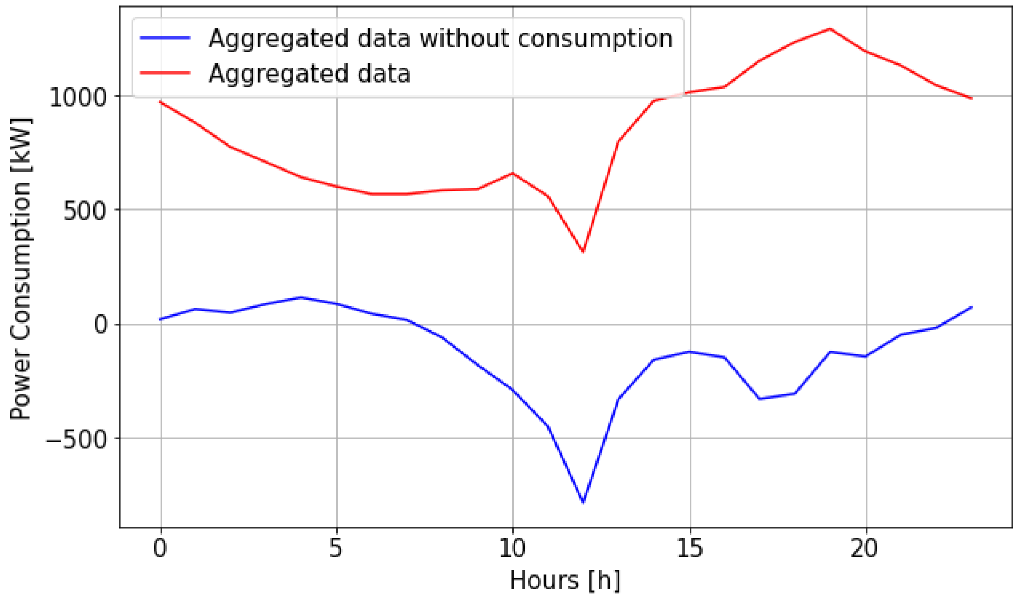
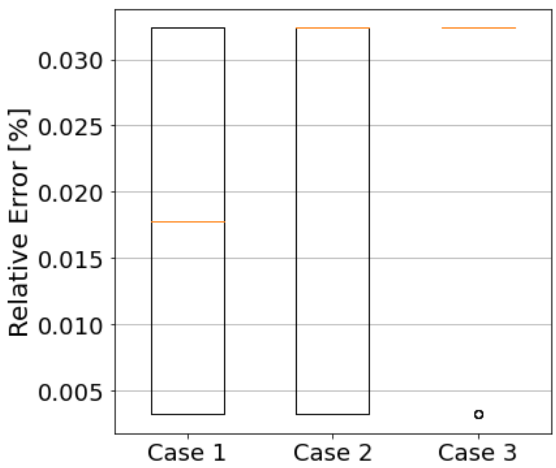
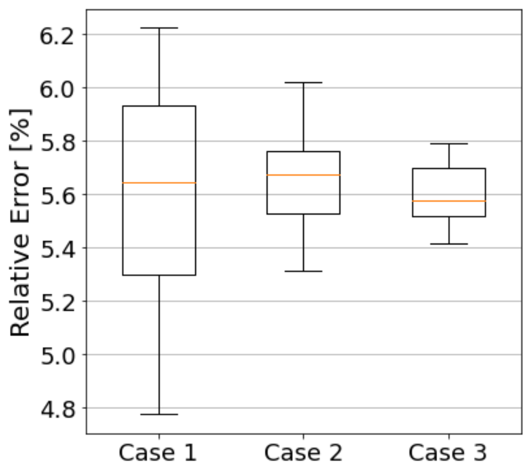


Figure 5.1: Aggregated data for PV identification

Fig.5.1 represents the data over a day from the substation level. It can be seen that the solar capacity from the neighbourhood implemented is prevalent compared to any trend. Once the optimization has been pursued we have the following boxplots :



(a) HP identification for scenario 1



(b) PV identification for scenario 1

Figure 5.2: Boxplots for the results of Scenario 1

Fig.5.2 presents the results from the PV identification and the HP identification. The results show the relative error of the estimation of the HP units in Fig5.2a and of the PV yearly energy produced in Fig5.2b. Fig.5.2a highlights that the optimization has a very low error in the estimation of the number of HP units. It identifies at most one heat pump in the mix in the worst result. The relative error interval is first within [0.0025%,0.035%] in case 1 and no relative error is witnessed. The HP identification can be considered very accurate in this case as the relative error is below 1% and the values are proved to be reproducible along the iterations, therefore proving also that the algorithm is precise.

As for Fig.5.2b a trend can be noticed concerning the noise implemented in the data set. The accuracy of the PV identification is in the interval of [94%,94.7%]. The relative error interval is first within [4.78%,6.2%] in case 1 and then [5.4%,5.8%] in case 3. The interquartile range is in case 1 [5.3%,5.9%]

and case 3 [5.5%,5.7%]. There is a decrease of 50% of the interval between cases 1&2 and then a decrease of 50% again for the range of the relative error between cases 2&3. The reduction of relative error is linear as noticed in Chap.4 proving that the linearity is conserved regardless of the number of assets. Moreover, there is very little error in the estimation of PV.

From the DSO's perspective, the results for the PV estimation are much more accurate in the case of a large quantity of PV panels implemented. As the range of the relative error values is not spread as in the Validation Scenario 1, the accuracy is enhanced thanks to the stronger profile of the asset in the aggregated data from the power substation. Hence the flexibility estimation can be pursued without large values of relative error. The high accuracy of the model allows the DSO to have an accurate estimation of the available flexibility helping with the reduction of the load on the power substation. The amount of flexibility that can be used consists of a range of [2%,157%] of the aggregated data's peak load that can be mitigated, on average 90% of the peak load could be alleviated using solar energy. Such an assumption requires a battery to be used to mitigate the peaks in consumption. Otherwise, the solar could be curtailed to avoid congestion in the grid.

5.2. Scenario 2

In this scenario, the HP units are considered to be the only asset in the grid. Therefore, the model should identify the HP energy consumed over the year using this model. In this perspective, the HP identification cannot be mitigated by other types of assets. 3089 units are being implemented in the data from the substation. The method to identify is the one presented in Alg.4. The goal is to identify the amount of HP and PV implemented into the mix, the accuracy is measured through the relative error concerning the number implemented. The range of accuracy values will be evaluated to gain more insights into the flexibility estimation. As heat pump behaviour can change from one household to another, a noise variation is introduced in the consumption data from the HP using the methodology introduced in Sec.3.2.6. The different values correspond to a proportional number to the standard deviation computed from the HP consumption data. You can see below the graph where the asset needs to be identified :

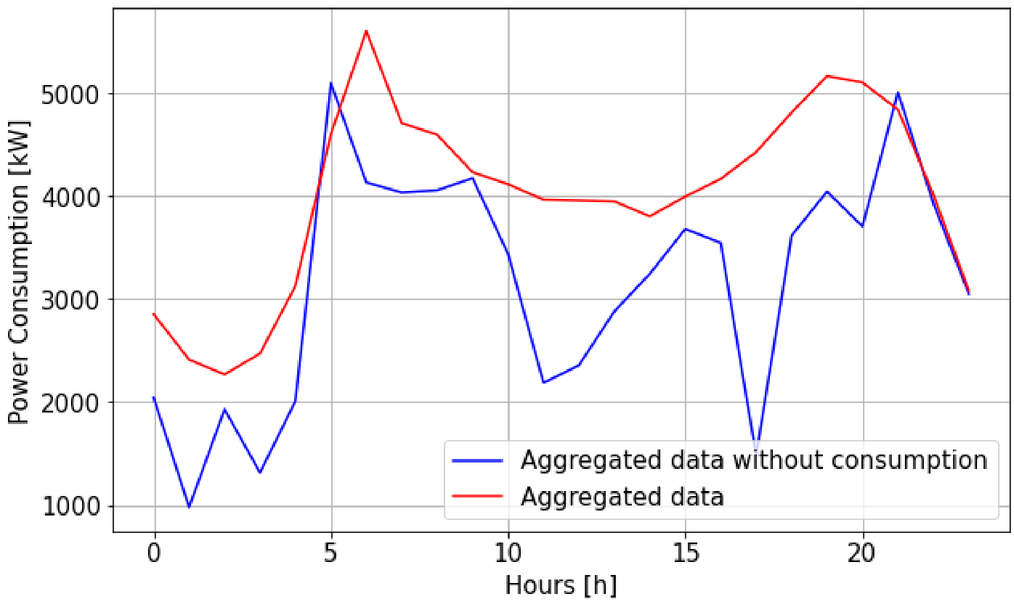
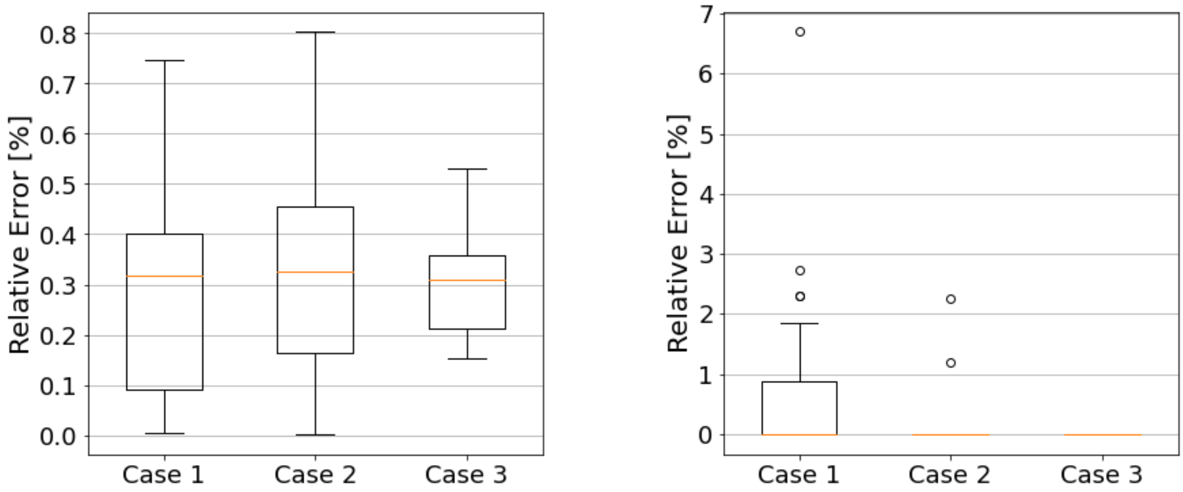


