



Modeling and design of connected objects for smart homes : Evaluation and optimization of their autonomy and QoS

Innocent Mpawenimana

► To cite this version:

Innocent Mpawenimana. Modeling and design of connected objects for smart homes : Evaluation and optimization of their autonomy and QoS. Electronics. Université Côte d'Azur, 2020. English. NNT : 2020COAZ4107 . tel-03223692

HAL Id: tel-03223692

<https://theses.hal.science/tel-03223692>

Submitted on 11 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE DE DOCTORAT

Modélisation et Conception d'objets connectés au service des maisons intelligentes :

Evaluation et Optimisation de leur autonomie et de leur
QoS

Innocent MPAWENIMANA

Laboratoire d'Electronique Antenne et Télécommunication

**Présentée en vue de l'obtention
du grade de docteur en Electronique
d'Université Côte d'Azur**

Dirigée par : Alain PEGATOQUET

Co-encadrée par : Cécile BELLEUDY

Soutenue le : 14 décembre 2020

Devant le jury, composé de :

Olivier Berder, Professeur, Université de
Rennes (Rapporteur)

Hervé Barthélemy, Professeur, Université
de Toulon (Rapporteur)

Benoît Miramond, Professeur, Université
Côte d'Azur (Examineur)

Valérie Roy, MCF HDR, Université Mines
ParisTech (Examineur)

Alain Pegatoquet, MCF HDR, Université
Côte d'Azur (Directeur de thèse)

Cécile Belleudy, MCF HDR, Université Côte
d'Azur (Co-directrice de thèse)

École doctorale STIC (Sciences et Technologies de l'Information et de la
Communication)

PhD Thesis

In Electronics Engineering

From the University Côte d'Azur

Presented on the 14th of December 2020

Author :

Innocent Mpawenimana

Supervised by Alain Pegatoquet
and Cécile Belleudy

**Modeling and design of connected
objects for smart homes : evaluation and
optimization of their autonomy and QoS**

Abstract

This PhD thesis is in the field of smart homes, and more specifically in the energy consumption optimization process for a home having an ambient energy source harvesting and storage system. The objective is to propose services to handle the household energy consumption and to promote self-consumption. To do so, relevant data must be first collected (current, active and reactive power consumption, temperature and so on). In this PhD, data have been first sensed using an intrusive load approach. Despite our efforts to build our own data base, we decided to use an online available dataset for the rest of this study. Different supervised machine learning algorithms have been evaluated from this dataset to identify home appliances with accuracy. Obtained results showed that only active and reactive power can be used for that purpose. To further optimize the accuracy, we proposed to use a moving average function for reducing the random variations in the observations. A non-intrusive load approach has been finally adopted to rather determine the global household active energy consumption. Using an online existing dataset, a machine learning algorithm based on Long Short-Term Memory (LSTM) has then been proposed to predict, over different time scale, the global household consumed energy. Long Short-Term Memory was also used to predict, for different weather profiles, the power that can be harvested from solar cells. Those predictions of consumed and harvested energy have been finally exploited by a Home Energy Management policy optimizing self-consumption. Simulation results show that the size of the solar cells as well as the battery impacts the self-consumption rate and must be therefore meticulously chosen.

Keywords : Smart Home, Renewable Energy, Internet of Thing, Deep Learning, Home Energy Management System, Load energy forecasting,

Résumé

Cette thèse s'inscrit dans le domaine des maisons intelligentes, plus précisément dans l'optimisation énergétique et l'utilisation d'un système de récupération et stockage de l'énergie ambiante. L'objectif est de proposer, après collecte d'un ensemble d'informations pertinentes (courant, puissance active et réactive, température, etc.), des services liés à la gestion de la consommation électrique domestique et favorisant l'autoconsommation. Dans cette thèse, la collecte des données a tout d'abord été basée sur une approche intrusive. A défaut de pouvoir construire notre propre base de données, nous avons utilisé une base de données disponible en ligne. Différents algorithmes d'apprentissage supervisés ont été évalués à partir de ces données afin de reconnaître un appareil électrique. Nos résultats ont montré que les puissances active et réactive seules suffisent à identifier de manière précise un appareil électrique. Afin d'améliorer l'identification des différents appareils, une technique basée sur une moyenne glissante a été utilisée pour le pré-traitement des données. Dans cette thèse, une approche non-intrusive consistant à mesurer la consommation électrique d'une habitation de manière globale, a finalement été privilégiée. A partir de cette mesure globale, des prédictions de l'énergie globale consommée à partir d'algorithmes d'apprentissage automatique (LSTM) a été proposée. L'algorithme LSTM (Long Short-Term Memory) a également été utilisé afin de prédire la puissance récupérée par des cellules photovoltaïques, ceci pour différents profils d'ensoleillement. Ces prédictions de l'énergie consommée et récupérée sont finalement exploitées par un algorithme de gestion de l'énergie favorisant l'autoconsommation.

Mots clés : Maison intelligentes, Energie Renouvelable, Objets connectés, Deep Learning, Gestion énergétique, prédiction de la charge électrique

Personal publications

International Conferences with peer review

1. Innocent Mpawenimana, Alain Pegatoquet, Valérie Roy, Laurent Rodriguez and Cécile Belleudy, A comparative study of LSTM and ARIMA for energy load prediction with enhanced data preprocessing, 15th IEEE Sensors Applications Symposium (SAS), Mar 2020, Kuala Lumpur, Malaysia.
2. Innocent Mpawenimana, Win Thandar Soe, Alain Pegatoquet, and Cécile Belleudy, Appliances identification for different electrical signatures using moving average as data preparation, in 9th International Green and sustainable computing conference, Pittsburgh, United States, 2018.
3. Win Thandar Soe, Innocent Mpawenimana, Mathieu DiFazio, Cécile Belleudy, and Aung Ze Ya, Energy Management System and Interactive Functions of Smart Plug for Smart Home, International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering, no. 11, 2017.
4. Win Thandar Soe, Innocent Mpawenimana, Mathieu DiFazio, Cécile Belleudy, and Aung Ze Ya, Energy Management System and Interactive Functions of Smart Plug for Smart Home, in International Conference on Smart Home Technology (ICSHT), Porto, Portugal, 2017.

National Conference with reading committee

1. Innocent Mpawenimana, Cécile Belleudy, Win Thandar Soe, Mathieu Difazio, Alain Pegatoquet. A wireless infrastructure to collect data for energy management in a Smart

Home, Journées scientifiques de l'URSI-France 2017, Radiosciences au service de l'humanité, Feb 2017, Sophia Antipolis, France.

2. Win Thandar Soe, Cécile Belleudy, Innocent Mpawenimana, Mathieu Difazio, Aung Ze Ya, Modeling and Architecture Exploration of Wireless Sensor Network to manage energy in Smart Home. Compas'2017 - Conférence d'informatique en Parallélisme, Architecture et Système, 2017, Sophia-Antipolis- Nice, France.

Remerciements

Les travaux de recherche présentés dans ce manuscrit ont été réalisés au sein du Laboratoire d'Electronique, Antennes et Télécommunications (LEAT), unité mixte de l'université Cote d'Azur (UCA) et du CNRS. Ils n'auraient jamais pu être accompli sans le soutien de certaines personnes pour lesquelles je suis infiniment reconnaissante.

Tout d'abord, je souhaite remercier mon directeur de these, Alain Pegatoquet, pour ses conseils , ses directives ainsi que pour ses critiques constructives qui m'ont permis de faire beaucoup de progrès sur ma qualité rédactionnelles et pour avoir toujours su se rendre disponible quand nécessaire. Je tiens également à remercier ma co-directrice de thèse, Mme Cécile Belleudy, Maître de conférences HDR à l'Université Côte d'Azur pour ces enseignement avant et et durant la thèse.

J'exprime aussi mes sincères remerciements à M. Hervé Barthélemy, Professeur à l'Université de Toulon, et à M. Olivier Berder, Professeur à l'Université de Rennes, d'avoir accepté de rapporter mon travail de thèse. C'est un grand honneur pour moi. Je tiens à remercier aussi les personnes qui m'ont honoré par leurs participations à mon jury de thèse, M. Benoît Miramond, Professeur à l'Université de l'Université Côte d'Azur, et Mme Valérie Roy, Maître de conférences HDR à l'Université Mines ParisTech.

Je souhaite remercier tous les membres du LEAT, BIATSS, Professeurs, Maîtres de Conférences, ATER, stagiaires, et surtout mes collègues doctorants. Merci à toutes les personnes que j'ai connues durant mon travail au LEAT. Merci pour toutes les discussions, les blagues, les événements qu'on a faits ensemble. Cela me permet de garder de beaux souvenirs sur cette période de ma vie.

Enfin, je tiens à remercier mes parents, mes sœurs, mes frères, toute ma famille et surtout ma chère épouse et ma fille qui m'ont dû supporter et soutenir tout au long de ce travail. Je n'aurai jamais pu avancer ou réussir ces étapes de ma vie sans votre soutien et votre aide. Le plus grand

merci est à vous, mes parents pour m'avoir encouragé, soutenu, pour avoir pensé à moi et pour avoir patienté toutes ces années avec moi. Tout ce que j'ai pu accomplir jusqu'à maintenant est grâce à vous, votre éducation, votre exemple, votre confiance et surtout votre amour. Je ne suis rien sans vous deux.

Du fond du cœur, merci à vous tous.

Table of content

Abstract	iv
Résumé	v
List of publications	vi
Remerciements	viii
Table of contents	xii
List of tables	xiii
List of Figures	xv
Abbreviations	xviii
1 Introduction	1
1.1 Context et Objectives	1
1.2 Contributions and thesis structure	4
2 State of the art	7
2.1 Energy Management in Smart Home	7
2.1.1 What is energy management system in smart home	7
2.1.2 Energy Monitoring approach in smart home	8
2.1.2.1 Energy Monitoring system based on intrusive approaches (ILM)	9
2.1.2.2 Energy Monitoring system based on non intrusive approaches (NILM)	10
2.1.3 Appliance control (ON / OFF)	10
2.2 Machine learning techniques	11
2.2.1 General presentation	11
2.2.2 K-NN	11

2.2.3	Random Forests	12
2.2.4	Artificial Neural Network	14
2.2.4.1	Introduction	14
2.2.4.2	Architectural description of a formal neuron	15
2.2.4.3	MultiLayer Perceptron	17
2.2.4.4	Recurent neural network	21
2.3	Energy Harvesting Technology	22
2.3.1	Introduction	22
2.3.1.1	Wind power generation	22
2.3.1.2	Power generation by a generator set	22
2.3.1.3	Production of electrical energy from solar energy	22
2.3.2	Grid-Connected Photovoltaic System	23
2.3.2.1	Introduction	23
2.3.2.2	The photovoltaic effect	24
2.3.2.3	Different types of solar cells	25
2.3.2.4	Types of Photovoltaic system	28
2.3.3	Battery Bank for PV system	32
2.4	Conclusion	34
3	Electrical Signatures and Analysis of smart home appliances	35
3.1	Introduction	35
3.2	Related work	38
3.3	Energy monitoring	40
3.3.1	Definition	40
3.3.2	A wireless infrastructure to collect data for energy management in a smart home	40
3.3.2.1	Types of sensors and protocols in SMART Home	40
3.3.3	Choice of a Home Automation System	43
3.4	Impact of electrical signatures on Appliances identifications	46
3.4.1	Data description	46
3.4.2	Data preprocessing with moving average	46
3.4.2.1	Feature scaling	48
3.4.3	Classification methods	48
3.4.3.1	K-NN	48
3.4.3.2	Random-Forest	49
3.4.3.3	Multilayer Perceptron	50
3.4.4	Appliances Recognition Experiments	51
3.5	Conclusion	54

4	Short and Medium load forecast in smart Home	56
4.1	Introduction	56
4.2	Related works	57
4.3	Energy load prediction methodology	58
4.3.1	Data set and Data preprocessing	59
4.4	Choice of energy load forecasting model	65
4.4.1	Linear model (ARIMA)	65
4.4.2	Non-Linear Model (LSTM)	66
4.4.3	Performance metrics	68
4.5	Results and discussion	68
4.5.1	LSTM and ARIMA prediction models comparison	68
4.5.2	LSTM prediction model validation on a new dataset	70
4.6	Conclusion	71
5	Coupling the grid and an energy harvesting system for smart home self-consumption	73
5.1	Introduction	73
5.2	Related works	74
5.3	Grid-connected PV system	76
5.3.1	Electric generation profile	76
5.3.2	Electric load profile	79
5.3.3	Power management (PM) policies for Grid connected PV System	80
5.3.3.1	Global overview of the PM	80
5.3.3.2	A slotted-based PM algorithm	81
5.3.3.3	Battery state of charge evolution	82
5.3.3.4	Variables used by the PM policy	84
5.3.3.5	Power management algorithm without battery	86
5.3.3.6	Power management algorithm with battery	87
5.4	Results and Discussion	92
5.4.1	Simulation setup	92
5.4.2	Household located in south of France (Nice)	92
5.4.3	Household located in North of France (Paris)	105
5.4.4	Impact of PV and battery size on blackout Error rate	114
5.5	Conclusion	116
6	Conclusion and Perspectives	117
6.1	Conclusion	117
6.2	Perspectives	118

A	121
Bibliographie	125

List of tables

3.1	Some smart plugs available on the market	37
3.2	Comparison among the different protocols	42
3.3	Existing dataset of load consumption energy [Monacchi et al., 2014]	45
3.4	APPLIANCE PRESENT IN DATASET	46
3.5	FEATURE USED ON RECOGNITION PATTERN	47
3.6	Classification algorithms	50
3.7	Obtained accuracy for different classification algorithms with features set F4	51
3.8	Different data preparation techniques used for appliance identification using ACS-F1 dataset and the K-NN classifier	54
5.1	Input parameters	80
5.2	Configurations used in this study	91
5.3	Results obtained for config.1	93
5.4	Results obtained for config.2	93
5.5	Results obtained for config.3	97
5.6	Results obtained for config.13	98
5.7	Results obtained for config.14	98
5.8	Results obtained for config.15	99
5.9	Results obtained for config.4	99
5.10	Results obtained for config.5	100
5.11	Results obtained for config.6	100
5.12	Results obtained for config.16	104
5.13	Results obtained for config.17	104
5.14	Results obtained for config.18	105
5.15	Results obtained for config.7	106
5.16	Results obtained for config.8	106
5.17	Results obtained for config.9	107

5.18 Results obtained for config.19	107
5.19 Results obtained for config.20	108
5.20 Results obtained for config.21	108
5.21 Results obtained for config.10	112
5.22 Results obtained for config.11	112
5.23 Results obtained for config.12	113
5.24 Results obtained for config.22	113
5.25 Results obtained for config.23	114
5.26 Results obtained for config.24	114

List of Figures

1.1	Primary energy world consumption (a), Share of globally primary energy consumption by fuel (b) [industry, 2019]	2
1.2	Average installed cost for solar photovoltaics worldwide from 2010 to 2019 (in U.S. dollars per kilowatt) [Statista, 2020]	2
1.3	Global solar manufacturing capacity vs. module price, 2010-2020E [Sun, 2019]	3
1.4	World Internet of Things market, 2013 - 2025 [Ropert, 2016]	4
1.5	Home Energy Management system	5
2.1	Approach used on smart home for monitoring appliances [Nguyn, 2015]	8
2.2	Example of a 3NN classification	12
2.3	Visualization of a random forest model making a prediction [Yiu, 2019]	13
2.4	Model of a formal neuron	15
2.5	Linear function	16
2.6	Sigmoid function	16
2.7	ReLU function	17
2.8	Hyperbolic tangent function	17
2.9	A multilayer perceptron with two hidden layers	18
2.10	Photovoltaic effect [B. Afework, 2018]	24
2.11	Monocrystalline silicon Photovoltaic panels	26
2.12	Polycrystalline silicon Photovoltaic panels	27
2.13	Amorphius silicon solar panel	28
2.14	Thin film silicon solar panel	29
2.15	Grid-connected solar PV system	30
2.16	Stand-alone PV system without battery storage	30
2.17	Stand-alone PV system with battery storage	31
2.18	Photovoltaic hybrid system	32
3.1	Architecture to collect data with WEMO insights	38

3.2	Wireless infrastructure diagram	43
3.3	Low-cost wireless infrastructure based on Domoticz, Grafana, InfluxDB and a Raspberry pi3	43
3.4	Accuracy Vs. K parameter for K-NN on F4 feature set	49
3.5	Accuracy per Class for K-NN on F1 features set	50
3.6	Accuracy per Class for K-NN on F2 features set	51
3.7	Accuracy per Class for K-NN on F3 features set	52
3.8	Accuracy per Class for K-NN on F4 features set	53
3.9	Average Accuracy for different features set	53
4.1	Machine Learning process	59
4.2	Curves of each features in the IHEPCD	60
4.3	Curves of global active power for each years in the IHEPCD	60
4.4	Curves of global active power for week of january the IHEPCD	61
4.5	Histograms of global active power of the IHEPCD	61
4.6	Histograms of global active power of the IHEPCD for each years	62
4.7	Algorithm proposal STDAN	62
4.8	Representation of data after filling missing values based on different data preprocessing	64
4.9	Representations of all the daily load profiles in IHEPCD	64
4.10	The internal structure of an LSTM	67
4.11	Comparison of LSTM and ARIMA prediction models for different preprocessing techniques	69
4.12	Evaluation of STDAN and Mean techniques accuracy for different missing values percentage	70
4.14	Validation of new data set	70
4.13	LSTM for short and medium term load forecasting	71
4.15	Electric Power consumer for an office located at Nice and IHEPCD	72
5.1	House model for Grid-connected PV System	74
5.2	Energy production P_H in Nice and Paris (for 8kwp PV system) VS. annual profile of energy load (P_C)	77
5.3	Zoom of energy production P_H in Nice and Paris (8kwp) vs. energy load profile P_C in winter condition (i.e. in December)	78
5.4	Zoom of energy production P_H in Nice and Paris (8kwp) vs. energy load profile P_C in summer condition (i.e. in July)	79
5.5	Global overview of the PM	81
5.6	Description of power management algorithm with or without batteries	82
5.7	Battery Management Strategy	83

5.8	Energy flows of a residential in Nice : Grid-PV Connected without (a) and with battery system for one day of December (b) (PV system size 2 kWp, battery size 20 kWh, annual load demand 9 MWh)	94
5.9	Energy flows of a residential in Nice : Grid-PV Connected without (a) and with battery system for one day of December (b) (PV system size 4 kWp, battery size 20 kWh, annual load demand 9 MWh)	95
5.10	Energy flows of a residential in Nice : Grid-PV Connected without (a) and with battery system for one day of December (b) (PV system size 8 kWp, battery size 20 kWh, annual load demand 9 MWh)	96
5.11	Energy flows of a residential : Grid-PV Connected without (a) and with battery system for one day of July (b) (PV system size 2 kWp, battery size 20 kWh, annual load demand 9 MWh)	101
5.12	Energy flows of a residential : Grid-PV Connected without (a) and with battery system for one day of July (b) (PV system size 4 kWp, battery size 20 kWh, annual load demand 9 MWh)	102
5.13	Energy flows of a residential : Grid-PV Connected without (a) and with battery system for one day of July (b) (PV system size 8 kWp, battery size 20 kWh, annual load demand 9 MWh)	103
5.14	Energy flows of a residential in Paris : Grid-PV Connected without (a) and with battery system for one day of December (b) (PV system size 2 kWp, battery size 20 kWh, annual load demand 9 MWh)	109
5.15	Energy flows of a residential in Paris : Grid-PV Connected without (a) and with battery system for one day of December (b) (PV system size 4 kWp, battery size 20 kWh, annual load demand 9 MWh)	110
5.16	Energy flows of a residential in Paris : Grid-PV Connected without (a) and with battery system for one day of December (b) (PV system size 8 kWp, battery size 20 kWh, annual load demand 9 MWh)	111
5.17	Impact of PV and battery size on blackout error rate during summer conditions (July) in Nice	115
5.18	Impact of PV and battery size on blackout error rate during summer conditions (July) in Paris	115
A.1	Computation of the dynamic coefficients for a synthetic signal. A given n-th sample value is extended through two new values computed on the delta and delta-delta windows	121
A.2	Example of features active power of a Fridge signature in the ACS-F1 database	122
A.3	Moving average algorithm	123

Abbreviations

<i>6LoWPAN</i>	IPv6 over Low-Power Wireless Personal Area Networks
Φ	phase of voltage relative to current
<i>AC</i>	Alternating current
<i>ACSF1</i>	Appliance Consumption Signature Fribourg1
<i>ANNs</i>	Artificial Neural Networks
<i>AR</i>	AutoRegressive model
<i>ARIMA</i>	AutoRegressive Integrated Moving Average
<i>ARMA</i>	Auto-Regression Moving Average
<i>BER</i>	Blackout Error Rate
<i>BER</i>	Blackout Error Rate
<i>BVAR</i>	Bayesian Vector AutoRegression
<i>DBN</i>	Deep Belief Networks
<i>DC</i>	Direct Current
<i>DoD</i>	Depth of Discharge
<i>E</i>	Active energy
<i>EDR</i>	Enhanced Data Rate
<i>F</i>	Frequency
<i>HEC</i>	Household Energy Consumption
<i>HIPHIP</i>	House Integrated Photovoltaic Hightech in Public
<i>HS</i>	High Speed

ABBREVIATIONS

<i>I</i>	Root mean square current
<i>IHEPCD</i>	Individual Household Electrical Power Consumption Dataset
<i>ILM</i>	Intrusive Load Monitoring
<i>IP</i>	Internet Protocol
<i>IPv6</i>	Internet Protocol version 6
<i>IWT</i>	Integrated Wireless Technologies
<i>K – NN</i>	k-nearest neighbors algorithm
<i>LSTM</i>	Long Short-Term Memory
<i>MAPE</i>	Mean Absolute Percentage Errors
<i>MLP</i>	multilayer perceptron
<i>NILM</i>	Non Intrusive Load Monitoring
<i>NREL</i>	National Renewable Energy Laboratory
<i>P</i>	Active power
<i>Q</i>	Reactive power
<i>RF</i>	Random Forests
<i>RMSE</i>	Root Mean Square Error
<i>S</i>	Apparent power
<i>SAM</i>	System Advisor Model
<i>SCR</i>	Self Consumption Rate
<i>SoC</i>	state of charge
<i>SSR</i>	Self sufficiency Rate
<i>STDAN</i>	Same Time a Day Ago or Next
<i>SVM</i>	Support Vector Machines
<i>TMY</i>	Typical meteorological year
<i>V</i>	Root mean square voltage
<i>WIFI</i>	Wireless Fidelity

Introduction

Sommaire

1.1 Context et Objectives	1
1.2 Contributions and thesis structure	4

1.1 Context et Objectives

Most of the energy produced in the world comes from fossil fuels (oil, gas and coal) and accounts for more than 80% of the world’s total primary energy production as shown in figure 1.1. However, the world’s energy reserves are not inexhaustible (at the current rate of consumption, oil will run out in 54 years, gas in 63 years and coal in 112 years) and contribute largely to global warming due to the greenhouse gases emitted by these fossils. Indeed, according to data published by the International Energy Agency, world energy demand could increase by 45% by the year 2030, in particular due to demographic development and the industrialization of countries such as China and India, which alone have more than 2 billion inhabitants.

Electricity consumption is expected to grow twice as fast as average energy consumption. In order to reduce this dependence on fossil fuel-based energy, which is largely responsible for greenhouse gas emissions, several initiatives have been put in place in this context around the world. These initiatives aim at stimulating energy efficiency, limiting the construction and use of the least efficient coal-fired power plants, reducing methane emissions in the upstream oil and gas sectors, and reforming fossil fuel subsidies. These measures will support the development of renewable energy.

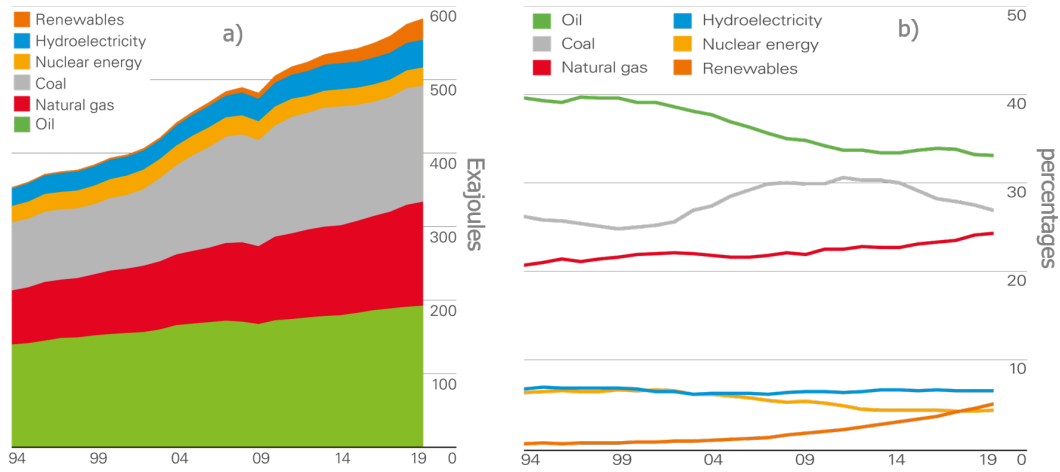


FIGURE 1.1 – Primary energy world consumption (a), Share of global primary energy consumption by fuel (b) [industry, 2019]

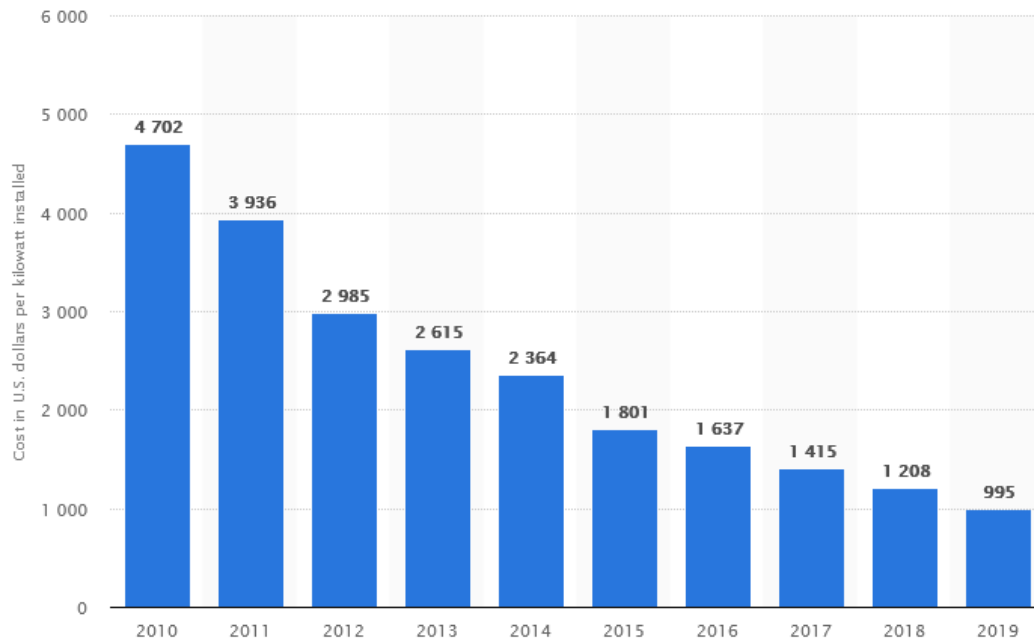


FIGURE 1.2 – Average installed cost for solar photovoltaics worldwide from 2010 to 2019 (in U.S. dollars per kilowatt) [Statista, 2020]

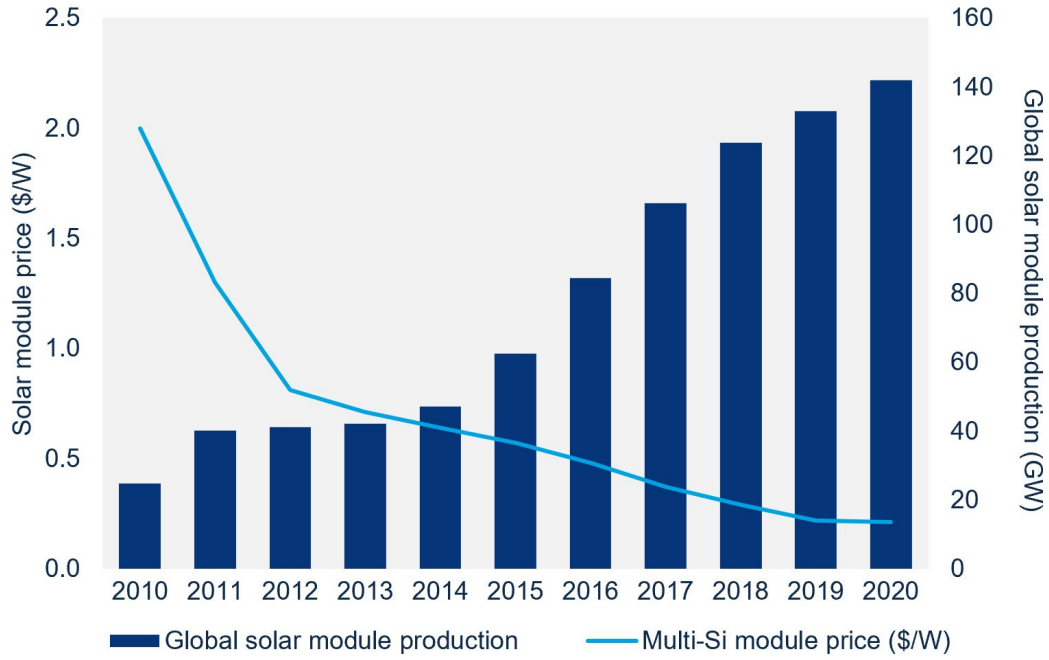


FIGURE 1.3 – Global solar manufacturing capacity vs. module price, 2010-2020E [Sun, 2019]

Today, these renewable energies are gradually becoming energies in their own right, competing with fossil fuels in terms of cost and production performance. They are clean and inexhaustible, and therefore capable of covering our energy needs in the long term. Among these renewable energies, solar energy offers high potential. Moreover, it is currently undergoing a strong development in the world. This development can be explained, by government measures aimed at reducing the use of fossil fuels and, on the other hand, by the impressive drop in the price of the solar module and the cost of installing solar panels, which have fallen by more than 90% [Sun, 2019] and 70% [Statista, 2020] respectively, as shown in figures 1.2 and 1.3. Despite rising energy consumption, global carbon emissions related to energy consumption have increased by 0.5% in 2019 compared to an average of 1.1% per year over the last decade. This slight increase is partly due to improved energy efficiency as well as better penetration of new technologies. These new technologies such as connected objects can bring significant potentials in the energy sector. Connected objects are already part of our everyday life and will grow strongly in the coming years. As can be seen in figure 1.4, a market study published by IDATE in 2015 predicts an increase from 42 to 155 billion connected objects worldwide by 2025 [Ropert, 2016]. These objects can meet multiple needs and provide a variety of services. As a result, they address a wide variety of application areas, such as e-health, agriculture, the environment, the automotive industry, smart cities and smart homes.

In this thesis, we are interested in the field of Smart Homes. As shown in figure 1.5, a smart home can be defined as a residence equipped with connected objects integrating communication and information systems to measure, monitor and automatically control energy consumption through an Internet connection using typically a mobile device. Using this new type of technology, smart home can also improve the safety and comfort of the occupants. smart homes are increasingly

equipped with renewable energy systems such as solar panels (as in the case of this thesis) connected to grid to maximize self consumption and energy saving. For energy efficient buildings, the challenge is to set up these objects not only to monitor energy consumption, but also to manage energy (load shedding, adjusting heating curves, etc.) either directly or through the involvement of the occupants. The aim of this thesis is to propose solutions to facilitate the integration, deployment and interoperability of connected objects linked by a wireless network within a house and also to design a Power Management which favors self consumption. This Power Management will take decisions according to predictions of the consumed and harvested energy. Obviously, a key principle of smart home is to collect relevant information and then to propose services, such as those related to energy control and management.

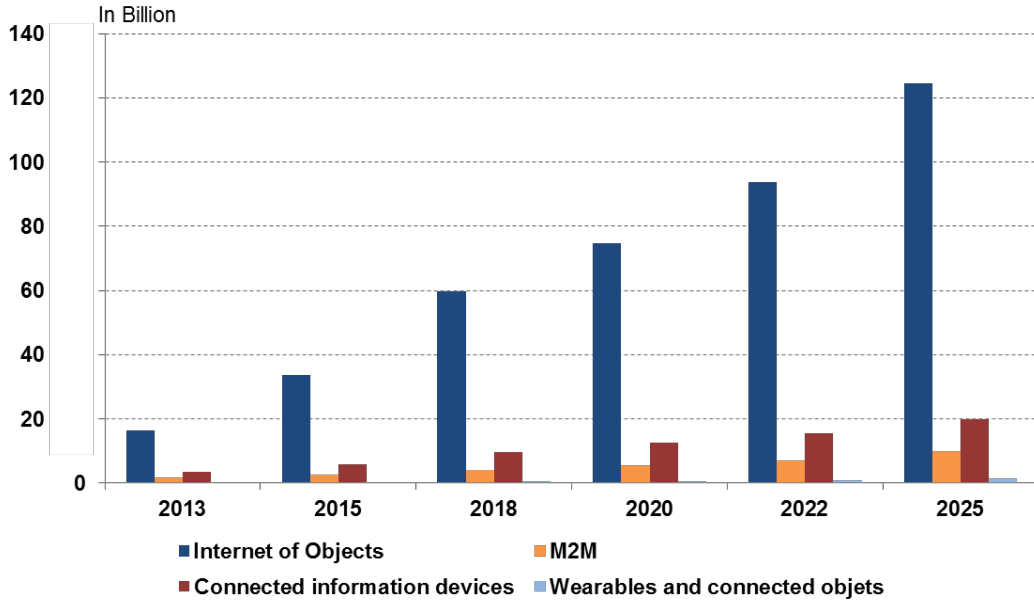


FIGURE 1.4 – World Internet of Things market, 2013 - 2025 [Ropert, 2016]

Indeed, according to [Florian, 2019], the residential sector is one of the main consumers of energy and accounted for 35 percent of total French energy consumption in 2018. Moreover, recent studies have shown that a continuous feedback (e.g. through a real-time visualization) coupled with an automated management system for domestic equipment could allow an energy reduction of 15 to 30% [B. Neenan, 2009], [Darby, 2006]. A data collection infrastructure related to energy consumption is therefore a necessary basic element of such a management system.

1.2 Contributions and thesis structure

Contributions

In this thesis, 4 main contributions are proposed in the context of smart home :

1.2 Contributions and thesis structure

- Study on the impact of the number and type of features for home appliance identifications (chapter 3)
- A data preparation technique to optimize classifiers (chapter 3)
- Prediction of overall power consumption using Deep Learning methods (LSTM) (chapter 4)
- Power Management policies for smart home equipped with a photovoltaic-based energy harvesting system (PV system), with and without a battery (chapter 5)

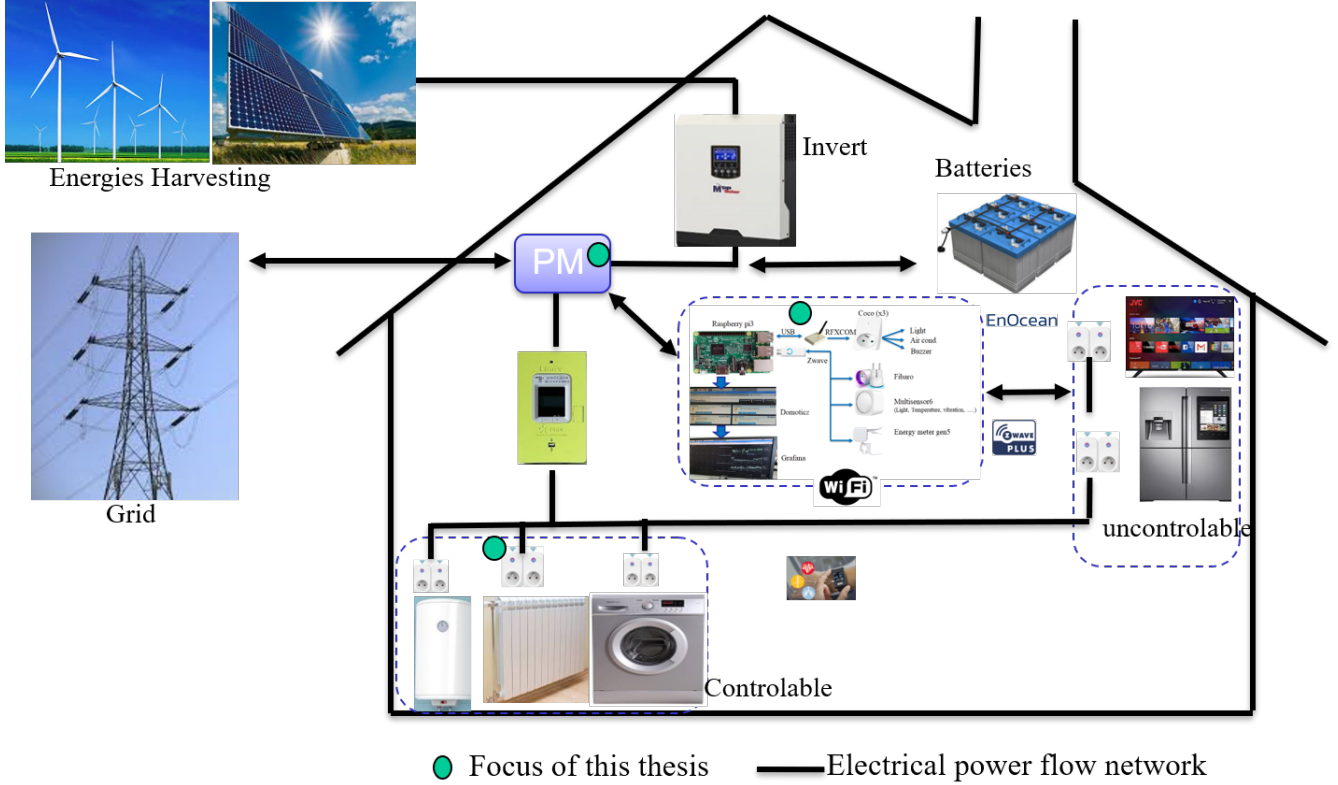


FIGURE 1.5 – Home Energy Management system

Thesis structure

This thesis is structured in six main chapters. Following this introductory chapter, Chapter 2 sets the context of the research topic by presenting the basics of the work carried out in the energy management of a house. It also examines the different data collection architectures for energy consumption. Challenges in achieving energy savings through the monitoring and control of energy consumption will be discussed. The second part of this chapter describes various Machine Learning and Deep Learning techniques commonly used in smart homes. Finally, this chapter proposes a section related to renewable energy, with a focus on solar panels modeling.

Chapter 3 presents our methodology for recognizing electrical appliances based on their consumption profiles called electrical signatures. In the framework of this thesis, a wireless

data collection environment based on an intrusive approach is presented. Despite our efforts to build our own database, a free online database ACSF1 was rather used in this thesis for our experiments. Different Machine Learning techniques such as Multilayer Neural Perceptron (MLP) and Random Forest were implemented and compared with K-NN, a discriminative and non-parametric approach. Variations in the recognition rate were analyzed by adjusting the system parameters. In addition, the impact on the numbers and the type of electrical signature were analyzed and evaluated.

Chapter 4 presents the short- and medium-term global energy consumption prediction model based on the recurrent LSTM algorithm. A comparative study based on an ARIMA linear model and a LSTM non-linear approach was carried out. This study is based on an online available dataset collected in a household near Paris during 4 years (IHEPCD dataset). A new approach for managing missing values in the dataset was proposed and compared with other existing techniques. The results obtained using these two approaches as well as the technique for managing missing data are analysed. Finally, the prediction model trained on the IHEPCD database was evaluated and validated on another dataset containing measures collected in a house located near to Nice in the south of France.

Chapter 5 presents the complete system integrating an energy harvesting system (solar panels), a storage device and a data collection environment based on a mixed approach. In order to favour household self-consumption, two power management algorithms taking as input the actual and predicted value of the global energy consumption as well as the energy produced by solar panels are proposed. The first Power Management policy considers that the household is not equipped with a battery, while the second one does. To evaluate the efficiency of our approach with different harvested energy profiles, two different locations are considered in our study (Nice and Paris). The number of solar panels as well as the battery sizing was also evaluated in this thesis.

This thesis ends with a general conclusion of our research work, detailing in particular the various achieved objectives and the prospects they open up.

State of the art

Sommaire

2.1 Energy Management in Smart Home	7
2.1.1 What is energy management system in smart home	7
2.1.2 Energy Monitoring approach in smart home	8
2.1.3 Appliance control (ON / OFF)	10
2.2 Machine learning techniques	11
2.2.1 General presentation	11
2.2.2 K-NN	11
2.2.3 Random Forests	12
2.2.4 Artificial Neural Network	14
2.3 Energy Harvesting Technology	22
2.3.1 Introduction	22
2.3.2 Grid-Connected Photovoltaic System	23
2.3.3 Battery Bank for PV system	32
2.4 Conclusion	34

2.1 Energy Management in Smart Home

2.1.1 What is energy management system in smart home

In general, the energy management system can be defined as a system that allows efficiently managing energy both for the consumers (smart home in our case) and for the suppliers (grid) by shifting the demand to off-peak energy prices. This can help to conserve fossil fuel resources and reduce energy consumption. In the case of a smart home, the energy management system allows to communicate, monitor and remotely activate home appliances in order to not only improve

the quality of life and comfort, but also to optimize the self-consumption, thus reducing the energy bill. According to [Leitão et al., 2020] Smart homes are a ubiquitous branch of computing that includes the integration of connected objects in homes for safety, comfort, healthcare and energy efficiency [Alam et al., 2012] [Zafari et al., 2016]. These connected objects are part of the ongoing technologies in power grids, driving the deployment of smart sensors and other advanced measurement devices that enable communication, monitoring and remote operation of household appliances. Smart homes offer a better quality of life and greater efficiency by taking advantage of remote monitoring and contextual self-adaptation mechanisms to identify occupants' needs and preferences and coordinate the operation of appliances. Wired and wireless networks of sensors and actuators are deployed in smart homes, with sensor data collected and stored on a local or remote central platform depending on the application context. This platform is also in charge of processing the acquired information, enabling the optimized management and operation of household appliances, in the interest of the comfort and energy efficiency of the occupants.

2.1.2 Energy Monitoring approach in smart home

In the literature, we often find two types of approach to monitor the power consumption of household appliances : the Non Intrusive (NILM) or Intrusive (ILM) approach as shown in figure 2.1.

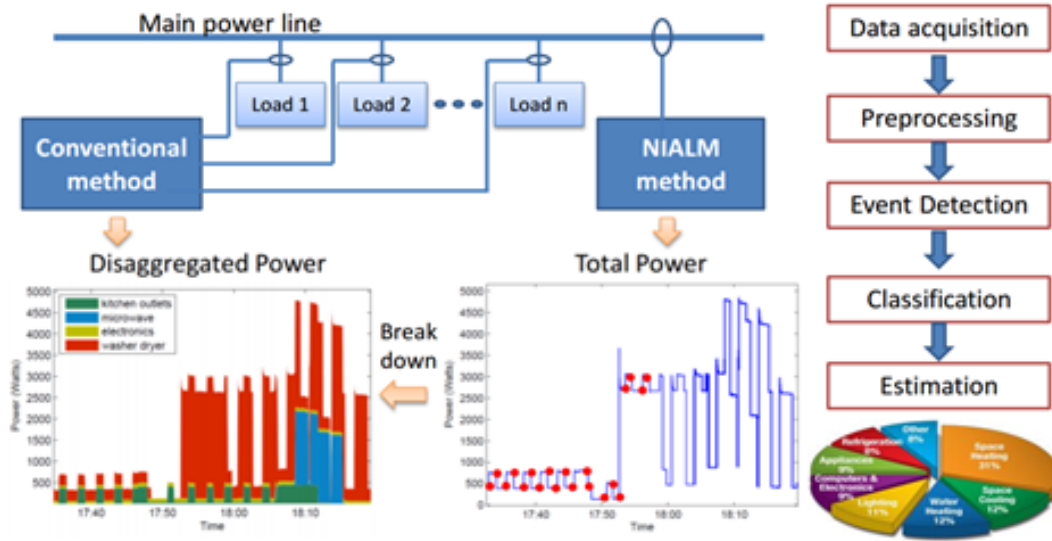


FIGURE 2.1 – Approach used on smart home for monitoring appliances [Nguyn, 2015]

The Non Intrusive approach [Hart, 1992] consists in analyzing the transient and stable signature appearing when appliances are switched on or off. This approach requires a single sensor with a high sampling rate placed at the point of entry to a house. It is used to identify each device connected to the network based on the overall consumption. According to [Nguyn, 2015] and as illustrated on the figure 2.1, the process of identifying a device is done in 5 steps :

- **Data acquisition** : The NIALM system collects values from current and voltage sensors at a defined sampling rate.
- **Preprocessing** : This is an important step to filter out electrical noise and to extract electrical characteristics such as total active power, reactive power and apparent power. It may contain some advanced tasks such as calculating the phase of the electrical signal, harmonics data and power factor. The more electrical characteristics are extracted, the more accurate the step-by-step classification.
- **Event Detection** works on detecting aggregate current or power changes to define an event to turn a device on or off in the power grid. This step also extracts transition signatures after a detected event.
- **Classification** groups the events after the event detection step and matches the on/off events to classify the devices.
- **Estimation** is the final process to summarize the total power consumption and power consumption ratios for each device.

