TÉCNICO LISBOA

Net Load Disaggregation and Forecast at Medium Voltage Level

Mateo Toro Cárdenas

Thesis to obtain the Master of Science Degree in

Energy Engineering and Management

Supervisors: Prof. Pedro Manuel Santos de Carvalho
Prof. Luís António Fialho Marcelino Ferreira

Examination Committee

Chairperson: Prof. Duarte de Mesquita e Sousa
Supervisor: Prof. Pedro Manuel Santos de Carvalho
Member of the Committee: Prof. Tiago Morais Delgado Domingos

December 2021

This work was created using LaTeX typesetting language
in the Overleaf environment (www.overleaf.com).

Acknowledgments

I want to deeply thank my supervisors, Professor Pedro Carvalho and Professor Marcelino Ferreira, for the trust they put on me when accepting me on board of the project at INESC-ID. I am highly pleased to have been able to develop my research work in the context of a real world engineering challenge. I would also like to acknowledge Professor Hugo Morais' participation in the development of this thesis. His invaluable support went far beyond his obligations aiming to achieve outstanding processes and high quality results. In addition, Ines Moreira deserves a recognition given that we worked together as a group supporting each other with our work.

Being at this point marks the end for an unbelievable phase of my life. I would like to thank EIT InnoEnergy and its wonderful staff and collaborators for arranging and partially financing the MSc in Energy Transition degree. Because of it, I achieved a life-goal about traveling, learning and getting experiences to improve myself personally and professionally. I cannot imagine how my life would have been without this program. I will gratefully remember every single one I met along this journey.

Nevertheless and on top of everything, my parents and my sister are the most important reason for me to be where I am right now. Even though we are far physically, I can always feel their love for me and I thank them from the bottom of my heart every effort they make for me and my wellness.

This work was partially supported by Portuguese national funds through Fundação para a Ciência e a Tecnologia with reference UIDB/50021/2020 and by E-REDES, the Portuguese Distribution System Operator.

Abstract

With more and more micro-generation being connected to the lower voltage levels, it becomes increasingly difficult for operators to anticipate the loading condition of their networks. Several facts concur to such difficulty, one being the frequent lack of knowledge about the generation component of the measured net load. The generation component is typically more volatile than natural load and is impossible to predict without reliable information on the grid renewable installed capacity.

In this context, the objective of this thesis is to propose a non-intrusive methodology to estimate distributed generation' installed capacity. Relying upon net load historical data obtained at the secondary substation level and on historical meteorological data, this thesis presents algorithms that proved effective in determining an approximation of the renewable embedded capacity. The accuracy and effectiveness of the algorithms presented are illustrated with real data, for real use cases in distribution grids.

Keywords

Distribution Network, Forecast, Net Load Disaggregation, Photovoltaic Generation, Secondary Substation.

Resumo

Com cada vez mais microgerações conectadas aos níveis de tensão mais baixos, torna-se cada vez mais difícil para as operadoras prever a condição de carregamento de suas redes. Vários fatos concorrem para tal dificuldade, sendo um deles o frequente desconhecimento do componente de geração da carga líquida medida. O componente de geração é normalmente mais volátil do que a carga natural e é impossível prever sem informações confiáveis sobre a capacidade renovável instalada da rede.

Neste contexto, o objetivo desta tese é propor uma metodologia não intrusiva para estimar a capacidade instalada de geração distribuída. Baseando-se em dados históricos de carga líquida obtidos no nível da subestação secundária e em dados meteorológicos históricos, esta tese apresenta algoritmos que se mostraram eficazes na determinação de uma aproximação da capacidade embarcada renovável. A precisão e eficácia dos algoritmos apresentados são ilustradas com dados reais, para casos de uso reais em grades de distribuição.

Palavras Chave

Desagregação de carga líquida, Geração Fotovoltaica, Previsão, Rede de Distribuição, Subestação Secundária.

Contents

1 Introduction **1**

1.1 Motivation . . . 3

1.2 Contributions . . . 4

1.3 Organization of the Document . . . 4

2 State of the Art **5**

2.1 Other Disaggregation Approaches . . . 8

2.2 Non-Intrusive Load Management . . . 9

2.3 Load Forecasting Literature . . . 11

3 Methodology **13**

3.1 Day Comparison . . . 15

3.1.1 Modified Bisection . . . 16

3.2 DG Size Estimation . . . 21

3.2.1 Mode/Cluster Algorithm . . . 21

3.2.2 K-means . . . 23

3.2.3 Disaggregated Profiles . . . 24

3.3 AutoRegressive Model with Exogenous variable . . . 24

4 Study Cases **27**

4.1 Data Collection . . . 29

4.2 Synthetic PV . . . 32

4.3 Real PV . . . 33

5 Results **37**

5.1 Disaggregation . . . 39

5.1.1 Synthetic PV . . . 39

5.1.2 Real PV . . . 39

5.1.3 Discussion . . . 40

5.2 Load Forecasting . . . 43

5.2.1 Synthetic PV . . . 44

5.2.2 Real PV 48
5.2.3 Discussion 52

6 Conclusion **55**

6.1 System Limitations 58
6.2 Future Work 59

Bibliography **61**

List of Figures

2.1 CASIO Duck chart 7

3.1 Inputs (a) and Outputs (b) of the proposed methodology 15
3.2 Net Load Dissagregation methodology 16
3.3 Day Comparison output 16
3.4 Epsilon Function $f(x_{d,i})$ = $\varepsilon_{d,i}$ 18
3.5 Epsilon Function $f(x_{d,i}) = \varepsilon_{d,i}$ with two x_d as input leading to big (a) and small (b) ε_d. . . 19
3.6 Graphical context for Stability Control Function 19
3.7 DG Size Estimation Output Obtained by Mode/Cluster Algorithm 22
3.8 DG Size Estimation Output Obtained by K-means Algorithm 23
3.9 AutoRegressive Model with Exogenous variable inputs and outputs 25
3.10 Forecast scheme 25

4.1 Original True Natural Load profile (a) and detailed (b) as given by a private source 29
4.2 Original True Solar Irradiance profile (a) and detailed (b) as given by the public source . . 30
4.3 Original True PV profile (a) and detailed (b) as given by the public source 31
4.4 RMSE in L by comparing two consecutive days (a) and two same days of consecutive weeks (b) 32
4.5 Performance of Mode/Cluster and K-Means in a Synthetic PV Scenario 33
4.6 Real PV generation on a 220 Wp panel against solar irradiance during one month 34
4.7 Solar Irradiance and Real PV generation on a 220 Wp panel during 3 days 34
4.8 Performance of Mode/Cluster and K-Means in a Real PV Scenario 35

5.1 True and Estimated Natural Load Profiles with 25% PV penetration 41
5.2 True and Estimated PV Generation Profiles with 25% PV penetration 41
5.3 True and Estimated Natural Load Profiles with 75% PV penetration 42
5.4 True and Estimated PV Generation Profiles with 75% PV penetration 42
5.5 Estimated DG Installed Capacity, $\hat{X}$, Average Errors according to PV penetration 43

5.6 Effect of Disaggregation on Net Load Forecast Results under Synthetic PV Scenario . . . 44
5.7 Estimated PV Generation (a) and Estimated Natural Load (b) Forecast Under Synthetic PV Scenario with 25% PV Penetration . . . 45
5.8 Estimated PV Generation (a) and Estimated Natural Load (b) Forecast Under Synthetic PV Scenario with 75% PV Penetration . . . 45
5.9 Net Load Forecast Under Synthetic PV Scenario with 25% PV penetration . . . 46
5.10 Net Load Forecast Under Synthetic PV Scenario with 75% PV penetration . . . 47
5.11 Solar Irradiance and PV Generation Under Synthetic PV Scenario . . . 47
5.12 Effect of Disaggregation on Net Load Forecast Results under Real PV Scenario . . . 48
5.13 Estimated PV Generation (a) and Estimated Natural Load (b) Forecast Under Real PV Scenario with 25% PV Penetration . . . 49
5.14 Estimated PV Generation (a) and Estimated Natural Load (b) Forecast Under Real PV Scenario with 75% PV Penetration . . . 49
5.15 Net Load Forecast Under Real PV Scenario with 25% PV penetration . . . 50
5.16 Net Load Forecast Under Real PV Scenario with 75% PV penetration . . . 51
5.17 Solar Irradiance and PV Generation Under Real PV Scenario . . . 51
5.18 Estimated PV Generation Prediction, $\hat{PV}^{pred}$, Average Errors under Synthetic PV and Real PV scenario . . . 52
5.19 Estimated Natural Load Predictions, $\hat{L}^{pred}$, Average Errors under Synthetic PV and Real PV scenario . . . 53

List of Tables

5.1 Estimated Installed DG Capacity Results at SS Level under Synthetic PV Scenario 39
5.2 Estimated Installed DG Capacity Results at SS Level under Real PV Scenario 40
5.3 Average Error [%] on Results of Estimated Installed DG Capacity at SS Level 40
5.4 Dates to execute the ARX model 43
5.5 Mean Errors [%] on Net Load Forecast 52

List of Algorithms

3.1 Modified Bisection 17
3.2 Epsilon Function 18
3.3 Stability Control Function 20
3.4 Mode/Cluster Algorithm 21
3.5 Iterative Histogram 22

Nomenclature

Methodology Symbols

Δ	Precision in Modified Bisection
Γ	Precision in Iterative Histogram
$\hat{L}$	Estimated Natural Load
$\hat{L}^{pred}$	Estimated Natural Load Prediction
$\hat{N}^{pred}$	Aggregated Net Load Prediction
$\hat{N}^{pred}_{partial}$	Partial Aggregated Net Load Prediction
$\hat{PV}$	Estimated PV Generation
$\hat{PV}^{pred}$	Estimated PV Generation Prediction
$\hat{X}$	Estimated installed DG capacity
$\hat{x}_d$	Estimated installed DG capacity per pair of day d
μ	Centroid in a K-means Algorithm
ω	Stability Criteria in Modified Bisection Algorithm
ε_d	RMS error between the estimate natural load $\hat{L}$ profiles of a pair of days d
$\varepsilon_{d,i}$	RMS error between the estimate natural load $\hat{L}$ profiles of a pair of days d per index i
D	All possible pair of days d in a data set
d	Pair of days from consecutive weeks (same weekday)
$freq$	Number of times that a value is repeated in a data set
G	Maximum number of iterations in Modified Bisection

I Solar irradiance

K Scale Factor of $\hat{PV}$ size

L True Natural Load

m Statistical mode Mo of estimated DG capacity $\hat{x}_d$ in all possible pair of days D

Mo Statistical Mode

N True Net Load

N^{pred} Net Load Prediction

PV True PV Generation

X True installed DG capacity

$x_{d,i}$ Installed DG capacity per pair of day d and index i

ARX Model Symbols

β_j Coefficient of the j^{th} endogenous variable lag

θ_f Coefficient of the f^{th} exogenous variable lag

ξ_t Stochastic term

I_{t-f} f^{th} value of exogenous variable

p Model order of Endogenous Variable

s Model order of Exogenous Variable

Y_t Value to predict for instant t

Y_{t-j} j^{th} previous value of the variable to predict

Acronyms

ANILM	Advanced NILM
CASIO	California Independent System Operator
DG	Distributed Generation
DSO	Distribution System Operator
HMM	Hidden Morkov Models
HV	High Voltage
HVAC	Heating, Ventilation and Air Conditioning
LSTM	Long Short-Term Memory
LV	Low Voltage
MV	Medium Voltage
NILM	Non-Intrusive Load Management
PV	Solar Photovoltaic
RES	Renewable Energy Source
RMSE	Root Mean Square Error
SIQCP	Segmented Integer Quadratic Constraint Programming
SS	Secondary Substation
TSO	Transmission System Operator

1

Introduction

Contents

1.1 Motivation 3

1.2 Contributions 4

1.3 Organization of the Document 4

Distributed Generation (DG) penetration is increasing mainly due to the deployment of wind and Solar Photovoltaic (PV) technologies. Most wind generation is being installed at Medium Voltage (MV) and High Voltage (HV) networks while PV is being installed across all voltage levels, including in Low Voltage (LV) networks [1]. One of the main operational challenges of the deployment of PV at LV networks is the change in the consumption profile it causes at the higher voltage levels by producing duck-shaped profiles [2] which will be further explained in Chapter 2. Those conditions need to be anticipated by Transmission System Operators (TSOs) and mainly by Distribution System Operators (DSOs) to avoid voltage constraints (under and over voltages) in the system.

Anticipating such profile changes with acceptable accuracy requires knowledge of the approximated PV installed capacity in order to infer about PV power active power generation and its dependencies on meteorological conditions. However, such capacity is frequently unknown to the DSO, either because it did not keep track on the licensed PV connections, or because such connections did not require any kind of licensing from the DSOs, which is frequently the case for small capacities and/or self-consumption in many EU countries. In [3], an overview of the current requirements for grid connections in several European countries is presented. Specifically, a given DG with less that 1 MW installed capacity is exempt of some requirements imposed by the nations grid codes such as frequency control, dynamic voltage support and remote control and communication, among others. In [4], the interconnection procedures of small distributed generation in USA are described. There, an installation of less that 50 kW is considered typical for a household. Precisely, the last is the type of PV expected in LV networks and the aggregation of such would produce the previously described challenges for a DSO.

In addition, if any DSO choose to install a smart meter in every consumption node of its distribution grid, they will face the challenge to deploy an extensive sensing infrastructure with the associated installation, operational and maintenance costs, plus the increased computational efforts as a consequence of processing the enormous amount of data collected.

1.1 Motivation

Given the previous context and challenges, it becomes obvious the beneficial effect of, first, anticipating the behaviour of a variable consumption profile and the impacts it can produce on the grid, and second, avoiding to incur on large capital investments.

In this sense, the objective of the thesis here presented is to develop an effective non-intrusive methodology that is capable of estimating the DG capacity installed in a LV distribution grid.

The methodology makes use of Secondary Substation (SS) historical data of active power measurements (Net Load) and available weather data to infer about the approximated installed capacity by disaggregating consumption (Natural Load) and PV generation active power profiles.

1.2 Contributions

The proposed thesis achieved successful results by effectively obtaining the size of the DG installed in the LV and the components of a Net Load profile by analyzing only historical data of the named variable and associated weather data. On average, over 90% accuracy was obtained when estimating the DG installed capacity and such value was used to created the estimated profiles of PV generation and Natural Load. The last represents the actual consumption by the loads connected in the feeder downstream the SS Transformer.

Moreover, the proposed work was based on classical mathematical concepts and statistical analysis, avoiding complex learning approaches found in the literature.

Furthermore, other application for the work here presented is addressed. The disaggregated profiles found, i.e. Natural Load and PV generation profiles, were used as inputs for Load Forecasting calculations. The mentioned process, which consisted on implementing disaggregation process before Load forecasting calculations, proved to be effective as well. In average, the accuracy on Net Load forecasting was improved by 0.41 pp in comparison with a reference value.

The work developed in this thesis, plus the collaboration with other authors, led to the publication of a paper called “Net Load Disaggregation at Secondary Substation Level” in the Electric Power Systems Research (EPSR) journal after successfully being accepted to deliver a presentation in the Power Systems Computation Conference (PSCC) 2022.

1.3 Organization of the Document

This thesis is organized in 6 chapters. Chapter 1 presents the context and challenges that motivate the development of this work, as well as the contributions obtained with its implementation. On Chapter 2 a deep look into the contributions made by other authors in the topics addressed in this work was done. The main explanation about all the steps taken so as to achieve the thesis’ objective is extensively detailed in Chapter 3. Chapter 4 presents the study cases developed on this work. The tests made on them and the evaluation of the developed work performance is described in Chapter 5. In addition, this chapter exemplifies how Load Forecasting was carried out by applying disaggregation calculations on historic Net Load profiles. At the end, conclusions and final remarks are presented in Chapter 6.

2

State of the Art

Contents

2.1 Other Disaggregation Approaches 8

2.2 Non-Intrusive Load Management 9

2.3 Load Forecasting Literature 11

As the context was mentioned in the previous chapter, so was introduced one of the main operational challenges that TSOs and DSOs will face in the near future. The deployment of PV at LV networks will produce Net Load profiles in the shape of a duck as seen in Figure 2.1 by [2], the California Independent System Operator (CASIO), where the March 31' Net Load in Megawatts during the day is presented in various years. In the mentioned report is said that *the "belly" of the duck represents the period of lowest net load, where PV generation is at a maximum*. It is seen the change in Net Load on a specific day, March 31, along the years, from 2021 to 2020, as more DG in the form of PV has been installed in the network.

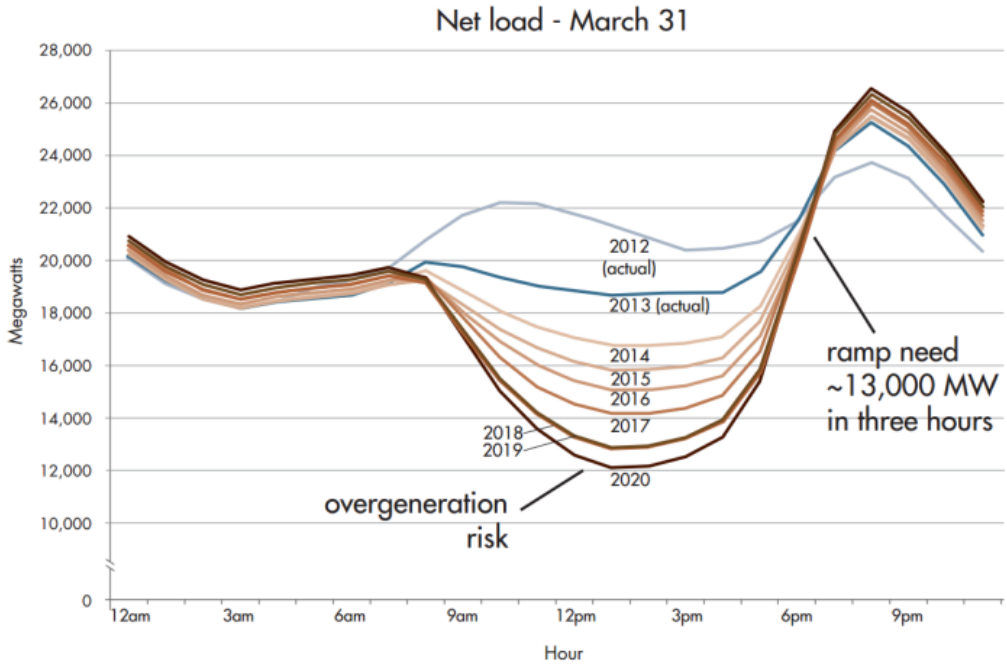


Figure 2.1: CASIO Duck chart

Given the previous, and in the pursuit to anticipate such behavior, this thesis presents an effective non-intrusive methodology that is capable of estimating the DG capacity installed in a LV distribution grid. The previous fits in the definition of Load Disaggregation. In essence, disaggregation is *the breakdown of observations within a common branch of a hierarchy to a more detailed level to that at which detailed observations are taken. With standard hierarchical classifications [...] categories can be split (disaggregated) when finer details are required and made possible by the codes given to the primary observations* [5]. In that context, to disaggregate load measurements is the action of breaking those observations down to a more detailed view where other categories of load can be seen. Those categories refer to what the original load signal is composed of. This way, finer details and individual profiles can be obtained and analyzed according to specific needs and goals.

