



UNIVERSITÀ DEGLI STUDI DI GENOVA

POLYTECHNIC SCHOOL

**Department of Electrical, Electronics,
Telecommunication Engineering
and Naval Architecture**

INTELLIGENT ELECTRIC ENERGY SYSTEMS LABORATORY

XXXIV Cycle Ph.D in Sciences and Technologies for Electrical
Engineering, Naval Engineering, and Complex Systems for
Mobility - Curriculum: Electrical Engineering

**Simulation, forecasting, and control
in power system analytics:
methodological aspects and applications**

Gabriele Mosaico

Supervisor

Prof. Ing. Stefano Massucco

Co-Supervisor

Prof. Matteo Saviozzi

February, 2022

Abstract

Climate change driven by fossil fuels consumption demands a quick energy transition to a carbon-neutral society. One of the most affected systems by this paradigm shift is, without a doubt, the electric power system.

Once designed with few, typically high pollutant production nodes on the transmission side, production is also located in the distribution side, with a very high number of Distributed Energy Resources (DERs), primarily renewable. This change, however, comes with a cost: many, intermittent small resources are hard to manage. They cannot always guarantee the meeting of demand, creating difficulties to the reliability, security, and power quality of the system.

As a result, the demand side is becoming more and more active, as its contribution will be essential in addressing the new challenges. Moreover, diverse storage systems have been recently researched, prototyped, and commercialized that allow even more flexible management of the system, also creating other types of loads such as Electrical Vehicles (EVs). Such loads will modify load patterns in ways difficult to anticipate, which poses serious questions on how to reinforce the current distribution and transmission assets. Moreover, the role of the electricity markets, once liberalized to ensure low energy prices to all the consumers, are becoming a central tool for system operators. Indeed, markets are expected to procure all the flexibility needed to meet unexpected generation, load, and infrastructural contingencies.

In this context, simulation, forecasting, and automated decision-making tools derived from business and data analytics are becoming crucial for modern power systems design, planning, and operation.

During the Doctorate (Ph.D.), the author has worked on research projects such as "Adaptive Energy Efficiency Platform For Consumption Reduction In Non-Residential Buildings" (PREDICT) and "Distribution Optimization Platform Through The Use Of Data From Electronic Meters And Distributed Storage Systems" (PODCAST), intending to develop data analytics tools better to manage distributed resources and all the related challenges.

This thesis homogenizes the work done by framing it in the context of two great revolutions: one is the modern power system revolution, and the other is the advent of the analytics revolution, which has been possible thanks to the emergence of Big Data (BD) paradigm.

The developed applications will be framed as particular cases of analytics subfields (descriptive, diagnostic, predictive, or prescriptive) after detailing the methodological aspects of the involved analytics branch that have much broader use than the single applications presented.

In **Chapter 1**, this work is introduced by presenting the two current paradigm shifts of energy transition and data analytics and explaining how the two are interacting in the context of power systems. Some key definitions are laid out that will help to clarify the concepts explained in the following chapters.

In **Chapter 2**, descriptive and diagnostics analytics and their possible applications to power systems are presented in the context of power systems. As an application, the proposed Instantaneous Growing Stream Clustering (IGSC) algorithm, devised during the Ph.D., is illustrated, which allows for flexible probabilistic modeling and simulation of quantities of interest via light online and adaptive clustering.

In **Chapter 3**, predictive analytics, with a particular focus on forecasting, is presented, by following the Cross-Industry Standard Procedure for Data Mining (CRISP-DM) framework, together with some of its many applications present in the literature of power systems.

Applications of distribution network load forecasting, building energy forecasting, and Photovoltaic (PV) short-term forecasting are presented, all developed in the context of the projects PREDICT and PODCAST.

In **Chapter 4**, the final step of analytics, that is, prescriptive analytics, is described together with some of the applications to the power system field. Battery profile optimization and optimal sizing and siting, developed as a functionality of a modern Distribution Management System (DMS), are the applications of this chapter and are drawn from the PODCAST project.

Afterward, in the **Conclusions**, the key findings are summarized together with some final comments.

Finally, the publications produced during the Ph.D., and contributed projects, collaborations, and attended courses, are listed, while references cited in this thesis conclude the dissertation.

Contents

Abstract	iii
1 Introduction: Power Systems and Data Analytics	9
1.1 The Evolution of Power Systems	9
1.1.1 Contemporary Power Systems	11
1.2 Data Analytics	17
1.2.1 Big Data	18
1.2.2 Artificial Intelligence	20
1.2.3 Machine Learning	21
1.2.4 Deep Learning	22
1.2.5 Expert system	23
1.2.6 Analytics	23
1.3 Smart Power Systems	24
1.3.1 Possible Applications	28
1.4 This Thesis	30
2 Descriptive and Diagnostic Power System Analytics	35

2.1	General Methodological Aspects	35
2.1.1	Descriptive & Diagnostic Analytics Approaches . .	36
2.1.1.1	Visualization	36
2.1.1.2	Interpretability	38
2.1.1.3	Simulation	39
2.2	Descriptive and Diagnostic Applications in Power Systems .	41
2.3	Application: Instantaneous Growing Stream Clustering . . .	47
2.3.1	Literature Review	48
2.3.2	Proposed Approach	50
2.3.3	Mathematical Background - Discrete Markov Chains	51
2.3.4	Designing the Algorithm	52
2.3.5	The Instantaneous Growing Stream Clustering Algo- rithm	53
2.3.6	Application	55
2.3.7	Study Cases	56
2.3.8	KPIs and Reference Methods	58
2.3.9	Experimental Results	61
2.3.9.1	First set of Experiments	61
2.3.9.2	Second set of Experiments	68
2.4	Chapter Conclusions	77
3	Predictive Power System Analytics	79
3.1	Cross-Industry Standard Procedure for Data Mining	81
3.1.1	Business Understanding	82
3.1.2	Data Understanding	86
3.1.3	Data Preparation	87
3.1.4	Modeling	90
3.1.5	Evaluation	93
3.1.6	Deployment	95
3.2	Predictive Applications to Power Systems	95
3.2.1	Common Issues	96
3.2.2	PV Forecasting Applications	96
3.2.3	Load Forecasting	98
3.3	Application: PV Forecasting	100
3.3.1	Key Performance Indicators - KPI	102

3.3.2	Proposed Approach	103
3.3.3	Clear Sky Model - CSM	105
3.3.4	Corrected Clear Sky Model - CCSM	107
3.3.5	Neural Network Technique	108
3.3.5.1	Input Variables	108
3.3.5.2	Hyperparameters Selection	110
3.3.5.3	Basic Ensemble Method - BEM	110
3.3.6	Hybrid Technique for the PV Forecasting Technique Selection	111
3.3.6.1	Selection with Decision Trees	112
3.3.6.2	Selection with Linear Regression	116
3.3.7	Test Site	119
3.3.7.1	Shading Modeling	119
3.3.8	Available Data	122
3.3.9	Results: Decision Tree Hybrid Procedure	123
3.3.10	Results: Linear Regression Based Hybrid Procedure	127
3.4	Application: Distribution Network Load Forecasting	130
3.4.1	PODCAST Grid description	130
3.4.2	KPI: Mean Absolute Percentage Error	131
3.4.3	The Load Forecasting Algorithm	131
3.5	Application: Non-Commercial Building Energy Prediction	138
3.5.1	Test Site	140
3.5.2	Monitoring and Control System	141
3.5.3	Building Occupancy	143
3.5.4	Occupancy Estimation	143
3.5.5	Occupancy Prediction	145
3.5.6	Building Energy Model	148
3.5.7	Inputs	148
3.5.8	Output	149
3.5.9	Key Performance Indicators	149
3.5.10	Experimental Validation	150
3.5.11	Occupancy Estimation	150
3.5.12	Building Energy Model	152
3.6	Chapter Conclusions	154

4 Prescriptive Power System Analytics	157
4.1 Approaches to Prescriptive Analytics	164
4.1.1 Mathematical Programming	165
4.1.2 Optimization under Uncertainty	167
4.1.3 Machine Learning Tools	169
4.1.4 Other Prescriptive Analytics Tools	172
4.2 Prescriptive Applications to Power Systems	173
4.3 Application: Optimal Storage Siting and Sizing	177
4.3.1 Optimization Problem Formulation	178
4.3.2 Real Network Model	180
4.3.3 Grid Description	181
4.3.4 Load Modeling for Monte Carlo simulation	183
4.3.5 Simulation Results	186
4.4 Application: Microgrid/Energy Community Profile Optimiza- tion Algorithm	189
4.4.1 Day-ahead Stage	190
4.4.2 Intra-day Stage	195
4.4.3 Experiments	196
4.4.3.1 Battery Energy Storage System of the POD- CAST project	196
4.4.3.2 Chosen Parameters	201
4.4.4 Round-Trip Efficiency Estimation	201
4.4.4.1 First Experiment	203
4.4.4.2 Second Experiment	207
4.5 Chapter Conclusions	210
5 Conclusions	213
Bibliography	217
List of Publications	253
List of Projects	255
List of Courses	259

Acronyms

ABB Lab ABB Marine Laboratory

AHU Air Handling Unit

AI Artificial Intelligence

AIC Akaike Information Criterion

AMI Advanced Metering Infrastructure

ANN Artificial Neural Network

API Application Programming Interface

AR Augmented Reality

aSAX Adaptive Symbolic Aggregate approXimation

AutoML Automated Machine Learning

BD Big Data

BEM Basic Ensemble Method of Neural Networks

- BEMS** Building Energy Management System
- BESS** Battery Energy Storage System
- BI** Business Intelligence
- BIC** Bayesian Information Criterion
- CART** Classification And Regression Trees
- CASH** Combined Algorithm Selection and Hyperparameter
- CBM** Condition Based Monitoring
- CCI** Cloud Cover Index
- CCO** Chance Constrained Optimization
- CCSM** Corrected Clear Sky Model
- CHI** Calinski-Harabasz Index
- CO** Combinatorial Optimization
- CO₂** Carbon Dioxide
- CRISP-DM** Cross-Industry Standard Procedure for Data Mining
- CSM** Clear Sky Model
- DBI** Davies-Bouldin Index
- DER** Distributed Energy Resource
- DeRu** Decision Rule
- DG** Distributed Generators
- DISFOR** Education Sciences Department
- DITEN** Department of Electrical Electronic Telecommunication and Naval
Engineering
- DL** Deep Learning
- DMC** Discrete Markov Chain

DMS	Distribution Management System
DNN	Deep Neural Network
DRL	Deep Reinforcement Learning
DRO	Distributionally Robust Optimization
DSO	Distribution System Operator
DT	Digital Twin
DV	Data Visualization
DyCl	Dynamical Clustering
EB	Electrical Bus
EDA	Exploratory Data Analysis
EMS	Energy Management System
EPM	Estimated Physical Memory
EU	European Union
EV	Electrical Vehicle
EVSE	Electrical Vehicle Supply Equipment
FA	Factor Analysis
FMS	Full Model Selection
GEFCom2017	Global Energy Forecasting COMpetition 2017
GHMM	Growing Hidden Markov Model
GNG	Growing Neural Gas
HMM	Hidden Markov Model
HOG	Histogram of Oriented Gradients
HPO	HyperParameters Optimization

- HVAC** Heat Ventilation Air Conditioning
- IEA** International Energy Agency
- IGSC** Instantaneous Growing Stream Clustering
- IID** Independent and Identically Distributed
- IMDELD** IEEE Industrial Machine Dataset for Electrical Load Disaggregation
- IoT** Internet of Things
- ISO** Independent System Operator
- ITM** Instantaneous Topological Mapping
- kNN** k Nearest-Neighbors
- KPI** Key Performance Indicator
- LOA** Level Of Automation
- LP** Linear Programming
- MAE** Mean Absolute Error
- MAPE** Mean Absolute Percentage Error
- MAR** Missing At Random
- MASE** Mean Absolute Scaled Error
- MBE** Mean Bias Error
- MC** Markov Chain
- MCAR** Missing Completely At Random
- MCP** Market Clearing Price
- MDP** Markov Decision Process
- MDS** Multi Dimensional Scaling
- MILP** Mixed Integer Linear Programming

-
- MIMO** Multi-Input Multi-Output
 - ML** Machine Learning
 - MLP** Multi-Layer Perceptron
 - MNAR** Missing Not At Random
 - MOS** Model Output Statistic
 - MP** Mathematical Programming
 - MPC** Model Predictive Control
 - MPP** Multi Parametric Programming
 - MPT** Maximum Processing Time
 - MVU** Medium Voltage User
 - NAS** Neural Architecture Search
 - NCI** Number of Clusters Identified
 - NILM** Non-Intrusive Load Modeling
 - NMT** Number of Markovian Transition
 - nRMSE** normalized Root Mean Square Error
 - NWP** Numerical Weather Prediction
 - OPC-UA** Open Platform Communications - Unified Architecture
 - OR** Operations Research
 - PCA** Principal Component Analysis
 - Ph.D.** Doctorate
 - PID** Proportional–Integral–Derivative
 - PLF** Piecewise Linear Function
 - PMU** Phasor Measurement Unit

PODCAST "Distribution Optimization Platform Through The Use Of Data From Electronic Meters And Distributed Storage Systems"

PREDICT "Adaptive Energy Efficiency Platform For Consumption Reduction In Non-Residential Buildings"

PSO Particle Swarm Optimization

PV Photovoltaic

RELRMSE Relative Root Mean Square Error

RES Renewable Energy Source

RL Reinforcement Learning

RMSE Root Mean Square Error

RO Robust Optimization

SCADA Supervisory Control And Data Analysis

SFTP Secure File Transfer Protocol

SG Smart Grid

SL Supervised Learning

SL Supervised Learning

SLR Stepwise Linear Regression

SM Smart Manufacturing

sMAPE Symmetric Mean Absolute Percentage Error

SoC State of Charge

SOM Self-Organizing Maps

SP Stochastic Programming

SS Skill Score

SUB MV/LV substation

SubM Sub Model

SVR Support Vector Regression

t-SNE t-Stochastic Neighbor Embedding

TCP Transmission Control Protocol

TL Transfer Learning

VR Virtual Reality

VVO Volt/Var Optimization

Introduction: Power Systems and Data Analytics

*AI is the new electricity.
- Prof. Andrew Ng*

This chapter will briefly outline the historical evolution of the power system and data analytics systems. Then it will be shown how recent years have nudged power systems into being more and more intelligent by artificial intelligence. Also, the terminology used in the rest of the thesis will be established, and a detailed outline of the thesis will be given.

1.1 The Evolution of Power Systems

Natural phenomena involving electricity have been observed since ancient times. Only in the modern age were they studied using the experimental method. The first applications of these discoveries were for games, pranks, or fraudulent advertising. But slowly, industrial applications began to emerge, to the point that "Electric city" soon became synonymous with progress [1].

At the end of the 1800s, gas turbines became very common in the industrial sector for self-production, together with hydropower plants. Urban transport was quickly revolutionized by the electrified tram lines in the same years. Also, private transport was electrified soon [2]. In the same years, With the Great War, it became evident that electricity was central to industrial development and military supremacy.

After the second world war, an unprecedented increase in demand began to occur due to demographic and economic growth. Demand roughly doubled every ten years in developing countries, without stopping with financial and energetic crises. Nuclear plants began to be built nearly everywhere because the amount of resources and reserves of oil and natural gas was estimated to last very little at the time.

Most importantly, demand for large blocks of power, increased reliability, and economic convenience suggested a broader interconnection, typically state-wide. More interconnection brought new problems: new types of breakers and control systems for ensuring frequency synchronicity among all generators were invented.

In 1973, the oil embargo crisis gave impulse to the search of Renewable Energy Source (RES) alternatives to fossil fuels, with states leading a series of long-term energy planning initiatives for achieving energy independence [3]. Moreover, also nuclear power was receiving less support from the public.

PV electricity, previously employed only for niche applications in outer space exploration, began to be applied to a wider range of applications, even though it took off only in the 2000s [4]. Also, wind power plants were an integral part of these plans [3, 5]. Wave and tidal energy conversion were experimented, although with much smaller success [6].

In the same years, the first scientific signs of human-induced global warming [7] made clean sources and efficiency at the generation, transport, and demand level very crucial policy topics.

In the 80s, thanks to the rise of pro-capita income, rigid prices for electricity were less and less relevant [1]. Also, the demand growth started to decrease, so new investments in generation plants with a high concentration of power (such as nuclear or coal plants) were less convenient. Consequently, new generation plants with capital intensity, like combine cycles, became affordable. More economic agents could now contribute to the evolution of the power systems. Together with a change of the dominant economic ideas, these facts lead to the liberalization of electricity markets, with the objective of decreasing energy prices.

The first liberalized electricity market was the UK, whose market started to be liberalized in 1988 and finished in 2005. In the European Union, the 96/92/CE directive of December 19th, 1996, introduced minimum elements of liberalization, which were:

- Division of balances of state-owned companies active in the sectors (not necessarily privatization);
- The presence of an Independent System Operator (ISO) of the transmission grid, needed for fair access to the market;
- Distinction in the market between final consumers or intermediates (vendors or distributors).
- A Unique Acquirer for agents not participating in the liberalized market.

Hence several drivers that lead to the evolution of contemporary power systems can be identified: industrial development, climate change, electricity markets, concerns about fossil fuel resources availability, and an increase in demand for a more electrified society.

1.1.1 Contemporary Power Systems

Nowadays, the energy sector (of which electric power systems are one of the major components) is deemed responsible for more than half of the emissions that caused global warming, with respect to the preindustrial era [8].

The current climate science consensus identified by the Intergovernmental Panel on Climate Change affirms that [9]:

- Human influence on recent climate change is unequivocal (see Figure 1.1);
- Climate change affects many weather and climate extremes in every region across the globe with increasing evidence;
- Many of these changes are irreversible or reversible on a century or even millennial-scale because they are linked to cumulative concentrations of Greenhouse gases in the atmosphere (especially, Carbon Dioxide (CO₂)).

- These changes have negative effects for all ecosystems;
- These changes are not uniform across the region of the globe;
- For limiting these changes, which have obvious negative effects for all ecosystems, reaching zero net emissions of CO₂ is crucial, together with limitations of other Greenhouse gasses;
- the decarbonization of the electricity demand is a key measure for a better outlook [10] (see also Figure 1.2).

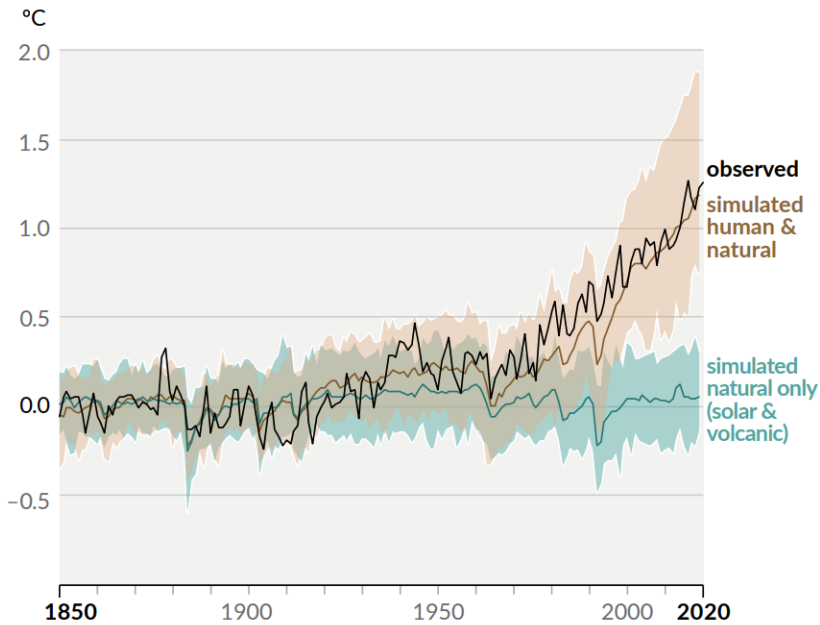


Figure 1.1: Changes in annually averaged global surface temperature over the past 170 years (**black** line) relative to 1850-1900, compared to climate model simulations [11] with human factors (in **brown**) and without human factors (in **green**). Solid colored lines show multi-model averages, while colored shades indicate the 90% central range of simulations [9].

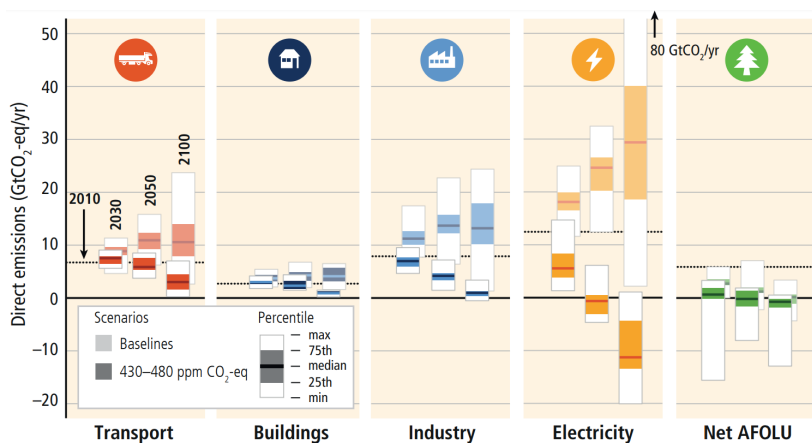


Figure 1.2: Direct CO₂ emissions by sector in baselines (solid bars) and mitigation scenarios that likely will limit warming to 2°C above preindustrial levels (faded bars). All end-use sectors have potential of reducing emissions, but the electric energy sector has the highest potential of all [10]. AFOLU stands for Agricultural, Forest and Other Land Uses.

On the other hand, the expansion of the power system is one of the foremost necessary conditions for the improvement of livelihoods in developing countries [12, 13]. Moreover, fossil fuels resources are not distributed uniformly across nations. This context calls for deployment and integration of greener, more widely available RESs, which, together with storage systems, can contribute to wider participation in flexibility markets.

The tremendous growth in wind and solar PV can be observed by the graphics reported in Figures 1.3–1.5.

Figure 1.3 reports the increasing share of new capacity expansion being nowadays dominated by renewables.

Also, Figure 1.4 testifies how PV and wind plants constitute the most consistent growth in capacity (in GW).

Finally, Figure 1.5 depicts the same phenomenon from the energy generated point of view. However, it points out that most of the share is still due to hydro energy.

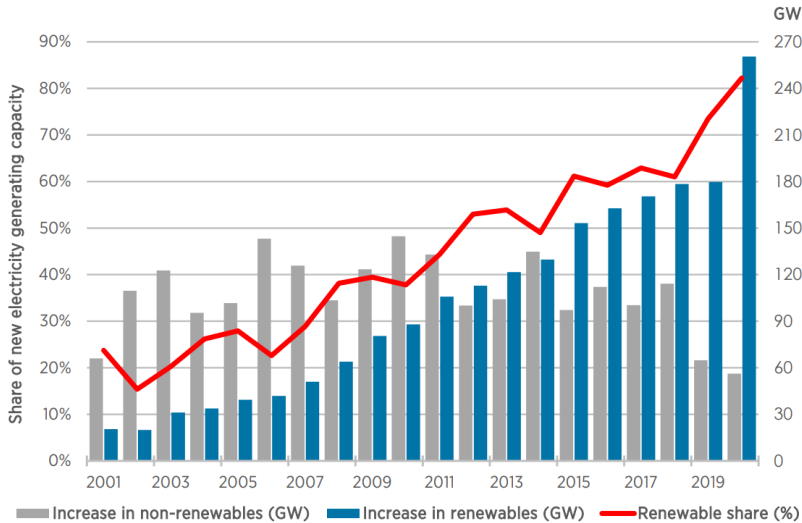


Figure 1.3: Renewable share of global annual power capacity expansion [14]

However, this decarbonization process, which pivots on the integration of RESs such as PV and wind power plants, is strictly tied with an increasing complexity [16].

This complexity has many causes.

Firstly, RESs have low or no inertial response since they are mostly interfaced with the grid via electronic inverters. On the other hand, main traditional power sources such as coal, nuclear, hydroelectric power, and natural gas are characterized by synchronous generators, which can change their rotational speed thanks to stored kinetic energy. This inertia can help meet frequency imbalances due to abrupt load or generation variations, as well as grid faults in an automatic way, via suitable control systems. With the increasing penetration of renewables, the inertia of the system is lower and lower, and grid operators have more challenges in keeping the system balanced [17].

Secondly, unlike traditional generation sources, which are dispatchable, RES are variable, uncertain, and undispachable, so backup sources (spinning reserves) must be ready to provide the energy missing from RES at any given

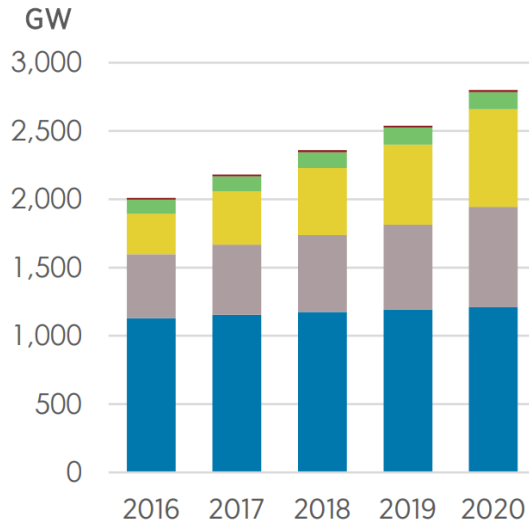


Figure 1.4: Renewable power worldwide growth. RES types: in blue hydropower, in gray wind power, in yellow solar, in green bioenergy, and in red geothermal [14]

time. Undispatchability can also lead to the opposite problem, as RES generation can happen to be higher than demand. In this case, either the RES is stored or curtailed. In the former case, the amount of storage capacity needed can be prohibitive; in the latter case, clean energy is lost.

Thirdly, most RES (especially solar) are connected at the distribution level, a portion of the power system that was built for serving passive clients, on the basis of a monodirectional flux of energy. This also poses power quality and safety concerns since more generation in the distribution systems raises the voltage, and protection systems must be recalibrated. This calls for better observability and more advanced control mechanisms at the distribution level.

Also, renewables have a low energy density: they need much more space in comparison with fossil-fuel sources and are site-dependent (convenient only where enough wind, solar, or wave energy is present), which can easily bring to a high concentration of RES in certain areas of the power system, with the result of possible congestions in the existing lines. For this, more thoughtful planning and grid reinforcement investments are needed.

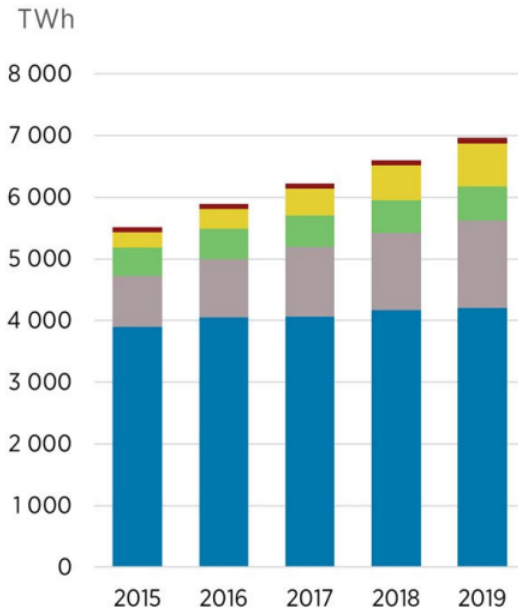


Figure 1.5: Renewable energy generation worldwide growth. RES types: in blue hydropower, in gray wind power, in yellow solar, in green bioenergy, and in red geothermal [15]

Moreover, while the historical power systems had to deal with few generation plants and owners, now the number of agents involved in the secure operation of the grid is higher than ever. For example, in Italy more than 900.000 PV plants are installed, for a capacity of 56 GW [18].

Thus, a new kind of agents, called aggregators, are being involved in the management of the system that can manage the flexibility requests of the Transmission System Operator by aggregating (hence the name) a great number of generation plants and loads in what are called Virtual Power Plants. But the aggregators themselves need tools for optimally managing the flexibility requests [19].

To have an idea of the challenges ahead, the framework developed by International Energy Agency (IEA), which characterizes the different phases of

renewable integration, can help. The framework captures the evolving challenges as countries transition to higher shares of RES and helps to prioritize actions that ensure continuity of supply. The framework puts each system in one of the following phases [8]:

- Phase 1: System integration is not a relevant issue;
- Phase 2: System integration draws on existing flexibility;
- Phase 3: System integration needs flexibility investments in all measures (like Battery Energy Storage Systems (BESSs));
- Phase 4: System integration requires advanced technologies because, in this phase, system operators need to intervene frequently in order to balance electricity demand and supply and to support power quality requirements;
- Phase 5: System integration brings to frequent periods of RES exceeding demand;
- Phase 6: System integration causes excess or deficit over months and seasons.

Currently, the European Union (EU) is considered in phase 3 by the IEA, with countries like Denmark and Ireland being already in phase 4. In 2030 all countries will be either in phase 4 or 5, while in 2050, they will be in phase 5 or almost in phase 6 (the uncertainty is due to whether to prediction is made using current policies or announced pledges) [8].

Some of these challenges can be addressed with specific technological advances (such as the implementation of synthetic inertia or improved BESS technologies). But all of them can take the benefit of an information layer over the assets on which insights and proactive actions can be deduced with data analytics.

1.2 Data Analytics

Outlining the history of data analytics is more complicated than the one of power systems since it is a discipline at the crossroad of many other fields and relatively younger than power systems. Indeed, the concept of discovering valuable information from massive collected data in commercial operation as

aiding knowledge for business decisions was only proposed in 1989 under the term "business intelligence" (BI) [20].

Moreover, the field is linked with many terms, that have no clear shared common meaning and are often misunderstood. Most of them will be described in the following, in order to better clarify the object of the thesis and to give an overview of this vast field of knowledge, technologies and processes. The discussed terms will be:

- Big Data;
- Artificial Intelligence;
- Machine Learning, Pattern Recognition, and Data Mining;
- Deep Learning;
- Expert System;
- Analytics.

1.2.1 Big Data

With the fast spread of the internet and in particular of the Internet of Things (IoT), data on the internet can be measured in exabytes (10^{18} bytes) and zettabytes (10^{21} bytes) [21]. To refer to the associated technological innovations, the BD term was coined [22], which might seem self-explanatory.

However, its definition is nonuniform, although some consensus can be identified [20]. Definition of BD is typically linked to three 'V's: volume, variety, velocity [23].

- Volume refers to the amount of data, in particular, whether the data is stored in one device or more than one device (via distributed computing frameworks like Hadoop [24]). In the case of distributed storage, as it is more and more common, communication issues also arise [20];
- Variety refers to the different sources of data (such as IoT, external Application Programming Interface (API), services, surveys) and types of data (structured, tabular data or unstructured data like audio, video, and images);

- Velocity refers to the frequency of acquisition of these data. Even small increases in data frequency sampling can result in a tremendous increase in storage requirements if all data is to be kept (see Table 1.1)

Table 1.1: Quantification of collected data in 1 year for 1 million devices for different sampling rates (estimate of 5 KB per record) [20, 25]

Collection Frequency	1/day	1/h	1/30 min	1/15min
Records Collected	365 millions	8.75 billions	17.52 billions	35.04 billions
Volume of Data	1.82 TB	730 TB	1460 TB	2920 TB

It should be clear that the concept is by itself a relative one, at least in three aspects:

- What is high depends on storage capacity, computing power, and algorithms. In recent decades, data storage has tremendously improved. Also, the increase in computing power has been remarkable, and high-velocity data has shifted meaning. Consequently, computing costs have dropped significantly, as investments that were only affordable by big companies and the military are now at hand for nearly every business and sector [26]. Finally, algorithms allowing easy analysis of diverse types of data (like neural networks) have changed meaning to the concept of high variety data. So what was once considered BD may not be seen as BD now.
- Some data may be big in only one or two of the three aspects. A large high-frequency group of IoT devices may be high in velocity and volume, but not in variety, for example, because the type of data involved is just one.
- Also, what is BD depends on the single sector taken into consideration and what data is considered (see Figure 1.6). For example, data considered big in the power sector may be fairly small in other industries or vice versa.

Another fundamental aspect of BD is that data is not collected for a single, premeditated aim, but it is considered as an asset that can be exploited in a number of unforeseen, novel ways in the future.

The emergence of the BD has been fundamental for the affirmation of Big Data Analysis, which most of the time can add increasing value (sometimes

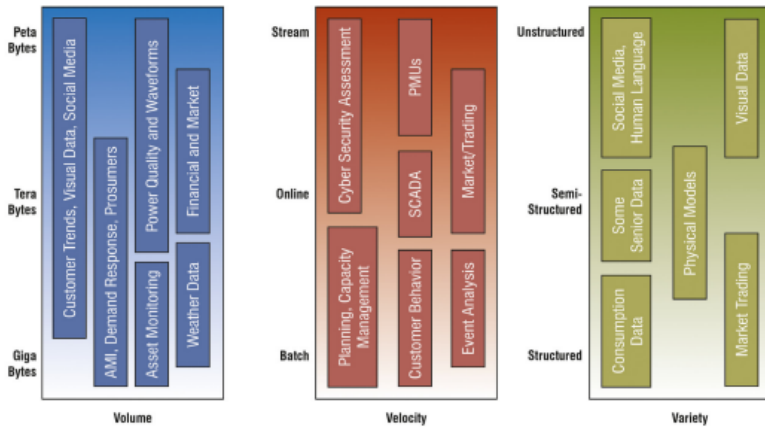


Figure 1.6: The three Vs (volume, velocity, variety) of power system data [23]

added to the three "Vs" of big data [20]) to processes as data dimension increases and hence information is more diluted in data. In any case, BD and the proper analysis tools are not sufficient: for handling the complex scenarios of power systems also guarding the quality of data against measurement errors and adversarial attacks (i.e., cybersecurity attacks) is fundamental (sometimes referred to as another V: Veracity) [16, 20]).

1.2.2 Artificial Intelligence

Artificial Intelligence (AI) has no standard definition. It is referred to as a subfield of computer science that investigates the creation of intelligent machines that can act like humans via the exploitation of the ever-increasing amount of data produced worldwide [16, 27].

Historically, AI was born in the middle of the 20th century, with the arrival of the first computers. Still, the strive of building machines with autonomous, human-like capabilities has a long history. Artificial Neural Network (ANN) and AI are related areas, and sometimes they are used as synonymous: actually, ANNs are just a tool to implement AI, but not the only one.

The confusion is due to the fact that ANN is arguably the most common way to build AI systems, given the impressive breakthroughs that ANN has achieved with Deep Learning (DL) (see Section 1.2.4).

ANN is also the oldest tool for realizing AI since the first ANN (the Perceptron) was invented in 1957 by Frank Rosenblatt, as a result of even older research.

The term AI was coined just the year before (1956), in the context of the Dartmouth Summer Research Project on Artificial Intelligence [28].

Other tools for implementing AI are expert systems (which were popular at the end of the 21st century) and Reinforcement Learning (RL), which may or may not use ANN as a subsystem. Also Machine Learning (ML) is a fundamental piece of most AI systems. Another field is Transfer Learning (TL), which deals with the problem of using developed models for unseen tasks.

Although the most astounding breakthroughs of AI were unthinkable until recently (for example, in the field of speech processing, videogames, and computer vision), whether this constitutes true 'understanding' is not clear. Indeed, intelligence is all about generalization, and dealing with entirely novel cases is very difficult even with state-of-the-art AI systems.

AI can be studied in two ways: one, more purely scientific, is by modeling intelligence in order to understand intelligence, and one more technological, that is, finding more clever ways to automate tedious or difficult tasks [29], by exploiting BD, which can be regarded as a prerequisite for modern AI [16]. In this thesis, the focus will be on the applicative side of AI.

1.2.3 Machine Learning

ML and AI are not the same thing, although they are often used interchangeably. Machine Learning is a form of AI that enables a system to learn from data rather than through explicit programming (while not all types of AI are automatic [16]).

ML aims to solve a wide range of problems (or more precisely, tasks), the most frequent ones being the following:

- Regression, where the system has to learn to map inputs to a continuous quantity;

- Classification, where the system has to classify the input into a pre-determined class. A very similar task is solved by Pattern Recognition, which is very similar to ML and sometimes overlapping;
- Clustering, where the input data has to be grouped into similar groups;
- Dimensionality Reduction, where the aim is to reduce the dimension of the input data;

These tasks can be solved in various ways, the main ones being:

- Supervised Learning, where inputs have right examples (labeled data) to learn from;
- Semi-supervised learning, where some inputs are not labeled, typically the great majority of them;
- Unsupervised learning, where no data is labeled;
- Reinforcement learning, which is used in cases where the system has to teach itself how to act in a dynamic environment by collecting rewards to learn which actions are better. It has an aspect of exploration that renders it quite peculiar compared to the other learning methods.

ML is strictly linked to many other disciplines like probability theory, statistics, information theory, algorithmic science, and other ones. Another related concept is Data Mining: the difference being in the mindset to uncover precious, new knowledge from corporate Databases, and can be seen as a part of the Knowledge Discover in Databases process.

1.2.4 Deep Learning

DL is a subset of ML, which is itself a subset of AI. DL goes beyond machine learning by creating more complex hierarchical models designed to mimic the way humans learn new information. Specifically, the models are called Deep Neural Network (DNN), which are deep versions of ANN. With time DNNs have become deeper and deeper, with more artificial neurons and complex architectures. Significant advances have been made in the last ten years (2012-2022), especially with unstructured data problems (such as images, audio, and video). Indeed, in 2012 for the first time, a DNN (called AlexNet) achieved state-of-the-art recognition accuracy against all traditional machine learning

and computer vision approaches. Also, for the first time, it was shown that training on a Graphic Processing Unit significantly improved training time [30].

Soon, DL applications spread everywhere. For example, it gave a tremendous impulse to the development of RL, a kind of AI which learns through experience rather than through examples.