Note that measurements are made on multiple independent devices to determine the contribution of each device. As a result, the cost of a high-frequency analyzer, which is used to detect transient periods, was found to be a disadvantage. Although the identification results are promising, some researchers have focused on using an analyzer with a sampling frequency of up to 15 kHz to reduce the cost. In spite of its ease of installation and setup of its data acquisition system, the Non Intrusive approach has the disadvantage of having inaccuracy defects (related to the accumulation of power consumption of the various devices that generate interference on the network). The Intrusive ILM approach [Antonio.Ridi et al., 2014], on the other hand, consists in analyzing in the short and medium term (from a few seconds to a few minutes) the evolution of the power consumption of one or several devices that are either used or inactive. It is an intrusive process since the smart meter (connected plug) is placed on a device that is being measured. This approach also has several advantages over a Non Intrusive approach. The information collected is more accurate and the electrical signatures are numerous thanks to the multiple sensors used in this approach, making the identification process easier and more flexible. Moreover, an intrusive approach allows appliances to be remotely control by the user.

2.1.2.1 Energy Monitoring system based on intrusive approaches (ILM)

Energy monitoring systems can influence residents behaviour by informing them through a graphical interface of the real-time home energy usage. If the breakdown energy usage of each home appliances and consumer electronics is displayed on a wall pad, a computer, or a television, residents are aware of their power consumption and can therefore make an effort to reduce their energy. Over the last years, web-based monitoring and control systems were proposed by energy

suppliers (such as ENEDIS in France) to enable users to see home energy data and control home devices remotely through the Internet.

2.1.2.2 Energy Monitoring system based on non intrusive approaches (NILM)

The training phase of an automatic recognition system often requires several series of measures to take into account the behaviour of electrical power consumption of the charges. These measurement phases reveal several degrees of intrusion. First, they require an intrusion of a specialist into the residence to place measuring devices on each appliance to collect the different signatures. Following the first step of state change detection, the second step consists in identifying the loads using the signatures of the devices stored in a database. These signatures can be obtained using different techniques such as genetic algorithms [Nguyen, 2015], integer programming methods [Suzuki et al., 2008], or probabilistic Hidden Markov Models [Nambi et al., 2013]. The NIALM methods can be classified into two categories :

- those using an automated learning phase (AS-NIALM : Automatic Setup)
- those using a manual learning phase (MS-NIALM : Manual Setup)

Manual learning : Manual NIALM methods are more accurate than automatic NIALM methods thanks to the collection of signatures of the devices present in the installation. However, this semi-intrusion can be seen as annoying for the customer subscriber and unattractive for the distributor or supplier of electrical energy. In the case of MS-NIALM methods, a library of device signatures is built, from intrusive measurements on the installation. The time tracking of the devices requires a certain duration of the intrusion. During this period, the signatures of the devices are observed (active power, reactive power, RMS current value, etc.) and manually qualified as a heating, lighting, washing etc. signature. - during commissioning and decommissioning.

Automatic learning : In the case of AS-NIALM methods, the signatures of the devices improving over time and are based on information collected a priori in laboratories as well as questionnaires collected in habitats. These libraries actually identify each electrical device and its consumption, and have encouraged a move towards identifying the uses of these devices, or even the habits of use or consumption of the latter by customers. MS-NIALM methods are a tool in the development of AS-NIALM methods. They have probably been used to analyze situations where AS-NALM methods failed. Given their much less intrusive nature, AS-NIALM methods should then dominate in most applications by building their own signature library observed during the state change steps.

2.1.3 Appliance control (ON / OFF)

The loads in a smart home can be classified into two categories : non-controllable and controllable loads. The controllable loads are defined as the loads that can be controlled (shifted in time for example) without noticing impacts on consumers' life styles. The controllable appliances

are more and more proposed by suppliers and already include air-conditioner (AC), water heater (WH), clothes dryer (CD), and dishwasher (DW). The non controllable loads contain loads that are either very important or likely to impact consumers' life style when shifted. It includes refrigerator, cooking and lighting loads.

2.2 Machine learning techniques

2.2.1 General presentation

Supervised machine learning is one of the most commonly used and successful type of machine learning. In this section, we introduce some popular supervised learning algorithms used in this work. Supervised learning is used to predict a certain output from a given input, and we have examples of input/output pairs. We build a machine learning model from these input/output pairs, which comprise the training set. The objective is to make accurate predictions for new, never before seen data. Supervised learning often requires human effort to build the training set, but afterward automates and often speeds up laborious or infeasible task. There are two major types of supervised machine learning problems, called classification and regression. In classification, the goal is to predict a class label, which is a choice from a predefined list of possibilities. Classification is sometimes separated into binary classification, which is the special case of distinguishing between exactly two classes, and multiclass classification, which is classification between more than two classes. For regression tasks, the goal is to predict a real number. Predicting energy consumed for weeks in a house given habitant's education, age, and where they live is an example of a regression task. When predicting income, the predicted value is an amount, and can be any number in a given range. Another example of a regression task is predicting the yield of a corn farm given attributes such as previous yields, weather, and number of employees working on the farm. The yield again can be an arbitrary number. In the following sections, the different Machine Learning techniques used in this thesis are introduced.

2.2.2 K-NN

The K nearest neighboring algorithm (K-NN) is a non-parametric supervised learning algorithm that is intuitive and easy to implement. This algorithm is often called "lazy learning" or memory-based because it is based only on training data. It can be used for both classification and regression. Its principle consists in calculating the distance between a new data to be classified and the referenced data that form the training database. The unseen data is classified by a vote to the plurality of its neighbors with data being assigned to the most common class among its k nearest neighbors. There are different types of distance calculation : Euclidean, Manhattan, Minkowski or Chebyshev distance. An example of classification by K-NN is shown in the figure [2.2](#).

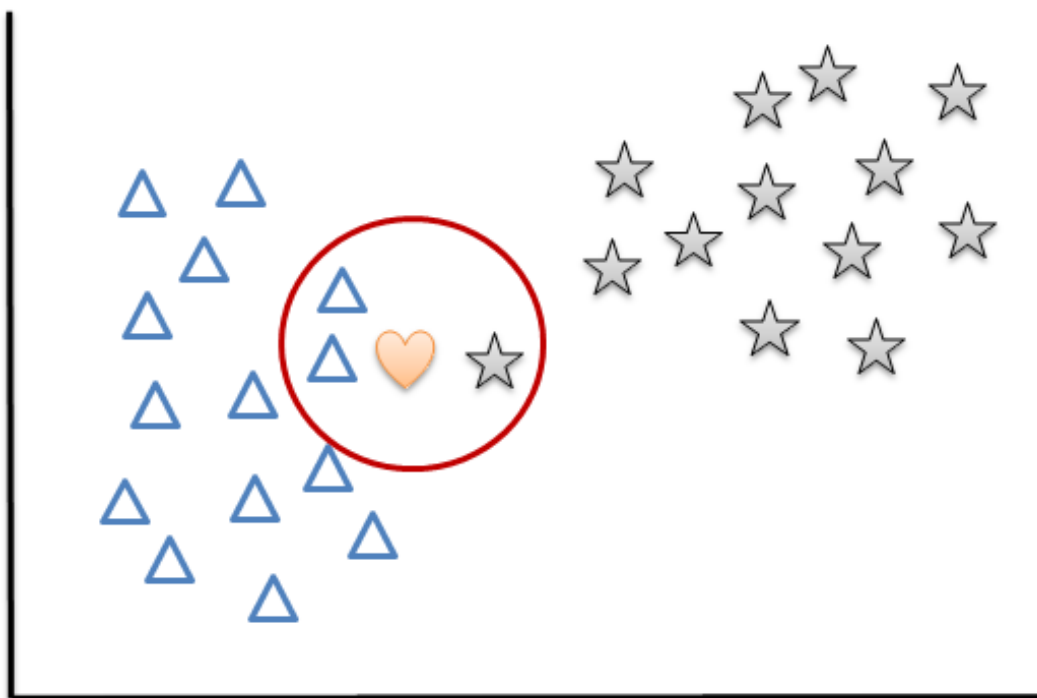


FIGURE 2.2 – Example of a 3NN classification

In this figure, the first class is represented by triangles and the second by stars. The new object to be classified is in the form of a small heart. Using a 3NN classifier ($k=3$), we find 2 nearest neighbors of the new object belonging to the triangles class against one of the stars class. The classifier then considers that this new object belongs to the triangles class.

2.2.3 Random Forests

Random forest is basically made up of a large number of decision trees that function as a set as illustrated by figure 2.3. Each tree is slightly different from the others. The idea behind random forests is to do an acceptable job of prediction for each tree, but will likely be overfitted on some of the data. If you build many trees, all of which work well and overfit in different ways, you can reduce the amount of overfitting by averaging their results.

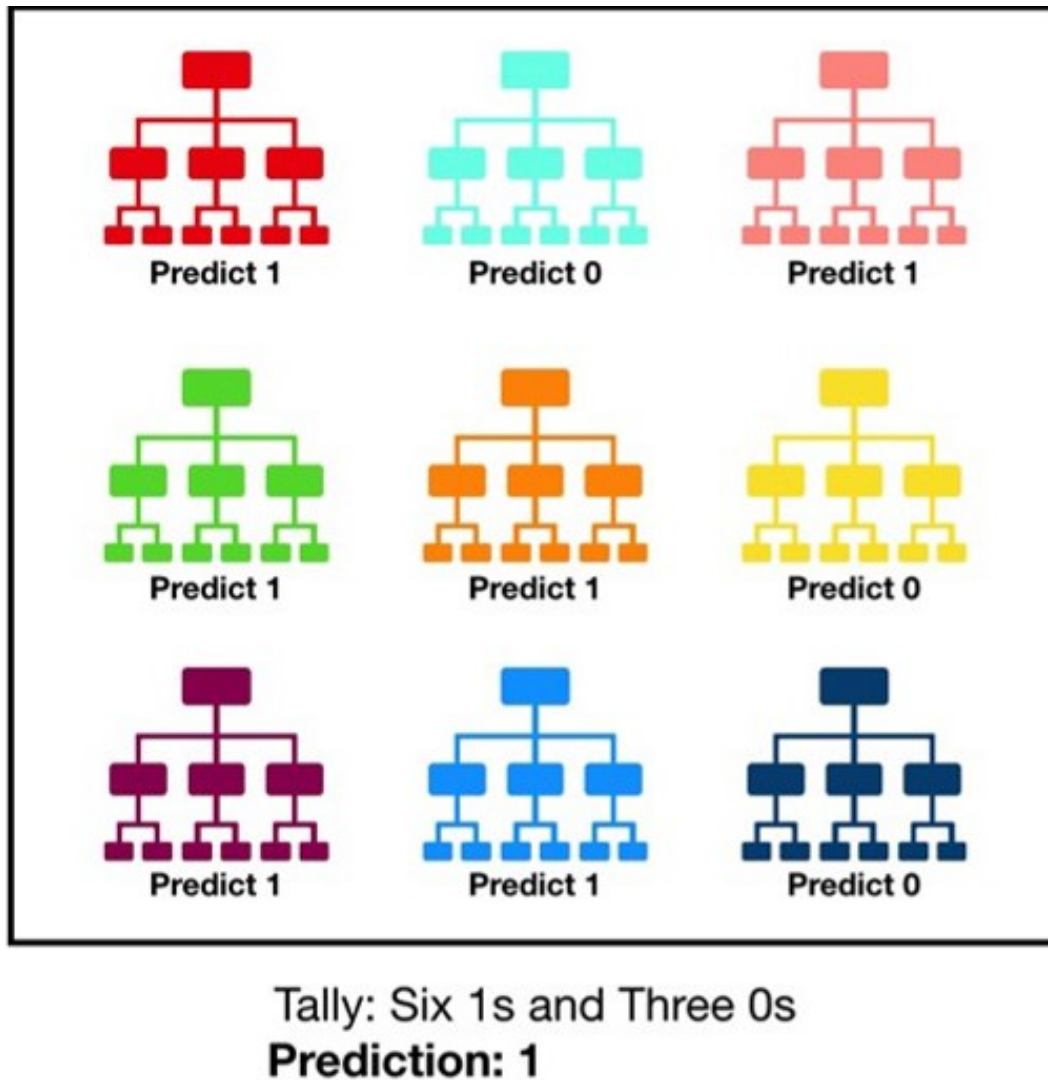


FIGURE 2.3 – Visualization of a random forest model making a prediction [Yiu, 2019]

To implement this strategy (reduction of overfitting), we will have to build a fairly large number of decision trees. Each tree should do an acceptable job of predicting the target and should also be distinct from other trees. Random forests get their name from the injection of randomness into the construction of the tree to ensure that each tree is different. There are two ways to randomize trees in a random forest : by selecting the data points used to build a tree and by selecting the features in each split test. To make a prediction using the random forest, the algorithm first makes a prediction for each tree in the forest. For the regression, we can average these results to get our final prediction. For classification, a “soft voting” strategy is used. This means that each algorithm performs a "soft" prediction, providing a probability for each possible output tag. The probabilities predicted by all trees are averaged and the class with the highest probability is predicted as can be seen in figure 2.3

2.2.4 Artificial Neural Network

2.2.4.1 Introduction

One of the main tasks of an energy management system in a household is the prediction and classification of electrical signatures. This task can be performed using various tools such as smart plugs as well as real-time visualization tools. When an anomaly occurs on a household appliance (e.g. an unexpected over-consumption or a power outage), the information changes and users must then identify their anomalies according to the type of architecture (IALM or NIALM) installed in the house. As it has been demonstrated [Zufferey et al., 2012] [Antonio.Ridi et al., 2014], artificial neural networks are well adapted tools to help users to better manage their energy consumption and also detecting anomalies (i.e. by closely monitoring an appliance that consumes so much energy compared to normal use) in the case of a smart home. An artificial neural network (ANN) can be defined as a set of small computing units linked together by communication links. The information conveyed by these connections is digital and can be coded in different ways. Each unit, which may have a small local memory capacity, performs a calculation based on the data of its connections and local data. Some neural networks model biological neural networks, others do not. Historically, the main objective of neural network research has been to increase our knowledge of the brain mechanism through the development of artificial systems capable of reproducing complex (even intelligent) calculations, similar to those performed by the human brain. Most neural networks use data-based learning rules to adjust connection weights. In other words, neural networks are usually developed from examples. They then have some ability to generalize data not present in the learning base. The neural network technique is therefore, in principle, a regression method, similar to linear or multilinear regression methods. Once the parameters (weights) have been adjusted, the neural network is a non-linear statistical model. The advantage of neural networks over conventional regression methods (i.e. linear regression, logistic regression) is that they generally require a larger number of adjustable parameters to obtain a non-linear model of a given precision.

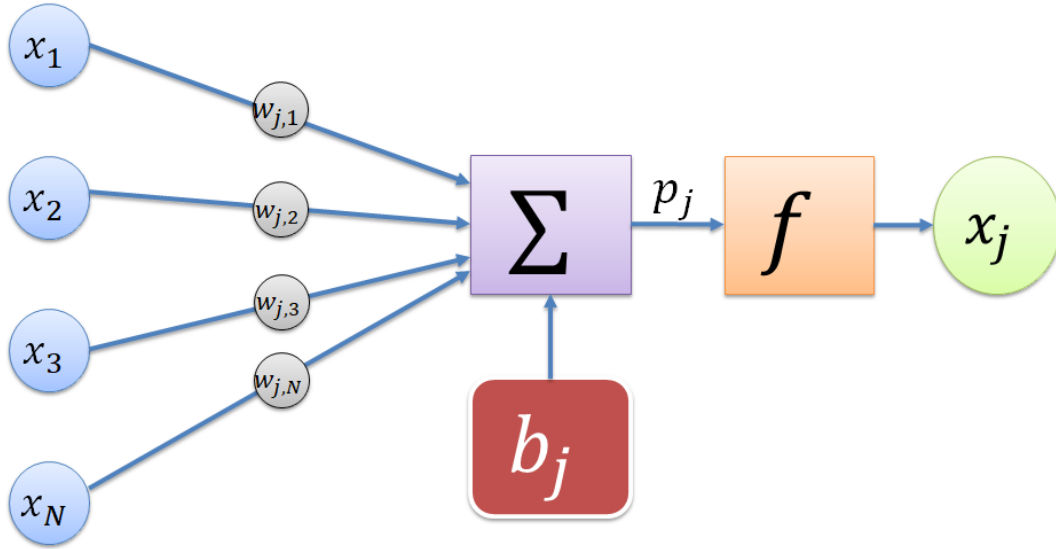


FIGURE 2.4 – Model of a formal neuron

2.2.4.2 Architectural description of a formal neuron

Architecturally, a neural network can be seen as a set of elementary units (formal neurons) interconnected to form a system with one or more inputs and one or more outputs. These neurons receive informations produced by other nodes through the input connections. The weights assigned to the inputs of a neuron are stored in a matrix w , where the value w_{ij} represents the weight of the input connection x_i of the neuron j . To this sum is added the threshold value b_j which represents the output of a "bias". This represents the biased post-synaptic potential p_j described by equation 2.1. The neuron can be modeled by two operators as shown in figure 2.4 .

- A summation operator that develops a "post-synaptic potential" p_j equal to the weighted sum of the cell inputs :

$$p_j = \sum_{i=1}^N w_{ij}x_i + b_j \quad (2.1)$$

with w_{ij} represents the weight, b_j the bias and x_i the input.

- A decision operator that calculates the state of the neuron's output x_j according to its potential p_j which can then be transmitted to other neurons. This operator is called "activation function".

$$x_j = f(p_j) \quad (2.2)$$

The calculation of the neuron state is obtained by calculating the post-synaptic potential and then applying the decision operator (activation function) to it.

Among the commonly used activation functions, we can mention :

- Linear function

$$x_j = f(p_j) = k.p_j, \forall k \in \mathbb{R} \quad (2.3)$$

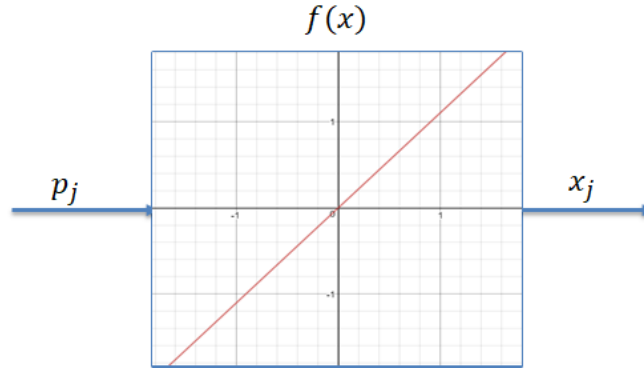


FIGURE 2.5 – Linear function

— Sigmoid or logistic function

$$x_j = f(p_j) = \frac{1}{1 + \exp(-p_j)} \quad (2.4)$$

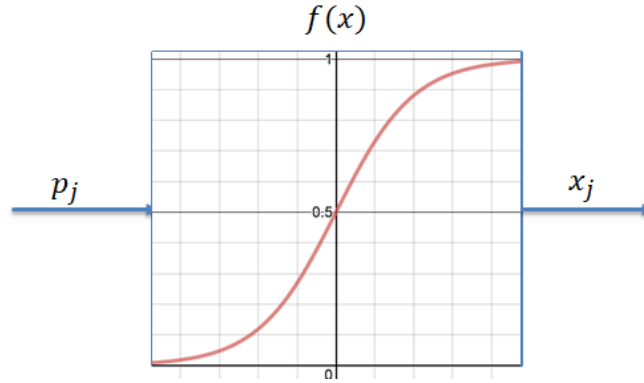


FIGURE 2.6 – Sigmoid function

— ReLU function

$$x_j = f(p_j) = \max(0, p_j) \quad (2.5)$$

— Hyperbolic tangent

$$x_j = f(p_j) = \frac{\exp(p_j) - \exp(-p_j)}{\exp(p_j) + \exp(-p_j)} \quad (2.6)$$

— Softmax function

$$x_j = f(p_j) = \frac{\exp(p_j)}{\sum_{m=1}^K \exp(p_m)}, j = 1, \dots, K \quad (2.7)$$

The sigmoid function which provides an output between 0 and 1 is mostly used as an activation function to perform a binary classification. The first formal neuron appeared in the 1940s [McCulloch and Pitts, 1943]. Since that, several types of neural networks with different

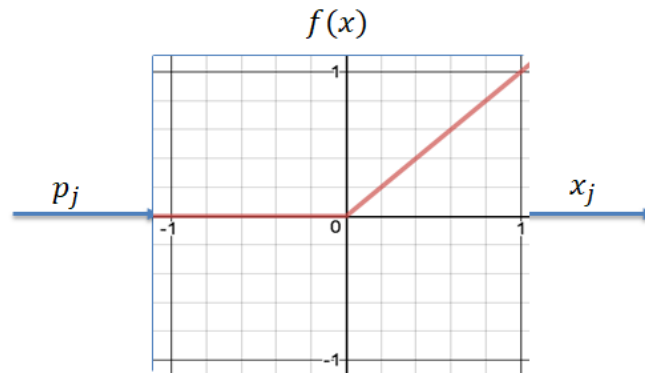


FIGURE 2.7 – ReLU function

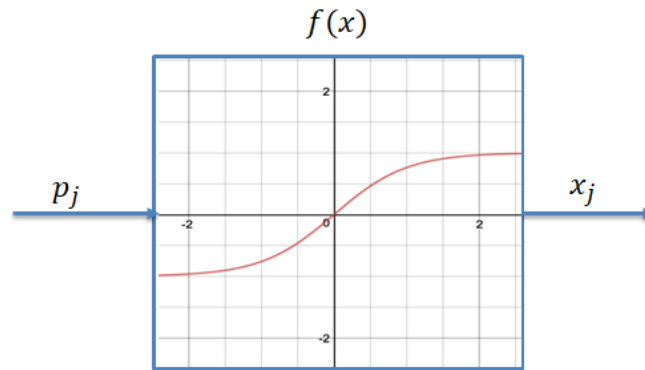


FIGURE 2.8 – Hyperbolic tangent function

properties were developed. Two types of networks can be distinguished : those whose connection graph has at least one cycle and those for which this is not the case. The first types of networks are called recurrent when the second are called acyclic (or feed-forward). Among the acyclic networks, we find perceptron [Rosenblatt, 1957], convolutional [Lecun et al., 1998], Radial Basis Function [Broomhead and Lowe, 1988] or Kohonen maps [Kohonen, 1982].

2.2.4.3 MultiLayer Perceptron

In 1957, F.Rosenblatt was inspired by the work on formal neurons [McCulloch and Pitts, 1943] and on the Hebb rule introduced in [Hebb, 2002] to develop the perceptron model but also the method which allows this model to learn [Rosenblatt, 1957]. Although this model has only one layer it already allows solving simple tasks of geometrical symbol classification. It is however impossible with the method formulated by F. Rosenblatt to train a system having several layers which turns out a few years later very restrictive. As a result, M. Minsky and S. Papert demonstrated in 1969 by a rigorous analysis that the perceptron is unable to learn functions if they are not linearly separable. They went a step further by demonstrating that it is necessary to have at least one additional layer of neurons in order to solve the problem. But there were at that time no way to train a multi-layered perceptron. It took more than ten

years for researchers to implement the learning method that allows adjusting the parameters of a multi-layer perceptron. Indeed it was P. Werbos who first proposed the idea of using the error gradient retro-propagation technique developed in the 1960s for the ANNs learning. During his thesis, which he defended in 1974 [Werbos, 1974], he analysed the relevance of this method but, given the lack of interest of the scientific community for ANNs following the publication of M. Minsky and S. Papert, he did not publish any results on the subject until 1982 [Werbos, 1970]. It was finally in the mid-1980s that the method was re-discovered by several research teams [Parker, 1985] [LeCun, 1985] [Rumelhart et al., 1986] and finally popularized. In 1986, D. Rumelhart, G. Hinton and R. Williams showed in [Rumelhart and McClelland, 1987] that using error gradient backpropagation applied to a multi-layer perceptron, it was finally possible to overcome the limitations of the perceptron that had been raised by M. Minsky and S. Williams. Papert in 1969. In particular, the multi-layer perceptron can handle complex non-linear problems and can approximate, with a single hidden layer and a sufficient number of neurons, any continuous non-linear function over a compact space with arbitrary precision [Hornik et al., 1989]. The multilayer perceptron is thus called a universal function approximator. The multilayer perceptron is part of the layered acyclic neural networks. Their neurons and their connections form an acyclic oriented graph in which information flows in one direction only (i.e. from input to output). The first layer is called the input layer, the intermediate layer(s) is called the hidden layer and the last layer is called the output layer. The figure 2.9 gives a representation of a multilayer perceptron with two hidden layers.

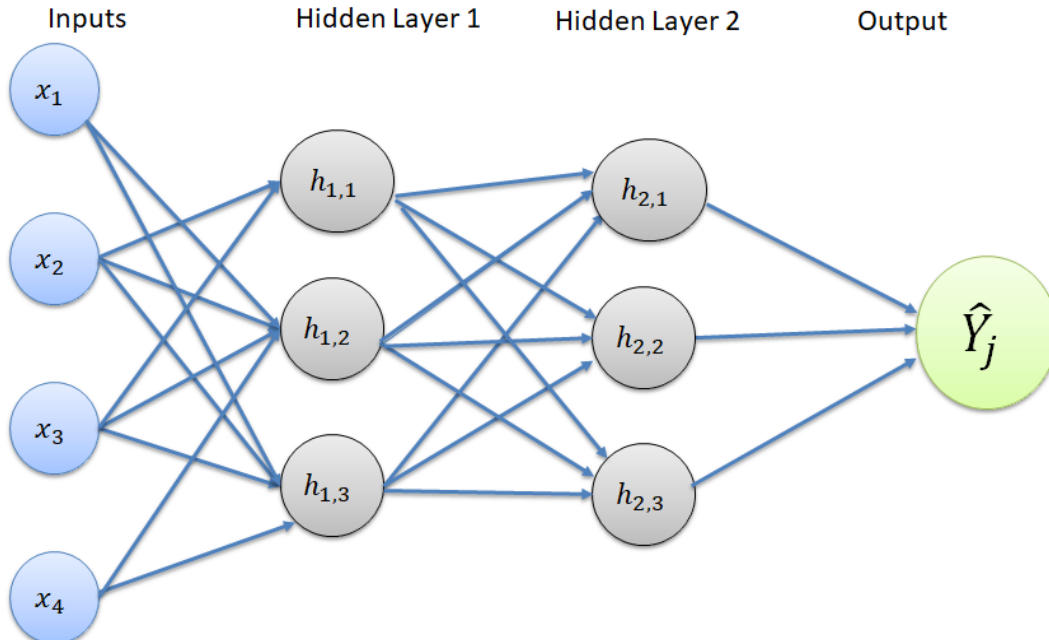


FIGURE 2.9 – A multilayer perceptron with two hidden layers

The activations of the input layer receive the information provided by the input vectors of each instance. This layer therefore does not have any input connections from other nodes. However, it is completely connected to the first hidden layer. In an MLP containing N hidden layers, each of the $(N-1)$ hidden layers is completely connected to the upper one. The N^{th} hidden layer is completely connected to the output layer. The activation of the neurons in the output layer represents the values of the MLP output vector. The neurons of these intermediate layers as well as those of the output layer apply two treatments : a linear combination of their inputs (whose weights are network parameters), followed by a non-linear function (i.e the activation function). Multilayer Perceptrons are generally used for supervised classification problems but less frequently for processing time series data. This implies the existence of a set of input-output pairs linked by a certain relationship, which the network will "learn" by adjusting its parameters. The learning is performed with the gradient back-propagation algorithm. These networks have nevertheless some disadvantages. In fact, the learning phase can be a few minutes to several hours long depending on the complexity of the problem. In addition, there is no formal methodology for the design and construction of this type of network. The choices of network hyperparameters need some experience to get the desired performance.

Gradient Backpropagation algorithm

According to [Bouaziz, 2017], the learning phase is called when the connection weights adapt to the prediction errors found in each classification of a new instance. The Gradient retropropagation (backpropagation) [Rumelhart and McClelland, 1987] is the most commonly used method for adjusting these weights. Its principle can be summarized in three basic steps :

- Routing information through the network
- Retropagating error and gradient calculation
- Adjust the parameters using the approximate gradient rule

The retropropagation algorithm is a method that calculates the error gradient for each neuron in the network from the last layer to the first hidden layer. The purpose of this algorithm is to adjust the connection weights to minimize a cost function, for instance the mean square error represented by the equation 2.8.

$$E = \frac{1}{2} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 \quad (2.8)$$

This error represents the difference between the expected output Y_i and the output estimated by the network $\hat{Y}_i$ corresponding to a given input vector N . N represents the size of output vectors. Despite its unavoidable role in optimization problems, backpropagation algorithm has limitations inherent in the gradient technique because of the risk of being trapped in a local minimum. It is enough that the gradients or their derivatives are zero for the network to be blocked in a local minimum. Additionally, there is a risk of the slow convergence especially when dealing with large

networks (i.e. when the number of connection weights to be determined is important). Considering that the error E is a function of the weights w , a local minimum is aimed by changing the weights in the opposite direction of gradient multiplied by the learning rate λ :

$$\Delta w_{i,j} = -\lambda \frac{\partial E}{\partial w_{i,j}} \quad (2.9)$$

In order to accelerate the convergence of back-propagation taking into account previous updates of the weights, a modified version of equation 2.9 has been introduced in [Plaut et al., 1986] by adding a second term called momentum :

$$\Delta w_{i,j}(m) = -\lambda \frac{\partial E}{\partial w_{i,j}}(m) + \alpha \cdot \Delta w_{i,j}(m-1) \quad (2.10)$$

where α is a weighting coefficient between 0 and 1, m being the number of iterations.

Let us take an example of a MLP network of P neurons containing D layers. Layer q ($2 \leq q \leq D$) contains M_q neurons to which a bias vector b_q and a weight matrix w_q is assigned. In this matrix, an element w_{ij}^q represents the weight of the connection from the neuron i ($1 \leq i \leq M_{q-1}$) of layer $(q-1)$ to the neuron j ($1 \leq j \leq M_q$) of layer q . For each new instance, a learning pass is performed in 3 steps :

- Forward propagation : The input vector has been copied into the x_i^1 activations of the first layer. Then, for each of K layers, leaving the first layer hidden and going up to output layer, the biased potential p_i^q and activation x_i^q of neuron i are calculated.
- Backward Propagation : Calculate the derivative Δi^d of error E for activation x_i^d of the output layer neuron i :

$$\Delta i^d = (Y_i - x_i^d) \frac{\partial f(p_i^d)}{\partial p_i^d} \quad (2.11)$$

Then, going down from the last hidden layer $h = d - 1$ to the first $h = 2$, calculate the term Δi^c for each neuron :

$$\Delta i^h = \sum_{j=1}^{U_{h+1}} \Delta j^{h+1} w_{i,j}^{h+1} \frac{\partial f(p_i^d)}{\partial p_i^d} \quad (2.12)$$

- Adaptation of the parameters. Update the bias and weights of nodes j for each layer q :

$$\hat{b}_j^q = b_j^q + \lambda \Delta i^q \quad (2.13)$$

$$\hat{w}_{i,j}^q = w_{i,j}^q + \lambda \Delta i^q x_j^q \quad (2.14)$$

The learning phase is much more expensive than the inference phase. The latter simply consists in propagating forward, as in step 1 of the learning algorithm, the data of each new instance in order to obtain the inference result.

2.2.4.4 Recurent neural network

In recent years there has been a resurgence in the use of recurring neural networks for temporal series processing, especially in predictive tasks. Unlike the MLP previously seen, a network of recurrent neurons is a network whose connection graph contains at least one cycle. The architecture of this type of network is intended to manipulate sequences of input vectors. They have a memory of what has been calculated in the past, thus making them particularly suitable for processing sequences. In theory, networks of recurring neurons can keep in memory the information seen in an arbitrarily large sequence but in practice lose their effectiveness on very long-term dynamics. It is in this sense that recent work has seen the emergence of recurrent neural network architectures with "gate" mechanisms that allow a considerable improvement in the memory capacities of models. Indeed, with cumulative calculations over the long term, the error obtained with the backpropagation of the gradient decreases or, less frequently, increases exponentially with respect to the time scale. These two problems are called "vanishing gradient" and "gradient exploding" respectively [Kolen and Kremer, 2001]. We also note that this type of problem also existed in deep non-looped architectures. The dissipation or explosion of the gradient worsens in this case depending on the number of layers. One of the most effective solutions to overcome this problem of calculating the gradient is manifested in an extension of the concept of RNNs, namely, the Long Short-Term Memory (LSTM) architecture [Hochreiter and Schmidhuber, 1997] which will be detailed in chapter 4.

2.3 Energy Harvesting Technology

2.3.1 Introduction

A local energy production system integrated into a house allows an additional energy source to be added to the system. In a connected site (the house is connected to the electrical grid), local energy sources can replace some (or all) of the energy from the grid. An additional energy source such as a generator can help to ensure security of energy supply when the main energy source is unavailable. Several renewable energy sources are detailed in the next sections.

2.3.1.1 Wind power generation

Wind is one of nature's renewable energy resources and can become an environmentally friendly and capable source of energy. The integration of wind systems in house is more difficult than photovoltaics because of its requirement of regularity on the speed of the wind. Large wind turbines are designed to "turn on" when the wind speed reaches 15km/h and to "turn off" when the wind speed is very high, so that the wind does not damage them. When determining whether the wind energy at a site is sufficient to run a wind energy system smoothly, it is very important to consider the average annual wind speed and the number of days that the wind speed is above the switch-on speed. In addition, to achieve the condition of stable wind speed, the installation should generally be about 100 meters away from obstacles of the same height and 10 meters higher than nearby. The weak point of solar and wind energy is their intermittent nature and their dependence on the weather elements. This makes it difficult to use them without an additional energy source and storage system, unless an energy management system is used such as the one proposed in chapter 5 are used.

2.3.1.2 Power generation by a generator set

A generator set is a self-contained device capable of producing electricity. Most generators consist of an internal combustion engine that drives an alternator. Generating sets are either used in isolated areas or in certain buildings to compensate for possible power cuts. In the second case, they are often used in parallel with a battery power supply or a power source inverter system. They generally operate from fossil fuel energy.

2.3.1.3 Production of electrical energy from solar energy

Photovoltaic solar cells are semiconductors capable of directly converting light into electricity. This conversion, called photovoltaic effect, is done without moving parts, without pressurized fluids, without pollution or waste generation. Due to the lack of sufficient energy production at the level of a cell, several cells are associated in series and/or in parallel to form modules in

2.3 Energy Harvesting Technology

order to provide the external receiver with adequate voltage and power. It is worth to note that the serial association of the cells increases the resulting voltage while the parallel association increases the resulting current [energie+, 2010]. The integration of photovoltaics is quite easy either on a sloping roof covered with discontinuous elements, or on a flat roof or a roof covered with a seal, or on an insulated wall or not, or on a balcony facade or a balustrade. The output current, and hence the power, is usually proportional to the surface of the installed module. An inverter converts the direct current produced by the photovoltaic panels into alternating current compatible with the power distribution network. The output of the inverter is square waves, which are filtered and transformed into sine wave shapes of alternating current. Any waveform, when analyzed, is essentially the superposition of many forms of sine waves known as harmonics. The first harmonic represents a pure sine waveform. Additional higher frequency waveforms, when superimposed on the base waveform, add or subtract amplitude from the base sine waveform. The combination of the basic waveform and the higher harmonics produces a distorted waveform that looks like a distorted sine wave. The output of converted direct current, derived from solar energy, is seen as a superposition of many odd and even harmonic numbers. To achieve a relatively clean sine wave output, most inverters use electronic circuitry to filter out a large number of harmonics. Filter circuits are made up of inductive circuits and capacitors specially designed to block certain unwanted harmonics. In general, DC-AC inverters are complex electronic power conversion equipment designed to convert direct current into single-phase or three-phase power that mimics regular electrical services provided by utilities. Most inverters, in addition to PV module input power, accept auxiliary input power to form a back-up generator, which is used to provide current when the battery voltage is lowered to a minimum level. A special type of inverter, called the grid-connected type, incorporates a synchronization circuit that allows the production of sine waveforms in harmony with the electrical utility grid. The electricity generated can be used by the building's energy services or exported to the grid, in whole or in part. The European HIP project [Cohen-Akne, 2014] conducted an experiment on the integration of photovoltaic systems in buildings in seven European countries. This project demonstrates the feasibility and profitability of photovoltaic systems. The integration of photovoltaic systems in buildings in the current context benefits from an investment and installation subsidy as well as a feed-in tariff for the energy produced. In the following, we will focus on the type of solar energy in more details, its operation and the different types of installations most commonly used.

2.3.2 Grid-Connected Photovoltaic System

2.3.2.1 Introduction

Photovoltaic modules or panels are made up of semi-conductors that convert sunlight directly into electricity. These modules are a source of electrical energy that is safe, reliable, maintenance free and non-polluting. The majority of solar modules on the market today come with warranties

of more than 20 years, and they will work well beyond that period. Millions of systems have been installed around the world, with different powers ranging from a fraction of a watt to several megawatts. For many applications, solar power systems are not only cost effective, but they can also be the cheapest option. In this chapter, we will start by briefly recalling the generality of the photovoltaic system as well as the different types of Grid connected photovoltaic system used in houses.

2.3.2.2 The photovoltaic effect

In Greek, the word photovoltaic means light, it is composed of two parts : "photos" (light) and Volta which is the surname of the Italian physicist (Alessandro Volta) who invented the electric battery in 1800 and gave his name to the unit of measurement of electrical voltage, the volt. When a semiconductor material is exposed to sunlight, as shown in figure 2.10, the atoms exposed to the radiation are "bombarded" by the photons constituting the light ; under the action of this bombardment, the electrons of the upper electronic layers (called valence shell electrons) tend to be "torn off". If the electron returns to its initial state, the stirring of the electron causes a heating of the material. The kinetic energy of the photon is transformed into thermal energy.

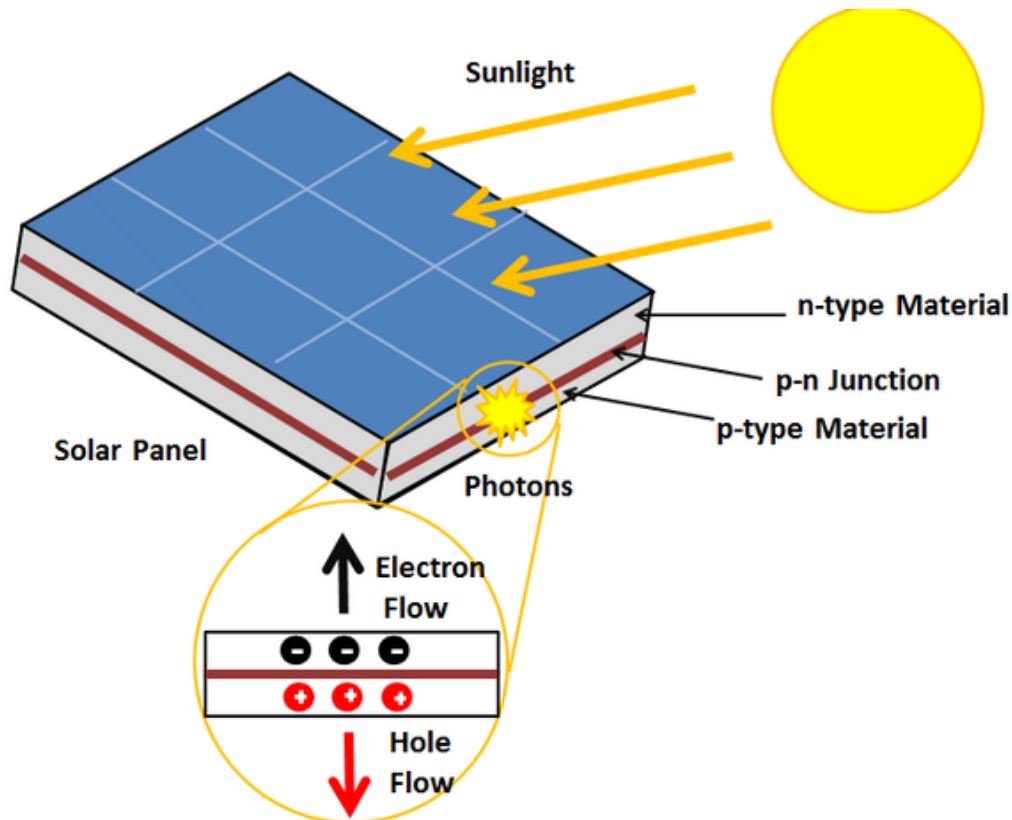


FIGURE 2.10 – Photovoltaic effect [B. Afework, 2018]

2.3 Energy Harvesting Technology

However, in photovoltaic cells, some electrons do not return to their initial state. The torn-off electrons create a low DC voltage. Part of the kinetic energy of the photons is therefore directly transformed into electrical energy : this is the photovoltaic effect. The photovoltaic effect is the direct conversion of the energy of solar radiation into electrical energy by means of cells generally based on silicon. In order to obtain sufficient power, the cells are connected to each other and form the solar module. The photovoltaic effect, i.e. the production of electricity directly from light, was first observed in 1839 by the French physicist Edmond Becquerel. However, it was not until the 1950s that researchers at Bell Laboratories in the USA succeeded in manufacturing the first photovoltaic cell, the main component of a photovoltaic system.

2.3.2.3 Different types of solar cells

There are different types of photovoltaic solar cells, each type of cell has its own efficiency and cost. However, regardless of their type, their efficiency remains quite low : from 5 to 22% of the energy they receive [Sendy, 2020]. There are three main types of solar cells :

a) Monocrystalline silicon PV panels

These are made using cells sliced from a single cylindrical crystal of silicon. This is the most efficient photovoltaic technology, typically converting from 17% to 22% of the sun's energy into electricity. The manufacturing process requires to produce monocrystalline silicon, which is complicated, resulting in slightly higher costs than other technologies.

b) Polycrystalline silicon Photovoltaic panels

Also sometimes known as multicrystalline cells, polycrystalline silicon cells are made from cells cut from an ingot of melted and recrystallised silicon. The ingots are then saw-cut into very thin wafers and assembled into complete cells. They are generally cheaper to produce than monocrystalline cells, due to the simpler manufacturing process, but they tend to be slightly less efficient, with average efficiencies of around 12%.

c) Thick-film silicon Photovoltaic panels

This is a variant on multicrystalline technology where the silicon is deposited in a continuous process onto a base material giving a fine grained, sparkling appearance. Like all crystalline PV, it is normally encapsulated in a transparent insulating polymer with a tempered glass cover and then bound into a metal framed module.

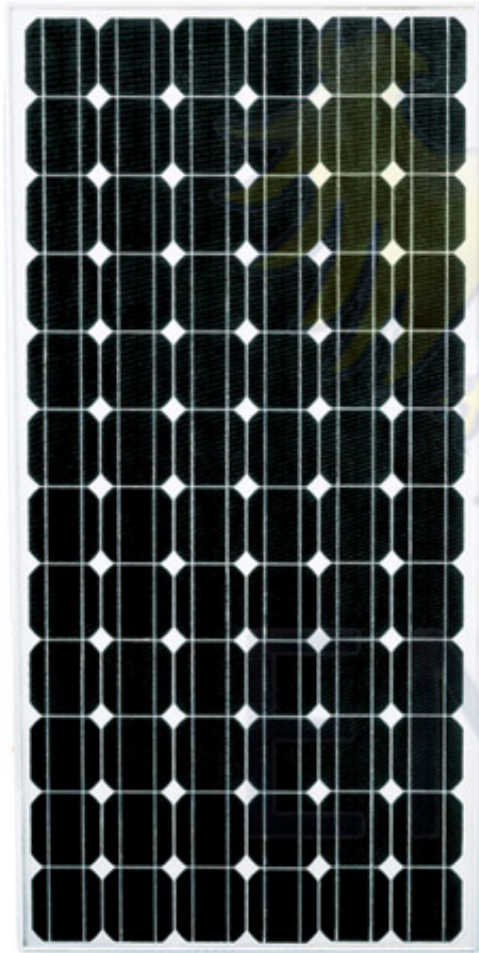


FIGURE 2.11 – Monocrystalline silicon Photovoltaic panels

d) Amorphous silicon Photovoltaic panels

Amorphous silicon cells are made by depositing silicon in a thin homogenous layer onto a substrate rather than creating a rigid crystal structure. As amorphous silicon absorbs light more effectively than crystalline silicon, the cells can be thinner - hence its alternative name of 'thin film' PV. Amorphous silicon can be deposited on a wide range of substrates, both rigid and flexible, which makes it ideal for curved surfaces or bonding directly onto roofing materials. This technology is, however, less efficient than crystalline silicon, with typical efficiencies of around 6%, but it tends to be easier and cheaper to produce. If roof space is not restricted, an amorphous product can be a good option. However, if the maximum output per square meter is required, design should choose a crystalline technology.



FIGURE 2.12 – Polycrystalline silicon Photovoltaic panels

e) Other thin film Photovoltaic panels

A number of other materials such as cadmium telluride (CdTe) and copper indium diselenide (CIS) are now being used for PV modules. The attraction of these technologies is that they can be manufactured by relatively inexpensive industrial processes, certainly in comparison to crystalline silicon technologies, yet they typically offer higher module efficiencies than amorphous silicon. Most offer a slightly lower efficiency : CIS is typically 10-13% efficient and CdTe around 8 or 9%. A disadvantage is the use of highly toxic metals such as Cadmium and the need for both carefully controlled manufacturing and end-of-life disposal ; it is worth to note that a typical CdTe module contains only 0.1% Cadmium, which is reported to be lower than is found in a single AA-sized NiCad battery.