Figure 5.3: Data used for the HP identification

Therefore, the heat pumps can be identified in the mix and no solar should be observed in the mix. The results for the different noises implemented can be seen below :



(a) HP identification for scenario 2

(b) PV identification for scenario 2

Figure 5.4: Boxplots for the results of Scenario 2

Fig.5.4a presents the results for the identification of HP in three different cases. It can be seen that the relative error is below 1% regardless of the noise in Fig.5.4a. The relative error interval is first within [0%,0.75%] in case 1 and case 3 in [0.15%,0.52%]. The interquartile range is in case 1 [0.1%,0.4%] and case 3 [0.2%,0.35%]. Therefore the identification of heat pumps is achieved with great accuracy reaching a level of 99%. Moreover, the range in which the relative error values are contained is narrow which underlines a high level of accuracy. Such results allow the DSO to use flexibility measures to alleviate load from the power substation level with very little error.

Fig.5.4b represents the relative error for the identification of solar in the mix. The relative error grows in this case with the increase in the noise implementation. As no solar is in the mix, the noise can be responsible for the false guess of solar. The relative error interval is first within [0%,2%] in case 1 and then 0% in case 3. Further increases in the relative error are expected if the noise is increased. The error of the pv identification in the foreseen cases shows relative errors within the DSO error margins. Hence, in this scenario, the DSO would be able to use this model as it can be considered accurate with an error below 2%.

The identification of only heat pumps in this scenario is achieved with high accuracy as the spread of the values is narrow and the relative error is low. However, the noise in the data can introduce errors in the identification of PV. The DSO with this model can alleviate an amount of [5.9%,97.65 %] of the peak load from HP units, on average 76.7% of the HP peak load can be curtailed.

5.3. Scenario 3

In this scenario, both of the assets are implemented in the mix in the quantity previously used Sec.5.1 and Sec.5.2. The problem in this case is that the different profiles can mitigate each other. Therefore the different methods and order of steps are important. The identification will follow the procedure proposed in method 1, which corresponds to the method introduced in Alg.4. The optimization process is pursued over the whole year without distinction to specific periods. The goal is to identify both assets with high levels of accuracy. The accuracy will be evaluated through the relative error taking as objectives 3089 HP units and for solar 2.05 GWh. It will also investigate the spread and the interval of the relative error. The data profile used for the mixed identification can be seen below :

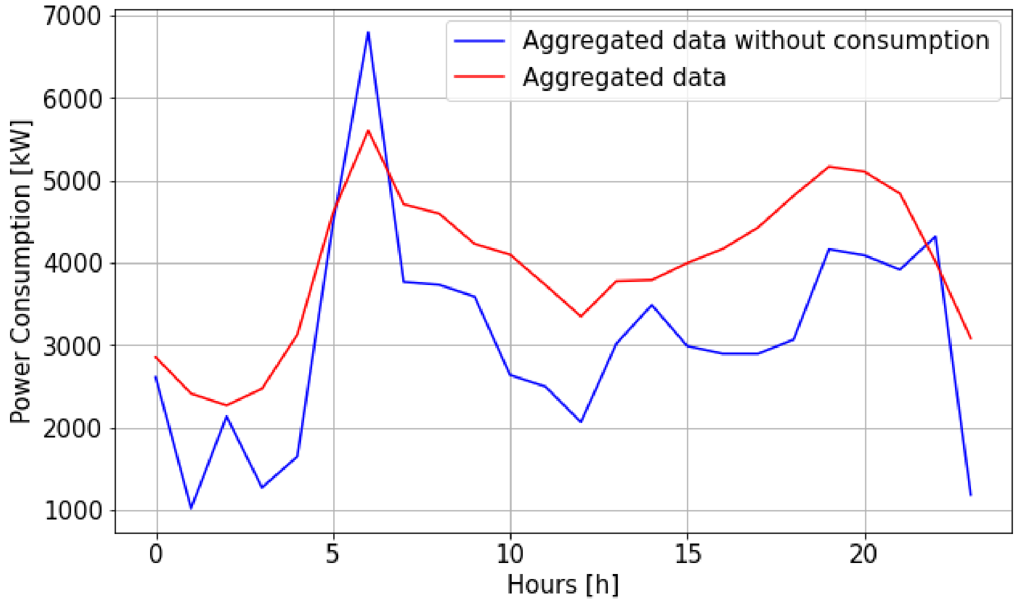
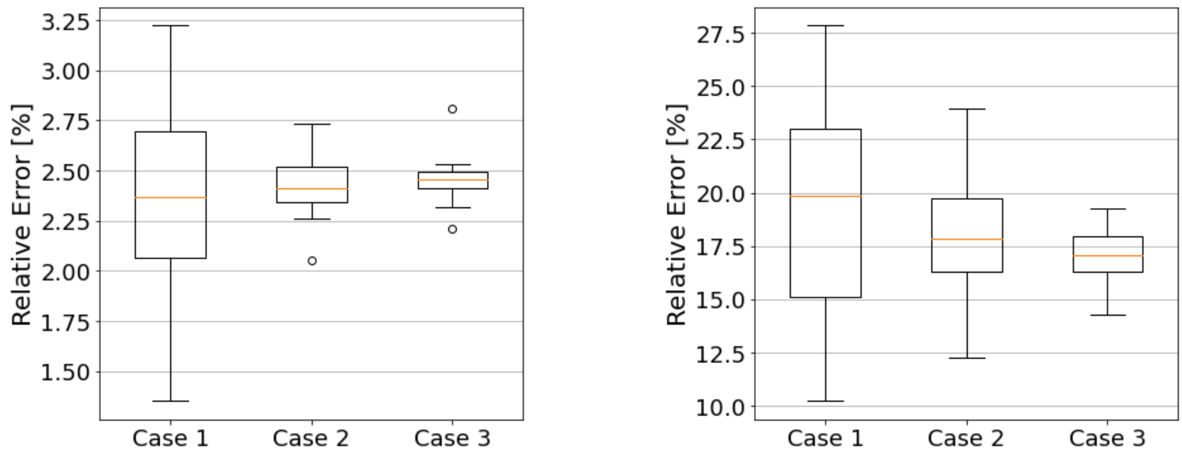


Figure 5.5: Data used for mixed identification

Fig.5.5 represents the profiles for the first day of the year. The results for the identification of the two assets in the profiles are presented below :



(a) HP identification for scenario 3

(b) PV identification for scenario 3

Figure 5.6: Boxplots for the results of Scenario 3

Fig.5.6a shows the result for the estimation of heat pumps in the mix, it can be noticed that the relative error decreases as the noise witnesses the same trend. The accuracy for the algorithm ranges in the interval [97.3%,97.8%]. The relative error interval is first within [1.25%,3.25%] in case 1 and [2.30%,2.55%] in case 3. The interquartile range is [2.1%,2.7%] in case 1 and [2.4%,2.5%] in case 3. The reduction of the spread between cases 1&2 corresponds to a reduction of 80% of the interval size and between cases 2&3 a reduction of 50%. This underlines that with multiple assets connected the linear relationship of the noise is not respected. The interquartile range of case 1 seems to cover the total relative error of case 3. It underlines that the guesses within the interquartile range of case 1 are rather correct than the ones outside of the interquartile range. This can help the DSO estimate the available flexibility with better accuracy. The accuracy is greatly improved, as the spread of the values is considerably reduced with the decrease of noise in the data. The accuracy of the HP identification is quite high and can help the DSO estimate the number of HP units with little relative error. Hence such results can allow accurate alleviation of the load on the substation level.

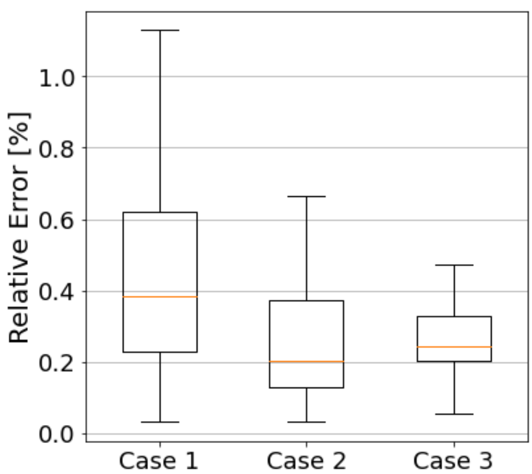
In Fig.5.6b the estimation of PV energy covers a broad range of relative errors. However, the error in the estimation of PV can reach up to 27.5%. The accuracy for the algorithm ranges in the interval [72.5%,90%]. The relative error interval is first within [10%,28%] in case 1 and then [14%,19%] in case 3. The interquartile range is in case 1 [15%,23%] and case 3 [16.5%,18%]. As observed for HP, the interquartile range of case 1 covers the total relative error of case 3. However, in Case 1 in Fig.5.6b the relative error is above 20% of relative error for the estimation of the energy produced by PV. Therefore, using the full range would not be accurate for the DSO. Only keeping the values within the interquartile range of large levels of noise would then grant more accuracy to the estimation of the asset in the mix. Such a phenomenon allows the DSO to get rid of errors in the case of a high level of noise implemented in the data, allowing the relative error of case 1 to be reduced. The DSO, which usually works with about 20% of relative error in their estimation of flexibility, would be able to use the results in most of the cases. Under the assumption that the relative error outside of the interquartile range only consists of the wrong estimations, the PV estimation would be more accurate for case 1. Therefore the DSO could use only the interquartile range in cases of higher levels of noise.