Recently, the transformer architecture revolutionized also tasks involving sequences, such as time-series and natural language processing [31].

DL has almost become synonymous with AI in the last years, although it still has some limitations on structured data.

This is relevant to this thesis because data in the power systems sector is highly structured. Anyway, the applications of DL to power systems have enormous potential.

1.2.5 Expert system

An expert system is a computer program that models the ability of human experts to solve problems. It can reach the same level of problem-solving skills as an expert can [27]. It was very popular in the 1970s and 1980s, with important applications in also in the energy sector [32]. Expert Systems are made of pieces of knowledge on a specialized field, expressed as if-then rules, together with a representation of particular conclusions. Examples of this kind of modeling are the ones based on fuzzy logic, which are also used for control [28]. Expert systems can be part of decision support systems, which manage knowledge in order to provide tools to make informed decisions about processes.

1.2.6 Analytics

Analytics can be traced back to Taylorism, a system of scientific management that emerged in the 20th-century Scientific management, based on the principles of gathering workers' knowledge to plan and optimize the time spent by shop workers in completing tasks, with subsequent monitoring and continuous improvement [33].

The analytics term refers to the steps needed to transform data into insights, predictions, and actions that bring value. There are several types of analytics structured around questions about the process being analyzed.

- **Descriptive Analytics.** The objective is to answer the question: *what has happened? what is happening?*. Connected with this step is **Diagnostic Analytics**, whose objective is to answer the question: *why it has happened? why it is happening?*;
- **Predictive Analytics.** The objective is to answer the question: *What will happen?*;
- **Prescriptive Analytics.** The objective is to answer the question: *What should I do?*.

All of them are related, with descriptive and diagnostic analytics leading to predictive analytics, which in turn aims to prescriptive, which is the part adding the most value to the process.

In order to fulfill each analytics aim, a wide range of tools can be used, like the ones described in the previous sections. Still, the way in which it is realized can vary tremendously, from a front-end tool to visualize data and perform simple what-if analysis to better inform manual corrective actions to fully-fledged end-to-end Deep Reinforcement Learning (DRL) automated solution.

The analytics process can be found in many places and can have many forms. This thesis will focus on the data analytics solutions that deal with AI solutions, informed, when appropriate, by a physical model of the power system or subsystem.

1.3 Smart Power Systems

On some level, power systems have always been "intelligent" because of automatic control systems installed at the generation, transmission, distribution, and end-use level [34]. But the energy transition, as outlined in Section 1.1 is putting pressure on the power system into being smarter than ever before.

To meet the new requirements, an information layer is now added to the physical electricity system at all levels (centralized generation, transmission, distribution, decentralized generation, and end-use [35]) with the purpose of collecting and analyzing data coming from smart meters and sensors [20].

Adding an information layer to the power grid gives birth to what is called a smart grid. Smart grid has been defined as an electricity network that can integrate cleverly the actions of all the users connected to it, for economic and

environmental objectives, all the while meeting security, policy, end-use, and societal requirements [36] (see Figure 1.7).

In this context, data is central to developing a successful integrated system. Data available are various, which can be sorted into three categories [20]:

- Measurement data, such as:
 - Telemetry and Supervisory Control And Data Analysis (SCADA) data;
 - Oscillographic and Synchrophasor data;
 - Consumption data (mostly from smart meters);
 - Grid metadata (describing the grids);
 - Vibration, Acoustic, Visual data streams of single assets [26];
- Business data, such as:
 - Electric Markets data;

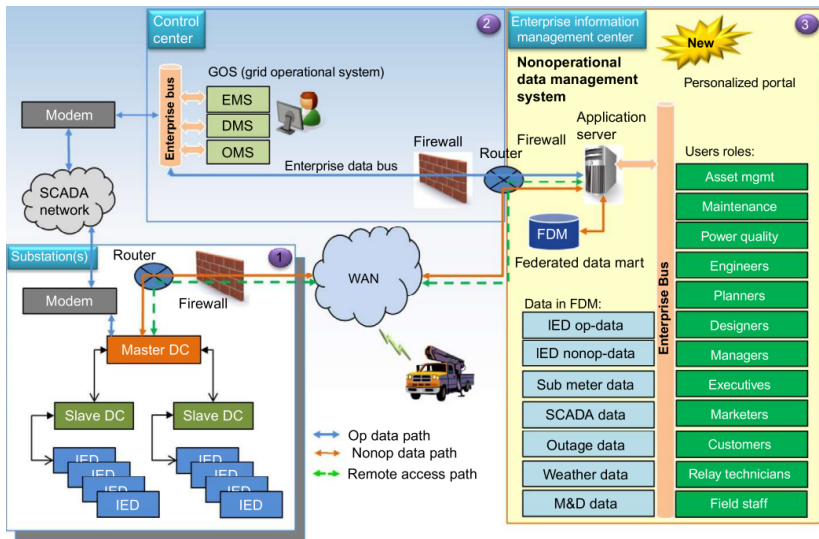


Figure 1.7: Smart Grid Infrastructure [37]

- Customer service data;
- Marketing system data;
- Geographic Information Systems;
- External but related data, such as:
 - Weather data;
 - Traffic data;
 - Social Media data;
 - Festival, sport events, popular tv shows data.

This is undoubtedly a lot of data, but its potential is significant. It was calculated that AI could reduce Greenhouse gases emissions by around 2% each year, for a total of 1.3 GtCo₂ by 2030 [38], with respect to a scenario without massive employment of AI.

Similarly has been reported that digitalization, driven by the massive deployment of smart meters (see Figure 1.8), could reduce global buildings sector demand by up to 10% until 2040. Also, transport CO₂ emissions could be reduced by over 50% in 2050 and increase demand response capacity more than ten-fold (although the exact scale of these impacts is uncertain) [39].

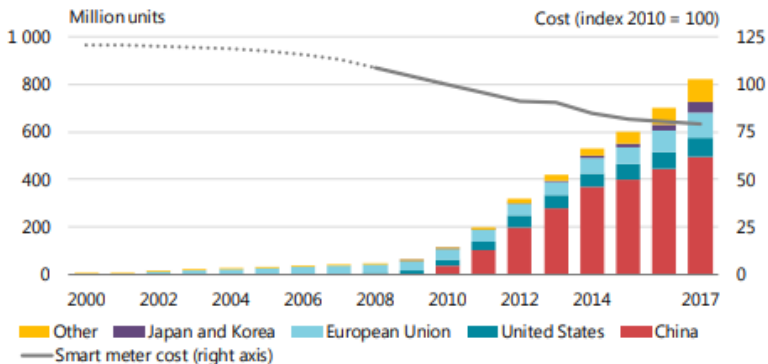


Figure 1.8: Smart meter deployment, cost, and penetration ©IEA 2019 [39]

Also, the infrastructure needed to collect the data and coordinate the analyses and the responses of the analytics algorithms presents their challenges,

Table 1.2: Smart data collection devices and systems in smart grids [20]

Smart Device	Acronym	Technology	Application
Advanced Metering Infrastructure	AMI	Integration of smart meters, data management systems and communication networks to provide bidirectional communication between customers and utilities	Remote meter configuration, dynamic tariffs, power quality monitoring and local control
Phasor Measurement Unit	PMU	Real-time measurements (30 to 60 samples/second) of multiple remote points with a common time source for synchronization	Electrical waves measurement of power grid
Wide Area Monitoring System	WAMS	An application server to deal with the incoming information from PMUs	Dynamic stability of the grid
Remote Terminal Unit	RTU	A microprocessor-controlled device to transmit telemetry data	Information collection of system operation status
Supervisory Control And Data Acquisition	SCADA	Data acquisition system for physical systems	System monitoring event processing and alarm
Smart Electronic Device	IED	Monitoring and recording status changes in the substation and outgoing feeders	Combination of different relay protection functions with measurement, recording and monitoring

as the number and variety of devices connected is significant (see Tables 1.2 and 1.3). For this reason, standards of communication are an important topic in the field [40].

On the other hand, the digitalization of the smart grid also drives up the electricity demand. Even more, it is known that in specific contexts, such as smart homes, efficiency solutions can lead to rebound effects, which is the tendency to consume more electricity because an efficiency solution has been implemented [41]. But there is also evidence that digitalization's impact

Table 1.3: Summary of communication network infrastructure in smart grid

Type of network	Acronym	Function	Deployment	Data Transmission Rate
Home Area Network	HAN	Enabling communication among smart home or office devices and smart meters for local energy management	Deployed at house or small office	Low (less than 1Kbps)
Neighborhood Area Network	NAN	Consisting of several HANs for energy consumption data aggregation and storage at Load Data Centers (LDCs)	Deployed within area of hundreds of meters	Medium (up to 2Kbps)
Wide Area Network	WAN	Enabling the communication of all smart grid's components	Deployed within tens of km	High (up to few Gbps)

on energy demand and emissions could be nearly ten times greater than any negative impacts [42].

Overall, making the grid smarter by Information and Communication Technology is a valuable option in meeting the sustainability objectives of the energy transition.

1.3.1 Possible Applications

Applications of analytics in the context of smart grids are copious. In the following, a nonexhaustive list is given [20]:

- **Renewable energy forecasting.** Being able to forecast RES generation is very important in all power systems, especially in low inertia ones. Good forecasting models could improve maintenance scheduling, dispatch planning, and frequency regulation [43];

- **Load forecasting.** Similarly, short-term load forecasting is essential for energy management and system operation, as well as for market analysis, with significant possible savings [44];
- **Load profiling.** Methods for describing typical behaviors of electric consumption can have important benefits on operations and capital planning, as well as tariff structure design [45];
- **Fault detection.** Uncertainty induced in the power grid by RESs, together with their much lower inertia, can lead to difficulties in detecting faults, especially in microgrids. Specific analytics solutions can be devised to help detect faults in active networks [46];
- **Predictive Maintenance.** The power system relies on equipment, such as transformers, whose abrupt failure can cause severe damages and from prolonged disservices to disastrous blackouts. Hence being able to monitor the condition of power transformers and possibly predict their future health is of immense value to the overall security of power systems. Similar methods can be used to proactively maintain equipment such as Heat Ventilation Air Conditioning (HVAC) systems and BESS, by prolonging their lifetime;
- **Transient stability analysis.** Power system transient stability is the ability of synchronous generators to continue to operate in synchronicity after severe contingencies such as lines disconnection and short-circuit faults [47]. In these delicate contexts, transient stability assessment methods are fundamental, especially in modern power systems, where increasing demand and higher RES penetration are pushing the system to work near its security limits. In smart grids, analytics solutions can provide a better security assessment via, for example, efficient summarization techniques [20];
- **Topology identification.** Analytics solution can provide system operators with pseudo-measurements that can help in the visualization of low voltage networks with limited metering [20, 48];
- **Power quality monitoring.** The increasing application of power electronics-based loads and generators harmonic distortions is leading to unstable and insecure states. Proactive identification of those states is possible with analytics solutions [20];

- **Load disaggregation.** Also called Non-Intrusive Load Modeling (NILM), it aims to disaggregate a load into its constituent appliances in order to, on the one hand, save on equipment, and on the other, acquire useful information about the subload usages. NILM works appeared in the nineties [49], and it is still an active field today, leveraging the most advanced analytics techniques [50];
- **Non-technical loss detection.** Losses due either by electrical theft or errors in accounting are one of the prominent financial concerns for utilities. They may also cause important imbalances in the system [20];
- **Cybersecurity.** Risk-informed strategies that rely on the understanding of the impact of a potential event are very important in contemporary smart grids, as attacks are becoming more frequent and more sophisticated [26].

It is in the contexts of these applications that the work of the Ph.D. and this thesis lies.

1.4 This Thesis

Among these significant changes and challenges that the power system faces, this thesis tries to contribute with several hopefully original works and perspectives.

The research in power systems analytics is considerable in size but suffers from several issues [20]. First of all, deployed applications are few, as access to data is difficult. In particular, in most countries, smart grid data are considered confidential, so the availability of real-world data for researchers is limited. Papers based on completely simulated data are not rare.

In addition, while the data analytics techniques have made astounding progress in previous years, their transfer to specific domains such as power systems is not immediate. The potentiality of such solutions is either not known or explored in full detail.

In this thesis, Applications presented stem from projects made in strict collaboration with Distribution System Operators (DSOs), power equipment manufacturers, BESSs manufacturers, and other actors that are highly aware of the value of data for advanced insights in power system analytics. Moreover, the replicability of the models developed was one of the main concerns in designing such solutions.

For the problem of technological transfer, this thesis groups the applications within the framework, firstly proposed in [51], of the three types of analytics presented in Section 1.2.6:

- Descriptive and Diagnostics Power System Analytics (Chapter 2);
- Predictive Power System Analytics (Chapter 3);
- Prescriptive Power System Analytics (Chapter 4).

Also, a comprehensive overview of the respective methodological aspects is given for each type of analytics, hoping to offer power system engineers and researchers the tools to appreciate analytics better and apply it in their work.

This division was preferred because it does not rely on single data analytics techniques but instead gives a goal-oriented perspective on such tools. Indeed, sometimes, also in scientific reviews, the difference of the single techniques receive more attention than the motivations for employing these tools, leaving the reader wondering why one should choose one method over the other.

An introductory methodological part on the specific analytics is given for each chapter. Then, an overview of selected applications is detailed. Finally, the corresponding applications are described together with empirical experiments.

In particular, in **Chapter 2**, Descriptive Power System Analytics and Diagnostic Power System Analytics are introduced. They try to answer the questions "*What has happened? What is happening?*" (**Descriptive Analytics**) and to the questions "*Why did it happen? Why is it happening?*" (**Diagnostic Analytics**). While Descriptive analytics relies mainly on visualization techniques, many types of Diagnostic Analytics approaches are possible, such as interpretability techniques and simulation models.

The applications reviewed span many subsectors of power systems, such as Phasor Measurement Unit (PMU) analytics, to smart meters analytics to energy efficiency dashboards. Another example is given from the PODCAST project, in which advanced functionalities for a DMS were developed and visualized on an interface informed with all the grid metadata.

The innovative application of this chapter is an algorithm called Instantaneous Growing Stream Clustering (IGSC), a new stream clustering algorithm,

devised during the Ph.D., which is able to dynamically adapt the number of clusters of incoming data all the while modeling them in a Discrete Markov Chain (DMC) [52, 53], for example for load modeling. Interesting features of this algorithm are: its suitability for application to IoT devices, given that it does not require a prohibitive amount of working memory; its probabilistic model, highly interpretable, easy to visualize, and capable of simulating the observed load; its applicability to multivariate settings, where the statistical relationship between different measurements may be needed.

Experiments are given for: modeling active and reactive power; modeling active power in time. The algorithm was tested on diverse data, such as a single HVAC unit serving a research laboratory; an HVAC composed of an Air Handling Unit (AHU) and a chiller unit serving an entire university department; several industrial devices for an industrial dataset.

Results show that the algorithm can give an immediate visualization and a meaningful model of the monitored loads. Its clustering capabilities are in line with comparable techniques drawn from the literature. Also, practical advice on the choice of IGSC hyperparameter is given.

In **Chapter 3, Predictive Power System Analytics** is introduced, which tries to answer the question *"What will happen?"*. Contrary to the other analytics, a ubiquitous framework is available for developing predictive analytics solutions: the CRISP-DM, developed by IBM [54, 55]. The single phases of these frameworks will be described in detail, taking advantage of the copious literature on the matter.

The subsequent review will focus on PV forecasting papers and load forecasting papers.

Next, three applications will be described:

- an innovative day-ahead PV forecasting hybrid approach, which leverages either a Clear Sky Model (CSM) based model or a data-driven, neural-network-based model based on a decision rule, constructed either by a decision tree algorithm [56], or a linear regression decision rule [57]. The algorithm is tested on a 20 kWp power plant located on the top of a university department in Genoa, Italy. The goodness of the proposed procedure is assessed;

- the application of a day-ahead load forecasting algorithm [58] to 68 MV/LV substations and 17 MV users of the Italian distribution network involved in the PODCAST project [59]. The experiments here ensure that the proposed solution beats the business-as-usual solution;
- Finally, an energy forecasting model for the university building involved in the PREDICT project [60]. The innovation of the model is the use of a deep transfer learning solution for inferring the occupancy data from a thermal imager. The estimated occupancy is then linked with room booking data analyzed via text analytics procedures, and future occupancy is forecasted using a k Nearest-Neighbors (kNN) model. Lastly, the forecasted occupancy is used as input, together with weather data and setpoint data, of a state-space model that constitutes the final energy forecasting model. The solution was thought of as part of a Model Predictive Control (MPC) solution for HVAC energy efficiency: the sensitivity of the model to various setpoints is therefore checked with success.

Lastly, in **Chapter 4, Prescriptive Power System Analytics** is introduced, which tries to answer the question *What should happen?*. This analytics is all about automating the decision-making process, often needed in a power system. As such, the previous analytics can be seen as the natural building blocks of this final type of analytics.

Then, the primary methods for realizing prescriptive analytics are given: first and foremost mathematical programming, together with its stochastic version. Also, more data-driven tools are presented, either enhancing classical mathematical programming or providing a completely different alternative to the problem. The many other prescriptive analytics approaches are also listed.

Some notable examples of prescriptive analytics are given: from HVAC optimal control to EV smart charging to microgrid and BESS optimization.

Finally, two prescriptive analytics applications realized during the Ph.D. (and in particular, during the PODCAST project) are described:

- A stochastic planning including DER regulation and sizing and siting of storage devices is described, with priority in reactive compensation distributed generators contributions. In contrast, the contributions of BESS are minimized while solving voltage problems and congestions

[61]. Probabilistic planning is based on daily Monte Carlo simulations, extracting load and RES profiles based on a gaussian probabilistic model. The proposed model is shown to be acceptable for the task by testing it on the model of the PODCAST project network;

- A day-ahead battery profile optimization algorithm that can help run a microgrid or an energy community. The algorithm considers several technical constraints and aims to maximize the profit of the system. Also, an intraday algorithm is devised in order to implement the declared profile. Two experiments from the PODCAST project are reported, which show that the batteries installed on the grid for the project are able to follow the declared profile with the proposed optimization algorithm.

The innovative contributions of this thesis can be therefore summarized in:

- A unified view of data analytics applied to power systems;
- Methodological considerations on all the types of analytics (descriptive, diagnostic, predictive, prescriptive), useful for power system professionals and researchers;
- Practical applications developed in strict conjunction with companies of the power system sector for each type of analytics.

Descriptive and Diagnostic Power System Analytics

*All models are wrong, but some are useful
- Prof. George E. P. Box*

2.1 General Methodological Aspects

In this chapter, the techniques that are used to answer the following questions will be presented:

- What has happened (or: is happening) to the system?
- Why did it happen (or: is it happening)?

These are questions relative to the past (or the present), and their answers, if present, are hidden in the data. The question may be about a particular event or a general feature of the system at hand. In this sense, descriptive analytics may be regarded as passive (historical) analysis [62]. Visualization of past or present data is the simplest form of Descriptive Analytics, but there are

others, as will be accounted for in the following. Also, visualizing data may not be trivial, especially when the dimensionality of the data is considerable.

When Descriptive Analytics tries to answer why a specific thing happened, it is more appropriately called Diagnostic Analytics, since in that case, reasons for events are needed.

Moreover, descriptive and diagnostic analytics are also valuable for the contexts of predictive and prescriptive analytics, where the description and diagnosis are needed for understanding the performances and the weaknesses of the forecasting and decision-making solutions.

At the end of the chapter, a method developed during the Doctorate (Ph.D.) (named Instantaneous Growing Stream Clustering (IGSC)) will be presented. The IGSC algorithm is helpful for compactly visualizing complex time series by identifying a series of observed states and for building a Discrete Markov Chain (DMC) over them, with the additional feature of being computationally light, thus making feasible its applicability on distributed Internet of Things (IoT) devices. Examples of possible applications will be given, such as load modeling on a variety of load types (i.e., building, Heat Ventilation Air Conditioning (HVAC), industrial machinery).

2.1.1 Descriptive & Diagnostic Analytics Approaches

We will start with a tour of how descriptive and diagnostic analytics may be implemented in general.

2.1.1.1 Visualization

Statistical graphics and Data Visualization (DV) are not a modern achievement of statistical sciences [63]. Indeed the concept of visualizing empirical measurements (i.e., data) can be traced back to centuries, and the idea of graphical coordinates even predate modern times.

With the development of statistical theory and the growing importance of empirical numerical data, between the 1700s and the 1900s, graphical representations were seen in new domains and with innovative pictorial concepts. At the same time, machines for producing graphs began to be invented and patented. Soon, an astounding explosion of use and innovation in statistical graphics happened in the most disparate contexts. The use of graphics for DV became common in scientific publications regarding natural and physical phenomena.

In the last century, DV became more and more popular: graphical methods entered textbooks and educational curricula, and their use in scientific activity was the norm. Also, Exploratory Data Analysis (EDA) (which comprises DV and other descriptive analytics techniques) gained academic dignity by being separated with mathematical statistics [64, 65]. Moreover, works on perceptual elements of graphics and their link with the data displayed and the message to be conveyed further refined the efficacy of DV. Also, new techniques for visualizing multidimensional were developed [63].

Finally, the advent of modern electronic calculators made data visualization easy and cheap to design, implement and distribute. Additionally, dynamic and interactive visualization was made possible. Also, DV techniques able to deal with vast amounts of data have been proposed, a problem that has its peculiar challenges and benefits of usability [66].

The reason why DV gained such prominence is that it helps see data from multiple viewpoints. That is, it helps with dealing with the multidimensionality of modern datasets. A common tool for navigating the data is the dashboard, which is a single screen visualization that shows several graphs and Key Performance Indicators (KPIs) to display a wide variety of information in a limited space. To build an effective dashboard, the most important pieces of information should be identified beforehand [67].

Dashboards are a special kind of interface. Interfaces connect a person to a machine, system, or device and allow people to create machine-readable data and instructions, for example, via a keyboard, touchscreen, voice recorder, or other input devices. They are also key technologies for converting data-driven analysis into real-world energy efficiency improvements, in conjunction with human intervention [39].

Another way to deal with multidimensionality, typically together with DV is to reduce the dimensionality of the data via some specific dimensionality reduction techniques, such as [68]:

- Principal Component Analysis (PCA), which exploits the fact that many datasets develop in certain directions rather than being distributed randomly. In a sense, often, there are more variables than needed: there is a redundancy of information. Hence the data can be represented with other coordinates that are a linear combination of the original variables. PCA is a specific algebraic method for building this new linear combination of variables. Specifically, principal components are orthogonal

to each other. In addition, the variance of data along the first principal component is highest among all the possible principal components. Also, the principal components are ordered according to the variance of the data along the principal components;

- Factor Analysis (FA), a way to try to estimate the underlying components of the multivariate data by assuming the data to be dependent on a linear combination of common factors.
- Multi Dimensional Scaling (MDS), which allows visualizing distance between points evaluated using distances different than euclidean distance;
- t-Stochastic Neighbor Embedding (t-SNE), which embeds (i.e., encodes) data in a lower-dimensional space (typically, 2D) in order to respect as much as possible the distances of the high-dimensional space in the low-dimensional space.

2.1.1.2 Interpretability

A model can be fitted to historical data to make the relationships interpretable. Although there is no shared mathematical definition of interpretability (also called explainability), it has been described as the ability to understand the cause of a model decision or the predictability of a model result. Explanations can be issued for a whole model or a single piece of data, in high stakes scenarios, for example, when security is involved [69]. Interpretability is also a prerequisite for fairness, privacy, robustness, causal inference, and trust.

Interpretability can be achieved in several ways:

- Intrinsic interpretability: simple models such as linear models and decision trees are intrinsically interpretable, so modeling data using these and similar types of models can guarantee a good level of interpretability by default;
- Post-hoc interpretability: apply specific interpretability approaches to arbitrary models (even highly complex like Deep Neural Network (DNN)).

The results of interpretability can be different:

- A summary statistic for each variable involved;

- A summary visualization for each variable;
- The model itself (for intrinsically interpretable models);
- A data point, which provides interpretation through an example;
- An intrinsic interpretable model developed for explaining the model (in the case of black-box models);

Moreover, some methods can be applied to any model (so-called "model-agnostic"), while others are model-specific.

An explanation can be thought of as an answer to a why question. In this sense, the link between explainability and diagnostic analytics is clear [70]. Good explanations may be contrastive (i.e., they refer to a meaningful baseline), as simple as possible, and social (i.e., they should be tailored to the target audience). They also focus on abnormalities, if present, and should be as truthful and general as possible [71].

Interpretability can help gain insight into causal relationships between variables, but to test causal relationships, more specific data-driven approaches can be used [72].

2.1.1.3 Simulation

From the data, a generative model can be built. A generative model is a mathematical model that, given a training set, can generate (i.e., simulate) data similar to the training instances [29].

Simulation can be used in contexts where collecting the data is costly in terms of time, money, or another relevant aspect. Nevertheless, even if data collection is cheap, it can be used for diagnostic purposes. Integrating data analytics tools with simulation software is a critical aspect that has its specific challenges [73].

A closely related concept is that of Digital Twin (DT), which aim is to recreate the digital version of a physical asset [26]. This renders anomaly detection applications possible, as well as many predictive and prescriptive analytics tasks. There are several types of DTs [74]:

- Asset model. In this case, the goal is a general performance assessment (by benchmarking), also with visualization of other connected assets;

- **Fault Model.** The goal is to do diagnostics on errors, faults, and security (included cyber-security) [75], in a timely manner, in order to understand how to avoid them. Faults may be due to external factors or to natural degradation, for which a specific mathematical model should be built;
- **Operational Model,** for operational planning activities, in order to optimize assets revenue on a day-to-day basis;
- **Business Model,** whose aim is to help in developing new services and business models.

The inputs for the DTs are the variables of interest (such as active power, reactive power, current, voltages, and flexibility) acquired with real-time IoT devices. The actual DT can be built using either physics or various data-driven methods, ranging from empirical models to statistics to Machine Learning (ML). Also, Virtual Reality (VR) and Augmented Reality (AR) can be incorporated to ease the interpretation of the virtual simulation. There is a lot of research going on in DT for Smart Grids (SGs), and a unified view is still missing [76].

However, DTs are a hot topic also in other related fields, such as Smart Manufacturing (SM), where manufacturers have widely used modeling and simulations as a step for improving their operations by achieving insights, patterns, trends, inefficiencies, and overall bottom-line [77].

Of course, no DT can ever be an actual replica of the physical asset. This is due to the presence of several uncertainties in the building of any DT or simulator.

The uncertainties can be of two types [77]:

- **Epistemic uncertainties,** that arise from ignorance about the problems;
- **Aleatory uncertainties,** that arise from problem-inherent variability.

Indeed, simulation for diagnostic analytics must be thoroughly validated and sensitivity analysis performed for giving meaningful and useful results [77]. In turn, simulation can be used for testing data-analytics solutions by means of generating data. Recently, also deep learning has been leveraged for generating realistic data [78–80].

2.2 Descriptive and Diagnostic Applications in Power Systems

In this section, some applications found in the literature are reported: all of them lie in the realm of descriptive and diagnostic analytics for a more intelligent and efficient power system.

Simulation is essential in the context of building energy modeling. Buildings have many environmental and energy efficiency requirements. Also, the possible design to fulfill them are numerous. Finally, not always the available data are sufficient to explore all the possible design scenarios. In that regard, simulation tools such as EnergyPlus [81] and specific software bridging simulation tools and data analytics tools (like eplusr R package [73]) are essential for applying data analytics tools in this area because the simulations can generate a significant amount of data that has to be post-processed, and the output is not always easy to be post-processed with standard tools. The benefits of such tools are better reproducibility, both as replicability of scientific studies and the repeatability of the analysis to different buildings.

As for DV, there are several examples of papers that try to enable advanced visualizations in the energy sector. For example, three different types of energy visualization dashboards are compared in [67], with an evaluation made by interviewing several users with experience with energy dashboards. Simple visuals emerged as key for reduced user response time and higher ratings (Figure 2.1).



Figure 2.1: A dashboard designed using gauges, a light indicator, alternating stoplights, a pie chart, and tables. (NG stands for natural gas) [67]

Indeed, Figure 2.1 was preferred to other dashboards that heavily relied on recent time-series visualization, without resorting to more straightforward gauges, which summarize in a rapid manner the current level of the measurements and possibly their abnormality. This was made possible by making questions to sample users with some knowledge about building energy management, usability, and performance of the dashboard (by performance, in this case, it is meant whether the user can get the right information from the single visualizations).

Other works apply descriptive and diagnostic analytics to check and improve forecasting tasks. For example, in [82] a statistical analysis is made on one and two days-ahead load forecasting errors for two Independent System Operator (ISO) over a one-year period. Normality of errors was investigated as well as an analysis to uncover which were the most difficult time periods to forecast. This type of analysis is vital because an imprecise forecast can lead, for example, to a suboptimal day-ahead unit commitment.

Another example is [83], where the role of data visualization and exploration in devising a top-notch load forecasting solution for the Global Energy Forecasting COMpetition 2017 (GEFCom2017) [84] is described. Descriptive analytics helped the understanding of the length and the causes of missing data, as well as outliers and general data distribution. The most useful of these was the so-called data heat map. The heat map alone was able to identify and summarize several characteristics of the time series, such as the weekly seasonality, the overall trend, daylight savings issues, and some pattern linkable to commercial or industrial consumption profiles. Some of the missing data visualized have been linked to an extreme weather event that occurred in that period, and some time series appeared to be moved from one location to another and therefore analyzed together.

In [85], a detailed error analysis of power output forecasting was performed on a Photovoltaic (PV) system of 960 kWp, over several forecasting horizons. The analysis was done by decomposing the Root Mean Square Error (RMSE) in bias, standard deviation bias, and dispersion, and by estimating skewness, and kurtosis, with some hints on how to use this analysis to improve the forecast.

Descriptive and diagnostic analytics on smart meter data is reviewed in [86] under the name of load analysis. Typical applications include volatility and uncertainty of load profiling, anomaly detection, bad data (i.e., missing/-failure data), and energy theft detection.

Visualization and scenario analytics for distributed energy resources in the context of stream data is studied in [87] for integrating it in standard Distribution Management System (DMS) since various distribution system data streams have different granularities, misalignment, uneven sampling, delays. The paper deals with the data fusion of different data sources. The system provides tools for descriptive analytics (called situational awareness) by making the best possible estimates of the state of the distribution system and some scenario analytics (which translates to predictive analytics in the framework of this thesis).

Three case studies on descriptive visual analytics are presented in [88], namely visualizing smart-meter data, Phasor Measurement Unit (PMU) data, and probabilistic forecasts. Smart meter data is represented in Figure 2.2, where the black dots represent the various energy measurements of two smart meters of two different households over the same time period in the Irish Smart-Meter Dataset [89], divided by day of the week.

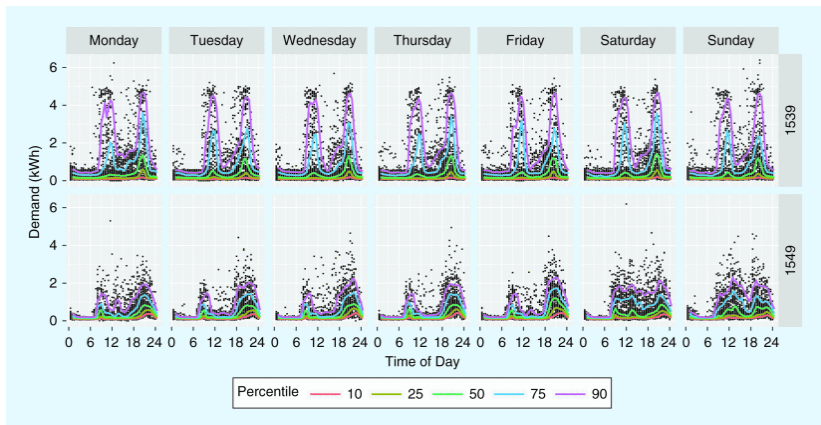


Figure 2.2: The demand plotted against the time of the week for two smart meters from [88]

As the data is variable over the dataset, in Figure 2.2 the percentiles over the hours of the single days are represented. This allows the quick identification of the different load shapes between the different days and the different smart meters, notwithstanding the anomalous and rare data present in the data. For example, for the smart meter 1549, weekend days have higher values around

midday than their working days counterparts, while smart meter 1539 is more regular throughout the week and registers a higher energy consumption.

Another kind of analysis on the same dataset is performed in the same work and reported in Figure 2.3.

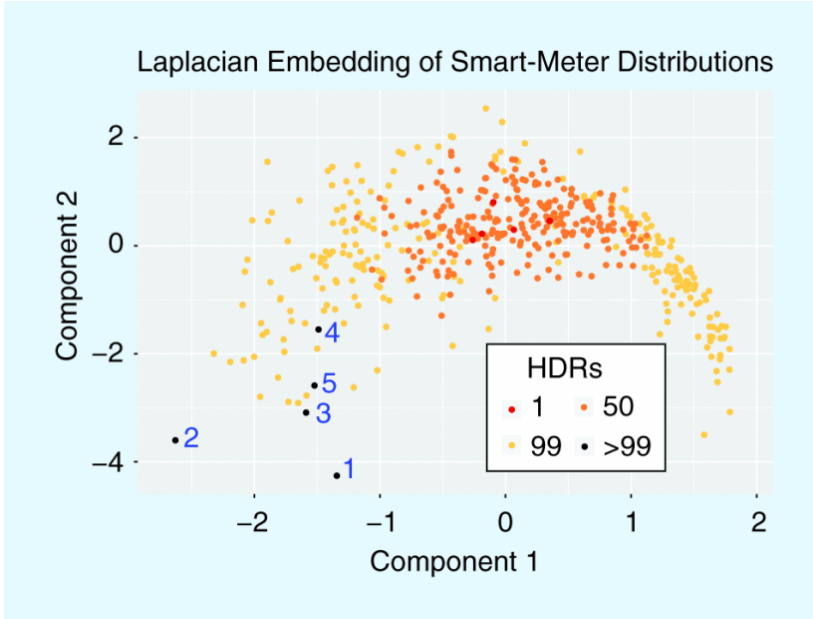


Figure 2.3: All 500 households time series represented on a 2-D plane. Points in red are the most typical households, while the black ones are the most atypical. (HDR stands for high density region) [88].

In Figure 2.3, each point represents a household smart meter time series. The time series is embedded in a 2-D plane via a laplacian eigenmap method (a technique whose aim is similar to PCA, MDS and t-SNE) in order to identify the most atypical households.

Big data visual analytics for PMU are the second study case that involves the use of PCA for visualizing big data generated from an islanding data event due to loss of load and visualizing post-disturbance voltage data.

The third case is for visualizing probabilistic forecasts through visualizing predictive distributions or by visualizing alternative trajectories in the future.

A framework for real-time energy visualization systems for light commercial businesses is studied in [90], with a critical evaluation of existing Electrical Vehicle (EV) systems. Energy audits were made at 15 light commercial businesses for benchmarking the proposed system.

A knowledge discovery method in building data is presented in [91] to discover associations, correlations, and intrinsic data structure in big data, in order to perform fault detection and diagnostics.

A paper describing Business Intelligence (BI) to manage energy on campuses is [92]. The research question was about how to manage power consumption efficiently, and consequently money spending by universities, using a managed process, especially on the data collection side.

A tool for immersive analysis of spatial energy data is introduced in [93], based on ImAxes [94], a software designed for immersive analytics of abstract data in VR, with some new interaction techniques, for easily adding arbitrary variables from the dataset being analyzed.

Power consumption visualizing across a range of arbitrary IT devices is reported in [95]. This kind of solution can help improve energy awareness in an organization. Two available software-based power metering solutions are compared, and a user interface is developed by bringing all the about a device dispersed across different data sources.

An anomalous detection algorithm is developed in [96], analyzing the whole building energy consumption and then performing a diagnosis on the sub-loads responsible for anomalous patterns. Evolutionary classification trees are used to detect infrequently aggregated energy patterns, transformed through an Adaptive Symbolic Aggregate approxImation (aSAX) process. Then a post-mining analysis based on association rule mining is carried out. The whole process is tested on data coming from a university campus [96].

As a final example, some visualizations taken from the participation to the "Distribution Optimization Platform Through The Use Of Data From Electronic Meters And Distributed Storage Systems" (PODCAST) project (happened during the Ph.D.) will be briefly commented.

The aim of the PODCAST project is to build a prototype of a modern DMS, capable of integrating Renewable Energy Source (RES) resources, monitoring a wide variety of assets using smart meter data, and with advanced functionalities like load forecasting (described in Section 3.4), PV forecasting (described in Section 3.3), Battery Energy Storage System (BESS) control (described in Section 4.4), state estimation, fault location, among others.

The DMS was developed and tested on the distribution system of the city of Sanremo (Italy). The view of the monitored portion of the grid under study is presented in Figure 2.4.

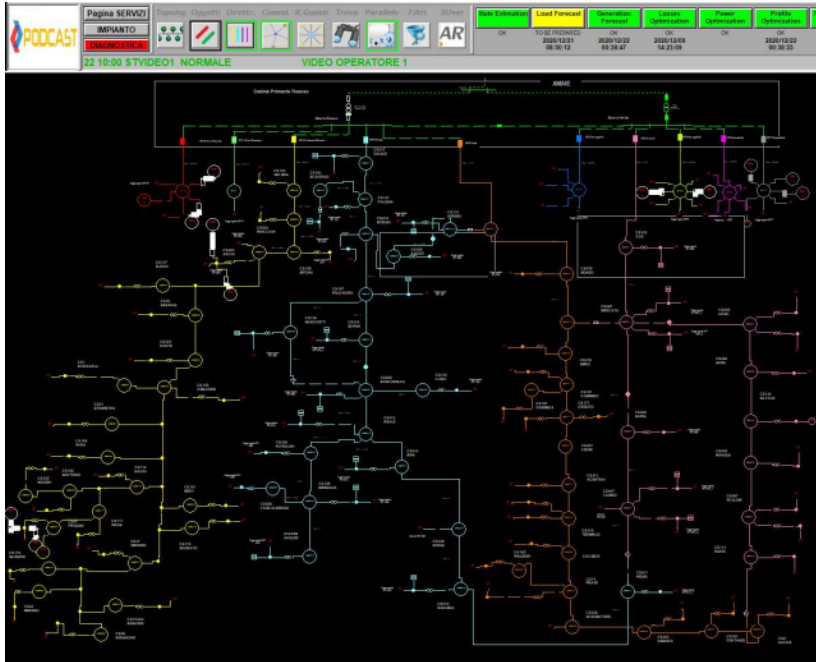


Figure 2.4: Monitored feeders of the DMS of Sanremo (Italy)

In Figure 2.4 the diverse monitored feeders are displayed in different colors in order to tell them apart. On the top of the image, the buttons for starting the advanced functionalities can be found, either as idle (green) or already started (red).