FIGURE 2.13 – Amorphous silicon solar panel

2.3.2.4 Types of Photovoltaic system

Photovoltaic power systems are generally classified based on their functional and operational requirements and how the PV system is connected to other energy sources and electrical loads. It can be classified into grid-connected and stand-alone systems. These photovoltaic systems provide a basic service of DC and AC current supply, may operate in interconnection with or independently of the grid. These systems may also be connected to other energy sources and energy storage systems.

Grid-connected or utility-interactive PV systems

The main component of grid connected PV systems is an inverter which converts DC power produced by the PV array into AC power in accordance with the voltage and power quality requirements of the power grid. A bi-directional interface is provided between the inverter output



FIGURE 2.14 – Thin film silicon solar panel

and the power grid, usually at a distribution panel or a field service entrance. This allows the AC power produced by the PV system to supply the site's electrical loads or to inject the surplus into the grid when the output of the photovoltaic system is greater than the demand of the site's load. Figure 2.15 shows the general block diagram of the grid-connected photovoltaic system. In grid-connected systems, the switching of the alternating current from the standby generator and inverter to the service bus or the connected load is carried out by internal or external automatic transfer switches. An important component of a network-connected system is net metering. Standard service meters are electromechanical meters that record energy consumption according to the speed of the rotating disc. Rotating discs operate according to an electro-physical principle called eddy current. Digital electricity meters use digital electronic technology that records power measurement by solid-state current and voltage sensing devices that convert analog measured values into binary values that are displayed on the meter using a liquid crystal display (LCD). Inverters are the main difference between a grid-connected system and a stand-alone system. Inverters must have line frequency synchronization capability to supply excess power to the grid. Net meters have the ability to record the power consumed or produced in a proprietary summation format. The power record is the net amount of power consumed - the total used power minus the amount of power that is produced by the solar cogeneration system. Net Meters are supplied and installed by utility companies that provide grid connection service systems. Net-metering PV solar power plants are

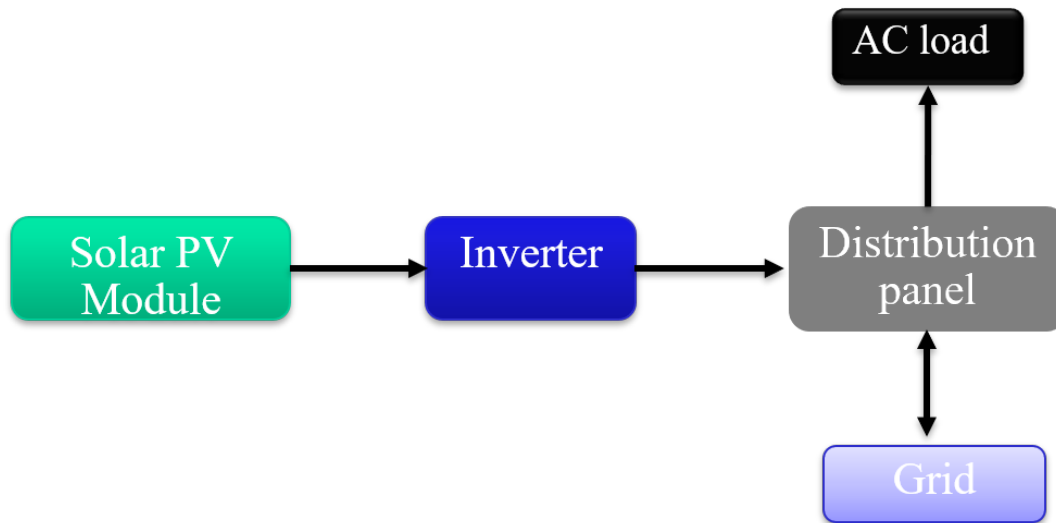


FIGURE 2.15 – Grid-connected solar PV system

subject to specific contractual agreements and are subsidized by state and municipal government agencies.

Stand-Alone Solar PV System

. Stand-alone PV systems or direct-coupled PV systems are designed and dimensioned to supply electrical loads with direct and/or alternating current. They are called direct-coupled systems because the DC output of a photovoltaic module or array is directly connected to a DC load. Direct-coupled systems do not have electrical energy storage (batteries), so charging only works during sunny hours. The MPPT (Maximum Power Point Tracker) is used between the generator and the load to help making a better use of the maximum available power from the generator and to adapt the impedance of the electrical load to the maximum power of the PV generator.



FIGURE 2.16 – Stand-alone PV system without battery storage

Figure 2.16 shows the general block diagram of the stand-alone PV system. A typical example of a direct-coupled solar photovoltaic system is in agricultural applications, where the solar photovoltaic module can be directly connected to operate the pump. Depending on the capacity of the pump, the module can be connected in series or parallel. In this type of application, a lightning arrester must be connected between the positive and negative power supply to provide protection against lightning surges. Batteries are used for energy storage in many stand-alone PV systems.

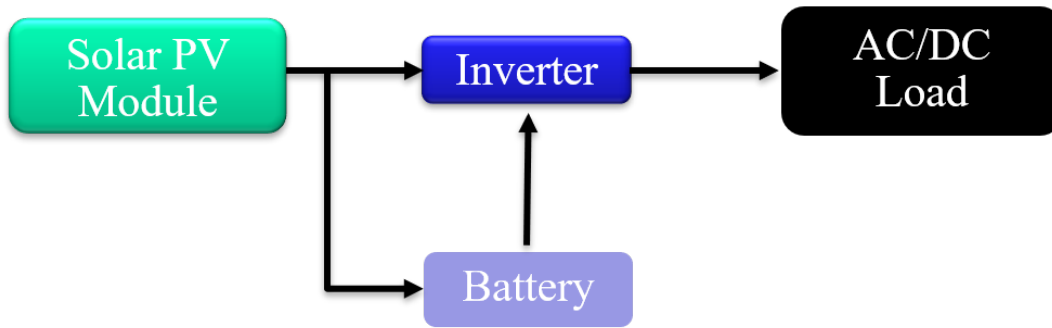


FIGURE 2.17 – Stand-alone PV system with battery storage

Figure 2.17 shows the block diagram of a typical stand-alone PV system supplying DC and AC loads with a battery storage option. The configuration of the PV solar array, a DC load with battery backup, is essentially the same as that without battery, except that there are some additional components that are necessary to ensure stable battery charging. The PV panels are connected in series to achieve the desired increase in DC voltage, e. g. 12, 24 or 48 V. The charge regulator regulates the current output and prevents the voltage level from exceeding the maximum value for charging the batteries. The output of the charge regulator is connected to the battery bank by means of a double DC cut-off switch. In addition, a cut-off switch can be provided, when it is switched off for safety reasons, to disconnect the load and the photovoltaic panels at the same time. During sunny hours, the load is supplied with direct current while simultaneously charging the battery. The charge controller will ensure that the continuous output power of the photovoltaic panels is sufficient to support the connected load in order to be able to size the required battery size. Battery sizing depends on a number of factors, such as the duration of uninterrupted power supply to the load when there is less or no solar radiation. The battery bank produces a 20-30% loss of power due to heat when in operation, which must also be taken into consideration. When designing a solar photovoltaic system with battery backup, the designer must determine the appropriate location for battery supports and room ventilation.

PV-Hybrid Systems

Hybrid systems generally refer to the combination of two input sources, for example a solar panel system integrated with a diesel generator, wind turbines, biomass or any other renewable or non-renewable energy source. Solar panel systems typically use a bank of batteries to store the energy produced by the panels for later use, when needed. There may still be exceptional periods of bad weather when an alternate source is needed to guarantee electricity production. Hybrid photovoltaic systems combine a solar panel module with other energy sources, usually a diesel generator, but sometimes also another renewable energy source such as a wind turbine. The photovoltaic generator is generally sized to meet the basic demand, the alternative source of supply being used only if it is essential. Hybrid systems can also be a wise approach in situations where

occasional peaks in demand are significantly higher than base demand. It makes little sense to size a system so that it can fully meet demand with PV if, for example, normal load is only 10% of peak demand. Likewise, a diesel generator set sized to meet peak demand would operate at inefficient part load most of the time. In such a situation, a PV-diesel hybrid would be a good compromise. Figure 2.18 shows the block diagram of a photovoltaic hybrid system.

Stand-Alone Hybrid AC Solar Power System with Generator and Battery Backup

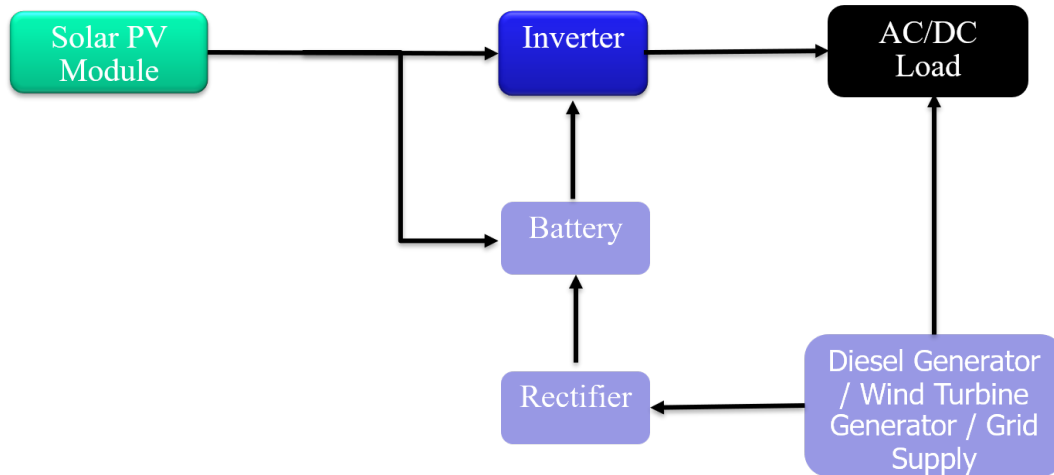


FIGURE 2.18 – Photovoltaic hybrid system

A stand-alone solar photovoltaic hybrid configuration is essentially identical to a DC solar power system. In this case, alternating current inverters are used to convert direct current into alternating current. When the inverter is connected to the power grid, it can act effectively as a source of AC power generation. Grid inverters used in grid-connected solar energy systems are strictly regulated by the utility agencies that provide net metering. Some inverters incorporate an internal AC transfer switch that is capable of accepting an output from an AC backup generator. In this type of design, inverters include special electronics that transfer energy from the generator to the load.

2.3.3 Battery Bank for PV system

In general, a “battery” is a combination of several electrochemical cells in series-parallel. There are two types of cells : primary and secondary cell. A cell is said to be primary when it is non-rechargeable. It is filled with a variety of chemicals which react irreversibly, which means that when the chemical reaction is exhausted, the battery is dead. Primary cells can be connected in series to obtain a specific voltage and should never be connected in parallel as there is the possibility that one cell will try to charge the other. A secondary cell is said to be rechargeable and commonly referred to as an accumulator. Secondary cells can be used in series, in parallel, or in a combination

of the two to achieve the required voltage and capacity. The only limitation is that each cell has the same voltage, capacity and chemical composition. Rechargeable batteries are the only ones used in homes with a solar panel system. Solar batteries are accumulators that store the electrical energy that solar panels produce. The charges can be electrochemical or electrostatic. The charge current stored in this way can be restored when connected to a receiver. There are several types of solar batteries on the market. Among these batteries we can cite lead acid batteries, lithium batteries, and salt water batteries. Increasingly, residential storage systems are made up of lead-acid or lithium-ion batteries.

Lead acid solar batteries

For decades, lead acid batteries have been regarded as the only residential energy storage solution. Among these lead-acid batteries, we distinguish two types :

- Flooded lead acid batteries
- Sealed lead acid batteries

The difference of these types of batteries is the number of maintenance for proper operation. Flooded lead-acid batteries require at least maintenance every 1 to 3 months to ensure proper operation, while sealed lead-acid batteries require no maintenance. These lead acid batteries are the most solar batteries that give the cheapest option. Despite its lower cost, they are much larger and heavier than other batteries (Lithium-ion) which will require more space per KWh of storage than other types of batteries. Their Depth of Discharge *DoDs* (a DOD is defined as the percentage of capacity that has been withdrawn from a battery compared to the total fully charged capacity) are getting lower and lower, usually around 50%. To maintain a home's energy deficit, they need to be recharged more frequently. They have a lifespan of 5 to 10 years.

Lithium-ion solar battery

As for lithium-ion batteries, they are a new type of energy storage technology. Currently, they have become popular among residential customers because of the benefits they provide. They are lighter than lead acid batteries so they take up less space for the same capacity (KWh) and have a lifespan of more than 10 years. This long lifespan is largely due to the fact that lithium batteries have a higher DoD. For this, the depletion of lithium-ion batteries is slower than lead acid batteries. Lithium-ion solar batteries have an average DoD of 90% or more. The only limitation of these batteries is that they can catch fire if there is an overload (thermal escape) or improper handling. Currently, these lithium batteries are equipped with advanced monitoring software to prevent catching fire.

2.4 Conclusion

This chapter introduced the concept of smart home in all its complexity, particularly that linked to the different technologies used. Machine learning techniques have been then presented, with a specific focus on methods used in this work. Finally, we discussed the benefits of using renewable energy for smart homes. We presented the different types of photovoltaic based systems as well as the typical technologies used to harvest and store ambient energy. In the following chapter, we describe our methodology for collecting data using a wireless infrastructure to determine the relevant information needed in a home and the impact of this information in the case of appliance recognition (section [3.4](#)).

Electrical Signatures and Analysis of smart home appliances

Sommaire

3.1 Introduction	35
3.2 Related work	38
3.3 Energy monitoring	40
3.3.1 Definition	40
3.3.2 A wireless infrastructure to collect data for energy management in a smart home	40
3.3.3 Choice of a Home Automation System	43
3.4 Impact of electrical signatures on Appliances identifications	46
3.4.1 Data description	46
3.4.2 Data preprocessing with moving average	46
3.4.3 Classification methods	48
3.4.4 Appliances Recognition Experiments	51
3.5 Conclusion	54

3.1 Introduction

While the electricity prices, greenhouse gas emissions and global average temperature are increasing, citizens are looking for solutions to reduce their environmental impacts. From this perspective, a considerable progress could be made in the management of the various and numerous domestic appliances. According to [Florian, 2019], residentials are one of the main consumers of energy and accounted for example in 35% of the total French energy consumption in 2016. Smart Home could be a possible solution for home energy saving. Appliance identification

plays an important role in smart home as it can provide optimized and automatic energy efficient solutions.

Every day, smart meters (such as Linky in France) or smart plugs monitoring devices are installed in houses. However, existing systems are expensive and usually complicated to implement and use. Recent studies have demonstrated that a continuous information feedback and fine-tuned automated management of home equipment's could allow an energy bill reduction from 15% to 30% [Neenan, 2009][Darby, 2006]. More and more mobile or web-based energy monitoring systems are proposed to provide pertinent information to the end users [M. Weiss and Roediger, 2009][D. Guinard and Trifa, 2009]. However, to be more adaptive and effective, they must integrate an automatic recognition of running appliances. Such identification can be made using automatic learning algorithms applied to the devices power consumption [Ridi Antonio and Jean, 2013][Ridi et al., 2014]. To do so, end users have to install monitoring systems on each appliance and manually label the associated appliances. Finally, users can visualize the energy consumption per appliance category and may perform optimization by activating or programming control rules. This recognition system can also present other utilities such as defects detection, localization of the appliances in offices or hospital for instance.

Every day, new smart plugs or smart meters appear on the market. The table 3.1 lists four of these plugs. For collecting information, an architecture based on smart plugs is required. Sensing information is then sent to a gateway and can be visualized on a dashboard (figure 3.1).

It is worth to note that these smart meters provide a variable number of electrical signatures (from 1 to 6) of different types (e.g. real power and reactive power). In this chapter, we propose to study an infrastructure to collect features which will be used to evaluate the impact of the number and the type of features used for appliance identification. Most of the appliances exhibit power features close to zero. This behavior corresponds to stretches of time where the appliances are not used, typically in stand-by or in off state. To take into account this kind of behavior, we propose a technique based on data smoothing to classify distinct appliances included in the dataset.

Manufacturer	Name	Description	References
EnOcean Alliance	NodOn	It is a smart socket using the EnOcean protocol. The role of this plug is to measure the instantaneous power (w) and the cumulative energy (Wh). However, it is remotely controlled by an EnOcean compatible home automation system or via other controllers using EnOcean such as Soft Remote or the NodOn wall switch	[NodOn, 2020]
Fibaro Fibaro	wall plug	It is a smart socket that allows you to control lighting or any other appliances via Z-wave controls. It is equipped with a visual indicator that shows the consumption where green indicates low consumption and red a very high consumption	[Fibaro, 2020]
Belkin Wemo Insight	Wi-Fi Smart Plug	It can monitor energy consumption from your phone or tablet. This Smart Plug provides real-time reports on how much energy your appliances are consuming. It uses a smart socket using the WI-FI protocol	[Belkin, 2020]
Zigbee Alliance	Smart Plug Mini	The wireless Smart Plug Mini with power metering feature is an intelligent, sharp and sophisticated, remotely controlled adapter. This smart plug can be applied wherever you want to control electrical devices, while monitoring the power consumption in a convenient and maintenance-free way	[Alliance, 2020]

TABLE 3.1 – Some smart plugs available on the market

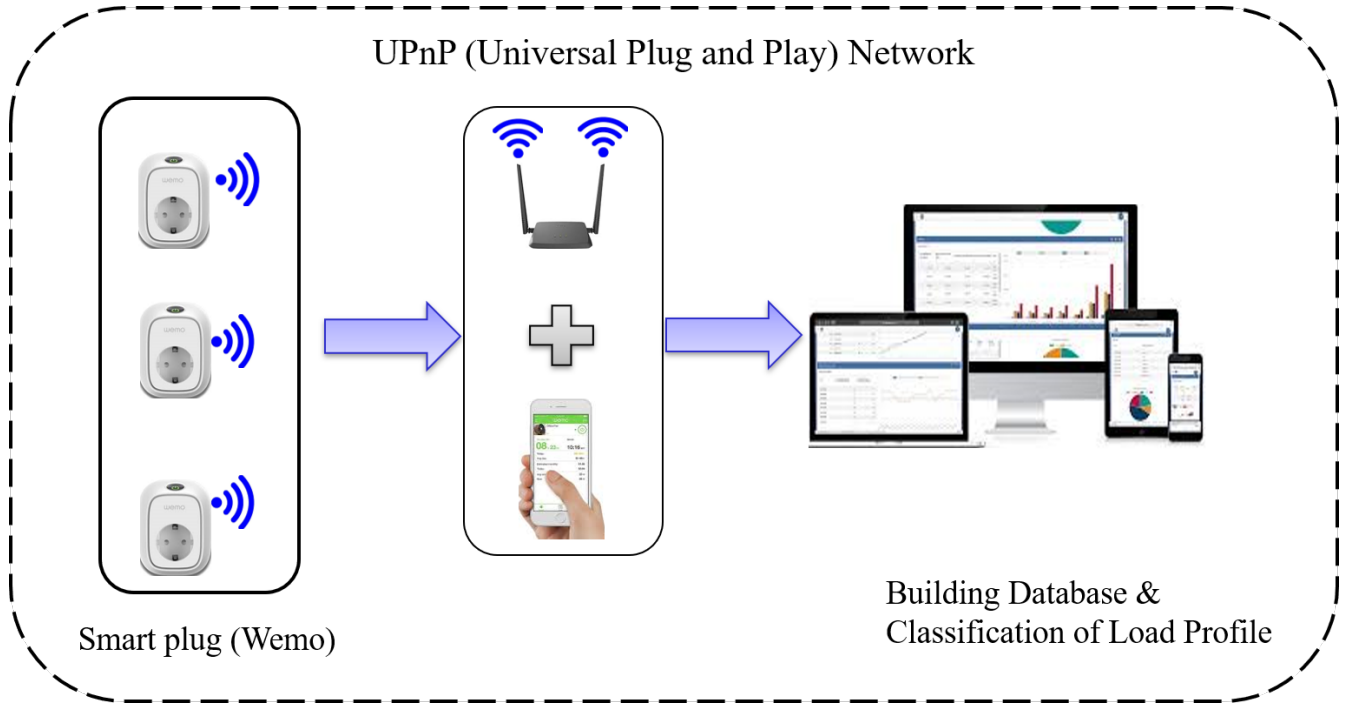


FIGURE 3.1 – Architecture to collect data with WEMO insights

3.2 Related work

Over the last few years, energy consumption analysis and appliance recognition received a growing attention from researchers. We can distinguish two main approaches according to the sampling rate of the energy measurements : high frequency and low frequency. High frequency measurement, also called Non-Intrusive Load Monitoring (NILM) [Hart, 1992], is based on the analysis of the transient signatures appearing when appliances are switched on or off. This approach, requiring a sensor with an expensive high sampling frequency placed at the house electricity's entry point, presents the problem of the sensibility to the noise from other appliances. On the other hand, a low frequency type of approach, called Intrusive Load Monitoring (ILM) [Antonio.Ridi et al., 2014], is based on the analysis of the short to medium term evolution of the electricity consumption, from few seconds to few minutes. Sensors used in this approach are in the home living space, typically close to monitored appliances and consequently to people using them. A NILM type of approach presents several advantages and drawbacks compared to ILM. NILM is based on a single sensor. The sensing system installation is therefore easier and data acquisition is quite trivial. However, appliance recognition using a NILM kind of approach generally suffers from inaccuracies due to the inherent problem of summation of consumption signals from different appliances. On the other hand, ILM type of approach is based on multiple sensors deployed in the house. Despite sensors are generally cheaper than for NILM, the overall cost increases linearly with the number of sensors. ILM also presents several advantages over

3.2 Related work

NILM. One of them is that more details in consumption signatures are available, thus facilitating the appliance identification. Typically, low power appliances or appliances in stand-by cannot be detected with NILM [A. Marchiori and Earle, 2011], while it is feasible with ILM. Yet, it is well-known that stand-by power was one of the largest source of domestic consumption [L. Harrington and Ellis, 2008], representing up to 26% of the total energy consumption in 2001 [Ross and Meier, 2001]. Since that time, constructors have made efforts to reduce the stand-by power of appliances, and we can therefore expect that the related energy consumption has been significantly reduced. Another drawback of NILM is the difficulty to detect appliances with multiple functioning states such as a dishwasher, or exhibiting continuously variable energy use such as an electric stove [A. Marchiori and Earle, 2011]. Finally, it is also not possible to remotely control an appliance (switch on or off) in the living space using NILM.

In this chapter, our aim is to propose a system which can automatically recognize home appliances. To do so, a dataset merging the power consumption profiles from different individual appliances needs to be constructed. In fact, we used the existing ACS-F1 dataset [Ridi Antonio and Jean, 2013] that contains electrical load consumption signatures of 100 appliances uniformly spread among 10 classes and acquired at a 10^{-1} Hz frequency. Each appliance is recorded on two acquisition sessions of one hour. In this dataset, samples have been collected at a low frequency using smart-outlet for electric signal analysis. Actually, few results have been reported on appliance identification task. A main reason for that is the lack of available public data and the difficulty to obtain a sufficient number of samples required for the analysis. The ACS-F1 dataset has been chosen in this work because it offers several features for a significant number of devices. Moreover, several studies have been carried on this dataset [Ridi Antonio and Jean, 2013] [Gisler et al., 2013], [Zufferey et al., 2012], which is not yet the case for the more recent version ACS-F2.

3.3 Energy monitoring

3.3.1 Definition

In this work, we propose to investigate a mixed approach combining intrusive and non intrusive approaches presented in section 2.1.2. To leverage advantages from both methods, we propose to use an intrusive approach for controllable appliances (i.e appliances that can be switch ON/OFF, shift it on peak off), and a non intrusive one to collect the global power consumption which will be used in the following chapters in order to predict the overall consumption (as it will be shown in chapter 4), as well as in chapter 5 in decision making by a power management policy.

3.3.2 A wireless infrastructure to collect data for energy management in a smart home

3.3.2.1 Types of sensors and protocols in SMART Home

At the beginning of this thesis, we first proposed an intrusive data collection infrastructure based on the wemo sockets as shown in figure 3.1. This infrastructure is designed to monitor energy consumption and remote control of domestic equipment's via the actuators included in these sockets. Wemo sockets use WI-FI as a communication protocol and only measure the instantaneous power (active power). Today, many plugs can be found on the market carrying a diversity of communication protocols as it can be seen in table 3.2. Electrical signatures measured by these plugs are diverse :

- Active power
- Reactive power
- Voltage
- Current
- Power factor
- Frequency

The challenge for a monitoring infrastructure is to offer a low expensive solution that largely takes into account this diversity of communication protocols illustrated in table 3.2. In this thesis, we propose a low cost system shown on figure 3.2 and capable of handling connected sockets based on different protocols (Zwave, RFXcom, EnOcean, etc.) and several sensors. This infrastructure allows a remote control of the connected sockets as well.

IWT (Integrated Wireless Technologies)	Max Transmission Speed/Operation Range	Transmission Distance	Adoption Rate	Strengths
6LoWPAN [G.Thomas, 2014] [X.Ma, 2008]	250 kbps, 2.4 GHz; 40 kbps, 915 MHz; 20 kbps, 868 MHz;	Up to 200 m	Medium	(1) Benefits of both IP and Bluetooth; (2) Low energy usage
Bluetooth [Bisdikian, 2001] [Davies, 2002]	721 kbps for v1; 2.1 Mbps for v2.0 + EDR; 24 Mbps for v3 + HS; 25 Mbps for v4, 2 400 - 2 483.5 MHz	10 m typical	Extremely high	Ease of access; No configuration requirement; Secure connection
DASH7 [Dash7-Alliance, 2020]	200 kbps; 433,92 MHz	0-500 m and 0.3-1 km	Medium	(1) It penetrates concrete and water; (2) It transmits and receives over very long ranges without requiring a large power drawn on a battery
EnOcean Technology [Anders, 2020]	868.3 or 315 MHz; 125kbps	30 m (in-door)	Medium	(1) Energy management and highly efficient energy storage; (2) It uses wireless standards optimized for solutions with ultra-low power consumption

IWT	Max Transmission Speed/Operation Range	Transmission Distance	Adoption Rate	Strengths
RFID [X.Jia, 2012]	Low 124-135 kHz; High 13.56 MHz; Ultra-high 860-960 MHz; 106 - 848 kbps	Low 30 cm; High 1.5 m; Ultra-high 1-15 m	Widely adopted	(1) Stable technology; (2) Continue evolution; (3) Open architectures becoming increasingly available
UWB [D. Porcino, 2003]	500MHz; 100 Mbps	10 m	Medium	(1) Indoor short-range high-speed; (2) Its bandwidth satisfies most of the multimedia applications
Wi-Fi (802.11n) [J.Lee, 2009]	2.4 GHz / 5 GHz; 450 Mbps	250 m	Extremely high	High speed mature standard
Z-Wave [Knight, 2007]	868.42 MHz; 908.42 MHz; 100 kbps	30m indoor; 100m outdoor	Medium	No interference from household devices
ZigBee [J.Lee, 2009] [Tsang, 2008]	250 kbps (2.4 GHz); 40 kbps (915 MHz)	10-75 m	Widely adopted	Low cost, low power usage, high number of nodes

TABLE 3.2 – Comparison among the different protocols

3.3.3 Choice of a Home Automation System

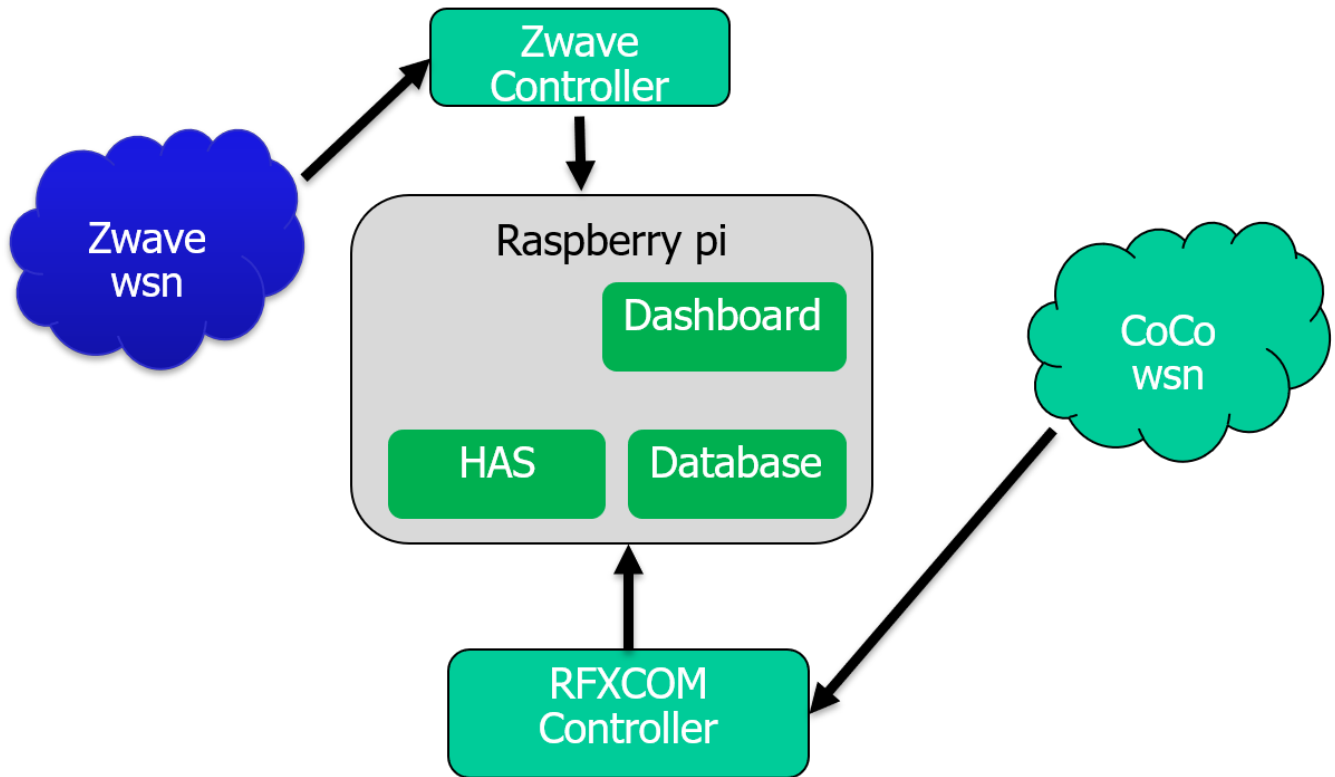


FIGURE 3.2 – Wireless infrastructure diagram

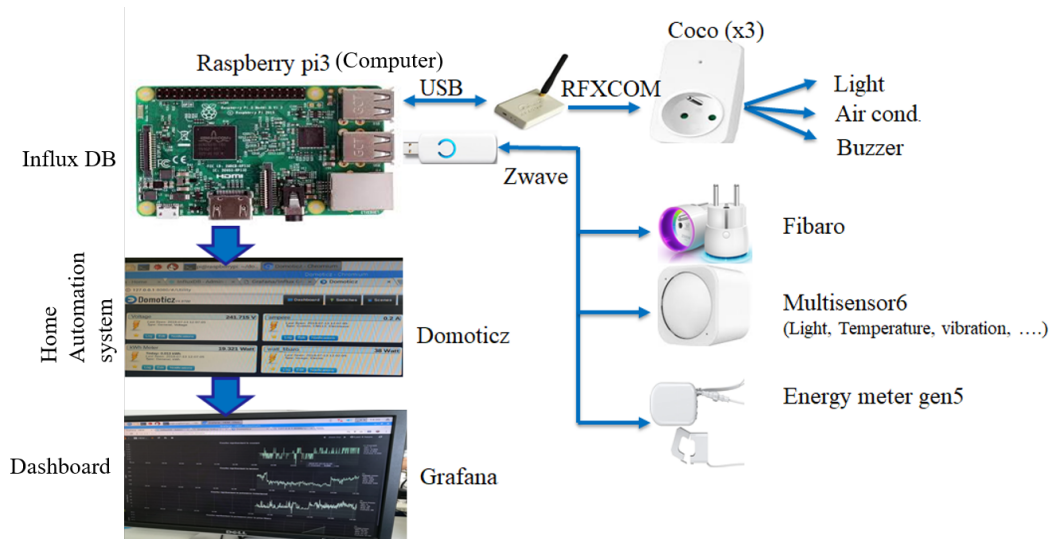


FIGURE 3.3 – Low-cost wireless infrastructure based on Domoticz, Grafana, InfluxDB and a Raspberry pi3

For collecting information in a household, an architecture based on smart sensors (or smart plugs) must be deployed. Through a wireless communication, sensing information are sent to a gateway connected to home automation system and can then be stored in a database. All data stored can then be visualized on a dashboard as shown in figure 3.2. As illustrated on figure 3.3, we developed an architecture to collect data in a household using different smart plugs such as Fibaro Plug [Fibaro, 2020], Coco plug [Chacon, 2018] and Multisensor6 [MultiSensor6, 2018]. This architecture is built around a Raspberry Pi3 computer [Pi, 2018] and an existing home automation system (HAS) called Domoticz [Domoticz, 2018]. This web front-end is interesting since various sensor devices (temperature, humidity, current, power instantaneous and so on) can be monitored and configured very easily. It is quite easy to install a home automation software such as Domoticz on a Raspberry pi since the development team has already prepared a ready-made image. Domoticz is able to communicate with a multi-protocol sensor network since it supports several radio technologies. To visualize stored data on a dashboard, we used an open source analytics and monitoring system called Grafana [Grafana, 2018]. Grafana is a powerful tool for the visualization of time series in real time supporting several databases such as graphite, influxdb and so on. In this work, we create dashboards from InfluxDB (a powerful tool to store sensed data from a home automation software) metrics collected by several sensors' devices. These metrics can be grouped together with mathematical expressions to form personalized graphs and dashboards more interactive and dynamic. They can also be grouped according to the metadata attached to each metric using dashboard variables of Grafana. This sensing prototype, though functional, has not been deployed and used to construct a dataset for different reasons. First, installing such a system includes electrical risks and requires some authorizations from both the home owner and the electrical supplier (ENEDIS in France). Moreover, having access to electrical wires may be difficult (it is not possible), leading sometimes to a difficult task to add a smart plug to each appliance.

Dataset	Location	Duration	Number of Sensors (per house)	Measured Features	Measurement Resolution	Reference
ACS-F1	Switzerland	1 hour session (2 session)	100 devices in total (10 types)	I, V, P, Q, Freq, Φ	10 secs	[Gisler et al., 2013]
BLUED	Pittsburg, PA, USA	8 days	aggregated	I, V, switch events	12 kHz	[Anderson et al., 2012]
Greend	Austira, Italy	1 year(3-6 months completed)	9	P	1 Hz	[Monacchi et al., 2014]
iAWE	India	73 days	33 sensors (10 appliances level)	V, I, F, P, S, E, ϕ	1 Hz	[Nipun et al., 2013]
IHEPCDS	France	4 years	3	I, V, P, Q	1 min	[G. Hebrail, 2012]
REDD	Boston, MA	3-19 days	9-24	Aggregate : V, P ; sub- metered : P	15 kHz (aggr.), 3 secs(sub)	[Kolter J., 2011]
Tracebase	Germany	N/A	158 devices in total (43 types)	P	1-10 sec	[Reinhardt et al., 2012]
UK-DALE	UK	499 days	5 (house 3) - 53 (house 1)	Aggregated P, Sub P, switch- status	16 kHz (aggr.), 6 sec (sub.)	[J. Kelly, 2014]
ACS-F2	Switzerland	1 hour session (2 session)	225 devices in total (15 types)	I, V, P, Q, F, Φ	10 secs	[Ridi et al., 2014]

TABLE 3.3 – Existing dataset of load consumption energy [Monacchi et al., 2014]

3.4 Impact of electrical signatures on Appliances identifications

3.4.1 Data description

Building a reliable database using an intrusive approach for device recognition is a complex task as it is necessary to collect metrics for different types of devices of different models over several days weeks or even months. As we could not build our own database, we used a freely available database. The table 3.3 presents an overview of existing available datasets of load consumption energy. As it can be seen, these databases come from different locations and have variable features and durations (from one hour to years). Moreover, a variable sampling rate can also be used to sense information. Depending on the type of approach (high or low frequency), the sampling rate varies from 120^{-1}Hz to 16 kHz. Finally, we can see that the number of sensors (or devices) as well as the type of appliances are also a variable parameter for these datasets. In this work, ACS-F1 (Appliance Consumption Signature Fribourg1) [Gisler et al., 2013] has been chosen as database of electric consumption signatures acquired from domestic appliances. It offers several features for a significant number of devices. Moreover, many results are available from this dataset [Ridi Antonio and Jean, 2013], [Gisler et al., 2013], [Zufferey et al., 2012]. This database was provided by a group of researchers on pattern recognition. It contains a total of 100 appliances of different brands and/or models as shown in table 3.4

Appliances categories(class)	Instances	Session
Coffee machines	10	2
Computers stations (with monitors)	10	2
Fridges and freezers	10	2
Hi-Fi systems (with CD players)	10	2
Lamps (compact fluorescent)	10	2
Laptops (via chargers)	10	2
Microwave ovens	10	2
Mobile phones (via chargers)	10	2
Printers	10	2
Televisions (LCD or LED)	10	2
Totals	100	200

TABLE 3.4 – APPLIANCE PRESENT IN DATASET

3.4.2 Data preprocessing with moving average

The electric sensors provide a sequence of observation $O = \{ O_1 \dots \dots O_N \}$ with O_N a vector of component corresponding to the measurements done with a 10 seconds period. A sequence of observation is composed of 6 measures : the Active Power (P in W), the Reactive Power (Q),

3.4 Impact of electrical signatures on Appliances identifications

No	Feature selected
Feature Set 1, F1	P (Active Power), Q (Reactive Power)
Feature Set 2, F2	I_{rms} (current), V_{rms} (Voltage), Φ (Phase)
Feature Set 3, F3	P, Q, I_{rms} , Φ
Feature Set 4, F4	P, Q, I_{rms} , Φ , V_{rms} , Frequency

TABLE 3.5 – FEATURE USED ON RECOGNITION PATTERN

the current (I_{rms} in Amps) and voltage (V_{rms} in Volts) Root Mean Square, the electric frequency (Hz) and the phase (ϕ). Some values included in ACS-F1 are dependent to other values according to equation 3.1 and 3.2. Our main objective is to determine the features (measures) to select for appliance recognition.

$$P = V_{rms} \times I_{rms} \times \cos \phi \quad (3.1)$$

$$Q = V_{rms} \times I_{rms} \times \sin \phi \quad (3.2)$$

Selecting the right features independent of the algorithm plays an important role in the performance of any classifier. An algorithm cannot find good features or create good features by itself. In this regard, four different types of feature sets shown in table 3.5 have been used to compare their performance. For the first feature set 1 (F1), only real power (P) and reactive power (Q) measures are considered. In F2, measurements include current root mean square (I_{rms}), voltage root mean square (V_{rms}) and the phase (Φ). The feature set F3 gathers 4 measures : real power (P), reactive power (Q), current root mean square (I_{rms}) and the phase (Φ). Finally, in F4 all the 6 measures are used by the classification algorithms. It is worth noting that the algorithm complexity increases with the number of features. Moreover, the selection of inappropriate features can result into complex decision boundary for classifiers, thus impacting the performance of the algorithm as shown in figure 3.9. The F4 features set has been used in this work to compare our results with dynamical coefficient data preparation described in annex A. A moving average is applied on these measures as a data preparation technique to create a smoothed version of the original dataset. Smoothing is useful as it reduces the random variation in the observations and better exposes the structure of the underlying causal processes. The moving average function is applied on each feature and class, and then a matrix is returned. It is worth to note that the size of the observation window also has an impact on the recognition accuracy. Our experiments show that a 33-sample sliding window (corresponding to 5 minutes and 30 sec, i.e. 33×10 secs) with an overlapping window over 32 samples allows obtaining the best accuracy. As usual for machine learning algorithms, the overall gathered data from the merging moving average matrix of each class is split into two disjoint sets, one for the training and another one for the testing phase. According to [Ridi Antonio and Jean, 2013], those sets are randomly selected with a ratio of 53%

and 47% respectively. The larger data set is used for training the classifier models, while the rest is used for testing the models.

3.4.2.1 Feature scaling

The different features of load consumption data as well as features from training and test sets are on different scales. For example, real power values in the dataset range from 0 to 1500 Watts, whereas the range is far less for current measurements (approximately from 0 to 7 Amps). However, if one feature has considerably a wider range than other features, the optimization function (for example a gradient descent) used for classification will be governed by that feature and makes the classifier unable to correctly learn from other features. Furthermore, the training phase will take much longer to converge according to the optimization objective. Feature scaling is therefore required to set all the features on the same scale, thus equally contributing to the classification algorithm. In our approach, a standardization method for feature scaling has been implemented, so that each feature has a zero mean and a unit variance. The general expression for feature scaling is defined as follows :

$$X' = \frac{\mu + M}{\sigma} \quad (3.3)$$

where X' is the new feature vector after scaling, M is the vector of moving average of each electrical signature, μ and σ are the feature vector mean value and variance, respectively.

3.4.3 Classification methods

Different types of classification algorithms have been implemented using electricity consumption data. In this study, the classification is performed using a supervised machine learning technique. Labeled data are therefore required to learn and recognize the patterns. In our case, labeled electricity consumption data with the ground truth acts as training data. Three classification algorithms used for pattern recognition and shown in table 3.6 have been evaluated in this work. The classification module from Python scikit-Learn has been used for all classification algorithms.

3.4.3.1 K-NN

The k-nearest neighbors algorithm (K-NN) is a conventional non-parametric neighbor-based classifier commonly used in data mining techniques. Neighbors-based classification is a type of instance-based learning or nongeneralizing learning : it does not attempt to construct a general internal model, but simply stores instances of the training data. Classification based on K-NN is computed from a simple majority vote of the nearest neighbors for each object : an object having the most representatives within the nearest neighbors of the object is assigned to the data class. Scikit-Learn has two different nearest neighbor's classifiers : KNeighborsClassifier

3.4 Impact of electrical signatures on Appliances identifications

and RadiusNeighborsClassifier. KNeighborsClassifier implements learning based on the k nearest neighbors of each query point, where k is an integer value specified by the user. On the other hand, RadiusNeighborsClassifier implements learning based on the number of neighbors within a fixed radius r of each training point, where r is a floating-point value specified by the user. The KNeighborsClassifier is the classification algorithm used in our recognition experiments. The optimal choice of the K value is highly data-dependent. In general, a larger k makes the classification boundaries less distinct [Ridi et al., 2014]. In this work, and as shown in the figure 3.4, we have used $k = 1$ as it provides the best accuracy.

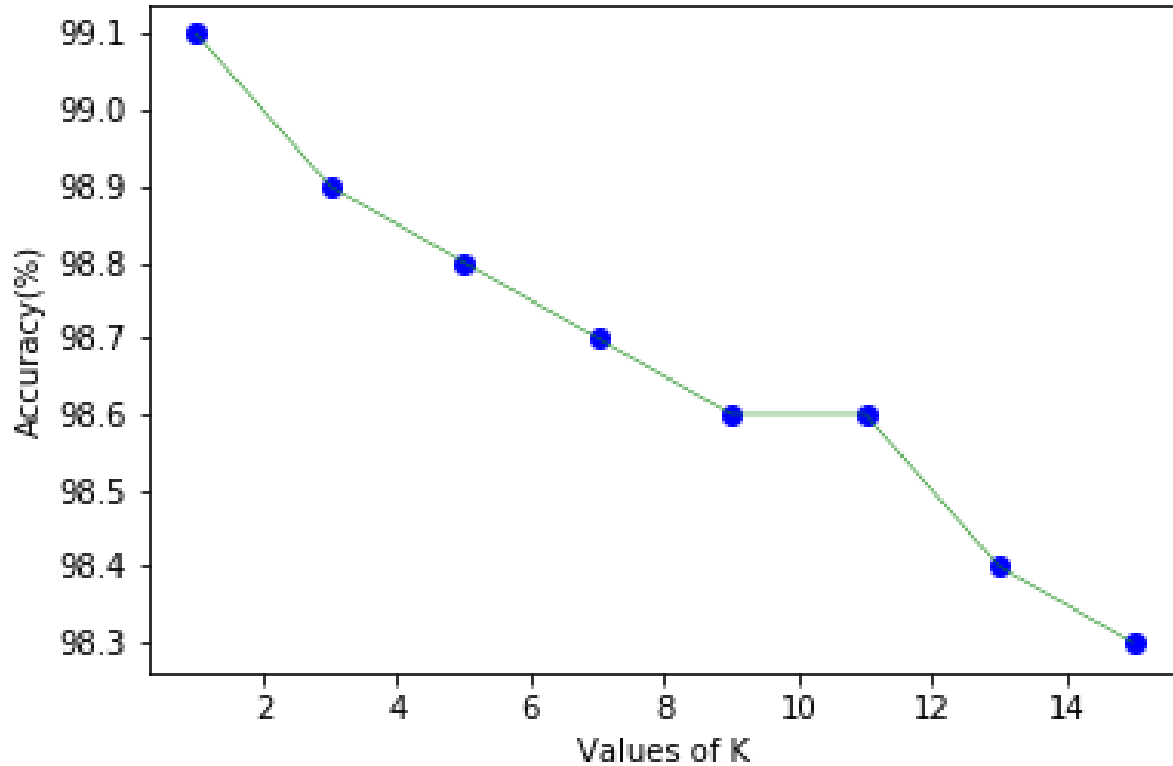


FIGURE 3.4 – Accuracy Vs. K parameter for K-NN on F4 feature set

3.4.3.2 Random-Forest

Random Forests (RFs) are ensemble classifiers used for classification and regression analysis on appliance identification data. RF works by creating various decision trees in the training phase and outputs class labels having the majority vote. RF exhibits high classification accuracy and can handle outliers and noise in the data.