With such results [5.7%,95.5%] of the aggregated data's peak load could be curtailed from HP, translating into an average of 75% of the peak load reduction. As for PV, the peak load can be mitigated by [0.25%,129%] and on average a reduction of 30% of the peak load can be witnessed if a battery is considered or with curtailment mechanisms.

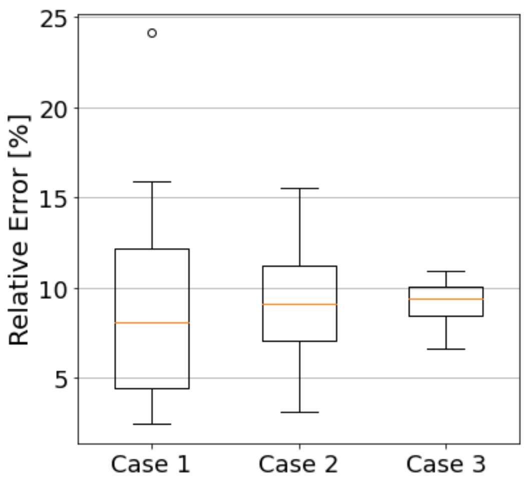
Method 1 has been proven to work in the conditions proposed in the methodology in Sec.3. However as seen in this section, certain conditions can pose a serious threat to the accuracy of the algorithm. Different methods will be investigated in the following sections.

5.4. Scenario 4

In this scenario, both of the assets are implemented in the mix in the quantity previously used Sec.5.1 and Sec.5.2. The problem in this case is that the different profiles can mitigate each other. Therefore the different methods and order of steps are important. As opposed to method 1, the heat pump identification is done by filtering the sample where solar is produced. The identification will follow the procedure proposed in method 2, which corresponds to the method introduced in Alg.5. The accuracy will be evaluated as previously done in Sec.5.3. The optimization takes place for the whole year similarly to previously in Sec.5.3. The goal is to identify the two assets using different time lapses for their identification.



(a) HP identification for scenario 4



(b) PV identification for scenario 4

Figure 5.7: Boxplots for the results of Scenario 4

Fig.5.7a represents the accuracy for the HP identification using method 2. It can be seen that the accuracy is higher with a lower level of noise. The accuracy is in the range of [99.3%,99.9%]. The relative error interval is first within [0.05%,1.15%] in case 1 and [0.1%,0.45%]in case 3. The interquartile range is [0.22%,0.62%] in case 1 and [0.2%,0.35%] in case 3. As the level of noise decreases, the values are less and less spread. The level of accuracy is then excellent and therefore is enhanced compared to the previous results of the accuracy from method 1. It can be seen in the figure that the relative error has diminished compared to Fig.5.6a, therefore method 2 is highlighting better accuracy than previously. The accuracy is improved and enhances the estimation of the flexibility from the DSO.

Fig.5.7b shows the relative error in the asset estimation as well as the accuracy levels for the PV identification. It can be noticed that the noise implementation impacts these values. The accuracy in the solar estimation is within [87.5%,96%]. The relative error interval is first within [2.5%,16%] in case 1 and [7.5%,11%]in case 3. The interquartile range is [4.5%,12%] in case 1 and [8%,10%] in case 3. Method 2 reduces the relative error of the estimation of PV by an average of 8% compared to method 1, therefore improving the accuracy of the model. Considering an 8% difference in the guess of solar, the DSO would be able to use 164 MWh that was previously not used. An interesting behaviour can be noticed, the median of the interquartile range seems to be increasing along the cases. Such a phenomenon can come from the overestimation at a higher level of noise of the number of heat pumps, therefore leading to an increase in the peak power of the solar peak. The PV identification is very sensitive to such parameters as there is only one daily peak of solar energy. As the following scenarios all use this method, similar behaviour can be noticed. With this improvement, the DSO is allowed to provide more flexibility to the grid and ease the congestion on the grid. As the relative error is below 20% regardless of the case, the DSO can use method 2 to estimate the available flexibility. If the range of only the interquartile range is used for estimating the available capacity the accuracy reaches at least 87.5%.

With such results [5.9%,97.6%] of the aggregated data's peak load could be curtailed from HP, translating into an average of 76.7% of the peak load reduction. As for PV, the peak load can be mitigated by [0.25%,135%] and on average a reduction of 32% of the peak load can be witnessed if a battery is considered or with curtailment mechanisms.

The method proposed in this section proves to be efficient and accurate. Compared to the previous method the relative error in the estimation of both assets has decreased leading to an increase in the accuracy. The flexibility estimations have gained accuracy compared to method 1. The interquartile range of case 1 contains the total range of the relative error of case 3. Therefore by estimating the interquartile range of case 1, the relative error can be reduced as method 2 has highlighted that the errors outside of the interquartile range of case 1 do not account for the relative error in case 3. The DSO can then only take the interquartile range of case 1 to determine the amount of available flexibility for the cases with high levels of noise.

5.5. Scenario 5

As opposed to Sec.5.3 and Sec.5.4, the goal of this identification process is to identify the period where solar and heat pumps are prevalent assets to evaluate with a higher level of accuracy the assets. In this scenario, the period of the chosen asset will cover 6 months. The accuracy will be evaluated as previously done in Sec.5.3. The same quantity of solar and heat pumps are used. The goal is to optimize where the assets are predominant, as shown below:

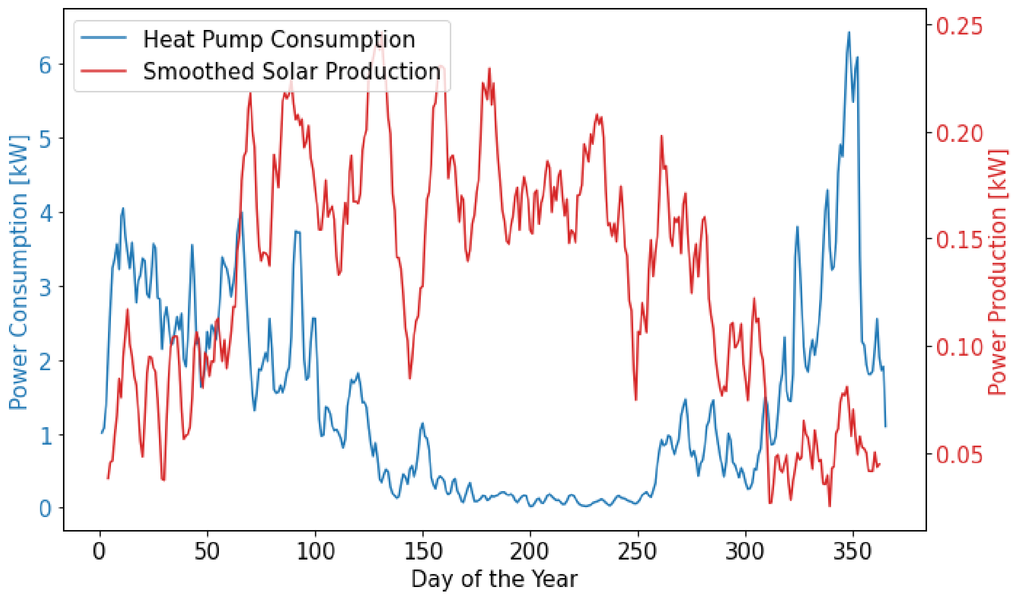
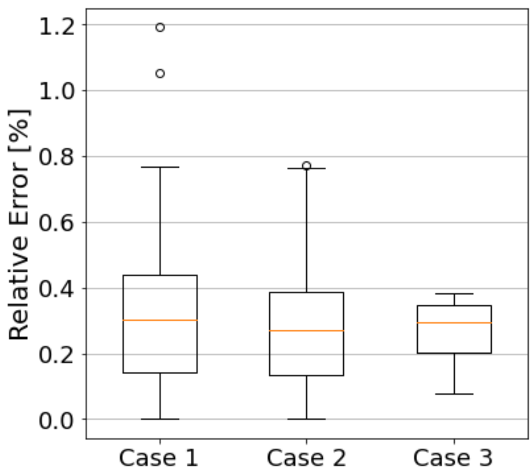
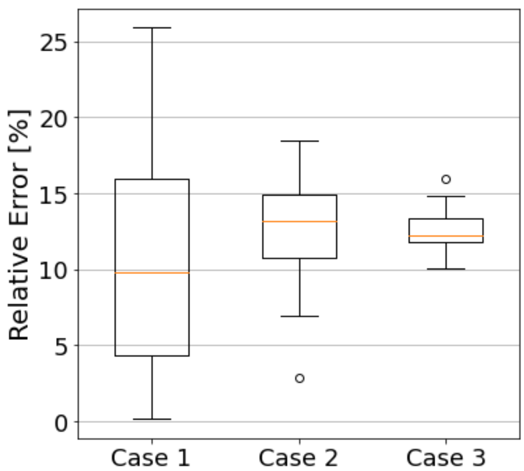


Figure 5.8: Daily Averages of Heat Pump Consumption and Solar Production

Fig.5.8 highlights the trend that can be seen from solar production and heat pump consumption. Therefore a specific period can be used for the identification of both assets. One corresponds to 6 months where solar is the predominant one and the other way around. The goal is to identify the two assets using different time lapses for the optimization. The amount of PV panels is identified during the summer when its influence is the most important, and the HP units are identified in the winter.