However, the disaggregation concept is mainly applied to residential loads as suggested in [6] when defines disaggregation as a *set of statistical approaches for extracting end-use and/or appliance level data from an aggregate, or whole-building, energy signal*. Some works regarding the household applications of disaggregation are presented and described in Section 2.1.

A major contribution to the previous definition of disaggregation in the energy context is given by [7] when appoints the absence of *plug-level sensors*. The last part defines the non-intrusive characteristic of the work presented in this thesis. Such attribute is deeply studied in the literature under the concept of Non-Intrusive Load Management (NILM) further described in Section 2.2.

After effectively developing and implementing the disaggregation concept and calculations, a use for the ouputs of the methodology is found is proposed. Among others, one use for the developed work is related to achieve improvements in Load Forecasting. Regarding such, related works are analyzed in Section 2.3.

2.1 Other Disaggregation Approaches

Several works have addressed challenges related to the disaggregation concept. This section focuses on providing a background on publications with the same goal as the work develop on this thesis about the decomposition of original observations. However, the authors here addressed used a wide variety of algorithms and input data.

For example, in [8] the objective is to disaggregate monthly energy data into its end-uses such as Heating, Ventilation and Air Conditioning (HVAC) systems, baseload, and variable loads. The previous is achieved by taking into account monthly energy use data and a localized weather to develop a highly Thermodynamics-Based Model. This because loads such as HVAC are highly depended on the climate of the area where the measurements are taken from and therefore, this is the main contribution of the referenced work. In the present thesis, such loads are considered to be included on the feeder's consumption and their seasonality behavior is expected inside the same demand curve.

Regarding the inputs which are used in [8] but not in the present thesis are temperature and residential billing prices. The first is not considered given that temperature does not affect DG in the LV when the only Renewable Energy Source (RES) consider is produced by PV. Residential billing prices are not considered given the difference in scope among [8] and this thesis. The first focuses on residential scale while this thesis is focus on distribution grids.

Other publication, [9], refers to the benefit of providing information about feeder's load and network demand to the DSOs. This concept is shared with the proposed thesis. However, [9] disaggregates observations into aggregate air conditioner demand, aggregate demand of non-air conditioning loads, losses and reactive power injections by considering temperature as the external variable that directly affects air conditioning demand. A statement that was previously explained why it is not considered in the present thesis. Furthermore, the authors in [9] make emphasis on their use of different resolution input data and implementation of online learning algorithms as the main contribution of their work.

This thesis aligns with [10] in its goal to estimate PV power from aggregate power measurements within the distribution grid. However, its approach involves the use of an optimization problem that fits linear regression models and includes regularization terms in the objective function plus an extensive study in tuning the objective function weighting factors in a way to maximizes performance and robustness. The result is a model that improves their previous work achievements and can be applied to networks with either a single PV system in a megawatt scale or many distributed ones in the residential scale connected downstream of the SS. This thesis' proposed methodology is much simpler without neglecting the quality and precision of the results.

2.2 Non-Intrusive Load Management

Regarding NILM, literature is extensive. From the first time it was presented in [11] its goal was to *determine the energy consumption of individual appliances turning on and off in an electric load*. From that point, authors have made valuable contributions by providing detailed description of operational states of consumer appliances as described in [12]. In addition, on energy disaggregation is consider to provide value to home owners by detecting aging or malfunctioning of their appliances. However, even the most significant approaches involving Hidden Morkov Models (HMM) to describe loads and Segmented Integer Quadratic Constraint Programming (SIQCP) to effectively solve the NILM problem are only focused on household power profiles [13]. On top of that, the applications for NILM are not fully described and addressed given various aspects such as constant market introduction of new appliances, reduced open common data bases, development and prototyping costs, scalability and lack of innovative business models [14].

A comprehensive review of NILM methods in the home energy management system is presented in [15]. There, the critical issues of NILM are discussed and the Advanced NILM (ANILM) concept is proposed, including the most important characteristics of this new method. The main issues with NILM

come from *appliances electrical characteristics, the methods of sampling their electrical signal, and the relationship between their operation patterns with occupants' behavior and environmental conditions*. Basically, the appliances can be divided in various groups according to different criteria. That criteria is not common among all the authors, causing misfits in the literature developed around NILM. In [15], the reviewed criteria to classifiy home appliances is: Customer perception, Operational state, Waveform features, and User interface. Sampling frequency criteria deals with the interval of time between every data point is collected. It can range from seconds until hours, or even days. The last issue with NILM is related non-electrical variables such as home occupancy and outdoor temperature that should be taken into account when disaggregation models are executed.

Another review paper, more centred on the performance evaluation of NILM methods is presented in [16] aiming the comparison of the performance of different NILM methods applied to several datasets. The performance evaluation metrics for NILM methods are also reviewed in [16]. Those are categorized under event detection or energy estimation. To name a few examples, event detection category performance metrics are true and false positives and negatives, failed detections, error rates, total number of detected event, among many more. Energy estimation category is more related to the disaggregation concept address in this thesis in the way that it search for the appliances behaviour in terms on energy consumed rather than operating states detection. In this category, some examples of performance metrics are Root Mean Square Error (RMSE), R-squared, energy error, fraction of energy explained, among many more as well.

The problem of Datasets used in NILM methods is also addressed in [17]. According to the authors, after a critical analysis of 42 NILM datasets, the performance and use of NILM methods are limited by the features that are available in each dataset as well as by the lack of standards used to describe these features. It is mentioned that *most of the energy disaggregation work is related to residential sector, only a few resources are available from commercial prospective*. The previous gives a hint that a disaggregation method able to be applied to various kinds of load is missing in the literature. Precisely, one of the benefits propose by this thesis is its implementation on any distribution feeder no matter the kind of load it serves.

What is retrieved from all the literature here reviewed is a handful of questions and topics that are left open to be validated in the future. Specifically, this thesis aligns with the opportunity, based on all the previous knowledge, to escalate the applications of the disaggregation concept merged with the increasing use of DG involving RESs and the option to provide a valuable final use for DSOs involving an improved methodology to forecast load.

2.3 Load Forecasting Literature

Load Forecasting is presented in this thesis as one of the possible applications for the outputs of the proposed disaggregation methodology. Then, it becomes necessary to analyze what literature has been develop around this topic.

In first place, [18] highlights the effect of weather data and forecast use and effect on load forecasting. The author states: “Typically, load forecasting models are constructed and tested using actual weather readings. However, online operation of load forecasting models requires the use of weather forecasts, with associated weather forecast errors”. In this context, the thesis here presented agrees on what is been said but extends such testing procedures for future work. One contrasting point among [18] and this thesis happens when the first develops a fusion model as a technique for minimizing the effect of weather forecast errors in load forecasting models while the model used in this thesis is classical Auto-regressive model with a single exogenous meteorological variable, i.e. an ARX model which will be explained in detail ahead in this document. The previous given that the main scope of this thesis is about the disaggregation methodology rather than a load forecasting modeling.

Finally, [19] effectively summarizes the nowdays approach to deal with the continuous increase of DG specially involving household small scale PV panels. It address “the need for transiting from the concept of load forecasting to net energy forecasting”. To achieve its goal, [19] uses a Long Short-Term Memory (LSTM) neural network applying it from household to aggregate level analyzing the effect of different resolutions in the input data. However, the main contrast among LSTM and this thesis, is that the first doesn’t consider any weather variable as input or as a factor that directly affects the energy production downstream the meter.

3

Methodology

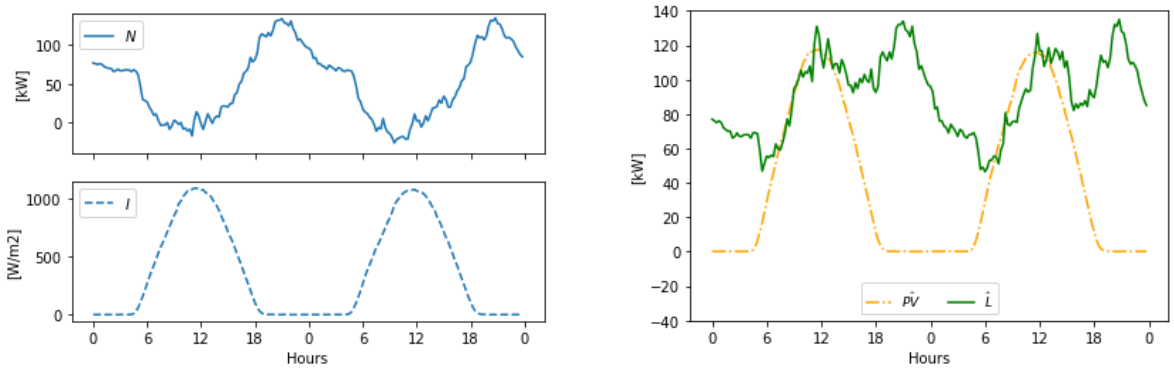
Contents

3.1 Day Comparison 15

3.2 DG Size Estimation 21

3.3 AutoRegressive Model with Exogenous variable 24

This thesis proposes a new methodology allowing the active power consumption and generation disaggregation at the SS level. The method uses two inputs. First, Net Load, N, i.e. active power measurements in kilowatts which could be in the downstream and/or upstream direction, from a meter installed in a SS. Second, weather information, specifically, Solar Irradiance in Watts per squared meter, I. Both inputs can be seen over two days in Figure 3.1(a). The method produces two outputs. First, the "estimated installed DG capacity", $\hat{X}$, that is an estimation of the capacity of the micro-generation installed in the LV network and connected at the SS. For the scope of this thesis, the only DG considered is coming from PV panels. The second output, the "Disaggregated Profiles", is the dissaggregated profiles in kilowatts of the historical N's components, i.e. Estimated Natural Load, $\hat{L}$, and Estimated PV generation, $\hat{PV}$. The previous can be seen over two days in Figure 3.1(b). $\hat{L}$ considers the power demand of consumers connected downstream from the meter, while $\hat{PV}$ refers to the active power injected in the LV network by the PV installed under the meter.



(a) Net Load, N (up) and Solar Irradiance, I (down)

(b) Estimated PV generation, $\hat{PV}$, and Estimated Natural Load, $\hat{L}$

Figure 3.1: Inputs (a) and Outputs (b) of the proposed methodology

As seen in Figure 3.2, the input data is processed by two different algorithms, "Day Comparison" and "DG Size Estimation", which will be explained in detail in Section 3.1 and Section 3.2, respectively. In addition, the outputs seen in Figure 3.2 can be used for different purposes. In the present thesis, the Disaggregated Profiles, i.e. $\hat{PV}$ and $\hat{L}$ profiles, will be used as inputs for a load forecasting method explained in Section 3.3.

3.1 Day Comparison

The inputs for this block of the methodology are Net Load, N, and Solar Irradiance, I, according to Figure 3.2 and are consider as global variables that can be access at any moment of the execution. The

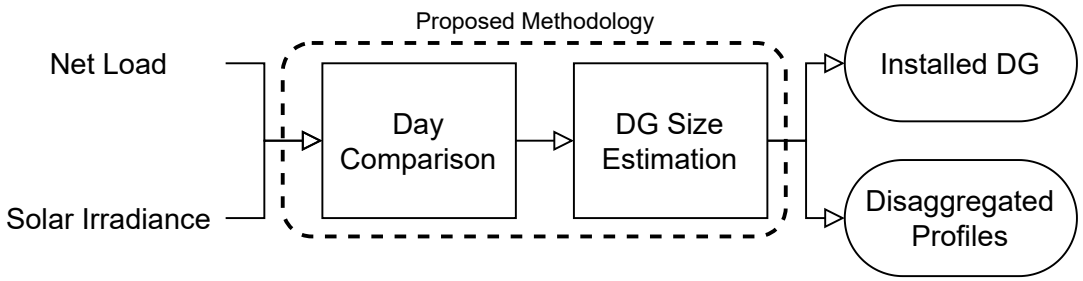


Figure 3.2: Net Load Dissagregation methodology

name of this block comes from the action of taking two days of data to compare them. Those days can be consecutive or not. Detailed explanation about it will be addressed in Section 4.1. The pair of days are referred as d and all the possible pairs of days is called D.

Each comparison will produce a value $\hat{x}_d$, the estimated installed DG capacity per pair of days d in kilowatts, and a value ε_d, the RMSE in kilowatts between the estimated natural load $\hat{L}$ profiles of a pair of days d. Day Comparison algorithm's output is the collection of $\hat{x}_d$ and ε_d for every d – see Figure 3.3.

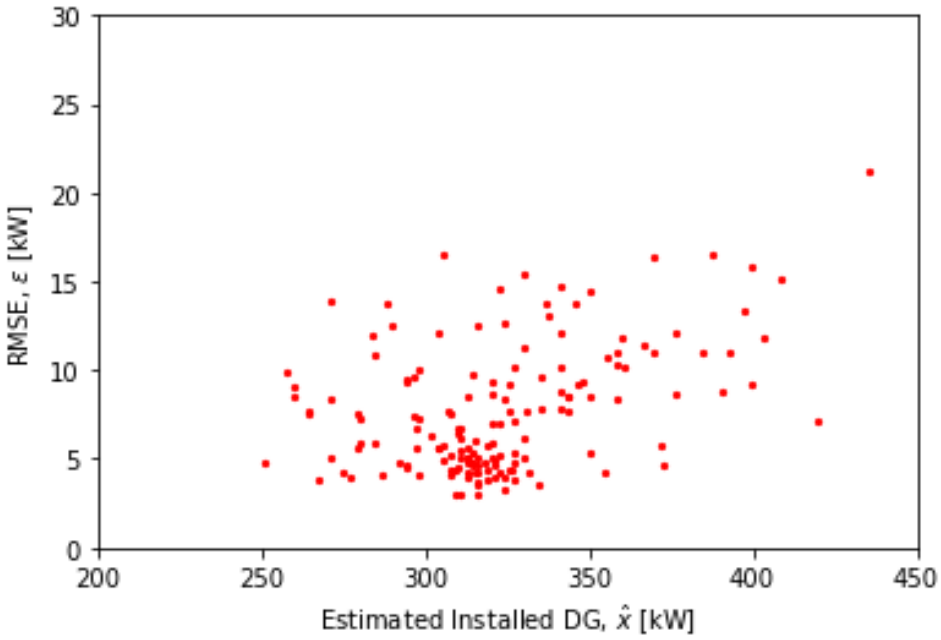


Figure 3.3: Day Comparison output

3.1.1 Modified Bisection

The comparison has the objective of finding the value $\hat{x}_d$ that minimizes the RMSE, ε_d, between the natural load profiles $\hat{L}$ of the pair of days d analyzed. The minimization problem is solved for every d with a Modified Bisection method presented in Algorithm 3.1. It is executed D times. Its goal is to find a

function's minimum point, in contrast with the original Bisection Method's goal which is to find a function's root [20].

Algorithm 3.1: Modified Bisection

begin

Input: $x_{d,0}$ and $x_{d,1}$

Parameters: G and Δ

1 Define "Epsilon Function" $\varepsilon_{d,i} = f(x_{d,i})$; $\forall i = 0, 1, 2$

2 Define "Stability Control Function" ω, $x_{d,0}$, $x_{d,1} = stabilityControl(x_{d,0}$, $x_{d,1}$, $left_{count}$, $right_{count})$

3 Declare $g = 0$, $\delta = inf$, $\omega = True$, $left_{count} = 0$, $right_{count} = 0$

4 **while** $g < G$ or $\delta > \Delta$ or ω **do**

5 $x_{d,2} = (x_{d,0} + x_{d,1})/2$

6 $\varepsilon_{d,i}$ = $f(x_{d,i})$; $\forall i = 0, 1, 2$

7 **if** $\varepsilon_{d,0} < \varepsilon_{d,1}$ **and** $\varepsilon_{d,2} < \varepsilon_{d,1}$ **then**

8 $x_{d,1} := x_{d,2}$

9 $left_{count} = left_{count} + 1$

10 $right_{count} = 0$

11 **else**

12 $x_{d,0} := x_{d,2}$

13 $right_{count} = right_{count} + 1$

14 $left_{count} = 0$

15 ω , $x_{d,0}$, $x_{d,1}$ = $stabilityControl(x_{d,0}, x_{d,1}, left_{count}, right_{count})$

16 $\delta = abs(x_{d,1} - x_{d,0})$

17 $g = g + 1$

18 $\hat{x}_d = x_{d,i} : min(\varepsilon_{d,i})$

19 $\varepsilon_d = \varepsilon_{d,i} : min(\varepsilon_{d,i})$

Return $\hat{x}_d, \varepsilon_d$

As written in [20], three initial points for $\hat{x}_d$ have to be taken. In this case, lower, upper and midpoint guesses are $x_{d,0}$, $x_{d,1}$ and $x_{d,2}$, respectively. The last is calculated every iteration in Line 5.