Zooming in, one can access the results of the algorithms for the single substations, as can be seen in Figure 2.5.



Figure 2.5: Result of the DMS functionalities on the single substations

The dropdown menu opened gives access to the results of the algorithms applied to the selected substation (e.g., load forecasting, in the Figure 2.4) as well as historical smart meter data and other relevant information.

In the next and last section of this chapter, an original contribution of the Ph.D. to descriptive and diagnostic analytics is described in detail.

2.3 Application: Instantaneous Growing Stream Clustering

As an example of a descriptive and diagnostic analytics tool, in this last section, an algorithm for characterizing loads monitored in real-time and with low computational power, developed during the Ph.D. course, is presented. The same method can be applied for other types of time series relevant to power system analytics, such as weather data, PV power output, voltage profiles, among others. The algorithm can be used for visualizing the measured data, as well as for simulation, new states detection, and can help the interpretation and the characterization of the monitored quantity.

Specific strategies and technologies for the management of uncertainties are crucial in the RES maximization [97–99]. These reasons have brought to large-scale adoption of Advanced Metering Infrastructure (AMI) in power systems, leading to improved system knowledge, monitoring, management,

and control. On the other hand, dedicated techniques are required to manage the large data flow generated by smart meters in order to use the full potential of the available information [100, 101].

Advanced load modeling methodologies [48, 58] can be considered important tools within this research area. Indeed, specific techniques have to be employed to face the challenges brought by this large amount of data. Traditional load modeling methodologies do not use the streams of data generated by AMI but provide only static load profiles, so advanced functionalities have to be implemented in order to deal with the massive data stream. The IGSC algorithm aims to deal with these limitations as it is an adaptive streaming algorithm capable of modeling model a load through a Discrete Markov Chain.

The proposed algorithm can cluster the load measurements with minimal computational effort, allowing real-time load modeling. The presented procedure's performance is evaluated by experimental validation and compared with two reference methodologies as benchmarks (Dynamical Clustering and k -Means) in terms of clustering optimality and computational time [52]. Additional discussion on the hyperparameter parameter selection was carried out in [53].

2.3.1 Literature Review

A load modeling process consists of the estimation of a typical load curve (also called load profile, load shape, or load pattern) of a generic customer (residential, commercial, or industrial) or a device (such as a thermal load, a pump, or a compressor) or even a whole process (such as a building, a factory, or a system) [58].

Among the various approaches on load modeling proposed by the scientific community, online and adaptive methods are gaining interest in both academic and industrial contexts because they can respond quickly to changes in load usage and state, as well as to component failures. In [101], a recursive and adaptive load profiling technique is presented for dynamically clustering loads of large customer databases. In [102] an adaptive online unsupervised load modeling technique based on Hidden Markov Models (HMMs) is shown in the context of Non-Intrusive Load Modeling (NILM). Authors of [103, 104] have developed dynamical load profiling methods for dynamic loads. Authors in [105] propose a dynamic clustering technique that applies to spatio-temporal data. This algorithm has been updated in [106] through the use of Particle Swarm Optimization (PSO).

Authors in [107] investigate three of the most widely used unsupervised clustering methods: k -means, k -medoid, and Self Organising Map for the load profiling of domestic consumption. In [108], an extension of the Kalman Filter is used for an online update of the parameters related to the exponential load model.

In this landscape, in [52, 53], the IGSC algorithm is proposed to model any load with a DMC through an adaptive streaming algorithm that clusters the load values with minimal computational effort. This technique allows to model the considered load in different modalities: in this thesis, a representation in the active-reactive power plane and another one in the active-power-time plane will be shown (see Section 2.3.7 and Figure 2.6), but many other representations are possible.

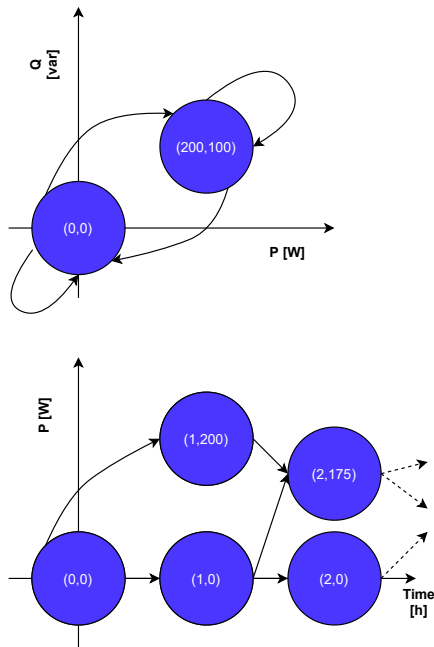


Figure 2.6: Different Load modeling modalities: P, Q (up) and P vs time (down)

Although other papers have tried to take advantage of the streaming data available from AMI, the proposed algorithm's main contribution is represented by its suitability for edge-computing (i.e., a distributed model on which computation is operated close to where data are gathered [109]). Indeed the results show that IGSC requires few computational resources while achieving satisfying clustering results. These properties allow IGSC to be applied within smart meters and other intelligent measurement and protection devices (e.g., [110]). Moreover, the resulting DMC represents a simplified model of the load, which can be sent to the cloud with few communication and memory resources. Also, the simplicity of the chosen probabilistic model makes the output highly interpretable. Overall, it is a good way of describing and visualizing the load history in a simplified way.

The main strong points of the presented algorithm can be hence summarized in:

- Exploitation of the streaming data provided by AMI;
- Low computational effort;
- Edge-computing applicability: data can be processed within smart meters or protection devices in a distributed configuration, rather than transferred to a data center (which is typically provided by a third-party that can charge for every additional computation);
- Limited request for memory and communication resources;
- Simplicity of the output. Indeed, a straightforward interpretation characterizes DMCs.

In addition, the proposed probabilistic load modeling procedure can be applied and used in different and important fields: from anomaly detection and condition monitoring to planning or sizing problems, as well as advanced control techniques (such as stochastic optimization or flexibility estimation/provision), data generation, and offline probabilistic modeling.

2.3.2 Proposed Approach

In this section, the background, the description, and some detailed applications of IGSC are presented.

2.3.3 Mathematical Background - Discrete Markov Chains

A Discrete Markov Chain (DMC) is a stochastic model that can describe a process through a sequence of possible events. It is composed of the following elements [111]:

- A set of N states $\{S_1, \dots, S_N\}$. The states in Figure 2.6 are represented by the blue circles;
- The transition matrix $A = \{a_{ij}\}_{i,j=1,\dots,N}$. This is a matrix of order N , wherein each element $a_{i,j}$ represents the probability of moving from state S_i to state S_j . Each row of A represents the probability distribution of the next state of the considered process conditioned on the current state. This implies that the sum of each row must be equal to 1. Looking at Figure 2.6, the presence of a blue arrow between two states S_i and S_j means that $a_{i,j} \neq 0$ and therefore that the transfer among S_i to S_j is possible. Otherwise, $a_{i,j} = 0$, i.e. the direct connection between the two states is impossible, and hence there is no arrow between the two states;
- the initial probability distribution related to the N states:

$$\pi = [\pi_1, \dots, \pi_2, \dots, \pi_N].$$

DMCs are a natural candidate for any temporal modeling. Despite their simple structure, they are a powerful tool for modeling several different temporal behaviors.

In load modeling and load profiling applications, a near real-time update of the transition probabilities can be of great interest.

To use DMCs for near real-time power system data streams modeling, it is crucial to face the following four issues, common to many data stream clustering problems [112]:

1. The continuity of power measurement is not trivially mapped into discrete states;
2. Data is not available all at once (and its order cannot be modified);
3. It is necessary to detect new states (which can arise at any given time);
4. Limited processing time. The computational time has to be lower than the data sampling time.

The proposed IGSC algorithm aims to resolve these issues with the following ideas:

1. The continuous measurements are discretized via a particular (and modified) topological mapping technique;
2. The algorithm is designed to process one value at a time (it is, therefore, an online algorithm);
3. New states can be added as soon as new data are available;
4. Its computational requirements are limited, and therefore the streaming data can be processed in real-time on simple hardware.

2.3.4 Designing the Algorithm

This section describes the proposed stream clustering algorithm, starting from a brief description of two similar algorithms present in the literature.

Since this approach is designed to perform load profiling in real-time, the algorithm must process one value at a time with minimal computational effort. Besides, the algorithm models the considered load with a DMC, clustering the continuous measurements through a topological mapping technique. Finally, the starting probability distribution is calculated by counting the number of samples in each cluster, while the transition matrix is evaluated by keeping track of all the transitions from one state to another one.

The algorithm is inspired by two other approaches, which serve as a starting point for the new algorithm, tailored for streaming applications:

- Instantaneous Topological Mapping (ITM)[113]. This algorithm aims to represent feature spaces, as an alternative to Self-Organizing Maps (SOM) [114] and Growing Neural Gas (GNG) [115].
- Growing Hidden Markov Model (GHMM), [116]. In this algorithm ITM is used for updating the structure of a HMM, in order to construct a GHMM.

The original ITM algorithm is described in the following.

Let ε and τ be positive real numbers. Let $d(\cdot, \cdot)$ be a distance measure between two data points. Let O_t be the current observation, i.e., the vector of the involved quantities. For instance, if active and reactive power are the features to be clustered, then $O_t = (P_t, Q_t)$, where P_t is the active power at

time t and Q_t is the reactive power at time t . The algorithm is composed of the following steps:

1. Matching step: find first and second centroids nearest to O_t (let them be f and s)
2. Weight adaptation: f centroid slides to O_t of a tiny amount (ε)

$$f_{new} = \varepsilon(O_t - f_{old}) \quad (2.1)$$

3. Edge Adaptation:

- (a) creates an edge between f and s , if it does not exist;
- (b) for each neighbour m of f_{new} : if O_t is in the circle having the segment $f_{new}m$ as diameter, i.e.:

$$d(O_t, \frac{1}{2}f_{new} + \frac{1}{2}m) < \frac{1}{2}d(f_{new}, m) \quad (2.2)$$

delete that edge (and delete also m if it has no other edges)

4. Node Adaptation:

- (a) if O_t is outside the circle having $\overline{f_{new}s}$ as diameter, that is:

$$d(O_t, \frac{1}{2}f_{new} + \frac{1}{2}s) > \frac{1}{2}d(f_{new}, s) \quad (2.3)$$

and $d(O_t, f_{new}) > \tau$, create new node with centroid equal to O_t and connect the new node and f_{new}

- (b) delete s if $d(f_{new}, s) < \frac{\tau}{2}$

As can be seen in the next section, the main ideas in this algorithm have been ported to IGSC, with suitable modifications for applying it to power system applications.

2.3.5 The Instantaneous Growing Stream Clustering Algorithm

The original ITM algorithm does not assign a probability to each link. Therefore, the IGSC algorithm modifies the ITM algorithm in order to create a DMC from the mapping, as described in the next paragraph.

The nodes (or states) dwell in the same space of the measurements (e.g. in the active power-time plane or the active power-reactive power plane) and are characterized by the mean of the measurements that happened to be closest to that state. For each state also, the number of measurements, their variance, and covariance between the variables are kept in memory.

The algorithm presents just one parameter: τ , a positive real number, which regulates how different from current states must be a new measurement in order to add a new state to the DMC. In the following, the steps of the algorithm are presented in detail, for the general case where n is the number of variables being measured (e.g., in Figure 2.6, $n = 2$).

Let

$$m(x) = [m_1(x), \dots, m_n(x)]$$

the mean vector of a state x ,

$$\sigma(x) = [\sigma_1(x), \dots, \sigma_n(x)]$$

the variances of a state x , and $\rho_{i,j}(x)$ the covariances within state x between variables i and j (with $i \neq j$), and $N(x)$ the number of measurements assigned to a state x . Let $O_t = [O_{1,t}, \dots, O_{n,t}]$ be the current observation, and x_{t-1} the last timestep state. The steps of the proposed algorithm for each incoming measurement are the following:

- Matching step: Find the two states closest to O_t . Let indicate these two states with F_t and S_t ;
- State Adaptation: create a new state with the same coordinates of O_t , connect it with F_t and assign it to x_t (current state) if:
 - O_t is outside the circle of diameter $F_t S_t$;
 - The Euclidean distance between O_t and F_t is greater than τ ;

otherwise do not create a new state, instead let $x_t = F_t$

- Weight adaptation: The state x_t , to which O_t has been assigned, is updated. For each variable $i = 1, \dots, n$, the mean is updated as follows:

$$m_i(x_t) := \frac{m(x_t) \cdot N(x_t) + O_t}{N(x_t) + 1} \quad (2.4)$$

With a similar formula the variance of each variable is updated for current state x_t :

$$\sigma_i(x_t) := \frac{\sigma_i(x_t) \cdot N(x_t) + (O_t - m_i(x_t))^2}{N(x_t) + 1} \quad (2.5)$$

And with another analogous formula the covariances of each distinct pair of variable $i, j = 1, \dots, n, i \neq j$ is updated for the current state x_t :

$$\begin{aligned} \rho_{i,j}(x_t) := & \frac{1}{N(x_t) + 1} \cdot \rho_{i,j}(x_t) \cdot N(x_t) \\ & + \frac{1}{N(x_t) + 1} (O_{i,t} - m_i(x_t)) \cdot (O_{j,t} - m_j(x_t)) \end{aligned} \quad (2.6)$$

Finally, $N(x_t)$ is increased by one.

- **Edge Adaptation:** A link between the past state x_{t-1} and the current state x_t is created if it does not exist. Moreover, the weights of all the links starting from the past state are changed to reflect the transition probability.

In the case of missing values, nothing is updated or checked, and at the next step, edge adaptation is omitted.

2.3.6 Application

Possible applications of the proposed IGSC in the context of the load modeling strategies may include:

- Visualizing the historical behavior of the load, and verify expected patterns or anomalous behaviors;
- Simulate likely scenarios for the load;
- Anomaly detection and condition monitoring by detecting newly found states and drifting probabilities. A newly found state can be linked to a new consumption pattern or some sort of failure. Also, the changes in probability are linked with usage changes;
- Extraction of probabilistic consumption patterns. Instead of storing all the consumption data, the Markov Chain can be used to summarize

an up-to-date probabilistic representation of the load behavior. With this approach, it is possible to reduce communication requirements and cloud computing memory;

- Near Real-Time load profiling. By having a probabilistic, updated, and reliable load profile, advanced stochastic optimization techniques can be applied to the system to which the load belongs (such as microgrid management, sizing, planning, flexibility provision);
- Having a probabilistic account of the process, a load disaggregation procedure could be devised if enough subloads are modeled.

2.3.7 Study Cases

In this section, two test cases used in the two published papers [52, 53] are summarized. For both study cases, active and reactive power data are sampled with one-second granularity.

The first test case is the external HVAC unit of the ABB Marine Laboratory (ABB Lab), located within the Department of Electrical Electronic Telecommunication and Naval Engineering (DITEN) of the University of Genova. The laboratory is used for both educational and research purposes. The rated power of this appliance is 3 kW, and the dataset spans the month of July 2019 (see Figure 2.7). Active and reactive power data are sampled with one-second granularity.

The second system is the HVAC system of the Education Sciences Department (DISFOR) of the University of Genoa. The electric HVAC system is composed of an Air Handling Unit (AHU), located at the top of the building, and a Chiller Unit, positioned in the surrounding area. Its aggregated rated power is around 100 kW. The dataset available was acquired between March and April 2018 for a total of 35 days (see Figure 2.8).

Another testing dataset has been the the IEEE Industrial Machine Dataset for Electrical Load Disaggregation (IMDELD) [117]. This database is composed of heavy-machine data collected in a Brazilian poultry feed factory. This plant produces at full-scale on the entire year, five days a week, Monday-Friday, and occasionally on Saturday, from 10 AM to 5 PM. Among all the measurements and devices, the active power measurements related to the following appliances have been used [117]:

- pelletizer II (rated power: 90 kW);

- exhaust fan I (rated power: 5 kW);
- milling machine I (rated power: 50 kW).

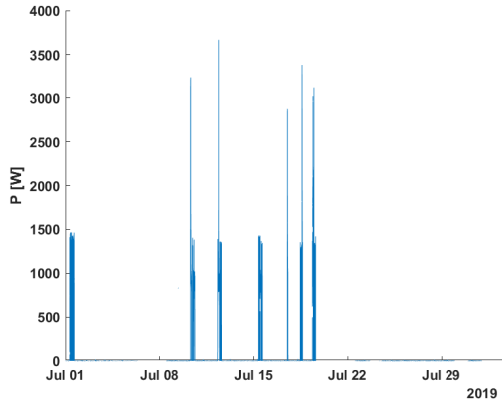


Figure 2.7: Active load profile of ABBLab study case

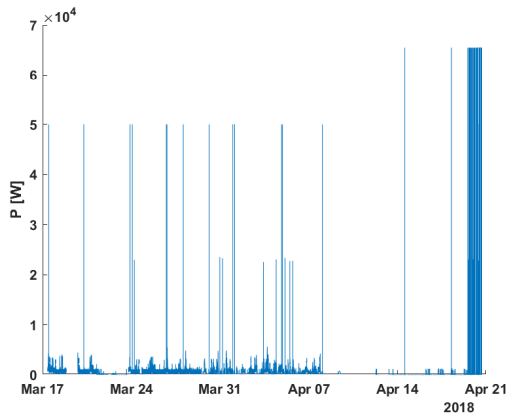


Figure 2.8: Active load profile of DISFOR study case

Samples have been collected from December 2017 to April 2018 for a total of 111 days. The considered profiles are illustrated in Figure 2.9.

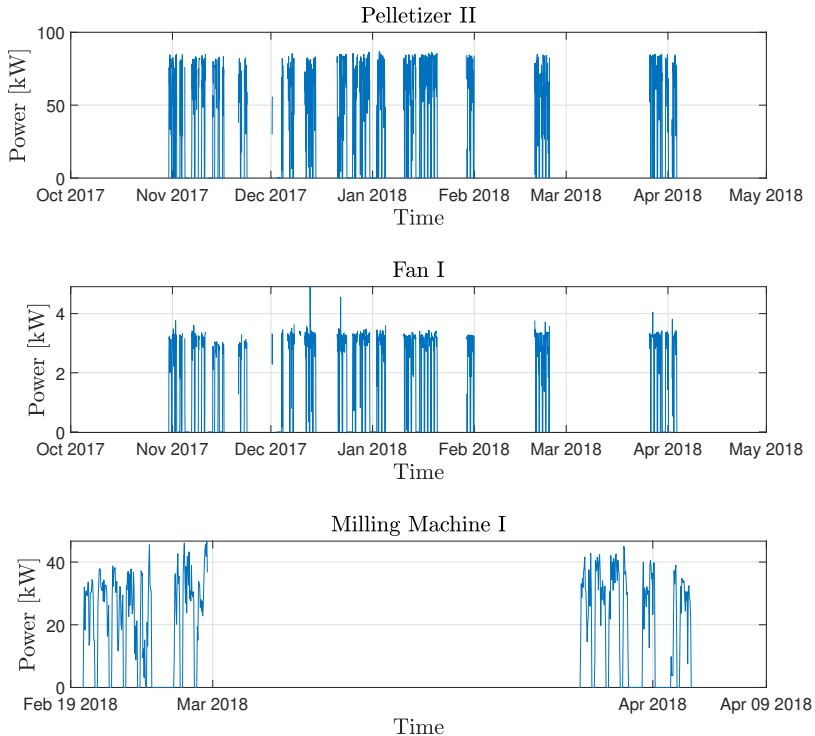


Figure 2.9: Active load profile from IEEE Industrial Machine Dataset for Electrical Load Disaggregation (IMDELD) dataset [117]

2.3.8 KPIs and Reference Methods

The results obtained by the proposed algorithm were evaluated through several Key Performance Indicators (KPIs):

- Maximum Processing Time (MPT), measured in milliseconds [ms]. MPT must be lower than the streaming rate, regardless of the number of

states identified. It evaluates the possibility of the proposed procedure to perform real-time modeling.

- Clustering Goodness of fit measures. Let \bar{x}_i be the cluster center of cluster C_i , and let N_C the total number of clusters. Let x a generic data point, N the number of data points, and $|C_i|$ the number of data points in the i -th cluster. Finally, let d^2 be the squared euclidean distance. The two goodness of fit indices employed in the published papers are the following:

- Calinski-Harabasz Index (CHI), defined as:

$$CHI = \frac{SSB}{SSW} \frac{N - N_C}{N_C - 1} \quad (2.7)$$

where SSB and SSW are respectively the between-cluster variance matrix and the within-cluster variance, respectively [118]. They are defined as:

$$SSB = \sum_{i=1}^{N_C} |C_i| \cdot d^2(\bar{x}_i, \sum_{j=1}^{N_C} \bar{x}_j) \quad (2.8)$$

$$SSW = \sum_{i=1}^{N_C} \sum_{x \in C_i} d^2(x, \bar{x}_i) \quad (2.9)$$

SSB measures how much the centroids are distant from each other, while SSW is related to how much the points in each cluster are concentrated. A good clustering requires (under the assumption that the clusters are convex) that SSB is as high as possible and SSW is as low as possible. Hence, the higher the CHI, the better the clustering.

- Davies-Bouldin Index (DBI):

$$DBI = \sum_{i=1}^{N_C} \max_{i \neq j} R_{ij} \quad (2.10)$$

where R_{ij} is the similarity between each pair of clusters, defined as

$$R_{ij} = \frac{s_i + s_j}{d^2(\bar{x}_i, \bar{x}_j)} \quad (2.11)$$

where s_i is the average distance between the samples in C_i , and s_j the same quantity in C_j . The DBI index is low if the R_{ij} are low, which is the case wherein the average distances within-clusters are low and the distance between the centroids are high. Hence, contrary to CHI, the lower the DBI index, the better the clustering.

- Skill Score (SS). For each of the previous KPIs, a skill score is defined:

$$SS_{CHI} = 1 - \frac{CHI_{ref}}{CHI} \quad (2.12)$$

$$SS_{DBI} = 1 - \frac{DBI}{DBI_{ref}} \quad (2.13)$$

where CHI_{ref} and DBI_{ref} are the CHI and the DBI attained by a reference method on the same dataset. The SS are a number in the interval $[-\infty, 1)$. They are 0 if the indexes are equal to the reference method; they are negative if the reference method scores better, and they are positive if the reference method has a worse score.

- Number of Clusters Identified (NCI) and Number of Markovian Transition (NMT) identified. This KPIs are important for studying the complexity of the algorithm as the granularity changes. It also affects the memory utilized by the algorithm.
- Estimated Physical Memory (EPM), measured in Bytes [B]. It is computed by assuming that each number of the Markov Chain is represented as a 64-bit (8 B). The EPM is computed as:

$$EPM = 8 \cdot N_C \cdot \alpha + 8 \cdot N_E \cdot \beta \quad (2.14)$$

where N_E is the number of non-zero elements of the transition matrix, and α and β are respectively the numbers stored for each cluster and for each non-zero element of the transition matrix, respectively. Assumed values are $\alpha = 6$ and $\beta = 4$.

For defining the SSSs, two other load modeling approaches have been implemented:

- *k*-Means [119]. Although it is not a streaming algorithm, it will serve as a reference for the skill scores. For each dataset, the *k* is chosen through the elbow method [120]. For the scenarios involving time, the clustering is run for each possible time instant.
- For [52], also Dynamical Clustering (DyCl), presented in [121] has been used. The comparison has been possible using the implementation made available by the authors on GitHub [122], with some modifications that reflect the need for computing the transition probability matrix of the Markov chain and the treatment of missing values in the datasets.

2.3.9 Experimental Results

The experiments carried out in the published papers are reported in this section. One set of experiments were performed on ABB Lab and DISFOR datasets, while another set on IMDELD. The two sets of experiments are reported separately in the following sections.

2.3.9.1 First set of Experiments

The data related to the DITEN and DISFOR test cases, described in Section 2.3.7, were derived from on-field measurements and stored in a time-series database in [52]. The data were then exported into daily files and concatenated in MATLAB. Abnormal data were treated as null values according to a threshold set on the raw data. (5 kW and 5 kvar for ABBLab, 10 kW and 10 kvar for DISFOR). Three different versions of each dataset were built for different time granularities: hourly, quarterly, minutely.

The algorithms were implemented on a Dell Precision Tower 3620, with an Intel(R) Core(TM) i7-6700 CPU @3.40GHz and 16 GB of RAM.

Two different feature sets, with different granularities, were tested: P and Q (i.e., active and reactive powers), as well as P and t (active power and time index: 1-24 for hourly scenarios, 1-96 for 15-minute scenarios, 1-1440 for scenarios with granularity equal to 1-minute: notice that 96 and 1440 are respectively the number of 15-minute intervals in a day and one-minute

Table 2.1: List of experimental scenarios

Scenario	Dataset	Granularity	Features	$[\tau, \delta, k_{max}]$
ABB-h-PQ	ABBLab	1h	P, Q	$[100, 10^4, 10]$
ABB-q-PQ	ABBLab	15 min	P, Q	$[100, 10^4, 10]$
ABB-m-PQ	ABBLab	1 min	P, Q	$[100, 10^4, 10]$
ABB-h-Pt	ABBLab	1h	P, t	$[100, 10^7, 10]$
ABB-q-Pt	ABBLab	15 min	P, t	$[100, 10^7, 10]$
ABB-m-Pt	ABBLab	1 min	P, t	$[100, 10^7, 10]$
DISFOR-h-PQ	DISFOR	1h	P, Q	$[100, 10^4, 10]$
DISFOR-q-PQ	DISFOR	15 min	P, Q	$[100, 10^4, 10]$
DISFOR-m-PQ	DISFOR	1 min	P, Q	$[100, 10^4, 10]$
DISFOR-h-Pt	DISFOR	1h	P, t	$[100, 10^7, 10]$
DISFOR-q-Pt	DISFOR	15 min	P, t	$[100, 10^7, 10]$
DISFOR-m-Pt	DISFOR	1 min	P, t	$[100, 10^7, 10]$

h=hourly, q=1 quarter, m=1 minute

intervals in a day). The complete list of the simulation scenarios is listed in Table 2.1.

Parameters τ (for IGSC), δ (for DC) and k_{max} (for k -Means) are also reported in Table 2.1. For each scenario, the τ for the IGSC has been set to 100, while the δ of the Dynamic Clustering has been set to 10^4 for PQ scenarios and to 10^7 for Pt scenarios. All other parameters of DC have been set to default.

In Table 2.2, for each scenario, the number of states identified and the number of transitions with positive probability are reported.

Table 2.2: Number of states and transitions identified

Scenario	States			Transitions		
	<i>IGSC</i>	<i>kmeans</i>	<i>DyCl</i>	<i>IGSC</i>	<i>kmeans</i>	<i>DyCl</i>
ABB-h-PQ	7	3	9	19	8	30
ABB-q-PQ	11	3	13	36	7	46
ABB-m-PQ	13	3	20	64	8	94
ABB-h-Pt	45	72	24	64	139	24
ABB-q-Pt	172	290	96	213	569	96
ABB-m-Pt	2305	4341	1440	2554	8711	1440
DISFOR-h-PQ	43	4	53	149	10	156
DISFOR-q-PQ	99	4	117	475	10	383
DISFOR-m-PQ	342	2	530	2455	4	2028
DISFOR-h-Pt	117	72	31	220	203	32
DISFOR-q-Pt	483	296	116	817	826	116
DISFOR-m-Pt	7415	4402	1671	11180	11317	1690

Firstly note that, as expected, the number of states and transitions identified increases as the granularity decreases. Moreover, the P, t scenarios have far more states and transitions identified: this is a consequence of the constraints in the case of clustering with the time variable. Also, the DISFOR scenarios present more states and transitions (a symptom of larger variability). The *k*-Means approach returns very few states and transitions in the PQ scenarios (see Figure 2.10). Finally, in all the ABB-Pt scenarios, the DyCl algorithm identifies exactly one state and transition for each time: it hence identifies a deterministic profile for those scenarios.

In Tables 2.3–2.5, the results of the experimental scenarios for each algorithm are presented. Notice that all the algorithms present a maximum time of computation lower than the sampling time, except for all the *k*-Means for the m-Pt scenarios (for which the sampling time is 60.000 ms).

Table 2.3: IGSC algorithm results (\uparrow : the higher, the better; \downarrow : the lower, the better. See Section 2.3.8 for KPI definitions)

Scenario	CHI \uparrow	DBI \downarrow	MPT [ms] \downarrow	EPM [kB] \downarrow
ABB-h-PQ	6.8E+03	0.3	77	1
ABB-q-PQ	1.3E+04	0.6	21	2
ABB-m-PQ	7.8E+05	0.5	23	3
ABB-h-Pt	8.3E+02	4.9	20	4
ABB-q-Pt	1.1E+03	5.0	22	15
ABB-m-Pt	9.9E+03	5.1	41	188
DISFOR-h-PQ	2.3E+05	0.5	15	7
DISFOR-q-PQ	1.5E+05	0.9	15	19
DISFOR-m-PQ	1.3E+06	1.3	21	93
DISFOR-h-Pt	6.8E+03	16.4	21	12
DISFOR-q-Pt	5.2E+03	16.2	21	48
DISFOR-m-Pt	7.3E+03	17.7	51	697

Table 2.4: DC algorithm results (\uparrow : the higher, the better; \downarrow : the lower, the better. See Section 2.3.8 for KPI definitions)

Scenario	CHI \uparrow	DBI \downarrow	MPT [ms] \downarrow	EPM [kB] \downarrow
ABB-h-PQ	1.2E+04	0.4	1	1
ABB-q-PQ	6.3E+04	0.3	57	2
ABB-m-PQ	1.0E+06	0.5	57	4
ABB-h-Pt	2.9E+00	20.4	44	2
ABB-q-Pt	3.2E+00	27.0	52	8
ABB-m-Pt	8.4E+01	54.4	52	113
DISFOR-h-PQ	5.1E+06	0.3	40	7
DISFOR-q-PQ	6.7E+06	0.3	40	17
DISFOR-m-PQ	3.0E+07	0.2	49	88
DISFOR-h-Pt	9.4E+02	63.2	13	2
DISFOR-q-Pt	7.0E+02	79.1	58	9
DISFOR-m-Pt	7.5E+02	125.2	52	131

In addition, the estimated physical memory, needed for the Markov Chain is very contained: it does not exceed the MegaByte after one month of continuous modeling.

Table 2.5: *k*-Means algorithm results (\uparrow : the higher, the better; \downarrow : the lower, the better. See Section 2.3.8 for KPI definitions)

Scenario	CHI \uparrow	DBI \downarrow	MPT [ms] \downarrow	EPM [kB] \downarrow
ABB-h-PQ	2.7E+03	0.3	271	0.4
ABB-q-PQ	8.9E+03	0.4	299	0.4
ABB-m-PQ	3.6E+05	0.3	2919	0.4
ABB-h-Pt	2.1E+02	2.3	4937	8
ABB-q-Pt	4.0E+02	3.9	23534	31
ABB-m-Pt	7.1E+03	2.4	14390781	476
DISFOR-h-PQ	1.3E+05	0.3	195	1
DISFOR-q-PQ	2.8E+05	0.4	210	1
DISFOR-m-PQ	1.9E+06	0.0	1921	0.2
DISFOR-h-Pt	2.6E+03	20.9	2505	10
DISFOR-q-Pt	2.3E+03	25.6	12741	40
DISFOR-m-Pt	2.8E+03	31.8	5300394	560

Moreover, Table 2.6 presents the Skill Scores of the IGSC and the DyCI algorithm for the *k*-Means results, as defined in Section 2.3.8.

We notice that the Skill Scores of IGSC are higher than those of DC in all the P, t scenarios, while DyCI outperforms IGSC in the P, Q scenarios, although when DyCI is outperformed, its SSs are worse than IGSC scores when IGSC is outperformed. Also, DBI is generally better for *k*-Means than for IGSC and DyCI, while the CHI is generally lower for *k*-Means. As for the granularities, there is no clear pattern, although the scenarios with granularity equal to one minute are typically lower than the other scenarios.

Figures 2.10–2.12 report the final Discrete Markov Chain and clusters of selected scenarios and methods.

While the *k*-means algorithms correctly identify the main three points of operations (idle, transient, and peak load), only the approaches involving DMC can characterize probabilistically how the state transitions happen, either during the day Figure 2.11 or together with the reactive power (Figure 2.12). In particular, the load has not been switched on during the night, while the peaks (at 7 AM, 9 AM, midday, and 2 PM) are identifiable in Figure 2.11; in addition, the peak is reached typically via the same path in the PQ plane, but sometimes the transition is more abrupt, which may indicate a particular shutdown or some missing data issue (Figure 2.12).

Table 2.6: Skill Scores (the higher, the better). For each KPI (See Section 2.3.8), **green** indicates best performance, **orange** worse than reference (k -means), but best between online methods, **red** worst performance between all methods, **black** better than reference but not best overall

Scenario	SS _{CHI}		SS _{DBI}	
	IGSC	dc	IGSC	dc
ABB-h-PQ	0.61	0.77	-0.06	-0.12
ABB-q-PQ	0.31	0.86	-0.73	0.21
ABB-m-PQ	0.54	0.65	-0.58	-0.49
ABB-h-Pt	0.74	-71.85	-1.15	-7.91
ABB-q-Pt	0.63	-124.30	-0.29	-6.00
ABB-m-Pt	0.28	-83.03	-1.16	-22.06
DISFOR-h-PQ	0.43	0.97	-0.66	0.06
DISFOR-q-PQ	-0.83	0.96	-0.92	0.37
DISFOR-m-PQ	-0.50	0.94	-36.69	-5.87
DISFOR-h-Pt	0.62	-1.74	0.22	-2.02
DISFOR-q-Pt	0.55	-2.33	0.37	-2.09
DISFOR-m-Pt	0.62	-2.68	0.44	-2.94

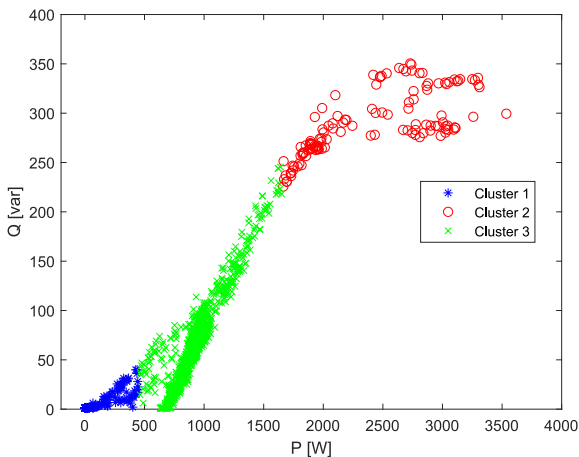


Figure 2.10: Result of k -Means algorithm in ABB-m-PQ scenario

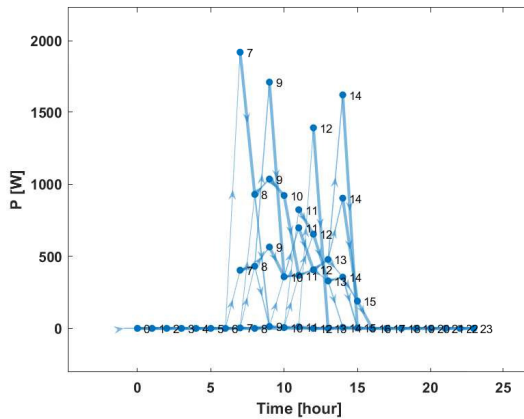


Figure 2.11: Markov Chain of IGSC algorithm in ABB-h-Pt scenario. Labels refer to the centroids time value

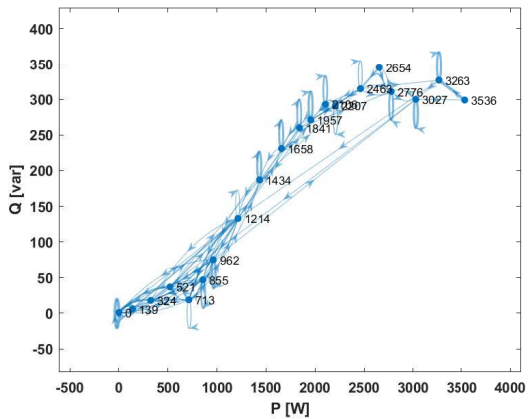


Figure 2.12: Markov Chain of DC in ABB-m-PQ scenario. Labels refer to the centroids active power value

2.3.9.2 Second set of Experiments

The dataset for the second set of experiments was downloaded from the IEEE Dataport site and preprocessed in MATLAB. For each considered device (pelletizer II, exhaust fan I, milling machine I), several scenarios were considered, with hourly granularity (see Table 2.7).