3.4.3.3 Multilayer Perceptron

A multilayer perceptron classifier (MLP) has also been implemented. The multilayer perceptron consists in a system of simple interconnected neurons. This system is a model representing a nonlinear mapping between an input and an output vector. The nodes are connected by weights and output signals. The outputs signals are a function of the sum of the weighted inputs to the node modified by a simple nonlinear transfer or activation function. A multilayer perceptron may have one or more hidden layers and finally an output layer.

Classifiers	Hyperparameters	Abbreviation
K-nearest neighbors	K=1, Euclidean distance, weights=uniform	K-NN
Random Forest	n estimators = 100, criterion='entropy'	RF
Multilayer Perceptron	4 hidden layers (300 neurons per layer) epoch=500, learning rate (alpha)=0.0001	MLP

TABLE 3.6 – Classification algorithms

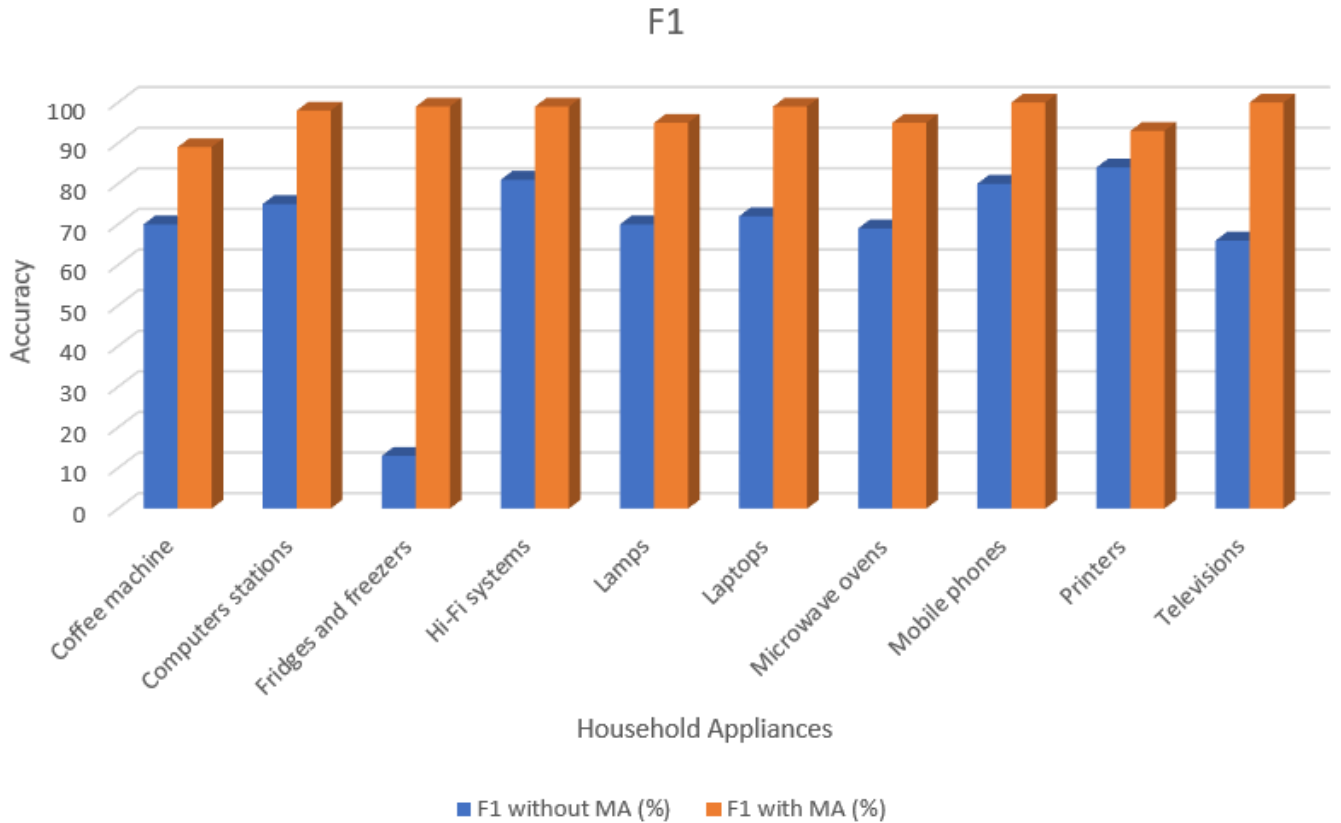


FIGURE 3.5 – Accuracy per Class for K-NN on F1 features set

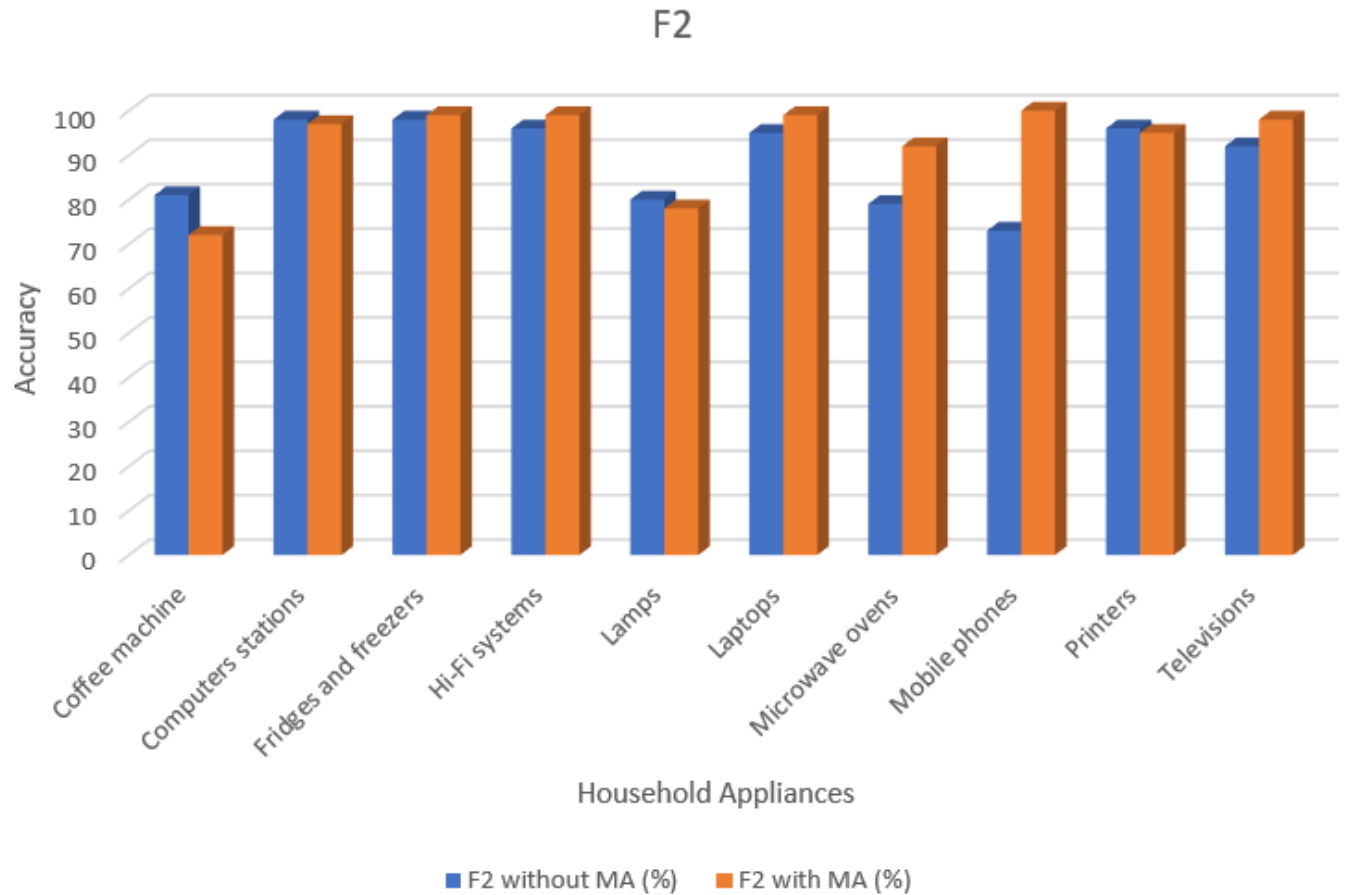


FIGURE 3.6 – Accuracy per Class for K-NN on F2 features set

3.4.4 Appliances Recognition Experiments

The objective of this study is to evaluate the effectiveness of machine learning techniques for appliances recognition and the impact of electrical signatures in the identification of household devices. As the complexity of the algorithm increases with the number of features, the selection of inappropriate features can lead to complex decision limits for the classifiers, thus affecting performance. First, a comparison based on the performance of the classifiers was carried out. Accuracy results obtained for the three classifiers on the F4 features set are shown on the table 3.7. As can be seen, the K-NN provides the best performance with a 89.1% accuracy. The

Data preparation technique	Classifier	Accuracy
Without Moving Average	K-NN	89.1%
	Random Forest	86%
	Multilayer perceptron	74%
With Moving Average	K-NN	99.1%
	Random Forest	99%
	Multilayer perceptron	98.8%

TABLE 3.7 – Obtained accuracy for different classification algorithms with features set F4

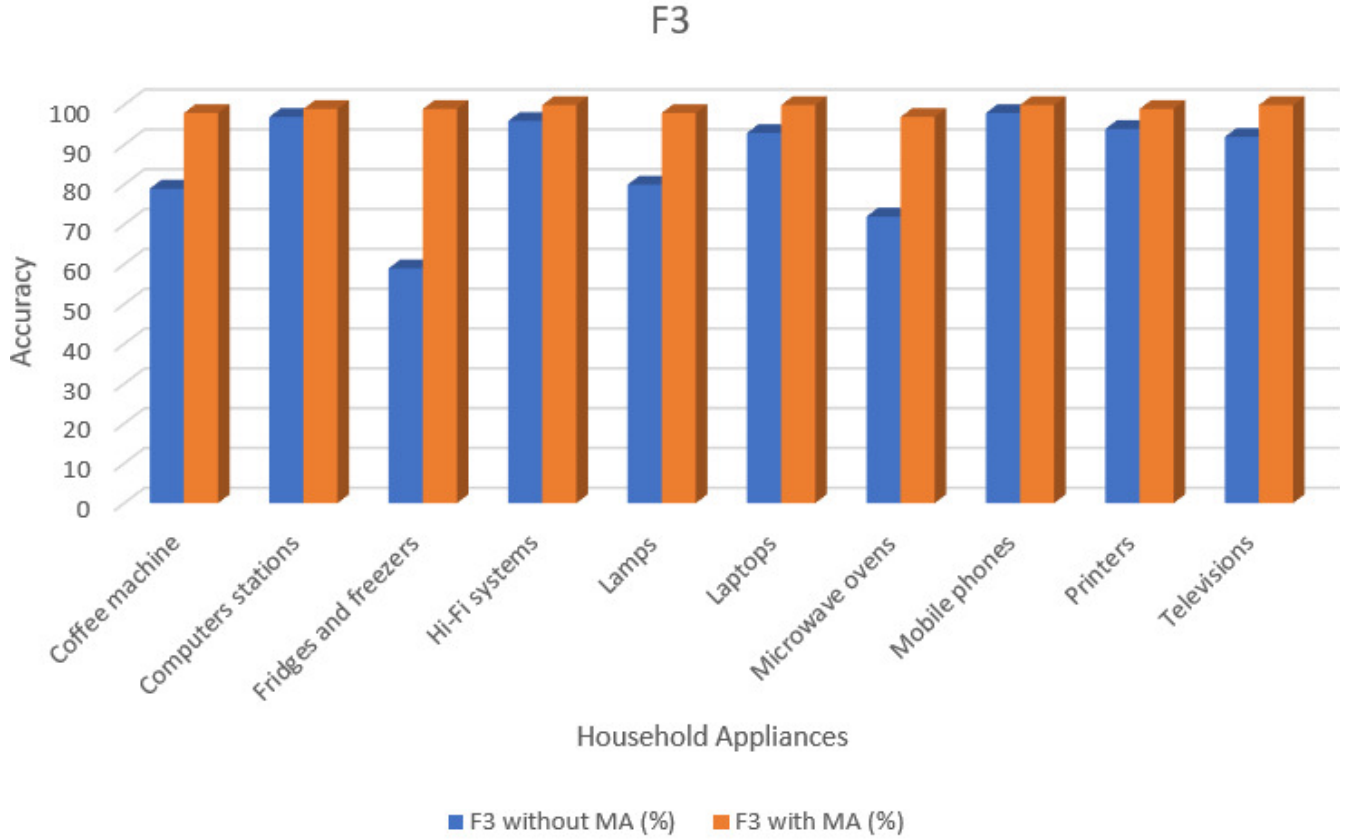


FIGURE 3.7 – Accuracy per Class for K-NN on F3 features set

RF and MLP provide an accuracy of 86%, and 74% respectively. We can also observed in table 3.7 that the moving average technique provides a clear improvement for all classifiers. RF provides 86% to 99% accuracy, while MLP has 74% to 98% accuracy without and with a moving average technique respectively. The dynamical coefficient data preparation technique described in annex A [Ridi Antonio and Jean, 2013], [Gisler et al., 2013] has also been evaluated for the K-NN classifier. The table 3.8 presents accuracy results obtained with different data preparation techniques when a K-NN classifier is performed using 10 classes (and 10 devices by class) and the F4 features set. The moving average function provides better performance than the dynamic coefficient based method. It is worth noting that this method provides two additional characteristics, the delta and delta-delta coefficients. In consequence, for each electrical signature the original dataset is changes to a feature space equal to 18, while the moving average function keeps the same dimension ($D = 6$) and reduces a little bit the number of samples.

The KNN classifier was then chosen to study the impact of electrical signatures in the recognition of household devices. In our study, F1, F2, F3 features set and F4 were used for that purpose. The figures 3.5, 3.6, 3.7 and 3.8 provide accuracy results obtained using K-NN, with or without moving average data preparation technique, for the different features set. As it can be

3.4 Impact of electrical signatures on Appliances identifications

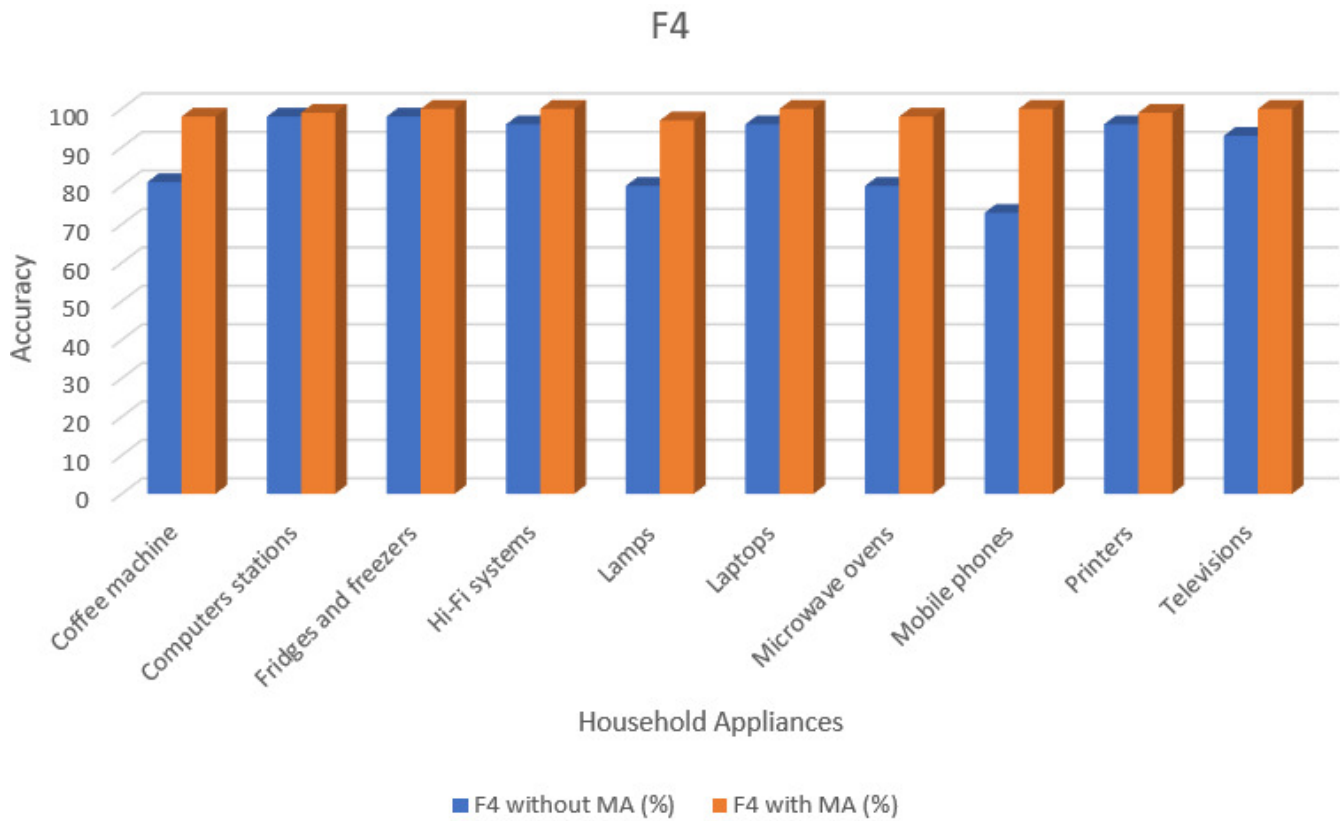


FIGURE 3.8 – Accuracy per Class for K-NN on F4 features set

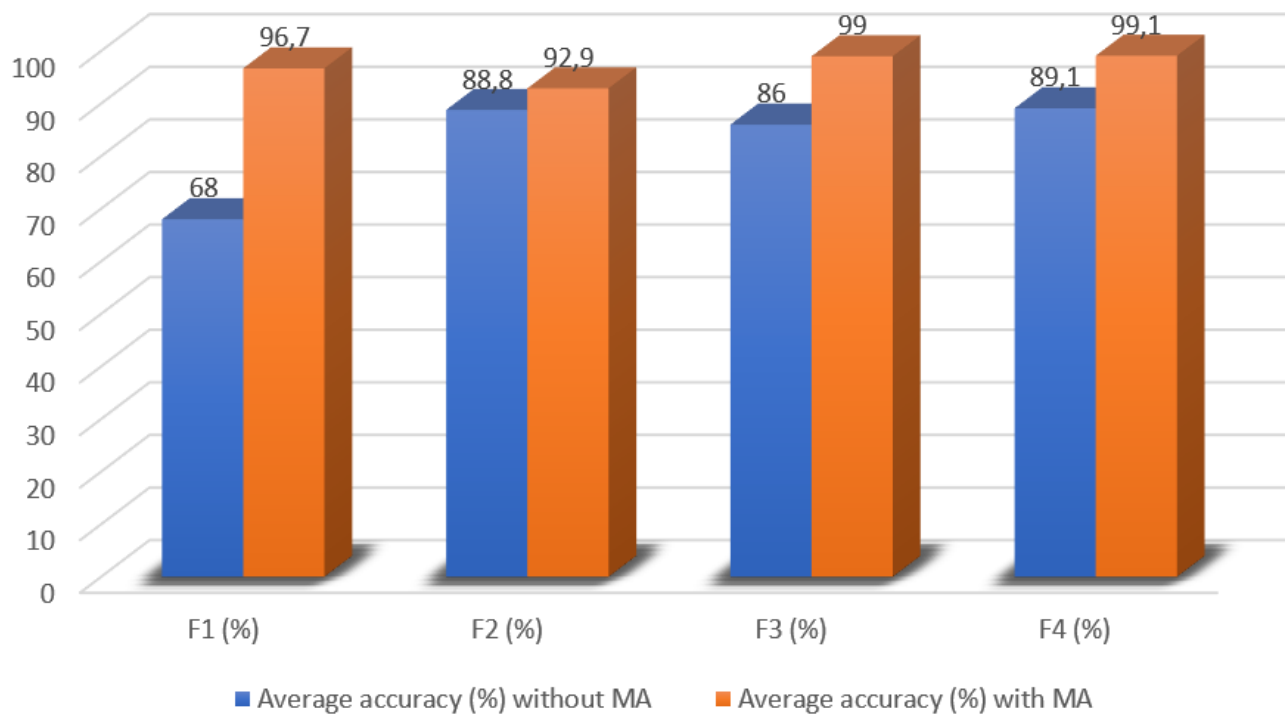


FIGURE 3.9 – Average Accuracy for different features set

seen, the number and type of electrical signatures impact the identification accuracy of household appliances. Considering the average accuracy obtained for different features set, we also observe on the figure 3.9 that the use of the moving average technique improves the precision from 68 % to 96.7 % for F1, from 88.8 % to 92.9 % for F2, from 86 % to 99 % for F3 and from 89.1 % to 99.1 % for F4. All feature sets containing power characteristics exhibit significant improvement when moving average is applied. The reason for this is that the moving average function removes the near zero values in the powerful entities (P and Q) from the dataset. Once the moving average has been applied, the real power and reactive power characteristics involved in F1, F3 and F4 do not overlap and are distinct. According to [Gregory Shakhnarovich and Indyk, 2006], K-NN works very well when different classes do not overlap in feature space. The overlapping nature of features in F2 likely explains the reduced performance of K-NN. To conclude, we can also say that applying a moving average function on input data provides an optimized accuracy for all the features set. The figure 3.9 shows that F1 provides a very good accuracy whereas it requires less electrical signatures (as only P and Q are used). This is interesting since, using less electrical signatures allows reducing the computing time and energy consumed during the learning and inference phases.

Features set	Data preparation technique	Accuracy
F4	Without Moving Average	89.1%
	Dynamical coefficient	90%
	Moving Average	99.1%

TABLE 3.8 – Different data preparation techniques used for appliance identification using ACS-F1 dataset and the K-NN classifier

3.5 Conclusion

In this chapter, we presented a study on the impact of the number and the type of features for home appliance identification. We also proposed to apply a moving average technique as data preparation to optimize classifiers. A comparison was performed with another technique using dynamic coefficients. The ACS-F1 database, freely available to the scientific community for the experiment reproducibility and algorithm comparison, was used. Using this dataset, we provided recognition results using machine learning algorithms based on K-NN, Random forest and Multilayer perceptron classifiers. At best, we obtained a recognition rate of 89.1% and 99.1% using k-NN without and with moving average respectively. We also showed that using only a reduced number of features, in our case only 2 power features (Active and Reactive power) is enough to recognize an appliance in household with high accuracy. Appliances identification is a complex task that remains difficult to generalize given the evolution of these appliances. A mixed approach combining the overall energy consumption as well as some key appliances measurements seems a good compromise. The global energy consumption can be used to build a dataset to predict the

3.5 Conclusion

energy consumed (chapter 4) and thus make decisions to optimize self consumption (chapter 5). The plugs installed on some key and controllable appliances would in other hand allow a manual or even an automatic control of these appliances. In the next chapter, we propose deep neural networks methods to predict the overall energy consumption of a household from a database collected using a non-intrusive approach.

Short and Medium load forecast in smart Home

Sommaire

4.1 Introduction	56
4.2 Related works	57
4.3 Energy load prediction methodology	58
4.3.1 Data set and Data preprocessing	59
4.4 Choice of energy load forecasting model	65
4.4.1 Linear model (ARIMA)	65
4.4.2 Non-Linear Model (LSTM)	66
4.4.3 Performance metrics	68
4.5 Results and discussion	68
4.5.1 LSTM and ARIMA prediction models comparison	68
4.5.2 LSTM prediction model validation on a new dataset	70
4.6 Conclusion	71

4.1 Introduction

In order to improve household energy efficiency, it is essential to have a system able to predict with accuracy its future energy consumption. Predictions are indeed useful to determine the adequate sizing of solar panels and battery for improving the self-consumption, thus reducing power flow in the grid. Predictions can also be used by power management policies as it will be shown in the chapter 5. However, energy load forecast are more and more based on algorithms requiring large scale energy consumption data.

Every day, new smart plugs appear on the market. These smart plugs have made energy consumption data available [Manic et al., 2016], making data statistical modeling possible [Jetcheva et al., 2014]. According to the prediction period, three different classes of energy load

forecasting can be considered : (1) Short-term-load when prediction varies from 1 hour to 1-week, (2) Medium-term load forecasting ranging from 1 week to 1 year, and finally (3) Long-term load when the model makes predictions longer than 1 year [Vahid Mansouri, 2014]. Whatever the category, energy load prediction has been shown to be a complex matter. Moreover, load prediction at individual household level has shown to be even more difficult than global load forecasting [Mocanu et al., 2016]. In this chapter, the aim is to compare the performance of two approaches, a linear ARIMA (AutoRegressive Integrated Moving Average) and a non-linear LSTM (Long Short-Term Memory) predictors for energy load forecast on short and medium term using the IHEPCD (Individual Household Electrical Power Consumption Dataset) dataset [G. Hebrail, 2012]. In [Thokala et al., 2018, Marinescu et al., 2013], the authors predicted the energy load in short interval without preprocessing the original dataset. However, a dataset usually contains some missing values due to system or network connection failures. The lack of this information may drastically reduce the accuracy of the prediction model. In this chapter, we therefore investigate the impact of different data preprocessing techniques to fill missing values. Prediction is then done using either a deep learning algorithm (LSTM) or a stochastic process (ARIMA). Both approaches are explored for performing energy load forecasting and tested on a dataset containing electricity consumption data. Our prediction was first only based on active power data collected within a household during a period of 47 months and a time resolution of one minute. To validate our training model, another data set collected in an office located in a different area and with a 5-second resolution is used.

4.2 Related works

Over the last few years, energy load forecasts have received an increasing attention from researchers. A lot of studies have used forecasting models based on time series dataset such as electrical power consumption. In [Saab et al., 2001], the authors studied a load forecasting model based on a one step ahead forecast for monthly electric energy consumption in Lebanon. Two approaches, ARIMA and AR (1), are used with a high pass filter. The best accuracy is obtained using AR (1) high pass filter. In [Zhu et al., 2012], the authors studied the issue of the sustained growth of household energy consumption in China from 1980 to 2009. ARIMA and BVAR were used as forecasting models and showed that both methodologies are appropriate to predict the sustained growth of HEC (Household Energy Consumption) trends.

Recently, new algorithms based on deep learning have been proposed to cope with the challenges related to the load forecasting models. Despite deep learning is quite a new approach to address energy load prediction problems, this kind of methods has gained popularity among private companies as well as academics over the last few years [Hochreiter and Schmidhuber, 1997]. In [Jetcheva et al., 2014], ANNs(Artificial Neural Networks) ensembles were used to perform energy load forecasting. In [Roldán-Blay et al., 2013] [Sulaiman et al., 2016], the authors have

explored in details ANNs for short, medium and long term periods of load forecasting. In [Ghelardoni et al., 2013], SVM (Support Vector Machines) coupled with empirical mode decomposition were used to perform long term energy load forecasting. In [Fiot and Dinuzzo, 2018], the authors proposed kernel based multi-task learning approaches to predict electricity demand. In [(Kanevce) Dedinec et al., 2016], DBN (Deep Belief Networks) were used to perform short term electricity load forecasting on a Macedonian hourly electricity consumption dataset. In [Thokala et al., 2018, Marinescu et al., 2013], the authors predicted electrical consumption using a forecasting model based on LSTM. However, their predictions were performed without any preprocessing of the raw data, thus including some missing values in the LSTM forecasting model [Kim et al., 2018]. In [Kingma and Ba, 2014][Siami-Namini and Namin, 2018], authors provided a comparison between ARIMA and LSTM for sales forecasting in retail, as well as for financial market price prediction. Their experiments showed that LSTM provides a better accuracy than ARIMA. In [Chujai et al., 2013], the authors studied the most suitable forecasting period for electrical load. Two stochastic approaches, ARMA (Auto-Regression Moving Average) and ARIMA, using a preprocessing technique based on the previous value to fill missing samples were compared. Their results demonstrated that daily and weekly forecasting period were the most suitable for ARMA, while monthly and quarterly periods were better for ARIMA. As energy load exhibits both linear and non-linear patterns, we propose a system which can predict a day ahead power consumption based on LSTM and ARIMA models. The choice of these models was motivated by the following reasons. LSTM is able to identify structures and patterns of data such as non-linearity and complexity in time series. On the other hand, ARIMA is known to perform well on linear time series data and stationary data. Finally, we propose a new technique to fill missing values from the original dataset called Same Time a Day Ago or Next (STDAN).

4.3 Energy load prediction methodology

In this section, short- and medium-term predictions will be considered. To do so, LSTM and ARIMA have been configured so that their outputs are able to predict one (which corresponds to the daily power consumption) or several values at the same time (for example 7 values (days) to represent the daily power consumption during one week). The main objective is to compare the performance of linear and non-linear models (ARIMA and LSTM respectively) for energy load forecasting. The proposed flow for load forecasting is shown in figure 4.1. This figure is a simplified schematic representation of the machine learning process. This process is composed of 3 main phases : Data preparation consists in pre-processing the data to be used by the learning machine models. It gathers different blocks such as data preprocessing, features extraction and dataset splitting. The learning phase generates a prediction model based on the training data. The testing phase consists in evaluating the generated model on the test data using different metrics.

4.3 Energy load prediction methodology

The main issues addressed in this work are two-fold : which algorithm has the best prediction accuracy for time series data ? Which data preprocessing technique to fill missing values is better for energy load forecasting ?

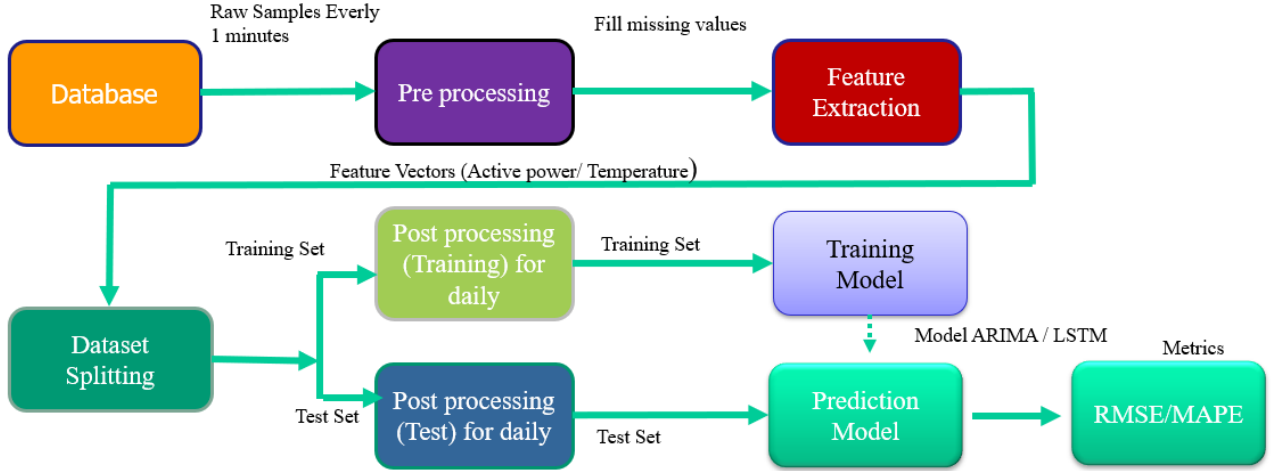


FIGURE 4.1 – Machine Learning process

4.3.1 Data set and Data preprocessing

For collecting information in household, an architecture based on smart sensors (or smart plugs) is required. Through a wireless communication, sensing information are sent to a gateway and can be then visualized on a dashboard. As it is illustrated on figure 3.3, we developed an architecture to collect data in a household using different smart plugs such as Fibaro Plug [Fibaro, 2020], Coco plug [Chacon, 2018], Energy meter gen5 [Gen5, 2018] and Multisensor6 [MultiSensor6, 2018]. This architecture is built around a Raspberry Pi3 computer [Pi, 2018] and a home automation system called Domoticz [Domoticz, 2018]. This web front-end is interesting since various sensor devices can be monitored and configured very easily. To visualize sensed data on a dashboard, we used an open source analytics and monitoring system called Grafana [Grafana, 2018]. This sensing prototype should have been deployed in a house located in the south of France (Nice), to get power measurements for different appliances as well as for the overall consumption. Unfortunately, the time required to get measures was too limited to construct a relevant dataset and apply our forecasting models. We therefore decided to use IHEPCD a freely available dataset of electric power consumption [G. Hebrail, 2012]. This dataset is the result of acquisitions performed within a French household between December 2006 and November 2010 (47 months), using a one minute sampling rate. The dataset contains therefore 2,075,259 measures of a household’s electrical energy consumption.

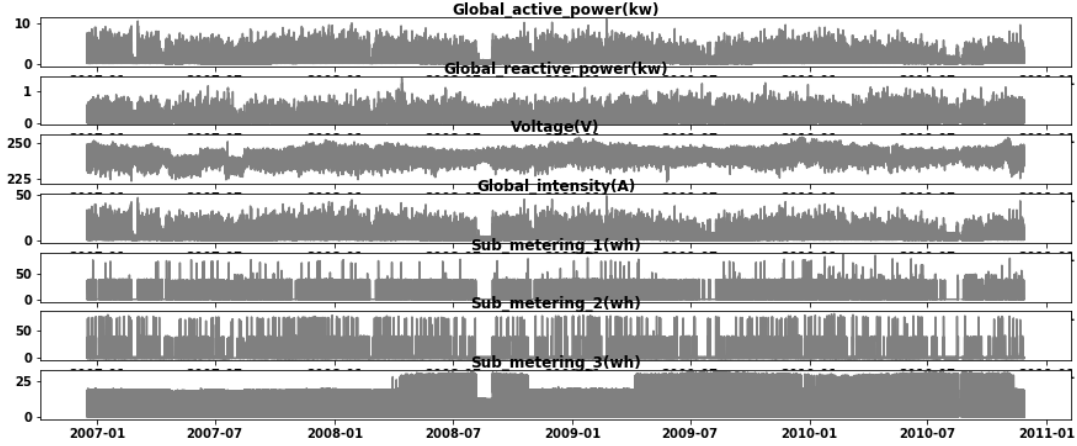


FIGURE 4.2 – Curves of each features in the IHEPCD

As illustrated in figure 4.2, this data set contains the following 7 household attributes (each attribute having a timestamp) :

- Global active power (kW) ;
- Global reactive power(kW) ;
- Global current (A) ;
- Energy sub-metering 1 (Wh) ;
- Energy sub-metering 2 (Wh) ;
- Energy sub-metering 3 (Wh) ;

Power, voltage and current measures are averaged over one minute. Energies sub-metering's have been used in this house to detect groups of appliances that consume more energy than others.

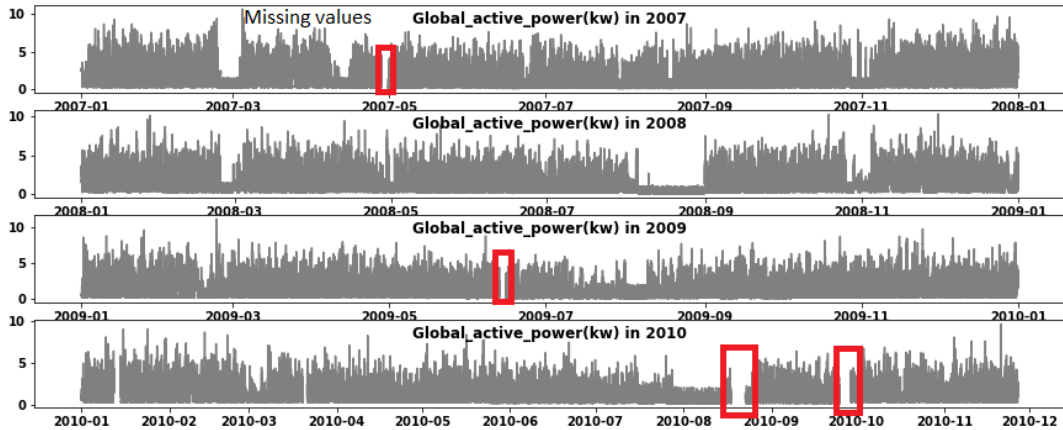


FIGURE 4.3 – Curves of global active power for each years in the IHEPCD

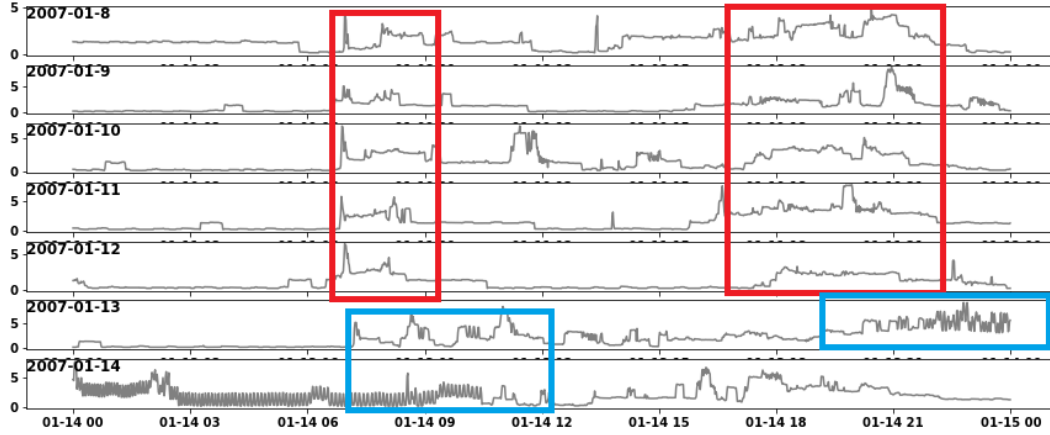


FIGURE 4.4 – Curves of global active power for week of january the IHEPCD

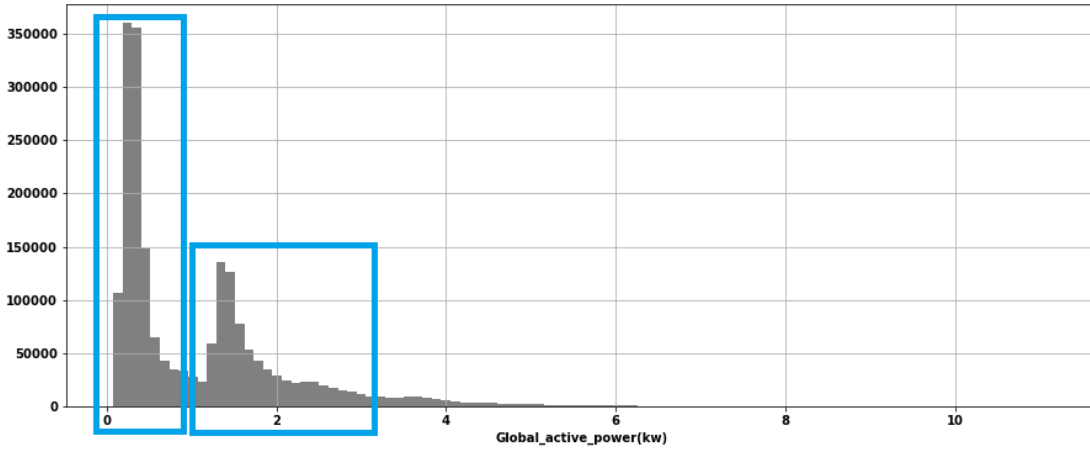


FIGURE 4.5 – Histograms of global active power of the IHEPCD

As mentioned in [G. Hebrail, 2012], all appliances in this household are considered to be supplied by the electrical network. In this dataset, appliances include a dishwasher, an oven, a microwave (hot plates are not electric but gas power), a washing machine, a tumble dryer, a refrigerator, a light, a water heater and an air conditioner. Some forecasting studies have been already performed in the past using this dataset. In [Amarasinghe et al., 2017], the authors compared CNN, ANN, SVM and LSTM models using power consumptions samples with one minute sampling rate.

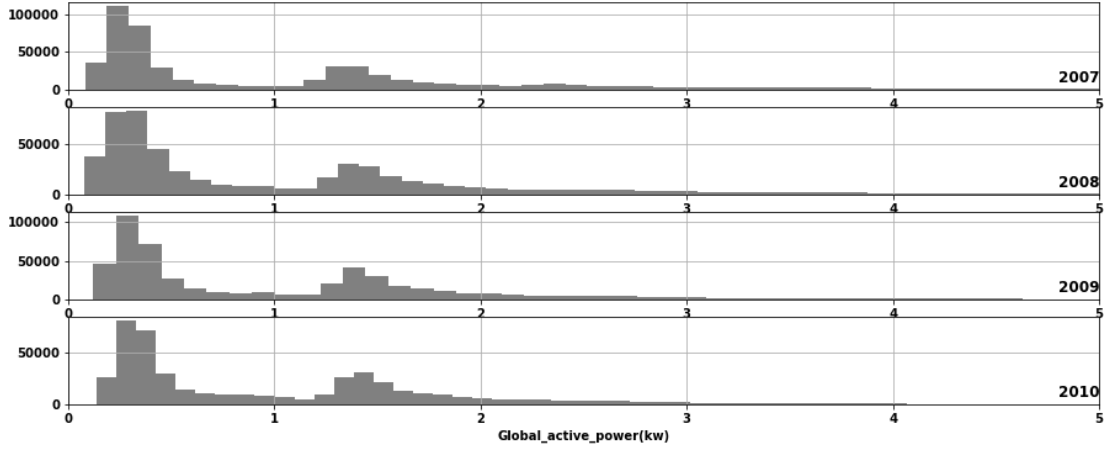


FIGURE 4.6 – Histograms of global active power of the IHEPCD for each years

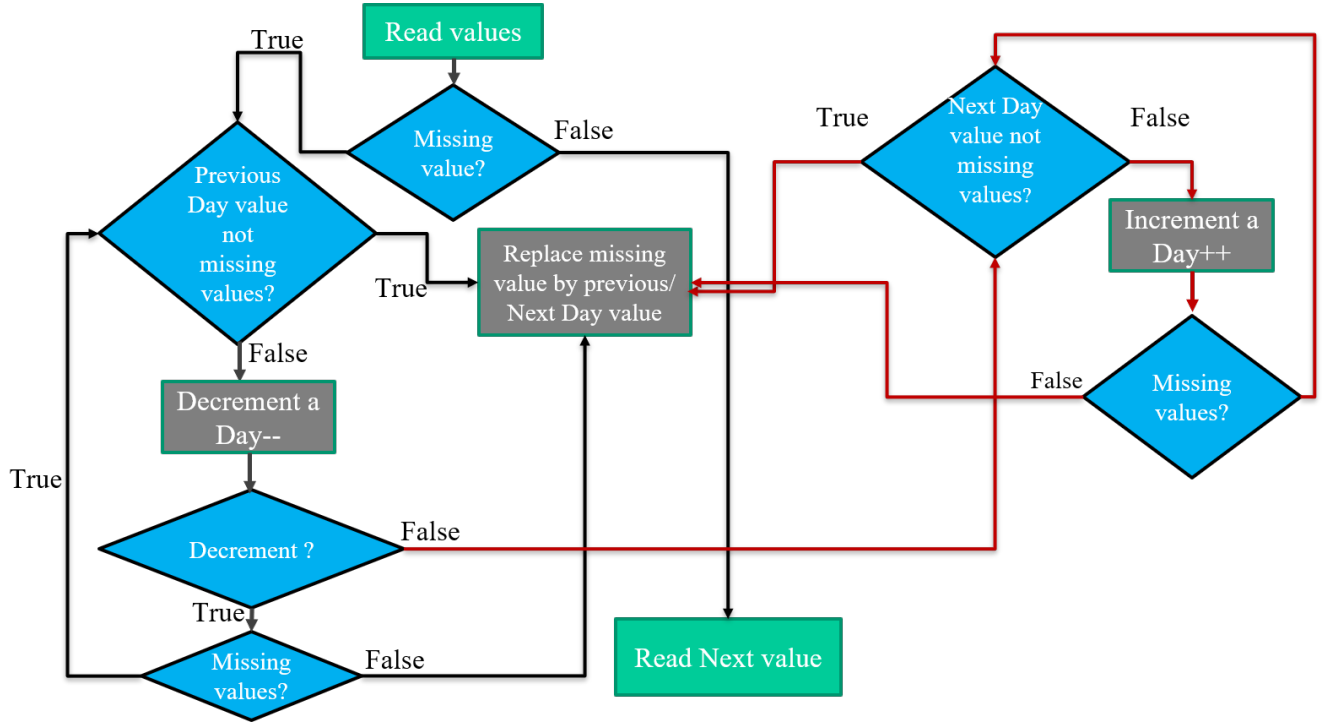


FIGURE 4.7 – Algorithm proposal STDAN

Their experiments showed that LSTM provides a better accuracy than others. It is worth noticing that, we first only consider active power as feature set. As it is illustrated in the figure 4.5, the active power distribution appears to be bimodal, i. e. it appears to have two groups of observation means. We can further explore this behaviour by looking at the distribution of active energy consumption over the four years of data (figure 4.6). As we can see on figure 4.6, the distribution of active energy consumption over these years seems very similar. The distribution is indeed bimodal, with a peak around 0.5 KW and possibly another around 1.5 KW. There is

4.3 Energy load prediction methodology

a long tail in the distribution towards higher kilowatt values which can correspond to a strong electrical activity of the occupants (use of electrical appliances which consume more). For this we can take a closer look at the power consumption at the daily level during a week for the month of January. As can be seen in figure 4.4, the high consumption is mainly around 6-7 a.m. and 5-10 p.m. (probably working days), which may indicate the presence of occupants in the house. But we also notice some time slots in the middle of the day when consumption drops. These slots should corresponds to hours of the day when occupants are at work. In this figure, we also see that the electricity consumption on January 13 and 14 differs from the other days, and should correspond to the weekend. These kind of information can be useful in developing a predictive model.