The function f in Algorithm 3.1 Line 1 is called "Epsilon Function". It is done according to Algorithm 3.2. It takes as input for a given value of i a value for $x_{d,i}$ and returns $\varepsilon_{d,i}$, the RMSE among the $\hat{L}_{d,i}$ profiles of the two days of d for a value of i. Load profile $\hat{L}_{d,i}$ is obtained in Line 2 based on the estimation of PV profile $\hat{PV}_{d,i}$ and the net-load profile N_d. The estimation of the PV profile $\hat{PV}_{d,i}$ is obtained in Line 1 using the irradiation measurements I_d. After finding $\hat{L}_{d,i}$ profiles of both days, $\varepsilon_{d,i}$ is calculated in Line 3 among the two days of d for a value of i.

Graphically, "Epsilon Function" can be seen in Figure 3.4. It was plotted by executing a given range of values of $x_{d,i}$ from 0 to 500 kW and the resulting $\varepsilon_{d,i}$ values. Algorithm 3.1's goal is to find the minimum

Algorithm 3.2: Epsilon Function

begin

Input: $x_{d,i}$

1 $\hat{PV}_{d,i} = x_{d,i} I_d$

2 $\hat{L}_{d,i} = N_d + \hat{PV}_{d,i}$

3 $\varepsilon_{d,i} = RMSE(\hat{L}_{d,i})$

Return $\varepsilon_{d,i}$

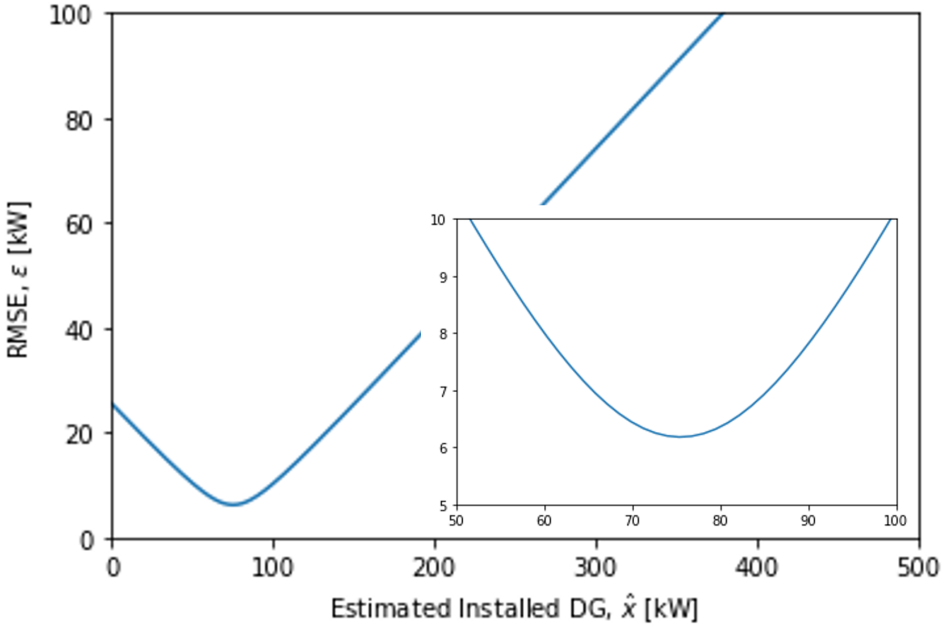
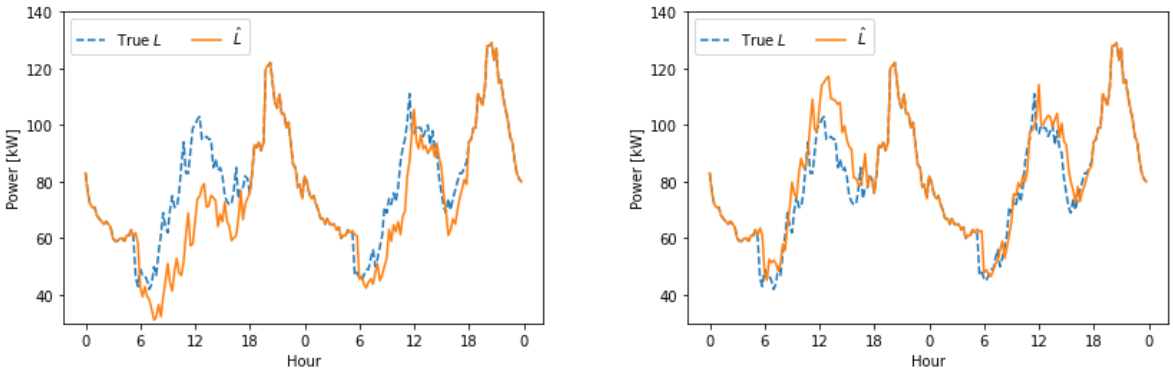


Figure 3.4: Epsilon Function $f(x_{d,i}) = \varepsilon_{d,i}$

point of such function.

Figure 3.5 shows two given cases of Algorithm 3.2 being executed for two different values of $x_{d,i}$ which generate a big (Figure 3.5(a)) and a small value (Figure 3.5(b)) of $\varepsilon_{d,i}$ using the same pair of days, d, in both cases. The RMSE in kilowatts is calculated among day 1 and 2 of $\hat{L}$ following Line 3. The true value, L, is plotted as reference and it represents $f(76) = 6.18$, i.e. the minimum point of the Epsilon Function in this case which is also plotted in Figure 3.4. Figure 3.5 plots the profiles L_d and $\hat{L}_d$ in kilowatts using a pair of days d and the shown cases according to Algorithm 3.2 are $f(34.13) = 14.93$ and $f(88.73) = 7.56$.

There are three stopping criteria for Algorithm 3.1: number of iterations, g, precision, δ, and stability control ω. The first two relate to the parameters G and Δ, the maximum number of iterations and the minimum distance accepted between $x_{d,0}$ and $x_{d,1}$, respectively. The last is calculated in Line 16. ω is a



(a) x_d = 34.13, ε_d = 14.93

(b) x_d = 88.73, ε_d= 7.56

Figure 3.5: Epsilon Function $f(x_{d,i}) = \varepsilon_{d,i}$ with two x_d as input leading to big (a) and small (b) ε_d.

boolean stopping criteria related with the stability of Algorithm 3.1 and is described as follows.

The Stability Control Function is defined in Line 2 and is executed in Line 15. It can be explain graphically by Figure 3.6. In it, star markers represent the testing points $x_{d,0}$, $x_{d,2}$ and $x_{d,1}$ from left to right of the first iteration of Algorithm 3.1 Line 4. The second iteration is the result of shrinking the points to the left according to Lines 7 to 10 and is presented in black point markers in the order $x_{d,0}$, $x_{d,2}$ and $x_{d,1}$ from left to right. This process will continue repetitively until a δ value in Line 16 triggers the precision stopping criteria in Line 4. However, the function's minimum point is still farther to that left edge.

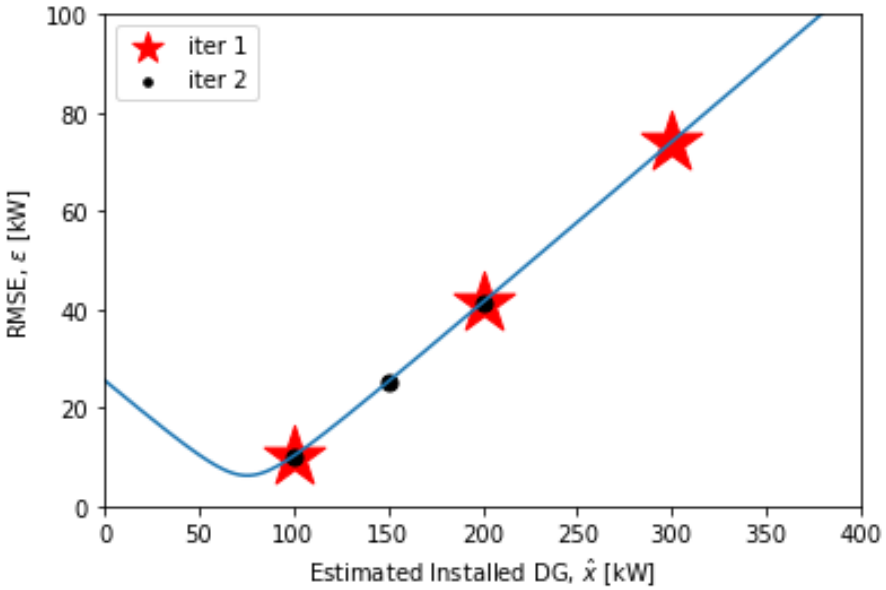


Figure 3.6: Graphical context for Stability Control Function

In this context, Stability Control Function takes action and follows Algorithm 3.3. It takes as input the points $x_{d,0}$ and $x_{d,1}$, and two counters which tell how many times an edge of the Modified Bisection have been shrinking towards a given direction. When the same edge have been shrunken more times than a S parameter towards the same direction, the edges are swamp. This is done either by taking a smaller or bigger value for $\hat{x}_{d,0}$ or $\hat{x}_{d,1}$ using Lines 2 and 3 or Lines 6 and 7, respectively. As one of the edges is fixed, the corresponding counter is reset to 0 in Line 4 or Line 8.

As an additional feature, Algorithm 3.3 remembers which edge it has corrected in the past. This is key given that the stability is lost when the corrections start jumping from side to side. The previous is recognized by Line 15 and will stop the execution of Algorithm 3.1. Otherwise, the process is consider to remain stable as in Lines 13 and 17.

Algorithm 3.3: Stability Control Function

```
   begin
       Input: x_{d,0} , x_{d,1} , right_count and left_count
       Parameter: S
1      if left_count > S then
2          x_{d,1} := x_{d,0}
3          x_{d,0} := x_{d,0}/10
4          left_count = 0
5      else if right_count > S then
6          x_{d,0} := x_{d,1}
7          x_{d,1} := x_{d,1} * 10
8          right_count = 0
9      else
10         x_{d,0} := x_{d,0}
11         x_{d,1} := x_{d,1}
12     if first correction then
13         ω = True
14     else if corrected right then left or vice versa then
15         ω = False
16     else
17         ω = True
       Return ω, x_{d,0}, x_{d,1}
```

Returning to Algorithm 3.1, when any of the stopping criteria in Line 4 is reached, the outputs of the modified bisection are calculated. These are $\hat{x}_d$, the point $x_{d,i}$ with the least corresponding $\varepsilon_{d,i}$ where $i = 0, 1, 2$, and ε_d, the least $\varepsilon_{d,i}$ itself. Those are calculated in Lines 18 and 19.

3.2 DG Size Estimation

The second block of the methodology presented in Figure 3.2 is called "DG Size Estimation". It groups the points obtained in the "Day Comparison" block by identifying a single one as the best representative of the whole set. The input of DG Size Estimation is the collection of points $\hat{x}_d$ and ε_d for $d = 0, .., D$. The output of the block is the "Estimated Installed DG", $\hat{X}$. Such identification process is done according to the origin of the data used in the methodology. Specifically, the PV data in this thesis can be considered synthetic or real as will be described in detail in Chapter 4. Given any of those cases, two algorithms were considered to accomplish the goal of this methodology's block. First a Mode/Cluster algorithm developed by the author and second a classic K-means.

3.2.1 Mode/Cluster Algorithm

This algorithm was developed by the author of this thesis and is done according to Algorithm 3.4. It uses as input the data set produced by "Day Comparison" in Section 3.1. The output of the algorithm is the Estimated Installed DG, $\hat{X}$.

Algorithm 3.4: Mode/Cluster Algorithm

```
begin
    Input: x̂_d ; ∀d = 0, .., D
    Parameter: M
1   m = Mo(x̂_d ; ∀d = 0, .., D)
2   if freq(m) ≥ M then
3       Return X̂ = m
4   else
5       X̂ = Iterative Histogram's Answer
    Return X̂
```

The execution starts in Line 1 by finding the statistical mode, Mo, of the input data. Line 2 checks if the frequency, $freq$, of such mode contents more than $M\%$ of the points. If so, the mode is given as DG Size Estimation's answer. Otherwise, an Iterative Histogram is executed in Line 5. Its answer is consider the Estimated Installed DG, $\hat{X}$.

An iterative histogram is the process of repetitively taking a histogram's most data frequent interval until the distance between the extreme values of such interval is less than a given threshold. In other words, in an recursive way, make a histogram of an initial histogram's most representative interval. The process can be seen in Algorithm 3.5.

Algorithm 3.5: Iterative Histogram

begin

Input: $\hat{x}_d$; $\forall d = 0, .., D$

Parameter: Γ

1 Declare $\gamma = inf$

2 **while** $\gamma > \Gamma$ **do**

3 $n = len(\hat{x}_d)$

4 $k \approx 1 + 3.3log_{10}n$

5 Histogram with k bins

6 k_{max} = Bin with most Frequency

7 $left_{edge} = min(k_{max})$

8 $right_{edge} = max(k_{max})$

9 $\gamma = right_{edge} - left_{edge}$

10 $\hat{x}_d$; $\forall[left_{edge} \leq d \leq right_{edge}]$

Return $(left_{edge} + right_{edge})/2$

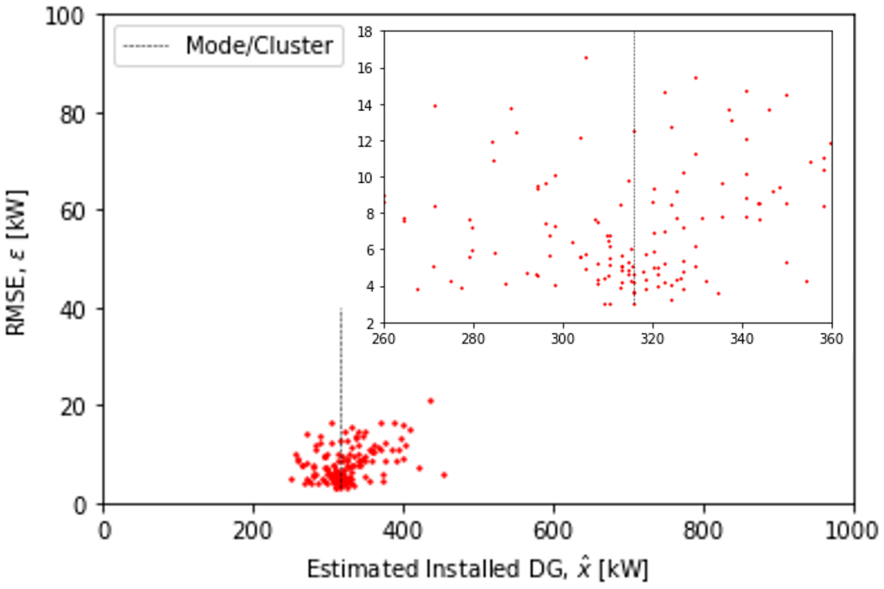


Figure 3.7: DG Size Estimation Output Obtained by Mode/Cluster Algorithm

The algorithm uses as input the data set produced by "Day Comparison" in Section 3.1. As the algorithm's name tells, a histogram is going to be created in Line 5. For it, it is needed the number bins, or classes, k, in which the points are going to be sorted. The previous is done according to [21] which proposes a statistical standard to do so by applying the equation seen in Line 4. It takes the length of the

data set in Line 3 and rounds the answer to the nearest natural number. After the histogram is created, Line 6 selects the interval with the most frequency, i.e. the interval that groups most data points. Lines 7 and 8 store the left and right edges of the k_{max} interval, respectively. The distance between those two points is calculated in Line 9 and is used as stopping criteria in Line 2 according to a precision Γ given as a parameter. Line 10 shortens the data set according to the left and right edges previously defined. This will reduce the size of the data set which will be used in the following iteration. When Algorithm 3.5 stops, it returns the midpoint of the resulting interval k_{max} as the answer.

At this point, Algorithm 3.4 is done and an answer has been returned. It is the "estimated installed DG capacity", $\hat{X}$, according to Figure 3.2. It can be seen graphically in Figure 3.7 given a very grouped data set of points which are the output of Day Comparison algorithm in Section 3.1.

3.2.2 K-means

On the other hand, when the output of Day Comparison algorithm is a more disperse data set of points, a classic K-means aiming for one centroid presents better performance. Its goal is to find a point that minimizes the distance among all the points being analyzed [22, 23]. The previous is done according to Equation (3.1) where a single centroid μ is found in relation with its distance to the values x_d for $d = 0, .., D$.

$$\underset{\mu}{\operatorname{argmin}} \sum_{d=0}^{D} \left\| \hat{x_d} - \mu \right\|^2 \tag{3.1}$$

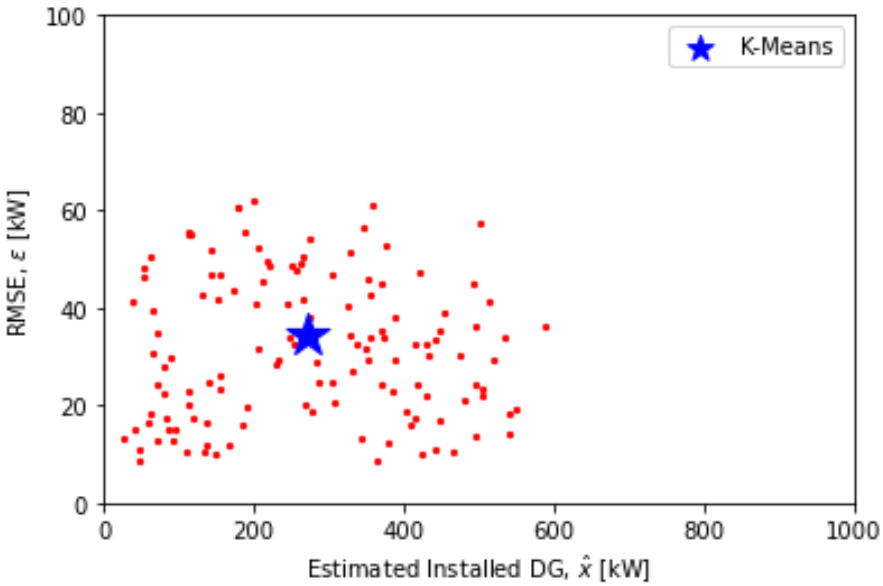


Figure 3.8: DG Size Estimation Output Obtained by K-means Algorithm

The mentioned larger dispersion in the points at the end of Day Comparison algorithm is seen in Figure 3.8. The solution in such case obtained by the K-means algorithm is plotted as well.