Table 2.7: List of experimental scenarios [53]

Scenario	Device	Granularity	τ
PEL-h-0.1	pelletizer II	1h	0.1 %
PEL-h-1	pelletizer II	1h	1 %
PEL-h-5	pelletizer II	1h	5 %
PEL-h-10	pelletizer II	1h	10 %
PEL-h-20	pelletizer II	1h	20 %
PEL-h-50	pelletizer II	1h	50 %
PEL-h-70	pelletizer II	1h	70 %
FAN-h-0.1	exhaust fan I	1h	0.1 %
FAN-h-1	exhaust fan I	1h	1 %
FAN-h-5	exhaust fan I	1h	5 %
FAN-h-10	exhaust fan I	1h	10 %
FAN-h-20	exhaust fan I	1h	20 %
FAN-h-50	exhaust fan I	1h	50 %
FAN-h-70	exhaust fan I	1h	70 %
MIL-h-0.1	milling machine I	1h	0.1 %
MIL-h-1	milling machine I	1h	1 %
MIL-h-5	milling machine I	1h	5 %
MIL-h-10	milling machine I	1h	10 %
MIL-h-20	milling machine I	1h	20 %
MIL-h-50	milling machine I	1h	50 %
MIL-h-70	milling machine I	1h	70 %

h=hourly

The focus of this set of experiments was the sensitivity analysis of the parameter τ . This has been performed to have practical guidelines on the algorithm implementation and exploit its full potential. In practice, τ represents the distance among two centroids belonging to the same time interval. Considering τ as a percentage of the rated power, small values of τ identify a great number of states. However, at the same time, it would lead to a slow down of the algorithm, threatening its scalability and the chances to perform

real-time modeling. On the other hand, higher values of τ (close to 100% of the rated power) may not identify meaningful patterns. The following values of τ have been tested: 0.1%, 1%, 5%, 10%, 20%, 50%, and 70% of the rated power, for a total of 21 scenarios.

In Tables 2.8–2.10 the KPIs related to the simulations results are reported.

Table 2.8: Results of IGSC for the second set of experiment scenarios
(↑: the higher, the better; ↓: the lower, the better. See Section 2.3.8 for KPI definitions)

Scenario	CHI ↑	DBI ↓	States ↓	Transitions ↓	MPT ↓ [ms]	EPM ↓ [kB]
PEL-h-0.1	732.82	38.74	499	940	281.81	52.76
PEL-h-1	1293.11	60.39	295	760	263.24	37.57
PEL-h-5	1528.37	52.26	154	469	187.85	21.87
PEL-h-10	819.58	95.51	106	276	150.02	13.59
PEL-h-20	478.09	83.98	65	140	40.61	7.42
PEL-h-50	339.99	156.40	48	85	83.37	4.90
PEL-h-70	95.18	99.34	40	60	36.20	3.75
FAN-h-0.1	1490.81	7.44	448	910	60.80	49.43
FAN-h-1	2164.56	12.60	223	542	36.12	27.39
FAN-h-5	1874.30	23.43	120	242	32.99	13.18
FAN-h-10	1082.93	25.07	76	124	32.47	7.43
FAN-h-20	934.61	37.20	55	89	29.52	5.35
FAN-h-50	482.17	43.19	47	70	31.65	4.39
FAN-h-70	36.22	203.52	24	25	30.31	1.90
MIL-h-0.1	229.12	19.67	191	277	37.15	17.60
MIL-h-1	286.76	17.48	166	261	38.44	15.93
MIL-h-5	280.02	19.99	119	219	35.25	12.42
MIL-h-10	214.72	27.69	93	167	30.69	9.57
MIL-h-20	119.94	68.25	70	113	38.63	6.81
MIL-h-50	34.05	45.38	45	63	34.10	4.07
MIL-h-70	8.88	116.70	28	31	31.39	2.28

Table 2.9: k-Means algorithm results (\uparrow : the higher, the better; \downarrow : the lower, the better. See Section 2.3.8 for KPI definitions)

Scenario	CHI \uparrow	DBI \downarrow	States \downarrow	Transitions \downarrow	MPT \downarrow [ms]	EPM \downarrow [kB]
PEL-h-0.1-70	683.56	81.94	73	168	4892.18	8.67
FAN-h-0.1-70	1544.06	12.61	80	159	3704.15	8.71
MIL-h-0.1-70	127.73	37.14	72	134	3774.88	7.56

Table 2.10: Skill Scores for the second set of experiments (the higher, the better). For each KPI (see Section 2.3.8), **green** indicates best performance across all alternatives of τ of the same load, **red** performance worse than k -means, **black** better than reference but not best overall

Scenario	SS _{CHI}	SS _{DBI}
PEL-h-0.1	0.06	0.52
PEL-h-1	0.47	0.26
PEL-h-5	0.55	0.36
PEL-h-10	0.16	-0.16
PEL-h-20	-0.42	-0.025
PEL-h-50	-1.01	-0.90
PEL-h-70	-6.18	-0.21
FAN-h-0.1	-0.03	0.40
FAN-h-1	0.28	0.00
FAN-h-5	0.17	-0.85
FAN-h-10	-0.42	-0.98
FAN-h-20	-0.65	-1.95
FAN-h-50	-2.20	-2.42
FAN-h-70	-41.62	-15.13
MIL-h-0.1	0.44	0.47
MIL-h-1	0.55	0.52
MIL-h-5	0.54	0.46
MIL-h-10	0.40	0.25
MIL-h-20	-0.06	-0.83
MIL-h-50	-2.75	-0.22
MIL-h-70	-13.38	-2.14

We discuss the results across three dimensions: KPIs, use cases, and different levels of τ . The KPIs are the same as the first set of experiments, that is, the same ones of Section 2.3.8. As the reference method, only k -means is used in this set of experiments.

All the tests have been implemented on a Dell Latitude E6540, equipped with an Intel(R) Core (TM) i7 @ 2.70 GHz and 8 GB of RAM.

As previously noted, MPT and EPM KPIs are important to assess the suitability of the algorithm to low-hardware streaming applications. Both KPIs are well under critical values (see Table 2.8). Clearly, for low levels of τ these two KPIs are higher: indeed, values of τ close to 0% of the rated power present a high number of states. This can be perceived graphically in Figures 2.13–2.15, where the Markov Chains (MCs) of three scenarios ($\tau = 1\%$, 5%, 10%) are depicted.

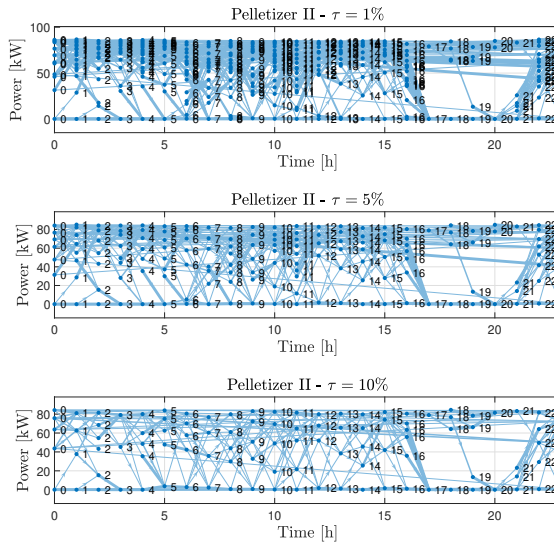


Figure 2.13: Results of the proposed algorithm - Pelletizer II - $\tau = 1\%$, 5%, 10%

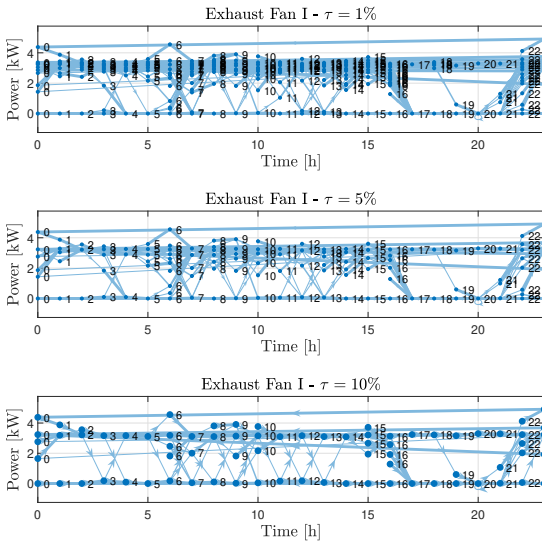


Figure 2.14: Results of the proposed algorithm - Exhaust Fan I - $\tau = 1\%, 5\%, 10\%$

In these figures, as expected (see Section 2.3.7), a frequent shutdown of the factory can be noticed in the late afternoon. The highest value for both MPT and EPM is achieved by PEL-h-0.1 scenario. This may be linked to the higher nominal power of the involved load. MPT is higher in the case of k -means while the EPM is limited (see Table 2.9), but this is to be expected, as it is an offline algorithm.

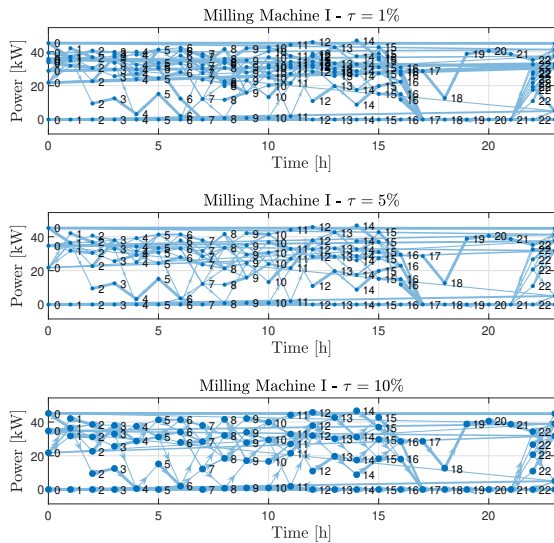


Figure 2.15: Results of the proposed algorithm - Milling Machine I - $\tau = 1\%$, 5% , 10%

Figures 2.16–2.18 report the result of the k -means clustering (it should be noted that the k -means approach is not influenced by τ , therefore only one plot per device is illustrated).

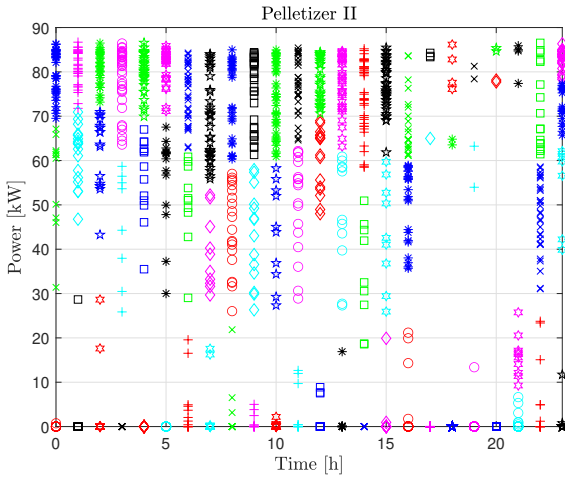
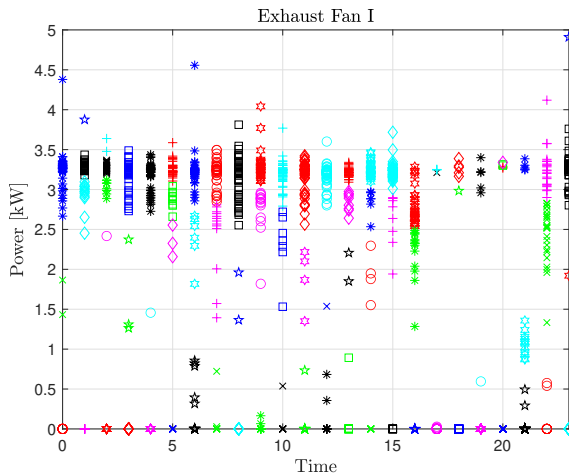
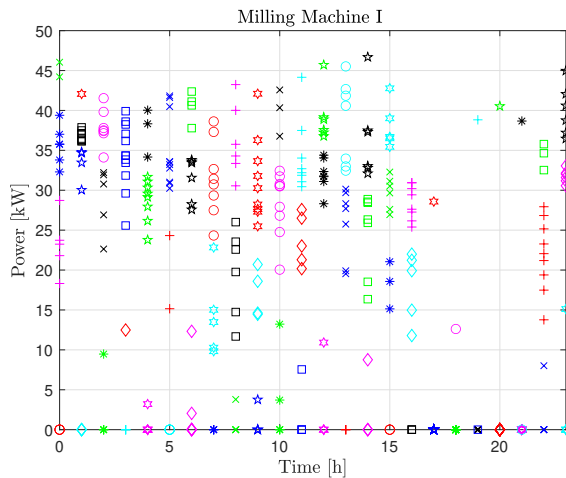


Figure 2.16: k -means results - Pelletizer II

Figure 2.17: k -means results - Exhaust Fan IFigure 2.18: k -means results - Milling Machines I

In Figures 2.16–2.18, the different colors in each time step indicate a different cluster. Notice that the number of clusters may vary for each interval.

From these experiments, some comments can be made on the choice of the τ parameter for IGSC.

From Table 2.10, it is possible to notice that the k -means algorithm is better than the proposed approach (negative SS) for values of τ that are greater or equal 20% of the rated power. For $\tau = 10\%$, results are mixed (positive for the milling machines, negative for the fan, and one positive and one negative for the milling machine). For the cases wherein $\tau = 5\%$, it is possible to observe positive results, except for the SS_{DBI} related to the fan. For τ under 5%, only positive skill scores can be noticed (except for the Fan-h-0.1, though the SS is only slightly negative). The highest SS_{CHI} is achieved twice by $\tau = 1\%$ (for the fan and the milling machines) and once by $\tau = 5\%$ (pelletizer). For SS_{DBI} , $\tau = 0.1\%$ emerges as optimal twice (pelletizer and fan), while $\tau = 1\%$ is the best choice in the remaining case. In general, the FAN subcase has lower scores than the other devices. This may be linked just to the fact that k -means achieves its best results with the fan database (see Table 2.5 and Figure 2.17).

2.4 Chapter Conclusions

This chapter introduced descriptive and diagnostic analytics, a broad set of analytic approaches to be aware of a system. Descriptive and diagnostic analytics leverage several tools, such as visualization, modeling for understanding, modeling for data generation, among others. Subsequently, some literature applications of this type of analytics have been introduced. Finally, a new streaming clustering algorithm, named Instantaneous Growing Stream Clustering (IGSC), and published during the Ph.D. [52, 53] has been proposed. The presented technique can detect new clusters in time, and it constructs a Discrete Markov Chain (DMC) for modeling the transitions between the identified clusters. The algorithm can model any pair of variables (including the time itself) to construct probabilistic profile models of relevant variables.

Applications of this algorithm are particularly interesting in the field of power systems. Indeed, smart meters and other intelligent devices can be exploited to construct synthetic power consumption models. These are interesting for utilities that currently resort to costly third-party cloud resources for processing smart meters data.

The proposed algorithm has been tested with two datasets, two different load modeling modalities (active power/reactive power and active power in time), and three granularities, for a total of 12 experimental scenarios [52]. The results of these tests have been compared with the outputs of similar algorithms found in the literature, in particular, Dynamical Clustering (DyCl) [121], modified for constructing the Markov chains, and with offline k -Means approach. The KPIs show that the proposed algorithm respects time constraints for all the investigated scenarios, and the resulting MC required a moderate memory. The proposed algorithm appears to have higher performance in the profiling scenarios, while DyCl achieves better results in the P, Q scenarios. These performances need further investigations to be generalized and explained.

In [53], a sensitivity analysis on the parameter (τ) of the IGSC algorithm has been presented for the case of load profiling. The study is performed on an industrial open dataset [117]. The loads present different ranges of active power, but normalizing τ through the rated power, an analysis of the parameter selection could be proposed. Values of τ greater than 5% never achieved the best results, while values greater than 10% never beat standard k -means. The smallest τ considered (0.1%) had good performances, but at the same time presented the highest Maximum Processing Time (MPT), which might

lead to scalability problems as the data keep coming. This is consistent across the loads considered, so one could argue that these conclusions are general, although more case studies should be considered to confirm this assertion. In addition, the optimal τ can vary according to the considered KPI. Thus, further works should concentrate on finding KPIs and benchmarks for specific applications of the algorithm, also by scoping the existing literature on similar algorithms developed in fields other than power systems. Also, the sensitivity of τ for other modeling problems such as PQ or even trivariate modeling and beyond could be investigated. Finally, the influence of the considered distance measure on the outcome of the algorithm could be explored.

Predictive Power System Analytics

*An approximate answer to the right problem
is worth a good deal more
than an exact answer to an approximate problem.
- Prof. John Tukey*

Predictive analytics deals with methods aimed at answering the question:

- What will happen to the system?

Hence it deals mainly with predicting the unknown, especially regarding the future. Therefore specific terminologies will be introduced for framing the general predictive analytics problem and related lines of research. Then a brief overview of forecasting applied to power systems will be given. Lastly, some case studies investigated during the Ph.D. will be presented.

Before starting, it should be noted that there are other ways to frame and define predictive analytics.

In particular, in [123] passive and active predictive analytics are distinguished. Passive predictive analytics does not involve intervention on the

system. It can be achieved by simply modeling the correlation between current and past data with future data (or between known and unknown data). On the other hand, active predictive analytics involves actively intervening in the system to obtain the desired outcome. Hence, in this case, modeling as precisely as possible the causal relationships between the past and present inputs and the future quantity (or between known and unknown quantities) is crucial. For example, predicting the Market Clearing Price (MCP) for a wholesaler that needs to submit bids to the market auction is different from predicting the same quantity for a retailer with no access to the bidding market since the wholesaler is contributing with its prediction in the formation of the predicted quantity.

Another example is weather forecasting as opposed to climate forecasting. Weather forecasting is a passive prediction since it is nearly impossible to change short-term local weather variables. On the other hand, climate forecasting is an active prediction since long-term temperature averages can be changed if proper policies at the global level are enforced.

In this chapter, passive predictive analytics will be discussed, leaving active tasks for the chapter of prescriptive analytics. It is worth mentioning this distinction, though, since predictive analytics is often tightly linked with prescriptive analytics: failing to recognize the final use for which predictive analytics solutions are developed can lead to suboptimal decisions. In other words, in this chapter, forecasting is intended as an open-loop problem [124]: subsequent decisions do not alter the quantity to be predicted.

Also, the same paper [123] points out correctly that the predicted quantity can involve not a future quantity but simply an unknown quantity. An example can be found in the third application (Section 3.5), where transfer learning is used to estimate the people present in a room monitored with a thermal camera. However, the same paper [123] is calling this type of predictive analytics "diagnostics", which was used in Chapter 2 to refer to tasks aimed at understanding the causes of observed events (see Chapter 2). So it would be confusing to use here in this thesis. Therefore, the remainder of the thesis will employ a specific name (forecasting) for predictive analytics regarding the future, leaving prediction as a generic term for both forecasting and other predictive analytics problems.

3.1 Cross-Industry Standard Procedure for Data Mining

While there is no common process for descriptive and diagnostic analytics to follow and come to a definitive solution, predictive analytics has standard procedures for continuously developing, deploying, monitoring, and improving models. In particular, a widely employed framework for predictive analytics is the Cross-Industry Standard Procedure for Data Mining (CRISP-DM) [54, 55], depicted in Figure 3.1.

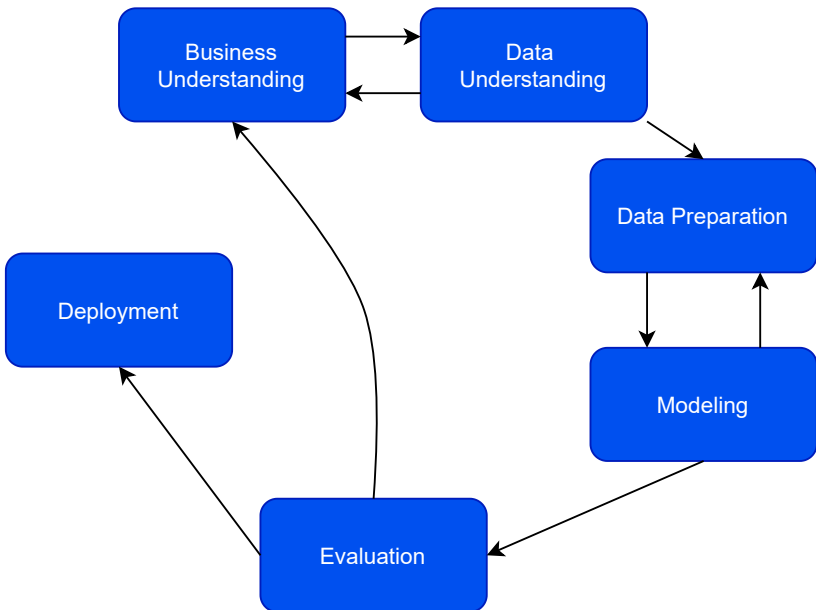


Figure 3.1: Scheme of the Cross-Industry Standard Procedure for Data Mining (CRISP-DM) [54, 55]

The steps of the CRISP-DM procedure are the following:

- **Business Understanding.** This is the phase where the business problem is analyzed and, if it can be framed as a data mining problem, the right dataset is collected in order to develop the solution. At the same time, the right KPIs to be used are selected in order to evaluate the model correctly;

- **Data Understanding.** In this phase, the chosen dataset is preliminarily analyzed in order to check inconsistencies, either by building simple models or using visualizations. It is mostly the realm of descriptive and diagnostic analytics;
- **Data Preparation.** In this phase, the data is preprocessed before being used to develop a predictive model;
- **Modeling.** In this phase, the model is developed. Typically, the final one is chosen among a pool of candidate models;
- **Evaluation.** The chosen model is evaluated using meaningful KPIs, the ones decided in the business understanding phase. If the evaluation is satisfying, then it can be deployed. Otherwise, the cycle restarts. It may well be the case that the evaluation can give another light to the problem to be solved: maybe the KPIs were not in line with the business problem, or even the data mining problem itself does not solve the business problem (or the business problem itself has to be reframed);
- **Deployment.** If the model passes the evaluation phase, the model is put into production. This is where the monitoring of the model begins, since the environment where the model runs can have characteristics that could not have been observed in the previous steps, either because of the limit of the experimental setting or because the environment is changing altogether.

In the following subsections, each phase will be broken down into several aspects that arise in most, if not all, predictive analytics tasks. Since their prominence in power system analytics, particular focus will be given to forecasting tasks since their prominence in power system analytics.

3.1.1 Business Understanding

In this phase, the business problem is dissected. It is a crucial phase: even the most rigorous data mining procedure is useless if it solves a non-existing problem. Hence, the real problem must be identified and framed as a data mining problem. There is no standard guideline, from an analytics perspective, on how to identify the right problem to solve. It typically involves maximizing some good measure such as profit or minimizing negative impacts, be it environmental, energetic, reputational, or financial.

However, some aspects of business problems translate easily into tasks for the predictive analytics task. For example, if the problem is inherently fast-changing or nonstationary, an online predictive model may be a clever solution. Online ML is the branch of ML that studies the models that update their parameters as soon as new data becomes available. This is opposed to offline (or batch) ML in which a model is trained using historical data and then, once validated, the same model is used, without a continuous update of parameters [125].

The forecasting problem, in particular, can be framed in various ways, as the quantities to be forecasted depend on the business problem specifics. For example, for a Distribution System Operator (DSO) the load forecasting problem can take very different characteristics, as exemplified by Figure 3.2. Data granularity (i.e., the frequency of the data) and forecasting horizon (i.e., how far in time must the model forecast), in particular, are very important in defining the subsequent CRISP-DM steps.

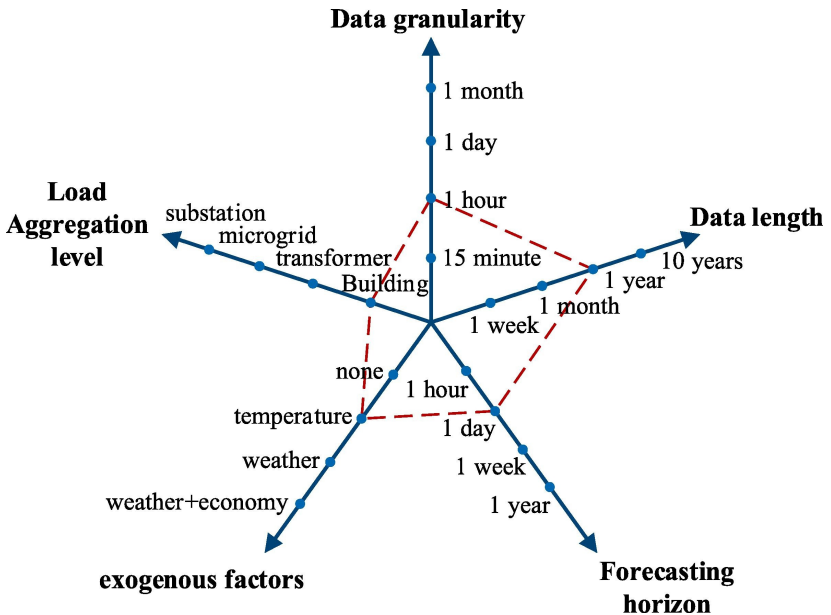


Figure 3.2: Various types of load forecasting tasks [126]

A typical distinction is made in various fields between very short-term, short-term, medium-term, and long-term forecasting, as shown in Figure 3.3 for power systems.

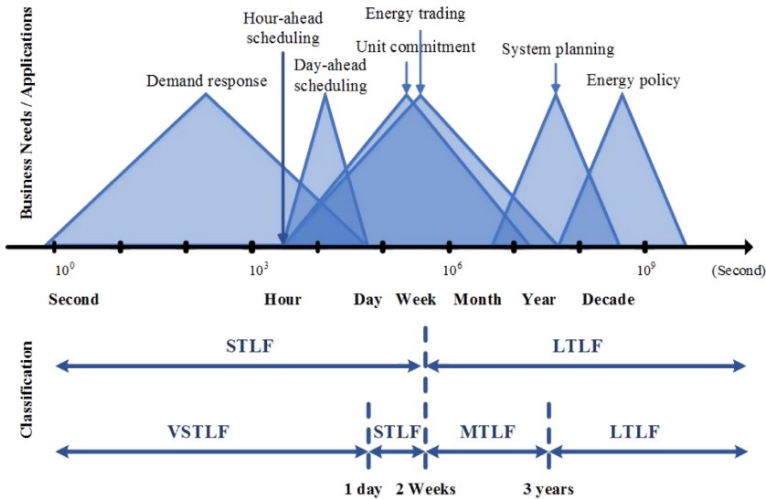


Figure 3.3: Different forecasting horizon definitions and their business and operational objectives [127]

The distinction is somewhat subjective, and it is mainly a matter of business objective: indeed very short-term forecasts are typically required for operations, short-term forecasts for operational planning, medium-term forecasts for risk hedging and maintenance, while long-term forecasts for technical and financial planning, as well as for policy design [128]. But from a technical point of view, it mainly affects the relative importance of historical data as inputs, with very short and long-term being the ones with the higher relative importance of historical data.

Furthermore, some tasks are induced by the business problem itself: for example, in electricity markets, an important task is the day-ahead forecasting, that is, the forecast issued for the next calendar day (which should not be confused with 24-hour ahead forecasting, which forecasts the value to be observed 24 hours after the forecast origin [124]).

Another important aspect is the level of aggregation for the forecasting task. Indeed, a quantity (such as load) can be split into subquantities (typically several subloads plus losses [124]), and the question is whether it is necessary to forecast the subquantities or it is sufficient to forecast the principal quantity. Indeed, typically the higher the aggregation level, the easier the forecasting tasks, both in terms of forecasting KPIs and in terms of complexity of the solution. On the other hand, a disaggregated forecast delivers more information, and if the subquantities at different levels can be reconciled, the accuracy of the forecast can increase ([86, 124]).

Also, the notion of "good" prediction must be defined, both in terms of meaningful KPIs and in terms of the magnitude of a satisfying predictive analytics solution. In power systems, a small percentage of error can lead to several hundred dollars loss for a utility financial bottom line [129].

Finally, as perfect predictions are impossible to obtain, one has to define whether a measure of the uncertainty regarding the predictions is needed. If a measure of uncertainty is given, forecasting is told probabilistic. Otherwise, the problem is a point forecasting one. Typically, probabilistic forecasting is to be preferred, as it has been shown to lead to better decisions [130]; however, it is also a less mature field, and the user of the forecast must be able to understand the output and act based on it to leverage its potentiality. In any case, the majority of use-cases (especially in power systems) benefit from the additional details given by probabilistic forecasts [131]. The importance of uncertainty qualification of point forecasts is paramount in the context of risk constraints, such as the ones involved in optimal bidding and bilateral contracts in liberalized markets [132].

Probabilistic forecasting can be in the form of quantiles, prediction intervals, or density functions [127], and can be issued in many ways:

- statistics of deterministic errors [131];
- conformal prediction [133];
- error variance modeling, as well as skewness and kurtosis modeling [131];
- quantile regression [134];
- bootstrapping [135];
- kernel dressing [136];

- specific models that give predictive intervals (such as xgboost) [137];
- copula models [138].

Finally, another aspect to be investigated is whether to frame the problem as a regression problem, a classification problem, or another type of supervised problem (or even a semisupervised or unsupervised problem).

This may arise, for example, when the quantity to be forecasted is difficult to predict, and hence a simplification of the problem is needed. The question here is whether the simplification still answers the business problem at hand. This is seldom encountered in power system analytics. However, there are instances where it may apply: for example, for ramp direction forecast [139], or in forecasting the future state of a balancing market. Indeed, balancing markets are very tough to predict but typically have four states: no regulation; only upward regulation; only downward regulation; or both [140].

3.1.2 Data Understanding

Once the type of predictive analytics task is identified and the corresponding data is collected, a closer look at the available information must be given. This can be done mainly with the approaches presented in Chapter 2. Moreover, additional care must be taken to understand whether the data has some biases that must be taken care. An example in load forecasting can be the case where a demand response program is in place. In that case, the load profile is not the "passive" load, but it is the load on which an action has been taken, and hence some caution must be taken in using the data without any modification [124].

Another aspect to observe is the quantity and the typology of data since many modeling techniques have different performances with different quantities of data (although, except in particular instances [141], the more the data, the better). Specific preprocessing techniques are needed for different typologies of data (such as unstructured data).

Also, the identification of bad data is an important aspect of this phase. Bad data is a concept relative to the question to be answered: a bad data instance for a task may not be bad for another one. Measurement errors in input data may be very important in estimating causal relationships but have less impact in predicting the response variable *per se* [142].

Another way in which data can be bad is if data is not up-to-date, meaning that the model developed may not be relevant for deployment.

It is not trivial to identify bad data, as expert knowledge and intuition are often necessary. Also, changes in measurement procedure or bad choices of missing value codes can give rise to bad data. This is further complicated by the fact that the presence of errors can be proven, but its absence is not [142].

3.1.3 Data Preparation

Once the data have been explored, they can be prepared for the modeling phase. There are many subphases here that can be tackled in different orders. This phase together can take a large percentage of time and effort involved in all the CRISP-DM process [143]. One is dealing with missing data.

Missing data can mean that an entire record is absent or that individual features of a record are missing [142]. It is very common in real-world datasets [144] and can severely hinder the performance of the desired models.

There are several types of missing data:

- Missing Completely At Random (MCAR) when the probability of a data point being missing is completely independent of any variables in the dataset;
- Missing At Random (MAR) when it is independent of its own (missing) value but probabilistically dependent on the value of another variable;
- Missing Not At Random (MNAR) when it is dependent on its own (missing) value.

Ways to deal with missing data are numerous as well:

- Complete Case Analysis (where data with partial information are discarded). This may shrink the dataset size sensibly and can produce biased results if data is MAR or MNAR;
- Imputation, where missing data are filled with simple substitutes (such as previous available value). However, this is still not a good choice if the missing data is a high proportion of all the data [144].

Other issues arise that deserve careful attention in developing a forecasting solution. For example, nonstationary data is typically difficult to model with ML techniques without applying various techniques of detrending. Similarly, seasonality has to be adequately addressed (such as with calendar features that

can capture seasonality), as well as calendar effects of holidays, weekends, and leap days. Also, peaks and high volatility should be properly treated (all very common characteristics of electricity markets [132]).

After bad, missing, and peculiar data is filtered out, inputs for predictive models are to be constructed. This step is called feature extraction, which is tightly connected with feature transformation techniques. This is necessary for optimal modeling, as predictive modeling can lead to poor results with some features but not with other features. This happens because different feature representations give different geometric shapes to the data, and hence they may have very different relationships with the desired output [145]. However, finding a suitable data representation is very difficult because it is very domain and dataset-specific [146]. Several common approaches exist, though. Since the type of data being analyzed influences the feature extraction possibilities to a great extent, the focus is on time series data feature extraction techniques. They involve:

- feature learning (such as standard or recurrent neural networks);
- concepts of signal enhancement, such as smoothing or de-noising. Fourier and Wavelet transforms in this category are among the most popular;
- local features methods, that encode local sequence characteristics into features (such as moving average values);
- linear and non-linear feature transformations, such as PCA, MDS, and t-SNE, described in Chapter 2;
- non-linear expansions, such as creating the square of existing features, adding interaction terms (i.e., multiplying existing features in new features)
- feature discretizations;
- normalization and standardization techniques;
- in general, basically any function of the input time series (see for example [147] for other types of features available for time series).

It is sometimes advised to include more than to exclude features since typically feature selection is the step done immediately after feature extraction [146].

Indeed, after extracting the features, feature selection is needed to retain the "relevant" features while discarding the irrelevant ones [148]. Also, in this context, the relevancy is problem, domain, and dataset-dependent. Moreover, two individually irrelevant features may be relevant if both are included [146]. The notion of relevance itself is related to the objective being pursued. An irrelevant feature for short-term forecasting may be relevant for long-term forecasting and vice versa. Also, a good feature selection may improve generalization capability and prevent overfitting (a phenomenon that happens whenever the model mimics too closely the training set, and thence the final model is unable to retain the signal helpful for predicting a new validation or test set).

Feature selection methods can be classified by their output or their relation with the predictive model. They can either produce a subset of the original feature or rank all the original features (in the latter case, a threshold is needed for actually selecting the features).

Nevertheless, the common distinction relies on their relationship with the predictive model. Three types of feature selection methodologies can be distinguished:

- Filter methods, which are independent of the predictive model;
- Wrapper methods, which use an evaluation of the predictive model trained with the selected features to evaluate the subset of features;
- Embedded methods, which perform feature selection in the training of the predictive model itself (examples of these kinds are decision trees and lasso regression).

While methods exploiting the predictive model (i.e., wrapper, embedded) lead typically to higher performances, they are also more computationally expensive. There are also approaches to use different feature selection methods by combining their output. They can be classified into homogeneous and heterogeneous ensemble methods. The former group uses the same feature selectors on different subsets of the data set, while the latter uses different selection methods on the same dataset. But many other possibilities of ensembling are possible [148].

All these steps have been named called "methodologies" in order to distinguish them from the actual predictive models, which are called techniques in [127]. The distinction is helpful because one can use the same exact predictive technique with many different methodologies and arrive at very different

accuracy levels and conclusions (although many times methods, methodologies, and techniques are used as synonyms in the literature). It is also a line of research less explored than using the same methodology with many different techniques. The different types of techniques are discussed in the following.

3.1.4 Modeling

In this step, the actual predictive model is developed. Here the possibilities are countless, as a lot of different choices can be made for the same task:

- Choice of the technique. There is a great number of techniques one can choose. Since also, in this case, the "best" technique is task, domain, and dataset-dependent, it is utopic to hope to find the "optimal" technique [127]. In any case, apart from empirical experimentation, several fundamental choices have to be done, for example:
 - Structured or Unstructured techniques. Structured techniques hypothesize a data generating process, while unstructured techniques try to learn it from data. Unstructured techniques can be competitive in the presence of a high quantity of data (typically the more, the better) [149]. Sometimes this distinction is referred to as statistical/machine learning techniques, although this is more a cultural, subjective distinction [149, 150]. The choice is not trivial, especially in forecasting, where typical assumptions made for unstructured techniques, such as that data are Independent and Identically Distributed (IID) are not held. Also, hybrid solutions are possible [145].
 - Physical/Statistical techniques. In instances where a physical model can capture the prediction, a physical model (i.e., derived from physical equations) can be preferred to a data mining solution. This is the case, for example, for PV forecasting. Also, in this case, hybrid techniques are available [56]. Physical consideration can also help in the data preparation phase [124], and in any case, statistical postprocessing is often used when using data that come from Numerical Weather Prediction (NWP) models.
- Global or Local approaches. If the task regards the forecasting of several time series (such as load forecasting at various levels of a distribution system), a choice must be made on whether the same model on

each time series is to be developed, or a single model, leveraging all the single time series is to be preferred [149];

- ”Integral” or Decomposition approaches. Especially in forecasting, the time series to be forecasted can be decomposed in a sum of more time series, capturing longer and shorter dynamics. Each decomposed time series is then predicted with one or more techniques and their results combined [151]. Similarly, data can be mapped into regions, for each of which a specific model is trained;
- In the case of multi-step ahead forecasting, several schemes can be devised [152]:
 - Multi-Input Multi-Output (MIMO) strategy;
 - Recursive strategy, where a one-step-ahead prediction model is developed and then used recursively, using its previous step forecast as new input;
 - Direct strategy, where a model is developed for each lead time to be forecasted.

With whichever technique and data preparation phase, a number of hyperparameters (a number that defines the parameters to be learned) is always to be chosen. For example, when using a fully-connected one hidden layer neural network, the number of hidden units, the transfer functions, the training algorithms, and other options are available before one can actually learn the neural network model from the data. With the advent of Big Data (BD) and in Deep Learning (DL) in particular, the number of hyperparameters can literally be immense. So, the problem of choosing in a reasonable amount of time and with limited model evaluations a good choice of the hyperparameters is not trivial. This also applies in modern ML systems, as there is a persistent push toward automatizing ML pipelines, in what is called Automated Machine Learning (AutoML), where the problem of hyperparameter choice is extended by including all the possibilities of the Data Preparation step. The problem is known as HyperParameters Optimization (HPO). In DL setting, the problem is also known as Neural Architecture Search (NAS). In the context of AutoML, also Full Model Selection (FMS) and Combined Algorithm Selection and Hyperparameter (CASH) concepts are encountered.