(a) HP identification for scenario 5



(b) PV identification for scenario 5

Figure 5.9: Boxplots for the results of Scenario 5

Fig.5.9a represents the relative error for the HP identification using method 2 for the identification process. It can be seen that the accuracy is higher with a lower level of noise. The accuracy is in the range of [99.3%,99%]. The relative error interval is first within [0%,0.75%] in case 1 and [0.1%,0.38%]in case 3. The interquartile range is [0.18%,0.43%] in case 1 and [0.2%,0.35%] in case 3. The accuracy improves as the relative error interval narrows. The method of the prevalent period does not impact the relative error of the HP identification. The accuracy of the HP estimation is high and allows the DSO to determine accurately the amount of available flexibility.

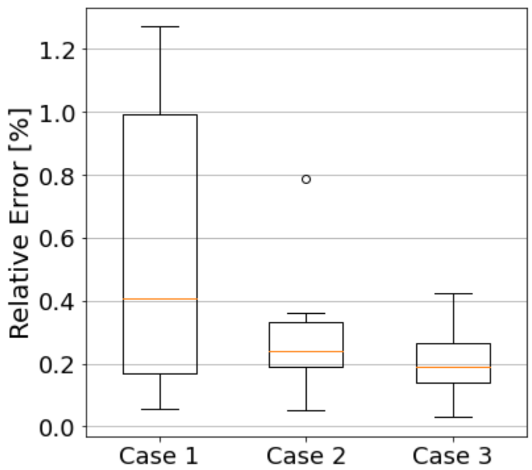
Fig.5.9b shows the accuracy levels for the PV identification. The accuracy in the solar estimation is within [87.5%,95%]. The relative error interval is first within [0%,26%] in case 1 and [10%,15%] in case 3. The interquartile range is [4.5%,16%] in case 1 and [12%,13.5%] in case 3. By using method 2 with a specific period, it can be seen that the relative error ranges outside of the 20% range. Therefore this scenario is less accurate for the estimation of PV in the aggregated data, as only half of the data from the year is used, it could be interesting to pursue the identification on the same period over years if data were available. Few conditions make this method tedious to implement as it requires to know if additional PV has been installed and if data are available from different years. However, in this case, the DSO can use the algorithm in the case of a low level of noise to keep the accuracy above 80%.

By using this method, [5.9%,97.6%] of the aggregated data's peak load could be curtailed from HP, translating into an average of 76.7% of the peak load reduction. As for PV, the peak load can be mitigated by [0.25%,129.5%] and on average a reduction of 30.5% of the peak load can be witnessed if a battery is considered or with curtailment mechanisms.

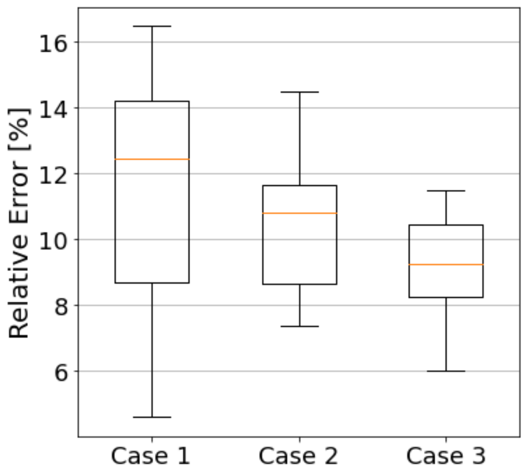
In this scenario, the use of a specific period for the identification of the assets has proven to be less accurate than method 2 using the data of the entire year. The flexibility estimation has been hindered and therefore is not advised to be used by the DSO. Method 2 as seen in Sec.5.4 is the best one yet, the number of iterations of the model will be studied in the next section.

5.6. Scenario 6

In this scenario, the goal is to evaluate the accuracy as the number of iterations increases. Three different steps will be investigated, for 10,30 and 50 iterations of the algorithms using method 2 as seen in Sec.5.4. The accuracy of the algorithm is assessed throughout the number of iterations that the algorithm is going through. The accuracy will be evaluated as previously done in Sec.5.3. The first set of results is the outcome of 10 iterations, then 30 iterations and finally 50 iterations.

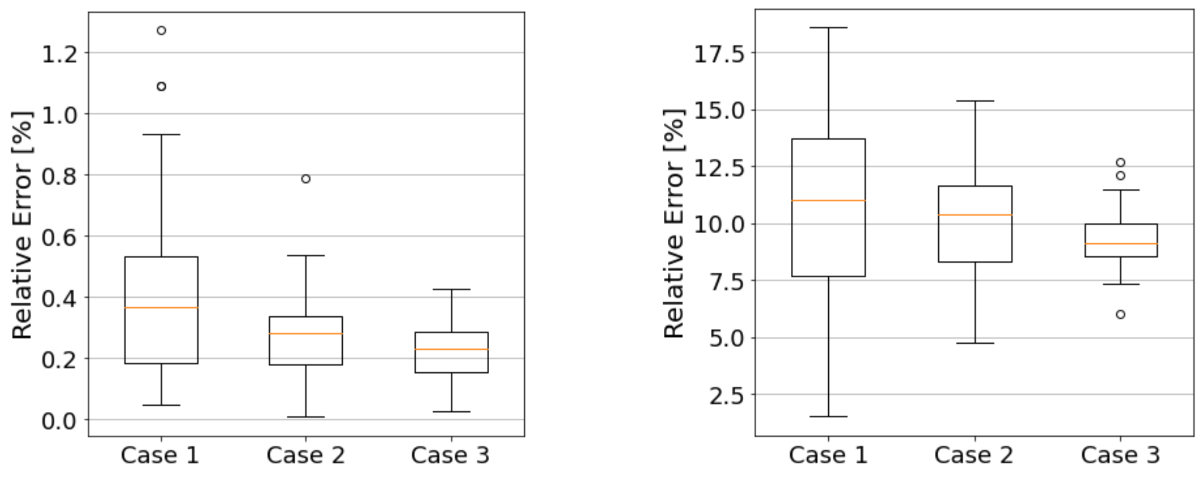


(a) HP identification for scenario 6



(b) PV identification for scenario 6

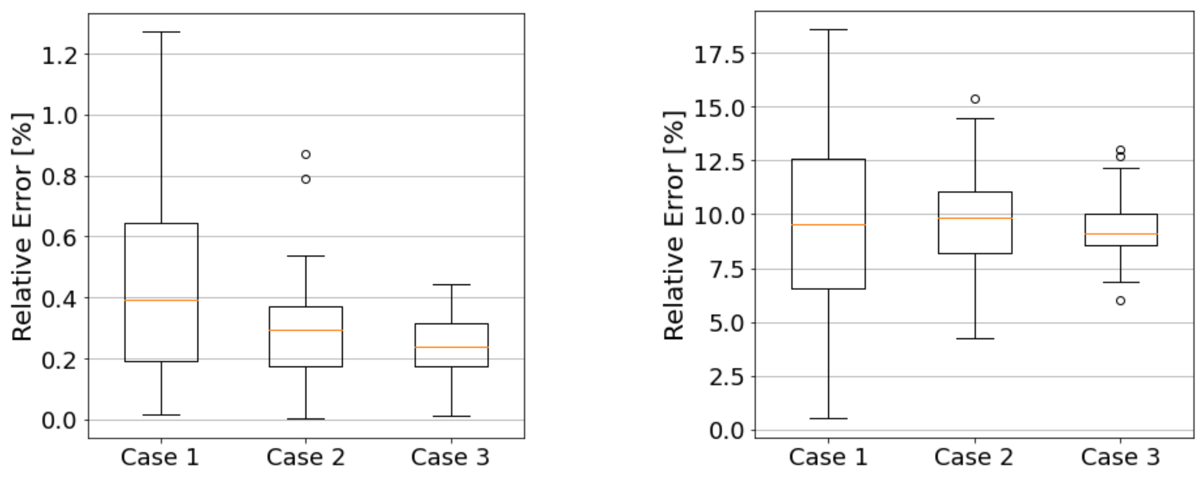
Figure 5.10: Boxplots for the results of 10 iterations in Scenario 6



(a) HP identification for scenario 6

(b) PV identification for scenario 6

Figure 5.11: Boxplots for the results of 30 iterations in Scenario 6



(a) HP identification for scenario 6

(b) PV identification for scenario 6

Figure 5.12: Boxplots for the results of 50 iterations in Scenario 6

The accuracy seems to be improving over the different iterations for HP in Fig.5.10a, Fig.5.11a and Fig.5.12a. However, the most impressive evolution can be found in the PV identification results. Over the evolution that can be witnessed from Fig.5.10b, Fig.5.11b and Fig.5.12b, the accuracy from cases 1 and 2 is increasing notably. For 10 iterations in Case 1, the median for the accuracy is 12.5% of error and the interval covers a range of error of [9%,14%]. After 50 iterations, Case 1 has a median of 9.5% of error and an interval for the relative error of [7%, 12.5%]. Whereas for case 3 the accuracy is increased over the iterations but with smaller variations, for 10 iterations the relative error is within [6%,11.5%] and after 50 iterations it is in [7%,11.5]. Using the interquartile range of case 1 to determine the total range of error reduces drastically the relative error for the PV identification.