At the end, the Estimated Installed DG, $\hat{X}$, is given by using either algorithm, Mode/Cluster or K-means in Section 3.2.1 or Section 3.2.2, respectively.

3.2.3 Disaggregated Profiles

At this point, one output in Figure 3.2 has been found: the Estimated Installed DG, $\hat{X}$. This subsection details the process of finding the remaining one: the Dissaggregated Profiles of the historical N's components, i.e. Estimated Natural Load, $\hat{L}$, and Estimated PV generation, $\hat{PV}$ as seen in Figure 3.1(b).

Net Load, N, is decomposed according to Equation (3.2) where Solar Irradiance, I, is multiplied by $\hat{X}$ to find $\hat{PV}$. $\hat{L}$ is the result of adding N and $\hat{PV}$ as seen in Equation (3.3).

$$\hat{PV} = I\hat{X} \tag{3.2}$$

$$\hat{L} = N + \hat{PV} \tag{3.3}$$

Finally, the two outputs in fig. 3.2 have been calculated and extensively explained. Those are the Estimated Installed DG, $\hat{X}$, and the Dissaggregated Profiles, i.e. the Estimated Natural Load, $\hat{L}$, and the Estimated PV generation, $\hat{PV}$.

3.3 AutoRegressive Model with Exogenous variable

For the proposed methodology, one possible application is the Net Load Forecast explained in this section. The objective is to achieve improvement in both consumption and generation forecasts, when compared to the performance obtained with the same methods while using the original Net Load measurements.

In first place, the model is introduced. Forecast, of any of the profiles, was carried out using a classical Auto-regressive model with a single exogenous meteorological variable, i.e. an ARX model – see (3.4) for an expression of the autoregressive equation [24] where all the symbols are included in the Nomenclature section at the beginning of this document. In the case here presented, the exogenous or external variable used was the Solar Irradiance, I.

$$Y_t = \beta_0 + \sum_{j=1}^{p} \beta_j Y_{t-j} + \sum_{f=0}^{s} \theta_f I_{t-f} + \xi_t \tag{3.4}$$

In ARX model, the variable to predict, Y, is obtained as a linear combination from the previous values

of Y and from the history of an external variable, I. The previous is done in such way to capture the autocorrelation of Y together with its dependence on I. Both coefficients, the autoregression β and the new coefficients for the exogenous variable, θ, are obtained with ordinary least-squares method [25], and are used to forecast future values of Y given future estimates for I.

A scheme representing the inputs and output variables in the ARX model is shown in Figure 3.9.

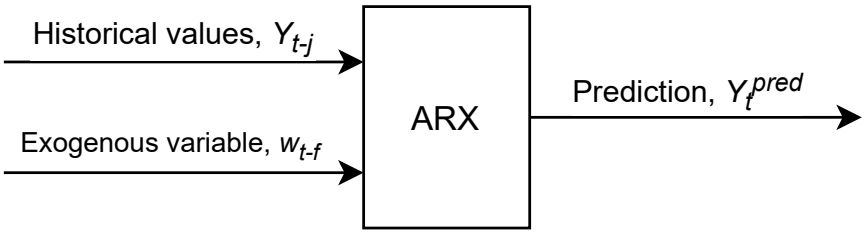


Figure 3.9: AutoRegressive Model with Exogenous variable inputs and outputs

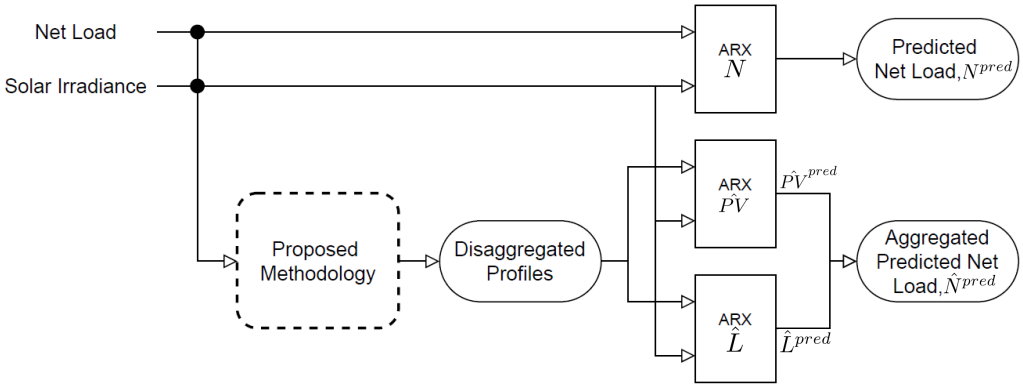


Figure 3.10: Forecast scheme

The ARX model was trained with around six months of data (15 minute granularity profiles) and was used to forecast seven days ahead with 15 minute resolution.

Figure 3.10 shows the use of ARX model after the execution of the proposed methodology. Having as inputs N and I, a first forecast is done. The variable used is N and the result, N^{pred}, represents the base to compare the remaining blocks of ARX model. The initial structure of blocks presented in Figure 3.2 is embedded in Figure 3.10. This by taking N and I, executing all the previously explained procedures and returning the dissaggregated profiles of the N's components, i.e. Estimated PV generation, $\hat{PV}$, and Estimated Natural Load, $\hat{L}$. Each is used as input in the ARX model along with I as the exogenous variable. The predictions are $\hat{PV}^{pred}$ and $\hat{L}^{pred}$, respectively. Both the generation and

consumption predictions were combined to create an aggregated Net Load prediction, $\hat{N}^{pred}$ according to Equation (3.5).

$$\hat{N}^{pred} = \hat{L}^{pred} - \hat{PV}^{pred} \tag{3.5}$$

At this point, there are two cases for forecasting, i.e. $\hat{N}^{pred}$ and N^{pred}, disaggregated and non-disaggregated forecast, respectively. However, a third case is developed. This is refer as partial disaggregation, $\hat{N}^{pred}_{partial}$, since it uses the basic not-disaggregated forecast during the night (18h to 06h) and the new disaggregated forecast during the day-light hours (06h to 18h). The previous is explained in detail in Chapter 5.

4

Study Cases

Contents

4.1 Data Collection 29

4.2 Synthetic PV 32

4.3 Real PV 33

This chapter presents the data and scenarios used to test the proposed methodology described in Chapter 3. Its performance and behavior is analyzed under certain conditions related to quality and variability of the input data.

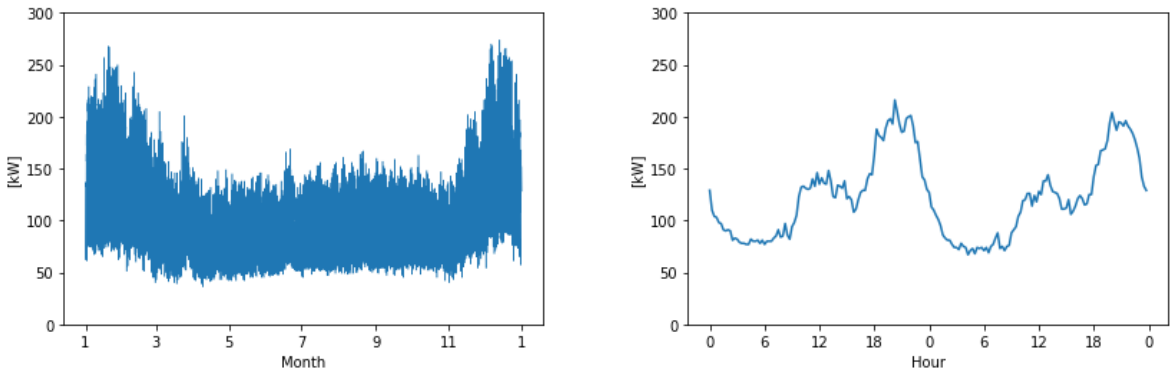
Section 4.1 explains all the matters related with the data being used, i.e. its source, characteristics and considerations to be taken into account when using them throughout the scenarios.

Sections 4.2 and 4.3 are key to understand the scenarios being used throughout the thesis. The two scenarios used in this work differ from each other given the way to obtain the PV data. Those are created according the relation between Solar Irradiance, I, and upstream PV Generation, PV, which can be linear or non-linear. The main remark to highlight is that both resulting Net Load profiles are constructed and therefore are fictitious.

In the first case, a synthetic PV generation, PV_{synt}, was created given a linear relation [26] obtained by scaling irradiance profile shape to meet with the user-defined DG installed capacity, X. Further details will be presented in Section 4.2. On the other hand, the second scenario considered a real PV Generation, PV_{real}, from an on-field laboratory in the south of Portugal [27]. This will be further discussed in Section 4.3 knowing that the relation between I and PV_{real} is non-linear.

Thanks to the way the work was developed, it is possible to test the methodology's performance under various PV penetrations magnitudes in the LV distribution grid. The previous given that the "estimated installed DG capacity", X, is a user-defined variable.

4.1 Data Collection



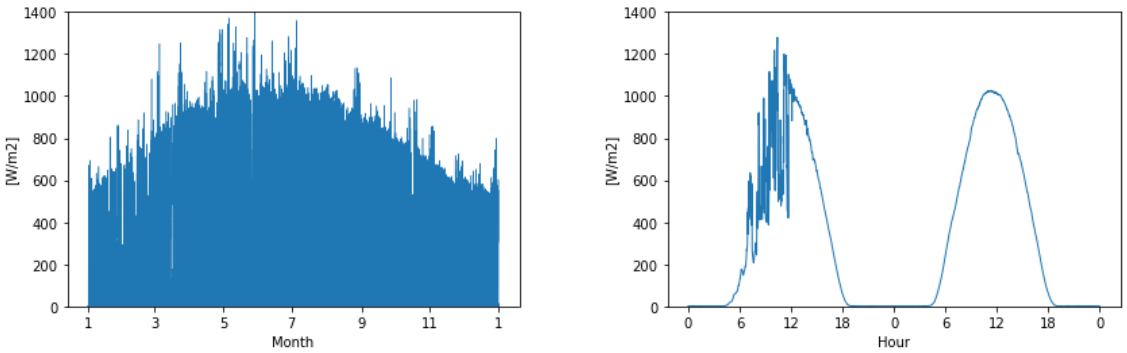
(a) 2017 **(b)** 01-02/02/2017

Figure 4.1: Original True Natural Load profile (a) and detailed (b) as given by a private source

Referring to the methodology's flow diagram in Figure 3.2, there are two input data to consider, Net Load, N, and Solar Irradiance, I. However in this thesis, N is constructed based on a Natural Load,

L, profile. Therefore, N is not collected from a real measurement device, it is fictitious. In contrast, all L profiles being used to construct the N profiles are real and measured from actual on-field devices. Such data was provided by a private source. A L profile is one that only considers the power demand of consumers connected in the LV and, in consequence, is independent from sun light. Such data in this thesis is composed of one-year length, 15-minute granularity measurements of various existing SS Transformers in the center region of Portugal with capacities ranging from 250 kW to 630 kW. As those are L profiles, they are composed of downstream active power measurements from various loads. One of the original and true L profile can be seen in Figure 4.1. The construction of N profile is explained on Sections 4.2 and 4.3 according to each scenarios' characteristics.

Solar Irradiance, I, data is obtained from [28], a weather station located in the city of Faro, Portugal. Such, is a full 2017, one-minute resolution data that can be seen in Figure 4.2.

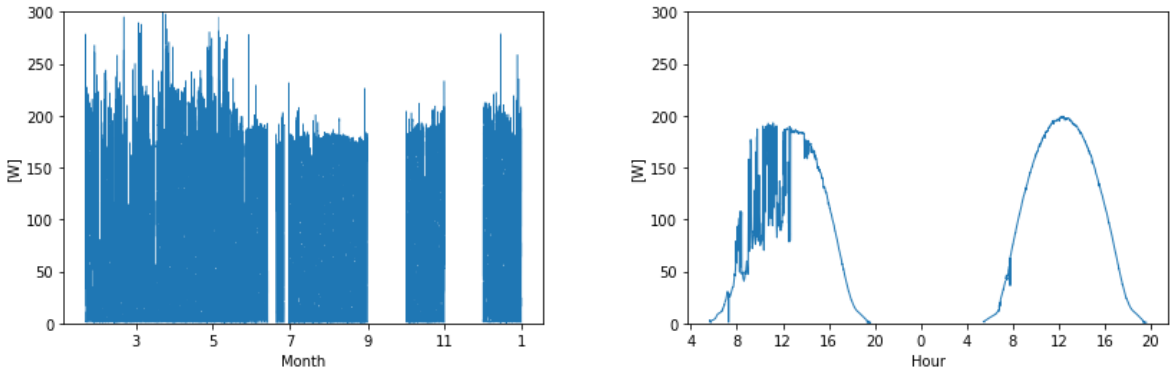


(a) 2017 **(b)** 22-23/05/2017

Figure 4.2: Original True Solar Irradiance profile (a) and detailed (b) as given by the public source

In order to fulfill the creation of the scenario related with real PV data which is further described in Section 4.3, measurements from an actual PV panel were required. A 220 Wp PV panel installed in an on-field laboratory was used and [27] provides its generation measurements. The data is 15-minute granularity from 2017 with the perk of not having data for a couple of months of the year. The previous can be seen in Figure 4.3. This condition limited the time span used during the tests performed as it had to be tailored to the longest possible non-interrupted data span. According to that, it was decided to work in this thesis with data from January to June plus the first week of July, avoiding the incomplete data seen in last part of June and during the second half of the year in Figure 4.3.

Other consideration to take into account while making the scenarios and executing the proposed methodology on them was mentioned in Section 3.1. There, it is stated the possibility to compare a pair days that can be consecutive or not. In that context, this subsection delivers an explanation about the best way to compare data of a pair of days. Specifically, the Net Load N and Natural Load L profiles in a given two days.



(a) 2017 **(b)** 22-23/05/2017

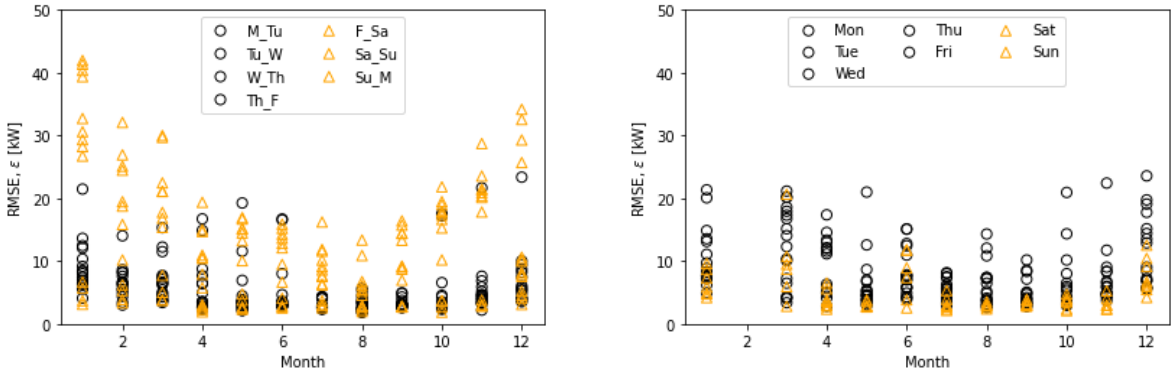
Figure 4.3: Original True PV profile (a) and detailed (b) as given by the public source

The methodology to achieve active power consumption and generation disaggregation at the SS level require the comparison of data. In particular, two days of information of N and Solar Irradiance, I, as inputs in order to execute the Modified Bisection Algorithm in Section 3.1.1. However N profile is very dependent on I given that I is directly related with PV generation which, in turn, changes N. The previous means that a contrasting behaviour of I among two given days will cause the N of the two days to be contrasting as well. Furthermore, such behaviour is expected as sun light in a given location is very unpredictable. Therefore, PV generation of one day can never be assumed similar or equal to other's day PV generation. In consequence, comparing N requires the use of external variables.

Then, the correct comparison might come from a profile which is I independent and is solely affected by the load consumption or the demand. Such profile is the Natural Load, L, profile. By comparing such profile the only remaining question to address is: how similar load consumption or demand is among two days? This can be answer by looking at Figure 4.4. There, the L profile of two days were compared using a RMSE among them. What is seen is a value ε that tells how similar the L or consumption profiles are in a pair of days.

In Figure 4.4(a) two consecutive days were used, i.e. pairs of days such as Monday and Tuesday, Tuesday and Wednesday, and so on. There is a different label for pairs of days involving a weekend day. The last proved to create the biggest values of ε. This tells that consumption is mostly different along the weekends and when comparing a week day with a weekend day. On the other hand, Figure 4.4(b) uses two same day of consecutive weeks, i.e. Monday Week 1 and Monday Week 2, Tuesday Week 1 and Tuesday Week 2, and so on, then Monday Week 2 and Monday Week 3, and so forth. There is also different label for weekend days. When pairing the days like this, all the possible pairs are less than the days in the original data set. As example, in a data set of 6 months, around 180 days, there will be 173 possible pair of days as the original number of days is subtracted in 7 days. In this case an

opposite result is achieved. The least difference in consumption is given among comparing Saturdays with Saturdays and Sundays with Sundays. However, the main difference given by Figure 4.4 is that the values of ε are less in average when the RMSE is calculated by using pairs of the same weekday of consecutive weeks.



(a) **(b)**

Figure 4.4: RMSE in L by comparing two consecutive days (a) and two same days of consecutive weeks (b)

The previous analysis is only possible given the opportunity to use real L measurements as previously described in this section. When dealing with feeders where DG is installed in the LV i.e. when forward and reverse power flows are present, the L profile becomes invisible for a DSO. In such case and since the DSO is only aware of the N measurements, the proposed methodology for disaggregation proves to be helpful to discover that hidden L profile.