There are four main strategies for solving the HPO problem [153]:

- Grid Search, also known as full factorial design. For each set of hyperparameters, a finite set of values is defined: the cartesian product of these sets is evaluated. It is the simplest method, but it is not convenient to use in the presence of many hyperparameters, as the number of evaluations quickly explodes as the number of hyperparameters increases;
- Random Search is a simple alternative, which samples configurations at random until a specific budget (such as number of evaluations or time) is exhausted. It has been shown to be more efficient than grid search. It is also easier to parallelize;
- Genetic Search methods. They can be thought of as guided random search methods. Indeed they start from a population (a set of configurations) and improve them by applying local perturbations (mutations) and combinations of different members (crossover) to obtain a new generation of better configurations;
- Bayesian Optimization. These approaches are black-box global optimization iterative algorithms with two key ingredients: a probabilistic surrogate model and an acquisition function to decide which point to evaluate next. This type of optimization can model integer, categorical, and conditional hyperparameters. Advanced versions include constraint handling (for example, involving memory consumption, training time, prediction time, energy usage) and ways to try configurations on a subset of the training data to speed up computations.

Finally, a decision must be made on whether to use one predictive model or an ensemble of predictive models. Strategies for ensembling are bagging, boosting, stacking [148]. Ensembling is convenient when different forecasts are comparable either in performance or diverse [154]. Diversity can be achieved using forecasts issued with different techniques or data preparation steps (methodologies). Ensembles will typically work worse if data are either too volatile or with a significant amount of outliers. Ensembles may indeed learn a poor model from the outliers. In general, the advantage is not that the performance is better than the best of the single forecast, but because it is much more difficult to issue a bad forecast if the predictive analytics solution does not rely on a single model [155].

Concluding this section, in Table 3.1 the data preparation and the modeling steps of the CRISP-DM framework are summarized.

Table 3.1: Summary of Data Preparation and Modeling steps

CRISP-DM phase	Subphase	Options
Data Preparation	Bad Data Management	Complete Case Analysis; Imputation; outlier removal
	Feature Extraction	Feature Learning; Signal enhancement; local feature methods; feature transformation; non-linear expansions; discretizations; normalizations; standardizations; etc.
	Feature Selection	Filter; Wrapper; Embedded methods.
	Selection Ensembles	Homogeneous; Heterogeneous.
Modeling	Choice of Technique	Structured or Unstructured; Physical or Statistical; or Hybrid.
	Approach	Global or Local; Integral or Decomposition; MIMO; recursive; or direct.
	Hyperparameter Selection	Grid Search; Random Search; Genetic Search; or Bayesian Optimization.
	Ensemble Method	Bagging; Boosting; Stacking; others.

Of course, Table 3.1 is not exhaustive, as the number of options is very large. Also, each option has many suboptions that enlarge the total number of possibilities a modeler has to choose from in these two phases. In any case, the main modeling decisions to be taken are listed.

3.1.5 Evaluation

Once a model has been developed, it must be evaluated. The task is not easy for several reasons. First, it would be advisable to measure the capacity of generalization of the model, that is, the distribution of the error (or some

functional of it) that is to be expected in applying the model to new data. This can be done only if a hold-out validation set is used or a cross-validation procedure. The cross-validation procedure is the ideal procedure to use, but it means running the same model several times.

In forecasting, the procedure must be taken with particular care since the main assumption of cross-validation standard procedures (i.e., the data being IID) are not met. There are several versions of cross-validation for time series [156], with respect to the forecast origin, which is the time point of the last known value:

- Fixed-Origin: only one forecast is issued for each time point of the test set (this is the same as standard training/test procedure)
- Rolling-Origin-Recalibration: forecast origin is moved in training set and model recalibrated;
- Rolling-Origin-Update: the same model is used while forecast origin is moved, and recalibration does not take place;
- Rolling-Window-Evaluation: same as Rolling-Origin-Recalibration, but the amount of training data is always the same (i.e., less recent data are discarded).

An alternative method is given by the techniques for which the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC) can be computed [157].

Also, particular care must be given to not using the same data for HPO and evaluation, as this may lead to an optimistic value for the evaluation. Indeed, typically, data used for HPO is included in a subset of the training set called validation set, separate from the final evaluation test set.

Second, a meaningful functional of the error measure must be chosen. Though it has been done in the business understanding phase, it is worth mentioning that every commonly used functional has some issue. Ideally, it should be scale-dependent (e.g., Mean Absolute Error (MAE) is not), but the most common scale-dependent measures (such as Mean Absolute Percentage Error (MAPE)), which rely on a reference value, have issues on zero or undefined values, or they have a skewed distribution towards the reference value [156]. Asymmetry can be corrected, but at the expense of the symmetry with respect to the magnitude of the error (e.g., Symmetric Mean Absolute Percentage Error (sMAPE) has this characteristic).

Problems with zero value can be mitigated by using a benchmark model as a reference, but again, in general, the problem persists (e.g., Relative Root Mean Square Error (RELRMSE)), and if the problem is overcome, this happens at the expense of interpretability (e.g., with Mean Absolute Scaled Error (MASE)).

A solution might be to evaluate the same model with different error measures. Nonetheless, another problem arises: since the model has been developed minimizing only one of them, how to arrive at a single conclusion on the model to select is not easy [158].

In general, it is very important to benchmark the model with simple models, naive models, classic models, established models, or state-of-the-art models, if possible: this helps to understand the potentiality of the selected models. It is also important to be fair in comparison by devoting the same resources (i.e., computing time) to each competing model [124]. In this regard, automatic HPO procedures are beneficial. Also, statistical tests exist to determine if two error measures are statistically different from each other [159], as also the possibilities that two models behave similarly (i.e., either is equally performant) must be taken into account [124, 128].

3.1.6 Deployment

Once the model has passed the evaluation phase, it is ready to be deployed (if deployment requirements have been taken into account, as they should have been in the business understanding phase). Now the monitoring phase takes place, where the actual performance is measured. This is to be done in order to understand whether the model is behaving correctly or some sort of concept drift (or persistent missing data) is happening that requires the model to be retrained [144, 160]. This is also important for online methods, as some of them, such as neural networks, can pick up new relationships at the expense of forgetting old ones [161].

3.2 Predictive Applications to Power Systems

Predictive analytics applications in power systems can be leveraged by all types of users, mainly energy market participants and power system operators. While the former is concerned with optimal purchase and selling of energy, the latter is concerned with reliability, security, and economic operation of the grid [131].

Because of the applications being presented afterward, this section will be mainly devoted to PV forecasting and load forecasting, although these are not the only applications of predictive analytics in this domain.

For example, price forecasting is quite a crucial task since the types of prices, markets (day-ahead, intra-day, balancing, and ancillary [128]), and uses are very large and also policy-dependent. Hence, a method developed for a specific market may not be transferable to a similar market of another nation, although similarities do exist. Price is also a complicated quantity to be forecasted because of spikes, and the fact that it is an inherently an active predictive analytics problem (in the sense of [123], meaning that forecasts contribute to the trading strategies of market participants [19, 132]).

Another important application that will not be reviewed in this section is wind power forecasting: it is a pretty mature field, especially in the context of probabilistic forecasting [162].

3.2.1 Common Issues

A big challenge facing the predictive analytics in power systems is how to deliver disaggregated forecasts and predictions with the highest possible performance, as this will greatly enhance the optimal operation of the whole system, a wider involvement in the market, and demand response programs (with, ideally, feasible net-load forecasting [131]). The quick adoption of smart meters is believed to be helpful in solving the challenge [23].

Also, at the retail level, efficiency in single facilities such as factories or buildings can be achieved by a comprehensive suite of predictive analytics products, integrating consumption patterns with weather predictions, energy prices, occupancy rates, and RES forecasting [39].

Since the challenges are all common and ultimately interconnected in the context of decarbonization objectives, a need for data sharing, all the while respecting the privacy of the data issuers, has been pointed out by many authors [131].

3.2.2 PV Forecasting Applications

Many PV power forecasting techniques have been developed through the years. A reliable forecast is the key for several smart grid applications [163], such as: optimal dispatch, [164], active demand response, grid regulation [165] and intelligent energy management [166].

PV forecasting represents a large research topic which can be characterized by the time horizon related to the prediction [167]:

- Very short/short term forecasting, wherein the time horizon varies from seconds to 24-48 hours;
- Medium-term forecasting, which analyzes periods up to one month;
- Long-term forecasting, wherein the horizon can be set to 1-10 years.

A first distinction is made between direct and indirect forecasting methods [167]. In an indirect forecasting approach, the solar irradiance is forecasted and then exploited in commercial PV simulation software to predict the PV power generation. Direct methodologies, on the other hand, aim at predicting the PV output directly. A comparison between the two strategies can be found in [168], wherein the results show that direct methods perform better.

Among direct methods, the most common classification is between physical models, statistical approaches, and hybrid methods that blend physical and statistical techniques. The physical methods rely on the fact that the PV power production is mainly affected by two physical variables: the amount of irradiance reaching the panel and the PV modules temperature [169, 170]. The main advantage of physical models is that they do not require any historical data to be implemented.

Among these, the day-ahead horizon is crucial for the scheduling of the conventional generation, and many national grid codes (e.g., [171], [172]) require a punctual and precise RES forecasting. In addition, in countries with a day-ahead electricity market, large RES plants can act as producers providing sale bids, wherein the actual production has to follow a scheduler offer that is provided through a forecasting approach.

Two examples of physical methods for the day-ahead horizon that represent the cell through an electrical circuit are described in [173]. The two models are compared with a hybrid technique. The results show that the hybrid model outperforms both physical models.

Still, the most popular approaches are the statistical ones. They consist of the exploitation of historical data to train any typology of data-driven techniques such as time series, regression models, machine learning, or deep learning architectures. Statistical models are more versatile than physical ones, and they are easier to set up. The most used statistical methods are the Artificial Neural Networks (ANN) [174].

This popularity derives from their effectiveness, which in turn stems from the high nonlinearity of relationships between PV production and its related variables [175]. Also, classical statistics models can lead to satisfactory results, possibly combined with ensemble techniques [176].

Any kind of mixture of physical and statistical approaches is a hybrid method. Physical methods can become hybrid if statistical techniques are used to correct their systematic errors. On the other hand, statistical approaches that use physical methodologies for the design of input variables can be considered hybrid. Examples of this last kind of hybrid method are described in [177] and [178]. In [179], an example of each of the two types of hybrid methods is developed. The literature also presents solutions that are called hybrid because of a particular combination of two statistical techniques. For example, [180] uses two different algorithms for the training of a feed-forward neural network, while in [181], nearest neighbor and ANN are exploited for fulfilling the forecasting task. However, in practice, they are static methods.

3.2.3 Load Forecasting

One of the most important quantities to be predicted and forecasted is the load demand: it is a common problem, and a key factor in electrical systems [127]. Load forecasting is throughout all electric power industry segments, including generation, transmission, distribution, and retail. Applications of load forecast include power supply planning, transmission, and distribution systems planning, demand-side management, power systems operations, maintenance, financial planning, and rate design. [127, 182]. In general, a reliable load forecasting functionality is essential for the coordination of the uncertainty brought by renewable generation units with actual demand, also considering the diffusion of active demand response programs, flexibility markets, and load shaping strategies [58]. In addition, load forecasting is useful for grid security, providing crucial information in order to determine in advance imbalanced and vulnerable scenarios [58]. Finally, in a microgrid scenario, accurate forecasting techniques and methodologies are crucial in the planning, design, and operation phases [183]. In this context, load forecasting plays a fundamental role in enhancing the management and usage of the conventional and renewable mix, and it helps to improve the economics of energy exchange for easy integration of renewables into utilities.

The number of methodologies, approaches, and techniques studied for tackling this task is remarkable. The number of load forecasting reviews is also considerable. Figure 3.4, taken from [184] gives a glimpse of how the various techniques are used for different horizons and granularity of the load forecasting problem.

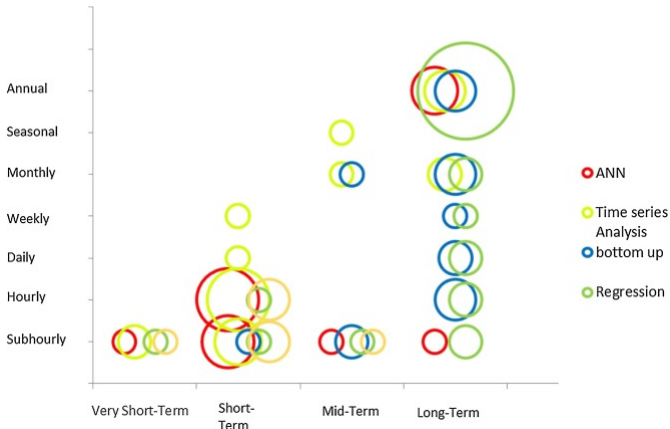


Figure 3.4: Techniques used for each horizon and data granularity for the load forecasting task [184]

While in short-term applications, fine granularities and time series analysis, as well as Artificial Neural Networks (ANNs) are typically used, in long-term applications, regression techniques and more detailed (not data-driven) approaches (“bottom-up” in Figure 3.4) are preferred.

Despite the great amount of research in the load forecasting realm, the problem is still worth to be studied. Indeed, new factors complicate now the study of the problem: for example, customers are becoming more price-responsive [124], and behind-the-meter RES, as well as demand response programs and increasing use of EVs, are rendering load forecasting more complicated than before. Also, the human activities served by the load play a major role [86], as well as weather, for which the optimal source of data to be used is not easy to find [185]. For this growing complexity, new approaches such as meta-learning [126], and deep learning [186] are being investigated, among many others.

Both PV and load forecasting are crucial for optimizing the grid operations (for DSOs) as well as maximizing the revenue and savings from energy efficiency solutions, as well as for minimizing the use of carbon-intensive energy sources by leveraging, for example, BESSs. As a matter of fact, the outputs of load and PV forecasting (and in general, RES) are a crucial input for energy management algorithms of the types described in Chapter 4.

In the next sections, two applications of PV forecasting and two applications of load (and energy) forecasting will be delineated, originating from works published during the Ph.D. [56, 57, 60] and from unpublished work done within the Ph.D. for the PODCAST project (see Projects section at Page 255).

3.3 Application: PV Forecasting

In this section, it will be presented an innovative PV forecasting procedure developed during the Ph.D. and published in [56, 57].

The contribution is about an innovative hybrid forecasting technique for the power output of a PV system. The proposed procedure has been validated on a real PV plant, located in Genoa, Italy. The innovation of the proposed hybrid forecasting method consists of the combined modality of physical and statistical approaches. In particular, two physical models and one statistical method are developed, while the proposed hybrid technique chooses among them for the PV prediction according to the day ahead weather forecasting.

Several works have already presented a technique based on the selection of different models, trained with data coming from different types of days. For example, [187] describes a method composed of four support vector regressors, one for each identified meaningful weather condition. In [188] five neural networks coupled with a harmony search algorithm are used according to a fuzzy k -means clustering technique. In that work, a fuzzy inference approach is used according to the weather prediction. In [189], days are divided into two groups: for each of them, an ensemble of ANNs is developed. In the forecasting phase, one of the two ensembles is utilized. In [190], data are grouped in 6 clusters using the variance of five differential sequences of weather Key Performance Indicators (KPIs). Each cluster is used to train, through a back-propagation algorithm, a neural network which is employed for the PV forecasting. In [191]. k -means clustering and gated recurrent unit are employed respectively for classification and prediction tasks.

Finally, the specific problem related to the classification of the day-ahead weather condition is addressed in [192] and faced through k -nearest neighbor and support vector machines.

In all these cases, weather prediction is used to select among different statistical methods. The innovative contribution of this technique lies in the selection between a physical model (Clear Sky Model (CSM)), a statistically corrected physical model (Corrected Clear Sky Model (CCSM)), and a statistical approach (Basic Ensemble Method of Neural Networks (BEM)). The proposed approach represents a novel hybrid method for the PV forecasting because it is neither a corrected physical approach nor a statistical technique that uses inputs from a physical model. It is a methodology that, according to the day-ahead weather forecast, may use a physical or a statistical approach, differently from all the above-mentioned hybrid strategies, which select among techniques of the same type.

The weather forecast can be used for the selection of the most appropriate method in different modalities. The selection is made with decision tree algorithm in [56] because of its easy implementation and straightforward interpretation, while in [57], a simplified procedure has been proposed: the decision rule is a single rule, identified through linear regression, that chooses between a physical approach and a data-driven technique. The decision in [56] is based on different weather variables, while in [57] it is based on the day-ahead forecasting of the sole Cloud Cover Index (CCI).

The CCI is a number that measures the percentage of the considered sky portion which is covered by the clouds at a given time. It ranges from 0 to 100, where CCI=0 indicates a cloudless sky, while CCI=100 indicates a weather condition of a completely covered sky.

Through the analysis of the resulting decision rule, it is possible to verify the rationality of the proposed approach. The proposed method is described in detail in Section 3.3.2 and depicted, for the integral case in Figure 3.5, while the simplified procedure is represented in Figure 3.6.

The main idea is that if the sky is predicted to be clear, a CSM, based on well-known sun equations, is exploited; alternatively, a data-driven methodology should work better. This approach, if implemented correctly, should achieve high accuracy since physical models are very precise on clear sky days, while ANNs and statistical methods are more reliable on the other days. The prediction is performed in MATLAB with a 15 minutes granularity and with a horizon of 24 hours, starting from midnight (thence being a day-ahead horizon).

3.3.1 Key Performance Indicators - KPI

This section collects all the KPIs used to assess the quality of the proposed forecasting procedure and for the selection of all the parameters of the methodologies described in subsequent sections.

The base for all the considered KPIs is the prediction error. It is defined as the difference between the forecast and the measured variable at time t :

$$\varepsilon(t) = x_{meas}(t) - x_{forec}(t) \quad (3.1)$$

From this equation, it is possible to observe that positive errors correspond to an underestimation of the actual value.

The second KPI proposed for this application is the Root Mean Square Error (RMSE) [169]. It is defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (\varepsilon(t))^2} \quad (3.2)$$

where N is the samples number. The RMSE can be normalized, obtaining an estimation of the percentage error. This KPI is called normalized Root Mean Square Error (nRMSE) and can be evaluated as:

$$nRMSE = \frac{RMSE}{\sqrt{\frac{1}{N} \sum_{t=1}^N (x_{meas}(t))^2}} \quad (3.3)$$

The Mean Bias Error (MBE) is defined as the mean difference between the prediction and the measurement:

$$MBE = \frac{1}{N} \sum_{t=1}^N \varepsilon(t) \quad (3.4)$$

It represents the systematic part (bias) of the error: if it is positive, the model has the tendency to underestimate the actual value (forecasts are on average over measurement values); if negative, it overestimates it (forecasts are on average under measurement values).

Finally, the Skill Scores (SSs) [193] measure the accuracy of a forecasting technique with respect to the precision of a reference methodology. The SS can be defined for different KPIs. In this application it has been calculated as:

$$SS = 1 - \frac{|KPI_{proposed}|}{|KPI_{reference}|} \quad (3.5)$$

where $KPI_{proposed}$ represents an estimation of the accuracy of the proposed approach, while $KPI_{reference}$ is the same KPI evaluated on a reference method.

The range of the SS is $[-\infty, +1]$. A positive value of SS implies that the proposed technique provides a better result with respect to the other approach, while a negative value corresponds to the opposite situation. Notice that a $SS = 1$ represents the perfect forecast.

3.3.2 Proposed Approach

The methodology of [56] will be firstly presented; then, the simplification introduced in [57] will be treated.

The method proposed in [56] consists of two Decision Rules (DeRus) and three Sub Models (SubMs). Figure 3.5 describes the flowchart of the proposed hybrid approach, where figure diamonds represent the decision rules, while rectangles are the available forecasting approaches. The implemented SubMs are:

- Clear Sky Model (CSM), based on well-known sun equations;
- Corrected Clear Sky Model (CCSM), a linear model which combines CSM and cloud cover index;
- Basic Ensemble Method of Neural Networks (BEM), which uses outputs of multiple ANNs.

As can be seen from Figure 3.5, first, a decision based on day ahead weather forecast is made on whether to use the BEM or a deterministic model. In case the choice ends up being a deterministic model, a second decision must be made between the CSM and the CCSM.

The forecasting has been designed with a day-ahead horizon and a granularity equal to 15 minutes (which represents the standard monitoring interval adopted in Italy, where the test site is located : see Section 3.5.1). Notice that the output of the proposed method is composed of 96 values representing the PV output for the 24 hours of the next day.

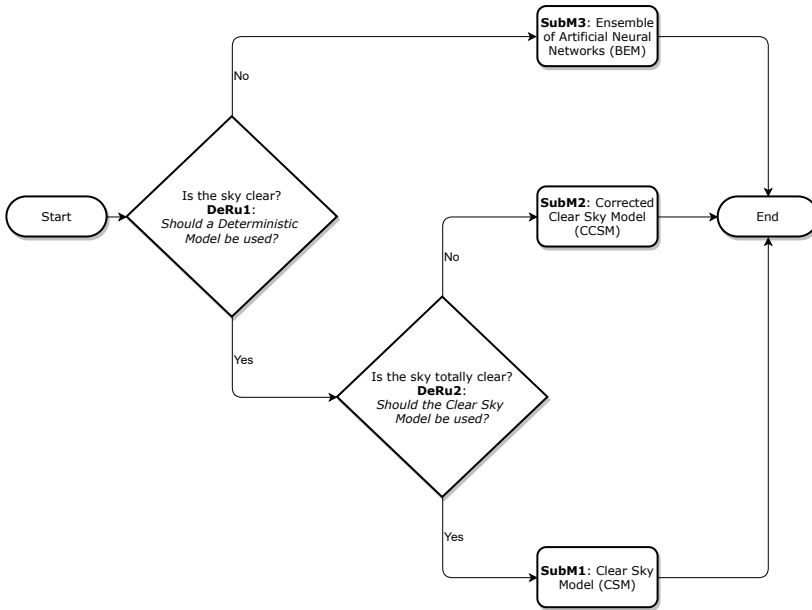


Figure 3.5: General scheme of the technique proposed in [56] (complete)

In [57] instead of a BEM, a single ANN is used, and a single decision rule chooses between the ANN and the CSM models. The proposed procedure is depicted in Figure 3.6.

The rest of this section describes the various part of the proposed strategy. In particular, in Sections 3.3.3 and 3.3.4 the deterministic approaches (CSM and CCSM) are described; in Section 3.3.5 the BEM is presented; finally, in Section 3.3.6 the proposed hybrid methodology is described.

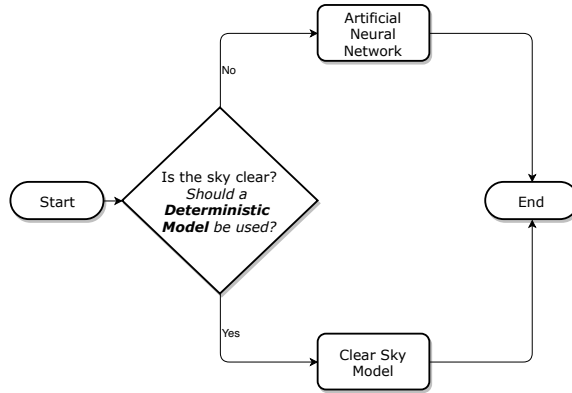


Figure 3.6: General scheme of the technique proposed in [57] (simplified)

3.3.3 Clear Sky Model - CSM

When the sky is clear, the PV system is not shaded by any cloud. In this case, there is little to no uncertainty in the PV output profile. Thus, a deterministic model can be set up for covering scenarios with this weather condition [194], [167]. Thus, the predicted PV output ($P_{system}(t)$) is modeled as follows [195] (notice that (t) indicates the time dependency):

$$P_{system}(t) = \frac{E_{g,pv}(t) \cdot P_{peak} \cdot \eta_{pan}(t) \cdot \eta_{inv} \cdot \omega_{DEG}(t)}{E_{STD}} \quad (3.6)$$

where: $E_{g,pv}(t)$ is the global irradiance on the plane of the array $\left[\frac{W}{m^2}\right]$, P_{peak} represents the total rated peak power of the solar panel [kW], $\eta_{pan}(t)$ is the relative efficiency factor of the panels [p.u.], η_{inv} indicates the relative efficiency factor of the inverter [p.u.], $\omega_{DEG}(t)$ represents the coefficient of degradation [p.u.] and E_{STD} is the irradiance of standard test conditions $\left[\frac{W}{m^2}\right]$.

The parameters P_{peak}, η_{inv} are related to technical data of the PV plant, while E_{STD} is a constant equal to $1000 \frac{W}{m^2}$. The parameter $\omega_{DEG}(t)$ can be either set to prescribed values taken from large scientific reviews (e.g., see [196]) or estimated through field measurements [197].

Thus, the main variables of this model are $E_{g,pv}(t)$ and $\eta_{pan}(t)$. Global irradiance is the sum of three components, weakened by the shading factors.

These take into account the possible shadows due to the surrounding buildings [198]:

$$E_{g,pv}(t) = (1 - S_{dir}(\alpha(t), \gamma(t))) \cdot E_{b,pv}(t) + (1 - S_{diff}) \cdot E_{d,pv}(t) + E_{r,pv}(t) \quad (3.7)$$

where:

- $E_{b,pv}(t)$ is the beam irradiance reaching the plane of the array $\left[\frac{W}{m^2}\right]$;
- $E_{d,pv}(t)$ represents the diffuse irradiance reaching the plane of the array $\left[\frac{W}{m^2}\right]$;
- $E_{r,pv}(t)$ is the reflected irradiance reaching the plane of the array $\left[\frac{W}{m^2}\right]$;
- $\alpha(t)$ denotes the sun azimuth [deg.];
- $\gamma(t)$ represents the sun elevation [deg.];
- $S_{dir}(\alpha(t), \gamma(t))$ is the direct component shading factor [p.u.];
- S_{diff} denotes the diffuse component shading factor [p.u.].

In particular, $E_{r,pv}(t)$ is calculated as follows:

$$E_{r,pv}(t) = \begin{cases} E_{g,hor}(t) \cdot \rho_g \cdot \left(\frac{1 - \cos(\beta)}{2}\right) & \text{for } \frac{1 - \cos(\beta)}{2} > 0 \\ 0 & \text{o/w} \end{cases} \quad (3.8)$$

where:

- $E_{g,hor}(t)$ represents the global horizontal irradiance $\left[\frac{W}{m^2}\right]$;
- ρ_g is the ground albedo [p.u.];
- β denotes the tilt angle of the PV array [deg.].

The shading factors calculation ($S_{dir}(\alpha(t), \gamma(t))$, S_{diff}) and the ground albedo (ρ_g) are specific to the site and described in Section 3.3.7.1, while the irradiance components are calculated as in [199].

The efficiency of the panel is mainly affected by modules temperature. In this application the following equation has been utilized for its estimation [200], [195]:

$$\eta_{pan}(t) = 1 + \beta_c(T_m(t) - 25^\circ C) \quad (3.9)$$

where β_c is the module temperature coefficient (see Table 3.5).

For the module temperature estimation, several models have been considered [201–206]. The performances of these models have been compared on a test set composed of clear sky days, and the following relation has been selected [201]:

$$T_m(t) = T_a(t) + \frac{E_{g,pv}(t)}{E_{STD}} \cdot (0.0712 \cdot W_s(t)^2 - 2.411 \cdot W_s(t) + 32.96) \quad (3.10)$$

where $T_a(t)$ is the ambient temperature [$^\circ C$] and $W_s(t)$ is the wind speed [$\frac{m}{s}$].

3.3.4 Corrected Clear Sky Model - CCSM

The model described in the previous paragraph supposes that the sky is completely clear. This means that its performance can be improvable if the presence of clouds is considered. Thus, a modified version of CSM has been developed. It consists of a Stepwise Linear Regression (SLR) model [207], trained on clear sky or almost clear sky days, with regressors composed of CSM output and CCI.

In order to estimate nonlinear behaviors, the variables have been taken up to the fifth power.

The SLR is a linear regression where the regressors are selected through an automated procedure that iteratively adds and removes regressors by testing their statistical significance through a hypothesis test on their corresponding coefficients, measured by the p-value of an F-statistic [207].

The proposed procedure is the following:

1. Fit an initial model with only the constant term;

2. Add to the model the candidate regressor with the smallest p-value, provided that it is smaller than a predetermined entrance tolerance. Repeat this step until no regressor can be added to the model;
3. Subtract to the model the regressor with the highest p-value, provided that it is higher than a predetermined exit tolerance. If there is no regressor with such a high p-value, end; otherwise, return to step 2.

The final result depends on the initial model and the predetermined tolerances. For this reason, there is no guarantee that the final result is the best possible model. However, checking all the possible model candidates would typically take a large amount of time (for p candidate regressors, there are 2^p possible models), and therefore this procedure is used as a compromise between optimality and feasibility.

In the proposed approach, the entrance and exit tolerance have been set respectively to 0.05 and 0.1.

3.3.5 Neural Network Technique

The architecture adopted in both [56, 57] is the Multi-Layer Perceptron (MLP). Figure 3.7 provides the general structure of a MLP.

Its architecture consists of three parts: an input layer, at least one hidden layer, and an output layer. Each layer receives the inputs from the preceding layer and, by means of weighting and translation, a nonlinear transformation passes them to the next layer. The input layer processes the original input vector, while the output layer passes the processed values to the user.

3.3.5.1 Input Variables

A crucial part of the design and the implementation of a reliable forecasting algorithm based on ANNs is represented by the input selection. The set of inputs chosen for the PV day-ahead forecasting is the following:

- quarter of an hour in the day (number from 1 to 96);
- day of the year (number from 1 to 366);
- ambient temperature [$^{\circ}C$];
- relative humidity [%];

- wind speed [$\frac{m}{s}$];
- cloud cover Index [%].

The format of days and quarters of an hour is chosen in order to take into account temporal autocorrelations of the target variable, as suggested in [208]. Temperature and wind speed are selected because they are involved in the panel efficiency estimation (see Section 3.3.3). Humidity is included because it influences temperature and irradiance [209], and it is exploited with interesting results in several literature works ([210], [178] and [208]). Finally, CCI represents a numerical index for the estimation of the sky covering. Notice that all the meteorological inputs have to be provided by weather predictions.

As a result, in order to forecast the day-ahead PV output, the ANN use as input a vector of 6 numbers, one corresponding to each of the input.

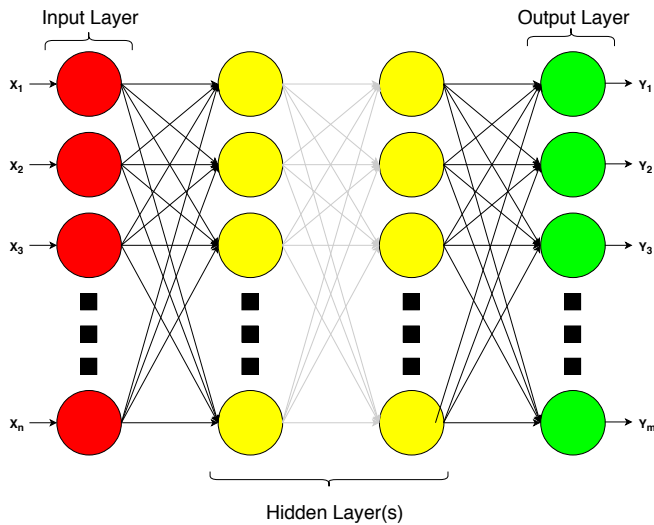


Figure 3.7: A multi-layer perceptron

3.3.5.2 Hyperparameters Selection

After the definition of the inputs for a single ANN, another fundamental step for the implementation of the BEM is represented by the parameters selection of the single MLPs. A generic MLP is characterized by the following parameters:

- Number of hidden layers;
- Neurons number in the hidden layers;
- Transfer functions between each layer. These functions define the relationship of inputs and outputs between each layer;
- Training algorithm. Any ANN has to be trained on a knowledge dataset. This database is composed of an input vector and a score vector. The training algorithm defines the training method.

For the selection of all these parameters, the approaches proposed in [211], [48] and [58] have been used. Table 3.2 reports the results of this phase. These parameters have been utilized in all the ANNs within the BEM.

Table 3.2: Selected hyperparameters for the single ANNs within the BEM

Hidden layers number	1
Transfer function	Hyperbolic Tangent Sigmoid Function $\text{tansig}(x) = \frac{2}{[(1+e^{-2x})]} - 1$
Training algorithm	Resilient Backpropagation

One can combine more MLPs to form an ensemble. The ensemble technique exploited in [56] is described in Section 3.3.5.3.

3.3.5.3 Basic Ensemble Method - BEM

Ensemble averaging methods are usually implemented in order to achieve more accurate results than a single ANN. The basic principle is to combine outputs of several ANNs in order to have a better forecast of the PV generation. Two main aspects can help to achieve a better prediction:

- the combined effect of different ANNs compensates the different random initialization of the weights;

- each MLP employs a slightly different number of hidden units.

There are several typologies of ensemble methods. For the PV forecasting application, a Basic Ensemble Method of Neural Networks (BEM) has been analyzed and implemented [212]. The BEM output is defined by:

$$f_{BEM}(t) = \frac{1}{n} \sum_{i=1}^n f_i(t); \quad (3.11)$$

where n is the total number of ANNs and $f_i(t)$ are the single networks outputs defined as a function of time index t .

In addition to these parameters related to a single ANN, for the BEM methodology is necessary to determine the number of MLPs (i.e. parameter n in (3.11)) within the ensemble. The selection of this parameter has been performed using the strategy presented in [58]:

1. This selection procedure consists of training and testing (on the same training and test set) single multi-layer perceptrons having a different number of neurons in the hidden layer starting from 3 and increasing by 1 for each iteration, until 99.
2. Next, five different perceptrons, with the top five neurons number identified, and with all the parameters reported in Table 3.2 have been considered. These five different networks have been trained on 14 different random initialization of the weights related to the transfer functions, thus 70 nets in total. The neural networks have been ordered with respect to the RMSE error on the common test set. Then, they are combined, starting from the best by adding to the BEM mean the best remaining net. The best combination of the n first networks is chosen as the best ensemble.

The final selected number of ANNs in [56] is $n = 6$: three with 52 neurons, two with 50 neurons and one with 88 neurons.

3.3.6 Hybrid Technique for the PV Forecasting Technique Selection

In this section, the proposed hybrid techniques for selecting between the methods presented in the previous sections are given. Firstly, the selection method with decision trees will be thoroughly described; then, the simplified version with linear regression will follow.

3.3.6.1 Selection with Decision Trees

The accuracy of the Sub Model (SubM) is strongly related to the weather conditions on the prediction window. Thus, the available information related to weather data could be exploited before the forecast execution in order to select the best methodology.

The selection steps can be two or one (see previous Figure 3.5):

1. DeRu1 consists in assessing whether the BEM or the CSM is more convenient;
2. DeRu2 takes place whenever in DeRu1 the BEM technique is not chosen. In that case, the second decision rule consists in the selection between the two other models: the CSM (see Section 3.3.3) or the CCSM (see Section 3.3.4).

The main idea is to compare, on a training set, the performance in terms of nRMSE of the two models under consideration in different climate conditions, evaluated through the forecast mean on the prediction horizon of several weather variables (CC, temperature, humidity, wind speed and pressure). This is performed in order to determine which forecasting techniques have to be selected (CSM, CCSM, or BEM) according to the different combinations of the considered weather variables. In principle, any binary classifier that takes as inputs multiple numerical variables could be used: in this section, the decision tree technique case is the object of focus.

Decision trees are composed of a series of If/Else rules on the regressors that lead to the output of the model. In order to predict a response, the user has to follow the decisions in the tree from the root node down to a leaf node. This last node contains the response. The If/Else rules are also known as splits, while the regressors are often called attributes in the context of this technique.

There are several ways to design and implement a decision tree. In this application, the Classification And Regression Trees (CART) technique has been employed [213].

CART is able to process nominal and continuous attributes both as targets and predictors. Given a training set, the algorithm grows the tree to its full size and then prunes it by eliminating the splits that give a little contribution to the overall performance and that could produce overfitting [213].

The splits are chosen by inspecting all the possible cases on each attribute. Each possible splitting value divides the data that has reached the node into two groups.

CART produces a sequence of nested pruned trees that are the final candidate trees. The final tree has to be chosen by comparison on a separate validation set. Other nice features of the algorithm include automatic handling of unbalanced classes and missing data, and possible adaptation for cost-sensitive learning [214].

Back to the PV forecasting case, since the time horizon is the day-ahead, also the two datasets built for the selection technique are composed of days. For the choice between the two deterministic approaches (DeRu2), only clear or almost clear sky days are included.

The general procedure for the dataset definition and the implementation of the proposed hybrid technique is depicted in Figure 3.8 and can be summarized as follows:

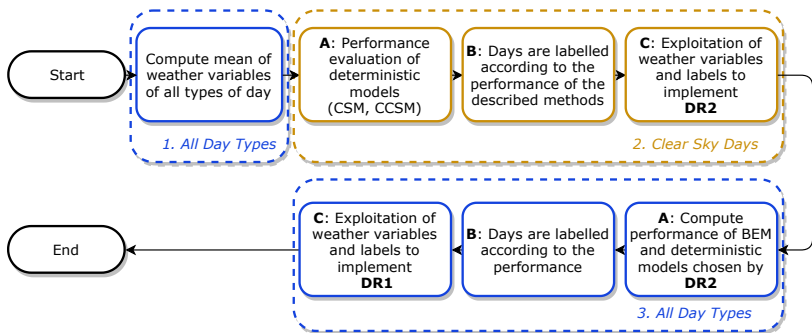


Figure 3.8: General procedure for the implementation of the hybrid selection method

1. the daily mean values of the aforementioned weather variables of all-day types are computed;
2. taking into account only clear sky days:
 - (a) performances in terms of nRMSE of the deterministic models are computed;

- (b) their difference in terms of nRMSE is computed: the sign of the difference tells which model has performed better. In this block, the numerical performance is transformed into a categorical label
 - (c) exploitation of weather variables and labels for the implementation of DeRu2 through a decision tree approach;
3. considering all-day types:
- (a) performances in terms of nRMSE of the BEM and the deterministic model selected by DeRu2 are calculated;
 - (b) the difference between the two performances for each day is computed: the sign of the difference tells whether the ensemble or the deterministic model has performed better. Also, in this case, the numerical performance is transformed into a categorical label;
 - (c) exploitation of weather variables and labels for the implementation of DeRu1 through a decision tree approach.