For the DSO, the number of iterations can only improve the identification's accuracy and therefore enhance the flexibility estimation. With this method, the DSO can provide accurate results and allows to minimize the relative error in their guess. The flexibility estimation can lead to alleviating [5.9%,97.6%] of the aggregated data's peak load from HP, translating into an average of 76.7% of the peak load reduction. As for PV, the peak load can be mitigated by [0.25%,133%] and on average a reduction of 31.5% of the peak load can be witnessed if a battery is considered or with curtailment mechanisms.

6 Sensitivity analysis

In this chapter, the different scenarios are evaluated and the important aspects of the optimization problems are discussed. The different subsections analyse the individual asset identification method in Sec.6.1 and then the two asset identification methods in Sec.6.2. Finally, in Sec.6.3 the different parameters are all compared to produce a conclusion from the model framework.

6.1. Individual asset identification

In Sec.5.1 and Sec.5.2, the assets are identified individually. The outcome for PV identification is accurate, the interval of the values of the relative error narrows as the level of noise decreases across cases. Such results correspond to the projections on which the estimation of assets is highly dependent on noise implementation. The median of the results is closer to the lower boundary as the noise is decreased. The interquartile range is also highly dependent on the noise implemented in the value. However, in the case of solar, the noise arises from the variability of consumption. Therefore, if stable consumption is observed, the accuracy of solar identification will only increase. Almost no heat pumps are identified during the identification of PV panels. To determine the full range of relative error, it has been seen that in the case of a low level of noise, the relative error full range corresponds to the interquartile range of a higher level of noise. Hence, DSO can use specifically the interquartile range at high levels of noise to estimate the flexibility. It will enhance the accuracy of the estimation.

In the other case, where heat pumps are the only asset present in the data, the accuracy is excellent, resulting in very little error from the algorithm. However, with noise implementation in the data, over-estimation in the case of HP can lead to errors in the identification of PV. The error remains low, but a higher level of noise could lead to a false estimation of solar. Therefore leading the DSO to errors that would hinder their flexibility mechanisms.

Overall, the accuracy of both PV and HP reach a confidence interval of at least 94%, which is excellent for a system operator, given that a DSO can operate with errors up to 20%. The accuracy criterion is highly dependent on the number of assets implemented into the mix, as seen in Chap.4 where the PV identification can reach an error of 100% in the worst-case scenario. The spread of the values is tremendously reduced as the number of assets is increased, the accuracy is then enhanced and the estimation more robust.

The next section digs into the results of mixed identification concerning the different scenarios.

6.2. Multiple asset identification

In Sec.5.3, Sec.5.4,Sec.5.5 and Sec.5.6, the assets are identified within the aggregated data. The outcome of the mixed identification using the first method in Sec.5.3 provides accurate results for the heat pumps with less than 2.75% error. However, an error of 2.75% can induce a wrong estimation in the solar profile. As the average consumption of heat pumps over a year is 4493 kWh [66], the low percentage of error can lead to a significant error in fitting the solar profile. To address this, method

2 was developed to reduce the error in solar estimation. As highlighted in Fig.5.6b, the relative error can reach up to 28%. By implementing method 2, the accuracy for solar improved by almost 10% on average, as seen in Fig.5.7b. Although the accuracy of the heat pump estimation improved by only 2% (Fig.5.7a), it led to a major improvement in solar identification. The identification of assets in the mix depends on when an asset can be identified. Identifying the specificities of the different profiles such as the number of peaks, amplitude and peak time allows the identification to gain accuracy. The criteria that allowed the identification of PV to reduce the relative error was to identify the heat pump consumption independently of the solar influence. The error arose from the peak of solar energy production mitigating the first peak of heat pump consumption. Therefore, the minimization process had to deal with a mismatch between the two peaks, resulting in a decreased number of heat pump units. Identifying assets in aggregated data must consider the specific profiles to improve the accuracy of the developed method.

To enhance the model's accuracy, the period during which the identification process takes place was investigated. Throughout Scenario 5, the method aimed to take a specific time slot where the assets would be prevalent, as seen in Fig.5.8. The solar energy curve is higher than the heat pump consumption curve from approximately the 100th day until the 280th day. Therefore, the identification process for solar occurs during this period, while the heat pump identification process takes place during the complementary period. The results in Fig.5.9 show better accuracy than those in Fig.5.6. However, the process in Scenario 4 remains better than that in Scenario 5. In Scenario 5, the accuracy is on average between [86%, 90%], while for Scenario 4, it is between [91%, 92.5%]. Choosing a specific period for identifying the different assets yields accurate results, but the period may need to be more specific to provide even more accurate results. Another aspect of the period method can be the number of samples available to the DSO, for this research work the data for one year were considered. If data from several years are available to the DSO and the PV installation has not been modified over the years, the accuracy can eventually improve and unlock better flexibility estimation.

As the previous paragraph investigates the dependency on the chosen period, this section explores the influence of the number of iterations on the spread of the relative error values. Since the method proposed in Scenario 4 is the most efficient, the number of iterations was investigated using its algorithm. In the results of Scenario 6, the number of iterations proves to be an important factor in increasing both the accuracy of the median and one of the relative error values. In Sec.5.6, the accuracy for a high level of noise improves with 50 iterations. Therefore, considering the high level of noise in the data, the algorithm requires more iterations to achieve the same level of accuracy as achieved with a low level of noise. The algorithm performs well with a low level of noise with fewer iterations, while a higher level of noise requires more iterations to achieve the same performance. The spread of the relative error values was evaluated throughout the evolution of the interquartile range and the relative error along the different scenarios. The boundaries of the interquartile range are being narrowed as the number of iterations is increasing.

The next section will combine the different findings from the previous chapters, to provide analysis on the overall process.

6.3. Comparisons and reflection

When looking at the results, the KPI has seen an increase along the different scenarios. The first step that improved the performance was to introduce a specific time to optimize the algorithm. Such a process can be helpful in various conditions, prior to the implementation of the right period in the optimization process, an analysis of the different characteristics of the assets is essential. For the solar profile, different parameters are important, first of all, the irradiance at the moment of power produced can be used as a filter to increase the accuracy of the algorithm. Furthermore, no positive influx in the profile can be possible therefore different conditions are lining up. As for the heat pumps, two peaks exist in the profile. The problem lies in the combination of assets, to identify behaviours there is a need to have windows where no influence of the other asset is degrading the profile. Once these conditions are fulfilled, different parameters such as the period of identification and the number of iterations can be investigated.

For the period on which the identification takes place, the optimization has limited samples. The KPI is better than in method 1 but is still lagging behind the results of method 2. One reason for this could be the lack of data available for the optimization to take place. The optimization took place over 6 months with an hourly resolution. Further investigation with the same period over several years could eventually improve the results. However, this process would require the DSO to have data from the past years. The different data required for the optimization would make this option tedious to accomplish.

As for the number of iterations, the accuracy of the results is increasing consequently when a higher number of iterations is pursued. However, for low levels of noise such as in Case 3, the results are not improving after 20 iterations. Hence the number of iterations shall be considered higher in the case of a high level of noise in the data. Such a process would guarantee the validity of the result for the DSO.

From the evolution of the median and the interquartile range, it can be noticed that scenarios 4 and 6 work the best for the results. The interquartile range in scenarios 4 and 6 is already highlighting the area of relative error in which the relative error in case 3 is located. Given the trend, the results then can be estimated from the interquartile range at a higher level of noise. The same trend can be noticed with case 2, most of the results in case 3 belong to the interquartile of case 2. In addition, the results that scenario 6 brings are that after a certain number of iterations, the results converge to a certain range. In that case, the interquartile range becomes more precise for solar.

Another important factor in the model is the proportion of the different assets compared to the consumption profile that can be derived from the MFFBAS profile. In Chap.4 the PV profile can be considered hindered by the consumption profile, whereas in Chap.5 the PV is much more important. It can be seen that the relative error is decreased and allows the identification to be more accurate. Reduced HP and PV implementation amounts would result in a significant loss of accuracy in asset identification. The consumption profile would then be proportionally more important than at higher levels of assets implemented. At a low number of assets, the accuracy is dropping rapidly. DSOs should then be careful of the use of the model in the case of a small number of assets connected to the grid.

6.4. Conclusion

In this chapter, the first section has validated the model which allowed the scenarios to be explored. Several scenarios have been tested and have highlighted in which cases the best accuracy was achieved. The accuracy is evaluated throughout the range of the values covered by the relative error and by the median of the different cases. A table with the results can be found below :

	Median PV[%]	Interval PV[%]	Median HP[%]	Interval HP[%]
Scenario 1	5.55	1.4	0.02	0.030
Scenario 2	0	2	0.03	0.75
Scenario 3	20	17.5	3.32	2.9
Scenario 4	9	14	0.4	1.1
Scenario 5	10	25	0.3	0.78
Scenario 6	9	17	0.4	1.3

Table 6.1: Table for the result of Case 1

	Median PV[%]	Interval PV[%]	Median HP[%]	Interval HP[%]
Scenario 1	6.7	0.7	0.03	0.030
Scenario 2	0	0	0.03	0.8
Scenario 3	18	11.25	3.4	0.5
Scenario 4	8.5	12.5	0.2	0.65
Scenario 5	13.5	10	0.18	0.78
Scenario 6	10	10	0.3	0.55

Table 6.2: Table for the result of Case 2

	Median PV[%]	Interval PV[%]	Median HP[%]	Interval HP[%]
Scenario 1	5.6	0.4	0.03	0
Scenario 2	0	0	0.3	0.34
Scenario 3	17	5	3.45	0.3
Scenario 4	9	4	0.15	0.4
Scenario 5	12.5	5	0.3	0.3
Scenario 6	9	5	0.25	0.4

Table 6.3: Table for the result of Case 3

High levels of accuracy have been witnessed in the case of a high number of iterations and specificity in the period of optimization. The best results have been obtained in scenarios 4 and 6 for the mixed identification model. In the case of one asset implemented, scenarios 1 and 2 provided as well high-level performances.