4.2 Synthetic PV

Synthetic PV, in this thesis, refers to a linear relation between Solar Irradiance, I, and PV active power generation, PV, i.e. take I data and multiply it by a factor K in order to create a PV_{synt} profile as seen in Equation (4.1). K is calculated in Equation (4.2) according to the user-defined DG installed capacity, X, and the maximum value of I. In such case, PV_{synt} has the same shape as I but scaled in order to reach the desired user-defined size X. After that, the N_{synt} profile is created following Equation (4.3). Profiles N_{synt} and I are the inputs for the methodology according to Figure 3.2.

$$PV_{synt} = IK \tag{4.1}$$

$$K = X/max(I) \tag{4.2}$$

$$N_{synt} = L - PV_{synt} \tag{4.3}$$

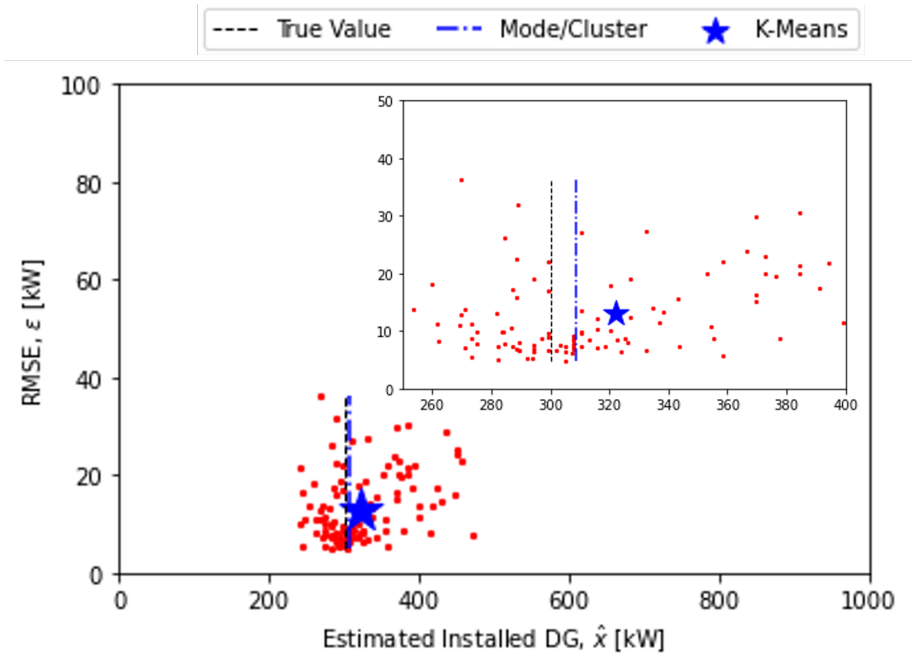


Figure 4.5: Performance of Mode/Cluster and K-Means in a Synthetic PV Scenario

Inputs get processed by the "Day Comparison" block as described in Section 3.1. However, in the second part, when the methodology reaches the "DG Size Estimation" block (Section 3.2), two paths can be taken. Either to use Mode/Cluster Algorithm (Section 3.2.1) or the K-means (Section 3.2.2) to find the Estimated Installed DG, $\hat{X}$. For a Synthetic PV scenario, the decision to use either of them is taken based on Figure 4.5. On it, a very grouped data set of points is seen which is being processed by the mentioned algorithms. On such case the algorithm that provides better results is the Mode/Cluster algorithm. Therefore, for all tests done for a Synthetic PV scenario, the algorithm used from now on in the second block of the methodology is the Mode/Cluster algorithm.

4.3 Real PV

In this scenario, real PV power generation measurements were taken into account, in other words, the relation between I and PV is non-linear. [27] provides the measurements from a 220 Wp PV panel during most of 2017 in 15-minute resolution. The power generation along the time can be seen in Figure 4.3. In addition, a dispersion graph is presented in Figure 4.6 comparing the panel's one month power generation according to Solar Irrandiance, I. It is seen that the panel's behavior is not perfectly linear as previously proposed. In this context, a component of delay and operational uncertainties was included in the disaggregation calculations. Furthermore, Figure 4.7 shows, in more detail, how a real

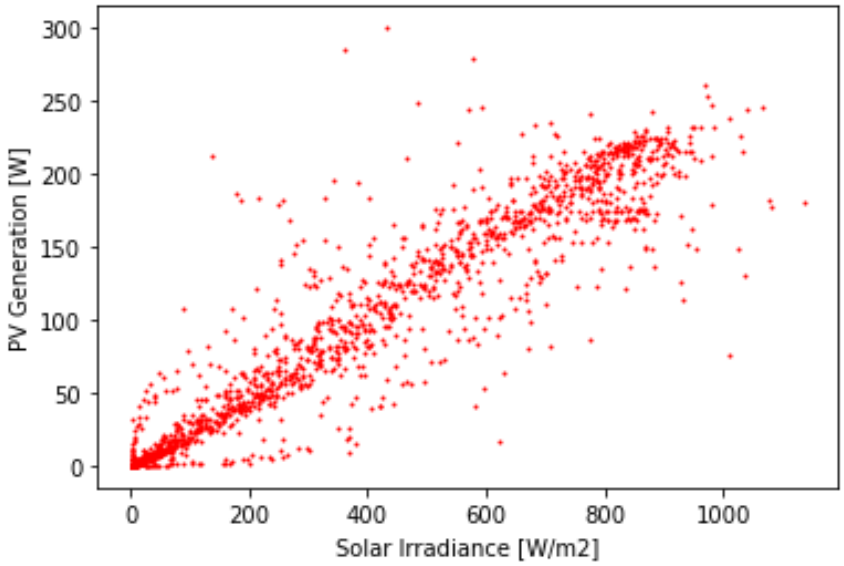


Figure 4.6: Real PV generation on a 220 Wp panel against solar irradiance during one month

PV profiles does not match exactly the I's profile shape. Even though the magnitudes seen in Figures 4.6 and 4.7 are in Watts, those will be consider in kilowatts from now on to match the magnitudes from the original L profiles seen in Figure 4.1.

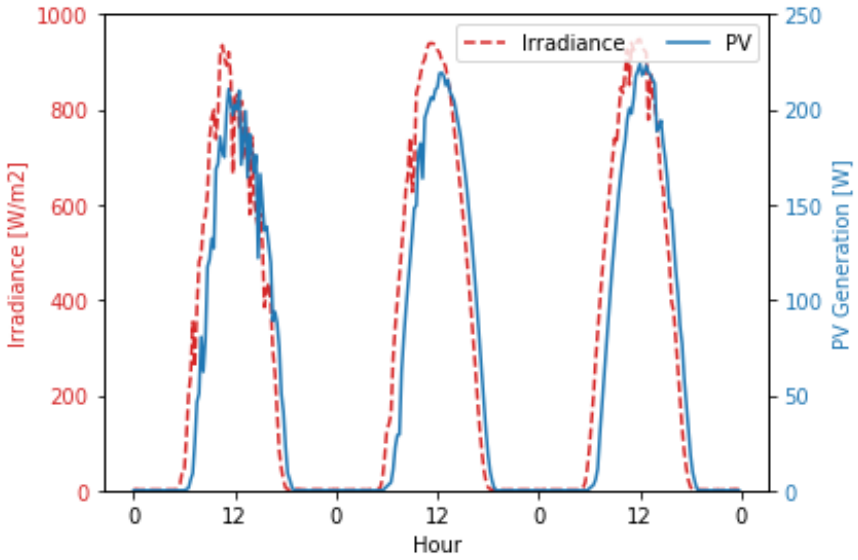


Figure 4.7: Solar Irradiance and Real PV generation on a 220 Wp panel during 3 days

The previous data, $PV_{original}$, was used to construct a PV generation, PV_{real}, profile that can be adjusted according to the user's desired DG installed capacity, X, and the PV panel's peak power. This can be seen in Equation (4.4). Equation (4.5) details the creation of N_{real}, the profile required along side with I to fulfill Figure 3.2 inputs.

$$PV_{real} = PV_{original} X/220 \tag{4.4}$$

$$N_{real} = L - PV_{real} \tag{4.5}$$

When the methodology reaches the "DG Size Estimation" block, once again, a decision about which algorithm to use has to be taken. Figure 4.8 illustrates for a Real PV scenario how dispersed the points in a data set are. Therefore, in this scenario the algorithm chosen to be used from now on is the K-means.

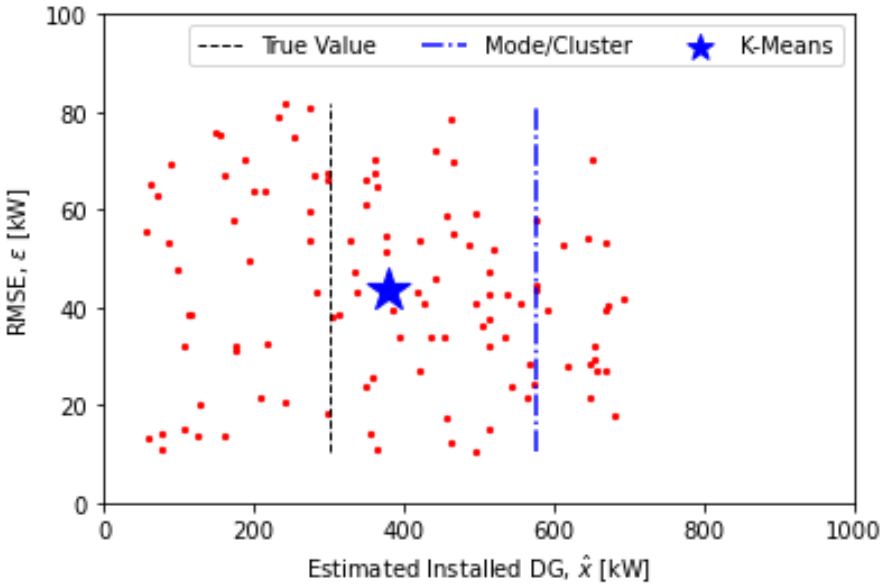


Figure 4.8: Performance of Mode/Cluster and K-Means in a Real PV Scenario

5

Results

Contents

5.1 Disaggregation 39
5.2 Load Forecasting 43

After describing how the scenarios are created in Chapter 4, this chapter presents and discusses the results obtained by executing the proposed methodology in Section 5.1. In addition, the results collected by computing load forecast along with the full, partial, or no use of disaggregation processes are presented and discussed in Section 5.2.

5.1 Disaggregation

Ten existing SS Transformers' data from the center region of Portugal were used by applying 5 different PV capacities according to the transformer's installed capacity for the purpose of recognizing that DG installed capacity and produce the disaggregated profiles as explained previously by seeing Figure 3.2 outputs. The corresponding results for each scenario are shared in Sections 5.1.1 and 5.1.2, respectively, and a discussion about them is presented in Section 5.1.3.

5.1.1 Synthetic PV

The methodology was tested in 10 SS by applying 5 different PV sizes corresponding to 10%, 25%, 50%, 75% and 100% of the SS Transformers' capacity. Table 5.1 shows the best, worst and average results, obtained by the proposed methodology, for the DG installed capacity estimation.

Table 5.1: Estimated Installed DG Capacity Results at SS Level under Synthetic PV Scenario

% of capacity	Best [kW]		Worst [kW]		Average error [%]
	True (X)	Estimated ($\hat{X}$)	True (X)	Estimated ($\hat{X}$)	
10	63.00	65.52	25.00	30.03	1.29
25	157.50	162.44	62.50	66.89	0.75
50	315.00	309.87	125.00	129.68	0.92
75	472.50	473.68	187.50	191.11	1.07
100	630.00	635.38	250.00	248.44	1.85

The algorithm proved to be effective and highly accurate in the tests it was subjected by applying different load profiles magnitudes and DG installed capacities, X. In average, the "estimated installed DG capacity", $\hat{X}$, deviates $\pm$5.65 kW from the true value, X. This represents a 98.82% accuracy in the DG installed capacity estimation.

5.1.2 Real PV

As done previously, the methodology was tested in the same 10 feeders by applying five different PV sizes corresponding to 10%, 25%, 50%, 75% and 100% of the SS Transformers' capacity.

The results can be seen in Table 5.2. As expected, the methodology's performance decreased in accuracy. In average, the "estimated installed DG capacity", $\hat{X}$, deviates $\pm$67.44 kW from the actual

Table 5.2: Estimated Installed DG Capacity Results at SS Level under Real PV Scenario

% of capacity	Best [kW]		Worst [kW]		Average error [%]
	True (X)	Estimated ($\hat{X}$)	True (X)	Estimated ($\hat{X}$)	
10	63.00	83.84	25.00	61.35	7.67
25	157.50	195.05	62.50	95.73	9.65
50	315.00	373.46	125.00	168.45	13.77
75	472.50	559.85	187.50	242.51	17.87
100	630.00	742.77	250.00	316.67	21.67

value of X. This represents a 85.86% accuracy in the DG installed capacity estimation. Even though the effectiveness of the methodology in this second study was reduced, it can be considered accurate given the real world imperfections.

5.1.3 Discussion

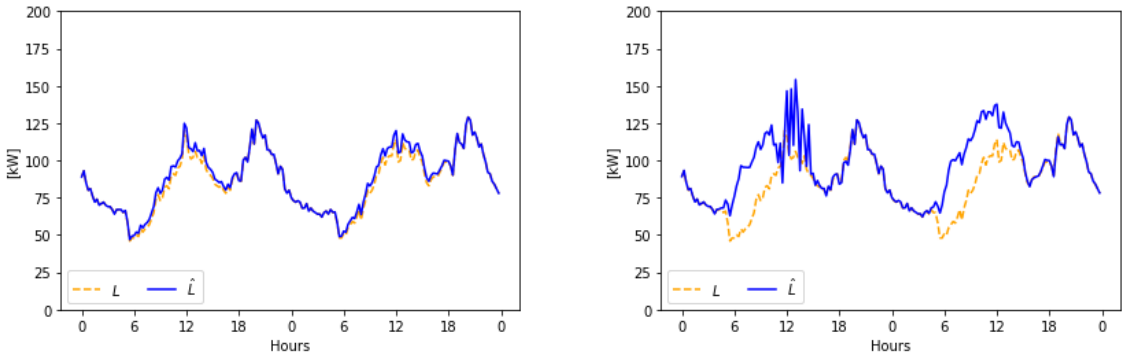
Table 5.3 presents a summary from the information given in Tables 5.1 and 5.2 about the results of the estimated installed DG capacity, $\hat{X}$. According to those, one can see that the PV panel's performance affects the final results. In a perfect information scenario, such as the one illustrated in Section 5.1.1 using synthetic PV, a 98.82% accuracy was achieved. This contrasts with the 85.86% accuracy seen in Section 5.1.2 when using imperfect information from real PV data. Nevertheless, the results achieved will bring value to a DSO and to other stakeholder avoiding the installation and maintenance costs related to the deployment of sensors across the grid.

Table 5.3: Average Error [%] on Results of Estimated Installed DG Capacity at SS Level

% of capacity	Synthetic PV	Real PV
10	1.29	7.67
25	0.75	9.65
50	0.92	13.77
75	1.07	17.87
100	1.85	21.67

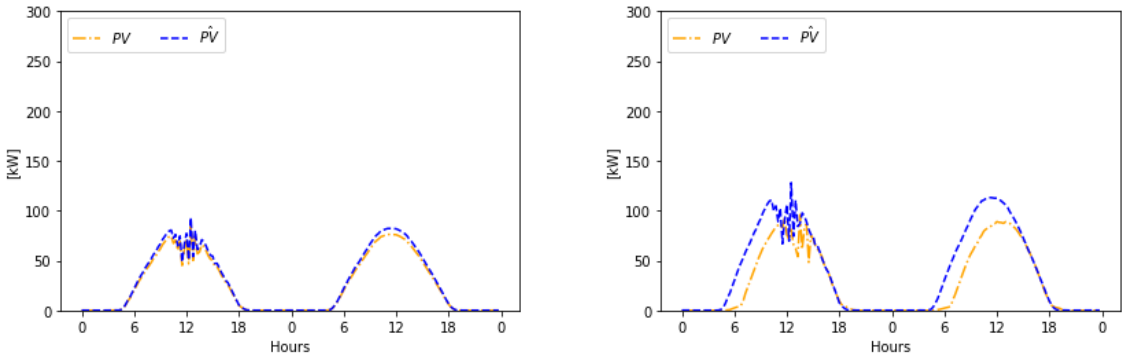
The other output of the methodology is the disaggregated profiles which are here presented in two specific cases. Both correspond to feeder a containing a SS Transformer rated at 400 kW but considering a 25% and 75% PV penetration. In the first case, the DG installed capacity, X, value was 100 kW and the found $\hat{X}$ were 107.84 and 148.43 kW under the Synthetic and Real PV scenario, respectively. In the second case a 300 kW DG installed capacity, X, was set and the found values of $\hat{X}$ were 308.51 and 377.78 kW under the Synthetic and Real PV scenario, respectively. This exemplifies what is described above about how Real PV scenario results are more distant from the actual value.