Table 3.3 collects an example of the results obtained with performance evaluation steps (steps A in Figure 3.8) related to DeRu2 for the definition of a dedicated database, which is used for the definition of the decision tree. Table 3.4 reports the same database but for DeRu1.

Table 3.3: Example of database definition for the decision tree - DeRu2

Day	Avg. Pressure [hPa]	Avg. Temp. [$^{\circ}$ C]	Avg. Hum [%]	Wind Speed [$\frac{m}{s}$]	CCI [%]	Label
8 May	1018	17.44	72.93	0.27	18.83	CCSM
14 May	1016	19.26	35.65	1.86	10.42	CSM
29 May	1012	18.75	70.93	0.45	24.69	CSM

In the end, two decision trees are grown on different training sets, i.e. one for the clear sky choice (DeRu2 in Figure 3.5) and another one for the selection among a deterministic approach or BEM (DeRu1 in Figure 3.5).

Table 3.4: Example of database definition for the decision tree - DeRu1

Day	Avg. Pressure [hPa]	Avg. Temp. [° C]	Avg. Hum. [%]	Wind Speed [m/s]	CCI [%]	Label
9 May	1017	17.91	74.06	0.30	21.54	DR2
10 May	1012	17.90	81.91	0.87	61.41	BEM
11 May	1007	18.61	53.82	1.55	25.63	BEM

In order to understand the process for the selection of the optimal pruning level, it is necessary to introduce two fundamental concepts:

- trivial tree, which is the tree that always labels the observations with the most frequent class. In DeRu1, for example, the trivial tree selects always the same method among the three proposed techniques (CSM, CCSM and BEM). For this reason, it is not a suitable tree: it makes useless the hybrid technique of this section;
- pruning levels. These represent the orders of the nested pruned trees produced by CART. Pruning level 0 is the complete tree, which achieves perfect performance on the training set (and therefore is affected by overfitting problems). The maximum pruning level corresponds to the trivial tree.

For the selection of the best pruning level, the tree corresponding to the smallest pruning level that improves the trivial tree on a validation set is chosen. Moreover, only the first split of the resulting tree is considered.

The decision tree technique applied to the considered test site (see Section 3.5.1) provides the hybrid method reported in Figure 3.9.

Notice that despite different weather variables have been considered during the implementation of the tree, both the decisions are only based on the Cloud Cover index. From Figure 3.9 it can be noticed that in DeRu1 the BEM is chosen when the CCI is particularly high, confirming that the BEM technique is better in cases wherein the sky is far from being clear (cloudy or rainy days), confirming that the intuition behind the utilization of deterministic models is correct.

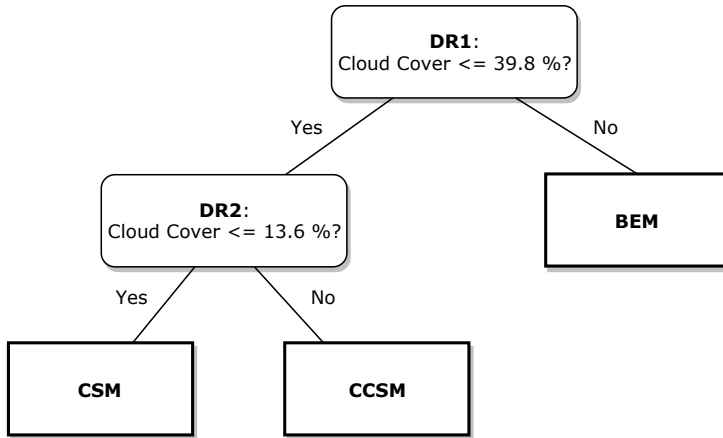


Figure 3.9: The final hybrid technique for the selection of the most appropriate model for the case of the investigated test site (complete procedure, [56])

3.3.6.2 Selection with Linear Regression

In Figure 3.10 the general technique for the implementation of the decision rule with linear regression is depicted.

The algorithm can be summarized in the following steps:

1. In a dataset containing clear and nonclear sky days, the daily mean values of the CCI for each day are computed;
2. The performances in terms of a meaningful KPI of the ANN and the CSM are evaluated;
3. The performance difference is computed: the sign of the difference tells which model has performed better;
4. The decision rule is implemented.

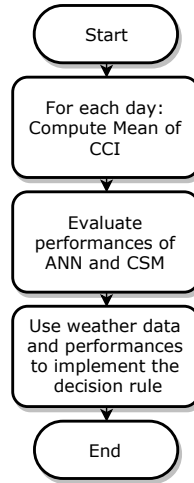


Figure 3.10: General algorithm for implementing the hybrid procedure

In detail, the decision rule is implemented as follows:

1. An ordinary least square regression line is computed with CCI as regressor and difference of performance between CSM and ANN as dependent variable. The difference of performance [p.u] of the day d is evaluated as:

$$perf_d^{diff} = nRMSE_d^{CSM} - nRMSE_d^{ANN} \quad (3.12)$$

where $nRMSE_d^{ANN}$ is the nRMSE of ANN for day d , while $nRMSE_d^{CSM}$ is the nRMSE of the CSM for day d . Therefore a positive value indicates that the ANN has performed better than the CSM, while a negative value means the opposite;

2. The value of CCI such that the regression line crosses the horizontal axis is found: this is the selected threshold for the selection between CSM and ANN approach.

Figure 3.11 plots the linear regression performed (in green) and the corresponding threshold value (dashed black line).

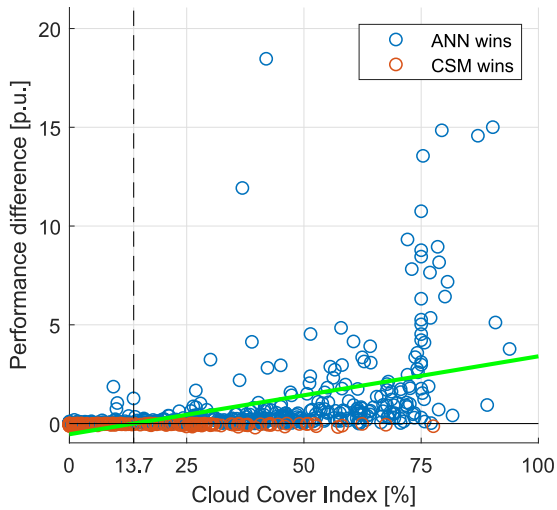


Figure 3.11: Linear Regression for the definition of the decision rule

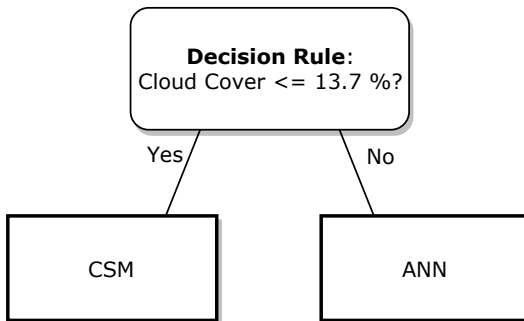


Figure 3.12: Final rule for the selection of the most appropriate method (simplified procedure, [57])

The decision rule exploited in [57] is represented in Figure 3.12.

In words, it can be summarized as follows:

- If the mean daily mean of CCI evaluated considering the 24 hours of the next day is less than 13.7 % use CSM;
- otherwise use the ANN.

With these figures, the presentation of the hybrid PV forecasting procedure is concluded. Next, the test site on which they have been benchmarked is being presented.

3.3.7 Test Site

The test system considered for this validation of the proposed hybrid forecasting procedure is a PV plant located in the harbor of Genova, Italy. In particular, the PV system is positioned on the rooftop of the Economics School of the University of Genova. The building is oriented with respect to the south of about 30° towards west.

The considered PV system presents a peak power of about 20 kW_p , and it is directly connected to the electric system of the underneath building.

The photovoltaic modules are supported by an aluminum structure of $51 \text{ m} \times 3.3 \text{ m}$, which has a 30° inclination (tilt angle) with respect to the horizon. The modules are composed of multi-crystalline silicon, and each of them can produce 180 W. The dimension of each panel is 1.3 m^2 . A total of 108 panels are installed on the structure. The modules are supplied by two inverters, with nominal power equal to 12.5 kW.

Table 3.5 collects all the main parameters of the test site considered in the Clear Sky Model. This table also reports the symbols related to the parameters.

In Figure 3.13 a picture of the PV array is proposed. It can be noticed that the surrounding buildings are near and present about the same height as the array. This implies the presence of shades when the sun is low on the horizon.

3.3.7.1 Shading Modeling

The PV modules are positioned on the roof in order to minimize the losses of irradiance due to the shadows of the higher surrounding buildings. Nevertheless, the considered PV system is slightly shaded by surrounding buildings (see Figure 3.14), especially in the morning and in the late afternoon.



Figure 3.13: Picture of the considered PV system, located on the rooftop of the Economics School of the University of Genoa



Figure 3.14: Top view of the considered PV system (©2018 Google)

The software used for this application for the shadows modeling is PVSyst 6.6.4 [215]. This software allows the user to draw the shape and the dimension of the buildings surrounding the PV plant (Figure 3.15).

The measurements needed for the modeling are obtained through Google Earth Pro. It is possible to have good measurements of both lengths and angles through the ruler functionality.

The shapes of the near buildings are various, and therefore the most appropriate geometrical model has to be carefully chosen for each of them.

Table 3.5: Technical parameters of the test site

Parameter	Value
Peak Power (P_{peak})	20 kW _p
Efficiency of the Inverter (η_{inv})	0.971
Temperature Coefficient (β_c)	-0.045 $\frac{1}{^\circ C}$
Tilt Angle (β)	30°
Latitude	44.4141°
Longitude	8.9221°
Local Time Zone	15°
Orientation w.r.t. S, positive W	30°
Number of Inverters	2
Inverters Nominal power	12.5 kW
Number of Panels	108
Nominal Power of each Panel	180 W
Surface of each Panel	1.3 m ²
Coefficient of Degradation (ω_{DEG})	1

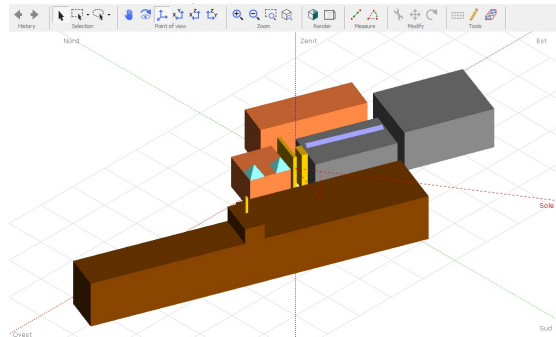


Figure 3.15: Shadows modeling of the surrounding buildings

In addition, also the colors of the building are important to set because they influence the outcome of the ground albedo, which in turn influences the value of $E_{r,pv}$ (see Equation (3.8)).

Once the drawing has been made, the software is able to estimate all the shadow-related parameters involved in (3.7) (S_{dir} , S_{diff} and ρ_g). In particular, the software provides a table that presents values of S_{dir} for different combinations of the sun position (see Table 3.6). Using linear interpolation,

these values can be exploited to compute the parameters for any combination of azimuth and solar altitude.

Table 3.6: S_{dir} approximation as a function of Azimuth (Az.) and Height (Hgt.), in degrees.

Az. Hgt.	-180	-100	-80	-60	-40	-20	0	20	40	60	100	180
90	0	0	0	0	0	0	0	0	0	0	0	0
65	0	0	0	0	0	0	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	0	0	0	0
30	1	0	0.1	0.1	0	0	0	0	0	0	0	1
20	1	0.1	0.2	0.2	0.1	0.1	0	0	0	0	0	1
10	1	1	0.4	0.4	0.4	0.2	0.1	0.4	0.3	0	0.1	1
2	1	1	1	1	0.5	0.2	0.3	0.8	1	1	1	1

Nonzero values are present only if the height of the sun is low. This is reasonable since shading over objects is longer at the beginning and the end of each day.

3.3.8 Available Data

This section describes the available data for the considered test site.

A historical database of the PV plant, described in this section, collects data from 2014 related to power production. For the weather variables, two different sources have been considered:

1. A weather station located just outside the PV plant location. This device can provide measurement data related to the actual temperature, humidity, and wind speed with a granularity of 15 minutes. In addition, the weather station has its own historical database that collects measurements since 2014;
2. A web weather provider, OpenWeatherMap [216]. From this website, it is possible to download a historical bulk dataset that contains all the weather information and, in particular, the crucial data related to the cloud cover index. This weather provider has been employed for the weather forecasting of the variables used by the proposed hybrid procedures. Weather data have been imported from the provider through the dedicated Application Programming Interface (API) in order to be

quickly stored and analyzed. The data obtained from [216] are related to a position that is about one kilometer away from the PV system.

In order to have a reliable dataset, several preprocessing actions have been performed (such as outliers identification and missing data management). The result of the preprocessing stage is a database with measurements related to more than 70,000 quarters of an hour (two years) of complete, reliable data. According to the literature, this represents a robust dataset for the implementation of an accurate forecasting procedure [178]. The preprocessed, historical data have been used in the training phase of all the described methodologies.

3.3.9 Results: Decision Tree Hybrid Procedure

In this section, the decision tree-based hybrid model (see Section 3.3.6.1) is applied on a two months dataset (November - December 2018) for a total of 61 days, located after the period used for training the single techniques and the proposed procedure.

All the proposed methodologies need weather forecast, provided for example by [216]. Thus, for each day of the testing period, a MATLAB routine has been launched at 11:45 pm in order to retrieve the meteorological predictions. The PV forecasting is then executed automatically at midnight by making use of the weather forecast.

Table 3.7 reports the performances of the base methodologies (CSM, CCSM, BEM), as well as the hybrid approach.

Table 3.7: Hybrid and non hybrid methods performance on online test set

	KPIs			days number		
	<i>nRMSE</i>	<i>RMSE</i> [kW]	<i>MBE</i> [kW]	BEM	CCSM	CSM
BEM	0.6090	2.4384	0.8092	61	0	0
CCSM	0.6702	2.6833	-0.7194	0	61	0
CSM	0.8143	3.2601	-1.1996	0	0	61
Hyb. Tree	0.3892	1.5583	0.0694	20	13	28
Ideal Hyb.	0.2761	1.1053	-0.0127	21	13	27

The best stand-alone method is the BEM, which outperforms the two clear sky methods. However, thanks to the decision rule, the BEM can be enhanced by the two clear sky models, whose individual performances are worse than

the BEM. The last line of Table 3.7 reports the results of an ideal hybrid method, which is a hypothetical technique that always chooses the most accurate methodology. Since its nRMSE is lower with respect to the error committed by the proposed approach, it is possible to understand that there is a margin for an improvement of the described decision rule. This can lead to an increment of the accuracy of the proposed hybrid methodology.

The proposed hybrid methodology can also be seen as a three-categories classification problem. Tables 3.8 and 3.9 report respectively the confusion matrices related to the hybrid model and to the BEM (which can be viewed as a hybrid model that always chooses the ensemble of ANNs).

Table 3.8: Confusion matrix for the hybrid technique

		Ideal		
		<i>CSM</i>	<i>CCSM</i>	<i>BEM</i>
Pred.	<i>CSM</i>	18	2	0
	<i>CCSM</i>	2	8	3
	<i>BEM</i>	1	3	24

Table 3.9: Confusion matrix for the BEM method

		Ideal		
		<i>CSM</i>	<i>CCSM</i>	<i>BEM</i>
Pred.	<i>CSM</i>	0	0	0
	<i>CCSM</i>	0	0	0
	<i>BEM</i>	21	13	27

Numbers on the main diagonal identify the days wherein the actual best model is chosen. The BEM is included because it is the best trivial classifier (i.e., a classifier that chooses the most populated category).

As a classifier, the BEM has been right 27 times out of 61 (44% accuracy, see Table 3.9) while the proposed hybrid method has been right twice more: 50 out of 61 times (85% accuracy, Table 3.8).

This analysis is very important to understand whether the proposed procedure actually improves on the standard procedures.

Table 3.10 reports the SSs for the comparison of the proposed hybrid approach with the single SubMs implemented in this application and other solutions proposed in the literature.

In addition, Table 3.10 collects also the different KPIs utilized in (3.5) for the evaluation of the SSs. Notice that for the literature comparison it has been considered the KPI average obtained in all the experimental tests.

Table 3.10: Skill Scores evaluation for a comparison with different day-ahead methodologies

Reference Method	SS	KPI
CSM	+0.5220	$nRMSE$
CCSM	+0.4192	$nRMSE$
BEM	+0.3609	$nRMSE$
[217]	+0.1147	$nRMSE$
[218]	+0.0137	$RMSE/P_{peak}$
[219]	+0.1018	$(Avg.(\varepsilon(t)))/P_{peak}$

In [217] a hierarchical approach based on machine learning methods has been implemented, while in [218] a similar day PV forecasting technique has been adopted. Authors in [219] present an ensemble of five methods (Grey-Box Model, ANN, K-Nearest Neighbours, Quantile Random Forest, and Support Vector Regression), proving that their strategy provides a more accurate forecast with respect to the single approaches.

As can be seen from Table 3.10, all the SSs are positive highlighting the precision of the PV forecasting approach.

This literature comparison, even if on different test sites/sets, suggests that the proposed hybrid technique is a robust and accurate procedure, representing a useful and reliable functionality for uncertainty management.

Focusing on the hybrid model, Figures 3.16 and 3.17 are useful to inspect the improvement of this approach with respect to the versatile BEM approach. As can be seen from these figures, the results of the proposed methodology are very satisfying, especially for the 26th and 27th of November and for the 1st and 4th of December.

As testified by Figures 3.16 and 3.17, the BEM can attain a large error in clear days, where in the deterministic models provide better results. For this reason, the proposed hybrid technique gives an important contribution to the improvement of the overall accuracy of the PV forecast.

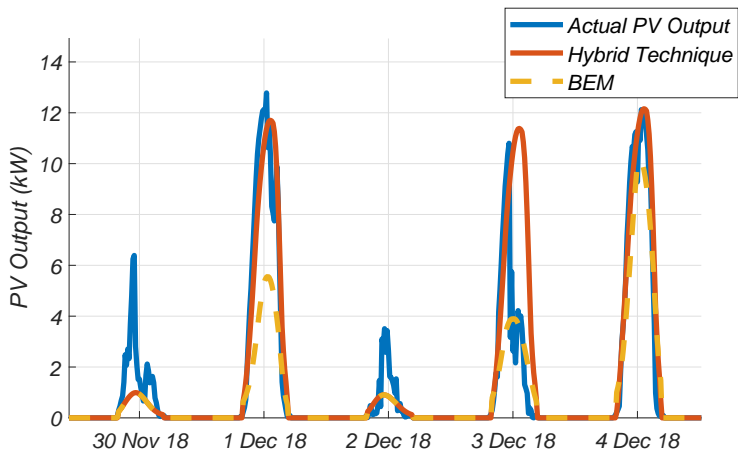


Figure 3.16: PV output, BEM (yellow dotted line) and proposed hybrid technique (red solid line) for five days of December 2018

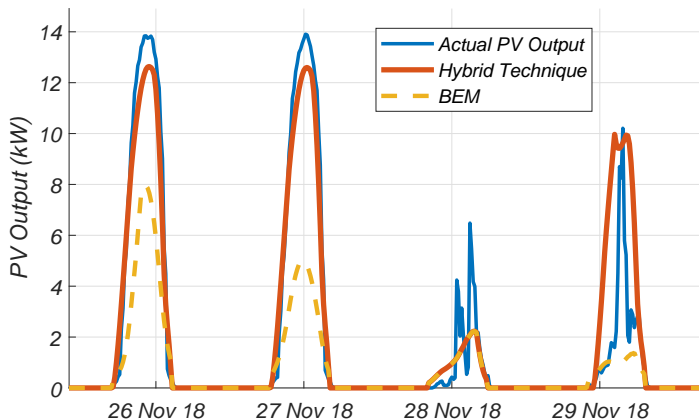


Figure 3.17: PV output, BEM (yellow dotted line) and proposed hybrid technique (red solid line) for four days of November 2018

3.3.10 Results: Linear Regression Based Hybrid Procedure

For hybrid model based on linear regression, in [57], the ANN and the CSM have been tested on 30 days between 31st October 2017 and 4th December 2017. The forecast weather variables have been retrieved from OpenWeatherMap [216], and they are characterized by a granularity equal to three hours. Thence they have been interpolated to the quarter of an hour because all the models have been designed with such granularity.

Table 3.11 shows the performances of all the tested models.

As can be seen in Table 3.11 the proposed hybrid model performs better than the ANN approach and CSM one, considering all the proposed KPIs. The last row of Table 3.11 reports an ideal hybrid method, i.e., an approach that always chooses the most accurate model between ANN and CSM. It demonstrates that the proposed hybrid technique can yield, in principle, very good results.

Table 3.12 shows the amount of days for which ANN (“ANN days”) or CSM (“CSM days”) are chosen as optimal methods by each of the considered approaches. In this context, ANN and CSM methods can be viewed as hybrid models that always choose the same method, regardless of the decision rule outcome.

Table 3.11: Hybrid and non hybrid methods performance

<i>Method</i>	<i>nRMSE</i>	<i>RMSE</i>	<i>MBE</i>
ANN	0.6994	2.6419	-0.8131
CSM	0.9896	3.7378	1.5077
Hybrid - Proposed	0.5041	1.9041	-0.3398
Hybrid - Ideal	0.3015	1.1388	-0.0234

Table 3.12: Summary of choices made by each method

<i>Method</i>	<i>ANN days</i>	<i>CSM days</i>
ANN	30	0
CSM	0	30
Hybrid - Proposed	22	8
Hybrid - Ideal	17	13

In Figures 3.18 and 3.19 predictions for six different days are reported. From these figures, it is possible to notice that the hybrid approach improves the forecasting results by selecting CSM for clear sky days and ANN in the other cases.

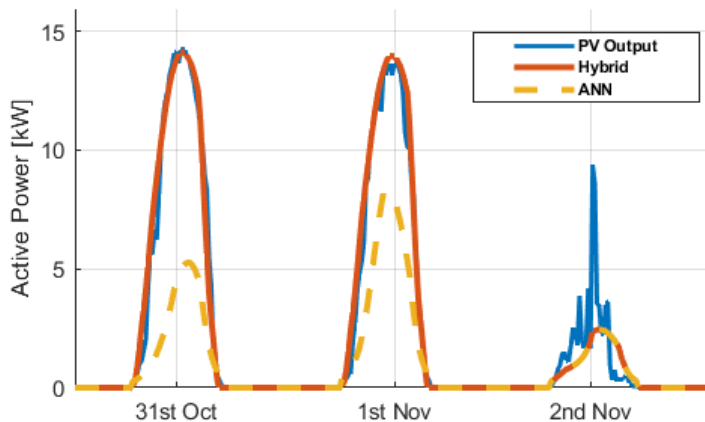


Figure 3.18: Performance of the hybrid algorithm for the PV output of the 31st October, 1st and 2nd November 2017

This hybrid technique can be seen as a classification problem too. Indeed the decision rule classifies the following day into CSM or ANN trying to select the most accurate method. Tables 3.13 and 3.14 present the confusion matrices of the hybrid model and the ANN (here interpreted as a hybrid model that always choose the ANN approach). The main diagonal of these Tables reports the number of days wherein the hybrid technique effectively selects the most accurate approach. As it can be seen from Tables 3.13 and 3.14, the hybrid method is right 25 times out of 30 (Table 3.13) while the ANN method is right 17 times out of 30 (Table 3.14). This testifies that the proposed hybrid methodology has behaved as a better classifier than the classic ANN methodology on the test set.

In the next two sections, applications of load forecasting developed during the Ph.D. are presented. The first regards the application of a modified BEM technique to large scale load forecasting of a distribution network at the medium voltage level in the context of the "Distribution Optimization

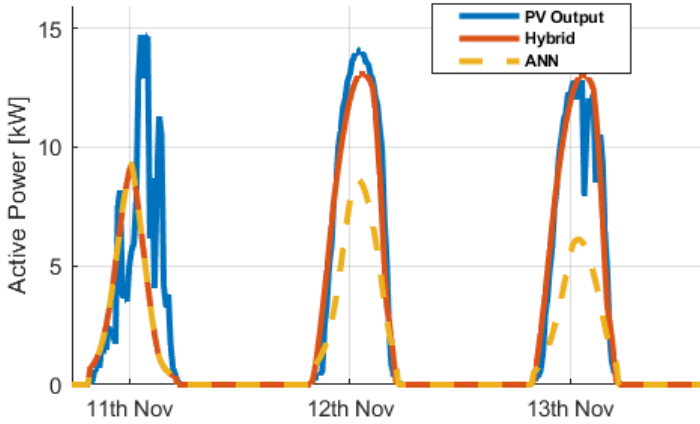


Figure 3.19: Performance of the hybrid algorithm for the PV output of the 11th, 12th and 13th November 2017

Table 3.13: Confusion matrix for the hybrid technique

		Actual	
		CSM	ANN
Pred.	CSM	8	0
	ANN	5	17

Table 3.14: Confusion matrix for ANN method

		Actual	
		CSM	ANN
Pred.	CSM	0	0
	ANN	13	17

Platform Through The Use Of Data From Electronic Meters And Distributed Storage Systems” (PODCAST) project, while the second one, in the context of the ”Adaptive Energy Efficiency Platform For Consumption Reduction In Non-Residential Buildings” (PREDICT) project, is an energy forecasting algorithm for a tertiary building, leveraging deep transfer learning and text mining for the detection and the prediction of the number of building occupants.

3.4 Application: Distribution Network Load Forecasting

A similar BEM technique as the one detailed in Section 3.3.5.3, was employed in the PODCAST project, within the developed DMS, for the load forecasting of the monitored MV/LV substations and MV users of the distribution system under study. Load forecasting represents a fundamental feature for modern DMS, as it provides crucial information to functionalities such as state estimation, optimization procedures, voltage support, optimal reconfiguration, and maintenance scheduling. Examples of papers that address load forecasting embedded in DMSs can be found in [58, 220], while an application of load forecasting applied to advanced DMS functionalities can be found in [221].

3.4.1 PODCAST Grid description

The distribution network of the project is the electric distribution grid of Sanremo, Italy, which covers urban and rural areas. The radial distribution grid is a Medium Voltage (MV) network that features:

- the HV/MV primary substation (132/15 kV/kV);
- 10 MV feeders, radially operated;
- 199 MV/LV substations (SUBs) (15/0.4 kV/kV);
- 20 Medium Voltage Users (MVUs) (2 producers, 3 prosumers and 15 consumers);
- a MV hydropower plant with a nominal power of 200 kW;
- 2 MV PV power plants, with a nominal power of 200 kWp and 20 kWp, respectively;

- 2 storage devices, with a nominal power of 30 kW and 70 kW, and nominal size of 35 kWh for both of them;
- More than 116 km of lines.

The local DSO of the considered test site, Amaie S.p.A., can rely on a DMS (see Figure 2.4 in Chapter 2) for the grid automation and the implementation of real-time functionalities through a dedicated Supervisory Control And Data Analysis (SCADA) system which allows for the data acquisition and management. Data from each measurement point are remotely acquired by the SCADA system, with a granularity of one minute [222].

Of the 199 MV/LV substations, 68 are monitored and served by the load forecasting algorithm, as well as 17 Medium Voltage Users out of 20. Together they form the 85 time series for which the forecasting output is required by the specifics of the DMS. Each SUB or MVU has a different BEM, trained on the respective data.

3.4.2 KPI: Mean Absolute Percentage Error

To read the results, Mean Absolute Percentage Error (MAPE) is used as KPI. It is defined as:

$$MAPE = \frac{1}{N} \sum_i^n \frac{|y_i - \hat{y}_i|}{y_i} \cdot 100\% \quad (3.13)$$

where N represent the number of observation in the dataset, y_i is the i -th measured load and \hat{y}_i the forecast for the i -th value.

3.4.3 The Load Forecasting Algorithm

The only difference of this application with respect to the BEM of the PV forecasting hybrid procedure presented in Section 3.3, is in the variable predicted in the output layer (load value, instead of PV output value) and in the input fed into the neural networks. The inputs are the following:

- quarter of an hour within the day (number 1...96);
- working day or holiday boolean;
- day of the week integer (1 being Sunday,...,7 being Saturday);
- load value with 24 hours lag;

- mean load average of previous 24 hours;
- load value with seven days lag.

An example of the described inputs can be found in Table 3.15

Table 3.15: Example of inputs for proposed forecasting approach

Quarter of hour	Weekday	Avg. day-ahead load	Load -24h	Load -7days	Working Day
1	7	144.20	111.94	110.11	0
2	7	144.22	112.36	105.18	0
⋮	⋮	⋮	⋮	⋮	⋮
96	7	153.24	90.21	96.26	0

It can be noted that the output of Table 3.15 dataset consists of 96 numbers, defining the forecasted load for the next 24 hours, with a granularity of 15 minutes.

Data is acquired through specific instrumentation, and have to be preprocessed, as explained in Section 3.1.3. In particular, missing data have been filled using previously available data (mean of previous minute, day, week), whenever available. Also, anomalous peaks (as identified by a threshold specific for each substation or customer).

Tables 3.16 and 3.17 reports the training, validation and test set limit days chosen for each time series. The validation set has been used for identifying the hyperparameters of the technique (in particular, the number of ensembles and the number of neurons for each neural network). The method for defining the hyperparameters to use for each of the forecasting points is the one explained in Sections 3.3.5.2 and 3.3.5.3.

Table 3.16: Dataset for each medium voltage time series forecasted
(Part 1 of 2)

ID	Training Set		Validation Set		Test Set	
SUB001	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB003	26/03/2019	15/08/2019	15/08/2019	08/09/2019	08/09/2019	18/11/2019
SUB005	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB006	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB007_1	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB007_2	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB009	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB010	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB011_1	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB012	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB014	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB018	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB019	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB020	12/03/2020	27/10/2020	27/10/2020	03/11/2020	04/11/2020	11/11/2020
SUB037	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB045	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB046	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB047	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB047_1	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB047_2	24/09/2020	27/10/2020	27/10/2020	03/11/2020	04/11/2020	11/11/2020
SUB048	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB049_1	29/07/2019	04/10/2019	04/10/2019	15/10/2019	15/10/2019	18/11/2019
SUB049_2	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB061	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB062	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB073	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB074	24/09/2020	27/10/2020	27/10/2020	03/11/2020	04/11/2020	11/11/2020
SUB079	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB086	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB087_1	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB087_2	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB089	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB097	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB101	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB103	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB107	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB111	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB114	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB121	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB124	24/09/2020	27/10/2020	27/10/2020	03/11/2020	04/11/2020	11/11/2020
SUB126	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB131	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB142	03/05/2019	30/08/2019	30/08/2019	19/09/2019	19/09/2019	18/11/2019
SUB155	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB157	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019

Table 3.17: Dataset for each medium voltage time series forecasted (Part 2 of 2)

ID	Training Set		Validation Set		Test Set	
SUB157	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB158	28/07/2020	08/09/2020	08/09/2020	15/09/2020	15/09/2020	22/09/2020
SUB159	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB165	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB168	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB177.1	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB177.2	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB179	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB200	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB203	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB211	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB215.1	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB215.2	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB217	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB218	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB220	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB222	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
SUB224	12/07/2019	27/09/2019	27/09/2019	10/10/2019	10/10/2019	18/11/2019
MVU002	28/05/2019	10/09/2019	10/09/2019	27/09/2019	27/09/2019	19/11/2019
MVU003	28/05/2019	10/09/2019	10/09/2019	27/09/2019	27/09/2019	19/11/2019
MVU047	08/11/2019	14/11/2019	14/11/2019	16/11/2019	16/11/2019	19/11/2019
MVU129	28/05/2019	10/09/2019	10/09/2019	27/09/2019	27/09/2019	19/11/2019
MVU166	28/05/2019	10/09/2019	10/09/2019	27/09/2019	27/09/2019	19/11/2019
MVU169	28/05/2019	10/09/2019	10/09/2019	27/09/2019	27/09/2019	19/11/2019
MVU171	28/05/2019	10/09/2019	10/09/2019	27/09/2019	27/09/2019	19/11/2019
MVU178	28/05/2019	10/09/2019	10/09/2019	27/09/2019	27/09/2019	19/11/2019
MVU180	28/05/2019	10/09/2019	10/09/2019	27/09/2019	27/09/2019	19/11/2019
MVU183	28/05/2019	10/09/2019	10/09/2019	28/09/2019	28/09/2019	19/11/2019
MVU204	28/05/2019	10/09/2019	10/09/2019	27/09/2019	27/09/2019	19/11/2019
MVU207	28/05/2019	10/09/2019	10/09/2019	27/09/2019	27/09/2019	19/11/2019
MVU212	28/05/2019	10/09/2019	10/09/2019	27/09/2019	27/09/2019	19/11/2019
MVU228	28/05/2019	10/09/2019	10/09/2019	27/09/2019	27/09/2019	19/11/2019
MVU229	28/05/2019	10/09/2019	10/09/2019	27/09/2019	27/09/2019	19/11/2019
MVU231	28/05/2019	10/09/2019	10/09/2019	27/09/2019	27/09/2019	19/11/2019
MVU234	28/05/2019	10/09/2019	10/09/2019	27/09/2019	27/09/2019	19/11/2019
SUB110	03/12/2019	31/01/2020	31/01/2020	10/02/2020	10/02/2020	10/03/2020
SUB137	18/11/2019	25/01/2020	25/01/2020	05/02/2020	05/02/2020	10/03/2020
SUB148	18/11/2019	25/01/2020	25/01/2020	05/02/2020	05/02/2020	10/03/2020
SUB150	18/11/2019	25/01/2020	25/01/2020	05/02/2020	05/02/2020	10/03/2020
SUB151	18/11/2019	25/01/2020	25/01/2020	05/02/2020	05/02/2020	10/03/2020
SUB159	08/10/2019	08/01/2020	08/01/2020	24/01/2020	24/01/2020	10/03/2020
SUB235	08/10/2019	08/01/2020	08/01/2020	24/01/2020	24/01/2020	10/03/2020

The date limits that can be seen in Table 3.16 were chosen on the basis of the availability of data and on the criterion of having at least one month of test data as well as the maximum possible training set for each time series.

In order to make sure that the BEM approach is beneficial to the system, it has been compared, on the same dataset, with a single ANN approach, that is, a BEM with only one ANN. This is considered to be a meaningful baseline because the single ANN technique is the one used currently by the experimental DMS, as it is a technique derived from a previous project SmartGen [223], in which a single ANN were applied for forecasting on a subset of the time series studied for the PODCAST project. The comparison can be appreciated in Table 3.18

From Table 3.18 it can be observed that the BEM technique implemented in the PODCAST project is better than the current one in all cases, apart from SUB151 and MVU228 (highlighted in red). Moreover, note that the medium voltage customers with a high MAPE will need further analyses since they are characterized by few data or communication issues (more than 30%).

Finally, in Figures 3.20–3.22 are reported the forecasts of selected MVUs.