7 Conclusion & Discussions

In this chapter, the limitations of the model built along this thesis will be detailed in Sec.7.1. Then the recommendations for future work will be proposed in Sec.7.2. Sec.7.3 will propose different advice and considerations regarding the problem's owner perspective being the DSO. Finally, a reflection will take place in Sec.7.4.

7.1. Limitations

The model that has been built in this work has different problems and limitations regarding its use. The different issues with the model are addressed regarding the scale on which it can be applied. Then the limits of the models are detailed with respect to the type of asset that can be used. Furthermore, the sample size of the data is discussed. Following, the limits of the method used for identification are exposed and then address the problem of the number of assets.

Even though the model was tested in the case of identification for solar and heat pumps, with promising results, the different assumptions may differ in the case of different demand response mechanisms. Hence the results would highly fluctuate in such cases. The consumption curve that was considered can fluctuate depending on the different types of houses and insulation labels. Therefore such a model can only be applied to residential neighbourhoods, as the consumption from commercial buildings has not been investigated. The model has not been tested with different profiles of consumption, and any options out of the frame defined for the optimization case have not been investigated.

The assets that can be used in the developed model are solar panels and heat pumps. Different assets such as wind turbines, and EVs have different profiles that require in-depth analysis and can be investigated in the future. The different specificities of the profiles of solar and heat pumps are key in the development of the algorithm. If different assets were to be considered in the future, their characteristics would need to be studied. The heat pump profiles were considered to always have the same pattern regardless of the different houses and commercial buildings. As for the solar profile, the configuration of the solar has not been investigated concerning the profile of energy production. As for the time resolution, it has only been coded for an hourly resolution. Most of the code is not versatile for the use of the different time steps. The different options will be further explained in Sec.7.2.

The method for the optimization process was always done separately, one asset after the other. Different optimization methods would be interesting to consider to solve such problems. Furthermore, no other type of algorithm has been compared to this method. Therefore paving the way for different options detailed in Sec.7.2.

The method that has been used for mixed identification is seen to have lower performance when a small number of assets are implemented. The accuracy can drop drastically, such events are however to be expected. When the number of assets diminishes, the profile from the selected asset becomes smaller and has less power amplitude. Therefore, it makes the identification tougher than at a high level of assets implemented where the characteristics are clearly distinct. The mixed process of identification is struggling to take into account these differences.

7.2. Recommendations and future work

In the future, different upgrades for the model can be considered. Different assets interesting to incorporate in the model are proposed. Then the different sample sizes that are worth being investigated are detailed. Finally, the different options for identifying flexibility are displayed.

Optimization has been the chosen method in the above research work. However, as many assets are to be considered in future work, it would be interesting to work on an equivalent model using machine learning. Machine learning can handle more complex data using several inputs for each type of asset. It is a promising approach to solving the problem that can deal with multiple inputs, yet the accuracy would have to be compared to the developed model.

In this model, the hourly format of the data was investigated as the variations on an hourly format would be less visible than on a smaller scale. The question remains on a smaller scale whether the method proposed above would still work, given that the different peaks of utilities would not be invisible anymore. Different works have been pursued in the literature identifying assets individually on small sample size, investigating such options would be interesting in the future. Methods have been considered as well in this research work such as simultaneous optimization of the two assets, which could be interesting in case of smaller intervals.

Besides, it would be interesting to make the model more versatile. In the future, different assets will be eventually considered flexible utilities, in that perspective, it would be interesting to create a function that would track and identify the key aspects of the considered profile and that would make constraints around it. Second, it would be interesting to allow the model to work with different types of resolution (hourly, daily, 15-minute intervals etc). The different assets would then need to be in the correct format and not interpolated. For solar and heat pumps the curves would need to have a higher level of detail if smaller time intervals were to be considered.

Future work shall consider methods to identify the assets even when low numbers of them are present in the data. Different methods shall be investigated to make the identification more robust. Machine learning can be considered to reduce uncertainty in the identification process.

7.3. DSO's perspective recommendations

From the DSO perspective, this model can help to estimate the capacity of the assets present in the grid. Such work can then lead to a reduction of the congestion in the area covered by the operator. Different advantages can be found in the short-term, and long-term for the DSO.

In the short term, the load could be consequently reduced during peak hours and the heat pump consumption could be shifted when there is solar energy produced. Hence the load would be decreased when power is needed the most. DSO could then electrify the heating and would not have an additional weight on the grid. By combining the two assets the electrification of the heating could be mitigated using solar energy. In terms of energy produced by solar to mitigate the peak consumption of HP, a ratio between the amount of solar and heat units shall be found. Given the previous results for an average heat pump consumption of 4493kWh, 4 solar panels of capacity 1kW would be needed to compensate for the energy consumption.

In the long term, knowing how many solar panels are present in the grid would be a major contribution to the grid operator. By forecasting how much power can be produced, the load on the system operator would be reduced. Therefore, allowing more reserve capacity to be available in the case of a surge of power in the grid. By estimating the number of heat pumps, the peak consumption along the year would be determined much clearer allowing the DSO to be more prepared for higher demand for power. Furthermore, combining assets contributes to reducing the energy demand of a neighbourhood, reducing the losses as well, leading to optimizing the efficiency of the overall grid. Throughout the use of storage, the energy produced during the summer that has not been used could be stored as a reserve capacity. Therefore by knowing the number of solar panels producing energy in the neighbourhood, the decrease in the load can help the DSO to forecast the demand more precisely. Throughout the use of storage, the fluctuations could then be mitigated to have a load as constant as possible.

Furthermore, having a constant load implies that the need for backup generators will be reduced. The emission of GHG can then be mitigated, yet solutions shall be found to store the energy. Moreover, by stabilizing the energy demand the energy price will be more constant, hence creating less costs for the people living in the neighbourhood. Thus allowing more incentives for people to invest in VRES. Such a process will only foster a better climate for people to adopt behaviours prone to facilitate the energy transition.

7.4. Reflection and contribution

This work contributes to the research on the estimation of flexibility. The creation of a model that can identify the different assets available in the grid grants system operators more tools to evaluate the state of the grid, and how to cope with it. It also reduces the cost that can create the smart metering of a neighbourhood. The development of such a model proves that identifying assets in the data from the transformer level can help with the estimation of flexibility. However, this work is limited to the identification of both solar and heat pump units. Considering the number of flexibility sources endless, the challenge remains in identifying them. The development of the model contributes to paving the way for future models aiming to identify multiple flexible options. Nevertheless, the more options included in the grid the more complex the problem will become. Such paradox shall be kept in mind when developing a similar model.

7.5. Conclusion

This research work aims to estimate the flexibility in distribution systems by identifying the assets. Flexibility can play an important role in the future by diminishing the load that DSOs have to manage. It can help operators deal with congestion management as well as make the most out of the VRES, therefore helping with the integration of RES in the power grid. However, the identification of these assets remains a struggle among many. The literature review provided different options to estimate the flexibility using different measurements. Most of the work that has been yet pursued, is using active and reactive power for the identification or profiles of consumption or production. However, none of the work investigated the case of different assets to be identified. In most cases, the contribution of assets was estimated individually to grant the best accuracy. In this study, a model that can identify different assets has been developed. To the knowledge of the author, this model is one of the first that can identify two different assets in aggregated transformer measurement profiles.

The main research question of this thesis is :

”How can the identification of assets based on transformers’ measurements provide an estimation of the available flexibility in the distribution system?”

To address this problem, four different sub-research questions were developed and are answered below to build up the solution to the main research question.

First sub-research question:
”How can the estimation of heat pump profiles be more accurate ?”

Accurate heat pump profiles have been derived using the evolution of COP of HP over the years. The COP is taken from a study [66], the study provides the coefficient of heat pumps for different types of households and countries. Three different types of COP have been chosen, ASHP, GSHP and WSHP with a multi-family house use. Throughout the use of these, the heat pump consumption profile has been derived from the heat demand profile. Temperature profiles are used to provide information on when the heat pumps shall be consuming energy. The consumption profile of heat pumps was determined using such an approach.

Second sub-research question:
”Which data and steps are needed for disaggregating the profiles with a 90% confidence interval?”

Different steps are required to be followed to disaggregate profiles with a high level of accuracy. First of all, the consumption curve shall be subtracted from the data issued from the substation level. Then, the identified assets shall be studied to understand their characteristics. Based on these specificities, the identification process is adapted to them for instance with the peaks in heat pump profiles. Depending on the asset needed to be identified, different steps shall be taken to prepare the data for the identification process. For solar, the disaggregation of the noise baseline is key to reaching a high level of accuracy. As for heat pumps, the identification period is the lever that shall be used. Then once done, the identification of assets can take place.

Third sub-research question:
"How can the uncertainty of the proposed method be minimized?"

Different levers have been efficient in the minimization of the relative error. A factor that has proven to be effective is the number of iterations, if the provided data are suffering from variability, a larger number of iterations can help identify the interquartile range for large levels of noise. In the case of a high level of noise, the interquartile range of relative error corresponds to the total relative error range at a smaller level of noise. Hence, a certain level of accuracy can be achieved by identifying the specificities of the profiles and using a high number of iterations. Therefore, the identification of assets based on transformer measurements can provide an estimation of flexibility with an accuracy of less than 1% for HP and less than 12.5% of error for PV with 75% of values with a relative error below 10%.

Fourth sub-research question:
"What is the correlation between the heat consumption profile and the PV production ?"