The estimated Natural Load, $\hat{L}$, and the estimated PV generation, $\hat{PV}$, profiles can be seen in Fig-



(a) Synthetic PV Scenario **(b)** Real PV Scenario

Figure 5.1: True and Estimated Natural Load Profiles with 25% PV penetration

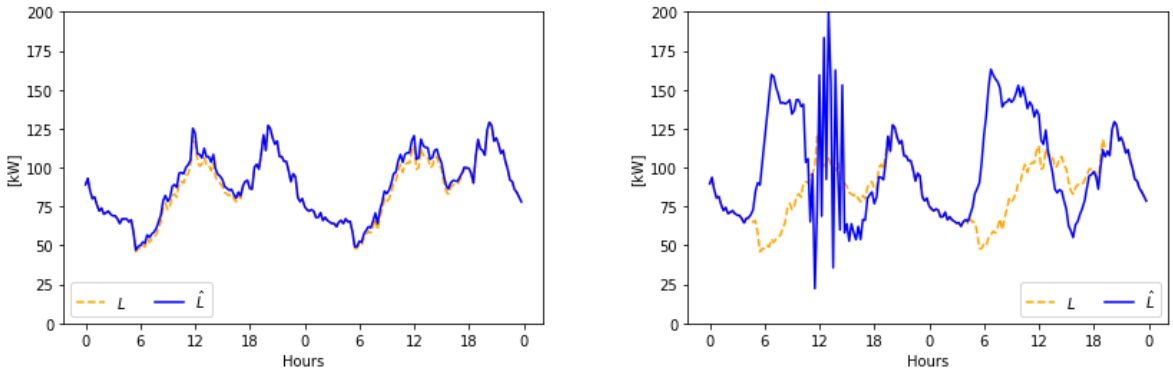


(a) Synthetic PV Scenario **(b)** Real PV Scenario

Figure 5.2: True and Estimated PV Generation Profiles with 25% PV penetration

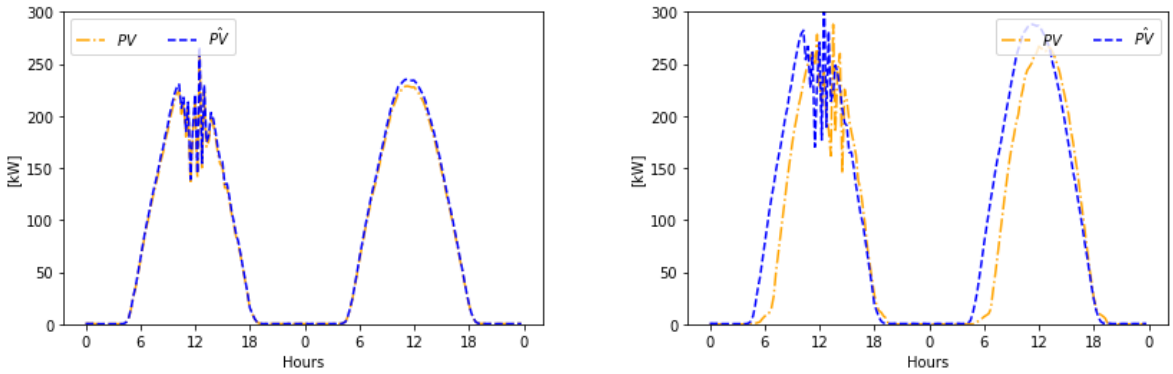
ures 5.1 and 5.2 under both scenarios, Synthetic and Real PV, for the feeder with 25% PV penetration. It is seen how the disaggregated feeders follow, in one scenario better than in the other, the true profiles of L and PV which are plotted as reference. In addition, it is seen the direct effect that the $\hat{PV}$ shape produces on the $\hat{L}$ shape. If the first varies a lot, the second will follow. Is also worth mentioning that the PV profiles seen in Figure 5.2 differ among the scenarios given the different way to obtain them. Recall that, in the Synthetic PV scenario, the PV profile is created with the same shape as the I profile, while, in the Real PV scenario, actual measurements from a PV panel were used.

Regarding the feeder with 75% PV penetration, the estimated Natural Load, $\hat{L}$, and the estimated PV generation, $\hat{PV}$, profiles can be seen in Figures 5.3 and 5.4 under both scenarios, Synthetic and Real PV, respectively. As the PV is bigger, the same variation in solar behavior will cause a much more noticeable effect in the $\hat{PV}$ profile and, in consequence, in the $\hat{L}$ profile. Again, the PV profiles seen in Figure 5.4 differ among the scenarios given the different way to obtain them as previously explained.



(a) Synthetic PV Scenario **(b)** Real PV Scenario

Figure 5.3: True and Estimated Natural Load Profiles with 75% PV penetration



(a) Synthetic PV Scenario **(b)** Real PV Scenario

Figure 5.4: True and Estimated PV Generation Profiles with 75% PV penetration

The presented disaggregated profiles, $\hat{L}$ and $\hat{PV}$, are used as inputs for the Load Forecast calculations shown in Section 5.2.

In addition, Figure 5.5 shows the estimation' errors in DG installed capacity according to the percentage of PV penetration in the corresponding SS Transformer. The DG capacity does not have effect when using synthetic PV information. In that case, the methodology is very stable showing high precision. However, it does affect the methodology when dealing with real PV data. The bigger the DG installed capacity the more average percent error is expected.

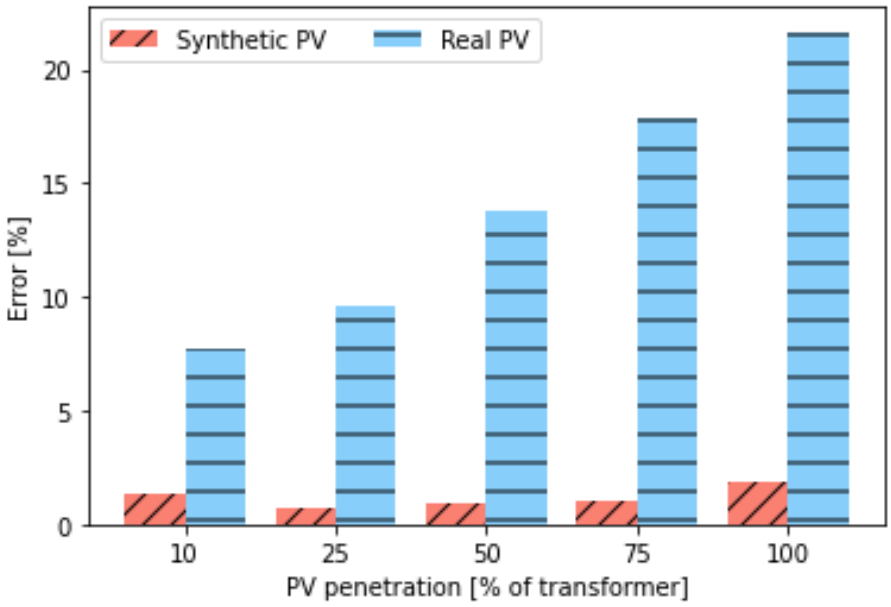


Figure 5.5: Estimated DG Installed Capacity, $\hat{X}$, Average Errors according to PV penetration

5.2 Load Forecasting

The main goal of this section is to evaluate if the proposed methodology led to an improvement in both the consumption and generation forecasts when compared to the performance obtained with the same methods while using the original Net Load measurements. The Estimated Natural Load, $\hat{L}$, and Estimated PV Generation, $\hat{PV}$, profiles obtained with the disaggregation process were used as input for the consumption and generation forecast methods. Forecast were carried out using a classical Auto-regressive model with a single exogenous meteorological variable, i.e. an ARX model – see Section 3.3. In this case, the exogenous external variable used was the Solar Irradiance, I.

The ARX model was executed according to the dates in Table 5.4. It can be seen that the model was trained with around six months of 15-minute data and was used to forecast seven days ahead with 15-minute resolution as well. As previously explained in Section 3.3 and Figure 3.10, three variables were forecast: N, $\hat{PV}$ and $\hat{L}$ and the results were: N^{pred}, $\hat{PV}^{pred}$ and $\hat{L}^{pred}$, respectively. Afterwards, the $\hat{PV}^{pred}$ and $\hat{L}^{pred}$ were combined to create aggregated Net Load prediction, $\hat{N}^{pred}$ according to Equation (3.5).

Table 5.4: Dates to execute the ARX model

Training dates		Forecasting dates	
21/01/2017	6/06/2017	7/06/2017	13/06/2017

At this point, there are two cases for forecasting, i.e. $\hat{N}^{pred}$ and N^{pred}, disaggregated and non-disaggregated forecast, respectively. However, a third case is developed. This is refered as partial

disaggregation, $\hat{N}^{pred}_{partial}$, since it uses the basic not-disaggregated forecast, N^{pred}, during the night (18h to 06h) when solar irradiance is zero or very small and the new disaggregated forecast, $\hat{N}^{pred}$, during the day-light hours (06h to 18h) when irradiance is not negligible.

The evaluation made to prove the effectiveness of forecasting Net Load by the use and partial use of disaggregation is through a comparison. It is made using as reference the base case of N forecasting which uses N and I as inputs to create N^{pred}. Therefore, the forecast errors produced by $\hat{N}^{pred}$ and $\hat{N}^{pred}_{partial}$ against N^{pred} are used to measure the performance of the proposed methodology when applied to Load Forecasting.

The previous evaluation is done under the two scenarios considered in this thesis, Synthetic PV and Real PV, in Sections 5.2.1 and 5.2.2, respectively. In each section, two detailed cases are presented by plotting all the Net Load predictions above explained, i.e. N^{pred}, $\hat{N}^{pred}$ and $\hat{N}^{pred}_{partial}$. The two cases are the same mentioned in Section 5.1.3. They correspond to a 25% and 75% PV penetration in a feeder with a SS Transformer rated at 400 kW.

5.2.1 Synthetic PV

The same 10 SS and 5 PV sizes used in Sections 5.1.1 and 5.1.2 were used to generate load forecasts according to Section 3.3. Therefore, three predictions, N^{pred}, $\hat{N}^{pred}$ and $\hat{N}^{pred}_{partial}$, were obtained per feeder for each PV size. The separate results were collected and averaged in order to get three values per feeder, i.e. the mean error among the PV sizes when disaggregation is not used, fully used and partially used.

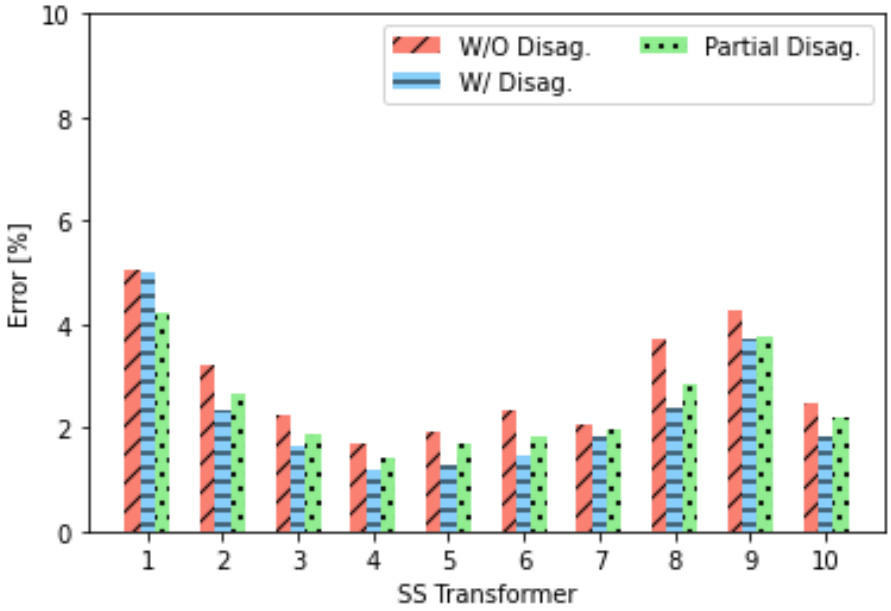
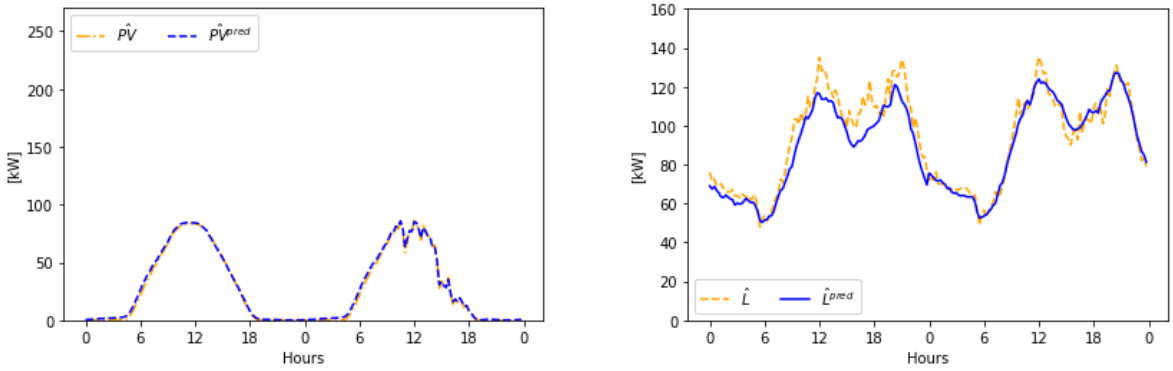


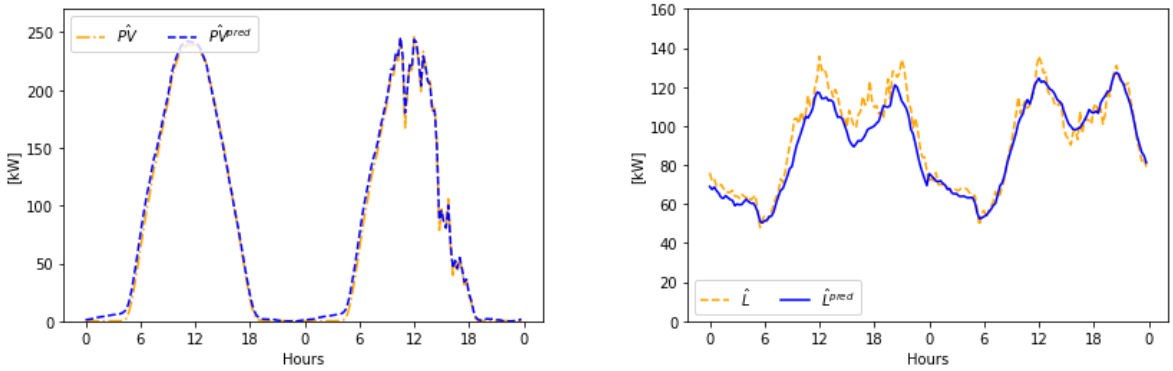
Figure 5.6: Effect of Disaggregation on Net Load Forecast Results under Synthetic PV Scenario

The previous led to prove that using disaggregation calculations before forecasting improves the predictions independently of the size of the associated Power Transformer and regardless if disaggregation is used fully or partially. Figure 5.6 provides a comparison between the average error in the 10 SS Power Transformers load forecast without, partially and fully using dissagregation calculations. A constant improvement is seen when using the methodology proposed in this thesis. Even though in this case, the largest change was achieved by using full disaggregated forecast resulting in an improvement of 0.63 pp compared with the 0.45 pp improvement achieved when using partial disaggregation. At the end, a trend about reducing predictions errors is achieved.



(a) **(b)**

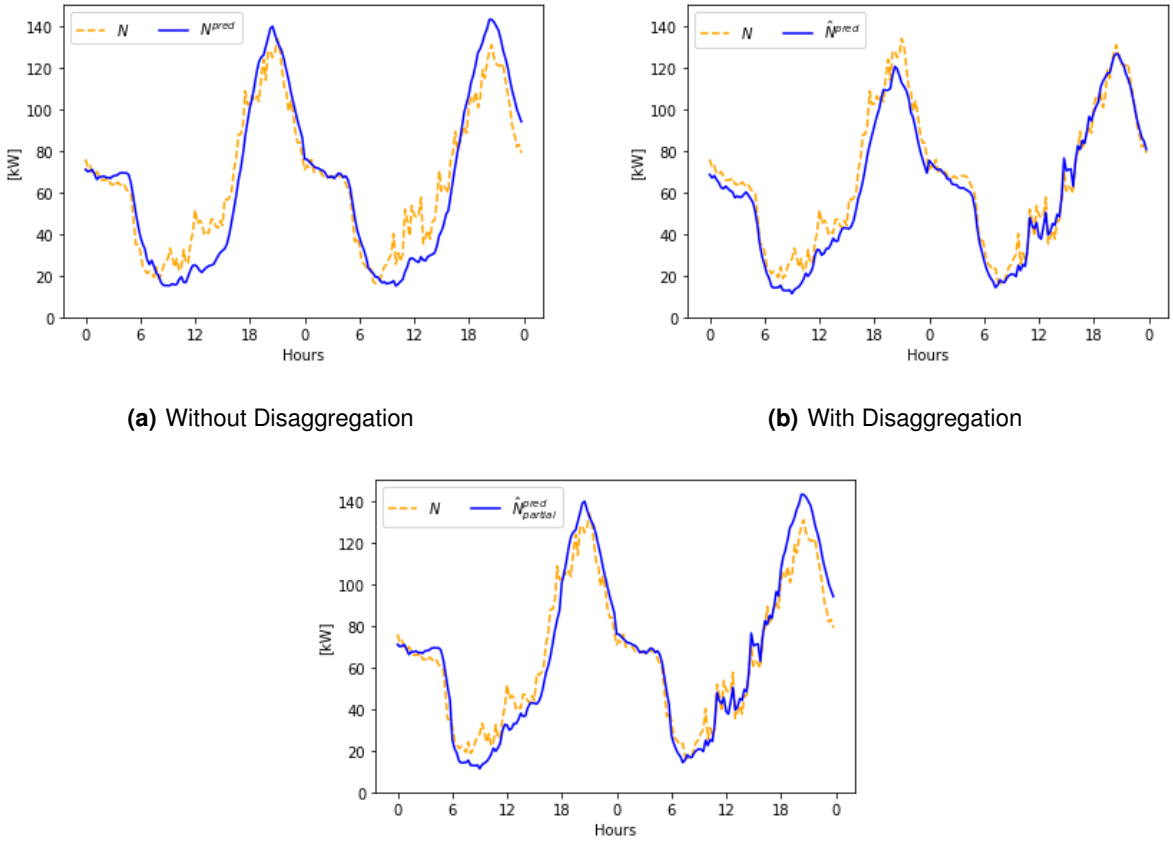
Figure 5.7: Estimated PV Generation (a) and Estimated Natural Load (b) Forecast Under Synthetic PV Scenario with 25% PV Penetration



(a) **(b)**

Figure 5.8: Estimated PV Generation (a) and Estimated Natural Load (b) Forecast Under Synthetic PV Scenario with 75% PV Penetration

Given the broad picture of the results shown, let some detailed results be presented. Figures 5.7 and 5.8 shows two predicted days of Estimated PV Generation and Estimated Natural Load in a 400 kW SS Transformer when 25% and 75% PV penetration happens, respectively. Notice that the Estimated Natural Load in both cases turned out to be very similar given that this profile is only affected by consumption which didn't change while applying more DG in the feeder.