Unfortunately, as it can be seen in the figure 4.3, this dataset contains around 1.25% missing values (indicated with red rectangles and represented with a '?' Character in the input CSV file) and is therefore not usable as this for a prediction model. The lack of these information, (or simply replacing missing values by zeros), may indeed reduce the predictive efficiency of the forecasting model as shown on figure 4.11. To cope with this issue, we investigated three techniques to replace missing values using 1) the previous sample, 2) the mean calculated over all the samples, or 3) the sample at the same time a day ago or next as illustrated by figure 4.8. Both previous and mean values are already implemented in panda's python library. The proposed STDAN (Same Time a Day Ago or Next) algorithm, dedicated for time series data is depicted on figure 4.7. As shown, STDAN consists in detecting a NaN (not a number) and then replacing it by the value at the same time a day ago, or the next day at the same time. When the returned value is still NaN, this mechanism is repeated iteratively until it finds a finite value as described in figure 4.7. To assess the benefit of this algorithm, we randomly generated 5% to 90% missing values from the original dataset. As we are interested in predicting a daily total power consumption, we aggregate the minute-by-minute dataset into daily observations by using 4.1.

$$P_{daily} = \sum_{i=1}^{N=24 \times 60} P_i \quad (4.1)$$

Where P_{daily} presents the daily total power consumed, P_i the instantaneous power consumed every minute and N the number of samples per day. We got therefore a new dataset with 1442 samples. Before applying the prediction models, let us have a look to the electrical activity patterns in the daily load profiles. In figure 4.9, all daily load profiles corresponding to 1442 days were plotted together. As we can see, two behaviors of high and low power consumption can be visualized. These behaviors correspond to the darker regions (where more curves are concentrated). The high consumption behavior is mostly found during early hours (5-7 a.m.) and in the evening (5-9 p.m.), while the hours when occupants are not at home (8 a.m. - 5 p.m.) on working days or are on vacations show a decrease in consumption. In the next section, we describe the different load forecasting models used in this work. The benefit of the STDAN algorithm will be demonstrated in the section 4.5.

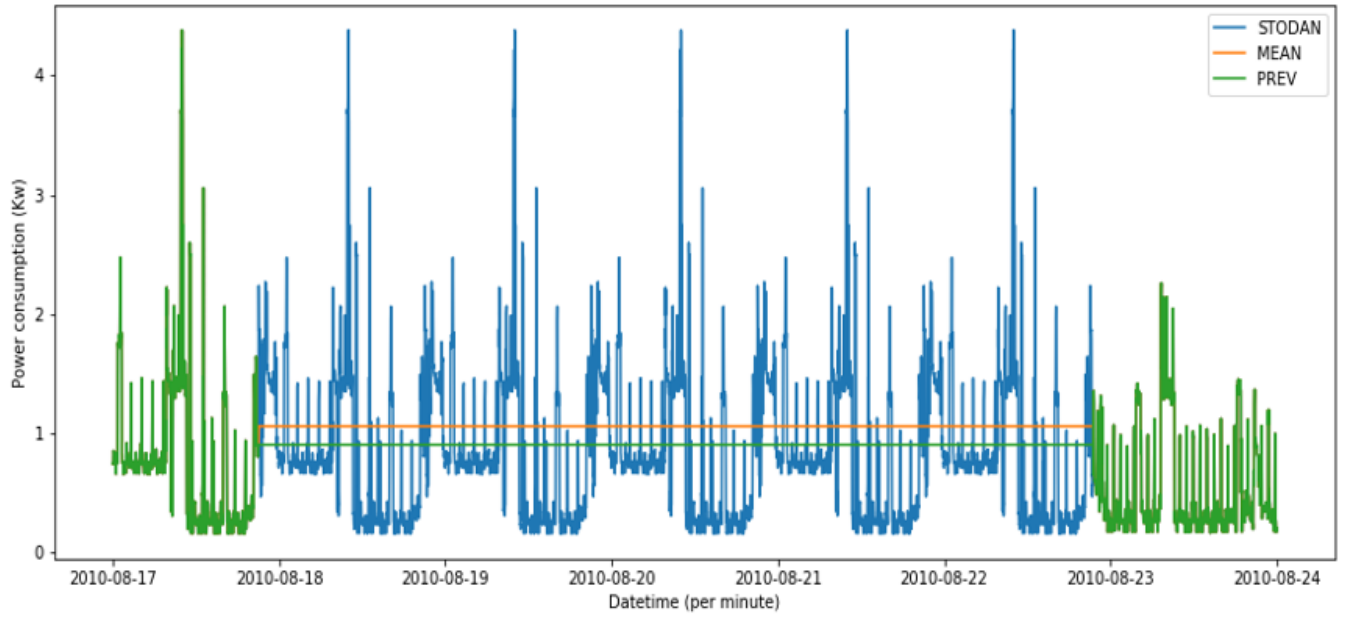


FIGURE 4.8 – Representation of data after filling missing values based on different data preprocessing

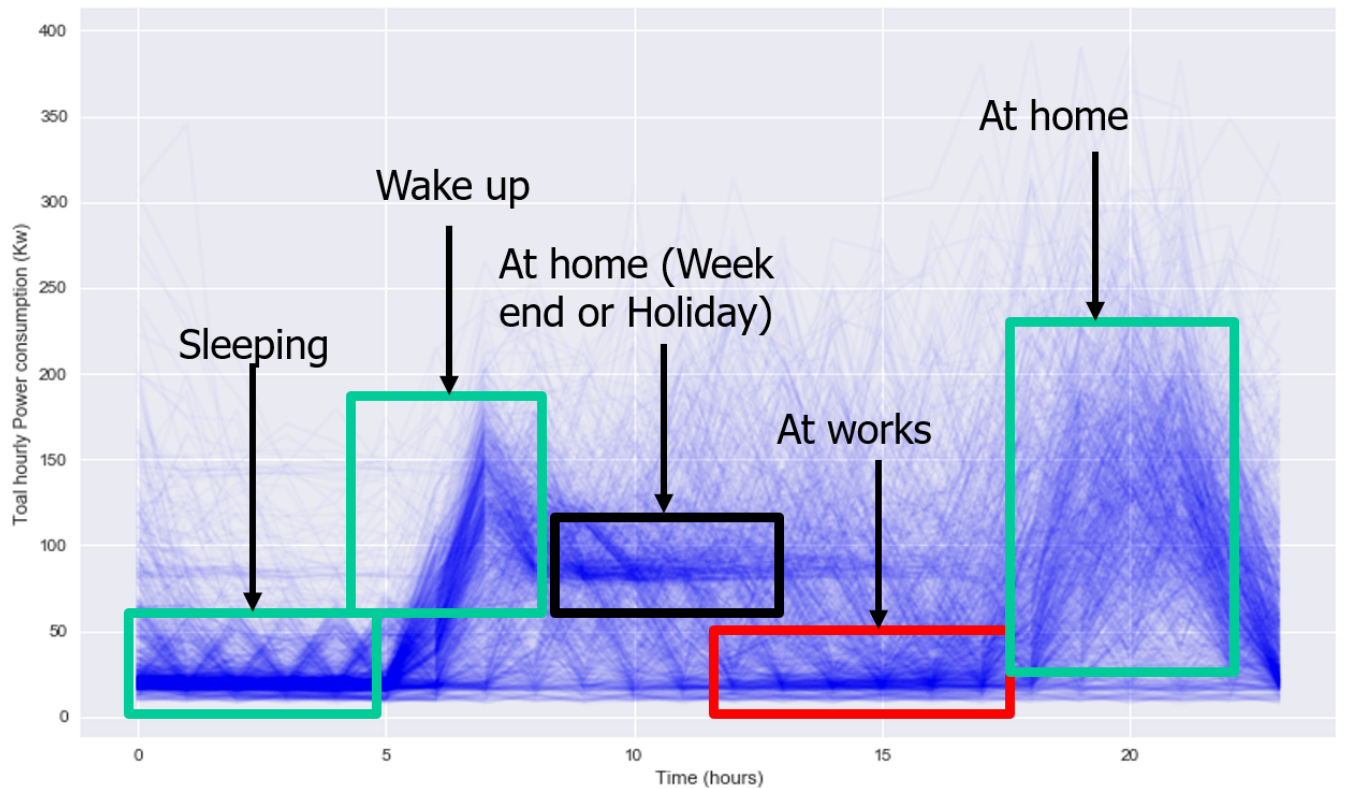


FIGURE 4.9 – Representations of all the daily load profiles in IHEPCD

4.4 Choice of energy load forecasting model

4.4.1 Linear model (ARIMA)

ARIMA is a time series prediction model initially proposed by Box and Jenkins [Box, 1979] to address the limitations related to auto-regressive moving average (ARMA). ARMA is only used for so-called stationary time series. A time series is stationary if the mean and variance are almost constant over the time. ARIMA is a generalization of the ARMA model to which a differencing process to make time series stationary is added. This differencing process is based on the difference between consecutive observations. An ARIMA model captures the following key elements :

- AR : Autoregression. It is a regression model that exploits the dependencies between an observation and a number of lagged observations (p).
- I : Integrated. This parameter is used to make the time series stationary. To do so, the differences of observations are measured at different times (d).
- MA : Moving Average. It is an approach able to consider the dependency between observed samples and the residual error terms when a moving average model is used with a number of lagged observations (q).

An AR model can be written as a linear regression given by [Kingma and Ba, 2014] :

$$X_t = b + \sum_{i=1}^p \phi_i x_{t-i} + \varepsilon_t \quad (4.2)$$

Where x_t is the stationary variable at time t, b is a constant, and ϕ_i are auto-correlation coefficients to be estimated from lags 1 to p. Finally, ε_t are the residuals.

An MA model of order q, i.e MA(q), can be written in the form :

$$X_t = \mu + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (4.3)$$

Where μ is the expected of x_t (usually assumed to be equal to zero), and θ_i coefficients to be estimated (with $\theta_0 = 1$) . We assume that ε_t is a Gaussian white noise series with mean zero and variance σ^2_ε . The ARIMA model of order (p,d,q) can be obtained by combining AR et MA models.

$$X_t = c + \sum_{i=1}^p \phi_i x_{t-i} + \varepsilon_t + \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (4.4)$$

In order to explore different combinations of parameters p (from 0 to 10), d (from 0 to 2) and q (from 0 to 5), we used a grid search method (an exhaustive search algorithm with a sub-set specified manually for hyperparameter optimization). For each combination of parameters, we evaluate the ARIMA model on a dataset subdivided into training set for the learning phase and the test set to validate the ARIMA model using the mean squared error as metric. The first 3 years of the dataset were considered as the training set while the last year (i.e. 2010) as the test

set. Upon exiting each combination, one should keep track of the lowest observed error score and the configuration that caused it. From this grid search method, the parameters having the lowest mean squared error score were identified as : $p = 1$, $d = 0$, and $q = 2$. Results obtained with this configuration are shown on figure 4.11.

4.4.2 Non-Linear Model (LSTM)

LSTM models are considered as a solution to overcome the problem of time series prediction due to their capacity of remembering patterns for short and long-term periods of time [Hochreiter and Schmidhuber, 1997]. A recurrent LSTM network is composed of different memory blocks called cells. A LSTM model can modify, remove or add information as it crosses different layers as shown in figure 4.10. Information flows through a mechanism known as cell states which allows memorizing or forgetting things in a selective way. At a particular cell state, the information has three different dependencies, called respectively the previous cell state c_{t-1} , the previous hidden state h_{t-1} , and the input at the current time step x_t . Three gates called forget, input and output, are responsible of saving informations and manipulations of a cell. To optimize the performance of the LSTM network, a forget gate (f_t) defined in Eq. 4.5 decides which information needs to be thrown away using a sigmoid layer by looking at h_{t-1} and x_t , and outputs a number between zero and one for each number in the cell state c_{t-1} . A value of 1 means “completely keep this”, while a 0 indicates “completely get rid of this”.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4.5)$$

The memory gate $\tilde{c}_t$ chooses which new information needs to be stored in the cell state. First, a sigmoid layer called the “input gate layer” chooses which values will be updated. Then, a tanh layer creates a vector of new candidate values, $\tilde{c}_t$, that could be added to the state. Then, it combines these two to create c_t an update of the state.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (4.6)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4.7)$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t \quad (4.8)$$

The output gate O_t decides what will be yield out of each cell. The obtained values are based on the cell state along with the filtered and newly added data. Then, the cell state is filtered through

4.4 Choice of energy load forecasting model

$\tanh$ (to get the values between -1 and 1) and multiply it by the output of the sigmoid gate, so that only the chosen parts are sent to the output layer.

$$O_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \quad (4.9)$$

$$h_t = O_t * \tanh(c_t) \quad (4.10)$$

Where W_f , W_c , W_i , and W_o represent rectangular weight matrices, b_f , b_i , b_c and b_o are bias vectors and σ is the logistic sigmoid.

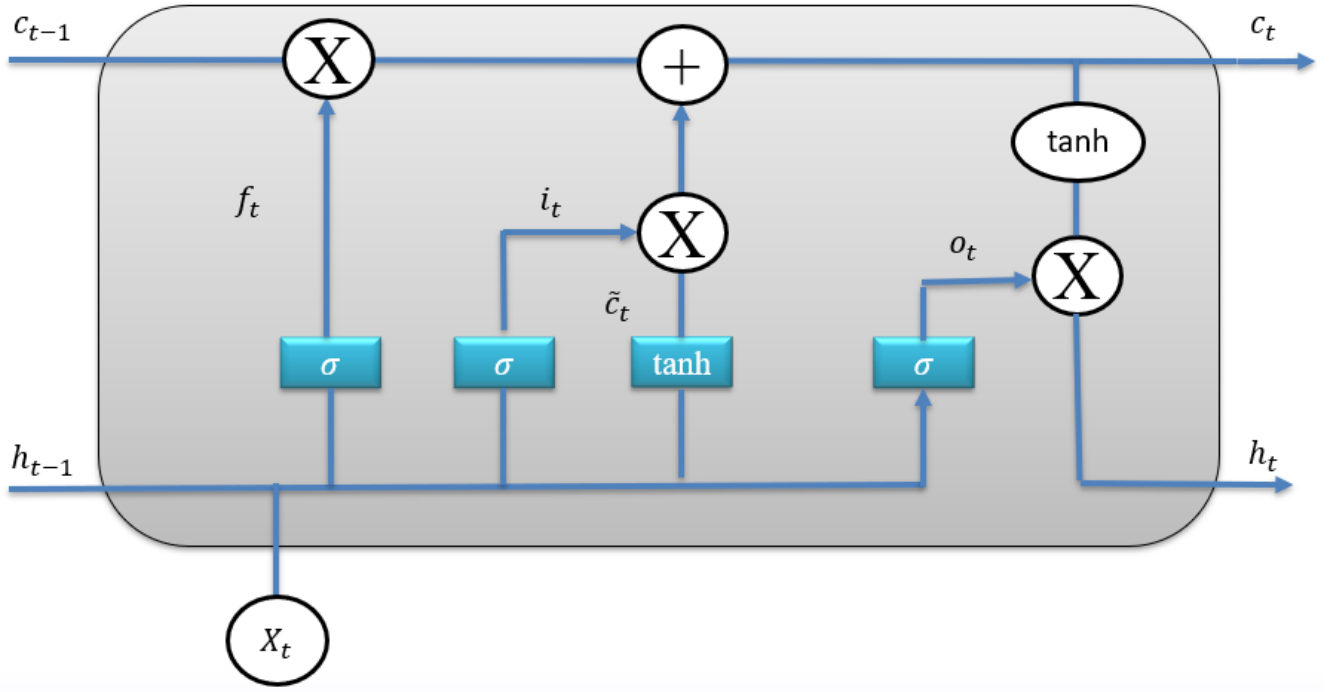


FIGURE 4.10 – The internal structure of an LSTM

A standard LSTM has been designed for forecasting household energy consumption. The network structure consists of 2 LSTM hidden layers with 50 memory units, one fully (dense) connected layer of 25 neurons with a linear activation function and an output layer that makes a single value prediction. The choice of the network structure has been performed following a trial and error approach. The network also uses the Mean Squared Error as a loss function and the ADAM [Kingma and Ba, 2014] algorithm as an optimizer. Networks were fitted with 70 epochs and a batch size of 16.

4.4.3 Performance metrics

To evaluate the performance of LSTM and ARIMA models, two different evaluation metrics were used : Root Mean Square Error (RMSE) and Mean Absolute Percentage Errors (MAPE) . RMSE and MAPE are defined as follows :

$$RMSE = \sqrt{\frac{1}{N} \sum_{j=1}^N (y_j - \hat{y}_j)^2} \quad (4.11)$$

$$MAPE = \frac{100\%}{N} \sum_{j=1}^N \left| \frac{y_j - \hat{y}_j}{y_j} \right| \quad (4.12)$$

Where N is the total number of observations, y_j is the actual value ; and $\hat{y}_j$ is the predicted value. The RMSE metric is used to measure, over a forecasting period, the global error between the actual energy consumed and the corresponding energy estimated by forecasting models. It is a quadratic scoring rule that also measures the average magnitude of the error. According to [Swamidass, 2000], the Mean Absolute Percentage Error (MAPE) is the mean or average of the absolute percentage errors of the forecast. The error is defined as the actual or observed value minus the forecast value. The percentage errors are summed regardless of sign to calculate MAPE. This measure is easy to understand because it provides the error in terms of percentages. In addition, because absolute percentage errors are used, the problem of positive and negative errors cancelling each other out is avoided. As a result, MAPE has managerial appeal and is a commonly used measure in forecasting. The smaller the MAPE, the better the forecast.

4.5 Results and discussion

4.5.1 LSTM and ARIMA prediction models comparison

To analyze the performance of ARIMA and LSTM models, experiments have been conducted using the IHEPCD dataset [G. Hebrail, 2012]. The data initially sampled every minute are grouped to obtain daily total power consumption and then split into two groups. The first group is considered as the training set : it contains data from 2006-12-16 to 2009-12-31 and represents 78% of the dataset. The second group is used for test and contains data from 2010-01-01 to 2010-11-26, representing 22% of the dataset. ARIMA and LSTM models were implemented using the python ecosystem. Statsmodel library was used to fit the ARIMA model by calling the ARIMA function along with the p, d and q parameters. Then *fit* and *predict* functions were called to train the model and make predictions respectively. SciPy environment with Keras deep learning library using the TensorFlow backend was used for the LSTM model.

Our objective is to perform a one-day ahead forecast of the consumed energy using ARIMA and LSTM models. The RMSE and MAPE metrics are used to evaluate both models using the three preprocessing techniques presented in section 4.3.1. As shown on figure 4.11, LSTM provides a significant improvement in accuracy (RMSE) compared to ARIMA model whatever the preprocessing technique. When STDAN is used as a preprocessing technique, the RMSE decreases from 7.844 to 3.991 kWh when LSTM is used. For LSTM and ARIMA models, better performance are obtained when data are preprocessed using our STDAN approach. To further evaluate STDAN, 5% to 90% missing values were randomly generated and then inserted into the test set. As it can be seen in figure 4.12, and as expected, the daily average error increases with the number of missing values for both preprocessing techniques. However, STDAN provides overall a better accuracy than the mean preprocessing technique.

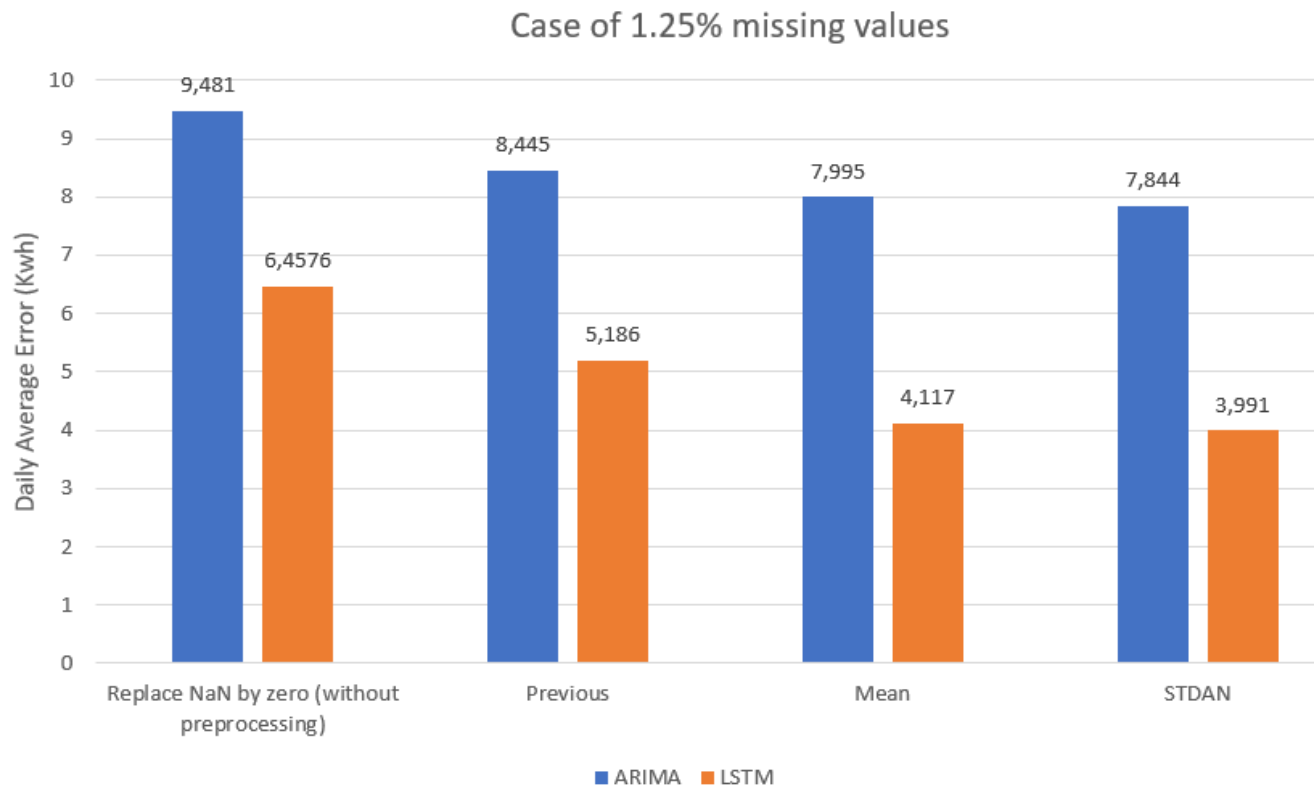


FIGURE 4.11 – Comparison of LSTM and ARIMA prediction models for different preprocessing techniques

Using the STDAN preprocessing technique, we then tested short (7 days) and medium (14 to 31 days) term load forecasting with the LSTM model. Obtained results are shown in figure 4.13. As it can be seen, LSTM performs better with short load term forecasting than with medium term when using STDAN. This can be explained by the k-step ahead out samples forecasts which accumulate the error terms, thus resulting in lower accuracy in medium-term forecasting performance. Overall, we can conclude that LSTM performs better than ARIMA for energy load forecasting. Results

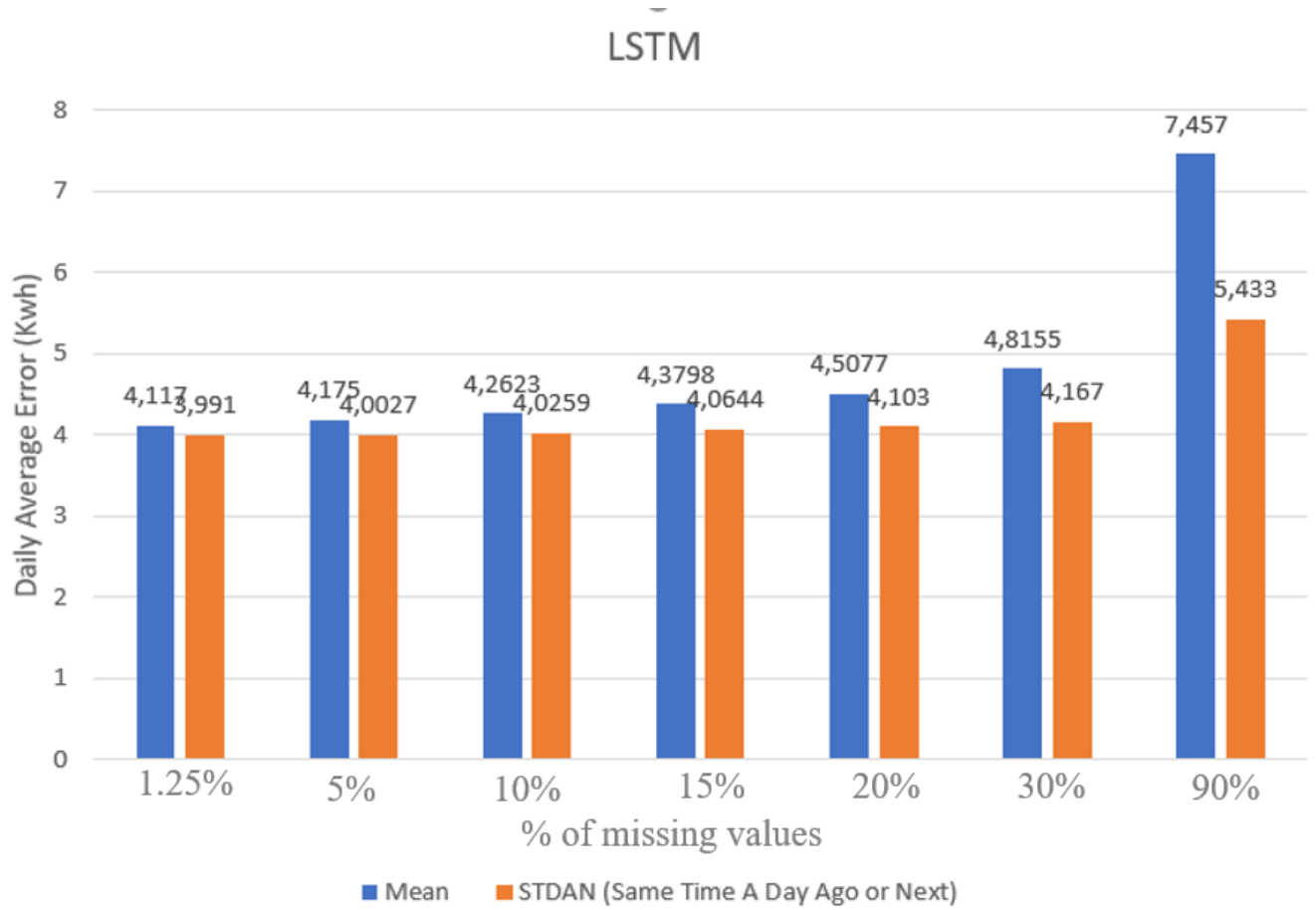


FIGURE 4.12 – Evaluation of STDAN and Mean techniques accuracy for different missing values percentage

also demonstrate that the STDAN preprocessing technique allows reducing the prediction error compared to existing techniques (previous and mean).

Algorithm	Data set	Prediction period	RMSE (KWh)	MAPE (%)
LSTM Univariate	IHEPCD (Paris)	Test set =	3.99	13.29
	New dataset (Nice)	322 days	5.99	20.94

FIGURE 4.14 – Validation of new data set

4.5.2 LSTM prediction model validation on a new dataset

In order to validate the proposed LSTM prediction model trained on the IHEPCD dataset, where samples were collected in an area close to Paris, we also performed experiments using another dataset collected from a building close to Nice in the south of France. This new dataset contains only active power measurements with a 5-second sampling rate. To get daily observations, samples are

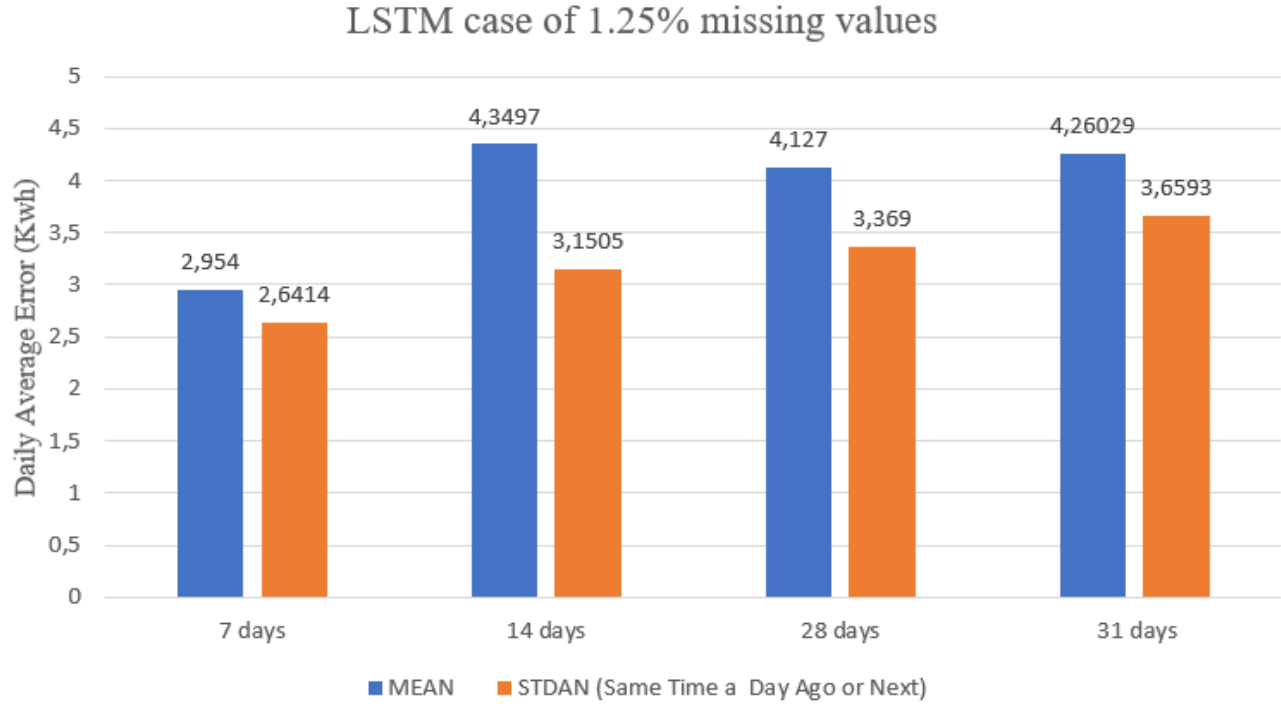


FIGURE 4.13 – LSTM for short and medium term load forecasting

aggregated using the Eq. (1) with $N=17280$ (i.e. $12 \times 60 \times 24$). As this new dataset contained missing values as well, all of them were replaced using our STDAN approach. As it can be observed in figure 4.15, both datasets present different energy consumption behaviours. It seems therefore that activities of both datasets are different. In Paris (i.e IHEPCD dataset), data have been collected in a place of residence and the energy consumption fluctuates over time. On the other hand, data collected close to nice have much less fluctuations and seem thus not to be a place of residence.

The figure 4.14 presents results obtained using LSTM prediction model for both dataset over 322 days. As it can be seen, the prediction error slightly increases with the new dataset. However, this increase remains acceptable and therefore validate our LSTM prediction model.

4.6 Conclusion

In this chapter, we first presented different techniques to replace missing values. Then, we evaluated the accuracy of ARIMA and LSTM, as representative models when forecasting time series data. Both models were implemented and trained on a set of residential power consumption data. The results showed that LSTM provides better performance than ARIMA whatever the preprocessing technique. To validate the effectiveness of the LSTM model, predictions were conducted using a new data test from a building located far away from the one used during the learning phase. Using this new dataset, the prediction error is around 20%, which can be considered as acceptable for a power management to take decision or for sizing an energy harvesting system.

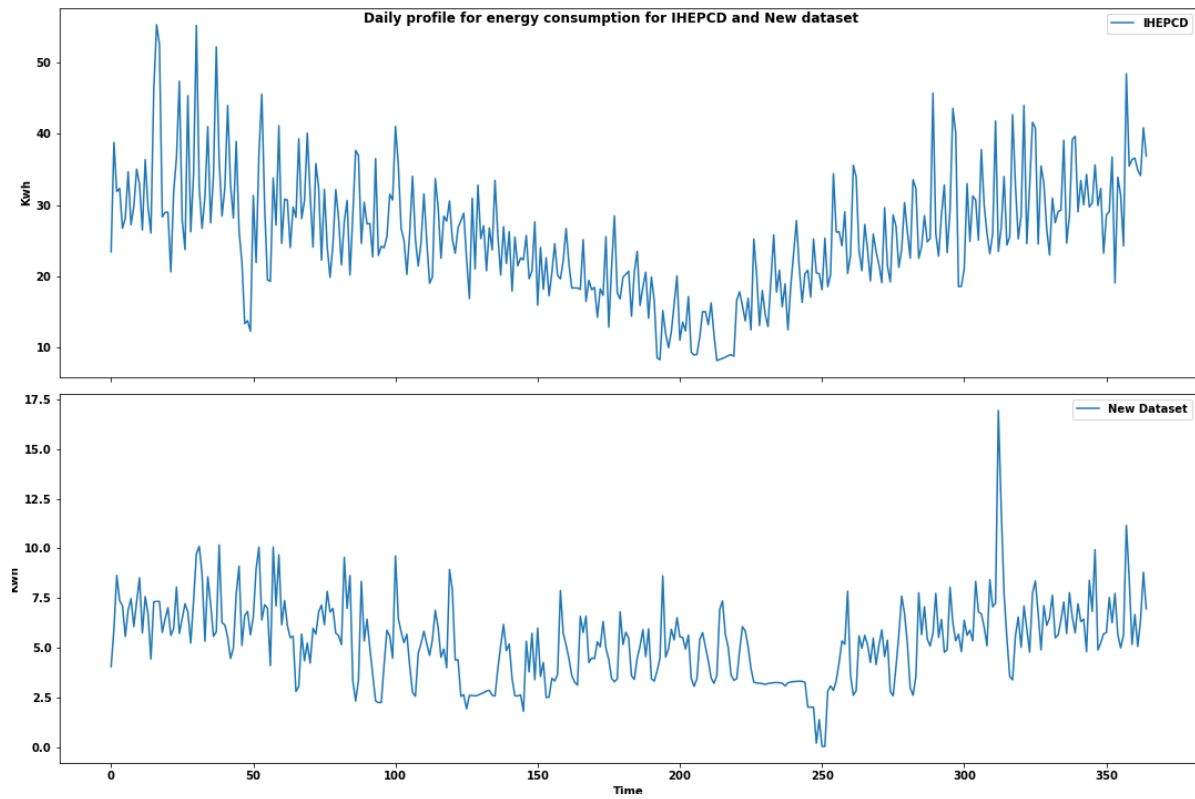


FIGURE 4.15 – Electric Power consumer for an office located at Nice and IHEPCD

Coupling the grid and an energy harvesting system for smart home self-consumption

Sommaire

5.1 Introduction	73
5.2 Related works	74
5.3 Grid-connected PV system	76
5.3.1 Electric generation profile	76
5.3.2 Electric load profile	79
5.3.3 Power management (PM) policies for Grid connected PV System	80
5.4 Results and Discussion	92
5.4.1 Simulation setup	92
5.4.2 Household located in south of France (Nice)	92
5.4.3 Household located in North of France (Paris)	105
5.4.4 Impact of PV and battery size on blackout Error rate	114
5.5 Conclusion	116

5.1 Introduction

Renewable energy technologies could play an important role in addressing societal challenges such as resource depletion and climate change. Among renewable resources such as wind, thermal, solar, it is proven that photovoltaic energy has a major potential for electricity production. However, two major obstacles have limited the widespread use of photovoltaics : first, their production depends on the level of sunshine during the day (local weather conditions) and secondly, they also exhibit strong fluctuations during the year. Initially, the main objective of photovoltaic

systems was to feed the generated output into the electricity grid, which was remunerated by feed-in tariffs. In France, the purchase price of photovoltaic cells (PV) varies according to the type of installation and its power [Gabriele, 2020]. However, during the 2nd quarter of 2020, the feed-in tariff for surplus PV has been set at 10 cents per kWh for a solar installation less than or equal to 9 kWp and is then relatively low compared to the price of electricity. It is indeed 0.06 euro per kWh injected by an installation with a power between 9.1 kWp and 100 kWp. As a result, it becomes more interesting to use the electricity produced by the solar panels on site at household level than to inject it into the grid. Nevertheless, in a household the simultaneity of electricity generated by the PV and consumed by the load remains limited. In [Cagigal et al., 2011], the authors proposed to shift the consumption of controllable loads to periods of energy surplus in order to increase self-consumption. Another way to increase self-consumption consists in combining the PV system with a battery. With such a system, the surplus energy can be stored and used later to reduce the energy drawn from the grid, thus increasing self-sufficiency. In this chapter, we will study through simulations and using prediction models presented in previous chapters, the energy efficiency of such a system illustrated in figure 5.1. To do so, two power management strategies based on predictions of the consumed and the harvested energy, will be in charge to select either the grid or the PV system as energy source for the household. The impact of the amount of harvested energy will also be evaluated through two different household localization.

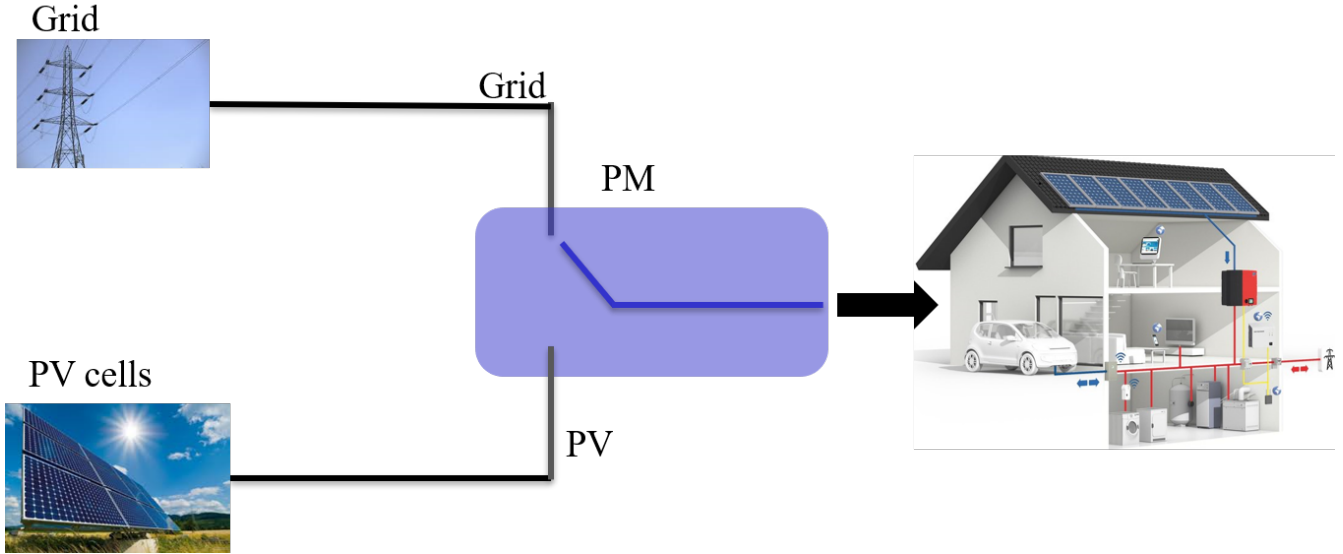


FIGURE 5.1 – House model for Grid-connected PV System

5.2 Related works

Many approaches and techniques have been used for the control and management of hybrid renewable systems and electrical storage systems. In [Hongbo et al., 2016], the authors developed an optimization model in a residence equipped with a hybrid system (using PV, fuel cell and

battery). By guaranteeing reliable operation of the system, the proposed model made it possible to determine the optimal operating strategies with the objective of minimizing the annual operating cost or annual CO₂ emissions. In [Sharma et al., 2016], a mathematical model for the optimal energy management of a residential building has been proposed. Their study determines the optimal decisions in real time taking into account models with realistic parameters and customer preferences. The results obtained showed that the proposed model can reduce the cost and energy consumption of customers by approximately 17% and 8%, respectively. However, these studies do not take into account the operation and real characteristics of the various devices (production, storage and conversion of energy) and their impact on losses. In [Song et al., 2011], a residential 10 kWh lithium-polymer battery energy storage system was implemented and simulated. A control algorithm has been proposed for the energy storage of batteries connected to the grid with photovoltaic production. However, their objective was to charge the battery overnight with energy coming from the grid and not to maximize self-consumption from local production. In [Ren et al., 2016], the impact of PV battery systems on peak demand and energy consumption is assessed in two existing studies and nine potential electricity tariffs, considering old and new homes. The study showed a potential for reducing peak demand by up to 50%, through the installation of 5.5 kW photovoltaic panels and 16 kWh batteries. With the adoption of PV battery systems, the biggest bill savings have been achieved at the household level with the critical peak tariff of retail energy pricing and grid capacity load. The main objective of this work was to ensure that consumption peaks were capped and to increase profitability and not to maximize self-consumption. In [Luthander et al., 2016], the self-consumption of residential photovoltaic energy in a community of several single-family houses was evaluated taking into account the reduction of photovoltaic energy and energy storage by individual or shared batteries. The results indicate that the self-consumption ratio increases when using shared storage instead of individual storage. In [Ranaweera and Midtgård, 2016] an energy management for a PV system coupled with battery energy storage is proposed. Such a system maximizes daily economic benefits while reducing energy injection into the grid to alleviate surge problems caused by reverse energy flow. The main objective of this work was to provide services to the electricity network and to ensure the quality of energy. In [Santos et al., 2014], an assessment of the impact of grid energy balancing obtained with or without energy storage is presented. A storage management strategy is proposed which coordinates the surplus photovoltaic production with the surplus demand over a daily period, and reduces their maximum peaks. However, the economic evaluation of these studies focuses on the perspective of network and not on the user.

In this thesis, power management algorithms are proposed in order to maximize self-consumption as well as the optimization of the blackout error rate. Our contribution consists in integrating a power management system in a residence equipped with a Grid connected PV system. The proposed algorithms take decisions periodically, every slot of time, Decisions are

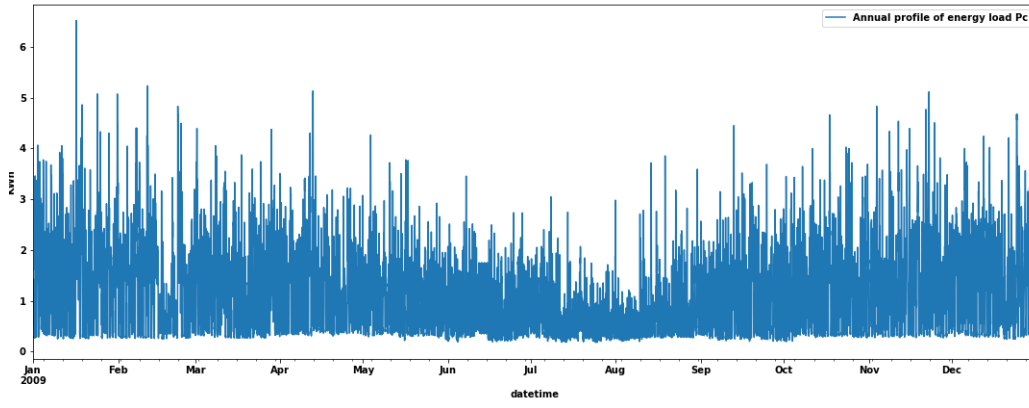
based on predictions and metrics are updated according to measures performed over the last time slot.

5.3 Grid-connected PV system

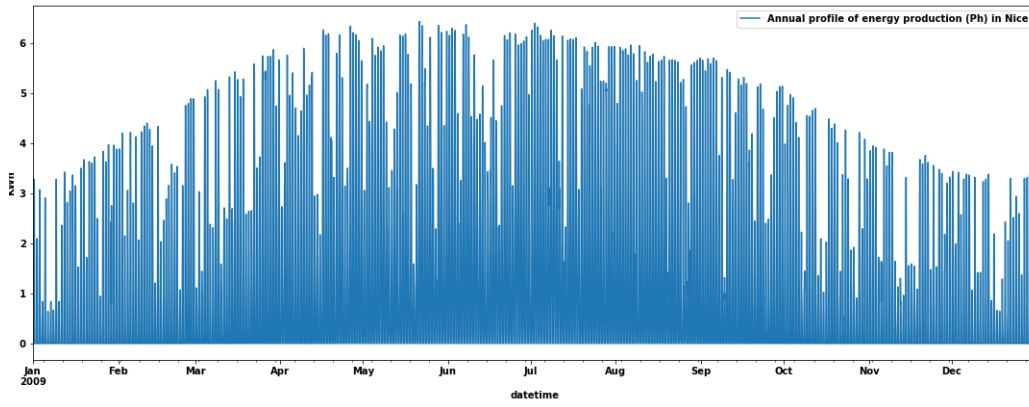
5.3.1 Electric generation profile

The household's energy production profile is simulated from an open source tool developed by Sam Borgeson available on GitHub [Borgeson, 2018]. The author has implemented a more simplified tool that provides a pythonic wrapper around the NREL's (National Renewable Energy Laboratory) SAM (System Advisor Model) distributed energy system simulation engine [NREL, 2020]. SAM is a simulation system for distributed energy projects, with the ability to configure, simulate PV systems and other renewable systems. It takes several input parameters such as PV size, module type, weather year (TMY3 (Typical meteorological year) files) of the desired location as well as other parameters such as azimuth and tilt angles. Due to the lack of information on the energy load profile of the household and to simplify our analysis, we consider a fixed azimuth and tilt angle for both locations (Nice and Paris). The rest of the parameters have been configured with default values shown in table 5.1. In practice, we know that the azimuth and the tilt angles can change from one household to another. The azimuth angle may change depending on the orientation of the roof and the tilt angle according to the shading. The optimal tilt angle ranges from 32 to 38 [HULD Thomas, 2008]. The generated DC power based on the input parameters provided in table 5.1 will be considered in the rest of this chapter. According to [Dobos, 2014], the "standard" option of module type represents typical poly-crystalline or mono-crystalline silicon modules with efficiency in the range of 14-17%. The "premium" option is suitable for modeling high efficiency mono-crystalline silicon modules (18-20%) that have anti-reflective coatings and lower temperature coefficients. The "thin-film" option assumes low efficiency (11%) and a significantly lower temperature coefficient, which is representative of most thin-film modules installed since 2013.