First of all the impact that heat pump identification has on the identification of solar can highly decrease the accuracy of the optimization. To address such a problem, the identification of the number of heat pumps is taking place outside of the window of influence from the solar energy profile. This step has allowed a gain of 10% of accuracy on average. This has proven to be an important lever to enhance the process throughout Alg.5. An ideal ratio to minimize the error would need to be found to grant more insights into the KPI of the optimization.

Having answered the different sub-research questions, asset identification through optimisation can take place and proves to be an accurate and precise method for the estimation of flexibility based on transformer measurement profiles. By identifying the different assets, the flexibility can be determined based on the energy identified from these assets. Therefore, this energy through either the use of external storage or by consuming the energy produced by PV could be a way to shift the behaviour of the heat pump. The estimation of flexibility throughout the use of optimization has reached accuracy levels of 90% and can be used to alleviate the load on the power infrastructures. This would allow to reduce the congestion on the substation level and allow more flexibility for the DSO.

References

[1] Fabien Chidanand Robert and Sundararaman Gopalan. "Low cost, highly reliable rural electrification through a combination of grid extension and local renewable energy generation". In: *Sustainable Cities and Society* 42 (Oct. 1, 2018), pp. 344–354. ISSN: 2210-6707.

[2] Danja. *Grid congestion: what is it and how can you avoid it?* CFP. May 2, 2023.

[3] Nurul Idayu Yusoff, Abdullah Asuhaimi Mohd Zin, and Azhar Bin Khairuddin. "Congestion management in power system: A review". In: *2017 3rd International Conference on Power Generation Systems and Renewable Energy Technologies (PGSRET)*. 2017 3rd International Conference on Power Generation Systems and Renewable Energy Technologies (PGSRET). Apr. 2017, pp. 22–27.

[4] Jaka Rober, Leon Maruša, and Miloš Beković. "A Machine Learning Application for the Energy Flexibility Assessment of a Distribution Network for Consumers". In: *Energies* 16.17 (Jan. 2023). Number: 17 Publisher: Multidisciplinary Digital Publishing Institute, p. 6168. ISSN: 1996-1073.

[5] International Energy Agency. *Unlocking Smart Grid Opportunities in Emerging Markets and Developing Economies*. OECD, June 6, 2023. ISBN: 978-92-64-86019-3. DOI: `10.1787/3cd7a4b0-en`. URL: `https://www.oecd-ilibrary.org/energy/unlocking-smart-grid-opportunities-in-emerging-markets-and-developing-economies_3cd7a4b0-en`.

[6] *REPowerEU*. May 18, 2022. URL: `https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal/repowereu-affordable-secure-and-sustainable-energy-europe_en`.

[7] Ministerie van Algemene Zaken. *Central government encourages sustainable energy - Renewable energy - Government.nl*. Last Modified: 2023-11-20T13:29 Publisher: Ministerie van Algemene Zaken. July 26, 2017. URL: `https://www.government.nl/topics/renewable-energy/central-government-encourages-sustainable-energy`.

[8] Lisa Göransson et al. "Linkages between demand-side management and congestion in the European electricity transmission system". In: *Energy* 69 (May 1, 2014), pp. 860–872. ISSN: 0360-5442.

[9] International Energy Agency. "Unlocking Smart Grid Opportunities in Emerging Markets and eveloping Economies". In: *IEA* (2023), pp. 123–456.

[10] *The Dirty Footprint of the Broken Grid*. IFC. DOI: `10/dirty-footprint-of-broken-grid`. URL: `https://www.ifc.org/en/insights-reports/2010/dirty-footprint-of-broken-grid`.

[11] Marta Poncela Blanco and Fabio Monforti-Ferrario. *The impact of power network congestion, its consequences and mitigation measures on air pollutants and greenhouse gases emissions. A case from Germany*. 2021.

[12] Ken Belson and Matthew L. Wald. "'03 Blackout Is Recalled, Amid Lessons Learned". In: *The New York Times* (Aug. 14, 2008). ISSN: 0362-4331. URL: `https://www.nytimes.com/2008/08/14/nyregion/14blackout.html`.

[13] Aishvarya Narain, S.K. Srivastava, and S.N. Singh. "Congestion management approaches in restructured power system: Key issues and challenges". In: *The Electricity Journal* 33.3 (Apr. 2020), p. 106715. ISSN: 10406190.

[14] E. Taibi et al. *Summary of power system flexibility for the energy transition. Part 2: IRENA FlexTool methodology*. Tech. rep. International Renewable Energy Agency, 2019.

[15] Peter D.Lund et al. "Review of energy system flexibility measures to enable high levels of variable renewable electricity". In: (2015).

[16] Jukka V. Paatero and Peter D. Lund. "A model for generating household electricity load profiles". In: *International Journal of Energy Research* 30.5 (2006), pp. 273–290. ISSN: 1099-114X.

[17] Yi Zong et al. “Model Predictive Controller for Active Demand Side Management with PV self-consumption in an intelligent building”. In: *2012 3rd IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*. 2012 3rd IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe). Berlin, Germany: IEEE, Oct. 2012, pp. 1–8. ISBN: 978-1-4673-2597-4 978-1-4673-2595-0 978-1-4673-2596-7.

[18] Nikolaos Savvopoulos and Nikos Hatziargyriou. “An Effective Method to Estimate the Aggregated Flexibility at Distribution Level”. In: *IEEE Access* 11 (2023), pp. 31373–31383. ISSN: 2169-3536.

[19] Matthias Huber, Desislava Dimkova, and Thomas Hamacher. “Integration of wind and solar power in Europe: Assessment of flexibility requirements”. In: *Energy* 69 (May 1, 2014), pp. 236–246. ISSN: 0360-5442.

[20] Alberto Bolzoni et al. “Model Predictive Control for optimizing the flexibility of sustainable energy assets: An experimental case study”. In: *International Journal of Electrical Power & Energy Systems* 129 (July 1, 2021), p. 106822. ISSN: 0142-0615.

[21] Zachary E. Lee and K. Max Zhang. “Scalable identification and control of residential heat pumps: A minimal hardware approach”. In: *Applied Energy* 286 (Mar. 15, 2021), p. 116544. ISSN: 0306-2619.

[22] Rajesh Retnamony and I. Jacob Raglend. “Congestion Management is to enhance the transient stability in a deregulated power system using FACTS devices”. In: *2015 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*. 2015 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT). Dec. 2015, pp. 744–752.

[23] Mohammad Mahmoudian Esfahani and Gholam Reza Yousefi. “Real Time Congestion Management in Power Systems Considering Quasi-Dynamic Thermal Rating and Congestion Clearing Time”. In: *IEEE Transactions on Industrial Informatics* 12.2 (Apr. 2016). Conference Name: IEEE Transactions on Industrial Informatics, pp. 745–754. ISSN: 1941-0050.

[24] Surender Reddy Salkuti and Seong-Cheol Kim. “Congestion Management Using Multi-Objective Glowworm Swarm Optimization Algorithm”. In: *Journal of Electrical Engineering & Technology* 14.4 (July 1, 2019), pp. 1565–1575. ISSN: 2093-7423.

[25] Jagadeeswar Chintam and Mary Daniel. “Real-Power Rescheduling of Generators for Congestion Management Using a Novel Satin Bowerbird Optimization Algorithm”. In: *Energies* 11.1 (Jan. 12, 2018), p. 183. ISSN: 1996-1073.

[26] S. Surender Reddy. “Multi-Objective Based Congestion Management Using Generation Rescheduling and Load Shedding”. In: *IEEE Transactions on Power Systems* 32.2 (Mar. 2017). Conference Name: IEEE Transactions on Power Systems, pp. 852–863. ISSN: 1558-0679.

[27] A. K. Singh and S. K. Parida. “A review on distributed generation allocation and planning in deregulated electricity market”. In: *Renewable and Sustainable Energy Reviews* 82 (Feb. 1, 2018), pp. 4132–4141. ISSN: 1364-0321.

[28] Jianwei Liu, M.M.A. Salama, and R.R. Mansour. “Identify the impact of distributed resources on congestion management”. In: *IEEE Transactions on Power Delivery* 20.3 (July 2005). Conference Name: IEEE Transactions on Power Delivery, pp. 1998–2005. ISSN: 1937-4208.

[29] Hong Tang, Shengwei Wang, and Hangxin Li. “Flexibility categorization, sources, capabilities and technologies for energy-flexible and grid-responsive buildings: State-of-the-art and future perspective”. In: *Energy* 219 (Mar. 2021), p. 119598. ISSN: 03605442.

[30] A. N. M. M. Haque et al. “Integrating Direct and Indirect Load Control for Congestion Management in LV Networks”. In: *IEEE Transactions on Smart Grid* 10.1 (Jan. 2019). Conference Name: IEEE Transactions on Smart Grid, pp. 741–751. ISSN: 1949-3061.

[31] Md Sarwar and Anwar Shahzad Siddiqui. “Congestion management in deregulated electricity market using distributed generation”. In: *2015 Annual IEEE India Conference (INDICON)*. 2015 Annual IEEE India Conference (INDICON). ISSN: 2325-9418. Dec. 2015, pp. 1–5.

[32] Antun Pfeifer et al. “Flexibility index and decreasing the costs in energy systems with high share of renewable energy”. In: *Energy Conversion and Management* 240 (July 2021), p. 114258. ISSN: 01968904.

[33] A. Yousefi et al. “Congestion management using demand response and FACTS devices”. In: *International Journal of Electrical Power & Energy Systems* 37.1 (May 1, 2012), pp. 78–85. ISSN: 0142-0615.

[34] Peter Palensky and Dietmar Dietrich. “Demand Side Management: Demand Response, Intelligent Energy Systems, and Smart Loads”. In: *IEEE Transactions on Industrial Informatics* 7.3 (Aug. 2011), pp. 381–388. ISSN: 1551-3203, 1941-0050.