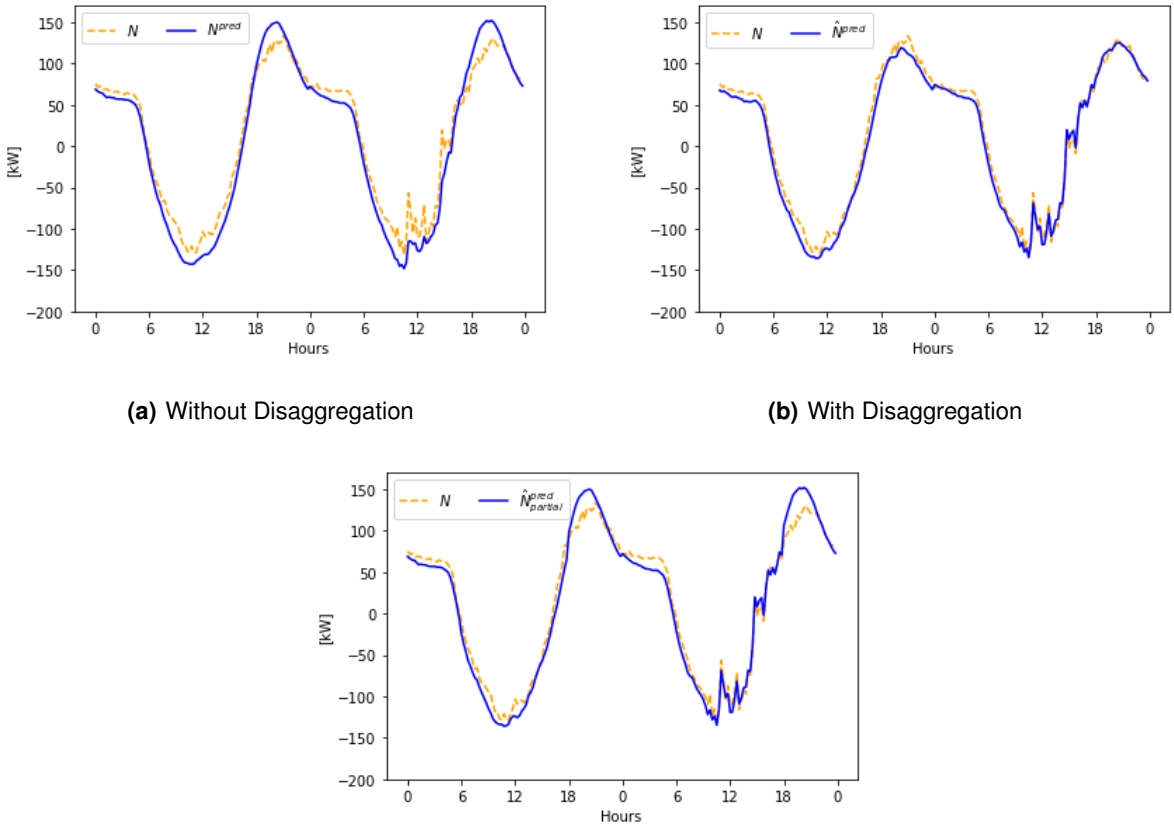
(a) Without Disaggregation

(b) With Disaggregation

(c) With Partial Disaggregation

Figure 5.9: Net Load Forecast Under Synthetic PV Scenario with 25% PV penetration

Figures 5.9 and 5.10 present the three predictions, N^{pred}, $\hat{N}^{pred}$ and $\hat{N}^{pred}_{partial}$ in the same 400 kW SS Transformer when 25% and 75% PV penetration happens, respectively. The aggregated Net Load prediction, $\hat{N}^{pred}$, is the result of combining $\hat{PV}^{pred}$ and $\hat{L}^{pred}$ according to Equation (3.5). Only reverse power flow is seen in the second case given the biggest magnitude of the associated PV generation in comparison with a relatively low magnitude in consumption by the users of the feeder.



(a) Without Disaggregation

(b) With Disaggregation

(c) With Partial Disaggregation

Figure 5.10: Net Load Forecast Under Synthetic PV Scenario with 75% PV penetration

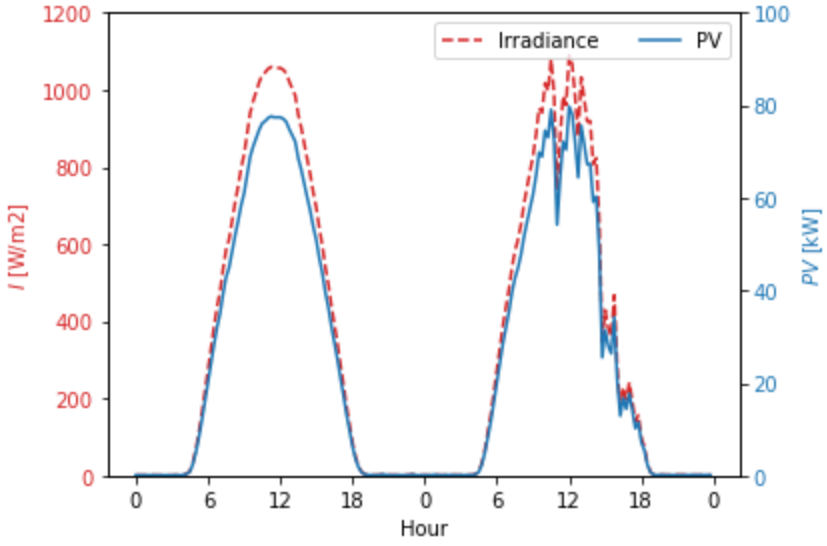


Figure 5.11: Solar Irradiance and PV Generation Under Synthetic PV Scenario

Also, there is a clear improvement in predictions when using either full or partial disaggregation seen in the way the predictions follow the actual profile plotted in dashed lines. Specially during the second day plotted as it happened to have a more distorted I profile and, in consequence a more changing PV profile. The previous can be seen in Figure 5.11.

5.2.2 Real PV

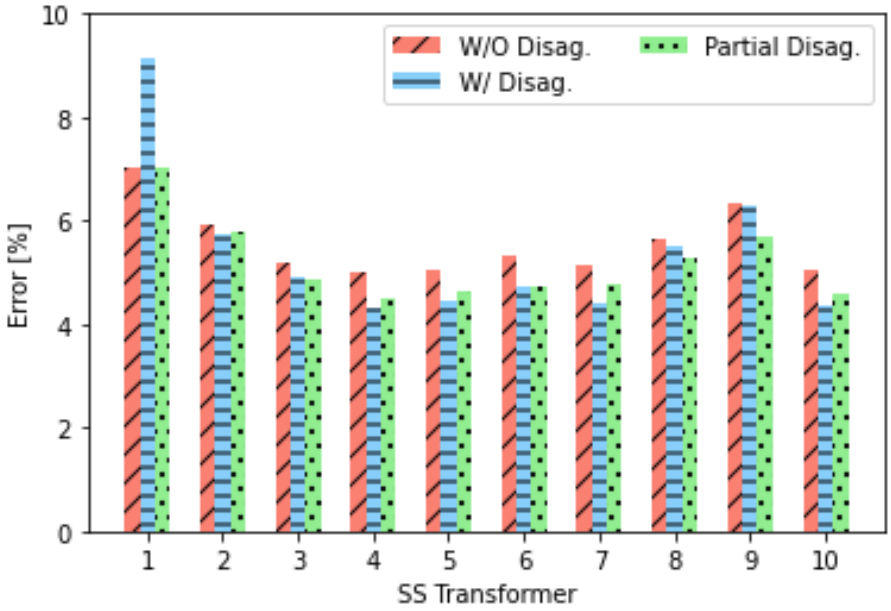
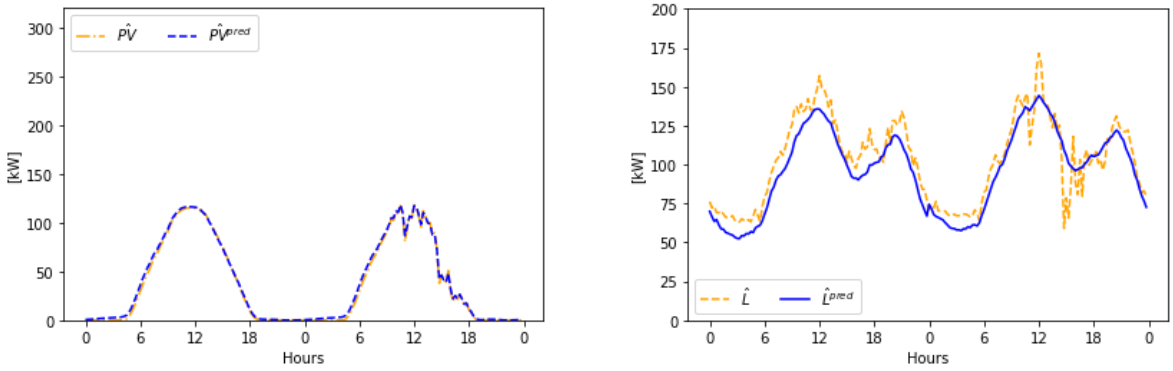


Figure 5.12: Effect of Disaggregation on Net Load Forecast Results under Real PV Scenario

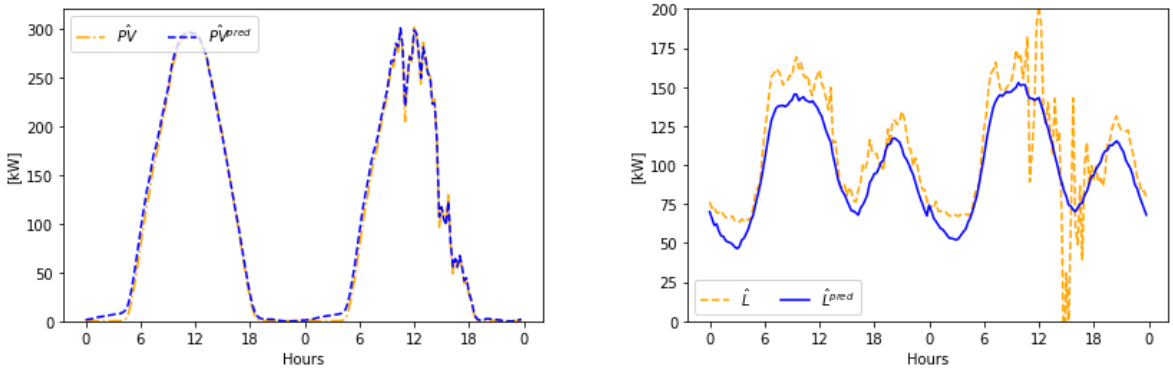
By using the same feeders and PV sizes as in Sections 5.1.1 and 5.1.2, load forecasting was run to illustrate the benefit from using disaggregation before predictions are made. Again, three predictions, N^{pred}, $\hat{N}^{pred}$ and $\hat{N}^{pred}_{partial}$, were obtained per feeder for each PV size. And after averaging those, three mean errors were obtained per SS concerning the non-use, full-use and partial-use of disaggregation. Those are presented in Figure 5.12.

At first sight, the relation among every SS Transformer illustrates the same relation as the one seen in Figure 5.6, however with bigger magnitude of prediction errors in the Real PV scenario. When considering a more realistic case study such as the Real PV scenario, the method that performed better involved the partial use of disaggregation. The improvement in percentage error was of 0.38 pp in comparison with the change of 0.19 pp given when full disaggregation is used. Although every value presented in Figure 5.12 is higher than those shown in Figure 5.6, a trend in decreasing the predictions' error value when using full or partial disaggregation in comparison with not using it is clearly seen in both figures.



(a) **(b)**

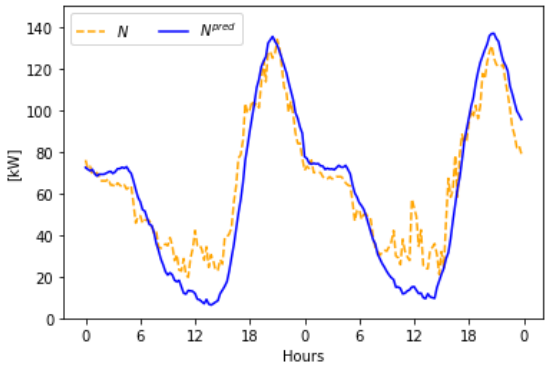
Figure 5.13: Estimated PV Generation (a) and Estimated Natural Load (b) Forecast Under Real PV Scenario with 25% PV Penetration



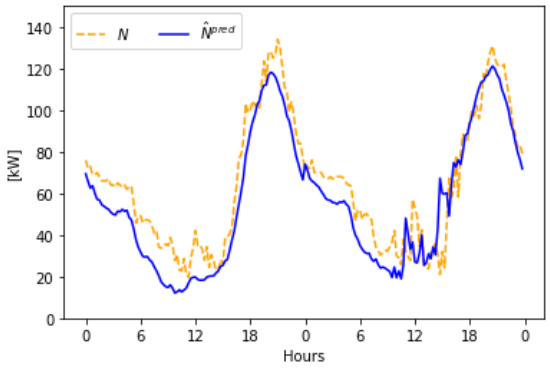
(a) **(b)**

Figure 5.14: Estimated PV Generation (a) and Estimated Natural Load (b) Forecast Under Real PV Scenario with 75% PV Penetration

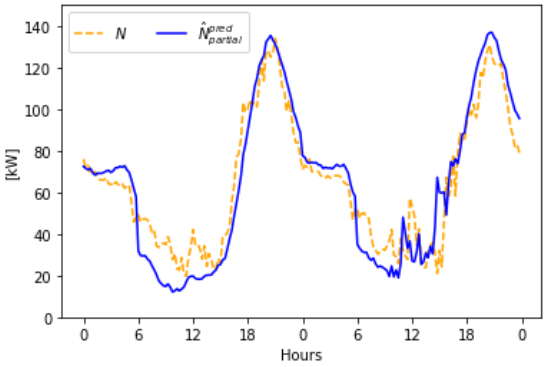
Again, more detailed results are presented. Figures 5.13 and 5.14 shows two predicted days of Estimated PV Generation and Estimated Natural Load in a 400 kW SS Transformer when 25% and 75% PV penetration happens, respectively. In this scenario, the similarity between the Estimated Natural Load in both cases is not seen at first sight. This given that the Estimated Natural Load is the result of disaggregation calculations which result in a shape highly related to the solar irradiance shape. In an ideal framework Estimated Natural Load is independent of solar irradiance.



(a) Without Disaggregation



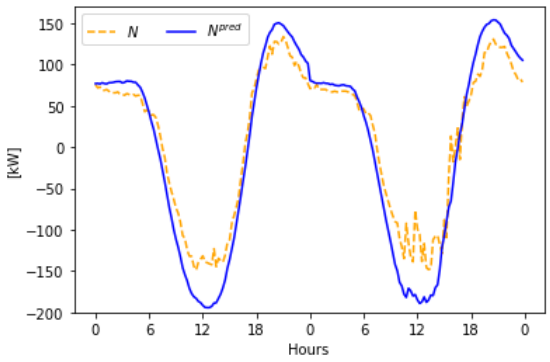
(b) With Disaggregation



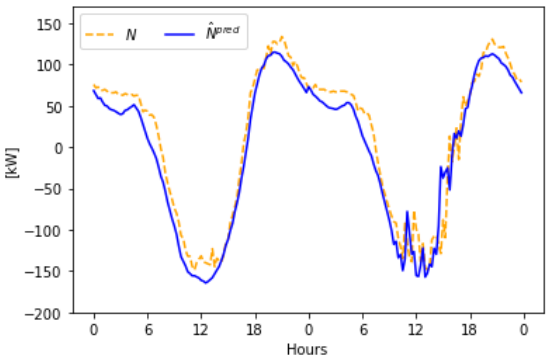
(c) With Partial Disaggregation

Figure 5.15: Net Load Forecast Under Real PV Scenario with 25% PV penetration

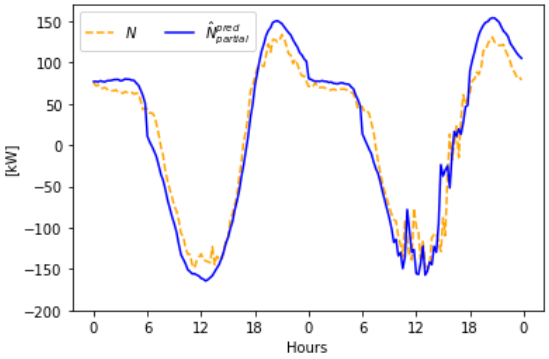
The same 400 kW SS Transformer is used and Figures 5.15 and 5.16 show how the predictions are when 25% and 75% PV penetration happens, respectively. As in the Synthetic PV scenario, reverse power flow only happens in this case when the associated PV generation overcomes drastically the consumption by the users of the feeder. Also, the predictions improve on the second day in the graphs presented given the same variation in the I profile which, in consequence, affects the PV profile. The previous is seen in Figure 5.17.



(a) Without Disaggregation



(b) With Disaggregation



(c) With Partial Disaggregation

Figure 5.16: Net Load Forecast Under Real PV Scenario with 75% PV penetration

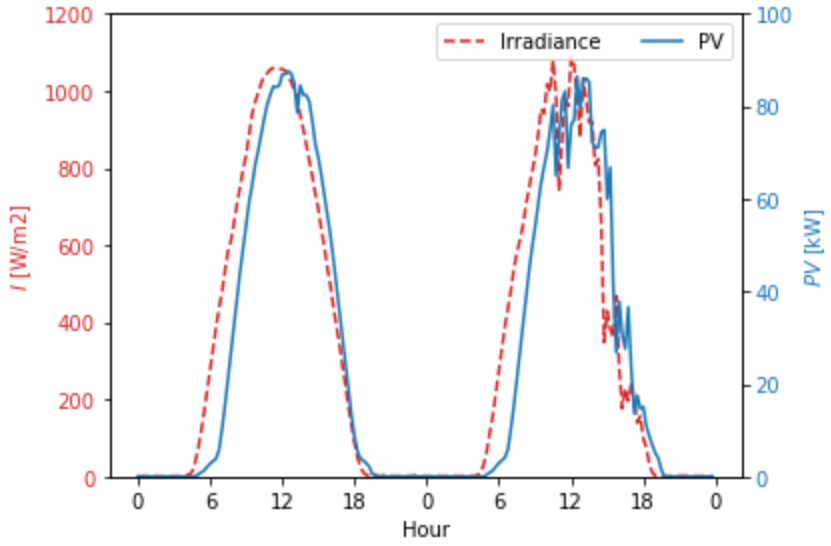


Figure 5.17: Solar Irradiance and PV Generation Under Real PV Scenario

5.2.3 Discussion

Regarding the application of disaggregation to load forecasting, Table 5.5 summarizes the previously presented results. It is clear from such results that, by applying the methodology presented in this thesis, a constant improvement is achieved either by using full or partial disaggregation.