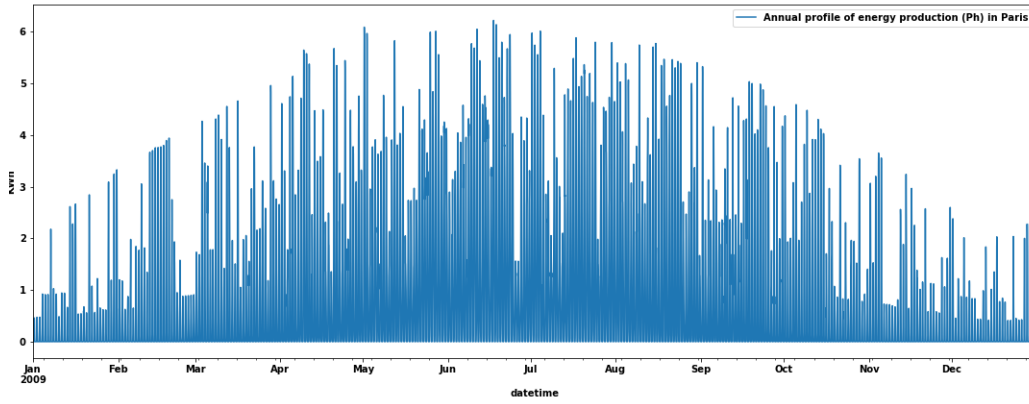
5.3 Grid-connected PV system



(a) P_C



(b) P_H in Nice



(c) P_H in Paris

FIGURE 5.2 – Energy production P_H in Nice and Paris (for 8kWp PV system) VS. annual profile of energy load (P_C)

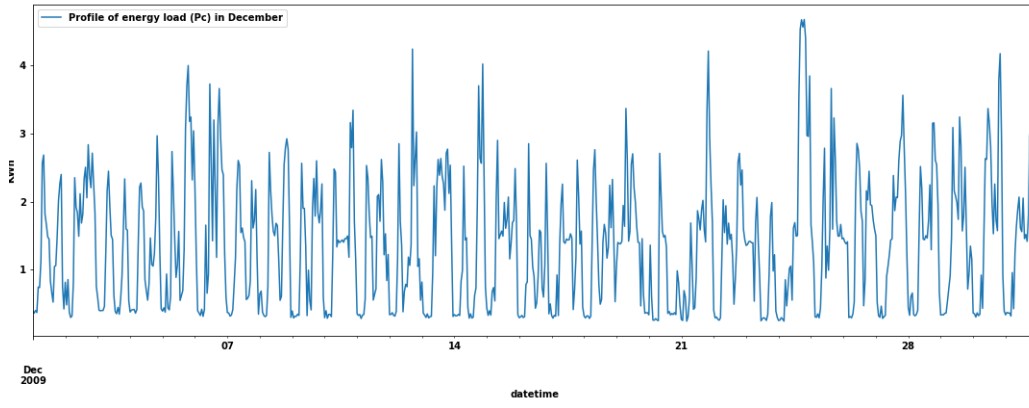
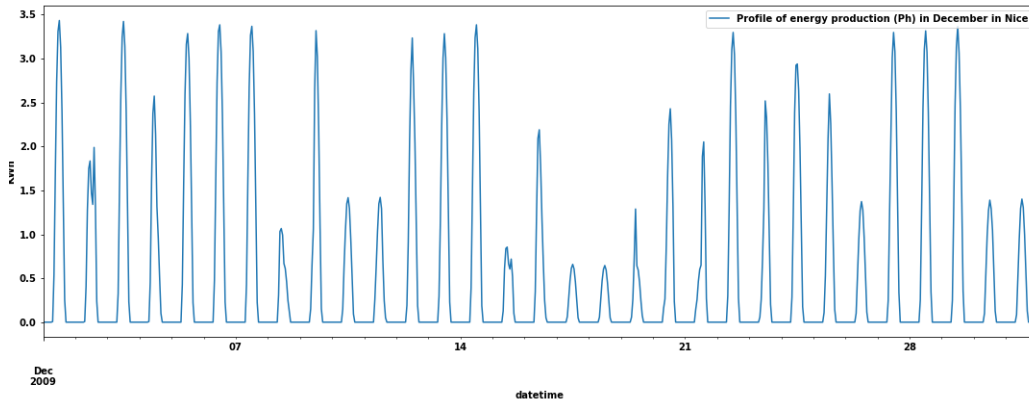
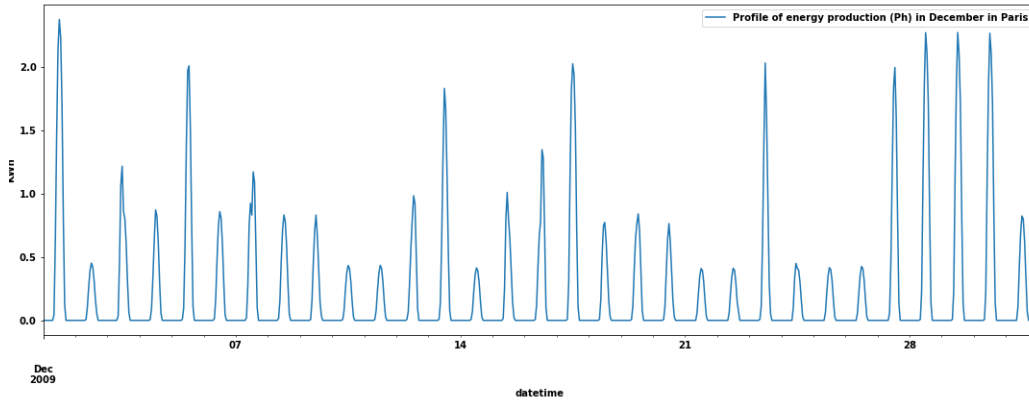
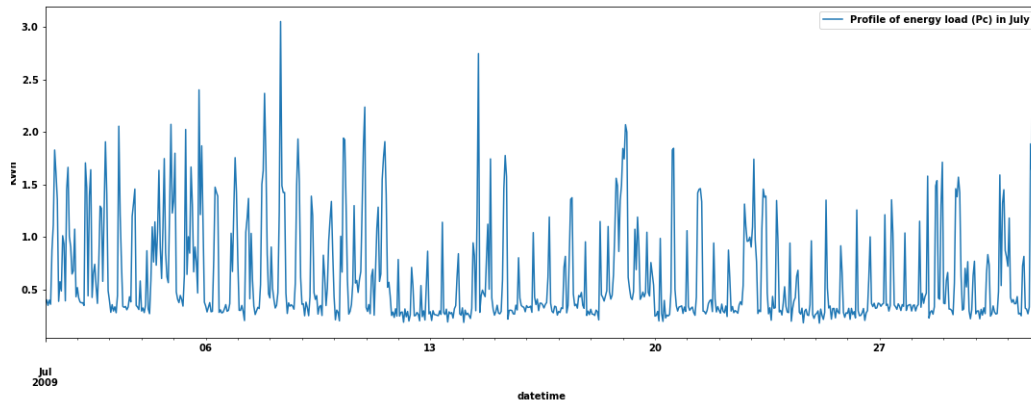
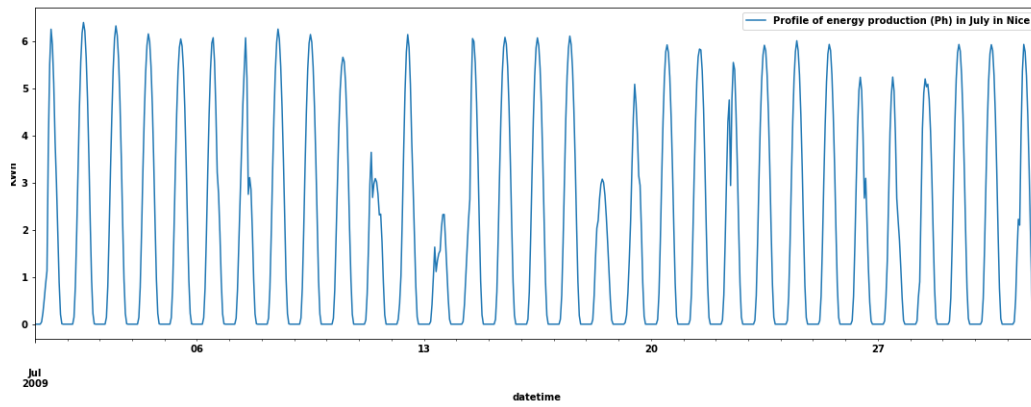
(a) P_C (b) P_H in Nice(c) P_H in Paris

FIGURE 5.3 – Zoom of energy production P_H in Nice and Paris (8kwp) vs. energy load profile P_C in winter condition (i.e. in December)

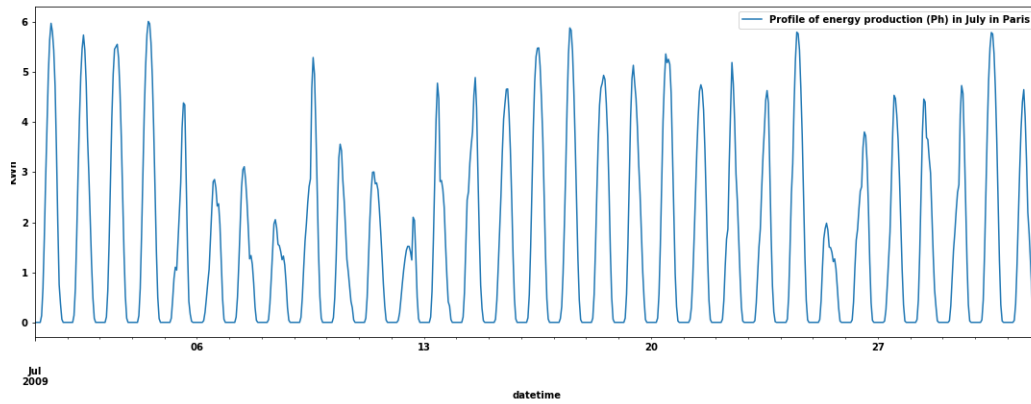
5.3 Grid-connected PV system



(a) P_C



(b) P_H in Nice



(c) P_H in Paris

FIGURE 5.4 – Zoom of energy production P_H in Nice and Paris (8kwp) vs. energy load profile P_C in summer condition (i.e. in JULY)

5.3.2 Electric load profile

For electric load profile, the IHEPCD [G. Hebrail, 2012] dataset available from a French household recorded during December 2006 to November 2010 is used. The power consumption

Parameters	Units	Default Value
System size	Kwp (DC)	8
Module type	Standard, Premium, Thin film	Premium
System losses	%	14
Array type	Fixed open rack, Fixed roof mount, 1-Axis	Fixed open rack
Tilt angle	degrees	35
Azimuth angle	degrees	180
Inverter efficiency	%	96

TABLE 5.1 – Input parameters

data for the year 2009 has been chosen as the reference year for domestic energy in our model. This year contains indeed very few missing data. It is important to note that missing values have been replaced using our STDAN approach presented in chapter 4. Time resolution has been changed to 1 hour to be aligned with energy profile produced by PV system. In figures 5.2, 5.3 and 5.4, the electrical load P_C is juxtaposed with the produced energy P_H . In these figures, the annual electricity production of the system is equal to the annual power consumption P_C of the household. It appears that there is a strong disparity between the electricity produced from solar panels and the electricity consumed during periods of strong sunlight. In the next section, two power management algorithms are proposed for a grid-connected PV system with or without a storage system.

5.3.3 Power management (PM) policies for Grid connected PV System

5.3.3.1 Global overview of the PM

The aim of this work is to propose a power management algorithm maximizing self-consumption for a Grid connected PV system. The figure 5.5 shows a functional overview of our power manager. As it can be seen, the harvested energy $P_H(n)$ and consumed energy $P_C(n)$ are measured during the time slot n using an energy monitoring system such as the prototype proposed in chapter 3. A measurements dataset is used by the LSTM model to predict the harvested $\hat{P}_H(n)$ and consumed

energy $\hat{P}_C(n)$ on the next time slot. Once the predictions are made, the power management takes a decision to switch to the Grid or the PV. If a battery is used, the decision is taken according to the state of charge of the battery at time t (i.e. at the end of time slot n).

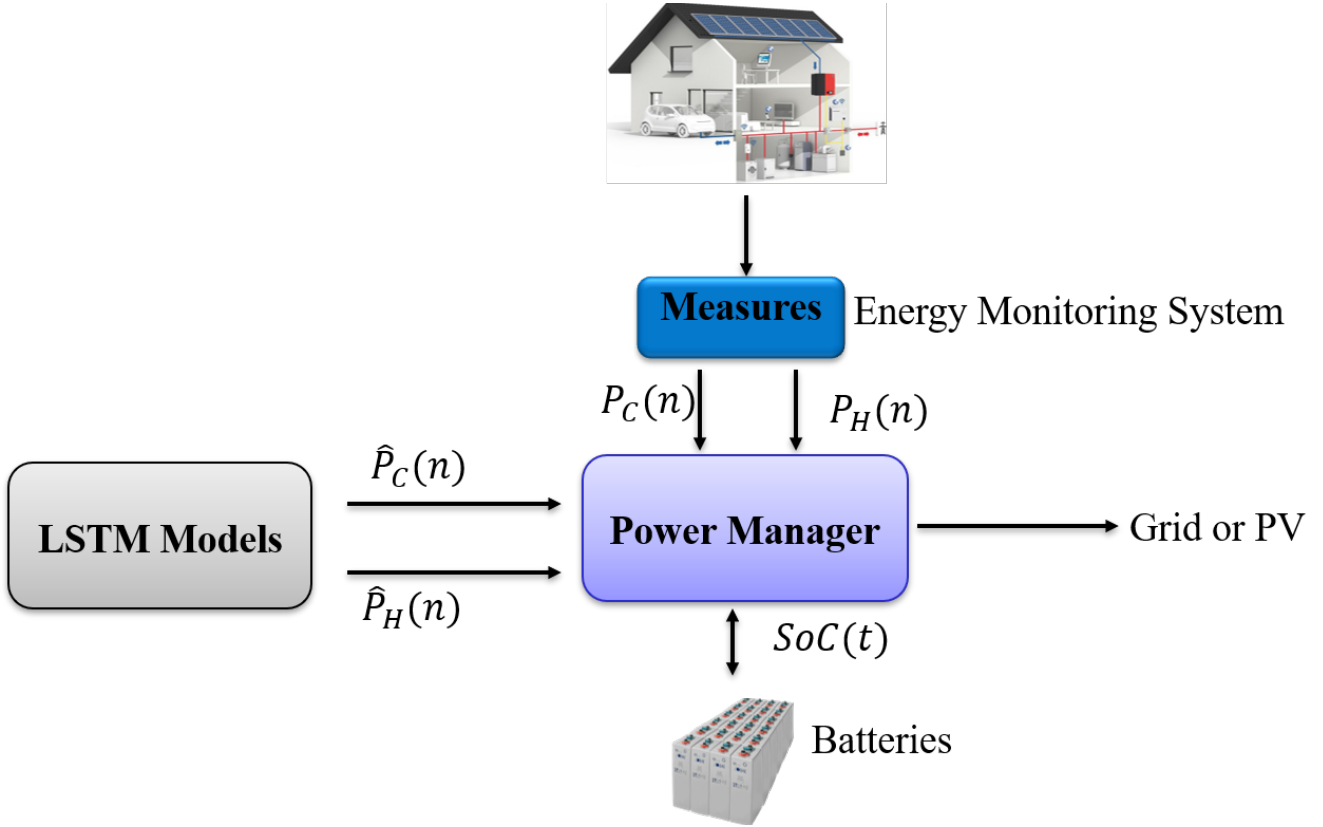


FIGURE 5.5 – Global overview of the PM

5.3.3.2 A slotted-based PM algorithm

The proposed PM algorithm is activated and take decisions periodically. This period is equal to a predefined time slot duration. In this work, each time slot n is equal to one hour. It means that the PM uses predictions to take a decision for the next time slot (i.e. next hour). This time slotted-based PM principle is illustrated on the figure 5.6.

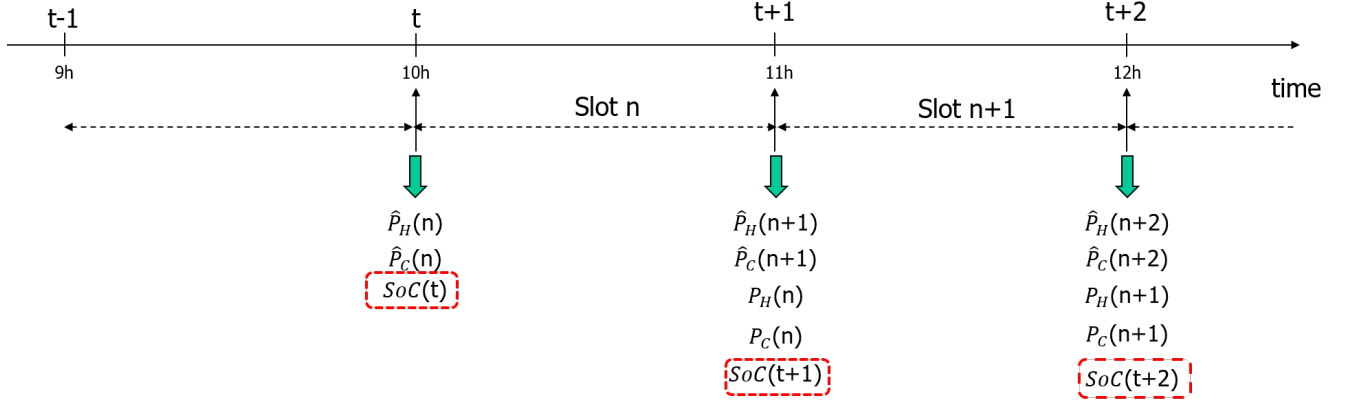


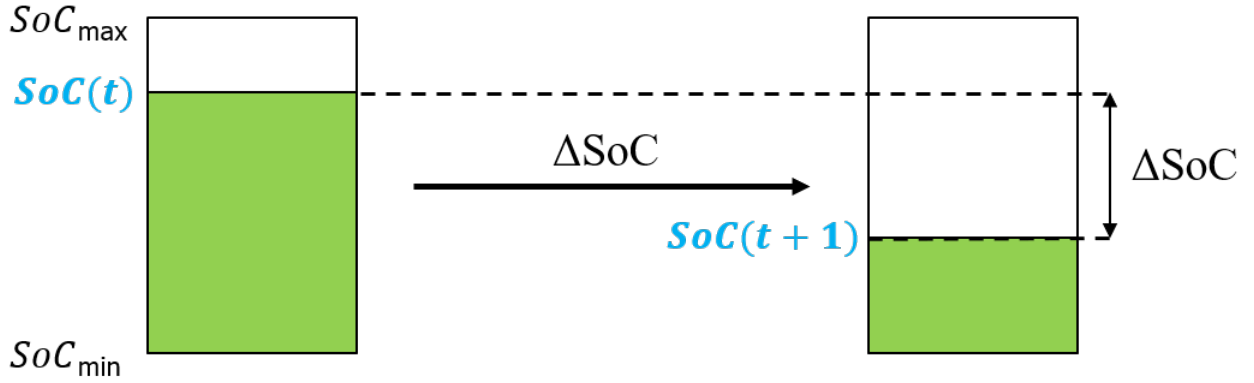
FIGURE 5.6 – Description of power management algorithm with or without batteries

Let us assume that time is equal to t (i.e. 10h00). In order to take a decision for the following time slot (slot n i.e. between 10h00 and 11h00), the PM use predictions of the amount of energy that should be consumed ($\hat{P}_C(n)$) and harvested ($\hat{P}_H(n)$) during that period n . In case a battery is used, the PM also takes into account the state of charge (SoC) of the battery at this time t ($SoC(t)$). One hour later at $t+1$, the PM has to check whether it took a good or a bad decision (PV or Grid) one hour before. To do so, the PM uses real measured values of the consumed ($P_C(n)$) and harvested energy ($P_H(n)$) during the last time slot n (between 10h00 and 11h00). These measures allow the PM to update the different parameters such as the self-consumption rate or the number of blackouts (these parameters will be defined in the section 5.3.3.4). In case a battery is used, the SoC of the battery is updated (i.e. $SoC(t+1)$). This process is repeated every time slot. For instance, at time $t+1$, $\hat{P}_H(n+1)$ and $\hat{P}_C(n+1)$ predictions (and $SoC(t+1)$ if a battery is used) are used by the PM to take decision for the next slot $n+1$.

5.3.3.3 Battery state of charge evolution

As shown on the figure 5.6, the PM policy needs to update the state of charge of the battery at every start of a time slot (i.e. at $SoC(t)$, $SoC(t+1)$, $SoC(t+2)$, ...). the SoC is then used by the PM to take decision for the considered time slot (i.e. $SoC(t)$ for time slot n).

Case 1: $P_C(n) > P_H(n)$ (Battery discharge)



Case 2: $P_C(n) < P_H(n)$ (Battery charge)

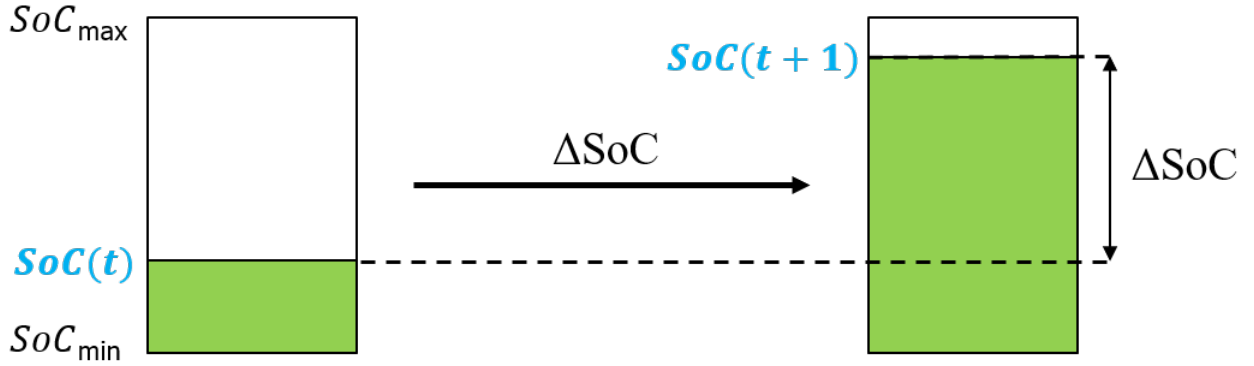


FIGURE 5.7 – Battery Management Strategy

As shown on the figure 5.7, during a time slot the battery can be either charged or discharged, depending on the real values of $P_C(n)$ and $P_H(n)$. The SoC must be updated accordingly for each time slot.

Let us assume that slot n is over and the SoC ($t+1$) must be updated. Two cases can occur :

— **Case 1 : When $P_C(n) \geq P_H(n)$, battery has been discharged**

The new state of charge of the battery is defined as follows :

$$SoC(t+1) = \max(SoC(t) - \Delta SoC, SoC_{min}) \quad (5.1)$$

where ΔSoC is calculated as follows :

$$\Delta SoC = \min(P_C(n) - P_H(n), SoC_{max} - SoC_{min}) \quad (5.2)$$

The ΔSoC represents the variation of the State of charge between time t and $t+1$ (i.e. during time slot n). It is worth to note that in case of discharge, the state of charge can obviously not be lower than the minimum capacity of the battery, i.e. SoC_{min} .

— **Case 2 : When $P_C(n) \leq P_H(n)$, battery has been charged**

In that case, the new state of charge of the battery is defined as follows :

$$SoC(t+1) = \min(SoC(t) + \Delta SoC, SoC_{max}) \quad (5.3)$$

where ΔSoC is calculated as follow :

$$\Delta SoC = \min(P_H(n) - P_C(n), SoC_{max} - SoC_{min}) \quad (5.4)$$

In case of charge, the state of charge can obviously not be higher than the max capacity SoC_{max} . The variation of the state of charge (ΔSoC) can therefore not be higher than the $SoC_{max} - SoC_{min}$. When $P_C(n) \leq P_H(n)$, the surplus of energy, defined as the amount of harvested energy that can be fed to the grid, has to be updated as well. This occurs only when more energy is harvested ($P_H(n)$) than consumed ($P_C(n)$) during a time slot. In such a case, our PM policy gives priority to the battery charge. However, when the battery is fully charged, the remaining harvested energy is then considered as surplus energy and is defined as follows :

$$Surplus+ = \max(SoC(t) + P_H(n) - P_C(n) - SoC_{max}, 0). \quad (5.5)$$

5.3.3.4 Variables used by the PM policy

Total monthly consumption

The Total monthly energy consumption is calculated as follows :

$$Totalmonthlyconsumption = \sum_{i=1}^{N=24 \times 31} E_L(i) \quad (5.6)$$

where E_L represents energy consumed during one hour.

Total PV production

The total photovoltaic production is the global energy produced during a month depending on the predefined configuration. It is calculated as follows :

$$TotalPVproduction = \sum_{i=1}^{N=24 \times 31} E_H(i) \quad (5.7)$$

5.3 Grid-connected PV system

where E_H represents the energy harvested during one hour.

Total surplus energy fed to the grid

The total surplus energy fed to the grid is defined as the total surplus energy that has been injected to the grid.

Total energy bought from the grid

This is the total energy purchased from the grid for the household when the PV system does not provide enough energy.

Self-consumption

The self-consumption can be defined according to two cases :

— **Case 1 : When an energy storage system is not used**

Direct self-consumption consists in using part of the electricity produced on site to supply the electrical load and injecting the surplus production into the Grid.

— **Case 2 : When an energy storage system is used**

Total self-consumption consists in using all the energy produced as well as the energy stored in the battery to supply the electric load.

Self-consumption rate

The self-consumption rate (SCR) is one of the evaluation criteria for a grid-connected PV system. The SCR can be defined according to two cases :

— **Case 1 : When an energy storage system is not used**

This is the ratio between the self-consumption energy used to feed the load divided by the overall production (Total PV production). The self-consumption energy used in this case is the one defined above without battery.

— **Case 2 : When an energy storage system is used**

This is the ratio between the self-consumption energy used to feed the load divided by the overall energy production (Total PV production). The self-consumption energy used in this case is the one defined with battery.

$$SCR = \frac{\sum_{PV Decision=1} \text{Self-Consumption}}{\text{Total PV production}} \quad (5.8)$$

Self-sufficiency rate (SSR)

The self-sufficiency rate is similarly defined as the self-consumption rate. The difference is that the self-consumption is divided by the total monthly energy consumption of the household's electrical load.

$$SSR = \frac{\sum_{PV\ Decision=1} \text{Self Consumption}}{\text{Total Energy Load } P_C} \quad (5.9)$$

Good Decision PH

The Good Decision PH Good Decision PH is defined as the total number of good decisions taken by the Power Management to switch to the PV system.

Bad Decision PH

The Bad Decision PH Bad Decision PH is defined as the total number of Bad decisions taken by the Power Management to switch to the PV system.

Good Decision Grid

The Good Decision Grid Good Decision Grid is defined as the total number of good decisions taken by the Power Management to switch to the grid.

Bad Decision Grid

The Good Decision Grid Bad Decision Grid is defined as the total number of bad decisions taken by the Power Management to switch to the grid.

Blackout error rate (BER)

The blackout error rate is defined as the ratio between the number of time the power management algorithm fails on PV decision (Bad Decision P_H) divided by the total number of time the PV system (Bad Decision P_H + Good Decision P_H) is used. It is calculated from the following equation :

$$BER = \frac{BadDecisionP_H}{BadDecisionP_H + GoodDecisionP_H} \quad (5.10)$$

5.3.3.5 Power management algorithm without battery

The algorithm 1 is the Power Management policy that we propose when the household is not equipped with a battery. The aim of this policy is to take decisions (switch to the grid or to the PV) at the beginning of each time slot based on predictions of consumed and harvested energy ($\hat{P}_C(n)$ and $\hat{P}_H(n)$ respectively). As shown, at the end of each time slot, the Power Management

5.3 Grid-connected PV system

checks if the decision taken at the beginning of the time slot was correct or not.

In case PV was chosen, the PM checks if the consumed energy $P_C(n)$ was actually less than the harvested energy $P_H(n)$. If it is not the case, $P_C(n) > P_H(n)$, then it means that a bad decision PV was taken, and therefore the bad decision PH counter is incremented. In that case, the number of blackout is also incremented and the self consumption is equal to the overall harvested energy $P_H(n)$ (as all the harvested energy has been consumed by the load).

When $P_C(n)$ is less than $P_H(n)$, i.e. the decision was good, the counter of good decision PH is incremented. In that case, the self-consumption is incremented by the overall consumed energy during this time slot $P_C(n)$, while the remaining energy ($P_H(n) - P_C(n)$) is considered as the surplus. If the Grid was chosen, there is obviously no risk of blackout. But, this decision could be a bad decision when the actual harvested energy is higher than the consumed energy ($P_C(n) < P_H(n)$). In such a case, the bad decision Grid is incremented. Otherwise, the counter of good decision grid is incremented and the surplus is incremented by all the energy harvested during this time slot. Then, a new decision must be taken for the next time slot. This process is repeated for each time slot.

5.3.3.6 Power management algorithm with battery

The algorithm 2 is the Power Management policy that we propose when a battery is used. The aim of this policy is to take decisions (switch to grid or to the PV) based on predictions of consumed and harvested energy ($\hat{P}_C(n)$ and $\hat{P}_H(n)$ respectively) and as well as the state of charge of the battery at time t ($SoC(t)$). As shown, at the end of each time slot, the Power Management checks if the decision taken at the beginning of the time slot was correct or not.

In case PV was chosen (PV decision = 1), the PM checks if the consumed energy $P_C(n)$ was actually less than $P_H(n) + SoC(t)$. If it is not the case, meaning that $P_C(n) > P_H(n) + SoC(t)$, then a bad decision PV was taken, and therefore the bad decision PH counter is incremented. In that case, the number of blackout is also incremented and the self consumption is equal to the overall harvested energy plus the state of charge of the battery ($P_H(n) + SoC(t)$). As all the harvested energy has been consumed, the new state of charge of the battery $SoC(t + 1)$ is updated to SoC_{min} .

When $P_C(n)$ is lower than $P_H(n) + SoC(t)$, i.e. the decision was good, the counter of good decision PH is incremented. In that case, the battery is either discharged (when $P_C(n) > P_H(n)$) or charged according to real measures of $P_C(n)$ and $P_H(n)$ as well as the state of the charge $SoC(t)$. In case the harvested energy is higher than the consumed energy, the remaining energy is considered as a surplus as described by the equation 5.5. In case of good decision PH, the

Algorithm 1 Power management algorithm without a battery

```

1: if  $\hat{P}_H(n) > \hat{P}_C(n)$  then
2:   PV decision  $\leftarrow 1$ 
3: else
4:   PV decision  $\leftarrow 0$ 
5: end if
6: for  $t : 1$  to  $\text{len}(N\text{steps})$  do
7:   if PV Decision = 1 then
8:     if  $P_C(n) > P_H(n)$  then
9:       Blackout  $\leftarrow \text{Blackout} + 1$ 
10:      Bad Decision PH  $\leftarrow \text{Bad Decision PH} + 1$ 
11:      Selfconsumption  $\leftarrow P_H(n)$ 
12:    else
13:      Selfconsumption  $\leftarrow P_C(n)$ 
14:      Surplus  $\leftarrow P_H(n) - P_C(n)$ 
15:      Good decision PH  $\leftarrow \text{Good Decision PH} + 1$ 
16:    end if
17:  else
18:    if  $P_C(n) < P_H(n)$  then
19:      Bad decision Grid  $\leftarrow \text{Bad Decision Grid} + 1$ 
20:    else
21:      Good decision Grid  $\leftarrow \text{Good decision Grid} + 1$ 
22:      Surplus  $\leftarrow P_H(n)$ 
23:    end if
24:  end if
25:  if  $\hat{P}_H(n + t) > \hat{P}_C(n + t)$  then
26:    PV decision  $\leftarrow 1$ 
27:  else
28:    PV decision  $\leftarrow 0$ 
29:  end if
30: end for

```

5.3 Grid-connected PV system

self-consumption is incremented by the overall consumed energy $P_C(n)$.

If the Grid was chosen (i.e. PV decision = 0), there is obviously no risk of blackout. But, this decision could be a bad decision if the actual harvested energy added to the state of charge of the battery is higher than the consumed energy ($P_C(n) < P_H(n) + SoC(t)$). Otherwise, the counter of good decision grid is incremented and the new state of the charge $SoC(t+1)$ is updated according to the equation 5.11 :

$$SoC(t+1) = \min(P_H(n) + SoC(t), SoC_{max}) \quad (5.11)$$

In that case, the surplus is also incremented by all the remaining harvested minus the maximum capacity of the battery (SoC_{max}).

$$surplus+ = \max(SoC(t) + P_H(n) - SoC_{max}, 0) \quad (5.12)$$

Then, a new decision is taken for the next time slot. This process is repeated for each time slot.

Algorithm 2 Power management algorithm with a battery

```

1: if  $\hat{P}_H(n) + SoC(t) - SoC_{min} > \hat{P}_C(n)$  then
2:    $PV\ decision \leftarrow 1$ 
3: else
4:    $PV\ decision \leftarrow 0$ 
5: end if
6: for  $t : 1$  to  $len(Nsteps)$  do
7:   if  $PV\ Decision = 1$  then
8:     if  $P_C(n) > P_H(n) + SoC(t)$  then
9:        $Blackout \leftarrow Blackout + 1$ 
10:       $Bad\ decision\ PV \leftarrow Bad\ decision\ PV + 1$ 
11:       $Selfconsumption(t) + = P_H(n) + SoC(t)$ 
12:       $SoC(t + 1) = SoC_{min}$ 
13:    else
14:      if  $P_C(n) > P_H(n)$  then
15:         $\Delta SoC \leftarrow \min(P_C(n) - P_H(n), SoC_{max} - SoC_{min})$ 
16:         $SoC(t + 1) \leftarrow \max(SoC(t) - \Delta SoC, SoC_{min})$ 
17:      else
18:         $\Delta SoC \leftarrow \min(P_H(n) - P_C(n), SoC_{max} - SoC_{min})$ 
19:         $SoC(t + 1) \leftarrow \min(SoC(t) + \Delta SoC, SoC_{max})$ 
20:         $Surplus + = \max(SoC(t) + P_H(n) - P_C(n) - SoC_{max}, 0)$ 
21:      end if
22:       $Selfconsumption + = P_C(n)$ 
23:       $Good\ decision\ PV \leftarrow Good\ Decision\ PV + 1$ 
24:    end if
25:  else
26:    if  $P_C(n) < P_H(n) + SoC(t)$  then
27:       $Bad\ decision\ Grid \leftarrow Bad\ Decision\ Grid + 1$ 
28:    else
29:       $Good\ decision\ Grid \leftarrow Good\ decision\ Grid + 1$ 
30:    end if
31:     $Surplus + = \max(SoC(t) + P_H(n) - SoC_{max}, 0)$ 
32:     $SoC(t + 1) \leftarrow \min(P_H(n) + SoC(t), SoC_{max})$ 
33:  end if
34:  if  $\hat{P}_H(n + 1) + SoC(t + 1) - SoC_{min} > \hat{P}_C(n + t)$  then
35:     $PV\ decision \leftarrow 1$ 
36:  else
37:     $PV\ decision \leftarrow 0$ 
38:  end if
39: end for

```

Configurations	PV size	Battery size	Period of the year	Geolocalisation
Config. 1	2kwp	20kwh	Winter	Nice
Config. 2	4kwp	20kwh	Winter	Nice
Config. 3	8kwp	20kwh	Winter	Nice
Config. 4	2kwp	20kwh	Summer	Nice
Config. 5	4kwp	20kwh	Summer	Nice
Config. 6	8kwp	20kwh	Summer	Nice
Config. 7	2kwp	20kwh	Winter	Paris
Config. 8	4kwp	20kwh	Winter	Paris
Config. 9	8kwp	20kwh	Winter	Paris
Config. 10	2kwp	20kwh	Summer	Paris
Config. 11	4kwp	20kwh	Summer	Paris
Config. 12	8kwp	20kwh	Summer	Paris
Config. 13	2kwp	10kwh	Winter	Nice
Config. 14	4kwp	10kwh	Winter	Nice
Config. 15	8kwp	10kwh	Winter	Nice
Config. 16	2kwp	10kwh	Summer	Nice
Config. 17	4kwp	10kwh	Summer	Nice
Config. 18	8kwp	10kwh	Summer	Nice
Config. 19	2kwp	10kwh	Winter	Paris
Config. 20	4kwp	10kwh	Winter	Paris
Config. 21	8kwp	10kwh	Winter	Paris
Config. 22	2kwp	10kwh	Summer	Paris
Config. 23	4kwp	10kwh	Summer	Paris
Config. 24	8kwp	10kwh	Summer	Paris
Config. 25	2kwp	5kwh	Winter	Nice
Config. 26	4kwp	5kwh	Winter	Nice
Config. 27	8kwp	5kwh	Winter	Nice
Config. 28	2kwp	5kwh	Summer	Nice
Config. 29	4kwp	5kwh	Summer	Nice
Config. 30	8kwp	5kwh	Summer	Nice
Config. 31	2kwp	5kwh	Winter	Paris
Config. 32	4kwp	5kwh	Winter	Paris
Config. 33	8kwp	5kwh	Winter	Paris
Config. 34	2kwp	5kwh	Summer	Paris
Config. 35	4kwp	5kwh	Summer	Paris
Config. 36	8kwp	5kwh	Summer	Paris

TABLE 5.2 – Configurations used in this study

5.4 Results and Discussion

5.4.1 Simulation setup

In order to evaluate the efficiency of the proposed power management policy and to quantify the evaluation criteria mentioned above, simulations for different configurations listed in table 5.2 have been performed. The following parameters are varied in those configurations :

- PV size : the size of solar panels from 2 to 8 kWp ;
- Battery size : the battery capacity is either equal to 5, 10 or 20 kWh ;
- Period of the year : winter and summer conditions are considered for the harvested energy (P_H)
- Geolocalisation : two different areas (Nice and Paris) are considered for the harvested energy ;

All the simulations are carried out over a period of 1 month, representing therefore 743 time slots and as many decisions to be taken by the Power Manager. In the next section, we present the results obtained for a house located either in Nice or Paris. For each locality, we simulate either winter (December) or summer (July) conditions. Then, for each weather conditions, the power management algorithm is evaluated using or not a battery, and considering different PV sizes. We assume that $P_C(n)$ does not change for investigating the impact of PV sizing and batteries on energy efficiency.

5.4.2 Household located in south of France (Nice)

Winter condition (December)

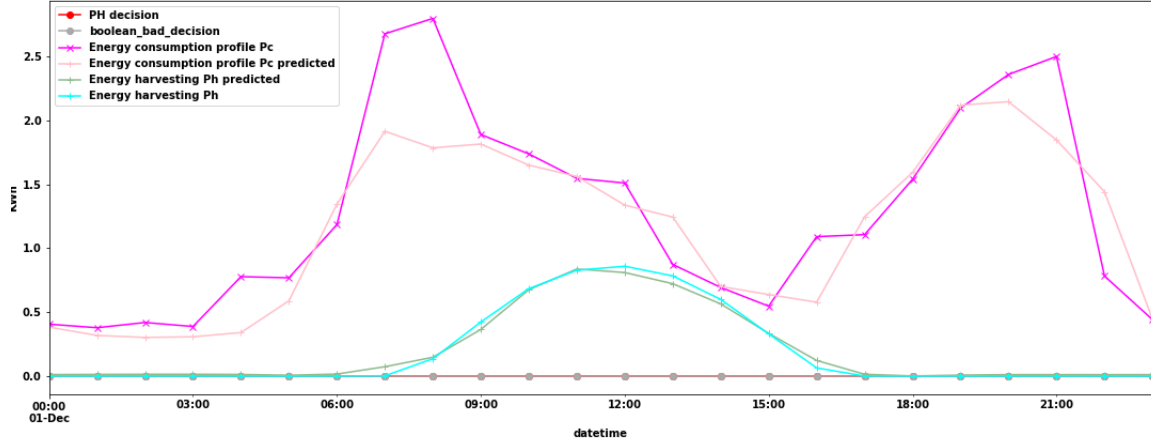
Tables 5.3, 5.4 and 5.5 are simulation results obtained for Nice in winter condition (December). As shown, the harvested energy (total PV production) in December is lower than the total monthly consumption. As the harvested energy is low, the self-consumption energy is low as well, even more when the battery is not used (because not enough energy is produced to cover the energy load P_C). This remaining energy produced will be considered as a surplus and injected into the grid. In case the battery is used, we notice a significant increase of the self-consumption energy. As the harvested energy P_H is stored in the battery, it can be used (during a time slot) when there is enough energy to cover the energy load P_C . This is the reason why zero energy is injected into the grid as shown in tables 5.3, 5.4 and 5.5.

TABLE 5.3 – Results obtained for config.1

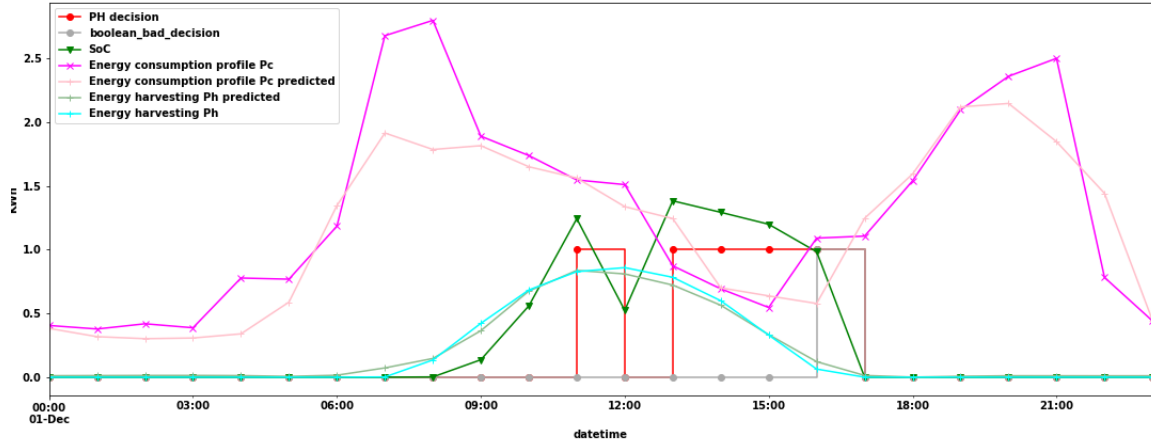
	System capacity 2 Kw	
	Without Batteries	With Batteries
Total monthly consumption	1015.132 kwh	1015.132 kwh
Total PV production	89.665	89.665
Self Consumption	5.505 kwh	89.44 kwh
Total fed to the grid	84.159 kwh	0 kwh
Total bought from the grid	1009.627 kwh	925.69 kwh
Self consumption rate	6.1399%	99.752%
Self sufficiency rate	0.542%	8.81%
Good Décision PH	9	89
Bad Décision PH	3	22
Good Décision Grid	719	599
Bad Décision Grid	12	33
Blackout error rate	25%	19.81%
Area (m^2)	10	10

TABLE 5.4 – Results obtained for config.2

	System capacity 4 Kw	
	Without Batteries	With Batteries
Total monthly consumption	1015.132 kwh	1015.132 kwh
Total PV production	179.330 kwh	179.330 kwh
Self Consumption	43.685 kwh	178.756 kwh
Total fed to the grid	135.645 kwh	0 kwh
Total bought from the grid	971.447 kwh	836.376 kwh
Self consumption rate	24.360%	99.68%
Self sufficiency rate	4.303%	17.609%
Good Décision PH	41	157
Bad Décision PH	11	21
Good Décision Grid	673	532
Bad Décision Grid	18	33
Blackout error rate	21.1%	11.7%
Area (m^2)	20	20



(a)

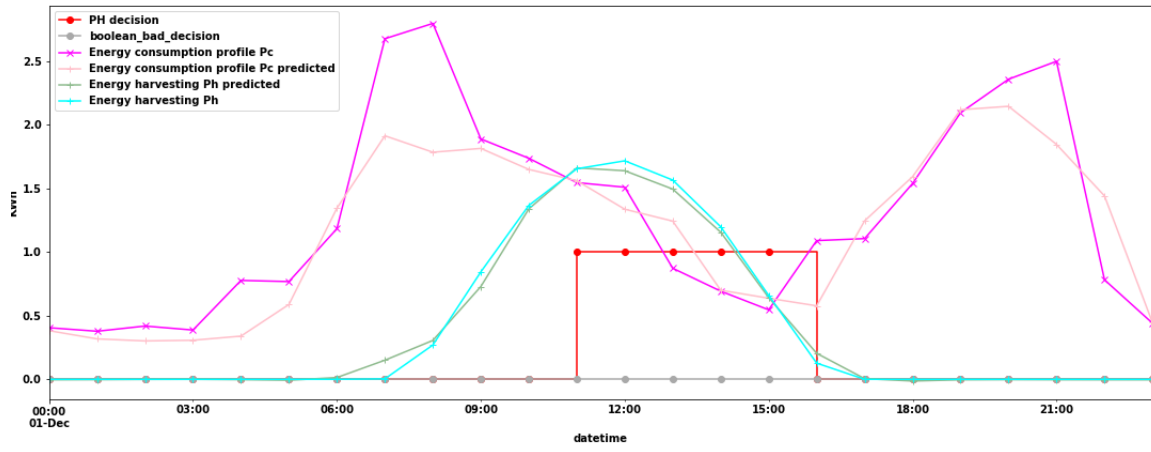


(b)

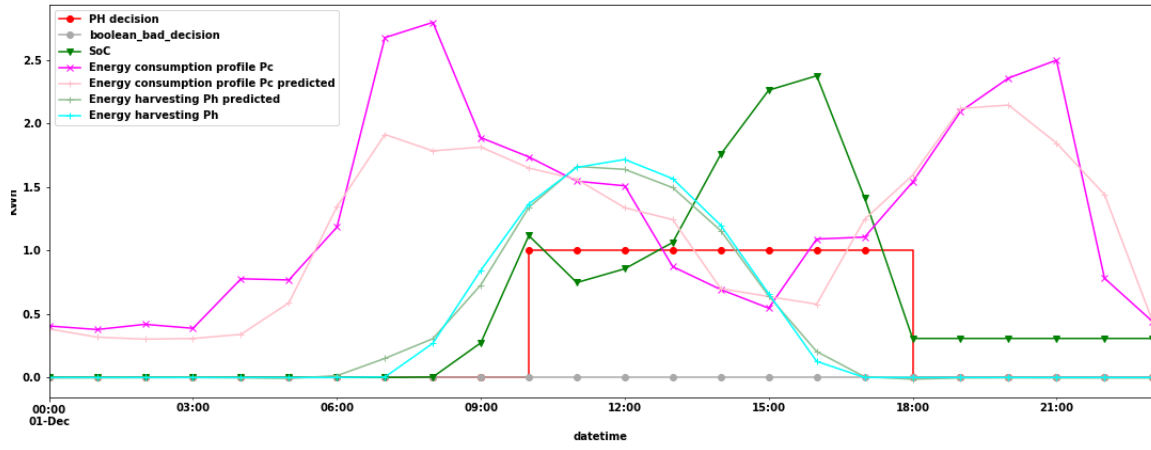
FIGURE 5.8 – Energy flows of a residential in Nice : Grid-PV Connected without (a) and with battery system for one day of December (b) (PV system size 2 kWp, battery size 20 kWh, annual load demand 9 MWh)

The strategy implemented in our power manager is to use the grid when the harvested energy P_H is low compared to the energy load P_C during a time slot. In the case the battery is not used, and as expected, the power manager decides to use the grid (719 against 599, 673 against 532, 622 against 435 respectively on the grid and PV as can be seen in tables 5.3, 5.4 and 5.5 respectively). In consequence, there is a strong demand for energy bought from the grid to cover the energy load P_C . We can also observe that, when the battery is used, the self-sufficiency rate doubles every time the PV size is doubled when the battery is used. On the other hand, the self-consumption rate tends to saturate with the increase of the PV size. The defined strategy indeed consists in recharging the battery rather than selling the surplus energy produced by the PV system when the battery is not fulfilled.