[35] Fariborz Zaeim-Kohan, Hadi Razmi, and Hasan Doagou-Mojarrad. “Multi-objective transmission congestion management considering demand response programs and generation rescheduling”. In: *Applied Soft Computing* 70 (Sept. 1, 2018), pp. 169–181. ISSN: 1568-4946.

[36] Jennifer Leisch and Jaquelin Cochran. “Sources of Operational Flexibility”. In: (2015).

[37] Manar Amayri et al. “Flexibility characterization of residential electricity consumption: A machine learning approach”. In: *Sustainable Energy, Grids and Networks* 32 (Dec. 1, 2022), p. 100801. ISSN: 2352-4677.

[38] Emil Nyholm et al. “Demand response potential of electrical space heating in Swedish single-family dwellings”. In: *Building and Environment* 96 (Feb. 1, 2016), pp. 270–282. ISSN: 0360-1323.

[39] Thomas Nuytten et al. “Flexibility of a combined heat and power system with thermal energy storage for district heating”. In: *Applied Energy* 104 (Apr. 1, 2013), pp. 583–591. ISSN: 0306-2619.

[40] Sebastian Pater. “Increasing Energy Self-Consumption in Residential Photovoltaic Systems with Heat Pumps in Poland”. In: *Energies* 16.10 (Jan. 2023). Number: 10 Publisher: Multidisciplinary Digital Publishing Institute, p. 4003. ISSN: 1996-1073.

[41] Lu Chen et al. “Optimal energy management of smart building for peak shaving considering multi-energy flexibility measures”. In: *Energy and Buildings* 241 (June 15, 2021), p. 110932. ISSN: 0378-7788.

[42] Yongbao Chen et al. “Measures to improve energy demand flexibility in buildings for demand response (DR): A review”. In: *Energy and Buildings* 177 (Oct. 15, 2018), pp. 125–139. ISSN: 0378-7788.

[43] Han Li et al. “Energy flexibility of residential buildings: A systematic review of characterization and quantification methods and applications”. In: *Advances in Applied Energy* 3 (Aug. 25, 2021), p. 100054. ISSN: 2666-7924.

[44] M. Heleno et al. “Estimation of the flexibility range in the transmission-distribution boundary”. In: *2015 IEEE Eindhoven PowerTech*. 2015 IEEE Eindhoven PowerTech. June 2015, pp. 1–6.

[45] Chinthaka Dinesh et al. “Non-intrusive load monitoring under residential solar power influx”. In: *Applied Energy* 205 (Nov. 1, 2017), pp. 1068–1080. ISSN: 0306-2619.

[46] Hafiz M. Usman, Ramadan ElShatshat, and Ayman H. El-Hag. “A novel non-intrusive framework for real-time disaggregation of behind-the-meter solar generation from smart meter data”. In: *Electric Power Systems Research* 225 (Dec. 1, 2023), p. 109831. ISSN: 0378-7796.

[47] Fankun Bu et al. “Disaggregating Customer-Level Behind-the-Meter PV Generation Using Smart Meter Data and Solar Exemplars”. In: *IEEE Transactions on Power Systems* 36.6 (Nov. 2021). Conference Name: IEEE Transactions on Power Systems, pp. 5417–5427. ISSN: 1558-0679.

[48] Emre C. Kara et al. “Disaggregating solar generation from feeder-level measurements”. In: *Sustainable Energy, Grids and Networks* 13 (Mar. 1, 2018), pp. 112–121. ISSN: 2352-4677.

[49] Charalampos Ziras et al. “Experimental flexibility identification of aggregated residential thermal loads using behind-the-meter data”. In: *Applied Energy* 242 (May 2019), pp. 1407–1421. ISSN: 03062619.

[50] Mark E.H. Dyson et al. “Using smart meter data to estimate demand response potential, with application to solar energy integration”. In: *Energy Policy* 73 (Oct. 2014), pp. 607–619. ISSN: 03014215.

[51] Jenny Love et al. “The addition of heat pump electricity load profiles to GB electricity demand: Evidence from a heat pump field trial”. In: *Applied Energy* 204 (Oct. 15, 2017), pp. 332–342. ISSN: 0306-2619.

[52] Rui Amaral Lopes et al. “A Literature Review of Methodologies Used to Assess the Energy Flexibility of Buildings”. In: *Energy Procedia*. Proceedings of the 4th International Conference on Solar Heating and Cooling for Buildings and Industry (SHC 2015) 91 (June 1, 2016), pp. 1053–1058. ISSN: 1876-6102.

[53] Joakim Widén. “Improved photovoltaic self-consumption with appliance scheduling in 200 single-family buildings”. In: (2014).

[54] Georgios Papaefthymiou, Bernhard Hasche, and Christian Nabe. “Potential of Heat Pumps for Demand Side Management and Wind Power Integration in the German Electricity Market”. In: *IEEE Transactions on Sustainable Energy* 3.4 (Oct. 2012). Conference Name: IEEE Transactions on Sustainable Energy, pp. 636–642. ISSN: 1949-3037.

[55] M. Razmara et al. “Building-to-grid predictive power flow control for demand response and demand flexibility programs”. In: *Applied Energy* 203 (Oct. 1, 2017), pp. 128–141. ISSN: 0306-2619.

[56] Shariq Riaz and Pierluigi Mancarella. “On Feasibility and Flexibility Operating Regions of Virtual Power Plants and TSO/DSO Interfaces”. In: *2019 IEEE Milan PowerTech*. 2019 IEEE Milan PowerTech. June 2019, pp. 1–6.

[57] Sandip Chanda and Abhinandan De. “Application of particle swarm optimisation for relieving congestion in deregulated power system”. In: *2011 IEEE Recent Advances in Intelligent Computational Systems*. 2011 IEEE Recent Advances in Intelligent Computational Systems (RAICS). Trivandrum, India: IEEE, Sept. 2011, pp. 837–840. ISBN: 978-1-4244-9478-1.

[58] Denis Mende et al. “Multiobjective optimization in congestion management considering technical and economic aspects”. In: *2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*. 2016 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC). Xi’an, China: IEEE, Oct. 2016, pp. 1061–1066. ISBN: 978-1-5090-5418-3.

[59] *Joint Estimation of Behind-the-Meter Solar Generation in a Community | IEEE Journals & Magazine | IEEE Xplore*.

[60] Kadir Amasyali, Mohammed Olama, and Aniruddha Perumalla. “A Machine Learning-based Approach to Predict the Aggregate Flexibility of HVAC Systems”. In: *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT). ISSN: 2472-8152. Feb. 2020, pp. 1–5.

[61] Rostislav Krč et al. “Machine Learning-Based Node Characterization for Smart Grid Demand Response Flexibility Assessment”. In: *Sustainability* 13.5 (Jan. 2021). Number: 5 Publisher: Multidisciplinary Digital Publishing Institute, p. 2954. ISSN: 2071-1050.

[62] Yi Wang et al. “Data-Driven Probabilistic Net Load Forecasting With High Penetration of Behind-the-Meter PV”. In: *IEEE Transactions on Power Systems* 33.3 (May 2018). Conference Name: IEEE Transactions on Power Systems, pp. 3255–3264. ISSN: 1558-0679.

[63] Kangping Li et al. “Capacity and output power estimation approach of individual behind-the-meter distributed photovoltaic system for demand response baseline estimation”. In: *Applied Energy* 253 (Nov. 1, 2019), p. 113595. ISSN: 0306-2619.

[64] Mohsen Kalantar-Neyestanaki et al. “Characterizing the Reserve Provision Capability Area of Active Distribution Networks: A Linear Robust Optimization Method”. In: *IEEE Transactions on Smart Grid* 11.3 (May 2020). Conference Name: IEEE Transactions on Smart Grid, pp. 2464–2475. ISSN: 1949-3061.

[65] Andres F. Moreno Jaramillo et al. “Distributed Energy Resources Electric Profile Identification in Low Voltage Networks Using Supervised Machine Learning Techniques”. In: *IEEE Access* 11 (2023). Conference Name: IEEE Access, pp. 19469–19486. ISSN: 2169-3536.

[66] Marlon Schlemminger et al. “Dataset on electrical single-family house and heat pump load profiles in Germany”. In: *Scientific Data* 9.1 (Feb. 15, 2022). Number: 1 Publisher: Nature Publishing Group, p. 56. ISSN: 2052-4463.

[67] Yang Song et al. “Estimating electric power consumption of in-situ residential heat pump systems: A data-driven approach”. In: *Applied Energy* 352 (Dec. 15, 2023), p. 121971. ISSN: 0306-2619.

[68] Oliver Ruhnau, Lion Hirth, and Aaron Praktiknjo. “Time series of heat demand and heat pump efficiency for energy system modeling”. In: *Scientific Data* 6.1 (Oct. 1, 2019). Number: 1 Publisher: Nature Publishing Group, p. 189. ISSN: 2052-4463.

[69] *Sectordocumenten*. MFFBAS. URL: `https://www.mffbas.nl/documenten/`.

[70] Reinaldo Tonkoski, Dave Turcotte, and Tarek H. M. EL-Fouly. “Impact of High PV Penetration on Voltage Profiles in Residential Neighborhoods”. In: *IEEE Transactions on Sustainable Energy* 3.3 (July 2012). Conference Name: IEEE Transactions on Sustainable Energy, pp. 518–527. ISSN: 1949-3037.

[71] *Intro Tutorial — pvlib python 0.10.5 documentation*. URL: `https://pvlib-python.readthedocs.io/en/stable/user_guide/introtutorial.html` (visited on 05/09/2024).

[72] *Welcome - KNMI Data Platform - KNMI Data Platform*. URL: `https://dataplatform.knmi.nl/`.