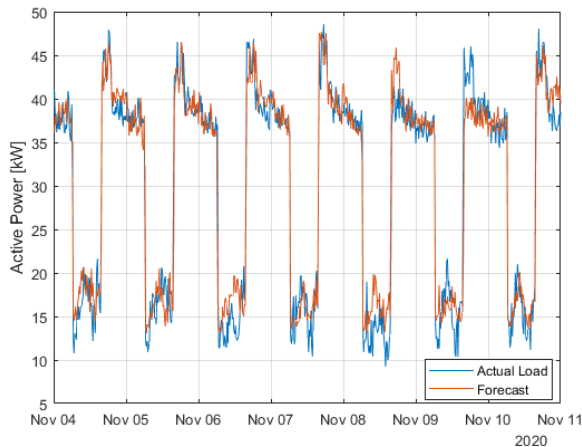


Figure 3.20: Forecast (in red) of BEM technique on the test set of SUB020 (9.4% MAPE)

Table 3.18: MAPE Results of BEM on load forecasting test sets (Tables 3.16 and 3.17). **Red**: cases where single ANN was better than BEM

ID	Single ANN	BEM	ID	Single ANN	BEM
SUB001	9.1%	8.9%	SUB142	10.5%	9.4%
SUB003	6.2%	6.1%	SUB148	10.4%	10.2%
SUB005	15.4%	10.5%	SUB150	7.3%	7.1%
SUB006	7.4%	7.0%	SUB151	12.3%	14.1%
SUB007_1	14.5%	10.3%	SUB155	11.1%	8.0%
SUB007_2	7.2%	5.9%	SUB157	14.9%	11.6%
SUB009	9.2%	8.2%	SUB158	23.3%	18.0%
SUB010	15.3%	13.2%	SUB159	13.3%	13.1%
SUB011_1	13.1%	9.5%	SUB165	18.0%	15.6%
SUB012	16.7%	16.0%	SUB168	15.2%	13.5%
SUB014	12.1%	9.8%	SUB177_1	9.0%	8.2%
SUB018	8.1%	7.2%	SUB177_2	18.1%	11.0%
SUB019	11.3%	9.8%	SUB179	51.2%	42.6%
SUB020	10.5%	9.4%	SUB200	8.4%	8.1%
SUB037	9.3%	7.9%	SUB203	7.1%	6.7%
SUB045	8.5%	5.9%	SUB211	13.1%	10.5%
SUB046	8.7%	7.7%	SUB215_1	12.1%	10.8%
SUB047	13.2%	8.5%	SUB215_2	8.9%	7.4%
SUB047_1	14.4%	8.3%	SUB217	4.3%	2.5%
SUB047_2	9.4%	7.3%	SUB218	14.2%	9.3%
SUB048	24.5%	13.2%	SUB220	66.6%	64.2%
SUB049_1	10.7%	9.7%	SUB222	10.3%	7.2%
SUB049_2	6.5%	5.8%	SUB224	8.2%	7.8%
SUB061	31.6%	22.2%	SUB235	7.0%	4.3%
SUB062	8.8%	8.1%	MVU002	49.7%	44.7%
SUB073	17.3%	16.1%	MVU003	48.6%	39.4%
SUB074	30.3%	27.2%	MVU047	24.5%	23.6%
SUB079	9.0%	7.9%	MVU129	24.3%	19.5%
SUB086	18.9%	9.1%	MVU166	33.4%	29.5%
SUB087_1	9.3%	7.0%	MVU169	32.3%	30.7%
SUB087_2	9.9%	8.8%	MVU171	6.0%	5.9%
SUB089	5.9%	5.6%	MVU178	13.8%	9.0%
SUB097	62.3%	57.3%	MVU180	6.5%	6.3%
SUB101	8.6%	6.1%	MVU183	7.7%	7.6%
SUB103	12.1%	10.3%	MVU204	10.5%	9.1%
SUB107	6.8%	6.2%	MVU207	8.5%	7.8%
SUB110	11.6%	9.4%	MVU212	12.0%	10.5%
SUB111	52.3%	49.3%	MVU228	43.3%	48.6%
SUB114	8.9%	8.1%	MVU229	25.9%	19.1%
SUB121	9.4%	8.9%	MVU231	53.3%	50.1%
SUB124	10.3%	8.0%	MVU234	15.6%	14.7%
SUB126	11.2%	11.1%			
SUB131	8.1%	7.5%			
SUB137	9.4%	9.0%			

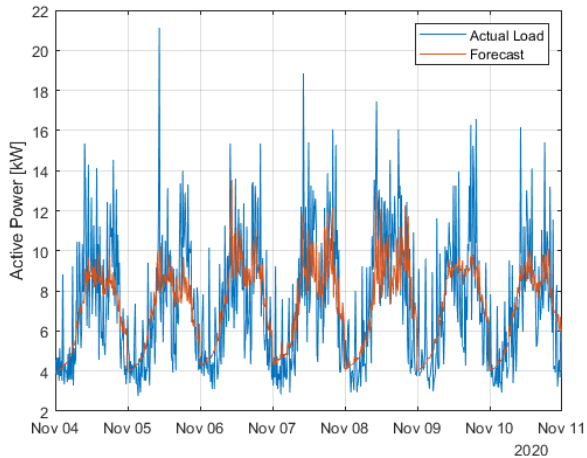


Figure 3.21: Forecast (in red) of BEM technique on the test set of SUB074 (27.2% MAPE)

As testified by Figures 3.20–3.22, the load forecasts can capture the daily seasonality of a wide range of MV/LV substations and MV customers. Also the peaks are forecasted with good precision, although it is a more challenging task in points of the network with a lower level of load (see Figure 3.21). The forecasts seen in Figures 3.20–3.22 can be accessed in the DMS via the procedure described in Section 2.2 (see Figure 2.4).

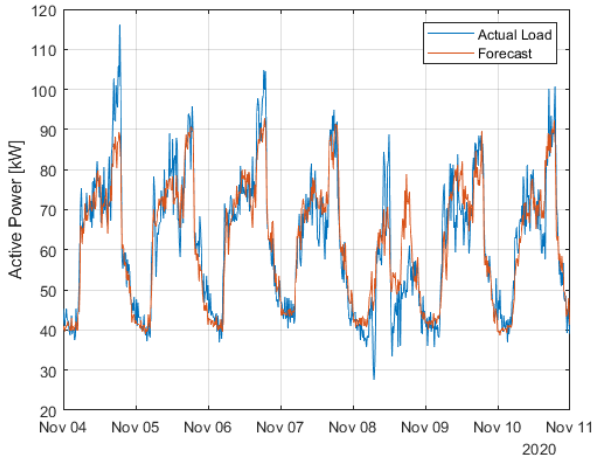


Figure 3.22: Forecast (in red) of BEM technique on the test set of SUB124 (8.0% MAPE)

3.5 Application: Non-Commercial Building Energy Prediction

According to the International Energy Agency (IEA), the building sector is responsible for 36% of global energy consumption and nearly 40% of total direct and indirect CO₂ emissions. Fossil fuels generate 84% of heating and cooling energy, while renewables produce only the remaining 16% [39]. In the European Union (EU), almost 75% of the buildings are energy inefficient, while only 0.4-1.2% (depending on the country) of them are renovated each year [224]. Hence there is space for significant improvements in terms of efficiency (up to 40% in 2040). Space heating alone offers over a quarter of the potential energy savings [39].

Structural improvements (e.g., outer envelope or window renewal) can achieve substantial savings, but they require high capital investments, which may not be affordable or may require a long payback period.

In this context, prescriptive analytics techniques based on Model Predictive Control (MPC) have been deeply analyzed and implemented by the scientific community in the last years. Several papers show promising results for the

HVAC systems with consumption reductions from 7% up to 50% [225]. Besides, MPC approaches can consider thermal comfort constraints.

In an MPC technique, a forecast model is employed to simulate the system evolution. This prediction is used to determine a control sequence for the system, wherein only the first element is implemented. At the next time step, new measurements are collected, and the process is repeated. Thus, a reliable system predictive model is a fundamental element for the performance of the MPC strategy.

The overall objective of the application is to have an easily replicable procedure for the identification of a building energy model. Obtaining a suitable building model is indeed the most time-consuming part of a MPC for the HVAC optimal control. This block of the process is preventing MPC from being widely exploited [226]. Thus, the technique presented in [227] has been followed in order to identify a simplified building energy model, able to consider occupancy as an input. This strategy is based on the utilization of operational information, obtained in the presence of an existing MPC, which in turn has been defined from historical data.

Among all the possible inputs of a simplified building model, occupancy is one of the most important. Indeed, it can be crucial in order to define thermal comfort constraints (e.g., when the building is empty for a considerable period, the HVAC can be switched off)[228].

The simplified model is based on state-space equations that have been identified through a black box strategy.

In this application, deep Transfer Learning (TL) to thermal camera images for the estimation of the building occupancy has been employed. TL leverages the predictive power of deep learning without the exploitation of costly hardware for the training of specific neural networks. An example of deep learning for occupancy estimation can be found in [229]. TL techniques avoid the training phase of the network and are easily applied to image data (e.g., see [230]). Prediction is performed through a historical database, derived from the TL technique and public textual data retrieved from the University website (which owns the building), together with a procedure based on a k Nearest-Neighbors (kNN) approach for the selection of the most similar historical occupancy.

First, a description of the test site will be given, then the algorithm for occupancy estimation and the algorithm for energy consumption prediction will be presented, and the results for a testing period will be given for last.

3.5.1 Test Site

The test site is a three-story building of the Department of Formation Sciences of the University of Genova (the same described in Chapter 2), located at $44^{\circ}25' \text{ N}$, $8^{\circ}53' \text{ E}$.

The construction was initially an office building. It was bought and refurbished in the early 2000s by the University of Genova.

Figure 3.23 shows a top view of the building, while Figure 3.24 reports a picture taken from the ground.



Figure 3.23: Top view of the test site, with AHU and Chiller (©2018 Google)

The building is used for educational activities, mainly lessons, seminars, and exams. The test site presents three rooms on the base floor and two rooms on each of the other two floors, and one restroom per floor.

The electric HVAC system is composed of an AHU, located at the top of the building, and a Chiller Unit, positioned in the surrounding area (see Figure 3.23).

Ducts going from North to South distribute conditioned air throughout the building, serving all three floors.



Figure 3.24: Picture of the test site

3.5.2 Monitoring and Control System

A monitoring system has been installed in the building as part of a massive energy monitoring program of the University of Genova. It is composed of:

- A pulse counter that measures global energy consumption;
- One energy meter (DIRIS I30) for the chiller;

while the following instruments have been added during the PREDICT project [231, 232]:

- Multi-configurable Regulator (Coster, model YLC 740) for the HVAC system, which allows the remote control of the temperature setpoints related to the three floors of the considered building;
- Two energy meters, one for the AHU and one for the general switch (DIRIS I30);

- One voltage meter (DIRIS U30);
- One thermal camera (FLIR AX8), installed in one room of the test site. An example of the captured images is reported in Figure 3.25;
- Four temperature sensors (three internal, one external).



Figure 3.25: Example of image downloaded from the thermal camera

Figure 3.26 represents the overall monitoring and control architecture.

The cloud server archives temperature, energy, and occupancy data. The utilization of a remote server makes the proposed system easily scalable. In addition, it hosts the MPC algorithm that has controlled the building during the "Adaptive Energy Efficiency Platform For Consumption Reduction In Non-Residential Buildings" (PREDICT) project (December 2018-February 2019), whose data make up the base of the algorithms that constitute this application. The data relating to thermal images are managed by the TCP protocol. Also, temperature and energy data are received through the SFTP protocol, while the temperature setpoints are sent to the controller via Modbus/TCP. Finally, the central server shares consumption data via web service with the monitoring system of all the buildings belonging to the University of Genova.

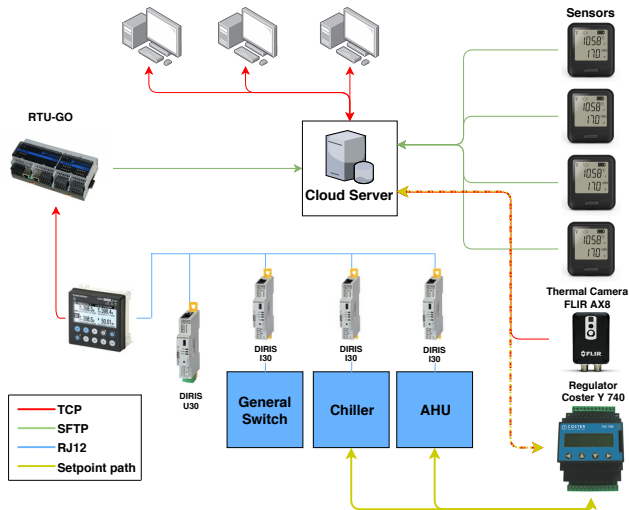


Figure 3.26: Scheme of the monitoring system

3.5.3 Building Occupancy

The images captured by the thermal camera can be used to estimate the occupancy of the monitored room. This is performed through TL (see Section 3.5.4). This estimation can, in turn, be exploited for the prediction of the occupancy related to the other rooms or future time instants (by kNN, Section 3.5.5). The prediction is then fed into the energy model as an input. The proposed approach can be employed even with few historical data.

3.5.4 Occupancy Estimation

The main idea of this algorithm is to leverage the predictive power of deep neural networks without resorting to costly hardware or cloud computing resources. For this reason, a pre-trained deep neural network (i.e., characterized by fixed weights) has been used to process each image into a vector of features, which are then fed into a supervised machine learning method. Indeed, deep networks learn general features in their first layers and problem-specific features in the last layers. The basic principle of TL is to exploit the first layers of pre-trained networks for the extraction of general features.

In Figure 3.27 the scheme of the proposed algorithm is reported. The chosen deep neural network is AlexNet, originally trained and employed for a considerable image classification task. The network is available as an open-source architecture (for more information about its composition, see [233]). The neurons activations of one of the last layers are the inputs of a Support Vector Regression (SVR) [234] methodology.

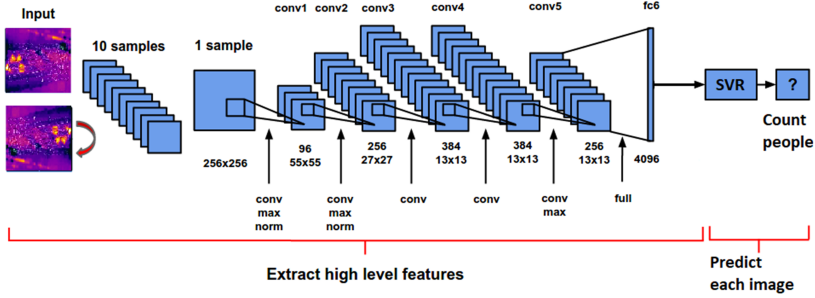


Figure 3.27: Scheme of the proposed algorithm [235]

The linear SVR, utilized in this application, consists in finding a function f of the form:

$$f(x) = x^T \beta + b \quad (3.14)$$

solving the following constraint optimization problem:

$$\begin{aligned} \min_{\beta} \quad & \frac{1}{2} \beta^T \beta + C \sum_{n=1}^N (\xi_n + \xi_n^*) \\ \text{subject to} \quad & \forall n : y_n - (x_n^T \beta + b) \leq \varepsilon + \xi_n \\ & \forall n : (x_n^T \beta + b) - y_n \leq \varepsilon + \xi_n^* \\ & \forall n : \xi_n \geq 0 \\ & \forall n : \xi_n^* \geq 0 \end{aligned} \quad (3.15)$$

where N is the number of samples in the training set; n is the generic sample; ε and C are hyperparameters. Finally, ξ_n and ξ_n^* are the slack variables of the optimization problem:

$$\xi_n = \begin{cases} 0 & |y - f(x_n)| \leq \varepsilon \\ |y - f(x_n)| - \varepsilon & \textit{otherwise} \end{cases} \quad (3.16)$$

$$\xi_n^* = \begin{cases} 0 & |y - f(x_n)| \geq \varepsilon \\ \varepsilon - |y - f(x_n)| & \textit{otherwise} \end{cases} \quad (3.17)$$

The layers number of AlexNet exploited in this application has been identified according to the following procedure. A training set of images manually labeled with the visible number of people is used for the computation of the features related to the different truncating layers. In this application, the last three hidden layers have been considered for the truncation process related to AlexNet. These features are then used to train the linear SVR. The performances of these three TL configurations are successively tested on a validation set. The configuration attaining the best performance is selected and again tested on another set for the estimation of the generalization error. Notice that images of the same day are employed only in one between training, validation, and test set in order to avoid the contemporary presence of similar pictures in the same dataset. After this process, the final hidden layer used within AlexNet is *fc6* (see Figure 3.27), which is the most internal among the considered candidates.

The results of the test are reported in Section 3.5.10.

3.5.5 Occupancy Prediction

Section 3.5.4 deals with the estimation of the visible number of people in an image. This algorithm is crucial for forming a historical database related to occupancy. However, it does not deal with cases wherein images are not available. Indeed, the test site (see Section 3.5.1) presents just one monitored room (Room 3), which has 100 seats framed over a total of 141 seats. Moreover, the other rooms do not have the same size as the one with the thermal camera. Finally, the application of the proposed building model to an MPC algorithm requires an occupancy forecasting technique.

In this application, the prediction of non-monitored rooms occupancy and the forecasting of the future people number are faced with the same approach. The idea is to use the lessons and exams schedule information, publicly available on the university website, to build a historical database of events associated with the occupancy of the monitored room.

Then, the occupancy related to a non-monitored room is approximated by the most similar monitored room image. Similarity has to be intended in terms of the time and scheduled events and computed by a nearest-neighbor algorithm. The retrieved occupancy is also scaled through correction factors, to consider the different dimensions of the test site rooms. These factors are represented by the ratio between the number of seats in the room and the number of visible seats in the monitored room (see Table 3.19). The same process is performed for the estimation of future occupancy.

Table 3.19: Correction Factors for each of room of the considered test site

Room ID	Floor n.	Seats	Correction Factor
1	3	99	99/100
2	3	180	180/100
3	2	141	141/100
4	2	150	150/100
5	1	108	108/100
6	1	99	99/100
7	1	46	46/100

Figure 3.28 reports the flowchart of the proposed prediction algorithm.

The historical database of processed images is composed of the following variables:

- Time variables: year; month; day of the month; day of the week; day of the year; hour; minute; the quarter of an hour within the day; information about holidays/working days and exam periods.
- Variables related to room activities: this kind of data can be retrieved on the university web portal, which contains a schedule for each room of the considered building. The information used by the proposed algorithm is: name of the activity, type of the activity (lesson, exam, and seminar), and its responsible (professor name). The information is represented by standard text mining techniques (one-hot encoding, bag of words).

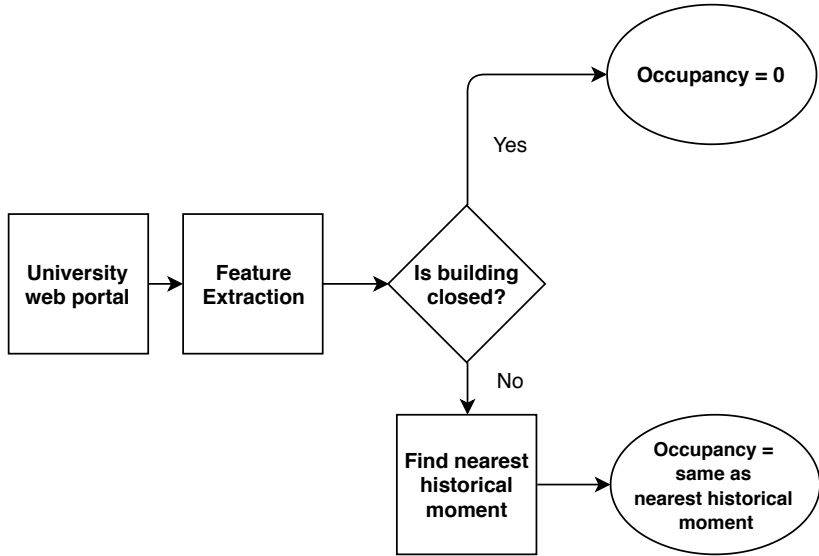


Figure 3.28: Scheme of the proposed algorithm

All the time variables are processed by the following formulas in order to take into account the seasonal behavior:

$$\sin\left(\frac{2\pi}{P_v} \cdot T_v\right) \quad (3.18)$$

$$\cos\left(\frac{2\pi}{P_v} \cdot T_v\right) \quad (3.19)$$

where T_v is the v -th time variable value and P_v is the period related to that variable (e.g., 7 for the day of the week).

The proposed algorithm has been used to estimate the occupancy of each floor of the test site. The results have been then summed floor by floor. Finally, the resulting time series is fed into the building energy model, described in the following section.

3.5.6 Building Energy Model

State-space models are a common choice for building energy modeling. The following equations define a state-space model:

$$\begin{cases} x_{k+1} = Ax_k + Bu_k + Ke_k \\ y_k = Cx_k \\ x_0 = x_{init} \end{cases} \quad (3.20)$$

where

- $k = 0, 1, \dots$ is a discrete time index;
- u is the input vector;
- x is the state vector;
- y is the output vector;
- e is the error vector;
- x_{init} is an initial value computed by the estimation algorithm;
- A, B, C, K are numerical matrices of appropriate dimensions.

The number of states for this kind of model is crucial. Here, it has been set to 2, according to [236].

There are several ways to estimate the matrices of a state-space model. They can be derived from the first principle relation between inputs and outputs (white-box approach), directly from data (black-box strategy), or through physics considerations and exploiting data/measurements (gray-box modeling). In this case, a black-box approach is utilized. In particular, a Prediction Error Minimization (PEM) procedure based on a subspace method (N4SID) is exploited [237].

3.5.7 Inputs

A meaningful input selection for a state-space model is essential for a satisfying performance. Different types of variables have been considered:

- weather inputs: external temperature, wind speed, humidity, and direct normal irradiation;

- HVAC setpoints: controlled temperature for each floor;
- occupancy: computed through the algorithms described in Section 3.5.3;
- information about the time of the day (in particular night or holiday).

The final inputs for the proposed state-space method are summarized in Table 3.20. In this table, column “Forecast Source” indicates how each variable is predicted or which service is employed for this task, while column “Actual Source” represents how the inputs are retrieved for the training of the algorithms.

Table 3.20: Input variables for the proposed building energy model with their corresponding sources

Variable	Actual Source	Forecast Source
External Temperature [$^{\circ}\text{C}$]	Sensors	[216]
Direct Normal Irradiance [$\frac{\text{W}}{\text{m}^2}$]	[238]	[238]
Setpoints Floor 1-2-3 [$^{\circ}\text{C}$]	Sensors	None (Controlled)
Humidity [%]	[216]	[216]
Occupancy Floor 1-2-3 [people]	AlexNet+SVR	kNN
Wind Speed [$\frac{\text{m}}{\text{s}}$]	[216]	[216]
Night/Sunday Mode [0 or 1]	Calendar	Calendar

3.5.8 Output

The output of the state-space model is represented by the HVAC energy consumption [kWh]. The accuracy of the proposed method can be verified with the energy sensors described in Section 3.5.1. The building energy model can be used within an MPC for the optimal control of the temperature setpoints of the HVAC system. Thanks to this model, it is possible to minimize the HVAC energy consumption considering thermal comfort constraints (e.g., limits on temperature setpoints).

3.5.9 Key Performance Indicators

The performances of the proposed methodologies (occupancy estimation and building energy model) are evaluated through the following Key Performance Indicators (KPIs):

- Root Mean Square Error (RMSE), defined as in Section 3.3.1;
- Normalized Root Mean Square Error (NRMSE), defined as in Section 3.3.1;
- Mean Absolute Error (MAE):

$$MAE := \frac{1}{N} \sum_{t=1}^N (|\varepsilon(t)|) \quad (3.21)$$

- Mean Bias Error (MBE), defined as in Section 3.3.1;
- Error Standard Deviation (ESD), defined as:

$$ESD := \sqrt{\frac{1}{N} \sum_{t=1}^N (\varepsilon(t) - MBE)^2} \quad (3.22)$$

3.5.10 Experimental Validation

This section presents the validation results of the proposed methodologies. Notice that the sampling time of all the exploited databases is equal to 15 minutes.

3.5.11 Occupancy Estimation

The TL method described in Section 3.5.3 has been trained on a database of 505 images (13 December 2018-8 January 2019), validated on 96 pictures (11 December 2018), and tested on 96 images (12 December 2018). A common feature extraction strategy Histogram of Oriented Gradients (HOG) has been trained for a comparison with the proposed approach (AlexNet+SVR). The validation and test set have been chosen in order to include images with both high and low occupancy.

Figure 3.29 presents the raw output of the proposed approach (red line), the result of the machine learning reference methodology (yellow curve), and the actual occupancy (blue line). Figure 3.30 proposes the same plot with the difference that the estimation of the night occupancy is set to zero. The presented approach follows more precisely the actual occupancy. The numerical KPIs, reported in Table 3.21, testify this conclusion.

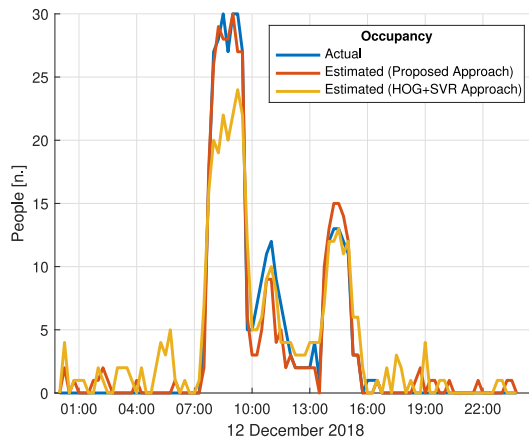


Figure 3.29: Performance of the occupancy estimation algorithm (case without smoothing)

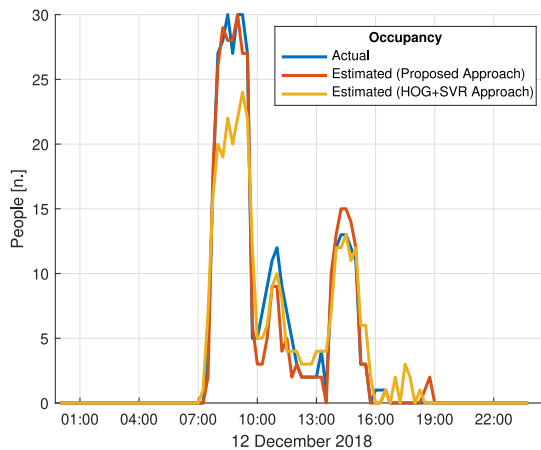


Figure 3.30: Performance of the occupancy estimation algorithm (case with night smoothing)

Table 3.21: KPIs for the occupancy estimation algorithms

	<i>Smoothed</i>	MBE	ESD	MAE	RMSE	NRMSE
AlexNet+SVR	No	0.06	1.3	0.7	1.3	14%
HOG+SVR	No	-0.09	2.6	1.6	2.6	29%
AlexNet+SVR	Yes	0.23	1.2	0.7	1.2	13%
HOG+SVR	Yes	0.35	2.3	1.1	2.3	27%

3.5.12 Building Energy Model

The proposed building energy model has been tested on a period of one week (2 February 2019-8 February 2019) using actual data as inputs. Figure 3.31 depicts the actual HVAC consumption of the considered building (blue line) and the output of the presented state-space model (red curve). As can be seen from this figure, the model output follows the HVAC behavior with satisfying accuracy, although the peaks are underestimated, probably for their high frequency. This result is also testified by all the KPIs that are reported in Table 3.22.

Since the developed model will be used as the forecasting module of an MPC algorithm, the correctness of the relation between the controlled variables and the output is crucial. Hence, a sensitivity analysis is performed. In Figure 3.32, the predicted consumption of the considered HVAC system is plotted for various values of temperature setpoints, which have been set identical for all the floors.

Table 3.22: KPIs for the proposed building energy model

KPI	Proposed State-Space Model
MBE	0.07 kWh
ESD	2.04 kWh
MAE	1.60 kWh
RMSE	2.04 kWh
NRMSE	26.0 %

As illustrated in Figure 3.32, for lower setpoints, lower energy consumption is predicted, which is the correct behavior for an HVAC operating in winter mode (notice that the test is performed in February). Thus, the proposed model can be exploited within an optimal HVAC control procedure.

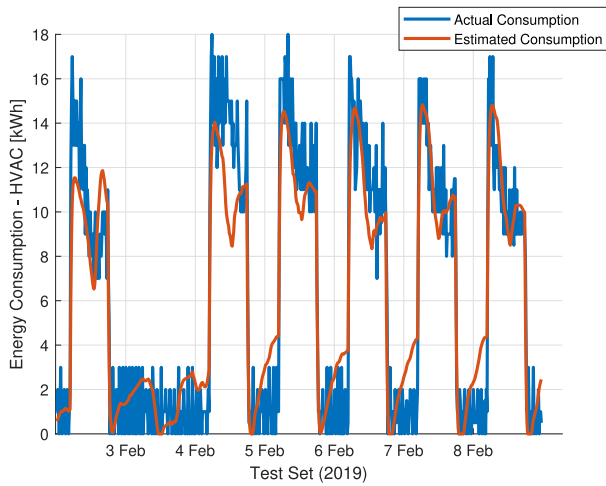


Figure 3.31: Results of the building energy model for the considered test set

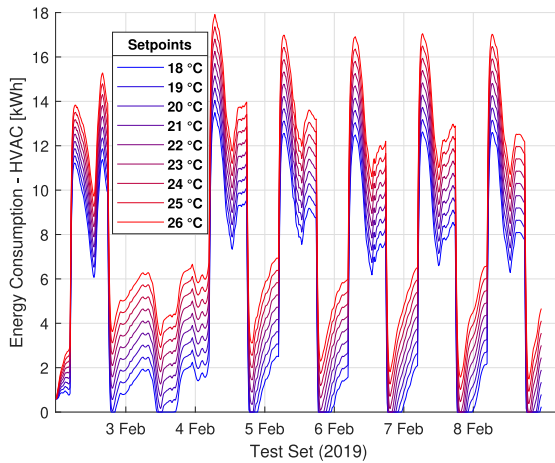


Figure 3.32: Output for different values of temperature setpoints

3.6 Chapter Conclusions

In this chapter, predictive analytics was introduced, employing a common cross-industry standard for explaining the various aspect to be analyzed in order to develop a predictive analytics solution. Forecasting was given a particular focus since predictions about the future are crucial in power systems, and a distinction was made between forecasting and other passive types of analytics that focus on predicting unknown (but not future) quantities. The chosen standard (the widely used CRISP-DM) was presented step by step, highlighting the relevant aspects and choices that an analyst has to make in order to develop a predictive analytics solution.

Next, examples of predictive analytics applied to power systems found in the scientific literature have been illustrated. A brief literature review was load forecasting and PV forecasting.

Finally, some applications of predictive analytics developed during the Ph.D. were exposed.

Firstly, a novel day-ahead PV forecasting procedure, based on a decision rule that, based on the weather forecast (and in particular on CCI), decides on whether to use a physical model or a neural network (in the simplified case [57]) or between a physical model, a hybrid model or an ensemble (Basic Ensemble Method of Neural Networks (BEM)) of neural networks (in the complete case [56]). Decision rules are implemented with linear regression in the simplified case and with two CART trees in the other case. The results observed for the test site (a PV located in Genoa, Italy) are promising since the hybrid approach improved the forecasting accuracy of both stand-alone physical models and ANN techniques on an extended test set. The proposed methodology presents all the good properties of an ensemble method, coupled with enhanced performance on clear sky days. A comparison in terms of accuracies with the literature seemed to confirm the goodness of the proposed procedure. Moreover, the final decision tree provides good results for the selection of the most appropriate method, and it reflects the intuition behind the utilization of deterministic models or an ANN-based approach.

The presented approach can be employed as an input of optimization or advanced algorithms within generation scheduling/unit commitment applications, energy management strategies, or grid regulation procedures.

Several parts of the proposed PV forecasting technique could be investigated in order to improve the forecasting accuracy:

1. the CCSM procedure could include more inputs (currently only CCI is considered as weather regressor for CCSM) in order to be more competitive with respect to the other methodologies;
2. the algorithm has been validated for the winter season but needs further tests for the other seasons;
3. more sophisticated ensemble techniques can be considered;
4. different architecture of ANNs could be analyzed;
5. the hybrid technique could be designed to select different methods within the same day;
6. a three-category classifier could substitute the two decisions that compose the hybrid approach;
7. Model Output Statistic (MOS) could be performed on weather forecasts in order to understand if they are affected by any bias [239];
8. Irradiance data could be used instead of CCI by the proposed hybrid approach;
9. The CSM could be used to derive other features for the ANN part of the forecasting procedure, for example, by computing a clear-sky index [240];
10. The procedure could be compared with other baselines, such as the ones derived from NWP.

Secondly, an unpublished application coming from the PODCAST project was presented, where 85 MV/LV substations and 17 MV users loads of the distribution network of Sanremo, Italy, were forecasted via BEM of neural networks. The results show that the BEM was able to enhance the currently used technique (a single neural network), developed in the course of a previous project (SmartGen, [223]) on the same distribution network, on an extensive dataset. This, together with the other DMS functionalities developed during the project, demonstrate that also DSOs with modest dimensions can be equipped with advanced algorithms for managing the increasing RES generation and the fluctuations in demand that are expected in the following

years. In that regard, the forecasting technique does not have many requirements other than historical data from smart meters at the secondary substation level, and hence it is easily repeatable in other contexts.

Finally, a noncommercial building HVAC energy forecasting procedure was presented. It was designed for an MPC application that makes use of a black-box building model based on state-space equations, with the goal of minimizing the energy absorption, applied to a University of Genoa building during the PREDICT project. As input, the proposed model uses the estimation of the building occupancy, evaluated through a deep transfer learning technique that exploits images taken by a thermal camera, as well as textual data of room booking coming from the University of Genoa website.

The results show that the occupancy estimation is accurate on the test set and that the state-space approach provides good results. Moreover, a sensitivity analysis of the controllable variable (HVAC temperature setpoints) confirms that the model is able to capture the realistic behavior of the considered HVAC system. Future works can exploit the proposed approach within an advanced MPC strategy.

Although good, these predictive analytics solutions per se require a human to extract meaningful insights for informing proactive decisions. A very powerful way to use them could be plugging them into a wider solution for automating the decisions that will lead to better management of energy assets. This can be done by prescriptive analytics, which is the topic of the next and final chapter of this thesis.

Prescriptive Power System Analytics

*Nothing is more difficult,
and therefore more precious,
than to be able to decide.
- Napoleon Bonaparte*

In this final chapter, prescriptive analytics is presented. The questions being answered are the following:

- What should happen to the system?
- What should be done to make it happen?

The first question (“what should happen to the system?”), in part, it is not the scope of analytics, as it involves a choice on what the system should do. But once the overall goal is set (such as financial, environmental, or other types of goals), it should be operationalized (i.e., transformed in a scalar, numerical number to be either maximized or minimized).

For example, the overall objective of the system could be maximizing profits.

But profits can be computed over different time horizons, thus giving more or less weight to long and short-term profits. As another example, one could aim to reduce the environmental impact of the system. As one might expect, "impact" is an imprecise word: it could mean to minimize the fossil fuel consumption, or maximize renewable generation, or minimize renewable curtailment, or even maximize profits, in contexts where they are strictly connected to overall positive actions towards the environment (e.g., in flexibility markets).

Even with the same choice of objective, different numerical formulations can be given, leading to very different actions and results.

The second question ("what should be done to make it happen?") is entirely in the scope of prescriptive analytics. It involves deciding which action to implement and then implementing it. There are mainly two dimensions along which decisions may be automated [241, 242]:

- One is the decision-making process stage;
- The other is the Level Of Automation (LOA);

In the taxonomies proposed in the literature, the decision-making process is typically divided into four stages that mimic the human information processing system itself:

- Phase 1. Firstly, there is the monitoring phase ([242]), or information acquisition phase ([241]), also tied with information presentation ([243]). Here, if automation takes place, it takes the form of Descriptive Analytics (Chapter 2). The human faculty being automated is the sensory system [241];
- Phase 2. Secondly, there is the generation of options phase ([243]), or simply generating phase ([242]), and involves information analysis ([241]). Automation made with analytics takes the forms of Diagnostic Analytics (Chapter 2) and Predictive Analytics (Chapter 3). The automated human faculty is the perceptual system and the working memory system;
- Phase 3. Thirdly, there is the selecting phase ([242]), which is the actual decision-making phase, or selection of course of action phase ([243]) or decision selection phase ([241]). This is the phase related to prescriptive analytics techniques and automates the decision-making mechanism in the human organism;

- Phase 4. Lastly, there is the implementing phase ([242]) implementation of action phase ([243]) or action implementation phase ([241]). This is mostly out of the scope of analytics. Indeed, the action is typically delegated to another agent, i.e., the user, or another automated system (such as with Proportional–Integral–Derivative (PID) control systems), via a setpoint to be reached or pursued. This phase may automate all other biological systems of the human being not listed before.

As for the level of automation, there is a variety of proposed taxonomies in the literature. However, all of them go from no automation to full automation. They also may differ in which phase of decision-making is being automated.

For example, in [241] decision-making and action phases LOAs are listed, based on [244]:

1. Computer offers no assistance: the human must take all decisions and actions;
2. Computer offers a complete set of decision/action alternatives;
3. Computer narrows the selection down to a few;
4. Computer gives one alternative;
5. Computer gives one suggestion and executes if human approves;
6. Computer allows the human a restricted time to veto before automatic execution
7. Computer executes automatically, then necessarily informs the human
8. Computer informs the human only if asked;
9. Computer informs the human only if the computer decides to;
10. Computer decides everything, acts autonomously, ignoring the human.

In this case, automation starts from level 5 to level 10.

The same papers point out that a system can have different LOAs in different decision-making stages (see Figure 4.1

Alternatively, in [242], the action implementation is automated to some extent already from level 2:

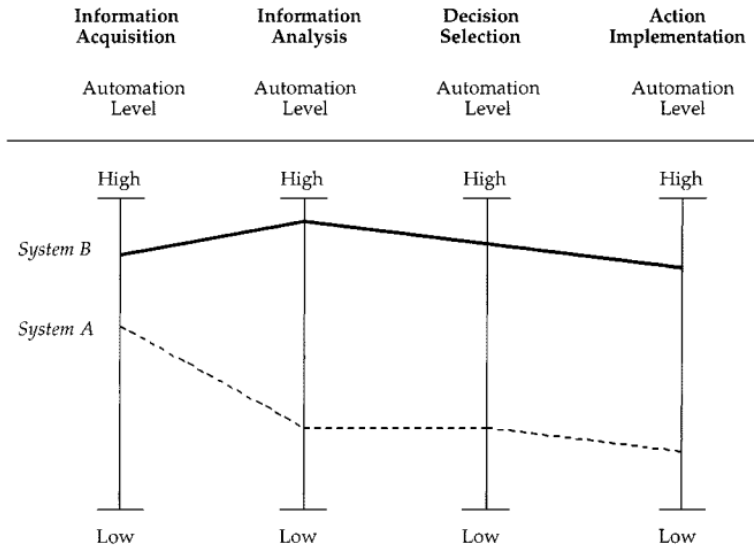


Figure 4.1: Systems having different levels of automation across decision-making stages [241]

1. **Manual Control.** The human performs all tasks;
2. **Action Support.** The system assists the operator with the performance of the selected action, although some human control actions are required.
3. **Batch Processing.** The human generates and selects the options to be performed, which are then passed to the system to be carried out automatically.
4. **Shared Control.** Both the computer and the human generate possible options for decision-making. The human retains full control over the selection of which option to implement. The actual realizations of the actions are shared between the human and the system.
5. **Decision Support.** The computer generates a list of decision options from which the human can select. Also, the operator may contribute to

generating additional options. The human selects an option, which is turned over to the computer for the final implementation.

6. **Blended Decision-Making.** The computer generates a list of decision options that it selects from and carries out the action if the human consents. The human may approve of the computer's selected option or select one from those generated by the computer or the operator.
7. **Rigid System.** This level represents systems presenting only a limited set of actions to the operator. The operator's role is to select from this set and may not generate any other options. The system will fully implement the selected actions.
8. **Automated Decision-Making.** The system selects and implements the best option based upon a list of generated alternatives (possibly augmented by suggestions from the human operator).
9. **Supervisory Control.** The system generates, selects, and implements an option automatically. If necessary, the human can intervene by selecting a different option.
10. **Full Automation.** The system carries out all actions at this level. Apart from shutting down the system, the human is completely out of the control loop and cannot intervene.