Table 5.5: Mean Errors [%] on Net Load Forecast

	Without Disaggregation	With Disaggregation	With Partial Disaggregation
Synthetic PV	2.90	2.27	2.45
Real PV	5.57	5.38	5.19

The contrast between errors obtained under synthetic and real PV scenario in Figures 5.6 and 5.12 deserves some notice. Disaggregation proved to be the most effective method to increase the accuracy of load forecast under the controlled environment such as the constructed synthetic PV scenario. However, when moving the testing procedures to a more realistic environment under the real PV scenario, disaggregation let to a decrease in N forecast accuracy in a few SS transformers. Such results led to the implementation of a partial disaggregation approach, in an effort to return to the improvement path found for the previous case study. At the end, either by using disaggregation calculations for daytime only or for the whole day, disaggregation was always capable of decreasing the average error under the real PV scenario.

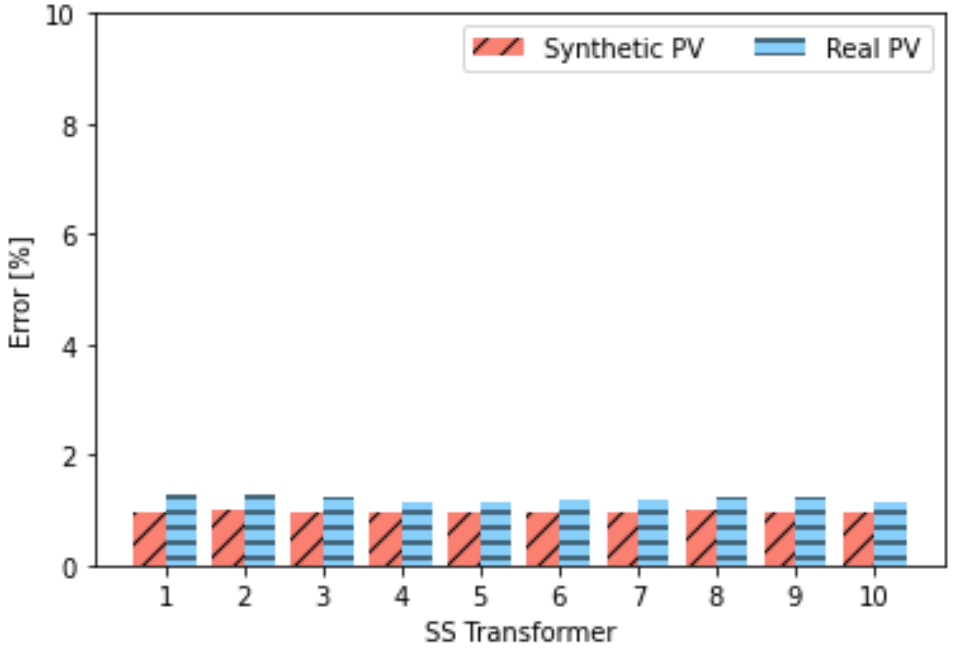


Figure 5.18: Estimated PV Generation Prediction, $\hat{PV}^{pred}$, Average Errors under Synthetic PV and Real PV scenario

After noticing the results on Net Load forecasting presented in Figures 5.6 and 5.12, a deeper investigation was done. In essence, to look at the separate forecast of the estimated PV generation, $\hat{PV}$, and the estimated Natural Load, $\hat{L}$. On first place, $\hat{PV}^{pred}$ average errors for the 10 SS Transformers are seen in Figure 5.18. Regardless of which scenario is used, the average errors are very low and constant in all the analyzed feeders. However, the same does not happen with $\hat{L}^{pred}$ average errors seen in Figure 5.19. There, the errors significantly increase when applying the Real PV scenario and the behavior among the 10 SS Transformers varies as well. That behaviour highly conditions the Net Load predictions observed Figures 5.6 and 5.12 no matter if disaggregation calculations are not used, partially or fully used. Nevertheless, the previous is only caused by the forecast model used and any improvement in such lies outside the scope of this thesis. What is worth mentioning is that, without taking into account the performance of the forecasting model itself, the use or partial use of the presented disaggregation methodology reduced the average prediction errors of Net Load as seen in Table 5.5.

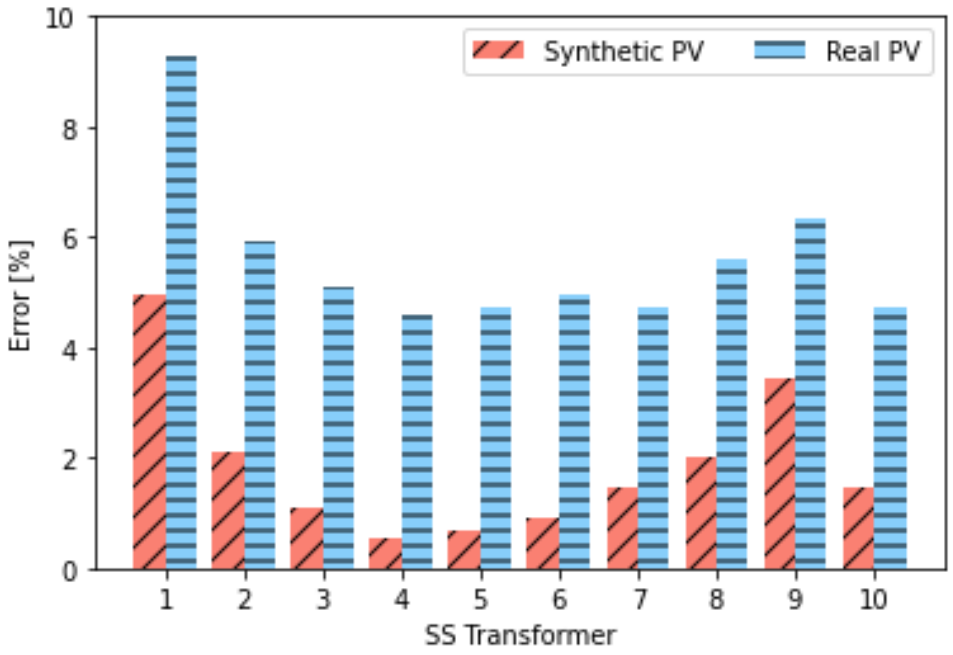


Figure 5.19: Estimated Natural Load Predictions, $\hat{L}^{pred}$, Average Errors under Synthetic PV and Real PV scenario

6

Conclusion

Contents

6.1 System Limitations 58
6.2 Future Work 59

This thesis developed a non-intrusive methodology to identify distributed generation' installed capacity in a distribution grid by analyzing the corresponding transformer's active power measurements and local meteorological data. The proposed methodology was based on classical mathematical optimization and statistical analysis concepts, avoiding complex learning approaches found in the literature. Results have proven to be very promising: in average, the estimated installed DG capacity will have an accuracy of about 99%, when considering a Synthetic PV scenario, and an accuracy of more than 85% when considering a Real PV scenario.

The previous shows that the proposed methodology can be considered accurate. However, it was proven that, under realistic scenarios, the bigger the DG installed capacity the more average percent error is expected. Hence, the referred big magnitudes in DG were reaching the associated transformer rated capacities. When a situation like that happens, the DSO will take preventive actions to not compromise the security of the network. Nevertheless, this will be a more common situation in the future.

DSOs and other stakeholders will consider the proposed methodology valuable in the sense that its implementation will avoid high installation and maintenance costs related to the deployment of sensors across the grid. Given that these actors are only aware of Net Load measurements, being able to observe hidden profiles such as the Natural Load and PV Generation profile will allow them to improve their daily basis operations and expansion plans.

After all the performed analyzes on Net Load, Natural and PV Generation profiles, is cleared the big dependence that Net Load has with the always variable PV Generation profile. As the size of the last tends to increase during time, so does its variability according to sun light, and consequently, the frequency of observing behaviours like forward and reverse power flows in the LV distribution grid.

The methodology was develop is such way that it was possible to test its performance under various PV penetrations magnitudes in the LV distribution grid. The previous was accomplished given that the estimated installed DG capacity, X, is a user-defined variable. Thus, the proposed methodology is ready to work with multiple DG capacities.

Furthermore, a practical use for the methodology was proposed as an illustration. Load forecast results were presented to illustrate the advantage of having good estimates of the PV capacity to accurately disaggregate Net Load into Natural Load and PV generation profiles as a prior step before forecasting them. Although results under Synthetic PV scenario were better than those under Real PV scenario, a trend in decreasing the predictions' error value when using full or partial disaggregation in comparison with not using it was clearly achieved in both cases.

6.1 System Limitations

Even though the proposed methodology and further application are very well structured and extensively explained, the system has some limitations in the actual version of development according to the scope of the thesis and the assumptions taken into account.

In first place, the DG assumed in this thesis only includes PV. However, it was decided to include the names "DG size estimation" and "installed DG capacity" in the methodology as it covers the various RES that can and will appear in a LV distribution grid. In such case, an improved version of the methodology needs to be developed including other features beyond solar irradiation.

Other system limitation lies in the amount of data required to execute the methodology. In this thesis around 6 months of data was used to perform disaggregation calculations and implement a forecast model, however the performance of the methodology will improve with the more historic data that can be available as input. Therefore, the recommended minimum amount of data to test the proposed methodology is the almost 6 months used in this thesis.

As mentioned along the explanation of the processes inside the methodology, the Net Load profiles seen in this thesis are constructed and, therefore, are fictitious. Those are based on true Natural Load profiles measured from actual on-field devices. In this way, the scenarios used to test the methodology's performance on were the result of calculations. The previous was decided in order to control the most as possible variables with the aim to only focus on the development of the disaggregation procedures.

Concerning the load forecast section of the work, there is one limitation to be mentioned. It was observed that the errors of Natural Load predictions were highly conditioning the errors of the Aggregated Net Load predictions. The previous happened indistinctly if disaggregation calculations were not used, partially or fully used. Certainly, the previous was only caused by the forecast model used. Then, the results were subjected to the performance of the ARX model used and any improvement on it lay outside the scope of this thesis.

The last system limitation covers the proposed methodology execution time. This criteria was not taken into account to generate any comparison among other disaggregation approaches. Then, it cannot be concluded that the work here presented is optimized regarding execution time. Similarly, any system's minimum requirements to execute the algorthims is not determined either.

6.2 Future Work

After the successful completion of the presented thesis, some points are left open as a window of opportunity to improve the work's performance in the future.

In first place, it can be considered that the presented work is ready to work with various kind of information, i.e. different DG capacities according to various PV penetrations among a wide range of SS Transformer rated capacities. In this context, future execution can be validated with experimentation on real Net Load profiles. The previous involves feeders where actual PV panels have been installed in the LV distribution grid. However, some requirements must be met. First, the historical Net Load profile must be accompanied by the corresponding historical Solar Irradiance profile. Second, and optionally, the actual DG installed capacity should be known with purpose of validating the results.

In addition, it would be preferable that the proposed methodology is executed on a larger data set in order to produce results with a greater livelihood.

Furthermore, the obtained information in such case, plus a good set of weather predictions will be appropriate to forecast the load components and afterwards aggregate them so, at the end, a Net Load forecast is achieved.

Regarding Load forecasting calculations and given the literature consulted in this work, the application of the proposed methodology could be tested among the work of those authors. This, searching to corroborate the here achieved conclusion that, by implementing disaggregation before predictions are made, the resulting aggregated signal resembles better the actual signal. Moreover, the implementation of other forecasting models such as LSTM and HMM could be investigated to improve the predictions accuracy.

At the end, it is always welcome to extensively test the proposed methodology along with other disaggregation approaches reaching and creating more knowledge around this thesis.

Bibliography

[1] T. Aziz and N. Ketjoy, "Pv penetration limits in low voltage networks and voltage variations," *IEEE Access*, vol. 5, pp. 16 784–16 792, 2017.

[2] P. Denholm, M. O'Connell, G. Brinkman, and J. Jorgenson, "Overgeneration from solar energy in california. a field guide to the duck chart," 11 2015. [Online]. Available: https://www.osti.gov/biblio/1226167

[3] R. Bründlinger, "Grid codes in europe - overview on the current requirements in european codes and national interconnection standards," 11 2019.

[4] IREC, "Model interconnection procedures 2019," 09 2019.

[5] U. Nations, "Iaeg-sdgs — sdg indicators," 2021. [Online]. Available: https://unstats.un.org/sdgs/iaeg-sdgs/disaggregation/

[6] K. Carrie Armel, A. Gupta, G. Shrimali, and A. Albert, "Is disaggregation the holy grail of energy efficiency? the case of electricity," *Energy Policy*, vol. 52, pp. 213–234, 2013, special Section: Transition Pathways to a Low Carbon Economy. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301421512007446

[7] A. Gupta and P. Chakravarty, *Bidgely White paper - Impact of Energy Disaggregation on Consumer Behavior*. Bidgely, 2016. [Online]. Available: https://www.bidgely.com/wp-content/uploads/2016/04/White_Paper_Savings__Engagement_v2_Case_Study.pdf

[8] K. Cetin, M. Siemann, and S. C, "Disaggregation and future prediction of monthly residential building energy use data using localized weather data network," in *ACEEE Summer Study on Energy Efficient Buildings*, Pacific Grove, CA, 2016, pp. 21–26,.

[9] G. S. Ledva and J. L. Mathieu, "Separating feeder demand into components using substation, feeder, and smart meter measurements," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3280–3290, 2020.

[10] E. Vrettos, E. C. Kara, E. M. Stewart, and C. Roberts, “Estimating pv power from aggregate power measurements within the distribution grid,” *Journal of Renewable and Sustainable Energy*, vol. 11, no. 2, p. 023707, 2019. [Online]. Available: https://doi.org/10.1063/1.5094161

[11] G. Hart, “Nonintrusive appliance load monitoring,” *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

[12] M. Sun, F. M. Nakoty, Q. Liu, X. Liu, Y. Yang, and T. Shen, “Non-intrusive load monitoring system framework and load disaggregation algorithms: A survey,” in *2019 International Conference on Advanced Mechatronic Systems (ICAMechS)*, 2019, pp. 284–288.

[13] W. Kong, Z. Y. Dong, J. Ma, D. J. Hill, J. Zhao, and F. Luo, “An extensible approach for non-intrusive load disaggregation with smart meter data,” *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3362–3372, 2018.

[14] M. Zhuang, M. Shahidehpour, and Z. Li, “An overview of non-intrusive load monitoring: Approaches, business applications, and challenges,” in *2018 International Conference on Power System Technology (POWERCON)*, 2018, pp. 4291–4299.

[15] S. S. Hosseini, K. Agbossou, S. Kelouwani, and A. Cardenas, “Non-intrusive load monitoring through home energy management systems: A comprehensive review,” *Renewable and Sustainable Energy Reviews*, vol. 79, pp. 1266–1274, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032117307359

[16] L. Pereira and N. Nunes, “Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools—a review,” *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, vol. 8, no. 6, p. e1265, 2018.

[17] H. K. Iqbal, F. H. Malik, A. Muhammad, M. A. Qureshi, M. N. Abbasi, and A. R. Chishti, “A critical review of state-of-the-art non-intrusive load monitoring datasets,” *Electric Power Systems Research*, p. 106921, 2020.

[18] D. Fay and J. V. Ringwood, “On the influence of weather forecast errors in short-term load forecasting models,” *IEEE Transactions on Power Systems*, vol. 25, no. 3, pp. 1751–1758, 2010.

[19] S. E. Razavi, A. Arefi, G. Ledwich, G. Nourbakhsh, D. B. Smith, and M. Minakshi, “From load to net energy forecasting: Short-term residential forecasting for the blend of load and pv behind the meter,” *IEEE Access*, vol. 8, pp. 224 343–224 353, 2020.

[20] R. L. Burden, J. D. Faires, and A. M. Burden, *Numerical analysis*, 3rd ed. PWS Publishers, 1985.

[21] H. A. Sturges, “The choice of a class interval,” *Journal of the American Statistical Association*, vol. 21, no. 153, pp. 65–66, 1926. [Online]. Available: https://doi.org/10.1080/01621459.1926.10502161

[22] J. Macqueen, “Some methods for classification and analysis of multivariate observations,” in *In 5-th Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281–297.

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Édouard Duchesnay, “Scikit-learn: Machine learning in python,” *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011. [Online]. Available: http://jmlr.org/papers/v12/pedregosa11a.html

[24] S. Kay and S. Marple, “Spectrum analysis—a modern perspective,” *Proceedings of the IEEE*, vol. 69, no. 11, pp. 1380–1419, 1981.

[25] B. Zdaniuk, *Ordinary Least-Squares (OLS) Model*. Dordrecht: Springer Netherlands, 2014, pp. 4515–4517. [Online]. Available: https://doi.org/10.1007/978-94-007-0753-5_2008

[26] S. Ibrahim, I. Daut, Y. Irwan, M. Irwanto, N. Gomesh, and Z. Farhana, “Linear regression model in estimating solar radiation in perlis,” *Energy Procedia*, vol. 18, pp. 1402–1412, 2012.

[27] EDP. Sunlab faro - pv 2017. [Online]. Available: https://opendata.edp.com/explore/dataset/sunlab-faro-pv-2017/information/

[28] ——. Sunlab faro - meteo 2017. [Online]. Available: https://opendata.edp.com/explore/dataset/sunlab-faro-meteo-2017/information/