5.4 Results and Discussion

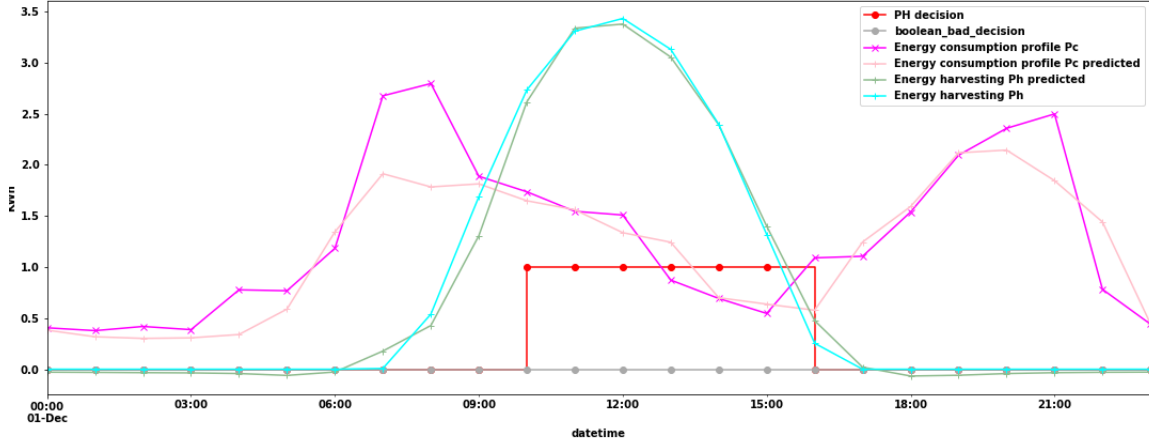


(a)

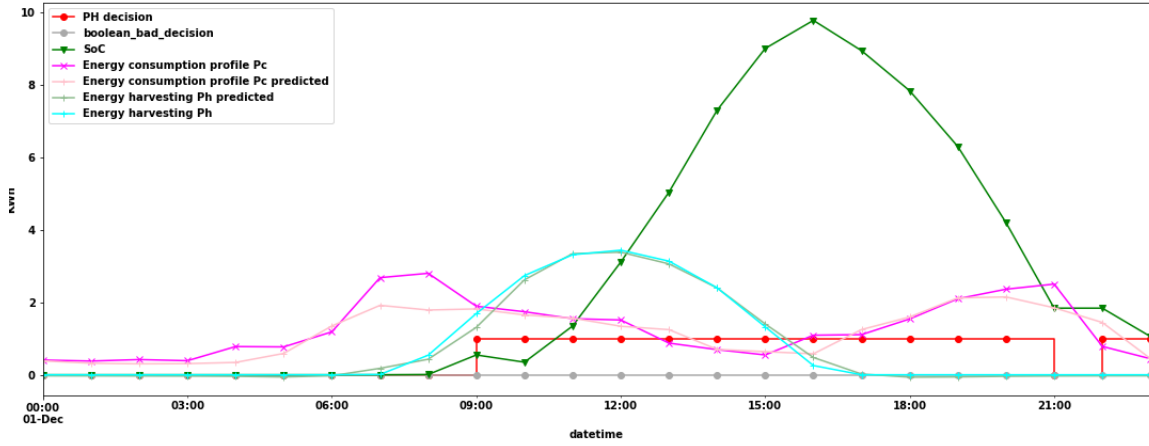


(b)

FIGURE 5.9 – Energy flows of a residential in Nice : Grid-PV Connected without (a) and with battery system for one day of December (b) (PV system size 4 kWp, battery size 20 kWh, annual load demand 9 MWh)



(a)



(b)

FIGURE 5.10 – Energy flows of a residential in Nice : Grid-PV Connected without (a) and with battery system for one day of December (b) (PV system size 8 kWp, battery size 20 kWh, annual load demand 9 MWh)

In order to analyze the decisions taken by our power manager in tables 5.3, 5.4 and 5.5, we present three figures (figures 5.8, 5.9 and 5.10) showing the evolution of the different parameters in a single day in December and illustrating the decisions taken by the power manager for each time slot. As can be seen in figure 5.8a, the predicted energy $\hat{P}_H(n)$ remains lower than the predicted energy $\hat{P}_C(n)$ throughout the day. For that reason, the power manager never takes the decision to switch to PV (PH decision remains at level 0). When a battery is used (figure 5.8b), we can see that the power management switches to PV on some time slots during the day (for example from 11 to 12 in the morning and 1 p.m to 5 p.m). During these time slots, the harvested energy and the energy stored in the battery seem to be enough to cover the consumed energy. However, among these 5 time slots where the power manager switched to the PV, the decision taken at 4pm was incorrect. As can be seen in figure 5.8b, during this time slot (from 4pm to 5pm), the predicted consumed energy $\hat{P}_C(n)$ is lower than the predicted harvested energy added to the energy stored

TABLE 5.5 – Results obtained for config.3

	System capacity 8 Kw	
	Without Batteries	With Batteries
Total monthly consumption	1015.132 kwh	1015.132 kwh
Total PV production	358.660 kwh	358.660 kwh
Self Consumption	125.221 kwh	358.321 kwh
Total fed to the grid	233.439 kwh	0 kwh
Total bought from the grid	889.911 kwh	656.8164 kwh
Self consumption rate	34.913%	99.905%
Self sufficiency rate	12.3355%	35.2979%
Good Décision PH	97	251
Bad Décision PH	14	27
Good Décision Grid	622	435
Bad Décision Grid	10	30
Blackout error rate	12.6%	9.7%
Area (m^2)	40	40

in the battery ($\hat{P}_H(n) + SoC(t)$). Predictions were therefore too optimistic for this time slot, as the real energy P_C is actually greater than the $P_H(n) + SoC(t)$. In consequence, a bad decision is raised.

In Figures 5.9 and 5.10, the amount of harvested energy increases with the PV size. In consequence, bad decision PH does not occur throughout the day, meaning that all decisions taken by the power manager are correct. We can observe that, despite the number of PH decision increases with the PV size, the blackout error rate decreases. When the battery is used, the blackout error rate is even more decreased compared to the case a battery is not used. This impact of the battery on the blackout error rate will be discussed in the next section.

After varying the size of PV for a battery capacity of 20kWh, we then reduced the battery size from 20kWh to 10kWh. Tables 5.6, 5.7 and 5.8 show the results obtained in this case. As can be seen, the results of tables 5.6 and 5.7 are the same as those of tables 5.3 and 5.4. Indeed, as the simulations take place in winter conditions with low sunshine (in December), the battery size has only an impact when the PV sizes are larger as the case of table 5.8. When the PV size is increased, the energy that can be stored in the battery (and used later) increases as well. For small PV sizes, it is not possible to saturate the battery capacity, and explains that the same results are obtained in tables 5.3, 5.4, 5.6 and 5.7, as well as the variation observed with table 5.8 compared to table 5.5.

Summer condition (July)

To simulate summer conditions, so less energy consumed and a lot of harvested energy, the energy manager was evaluated for the month of July. Tables 5.9, 5.10 and 5.11 are the simulation

TABLE 5.6 – Results obtained for config.13

	System capacity 2 Kw	
	Without Batteries	With Batteries
Total monthly consumption	1015.132 kwh	1015.132 kwh
Total PV production	89.665 kwh	89.665 kwh
Self Consumption	5.505 kwh	89.44 kwh
Total fed to the grid	84.159 kwh	0 kwh
Total bought from the grid	1009.627 kwh	925.69 kwh
Self consumption rate	6.1399%	99.752%
Self sufficiency rate	0.542%	8.81%
Good Décision PH	9	89
Bad Décision PH	3	22
Good Décision Grid	719	599
Bad Décision Grid	12	33
Blackout error rate	25%	19.81%
Area (m^2)	10	10

TABLE 5.7 – Results obtained for config.14

	System capacity 4 Kw	
	Without Batteries	With Batteries
Total monthly consumption	1015.132 kwh	1015.132 kwh
Total PV production	179.330 kwh	179.330 kwh
Self Consumption	43.685 kwh	178.756 kwh
Total fed to the grid	135.645 kwh	0 kwh
Total bought from the grid	971.447 kwh	836.376 kwh
Self consumption rate	24.360%	99.68%
Self sufficiency rate	4.303%	17.609%
Good Décision PH	41	157
Bad Décision PH	11	21
Good Décision Grid	673	532
Bad Décision Grid	18	33
Blackout error rate	21.15%	11.79%
Area (m^2)	20	20

results obtained in summer (July) for a household located in Nice. As can be seen, for a PV size of 4 or 8 kWp, the harvested energy (total PV production) is greater than the total monthly consumption. As the harvested energy is important, the self-consumed energy is also much more significant even though the battery is not used because the PV system produces enough energy to cover the energy load. The remaining harvested energy is considered as a surplus and is fed into the grid.

When the battery is used, the self-consumption increases significantly because the harvested energy can be stored in the battery, so that it can be used when there is enough energy to cover the

TABLE 5.8 – Results obtained for config.15

	System capacity 8 Kw	
	Without Batteries	With Batteries
Total monthly consumption	1015.132 kwh	1015.132 kwh
Total PV production	358.660 kwh	358.660 kwh
Self Consumption	125.221 kwh	336.18384 kwh
Total fed to the grid	233.439 kwh	22.1377 kwh
Total bought from the grid	889.911 kwh	678.9541 kwh
Self consumption rate	34.913%	93.73307%
Self sufficiency rate	12.3355%	33.1172%
Good Décision PH	97	239
Bad Décision PH	14	22
Good Décision Grid	622	453
Bad Décision Grid	10	29
Blackout error rate	12.6%	9.7%
Area (m^2)	40	40

TABLE 5.9 – Results obtained for config.4

	System capacity 2 Kw	
	Without Batteries	With Batteries
Total monthly consumption	459.952 kwh	459.952 kwh
Total PV production	330.199 kwh	330.199 kwh
Self Consumption	111.306 kwh	329.803 kwh
Total fed to the grid	218.892 kwh	0 kwh
Total bought from the grid	348.645 kwh	130.1486 kwh
Self consumption rate	33.708%	99.88%
Self sufficiency rate	24.199%	71.7038%
Good Décision PH	185	552
Bad Décision PH	30	14
Good Décision Grid	470	131
Bad Décision Grid	58	46
Blackout error rate	13.9%	2.5%
Area (m^2)	10	10

energy load. This is the reason why zero energy is fed into the grid for a PV size of 2kWp (table 5.9). We also note a decrease in energy demand bought from the grid for both cases (with or without battery) compared to the results obtained in December (tables 5.3, 5.4 and 5.5). As there is a strong energy production during the month of July, the power manager frequently switches to PV. The results obtained in tables 5.9, 5.10 and 5.11 also show an increase in the self-sufficiency rate compared to December (tables 5.3, 5.4 and 5.5). The self-sufficiency rate SSR tends to saturate when the PV size increases. This information can be used to determine if the system is over-sized. On the other hand, the self-consumption rate SCR tends to decrease when the PV size increases

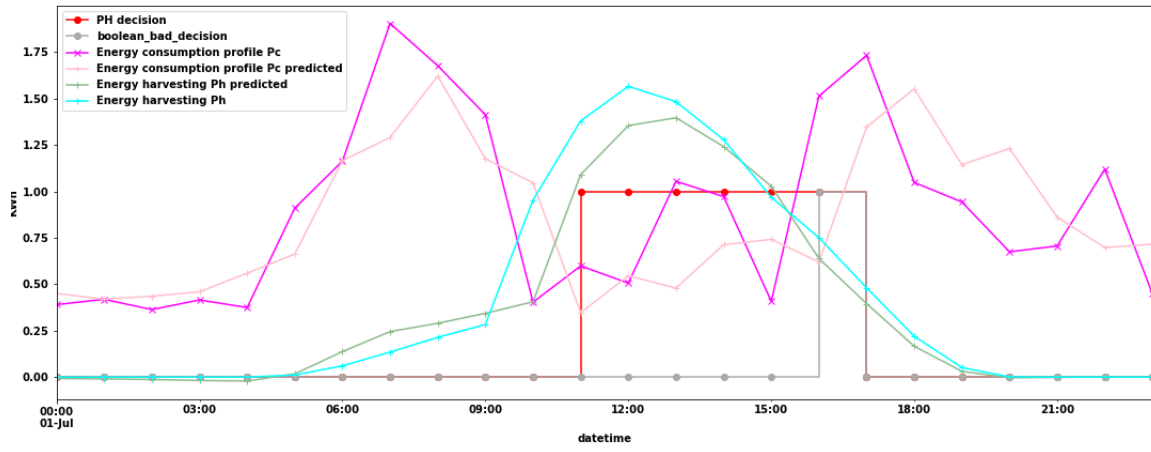
(tables 5.9, 5.10 and 5.11). Despite the harvested energy increases significantly with the PV size, the self-consumed energy decreases compared to the energy produced.

TABLE 5.10 – Results obtained for config.5

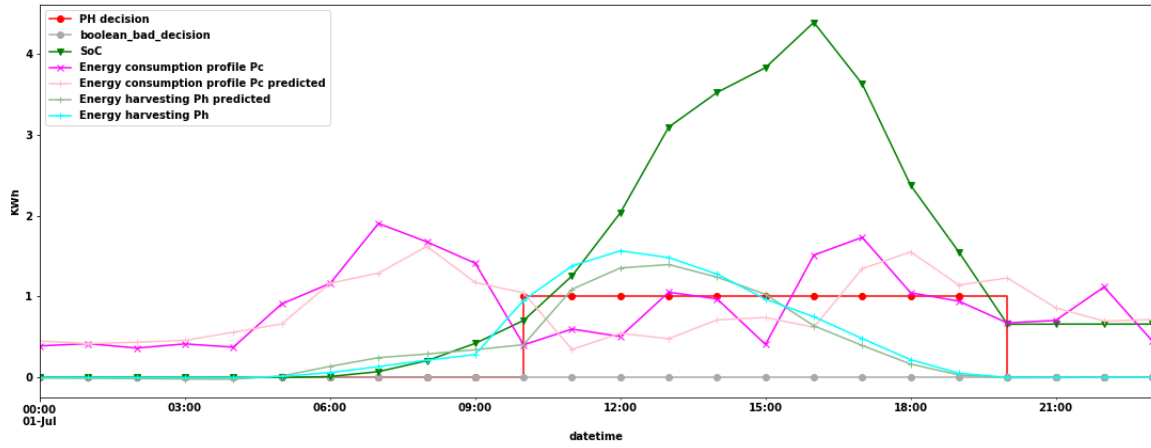
	System capacity 4 Kw	
	Without Batteries	With Batteries
Total monthly consumption	459.952 kwh	459.952 kwh
Total PV production	660.399 kwh	660.399 kwh
Self Consumption	185.912 kwh	450.339 kwh
Total fed to the grid	474.4865 kwh	201.421 kwh
Total bought from the grid	274.039 kwh	9.612 kwh
Self consumption rate	28.151%	68.192%
Self sufficiency rate	40.420%	97.910%
Good Décision PH	301	729
Bad Décision PH	17	1
Good Décision Grid	399	12
Bad Décision Grid	26	1
Blackout error rate	5.3%	0.137%
Area (m^2)	20	20

TABLE 5.11 – Results obtained for config.6

	System capacity 8 Kw	
	Without Batteries	With Batteries
Total monthly consumption	459.952 kwh	459.952 kwh
Total PV production	1320.7986 kwh	1320.7986 kwh
Self Consumption	230.773 kwh	452.033 kwh
Total fed to the grid	1090.024 kwh	858.790 kwh
Total bought from the grid	229.1783 kwh	7.919 kwh
Self consumption rate	17.47%	34.224%
Self sufficiency rate	50.173%	98.278%
Good Décision PH	355	732
Bad Décision PH	13	1
Good Décision Grid	350	10
Bad Décision Grid	25	0
Blackout error rate	3.5%	0.136%
Area (m^2)	40	40

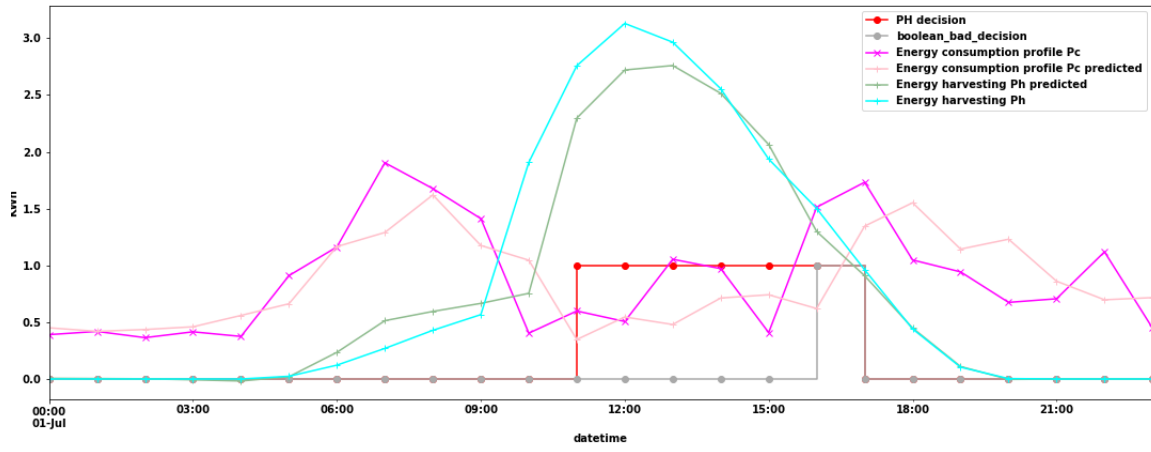


(a)

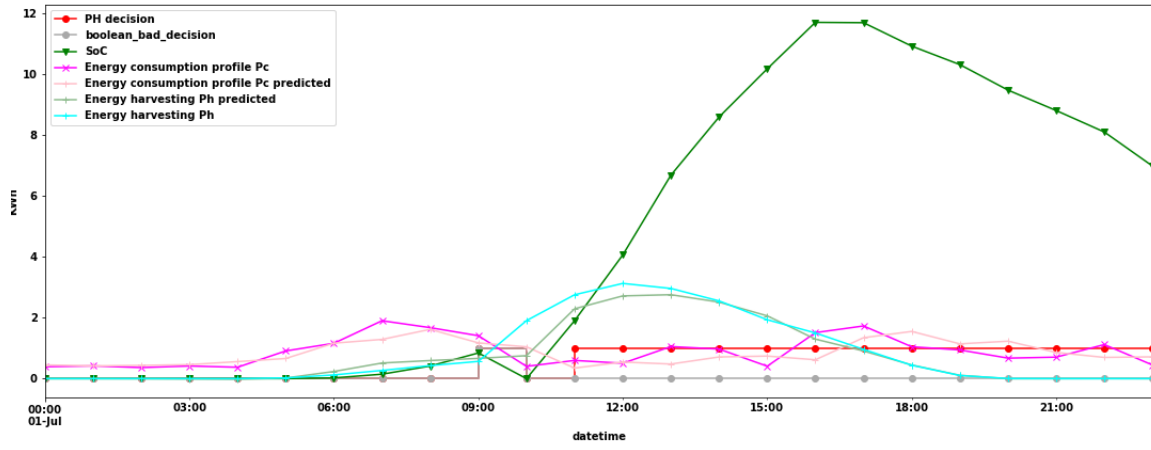


(b)

FIGURE 5.11 – Energy flows of a residential : Grid-PV Connected without (a) and with battery system for one day of July (b) (PV system size 2 kWp, battery size 20 kWh, annual load demand 9 MWh)

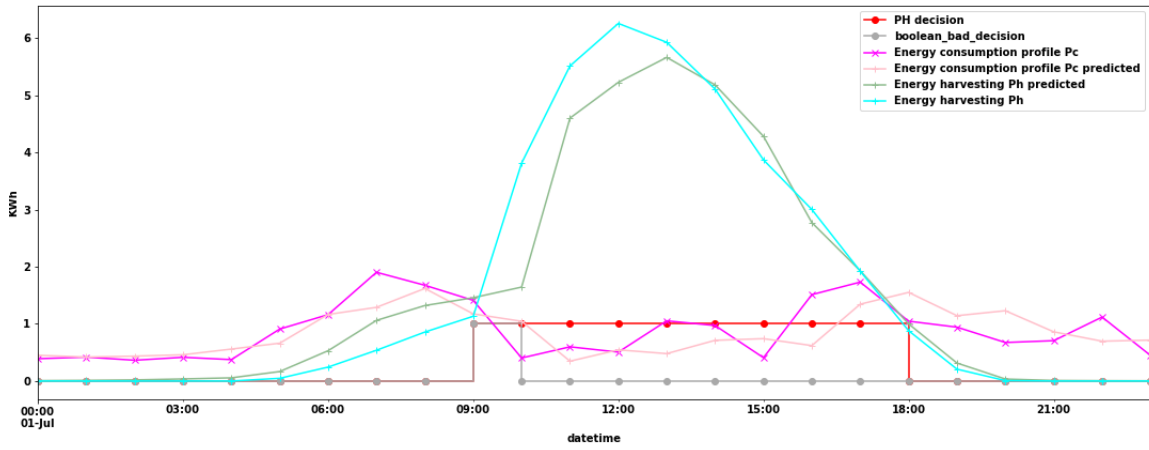


(a)

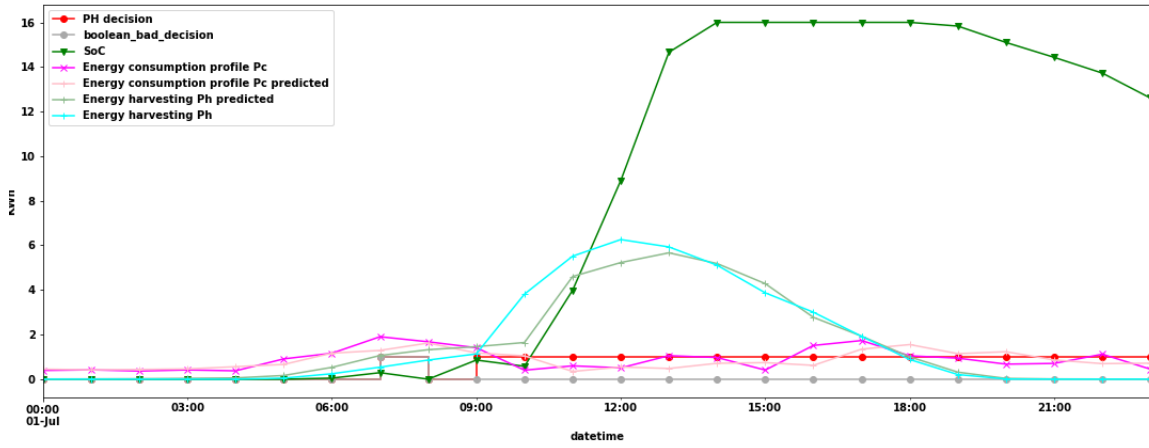


(b)

FIGURE 5.12 – Energy flows of a residential : Grid-PV Connected without (a) and with battery system for one day of July (b) (PV system size 4 kWp, battery size 20 kWh, annual load demand 9 MWh)



(a)



(b)

FIGURE 5.13 – Energy flows of a residential : Grid-PV Connected without (a) and with battery system for one day of July (b) (PV system size 8 kWp, battery size 20 kWh, annual load demand 9 MWh)

In order to analyze the decisions taken by the power manager in tables 5.9, 5.10 and 5.11, figures 5.11, 5.12 and 5.13 show the results obtained on a single day in July. As can be seen in figure 5.11a, the power management has switched 5 times to PV with a Bad decision PH at the slot time from 4 p.m to 5 p.m. As noticed in section 5.3.3, the power manager takes the decision to switch to PV because the predicted harvested energy $\hat{P}_H(n)$ was greater than the predicted consumed energy $\hat{P}_C(n)$ at time t . When the battery is used, the power manager switches 10 times on the PV without any Bad Decision throughout the day. This shows that using a battery improves the decision-making of the power manager. We also notice that the battery SoC reached during July is much more higher when compared to the month of December (figures 5.8, 5.9 and 5.10). In Figures 5.12 and 5.13 we can even observe that the battery reaches its maximum capacity (i.e. its saturation level). This battery saturation can be explained by the fact that the household profile

used in this study exhibits a decrease in energy demand during the month of July. This is likely because the occupants are on vacation during that period.

TABLE 5.12 – Results obtained for config.16

	System capacity 2 Kw	
	Without Batteries	With Batteries
Total monthly consumption	459.952 kwh	459.952 kwh
Total PV production	330.199 kwh	330.199 kwh
Self Consumption	111.306 kwh	322.83977 kwh
Total fed to the grid	218.892 kwh	6.94037 kwh
Total bought from the grid	348.645 kwh	137.1123 kwh
Self consumption rate	33.708%	97.7710%
Self sufficiency rate	24.199%	70.1898%
Good Décision PH	185	543
Bad Décision PH	30	15
Good Décision Grid	470	140
Bad Décision Grid	58	45
Blackout error rate	13.9%	2.68%
Area (m^2)	10	10

TABLE 5.13 – Results obtained for config.17

	System capacity 4 Kw	
	Without Batteries	With Batteries
Total monthly consumption	459.952 kwh	459.952 kwh
Total PV production	660.399 kwh	660.399 kwh
Self Consumption	185.912 kwh	406.7182 kwh
Total fed to the grid	474.4865 kwh	252.7231 kwh
Total bought from the grid	274.039 kwh	53.2338 kwh
Self consumption rate	28.151%	61.5867%
Self sufficiency rate	40.420%	88.42621%
Good Décision PH	301	662
Bad Décision PH	17	6
Good Décision Grid	399	48
Bad Décision Grid	26	27
Blackout error rate	5.3%	0.89%
Area (m^2)	20	20

Tables 5.12, 5.13 and 5.14 show the results obtained when the battery capacity is reduced from 20kWh to 10kWh. As can be seen, the results slightly differ from those obtained in tables 5.9, 5.10 and 5.11. As the simulations take place during a month of strong sunshine (July), it is expected that the performance will decrease with a 10kWh battery as can be seen in the tables 5.12, 5.13 and 5.14. If the battery is reduced, the battery saturation level is reached more quickly and the energy fed

TABLE 5.14 – Results obtained for config.18

	System capacity 8 Kw	
	Without Batteries	With Batteries
Total monthly consumption	459.952 kwh	459.952 kwh
Total PV production	1320.7986 kwh	1320.7986 kwh
Self Consumption	230.773 kwh	427.5020 kwh
Total fed to the grid	1090.024 kwh	891.001 kwh
Total bought from the grid	229.1783 kwh	32.4501 kwh
Self consumption rate	17.47%	32.3669%
Self sufficiency rate	50.173%	92.9448%
Good Décision PH	355	694
Bad Décision PH	13	3
Good Décision Grid	350	35
Bad Décision Grid	25	11
Blackout error rate	3.5%	0.43%
Area (m^2)	40	40

into the grid increases. In the meantime, the energy bought from the grid is also increased when the battery size is reduced by 2 (from 20kWh to 10kWh).

5.4.3 Household located in North of France (Paris)

Winter condition (December)

Tables 5.15, 5.16 and 5.17 are the simulation results obtained for a household located in Paris during winter conditions (December). As can be seen, the harvested energy (total PV production) in December is less than the total monthly consumption. This is due to the weather conditions used in the simulation (December). As the harvested energy is low compared to Nice for the month of December, the self-consumed energy is also much lower when the battery is not used. We can observe a slight increase in this self-consumed energy when the battery is used. This is due to the fact that the harvested energy is stored in the battery when the power management switches to the Grid. This stored energy will then be used when there is enough energy to cover the energy load, which justifies zero energy injected into the grid as shown in tables 5.15, 5.16 and 5.17.

We can also observe that the self-sufficiency rate doubles every time the PV size is doubled when the battery is used. On the other hand, the self-consumption rate tends to saturate when the PV size increases. The reason is that the defined strategy consists in recharging the battery rather than selling the surplus energy produced by the PV system when the battery is not full.

In order to analyze the decisions taken by the power manager in tables 5.15, 5.16 and 5.17, figures 5.14, 5.15 and 5.16 show the results obtained during a single day in December. As can be seen in figure 5.14a, the predicted harvested energy $\hat{P}_H$ remains lower than the predicted consumed

TABLE 5.15 – Results obtained for config.7

	System capacity 2 Kw	
	Without Batteries	With Batteries
Total monthly consumption	1015.132 kwh	1015.132 kwh
Total PV production	37.868 kwh	37.868 kwh
Self Consumption	0.576 kwh	36.654 kwh
Total fed to the grid	37.291 kwh	0 kwh
Total bought from the grid	1014.556 kwh	978.47 kwh
Self consumption rate	1.521%	96.79%
Self sufficiency rate	0.0567%	3.610%
Good Décision PH	2	48
Bad Décision PH	0	13
Good Décision Grid	740	644
Bad Décision Grid	1	38
Blackout error rate	0%	21.3%
Area (m^2)	10	10

TABLE 5.16 – Results obtained for config.8

	System capacity 4 Kw	
	Without Batteries	With Batteries
Total monthly consumption	1015.132 kwh	1015.132 kwh
Total PV production	75.736 kwh	75.736 kwh
Self Consumption	2.763 kwh	75.228 kwh
Total fed to the grid	72.972 kwh	0 kwh
Total bought from the grid	1012.369 kwh	939.904 kwh
Self consumption rate	3.649%	99.32%
Self sufficiency rate	0.272%	7.410%
Good Décision PH	5	71
Bad Décision PH	1	22
Good Décision Grid	731	616
Bad Décision Grid	6	34
Blackout error rate	16.6%	23.6%
Area (m^2)	20	20

energy $\hat{P}_C$ throughout the day. For this reason, the power manager does not take the decision to switch to PV. PH decision in figure 5.14a remains therefore at level 0. If the battery is used (figure 5.14b), we can see that the power manager switches to PV for three time slots during the day (from 12h to 13h and 14h to 16h). These decisions are correct since there is no Bad Decision PH. In figure 5.15b, a Bad Decision PH has been taken by the power manager at 4p.m.

Tables 5.18, 5.19 and 5.20 show the results obtained when the battery is set at 10kWh instead of 20kWh. As can be seen, the results are the same as those of tables 5.14, 5.15 and 5.16. This

TABLE 5.17 – Results obtained for config.9

	System capacity 8 Kw	
	Without Batteries	With Batteries
Total monthly consumption	1015.132 kwh	1015.132 kwh
Total PV production	149.363 kwh	149.363 kwh
Self Consumption	26.552 kwh	148.513 kwh
Total fed to the grid	122.811 kwh	0 kwh
Total bought from the grid	988.580 kwh	866.619 kwh
Self consumption rate	17.777%	99.431%
Self sufficiency rate	2.6155%	14.6299%
Good Décision PH	23	127
Bad Décision PH	3	20
Good Décision Grid	696	568
Bad Décision Grid	21	28
Blackout error rate	11.5%	13.6%
Area (m^2)	40	40

TABLE 5.18 – Results obtained for config.19

	System capacity 2 Kw	
	Without Batteries	With Batteries
Total monthly consumption	1015.132 kwh	1015.132 kwh
Total PV production	37.868 kwh	37.868 kwh
Self Consumption	0.576 kwh	36.654 kwh
Total fed to the grid	37.291 kwh	0 kwh
Total bought from the grid	1014.556 kwh	978.47 kwh
Self consumption rate	1.521%	96.79%
Self sufficiency rate	0.0567%	3.610%
Good Décision PH	2	48
Bad Décision PH	0	13
Good Décision Grid	740	644
Bad Décision Grid	1	38
Blackout error rate	0%	21.3%
Area (m^2)	10	10

results indicates that, in these conditions (i.e. during winter in Paris) a battery size of 10kWh is enough.

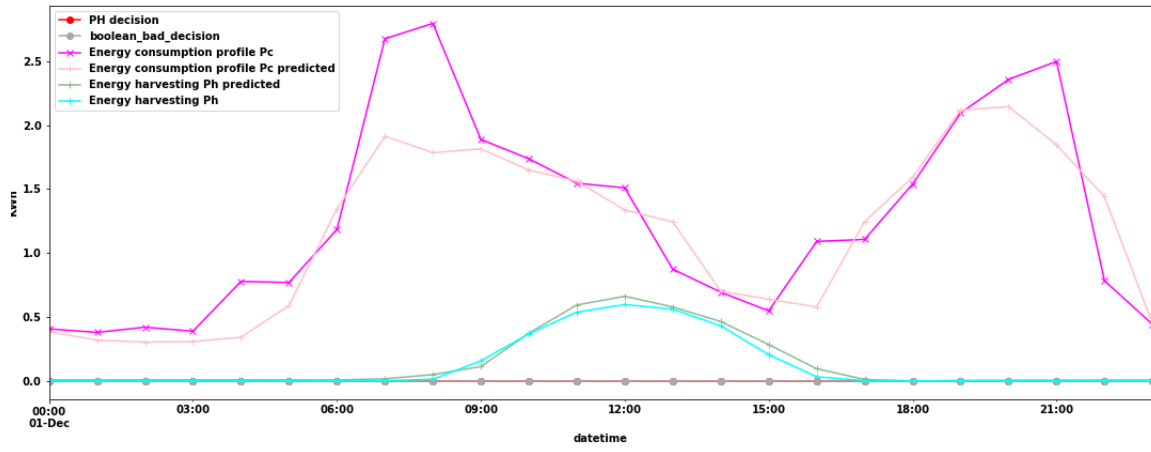
TABLE 5.19 – Results obtained for config.20

	System capacity 4 Kw	
	Without Batteries	With Batteries
Total monthly consumption	1015.132 kwh	1015.132 kwh
Total PV production	75.736 kwh	75.736 kwh
Self Consumption	2.763 kwh	75.228 kwh
Total fed to the grid	72.972 kwh	0 kwh
Total bought from the grid	1012.369 kwh	939.904 kwh
Self consumption rate	3.649%	99.32%
Self sufficiency rate	0.272%	7.410%
Good Décision PH	5	71
Bad Décision PH	1	22
Good Décision Grid	731	616
Bad Décision Grid	6	34
Blackout error rate	16.6%	23.6%
Area (m^2)	20	20

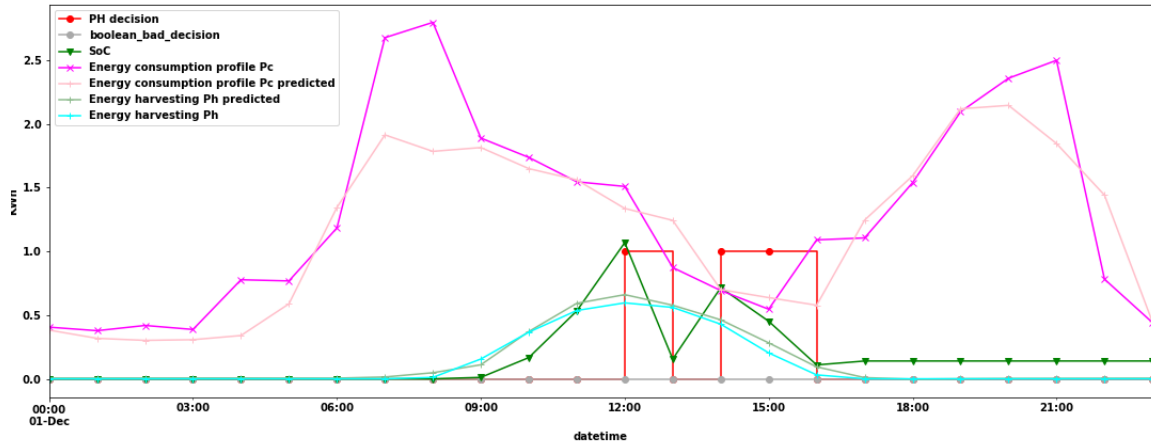
TABLE 5.20 – Results obtained for config.21

	System capacity 8 Kw	
	Without Batteries	With Batteries
Total monthly consumption	1015.132 kwh	1015.132 kwh
Total PV production	149.363 kwh	149.363 kwh
Self Consumption	26.552 kwh	148.513 kwh
Total fed to the grid	122.811 kwh	0 kwh
Total bought from the grid	988.580 kwh	866.619 kwh
Self consumption rate	17.777%	99.431%
Self sufficiency rate	2.6155%	14.6299%
Good Décision PH	23	127
Bad Décision PH	3	20
Good Décision Grid	696	568
Bad Décision Grid	21	28
Blackout error rate	11.5%	13.6%
Area (m^2)	40	40

5.4 Results and Discussion

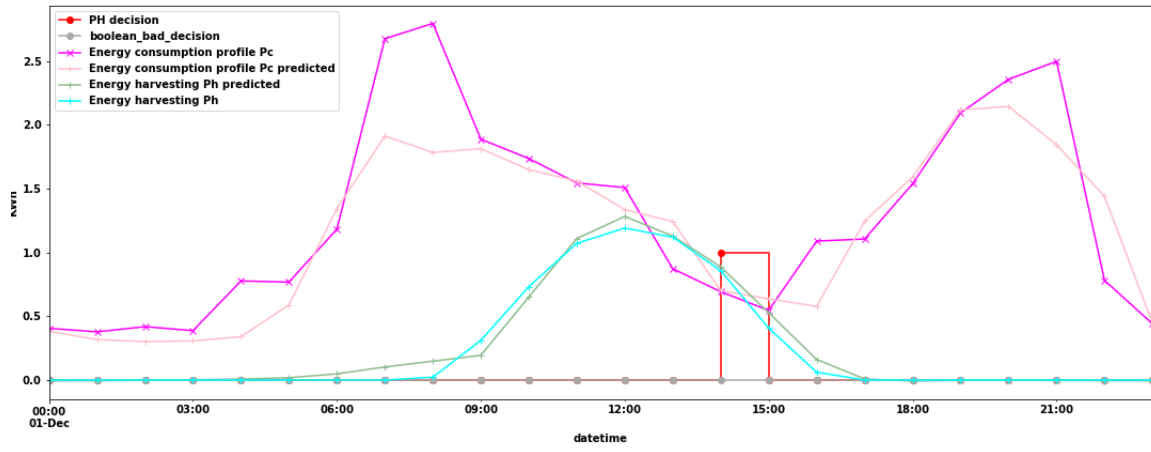


(a)

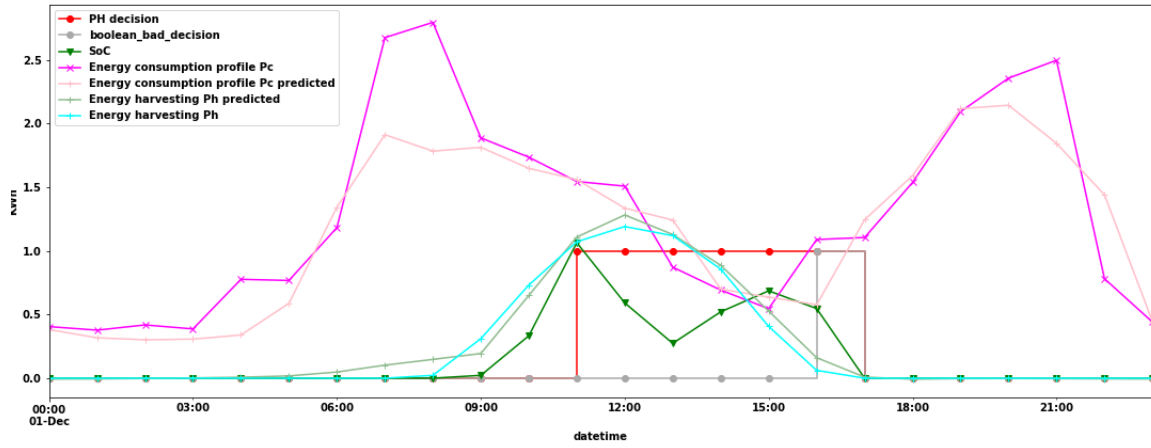


(b)

FIGURE 5.14 – Energy flows of a residential in Paris : Grid-PV Connected without (a) and with battery system for one day of December (b) (PV system size 2 kWp, battery size 20 kWh, annual load demand 9 MWh)

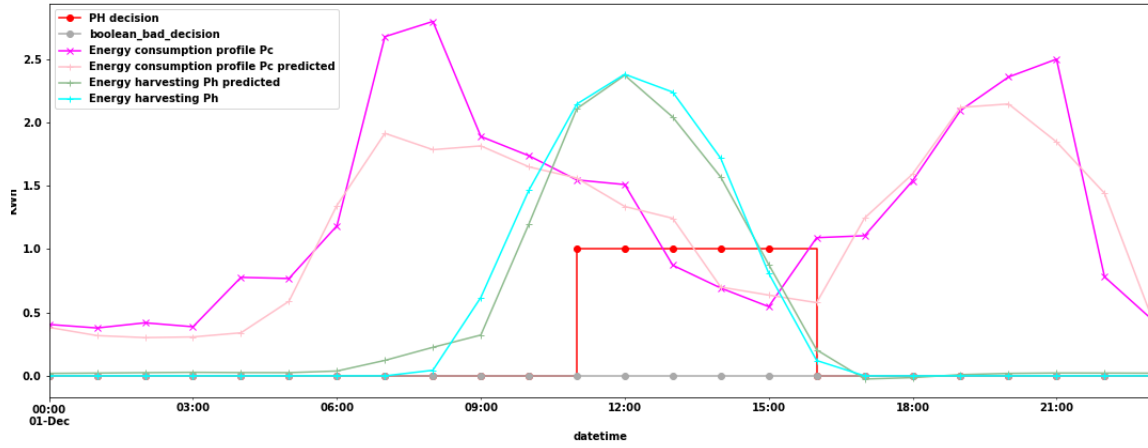


(a)

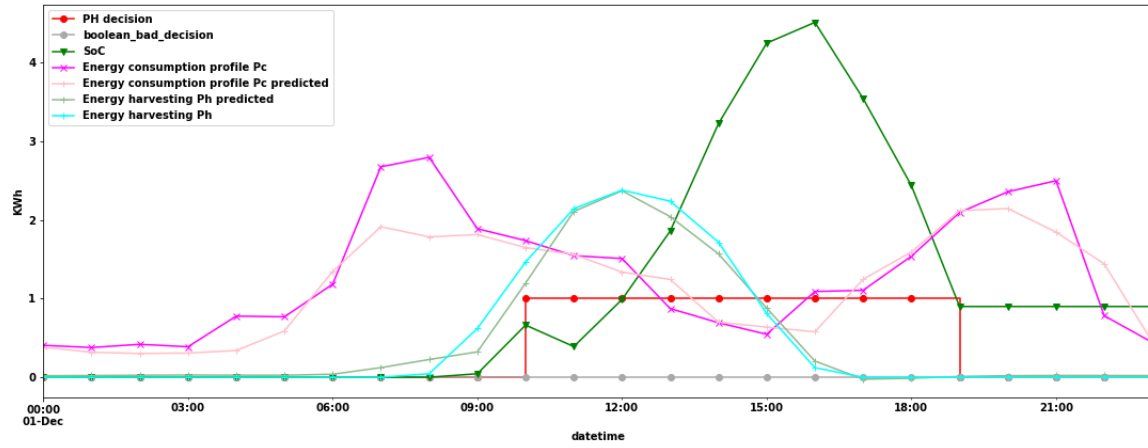


(b)

FIGURE 5.15 – Energy flows of a residential in Paris : Grid-PV Connected without (a) and with battery system for one day of December (b) (PV system size 4 kWp, battery size 20 kWh, annual load demand 9 MWh)



(a)



(b)

FIGURE 5.16 – Energy flows of a residential in Paris : Grid-PV Connected without (a) and with battery system for one day of December (b) (PV system size 8 kWp, battery size 20 kWh, annual load demand 9 MWh)

Summer condition (July)

To simulate summer conditions (i.e. less energy consumed and a lot of harvested energy), the power manager has been evaluated during the month of July. Tables 5.21, 5.22 and 5.23 are the simulation results obtained for Paris in this condition (i.e. in July). As can be seen, the harvested energy (total PV production) in July is greater than the total monthly consumption when the PV size is from 4kWp to 8kWp (as in Nice). As the harvested energy is enough, the self-consumed energy in this case is also much more important even if the battery is not used. In other words, the energy produced is enough to cover the energy load. This remaining energy produced will be considered as a surplus and will be fed into the grid.

In the case the battery is used, the self-consumed energy significantly increases because the energy produced can be stored in the battery, so that it can be used when there is enough energy

TABLE 5.21 – Results obtained for config.10

	System capacity 2 Kw	
	Without Batteries	With Batteries
Total monthly consumption	459.952 kwh	459.952 kwh
Total PV production	263.148 kwh	263.148 kwh
Self Consumption	80.852 kwh	262.542 kwh
Total fed to the grid	182.295 kwh	0 kwh
Total bought from the grid	379.099 kwh	197.409 kwh
Self consumption rate	30.725%	99.7698%
Self sufficiency rate	17.578%	57.0804%
Good Décision PH	159	469
Bad Décision PH	20	12
Good Décision Grid	509	181
Bad Décision Grid	55	81
Blackout error rate	11.2%	2.5%
Area (m^2)	10	10

TABLE 5.22 – Results obtained for config.11

	System capacity 4 Kw	
	Without Batteries	With Batteries
Total monthly consumption	459.952 kwh	459.952 kwh
Total PV production	526.296 kwh	526.296 kwh
Self Consumption	155.458 kwh	417.7983 kwh
Total fed to the grid	370.8382 kwh	100.2057 kwh
Total bought from the grid	304.494 kwh	42.153 kwh
Self consumption rate	29.538%	79.384%
Self sufficiency rate	33.798%	90.835%
Good Décision PH	260	677
Bad Décision PH	23	1
Good Décision Grid	416	44
Bad Décision Grid	44	21
Blackout error rate	8.1%	0.15%
Area (m^2)	20	20

to cover the energy load. This explains that zero energy is fed into the grid for a PV size of 2kWp (table 5.21).

We also note a decrease in energy demand bought from the grid in both cases (with or without battery) compared to the results obtained in December. This is due to the fact that the weather conditions favor a high energy production during July. In that case, the power manager frequently switches to PV. The results obtained in Tables 5.21, 5.22 and 5.23 also show an increase in the self-sufficiency rate compared to December. On the other hand, the self-sufficiency rate does not tend to saturate like the case of Nice. This shows that a location exhibiting less sunshine requires

TABLE 5.23 – Results obtained for config.12

	System capacity 8 Kw	
	Without Batteries	With Batteries
Total monthly consumption	459.952 kwh	459.952 kwh
Total PV production	1071.203 kwh	1071.203 kwh
Self Consumption	216.429 kwh	455.712 kwh
Total fed to the grid	854.7745 kwh	606.211
Total bought from the grid	243.5227 kwh	4.240 kwh
Self consumption rate	20.204%	42.542%
Self sufficiency rate	47.0547%	99.078%
Good Décision PH	342	736
Bad Décision PH	14	0
Good Décision Grid	354	7
Bad Décision Grid	33	0
Blackout error rate	3.9%	0%
Area (m^2)	40	40

a larger battery size. The self-consumption rate tends to decrease with the increase of the PV size (tables 5.21, 5.22 and 5.23. The reason is that the energy produced significantly increases with the PV size while the self-consumed energy decreases according to the energy produced.

TABLE 5.24 – Results obtained for config.22

	System capacity 2 Kw	
	Without Batteries	With Batteries
Total monthly consumption	459.952 kwh	459.952 kwh
Total PV production	263.148 kwh	263.148 kwh
Self Consumption	80.852 kwh	262.542 kwh
Total fed to the grid	182.295 kwh	0 kwh
Total bought from the grid	379.099 kwh	197.409 kwh
Self consumption rate	30.725%	99.7698%
Self sufficiency rate	17.578%	57.0804%
Good Décision PH	159	469
Bad Décision PH	20	12
Good Décision Grid	509	181
Bad Décision Grid	55	81
Blackout error rate	11.2%	2.5%
Area (m^2)	10	10

Tables 5.24, 5.25 and 5.26 show the results obtained when the battery capacity is set to 10kWh. As can be seen, these results differ from those obtained in tables 5.21, 5.22 and 5.23 when a 20kWh battery was used. As the simulations take place in July, it is expected that the performance decreases when a battery of 10kWh is used. Indeed, with a smaller battery the saturation level is reached more quickly, and will then increase the energy injected into the grid. In addition, the

TABLE 5.25 – Results obtained for config.23

	System capacity 4 Kw	
	Without Batteries	With Batteries
Total monthly consumption	459.952 kwh	459.952 kwh
Total PV production	526.296 kwh	526.296 kwh
Self Consumption	155.458 kwh	378.8213 kwh
Total fed to the grid	370.8382 kwh	146.0404 kwh
Total bought from the grid	304.494 kwh	81.1307 kwh
Self consumption rate	29.538%	71.97871%
Self sufficiency rate	33.798%	82.36103%
Good Décision PH	260	629
Bad Décision PH	23	10
Good Décision Grid	416	78
Bad Décision Grid	44	26
Blackout error rate	8.1%	1.56%
Area (m^2)	20	20

TABLE 5.26 – Results obtained for config.24

	System capacity 8 Kw	
	Without Batteries	With Batteries
Total monthly consumption	459.952 kwh	459.952 kwh
Total PV production	1071.203 kwh	1071.203 kwh
Self Consumption	216.429 kwh	427.04542 kwh
Total fed to the grid	854.7745 kwh	642.5580
Total bought from the grid	243.5227 kwh	32.9067 kwh
Self consumption rate	20.204%	39.8659%
Self sufficiency rate	47.0547%	92.8456%
Good Décision PH	342	689
Bad Décision PH	14	6
Good Décision Grid	354	35
Bad Décision Grid	33	13
Blackout error rate	3.9%	0.86%
Area (m^2)	40	40

energy consumed from the grid is also increased when the battery capacity is reduced by 2 (from 20kWh to 10kWh).

5.4.4 Impact of PV and battery size on blackout Error rate

After evaluating the impact of energy harvesting profiles and battery sizes on self-consumption rate and self-sufficiency, we analyze the behavior of the power management policies on the blackout error rate defined in equation 5.10. The lower the blackout error rate, the more efficient the PV system with or without battery.

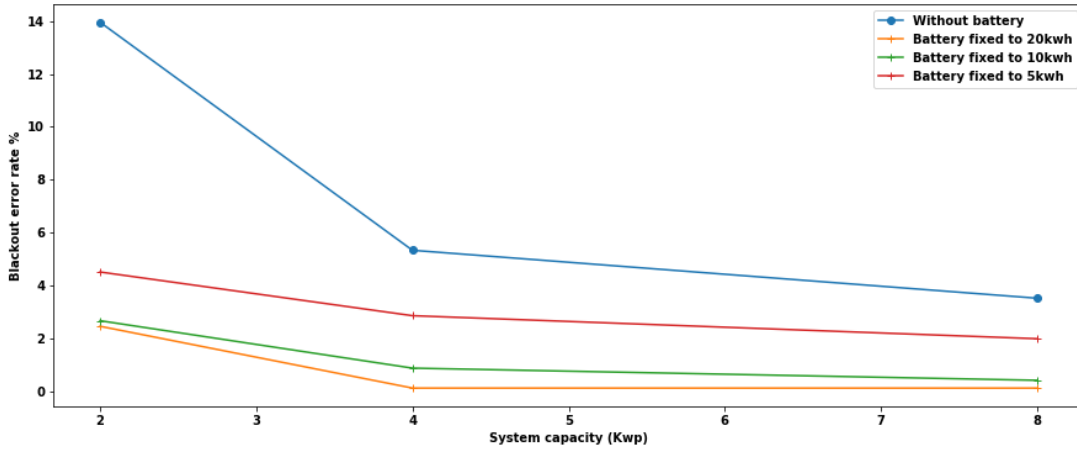


FIGURE 5.17 – Impact of PV and battery size on blackout error rate during summer conditions (July) in Nice

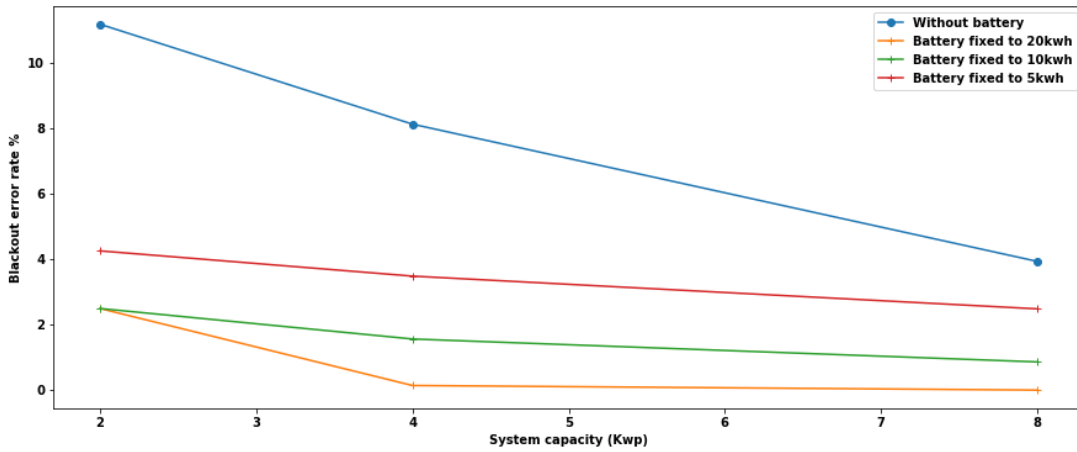


FIGURE 5.18 – Impact of PV and battery size on blackout error rate during summer conditions (July) in Paris

Figures 5.17 and 5.18 show the blackout error rate obtained in simulation by varying the PV size (2 to 8kwp) and the battery capacity (0 to 20kwh) in summer condition (i.e. July) but for two different profiles of harvested energy (i.e. in Nice and Paris). As expected, PV systems equipped with a battery have less blackout than those without battery. Moreover, there are less blackouts for large PV systems. However, for a household located in Nice, a battery size of 10 kWh with a 4kwp PV system can be considered as well adapted to cover the energy demand, while it is not the case for a house in Paris as shown in figure 5.18. These results show that how the household's localization impacts the sizing of the PV system.

5.5 Conclusion

In this chapter, household power management algorithms have been proposed to evaluate the self-consumption and optimize the blackouts error rate for a grid-connected PV system, with or without battery. In both cases, the power manager was simulated for two weather conditions (December and July). Predicted energy production and load profiles have been used to take decisions every time slot. The impact of PV and battery size has also been studied.

Based on the results obtained, we can conclude that the self-sufficiency rate strongly depends on the PV size as well as the battery capacity. However, the self-sufficiency rate makes more sense for PV systems without a storage system as a saturation is observed for PV systems with an energy storage system.

Conclusion and Perspectives

6.1 Conclusion

In this thesis work, we focused on smart homes energy management leveraging deep machine learning techniques and energy harvesting. The objective was first to develop an environment for collecting a set of relevant information (active power, reactive power, current, etc.) to propose services related to the management of household power consumption. For that purpose, a bibliographical study on data collection architectures was performed, with a particular attention to smart plugs using wireless communication protocols. The state-of-the-art reveals that two main approaches are used for data collection, either intrusive or non-intrusive.

We first proposed and developed a data collection environment based on an intrusive approach. Constructing a reliable database using this type of approach is however not a trivial task since it is required to collect measurements for different types of devices from various models over a long period. As we did not manage to build a relevant database for our purpose, we decided to use an existing dataset available online (ACSF1). To improve the identification of the different devices, a pre-processing technique based on a moving average sliding window has been proposed. Different supervised learning algorithms such as K-NN, Multilayer Perceptron as well as Random Forests were implemented to identify an electrical device with a recognition rate higher than 95%. We then also studied the impact of electrical signatures on devices identification. We were able to demonstrate that only a subset of electrical signatures (i.e. active and reactive power) can be considered to identify with good accuracy an electrical device. In a second phase, a data collection environment based on a non-intrusive approach was proposed. This method consists in measuring the overall power consumption of a household rather than each individual appliance. A non-intrusive approach thus allows visualizing the global electrical consumption of a house.

In order to improve the power management of smart home, predictions are needed. Therefore, predictions of the global energy consumption based on Deep Learning algorithms have been proposed in chapter 4. In this study, the Long Short Terms Memory (LSTM) Recurrent Neural Network algorithm has been used and compared with the ARIMA prediction model. We proposed a new technique to fill missing values from the original dataset called Same Time a Day Ago or Next (STDAN). Using this preprocessing technique, our models allowed us to accurately estimate the power consumption in the short and medium terms, from one day to one month ahead. Predictions in the short and medium terms are useful to optimize the sizing of the solar panels and the battery as well as to improve the self-consumption. We believe that a mixed approach, combining intrusive and non-intrusive techniques, represents a good compromise to manage the overall energy consumption while controlling some key appliances.

The last chapter of this thesis deals with power management algorithms with the aim at optimizing the self-consumption while reducing the number of blackouts. To do so, the algorithm switches between the grid and the energy harvesting system every time slot according to the predictions of the consumed and harvested energy. Two management policies have been proposed, one for a house without a battery, and another one when the house is equipped with a battery. The results obtained in simulations show the impact of the number of solar panels and the capacity of the battery on the self-consumption and the number of blackouts.

6.2 Perspectives

This thesis work opens several perspectives. The proposed power management policies presented in the chapter 5 focus so far on switching between the grid and the PV system. It could be also relevant to add the capability for this power manager to control some key appliances, for instance to switch them ON/OFF or shifting their activation time on advantageous periods. To do so, machine learning techniques proposed in chapter 3 for the identification of household devices could be useful. We can indeed imagine using them to automatically recognize a device plugged into a connected outlet, but also to determine whether this device is controllable or not. If so, the power consumption of these devices could be taken into account by our power management policy to increase the self-consumption.

A mixed approach, combining intrusive and non-intrusive monitoring methods, seems therefore a promising solution to improve the efficiency of power management policies. This kind of approach allows inhabitants to be informed about their global energy consumption as well as the energy consumed by some key appliances. It also offers the possibility to shift some key controllable devices on periods of time with cheaper energy prices or intervals of energy surplus.

6.2 Perspectives

So far, simulations of our power management policies have been performed only for one month. This study needs to be extended with longer periods of time. Simulating scenarios of harvested and consumed energy over a year or even several years could allow better evaluating the long-term compromise between the self-consumption and the blackout error rate of our power management policies. Additionally, only a one-hour time slot has been used so far in our prediction models and power management policies. It would be interesting to evaluate different time slot intervals, both for the predictions and the decision-making process. A perspective could be to use different time slots or even dynamically adapt the time slot duration according to the weather conditions for instance.

Smart homes will be adopted by population if the cost of the harvesting system (including a battery) is absorbed by energy savings. This study needs therefore to be extended by taking account the return on investment (ROI) related to such Grid connected PV systems. The ROI is of course tightly related to the sizing of the harvesting system.

Another perspective of this work would be to use more recent prediction algorithms (for the consumed and harvested energy) such as the combination of convolutional neural networks with LSTM (CONV-LSTM) [Shi et al., 2015] or other variations of LSTM (Gated Recurrent Unit [Cho et al., 2014], Mode Variational LSTM [Baddar and Ro, 2019]), and to compare these models in terms of performance and complexity.

In this thesis, only one energy load profile has been considered. Our prediction models as well as the power management policies need therefore to be validated and generalized with more diverse energy load profiles, considering for instance households of different size, different number of occupants, but also more geographical localities. Moreover, more parameters such as indoor and outdoor temperature, humidity or presence sensor could be considered and evaluated for the prediction models as well as for the decisions taken by the power management policies.

In the chapter 5, only mono-crystalline PV has been considered since this is currently the type of PV widely installed in homes. However, thermal solar panels also have some advantages as they are offering a mixed solar panel system. While mono-crystalline panels can generate energy from the sun to power a house, thermal panels can be used to produce domestic hot water that can be later used by a cumulus or to heat the house.

Finally, smart home power management policy can also be envisaged within a community micro-grid. Micro-grid offers indeed several benefits compared to a single distributed energy harvesting system per inhabitant. A community micro-grid allows indeed dispatching or sharing the excess of energy produced by solar panels. By sharing the energy storage device (i.e. the

battery), the installation cost could also be reduced, thus optimizing the return on investment for inhabitants. Community micro-grid seems therefore a key point for the adoption of such system by the population.

Dynamical coefficient

In [Ridi et al., 2014] [Ridi Antonio and Jean, 2013], the authors proposed a dynamical coefficient as data preparation for classification task. A dynamical coefficient has been used to extract information about the signatures dynamic evolution through the computation of velocity and acceleration coefficients.

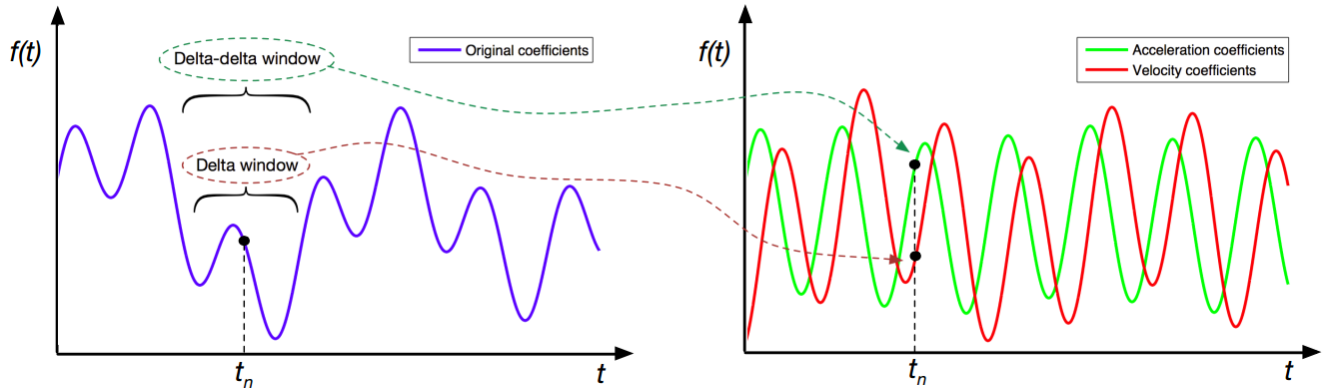


FIGURE A.1 – Computation of the dynamic coefficients for a synthetic signal. A given n-th sample value is extended through two new values computed on the delta and delta-delta windows

These two parameters are also called respectively delta and delta-delta coefficients. They are also able to capture information about transitions between the operating modes of the appliances. They are commonly used in many fields, as in speech recognition [Hennebert, 1998]. The velocity coefficient is computed as follows :

$$\Delta O_n = \sum_{w=-W}^W w \times O_{n-w} \quad (\text{A.1})$$

where W is the window length.

The acceleration coefficients can be computed from the delta coefficients :

$$\Delta\Delta O_n = \Delta O_{n+1} - \Delta O_{n-1} \quad (\text{A.2})$$

The dynamic coefficients need a window of time for their computation, as illustrated in figure A.1.

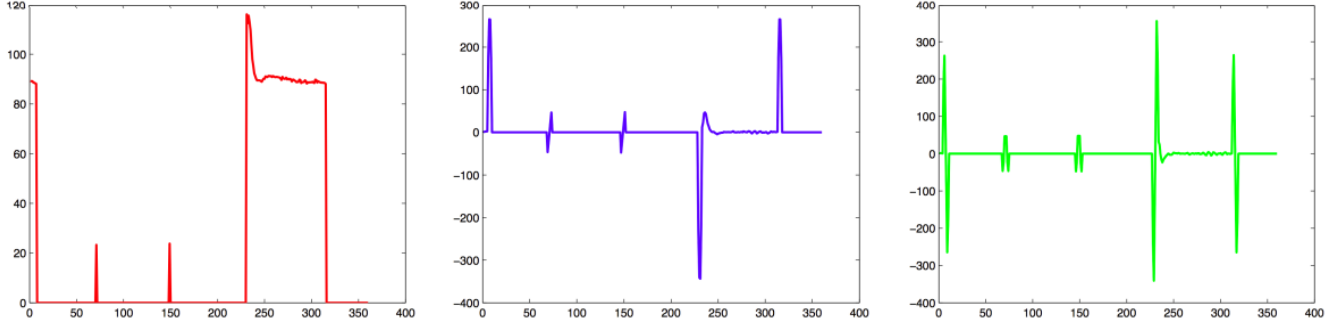


FIGURE A.2 – Example of features active power of a Fridge signature in the ACS-F1 database

The window length ($W = 2$) has been chosen in their study, which corresponding to a window of 50 seconds after some tests. The feature space has been artificially increment by three times and delta , delta- delta coefficients has also been added to the observations. The individual sample O_n is extended as follows : $O_n = \{O_n, \Delta O_n, \Delta\Delta O_n\}$. In figure A.2, the columns represent in the order the original features, the velocity coefficients and the acceleration coefficients. The velocity and acceleration coefficients are not present in the ACS-F database. However, given their extensive use in the classification task, they have been computed and added to the this figure.

Moving average

According to [Wikipedia, 2020] In statistics, a moving average (rolling average or running average) is a calculation to analyze data points by creating a series of averages of different subsets of the full data set. It is also called a moving mean (MM)[1] or rolling mean and is a type of finite impulse response filter.

Given a series of numbers and a fixed subset size, the first element of the moving average is obtained by taking the average of the initial fixed subset of the number series. Then the subset is modified by "shifting forward" ; that is, excluding the first number of the series and including the next value in the subset.It is calculated as follows :

$$MA(k) = \frac{1}{T} \sum_{n=0}^{T-1} X_{n+k} \quad (\text{A.3})$$

Where T and X_{n+k} represent the observation window size and the sampling number of an individual signature for each class, respectively in our case.

Moving Average (X):

```
Begin
    i=0
    step=1
    window_size =33
    New_matrix=[]
    Moving_average_matrix=[]
    while (i<(length(X)-window_size + step))
        begin
            Moving_average_matrix= mean(X[i:i+window_size])
            New_matrix.add(Moving_average_matrix)
            i=i+step
        End while
    final_matrix = New_matrix
return final_matrix
End
```

FIGURE A.3 – Moving average algorithm

The moving average function presented in figure A.3 was developed and applied on each feature and class and a matrix is returned. It is worth to note that the size of the observation window also has an impact on the recognition accuracy. Our experiments show that a 33-sample sliding window (corresponding to 5 minutes and 30 sec, i.e. 33×10 secs) with an overlapping window over 32 samples allows obtaining the best accuracy. A moving average is commonly used with time series data to smooth out short-term fluctuations and highlight longer-term trends or cycles.

Bibliographie

- [A. Marchiori and Earle, 2011] A. Marchiori, D. Hakkarinen, Q. H. and Earle, L. (2011). Circuitlevel load monitoring for household energy management. *IEEE Pervasive Computing*, 10 :40–48. 39
- [Alam et al., 2012] Alam, M. R., Reaz, M. B. I., and Ali, M. A. M. (2012). A review of smart homes—past, present, and future. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6) :1190–1203. 8
- [Alliance, 2020] Alliance, Z. (2020). Smart plug mini. https://zigbeealliance.org/zigbee_products/smart-plug-mini/. 37
- [Amarasinghe et al., 2017] Amarasinghe, K., Marino, D., and Manic, M. (2017). Deep neural networks for energy load forecasting. pages 1483–1488. 61
- [Anders, 2020] Anders (2020). EnOcean technology—energy harvesting wireless. <https://www.enocean.com/en/>. 41
- [Anderson et al., 2012] Anderson, K., Ocneanu, A., Benitez, D., Carlson, D., Rowe, A., and Berges, M. (2012). Blued : A fully labeled public dataset for event-based non-intrusive load monitoring research. *Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability (SustKDD)*, pages 1–5. 45
- [Antonio.Ridi et al., 2014] Antonio.Ridi, Gisler, C., and Hennebert, J. (2014). A survey on intrusive load monitoring for appliance recognition. In *2014 22nd International Conference on Pattern Recognition*, pages 3702–3707. 9, 14, 38
- [B. Afework, 2018] B. Afework, J. Hanania, K. S. B. Y. J. D. (2018). Energy education - photovoltaic cell. https://energyeducation.ca/encyclopedia/Photovoltaic_cell#cite_note-4. xv, 24
- [B. Neenan, 2009] B. Neenan, J. Robinson, R. B. (2009). Residential electricity use feedback : A research synthesis and economic framework. Technical report, United States Electric Power Research Institute. 4

- [Baddar and Ro, 2019] Baddar, W. and Ro, Y. (2019). Mode variational lstm robust to unseen modes of variation : Application to facial expression recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33 :3215–3223. 119
- [Belkin, 2020] Belkin (2020). Wemo insight smart plug 2-pack. <https://www.belkin.com/us/p/P-F7C029-BDL/>. 37
- [Bisdikian, 2001] Bisdikian, C. (2001). An overview of the bluetooth wireless technology. *IEEE Communications Magazine*, 39(12) :86–94. 41
- [Borgeson, 2018] Borgeson, S. (2018). <https://github.com/sborgeson/SAMwrapper>. 76
- [Bouaziz, 2017] Bouaziz, M. (2017). *Réseaux de neurones récurrents pour la classification de séquences dans des flux audiovisuels parallèles*. Theses, Université d’Avignon. 19
- [Box, 1979] Box, G.E.P., J. G. (1979). Time series analysis forecasting and control. *san Francisco holdenay*. 65
- [Broomhead and Lowe, 1988] Broomhead, D. and Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Syst.*, 2. 17
- [Cagigal et al., 2011] Cagigal, M. C., Martín, E., de Ávila, E. M., Bote, D., Martín, Á., Macia, F., and Leube, F. (2011). Pv self-consumption optimization with storage and active dsm for the residential sector. 74
- [Chacon, 2018] Chacon (2018). <https://chacon.com/fr/40-prises-telecommandees>. 44, 59
- [Cho et al., 2014] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. 119
- [Chujai et al., 2013] Chujai, P., Kerdprasop, N., and Kerdprasop, K. (2013). Time series analysis of household electric consumption with arima and arma models. *Lecture Notes in Engineering and Computer Science*, 2202 :295–300. 58
- [Cohen-Akne, 2014] Cohen-Akne, J. (2014). L’Électricité solaire photovoltaïque dans le bâtiment. <http://gui.sabatier.free.fr/Dataz32/Formation%20FCPV02%20161007/ADEMEHIPHIP201405.pdf>. 23
- [D. Guinard and Trifa, 2009] D. Guinard, M. W. and Trifa, V. (2009). Are you energy-efficient : Sense it on the web! *Proceedings of International Conference on Pervasive Computing, Nara, Japan*, pages 3–6. 36
- [D. Porcino, 2003] D. Porcino, W. H. (2003). Ultra-wideband radio technology : potential and challenges ahead. *IEEE Communications Magazine*, 41(7) :66–74. 42
- [Darby, 2006] Darby, S. (2006). The effectiveness of feedback on energy consumption. *A Review for DEFRA of the Literature on Metering, Billing and direct Displays*, 486. 4, 36

- [Dash7-Alliance, 2020] Dash7-Alliance (2020). Dash7 technology. https://fr.wikipedia.org/wiki/DASH7#cite_note-Mon-1. 41
- [Davies, 2002] Davies, A. C. (2002). An overview of bluetooth wireless technology/sup tm/ and some competing lan standards. In *ICCSC'02. 1st IEEE International Conference on Circuits and Systems for Communications. Proceedings (IEEE Cat. No.02EX605)*, pages 206–211. 41
- [Dobos, 2014] Dobos, A. P. (2014). *PVWatts Version 5 Manual*. 76
- [Domoticz, 2018] Domoticz (2018). <https://www.domoticz.com/>. 44, 59
- [energie+, 2010] energie+ (2010). Caractéristiques électriques des cellules et des modules photovoltaïques. <https://energieplus-lesite.be/theories/photovoltaique6/caracteristiques-electriques-des-cellules-et-des-modules-photovoltaiques/>. 23
- [Fibaro, 2020] Fibaro (2020). Wall plug. <https://www.fibaro.com/en/products/wall-plug/>. 37, 44, 59
- [Fiot and Dinuzzo, 2018] Fiot, J. and Dinuzzo, F. (2018). Electricity demand forecasting by multi-task learning. *IEEE Transactions on Smart Grid*, 9(2) :544–551. 58
- [Florian, 2019] Florian (2019). Consommation Électrique 2018 en france : Statistiques et analyses. 4, 35
- [G. Hebrail, 2012] G. Hebrail, A. B. (2012). Individual household electric power consumption data setx. <https://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption>. 45, 57, 59, 61, 68, 79
- [Gabriele, 2020] Gabriele, F. (2020). <https://www.insunwetrust.solar/blog/le-solaire-et-vous/tarif-rachat-photovoltaique/sssssss>. 74
- [Gen5, 2018] Gen5, H. E. M. (2018). <https://aeotec.com/z-wave-home-energy-measure/>. 59
- [Ghelardoni et al., 2013] Ghelardoni, L., Ghio, A., and Anguita, D. (2013). Energy load forecasting using empirical mode decomposition and support vector regression. *Smart Grid, IEEE Transactions on*, 4 :549–556. 58
- [Gisler et al., 2013] Gisler, C., Ridi, A., Zufferey, D., Khaled, O. A., and Hennebert, J. (2013). Appliance consumption signature database and recognition test protocols. In *2013 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA)*, pages 336–341. 39, 45, 46, 52
- [Grafana, 2018] Grafana (2018). <https://grafana.com/>. 44, 59
- [Gregory Shakhnarovich and Indyk, 2006] Gregory Shakhnarovich, T. D. and Indyk, P. (2006). Nearest-neighbor methods in learning and vision : Theory and practice. 54
- [G.Thomas, 2014] G.Thomas, J. (2014). User defined interactions between devices on a 6lowpan network for home automation. pages 1–4. 41

- [Hart, 1992] Hart, G. W. (1992). Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12) :1870–1891. 8, 38
- [Hebb, 2002] Hebb, D. (2002). *The Organization of Behavior : A Neuropsychological Theory*. Taylor & Francis. 17
- [Hennebert, 1998] Hennebert, J. (1998). *Hidden Markov models and artificial neural networks for speech and speaker recognition*. PhD thesis. 121
- [Hochreiter and Schmidhuber, 1997] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9 :1735–80. 21, 57, 66
- [Hongbo et al., 2016] Hongbo, Wu, Q., Gao, W., and Zhou, W. (2016). Optimal operation of a grid-connected hybrid pv/fuel cell/battery energy system for residential applications. *Energy*, 113 :702 – 712. 74
- [Hornik et al., 1989] Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5) :359–366. 18
- [HULD Thomas, 2008] HULD Thomas, SURI Marcel, C. T. D. E. (2008). Optimal mounting strategy for single-axis tracking non-concentrating pv in europe. 76
- [industry, 2019] industry, B. (2019). <https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy/primary-energy.html>. xv, 2
- [J. Kelly, 2014] J. Kelly, W. J. K. (2014). 'uk-dale' : A dataset recording UK domestic appliance-level electricity demand and whole-house demand. *CoRR*, abs/1404.0284. 45
- [Jetcheva et al., 2014] Jetcheva, J., Majidpour, M., and Chen, W.-P. (2014). Neural network model ensembles for building-level electricity. *Energy and Buildings*, 84 :214–223. 56, 57
- [J.Lee, 2009] J.Lee, C.Chuang, C. S. (2009). Applications of short-range wireless technologies to industrial automation : A zigbee approach. In *2009 Fifth Advanced International Conference on Telecommunications*, pages 15–20. 42
- [(Kanevce) Dedinec et al., 2016] (Kanevce) Dedinec, A., Filiposka, S., Dedinec, A., and Kocarev, L. (2016). Deep belief network based electricity load forecasting : An analysis of macedonian case :. *Energy*. 58
- [Kim et al., 2018] Kim, N., Kim, M., and Choi, J. (2018). Lstm based short-term electricity consumption forecast with daily load profile sequences. pages 136–137. 58
- [Kingma and Ba, 2014] Kingma, D. and Ba, J. (2014). Adam : A method for stochastic optimization. *International Conference on Learning Representations*. 58, 65, 67
- [Knight, 2007] Knight, M. (2007). Wireless security - how safe is z-wave? *Computing Control Engineering Journal*, pages 18 – 23. 42

- [Kohonen, 1982] Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1) :59–69. [17](#)
- [Kolen and Kremer, 2001] Kolen, J. F. and Kremer, S. C. (2001). *Gradient Flow in Recurrent Nets : The Difficulty of Learning LongTerm Dependencies*, pages 237–243. [21](#)
- [Kolter J., 2011] Kolter J., J. M. (2011). Redd : A public data set for energy disaggregation research. *Artif. Intell.*, 25. [45](#)
- [L. Harrington and Ellis, 2008] L. Harrington, H.-P. S. and Ellis, M. (2008). Standby power : Building a coherent international policy framework. *Proc. of American Council for an Energy-Efficient Economy Summer Study. Pacific Grove, California*, pages 1–16. [39](#)
- [LeCun, 1985] LeCun, Y. (1985). Une procedure d’apprentissage pour reseau a seuil asymmetrique (a learning scheme for asymmetric threshold networks). [18](#)
- [Lecun et al., 1998] Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11) :2278–2324. [17](#)
- [Leitão et al., 2020] Leitão, J., Gil, P., Ribeiro, B., and Cardoso, A. (2020). A survey on home energy management. *IEEE Access*, 8 :5699–5722. [8](#)
- [Luthander et al., 2016] Luthander, R., Widén, J., Munkhammar, J., and Lingfors, D. (2016). Self-consumption enhancement and peak shaving of residential photovoltaics using storage and curtailment. *Energy*, 112 :221 – 231. [75](#)
- [M. Weiss and Roediger, 2009] M. Weiss, T. Staake, D. G. and Roediger, W. (2009). emeter : An interactive energy monitor. *Adjunct Proceedings of UbiComp 2009, 11th International Conference on Ubiquitous Computing (UbiComp) UbiComp 2009 Orlando FL USA*, pages 3–4. [36](#)
- [Manic et al., 2016] Manic, M., Amarasinghe, K., Rodriguez-Andina, J., and Rieger, C. (2016). Intelligent buildings of the future : Cyberaware, deep learning powered, and human interacting. *IEEE Industrial Electronics Magazine*, Vol. 10 :32–49. [56](#)
- [Marinescu et al., 2013] Marinescu, A., Harris, C., Dusparic, I., Clarke, S., and Cahill, V. (2013). Residential electrical demand forecasting in very small scale : An evaluation of forecasting methods. pages 25–32. [57](#), [58](#)
- [McCulloch and Pitts, 1943] McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4) :115–133. [16](#), [17](#)
- [Mocanu et al., 2016] Mocanu, E., Nguyen, H., Gibescu, M., and Kling, W. (2016). Deep learning for estimating building energy consumption. *Sustainable Energy, Grids and Networks*, 6 :91–99. [57](#)
- [Monacchi et al., 2014] Monacchi, D. E., Elmenreich, W., D’Alessandro, S., and Tonello, A. M. (2014). GREEND : an energy consumption dataset of households in italy and austria. *CoRR*, abs/1405.3100. [xiii](#), [45](#)

- [MultiSensor6, 2018] MultiSensor6 (2018). <https://aeotec.com/z-wave-sensor/>. 44, 59
- [Nambi et al., 2013] Nambi, A., Papaioannou, T., Chakraborty, D., and Aberer, K. (2013). Sustainable energy consumption monitoring in residential settings. pages 3177–3182. 10
- [Neenan, 2009] Neenan, B. (2009). Residential electricity use feedback : A research synthesis and economic framework. 36
- [Nguyn, 2015] Nguyn, T. K. (2015). Conception et réalisation d’un système de gestion intelligente de la consommation électrique domestique. xv, 8, 10
- [Nipun et al., 2013] Nipun, B., Gulati, M., Singh, A., and Srivastava, M. (2013). It’s different : Insights into home energy consumption in india. 45
- [NodOn, 2020] NodOn (2020). Wall plug. <https://nodon.fr/nodon/prise-intelligente-enocean/>. 37
- [NREL, 2020] NREL (2020). <https://www.nrel.gov/>. 76
- [Parker, 1985] Parker, D. (1985). *Learning-logic : Casting the Cortex of the Human Brain in Silicon*. Technical report : Center for Computational Research in Economics and Management Science. Massachusetts Institute of Technology, Center for Computational Research in Economics and Management Science. 18
- [Pi, 2018] Pi, R. (2018). <https://www.raspberrypi.org/>. 44, 59
- [Plaut et al., 1986] Plaut, D., Nowlan, S., and Hinton, G. E. (1986). Experiments on learning by back propagation. 20
- [Ranaweera and Midtgård, 2016] Ranaweera, I. and Midtgård, O.-M. (2016). Optimization of operational cost for a grid-supporting pv system with battery storage. *Renewable Energy*, 88 :262 – 272. 75
- [Reinhardt et al., 2012] Reinhardt, A., Baumann, P., Burgstahler, D., Hollick, M., Chonov, H., Werner, M., and Steinmetz, R. (2012). On the accuracy of appliance identification based on distributed load metering data. *Proc. SustainIT*, pages 1–9. 45
- [Ren et al., 2016] Ren, Z., Grozev, G., and Higgins, A. (2016). Modelling impact of pv battery systems on energy consumption and bill savings of australian houses under alternative tariff structures. *Renewable Energy*, 89 :317 – 330. 75
- [Ridi et al., 2014] Ridi, A., Gisler, C., and Hennebert, J. (2014). Acs-f2 — a new database of appliance consumption signatures. In *2014 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR)*, pages 145–150. 36, 45, 49, 121
- [Ridi Antonio and Jean, 2013] Ridi Antonio, G. C. and Jean, H. (2013). Automatic identification of electrical appliances using smart plugs. pages 301–305. 36, 39, 46, 47, 52, 121
- [Roldán-Blay et al., 2013] Roldán-Blay, C., Escrivá, G., Alvarez, C., Roldán-Porta, C., and Rodríguez-García, J. (2013). Upgrade of an artificial neural network prediction method for

- electrical consumption forecasting using an hourly temperature curve model. *Energy and Buildings*, 60 :38–46. 57
- [Ropert, 2016] Ropert, S. (2016). <https://en.idate.org/internet-of-things-2/>. xv, 3, 4
- [Rosenblatt, 1957] Rosenblatt, F. (1957). *The Perceptron, a Perceiving and Recognizing Automaton Project Para.* Report : Cornell Aeronautical Laboratory. Cornell Aeronautical Laboratory. 17
- [Ross and Meier, 2001] Ross, J. P. and Meier, A. (2001). Whole-house measurements of standby power consumption. In Bertoldi, P., Ricci, A., and de Almeida, A., editors, *Energy Efficiency in Household Appliances and Lighting*, pages 278–285, Berlin, Heidelberg. Springer Berlin Heidelberg. 39
- [Rumelhart et al., 1986] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088) :533–536. 18
- [Rumelhart and McClelland, 1987] Rumelhart, D. E. and McClelland, J. L. (1987). *Learning Internal Representations by Error Propagation*, pages 318–362. 18, 19
- [Saab et al., 2001] Saab, S., Badr, E., and Nasr, G. (2001). Univariate modeling and forecasting of energy consumption : The case of electricity in lebanon. *Energy*, 26 :1–14. 57
- [Santos et al., 2014] Santos, J. M., Moura, P. S., and de Almeida, A. T. (2014). Analysis of the distributed residential energy storage impact on the grid operation. In *2014 IEEE/IAS 50th Industrial Commercial Power Systems Technical Conference*, pages 1–7. 75
- [Sendy, 2020] Sendy, A. (2020). Solarreviews - types of solar panels : which one is the best choice. <https://www.solarreviews.com/blog/pros-and-cons-of-monocrystalline-vs-polycrystalline-solar-panels>. 25
- [Sharma et al., 2016] Sharma, I., Dong, J., Malikopoulos, A. A., Street, M., Ostrowski, J., Kuruganti, T., and Jackson, R. (2016). A modeling framework for optimal energy management of a residential building. *Energy and Buildings*, 130 :55 – 63. 75
- [Shi et al., 2015] Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W. K., and WOO, W.-c. (2015). Convolutional lstm network : A machine learning approach for precipitation nowcasting. 119
- [Siame-Namini and Namin, 2018] Siame-Namini, S. and Namin, A. S. (2018). Forecasting economics and financial time series : Arima vs. lstm. 58
- [Song et al., 2011] Song, I., Jung, D., Ji, Y., Choi, S., Lee, S., and Won, C. (2011). A residential 10kwh lithium-polymer battery energy storage system. In *8th International Conference on Power Electronics - ECCE Asia*, pages 2625–2630. 75
- [Statista, 2020] Statista (2020). <https://www.statista.com/statistics/809796/global-solar-power-installation-cost-per-kilowatt/>. xv, 2, 3
- [Sulaiman et al., 2016] Sulaiman, S., Jeyanthi, A., and Devaraj, D. (2016). Artificial neural network based day ahead load forecasting using smart meter data. pages 1–6. 57

- [Sun, 2019] Sun, X. (2019). <https://www.greentechmedia.com/articles/read/solar-pv-has-become-cheaper-and-better-in-the-2010s-now-what>. xv, 3
- [Suzuki et al., 2008] Suzuki, K., Inagaki, S., Suzuki, T., Nakamura, H., and Ito, K. (2008). Nonintrusive appliance load monitoring based on integer programming. volume 128, pages 2742 – 2747. 10
- [Swamidass, 2000] Swamidass, P. M., editor (2000). *MAPE (mean absolute percentage error) MEAN ABSOLUTE PERCENTAGE ERROR (MAPE)*, pages 462–462. Springer US, Boston, MA. 68
- [Thokala et al., 2018] Thokala, N. K., Bapna, A., and Chandra, M. G. (2018). A deployable electrical load forecasting solution for commercial buildings. In *2018 IEEE International Conference on Industrial Technology (ICIT)*, pages 1101–1106. 57, 58
- [Tsang, 2008] Tsang, K.F., W. L. L. K. T. H. X. K. (2008). An integrated zigbee automation system : An energy saving solution. pages 252 – 258. 42
- [Vahid Mansouri, 2014] Vahid Mansouri, M. E. A. (2014). Efficient short-term electricity load forecasting using recurrent neural networks. *Artificial Intelligence in Electrical Engineering*, 3. 57
- [Werbos, 1970] Werbos, P. (1970). *Applications of advances in nonlinear sensitivity analysis*, volume 38, pages 762–770. 18
- [Werbos, 1974] Werbos, P. (1974). *Beyond Regression : New Tools for Prediction and Analysis in the Behavioral Science. Thesis (Ph. D.). Appl. Math. Harvard University*. PhD thesis. 18
- [Wikipedia, 2020] Wikipedia (2020). https://en.wikipedia.org/wiki/Moving_average. 122
- [X.Jia, 2012] X.Jia, Q. Feng, T. F. Q. L. (2012). Rfid technology and its applications in internet of things (iot). In *2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*, pages 1282–1285. 42
- [X.Ma, 2008] X.Ma, W. (2008). The analysis of 6lowpan technology. In *2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, volume 1, pages 963–966. 41
- [Yiu, 2019] Yiu, T. (2019). Understanding random forest. xv, 13
- [Zafari et al., 2016] Zafari, F., Papapanagiotou, I., and Christidis, K. (2016). Microlocation for internet-of-things-equipped smart buildings. *IEEE Internet of Things Journal*, 3(1) :96–112. 8
- [Zhu et al., 2012] Zhu, Q., Guo, Y., and Feng, G. (2012). Household energy consumption in china : Forecasting with bvar model up to 2015. pages 654–659. 57
- [Zufferey et al., 2012] Zufferey, D., Gisler, C., Khaled, O. A., and Hennebert, J. (2012). Machine learning approaches for electric appliance classification. In *2012 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA)*, pages 740–745. 14, 39, 46