A summary of this taxonomy can be found in Table 4.1

In [243] a simplified LOA spectrum is reported, where there are four LOAs, each characterized by which decision-making stage is automated (see Figure 4.2):

- In LOA1, only the monitoring phase is automated;
- In LOA2, both monitoring phase and option generation phase are automated;
- In LOA3, monitoring, generation, and decision phase are automated. Only the implementation phase is left to the user;
- In LOA4, the system is autonomous, and the human is out of the loop.

Table 4.1: Level Of Automation (LOA) taxonomy taken from [242]
(H=Human, C=Computer, H/C=Human/Computer)

LOA	FUNCTIONS			
	MONITOR	GENERATE	SELECT	IMPLEMENT
1. Manual Control	H	H	H	H
2. Action Support	H/C	H	H	H/C
3. Batch Processing	H/C	H	H	C
4. Shared Control	H/C	H/C	H	H/C
5. Decision Support	H/C	H/C	H	C
6. Blended Decision-Making	H/C	H/C	H/C	C
7. Rigid System	H/C	C	H	C
8. Automated Decision-Making	H/C	H/C	C	C
9. Supervisory Control	H/C	C	C	C
10. Full Automation	C	C	C	C

















		Task processing stages			
		Monitoring & Information Presentation	Generating of Options	Decision Making/ Selection of Course of Action	Implementation of Action
Level of Automation	LOA 1				
	LOA 2				
	LOA 3				
	LOA 4				

Figure 4.2: Levels of Automation according to [243]

Complete LOA taxonomies that span across all stages also exist, for example in [245, 246]

Moreover, specific applications may have their LOA taxonomy, which may be more specific about the functions being automatized (e.g., autonomous driving [247]).

Apart from clarification, these nomenclatures can help in designing systems. For example, [241] provides a scheme (reported in Figure 4.3) that can help in evaluating which part of the system should be automated and to which level.

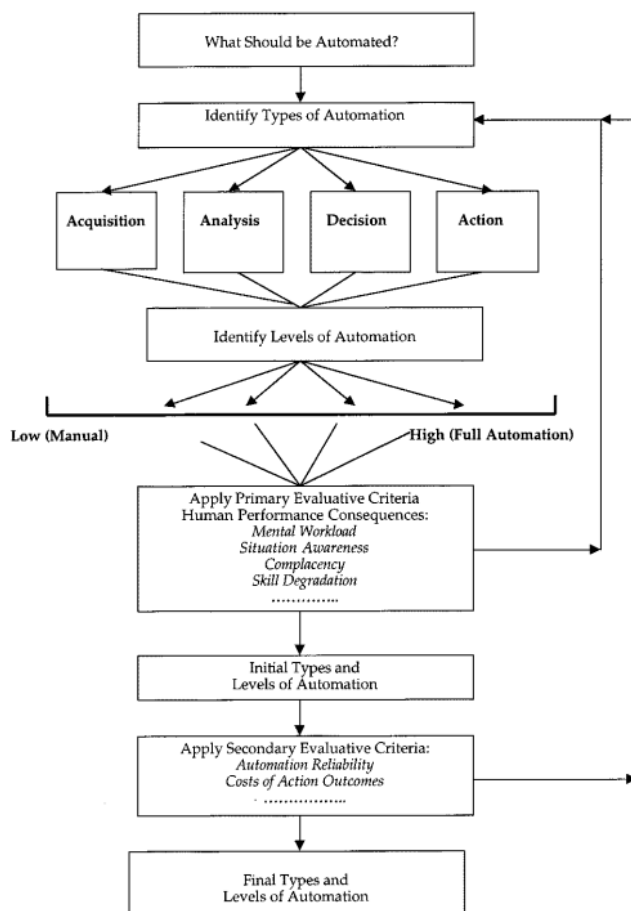


Figure 4.3: An example of automation design tool [241]

Framework in Figure 4.3 also incorporates the criteria thanks to which a design can be specified. In particular, they are:

- Human performance Consequences, such as:
 - mental workload;
 - situation awareness;
 - complacency (e.g. user trusts the system too much)
 - skill degradation
- System performance consequences, such as:
 - Automation reliability
 - Cost of incorrect decision and action outcomes

In the following of the chapter, the focus will be on the prescriptive analytics practical tools to achieve automation in phase 3 of the decision-making process. Then a brief review of how it is being applied in power systems will follow. Afterward, two applications of prescriptive analytics being developed during the Ph.D. will be detailed.

4.1 Approaches to Prescriptive Analytics

There are few literature reviews on prescriptive analytics tools [248]. It does not mean that such tools are rare. Indeed, they bring the most value to business across a wide range of industries (see Figure 4.4). Probably this is due to the wide variety of tools that can be used for prescriptive analytics.

A recent review identified 6 classes of tools to do prescriptive analytics (Figure 4.5) [248]. It also notes that most works use most than one method to implement prescriptive analytics. At the same time, it recognizes that Mathematical Programming (MP) is widely the most used tool to implement it.

MP is one of the main tools in the broad area of Operations Research (OR). OR started first in the United Kingdom and then in the USA during the Second World War as a discipline for contributing to the war effort on the part of prominent scientists. After the war, scientists that participated in OR projects started to apply the developed tools also in the engineering and the commercial fields.

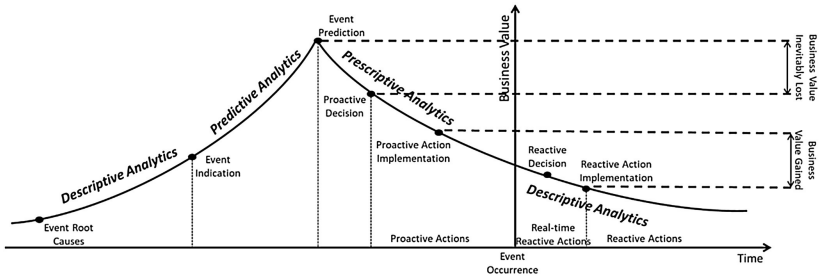


Figure 4.4: Various analytics value as function of time [241]

As such, it is a broad and interdisciplinary field, involving mainly mathematics and computer science, which is still hard today to define [249]. Some authors even use it as a synonym for prescriptive analytics [250].

Consequently, in the following MP will be defined and presented. Afterward, its intersections with Machine Learning (ML) methods will be outlined, and finally, a brief overview of all the other possible approaches will be described.

4.1.1 Mathematical Programming

Mathematical Programming (MP) was born in the context of OR, precisely in the US Air Force. In that context, a "program" was a schedule for logistical activities, which involved a huge amount of data to be laid out [249]. Indeed, months were needed to calculate those programs. Scientists involved in OR used both mathematics (in particular, optimization techniques) and the first computers to compute as quickly as possible those programs. The earliest form of mathematical program was a linear function to be minimized (the objective) over some variables (the decision variables) subject to linear equations and inequalities (the constraints). The same people also devised the simplex algorithm for solving those problems. In Equation (4.1) the Linear Programming (LP) is specified.

$$\begin{aligned} \min_x \quad & c^T x \\ \text{s.t.} \quad & Ax \leq b \end{aligned} \quad (4.1)$$

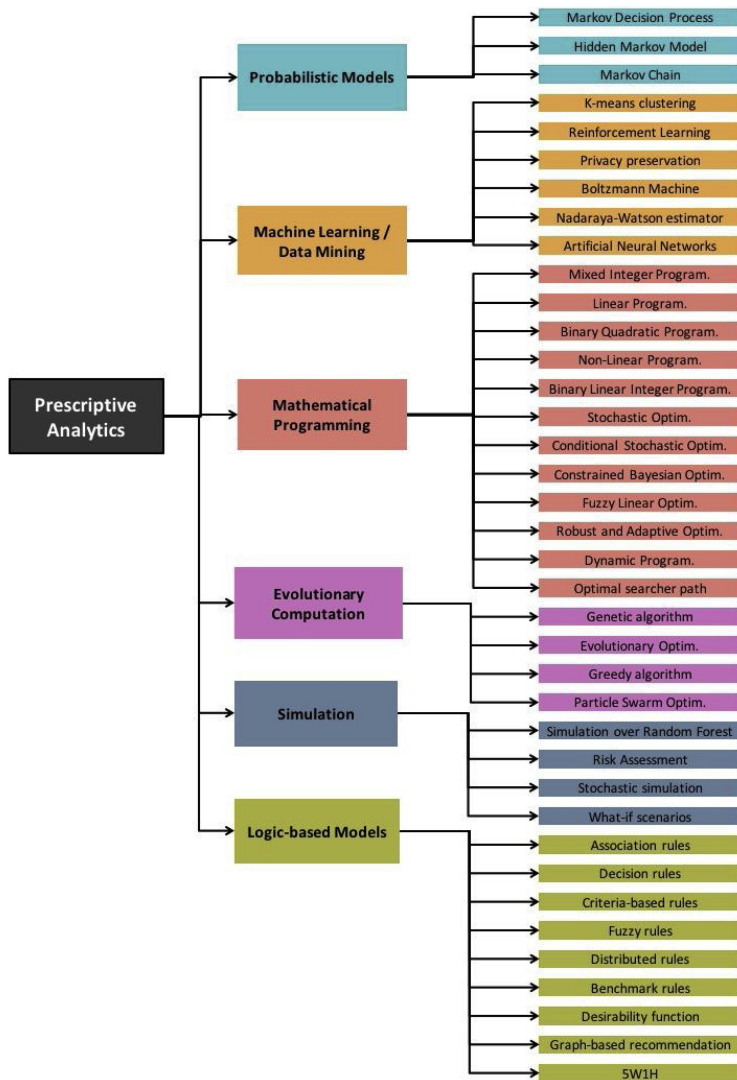


Figure 4.5: Methods for prescriptive analytics according to review [248]

Nowadays, the possibilities of MP are much wider: many types of functions (such as quadratic) can be optimized, and many types of constraints (such as conic constraints) and decision variables (notably, integer variables) can be handled. MPs can be divided into classes according to the objective. For example:

Also, not only planning and scheduling but also control and design problems can be solved through mathematical programming.

A MP nowadays can be solved by the use of the following components:

- **Model.** The model is the programming problem itself. For example, the Equation (4.1) is the model of a generic LP problem. In a model, the parameters to be specified by the data are defined, e.g., in Equation (4.1), c , A , b are all defined by the model. The language with which the model is defined is usually as close as possible to the standard mathematical notation (algebraic modeling languages);
- **Data.** Typically in a separate file, the model's parameters are specified for the problem at hand. This allows for calling the same model with different data, enhancing the portability of the problem;
- **Solver.** A solver is a program that takes a model and the data and computes (or at least tries to compute) the solution to the program. Solvers are typically specialized in some types of optimization problems: for example, there are LP solvers, Mixed Integer Linear Programming (MILP) solvers, among others.

Typically these components are assembled by a standard script program, such as python or MATLAB, which can also incorporate the logic on how and when to call the optimization problem. This is important in the context of the optimization over time since techniques such as Model Predictive Control (MPC) rely on a sliding window to carry out each optimization that constitutes the control problem solution.

4.1.2 Optimization under Uncertainty

In the presence of uncertainty, specific techniques are to be devised to solve mathematical programs.

In particular, those techniques fall collectively under the name of Stochastic Programming (SP), in which some quantity (parameter) is uncertain and

so not the objective but rather the expected objective is optimized across all the uncertainty realizations [251].

Randomness in uncertain parameters is modeled with probability distributions. It can be the case that the decisions have to be taken before knowing uncertainty realizations (single-stage SP). Additionally, corrective actions can be taken after uncertainty realizations (SP with recourse). Two-stage stochastic programming problems can be computationally expensive to solve: in this case, decomposition techniques, such as Benders decomposition, can be successfully used [251].

Another framework to deal with uncertainty is Chance Constrained Optimization (CCO). In CCO, an objective is minimized, while the constraints are satisfied with a certain probability, with the aim of trading-off performance (minimization) and reliability (constraints satisfaction) [251].

While a typical constraint is expressed in the form

$$G(x, \xi) \leq 0 \quad (4.2)$$

where x is the decision variable and ξ the uncertain parameter, chance constraint is expressed in the following form:

$$\mathbb{P}[G(x, \xi) \leq 0] \geq 1 - \varepsilon \quad (4.3)$$

with ε being a predefined risk level (typically small, so the probability of constraint satisfaction ε is high). Hence, to be defined, a CCO problem needs a risk level to be sustained by the user [251].

There are two main problems in applying CCO [251]:

1. the computation of probability of constraint satisfaction is typically intractable, although some ways to overcome this have been devised;
2. In general, feasible sets are not convex. Convex reformulations are possible in some cases.

Another paradigm for optimization under uncertainty is Robust Optimization (RO). RO is based on the concept of uncertainty set, a set in which uncertain realization can occur [251].

The idea of robust optimization is to hedge against the worst-case within the uncertainty set [251], for example:

- realization giving rise to the largest constraint violation;

- realization leading to the lowest objective value;
- the one with the highest regret.

There are two main kinds of RO: static and adaptive RO, each one corresponding to single-stage problems and problems with recourse. In particular, in adaptive RO the uncertainty is treated as a function of an ellipsoid (or another convenient figure) [251], that can depend on, for example, the probability distribution of a forecast [131].

Each approach presented (SP,CCO,RO) has its own limitation, and hybrid solutions are possible [251]. Nevertheless, SP is difficult to implement in practice [248].

4.1.3 Machine Learning Tools

With the rise of Artificial Intelligence (AI), the methods for doing optimization under uncertainty have known a renewed research interest. These researches try to do SP, CCO and RO by leveraging data-driven methods, especially in modeling the probabilistic distribution involved. Also, other approaches have emerged, such as scenario-based optimization [251]. Common to all these approaches is that they do not assume the uncertainty model to be given a priori, but they learn it from data. Such adaptivity can prevent sub-optimal solutions [251].

In this context, Distributionally Robust Optimization (DRO) is an approach that is based on the concept of ambiguity set, that is, a set of probability distributions, constructed based on available data, typically using first and second-order information [252]. DRO hedges against the worst-case distribution. By incorporating data into the definition of ambiguity set, DRO generalizes better, and it is more tractable than SP [251]. Moreover, it is often more interpretable than standard approaches [248]. However, ambiguity set constructions are not always guaranteed to converge to the true probability distribution.

Also, data-drive CCO is based on the ambiguity set concept. Nonetheless, it does not hedge against the worst-case distribution but rather uses the ambiguity set in the constraint definition [251]. But also in this version of CCO, computability is a non-trivial issue [251].

An interesting version of data-driven optimization under uncertainty is the scenario-based optimization approach [253]. This framework does not require the explicit knowledge of probability distributions of uncertain parameters [251]. Instead, using a sample of possible realizations, it seeks an optimal solution having a high probabilistic guarantee of constraint satisfaction [251]. It can be regarded as a RO with a discrete uncertainty set [251]. Also, it is a practical way to achieve approximate solutions to CCO problems. The scenario-based optimization approach has gained popularity, mainly because it reduces the stochastic optimization problem to a deterministic one. As such, scenario-based optimization leverages all the tools developed for deterministic programming.

All of these approaches are open-loop, in the sense that the optimization model and the data-driven system (usually a forecasting model) do not communicate. If the data-driven system induces bad solutions, typically, it has no feedback from it and cannot self-correct. Indeed, the training of the forecasting model is made independently of the optimization model it concurs to solve. This is a problem since it is known that closed-loop control systems are better in performance than open-loop systems. Although closed-loop solutions do exist (in particular Reinforcement Learning (RL) solutions), closed-loop versions of the data-driven stochastic optimization approaches are rare. The key could be to render the objective functions of the forecasting model training procedure dependent on the mathematical programming to be solved, for example, by using a weighted sum of the objectives. Alternatively, iterative schemes between forecasting and mathematical programming could be devised [251].

Another way data-driven techniques can enhance standard mathematical programming is when exact solutions cannot be found (as for many classes of non-linear problems) or cannot be found in a reasonable amount of time. Also, most of the best solvers are tied to an expensive license, hindering their use in cases where strict budget limits are present in building solutions. In those cases, approximate solutions can help in building prescriptive analytics solutions.

One example can be the Multi Parametric Programming (MPP) approach. MPP aims at estimating a function that outputs the optimal solution of a mathematical program given the parameters of the program [254]. Loosely speaking, it predicts the optimal solution based on parameters of the problem, employing Supervised Learning (SL), where the example inputs are the parameters, and the output is the optimal solution.

The dataset used is derived from a number of problems for which the optimal solution has been found. The estimated function can then be used to “predict” the optimal solution to the prescriptive problem, even for parameters never seen in the original dataset. Of course, in general, the resulting optimal value is not guaranteed to be actually optimal, nor to satisfy all the constraints of the original problems. However, the prescriptive analytics problem is reduced to a simple function evaluation instead of a programming problem for which a solver is needed. This can be very useful when solution time is critical, for example, in fast online applications.

Moreover, also metaheuristic techniques, such as genetic algorithms, can be used for such optimization problem instances. Indeed, they are powerful global optimizers [248]. However, their convergence to a nearly-optimal solution is not guaranteed in a fixed time, and the metaheuristic solver is still needed.

Another kind of approximate solution is given by Bayesian optimization, already shortly introduced in Section 3.1.4 in the context of hyperparameter optimization.

Finally, more and more ML is being exploited for difficult classes of problems in the solver themselves, such as those involving Combinatorial Optimization (CO) [250]. Even in those cases, challenges of feasibility are present, as well as scaling issues and how to actually generate an ML model that generalizes to unseen data.

A specific kind of ML tool for prescriptive analytics is RL. RL algorithms require that the problem be expressed as a multi-stage decision problem known as Markov Decision Process (MDP). The MDP model offers a formal mathematical language describing sequential control operations. In such operations, the outcomes are partly uncertain and partly informed by the actions of the decision-making agent [255].

An MDP consists of a set of states within a definite state space, a set of possible actions within a definite action-space, a reinforcement function, and a state transition function or probability. The agent’s objective is to maximize the total reward. The reward is any scalar quantity that can be used to implicitly communicate the objective of the learning activity to the agent. Thus, suitable reward shaping is essential to achieve the desired objective.

As one of the few closed-loop optimization frameworks, RL is promising in providing a wide application of prescriptive analytics. However, despite its ground-breaking results in recent years, RL still poses challenges to possible applications.

Indeed, the efficiency of the solution depends on the method being used, with no easy criteria. Also, conceiving an MDP that best represents the optimization environment is a complex task. Indeed mapping the state space in a mathematical space is not trivial, especially in industrial applications. Moreover, simulators are used for training RL, but transferring RL agent from a simulated environment to the real world is not guaranteed to achieve good policies. Although RL is suited for self-adaptation in dynamical environments, stability is still an issue.

4.1.4 Other Prescriptive Analytics Tools

In the following other possibilities for prescriptive analytics will be briefly introduced.

- **Simulation.** While in this thesis simulation has been located in Section 2.1.1.3, within Diagnostic Analytics, it is also widely used for decision-making. Modeling the dynamics of the system can be a hard and long process (and inherently imprecise). Yet, if the modeling is carried out, it can help make decisions even more than other tools. Indeed, what-if scenarios can be simulated and then analyzed for arriving at a good decision. It is even more convenient if the system to be studied is complicated or data gathering is expensive or risky. Note also that most simulators rely on first principles rather than analytics [248]. Examples of widely used simulators in power systems are EnergyPlus [81], and DIGSilent [256];
- **Logic-based models.** While it is hard to arrive at a mathematical program formulation, an RL solution, or even more to a simulator, it is easier and more common to have an idea on what should be the course of action to be taken. These broad types of tools may include simple rule-based systems to more complex expert systems, possibly involving interaction between the expert and a rule learning system [248];
- **Recommender systems** are a particular kind of prescriptive analytics tool aimed at providing suggestions to the decision-maker, rather than selecting a definitive action to be taken. They can be based on unsupervised techniques such as clustering and collaborative filtering, as well as predictive analytics techniques that leverage past decision data in order to predict what actions should be suggested [257];

- Risk assessment frameworks. Risk assessment tools can be used in the context of decisions where risk is the main uncertainty factor. Risk is defined as the probability of an undesirable event multiplied by the impact of that event. These methods are mostly a guideline for experts in order to decide, but they may include some analytical and stochastic aspects in order to be as less subjective as possible [258, 259].
- Stochastic Processes. Probabilistic models such as stochastic processes may be used to model volatile quantity such as market prices in order to estimate the probability of undesirable events to be hedged against [128].
- Statistical Inference. Other statistical tools such as sampling techniques and design of experiments can be thought of as prescriptive analytics tools in contexts where specialized decisions have to be made (typically in situations of small data, where the data that can be gathered is small or when it is unfeasible or too costly to use all data to draw conclusions).

4.2 Prescriptive Applications to Power Systems

While power systems are heading towards a complete RES integration, a high number of new and important decisions have to be made at all levels: investment, planning, and operational. Plenty of expectations are put on analytics in order to prescribe an optimal course of actions at all levels [16].

In the future, AI and big data will further enhance decision-making and planning, condition monitoring, inspections, certifications, and supply chain optimization and will generally increase the efficiency of energy systems [16].

Examples of prescriptive analytics applications abound: from customer characterization, to demand response program design, to retail pricing [86], to predictive maintenance, HVAC optimal control, EV smart charging, microgrid optimization, including BESS, as well as all kinds of decisions to be made by systems operators in order to keep the system safe and with quality service.

Much like prescriptive analytics is about automating decision-making processes, also predictive maintenance is the final step of a more and more automated approach to maintenance.

- The first step is reactive (or run-to-failure) maintenance, where repairs are made after the damage occurred only [260];
- The second step is preventive maintenance, where a machine is inspected at predefined time intervals [260];
- Next is Condition Based Monitoring (CBM), where maintenance actions are dynamic but based only on current states (the maintenance system is also called diagnostic system);
- Whenever also future states are estimated, and maintenance actions are made before damage can occur, predictive maintenance (or prognostics) is involved [260].

While diagnostic and partly the prognostic part of predictive maintenance are applications of predictive analytics, the action to take in order to be proactive with respect to the damage is an application of prescriptive analytics. In some cases, highly data-driven techniques are not suited because of the few failure data. For these reasons, efforts to share data on failure are carried out.

Predictive maintenance can significantly enhance the useful life of RES asset. For example, wind turbines operations contribute up to the 20% of the total cost of wind energy over a period of 20 years. In this sense, predictive maintenance can greatly help reduce the cost of renewable energy and thus contribute to the energy transition [260]

Not few predictive maintenance techniques integrate physical models and simulation into the final solution [261]

As already noted in Section 3.5, buildings contribute a lot to energy demand and Carbon Dioxide (CO₂) emissions. A smart Building Energy Management System (BEMS) (that controls HVAC and other building components such as lights and windows) can maximize the consumption of (possibly self-produced renewable energy, while also balancing building occupants' comfort requirements.[39].

Typically HVAC are run with simple Supervised Learning (SL), and many studies have suggested that substituting them with MPC can significantly benefit the current efficiencies KPIs [262].

While being efficacious on paper, MPC can have significant development and operational requirements. Hence in [262] a simplified control law procedure learned from MPC has been proposed (similar to the aforementioned MPP approach to an approximate solution to programming problems).

Also, work exploring the links between the energy efficiency improvement and the flexibility services of single and multiple buildings were studied, using both SL and RL [263, 264]. Heating, cooling, and ventilation optimization were investigated in [265], also involving occupancy data.

While today the number of EV is manageable, its number is predicted to increase tremendously in the next few years. Apart from infrastructural upgrades, new techniques for scheduling and control Electrical Vehicle Supply Equipment (EVSE) will be needed because the concurrent charging of millions of EVs can have a number of adverse effects on power grids [266]

Conflicting goals are at stake: while vehicles should be charged as quickly as possible, utilizing RES for the charge is important, as well as giving flexibility to the system, and also economic profits may be taken into consideration. Also balanced charging between near EVs is advisable [267].

In [268], a RL-based optimal charging strategy model for a DSO to address the voltage violation problem is proposed. A review of RL approaches on the problem of EV charging is carried out in [269]. A simulator for developing such algorithms is introduced in [267], while a more classical optimization solution is proposed to illustrate the goodness of the simulator. In [270] a modified version of laxity algorithm (a sort of heuristic way to prescribe smart charging) is reported. The problem of Electrical Bus (EB) Fleet smart charging is addressed in [271]. The simulation results show that the proposed algorithm can significantly reduce the battery degradation cost during the operating time of EBs while giving relatively small errors when considering randomnesses of EBs charging processes. Moreover, the proposed algorithm can reduce the computational complexity significantly for practical applications [271].

In the future, the real-time electric price can be considered as an external input for the state of MDP. Since the price is time-variant, forecasting the price is a great challenge. Further, the model-free approach can be potentially improved by combining the reinforcement learning with the EBs mechanical model to increase the accuracy of the charging schedule, which still requires extensive research [271].

Works on optimizing a microgrid are also very common in the literature. Batch RL is proposed to manage a battery in a microgrid by taking the inverter's efficiency into account [272]. Deep Reinforcement Learning (DRL) is applied in [273] to enhance the Energy Management System (EMS) of a microgrid. A review on the same problem is given in [255]. A simulator for benchmarking different RL solutions have been developed and presented

in [274]. A genetic algorithm is used to optimize an industrial microgrid in [275]. A mixed-integer algorithm written in GAMS [276] and developed in MATLAB is used in [277] to optimally dispatch a BESS based on day-ahead PV forecasting. A chance-constrained method for maximizing economic gain in an integrated PV-BESS system is presented in [278].

Another line of prescriptive analytics applications in power systems is the sizing and siting of BESSs. Both are planning problems: in one case, the optimal size according to some objective in terms of capacity or power is to be decided, while in the siting problem (also called allocation, site selection, or placement), the decision regards the site of the BESS. There are two types of BESS owners: system operators and demand-side owners. System operators may use BESS for ensuring grid operation under the pressure of high penetration of RES, while demand-side owners are interested in self-consumption or participation in the markets. Services offered by BESS are numerous. Among them are frequency regulation, black start services, flexible ramping, congestion relief, and peak shifting, RES capacity firming, which lead to reduced curtailment of RES, reduced reliance on thermal generators as well as capacity investment deferral [279]. For each of them, investments are needed, and suitable sizing (and siting when possible) must be found.

Examples of prescriptive analytics solutions for system operators include: optimal sizing considering a future prediction about PV penetration [280]; joint transmission and distribution side benefits satisfying chance-constrained security conditions [281]; BESS planning under high penetration of wind generation at the distribution level [282]; multiple services such as voltage constraints, system losses and total cost of energy minimization [283]; sizing and siting for minimizing costs of voltage deviations, power losses and peak demands [284]; siting and sizing for decreasing the voltage impact of PV systems in a distribution system [285]; allocation and optimal size for dealing with over-frequency in transmission systems [286]. Also demand-side owners need planning, for example, for minimizing PV grid export and increase self-consumption [287], for jointly sizing PV systems [288], and for minimizing investment costs and ensuring continuous operation [289].

Other types of objectives can be found in BESS siting and sizing literature: for example, for protecting the grid against adversarial attacks [290]; in conjunction with HVAC optimal operation, so that size of BESS can be minimized [291]; and allocation and sizing on shipboards [292].

Most works rely on deterministic programming, or their stochastic counterparts, and rely on simulations of the future state of the market, the grid, and of

the demand [293]. Computations for these kinds of problems are intensive, so many examples of alternative solutions to mathematical programming, such as genetic algorithms [282, 284, 294, 295], and reinforcement learning [296] can be found in the literature.

Before concluding this chapter, two applications drawn from the Ph.D. work are reported. The first application [61] presented is about an optimal storage siting and sizing methodology with the objective of helping an Italian DSO, involved in the PODCAST project [59], in respecting voltage and current constraints required by grid codes. The second is a Microgrid/Energy Community profile optimization algorithm, with day-ahead and intra-day stages, developed during the PODCAST project.

4.3 Application: Optimal Storage Siting and Sizing

Distributed Energy Resources (DERs) are changing the paradigms of network planning and management: the old "fit and forget" philosophy cannot be applied to modern active distribution networks, whereas smart technologies cover a key role in the integration of these actors in the system operation. In particular, in order to face new bottlenecks of power transmission, a European Directive [297] requires to evaluate all possible smart solutions before investing in grid reinforcement. Advanced communication technologies together with robust and efficient control strategies must be exploited by DMSs in order to coordinate simultaneous DER contributions to regulation, which include RESs, flexible loads and BESSs. Italian requirements for DER reactive power support and communication infrastructure are defined in [298, 299].

In this scenario, Volt/Var Optimization (VVO) methodologies represent crucial functionalities for the optimization of voltage profiles and reactive power. These techniques can help evaluate the most efficient configuration for the controllable devices, reduce operating costs, and minimize power losses. In addition, a robust VVO algorithm can contribute to congestion relief. Storage systems can be considered one of the main tools inside these strategies and, in order to maximize their contribution, DSOs should exploit advanced methodologies for the optimal sizing and allocation [300, 301].

In recent years, the scientific community has proposed a great number of VVO techniques. Authors in [302] face the problem of VVO for distribution systems with RES penetration. In [303, 304] multi-agent methodologies are

designed for loss reduction and voltage control, while in [305] a holistic approach for integrated Volt/Var control in medium and low voltage networks is presented. Finally, authors in [306] describe a strategy for the optimal sizing and allocation of Distributed Generators (DGs) within a VVO strategy, while in [307] also the grid reconfiguration is embedded.

This application aims to define and test an algorithm for DER regulation and optimal allocation of storage devices in a distribution grid. The presented formulation finds the optimal sizing and siting of storage devices necessary to guarantee the respect of voltage and current constraints when dispatchable DERs in the considered network are not sufficient.

Unlike most works in literature, a linearized technique has been chosen for the mathematical formulation, which leads to minor approximation errors (as testified by simulations). Moreover, it also leads to several benefits such as computational speed, problem scalability, working point flexibility (since the linearization is performed around nominal voltage), and the possibility to pursue multiple tasks simultaneously. The proposed procedure is tested on the real distribution grid model of the city of Sanremo, using Monte Carlo extractions from measured data. In fact, available real data from smart meters are essential to derive load and RES profiles suitable for integration in stochastic planning.

4.3.1 Optimization Problem Formulation

The proposed formulation uses MILP computation and is implemented in Matlab-GAMS environment with CPLEX solver [308]. It is an upgrade and extension of [309] and [59], where slack voltage regulation has been added, and robust probabilistic planning is performed using Monte Carlo technique including load and RES profile modeling. The presented procedure leverages a linearized approach to load flow equations (reworked from [310]) and magnitude constraints (revised from [311]), which require integer variables to describe Piecewise Linear Functions (PLFs). The problem anyway exploits integer variables to identify the optimal number of battery modules that should be installed at each node along the grid. Therefore, all network elements are correctly represented by the linear model, including their operational limits.

Briefly showing the upgrades with respect to [59], linear power flow equations compute Cartesian coordinates of voltages (V_N^{re} and V_N^{im}) as a function of nodal injections (using ZIP model: constant impedance (S_{ZN}), constant current (S_{IN}) and constant power (S_{PN})), defining in origin the constant

matrices A , B and C that depend on the linearization point. In order to isolate also slack voltage V_s and make it adjustable, equations have been further decomposed by adding slack voltage coefficient D .

$$A_{re} = (-B_{re} - C_{re})V_N^{re} + (C_{im} - B_{im})V_N^{im} + S_{IN}^{re} - D_{re}V_s \quad (4.4)$$

$$A_{im} = (-B_{im} - C_{im})V_N^{re} + (B_{re} - C_{re})V_N^{im} + S_{IN}^{im} - D_{im}V_s \quad (4.5)$$

$$A = -2S_{PN}^* \quad (4.6)$$

$$B = \text{diag}(S_{PN}^*) \quad (4.7)$$

$$C = Y_{NN} - \text{diag}(S_{ZN}^*) \quad (4.8)$$

$$D = Y_{NS} \quad (4.9)$$

It is recalled that since linearization is performed on voltage deviations, the approximation is accurate when voltage magnitude is close to 1 p.u. (which is true in the performed scenarios), but it is valid for any working point.

Voltage and current constraints are linearised as in Figure 4.6, fully defined in [309].

The objective function to be minimized, as already said, is the total number of battery modules employed to satisfy operational constraints. Assuming that installation cost is equivalent at all substations, a simple summation on all k nodes of the network is computed:

$$\min \sum_{k=1}^N (N_{BESS}(k)) \quad (4.10)$$

The problem can be reduced if only particular nodes are available for the installations due to logistic or space requirements. The slack variables which the formulation leverages for regulations are nodal constant power injections, S_{Ik}^{re} and S_{Ik}^{im} , here represented by DGs reactive injections and both active and reactive contributions of BESSs (indeed modeled as current-controlled sources). DG reactive regulation is free, as long as it respects power factor limits in [298].

Concerning BESS modeling, similarly to [309], its sizing parameters are defined by the integer variable N_{BESS} multiplied by the power rating and energy capacity of each module (which are both fixed values). A quadratic capability curve is assumed. The State of Charge (SoC) is computed taking into account charge and discharge currents ($S_{I,t}^{C,re}$, $S_{I,t}^{D,re}$ with corresponding efficiencies η_c, η_d), defining the relationship between the two parameters.

In order to perform non-correlated daily stochastic scenarios, everyday at midnight the SoC is bounded to return to the initial status, as in Equation (4.11). This constraint is introduced in order to ensure the continuous operation of the BESS on a day-to-day basis. For the initial SoC, in $t = 0$, an empirical optimal value is chosen (70%, optimal at midnight for systems with PV penetration).

$$SOC_{k,t \times 24} = N_{BESS}(k) \cdot SOC_{k,t=0} \quad (4.11)$$

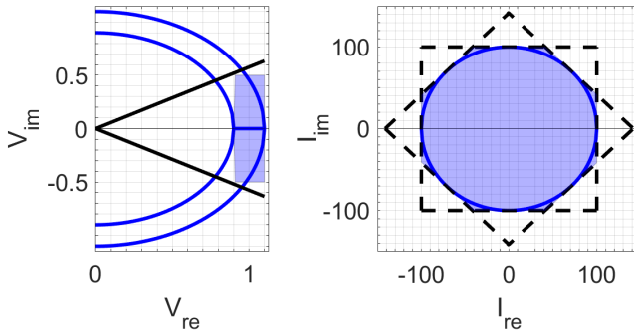


Figure 4.6: Modeling voltage (pu) and current (%) limits [309]

4.3.2 Real Network Model

The presented algorithm is applied to the real distribution network of Sanremo city (Italy) within the PODCAST project. The local DSO, AMAIE S.p.a., allows data acquisition through a dedicated SCADA system and various functionalities provided by the DMS [59, 171].

4.3.3 Grid Description

The complete distribution system is an MV network, composed of 186 MV/LV substations (15/0.4 kV/kV) and 17 MV prosumers, operated radially, with ten feeders starting from the HV/MV primary substation of Tinasso (which is 132/15 kV). Four feeders (shown in Figure 4.7) are completely monitored through measurement devices installed at the secondary winding of each MV/LV substation. The grid model of this network (represented in MATPOWER, a MATLAB toolbox for power system simulations, which is used for non-linear validation of the results) is used as input for the study, adding an equivalent load of non-monitored feeders at the primary substation.

The overall power related to the RES distributed generation in the network of Sanremo is close to 1.98 MW. According to the current standards [298, 299], DERs are connected at the LV or the MV grid depending on their size. In particular, two PV plants with nominal power of 200 kW_p and 20 kW_p are connected respectively at the MV and the LV level of substations 155 and 148 (see Figure 4.7)), while a 200 kW hydropower plant is connected at the MV side of the substation 97. The non-monitored feeders are modeled through six equivalent loads, whose actual power is monitored at the primary substation. In addition, equivalent loads connected at the LV level of secondary substations are used to model LV networks downstream MV/LV transformers. Finally, a SCADA system acquires data from each measurement point with a granularity equal to one minute (see also Section 3.4.1).

4.3.4 Load Modeling for Monte Carlo simulation

The load profiles of the substations are derived from active power historical data, using a sampling time of 15 minutes. The approach is derived from [171], with some adaptations to make it suitable for Monte Carlo simulation. The procedure is the following for each substation:

- The available data are divided into a working day dataset, a pre-holiday dataset, and a holiday dataset. This aggregation is defined according to the different load behavior for the different day types;
- For each dataset, the mean daily active power profile (composed of 96 values) is computed, as well as the standard deviation for each quarter of an hour. Missing values are omitted. Both yearly and monthly profiles are computed;
- Each raw profile is then smoothed using Fourier Decomposition [171]:

$$\hat{y}(t) = a_0 + \sum_{i=1}^6 a_i \cdot \cos(t \cdot w) + b_i \cdot \sin(x \cdot w)$$

where $w = \frac{2\pi}{96}$, α_i and β_i are the Fourier coefficients and $t = 1, \dots, 96$.

- In cases where no data are available for the monthly profile computation, the yearly profile is used, scaled by a convenient factor.

In Figures 4.8–4.10 some examples of the resulting load profiles are reported for the same substation, for the same period (July), and the three different types of day. In each plot, the raw mean profile, the profile smoothed with Fourier decomposition, the 95% confidence interval for each point of the curve (assuming gaussianity), and an example of Monte Carlo Simulation for the profile are reported. The simulation takes a random gaussian value for each quarter of an hour: the mean and the standard deviation are taken from the smoothed profile of the same time instant. In most cases, the Monte Carlo draw falls inside the confidence band region.

By comparing the three profiles of this specific substation, it is possible to see that the various types of the day have different behaviors. The holiday profile is lower in magnitude and less variable than the other two. In addition, the pre-holiday profile is slightly lower in magnitude and less variable than

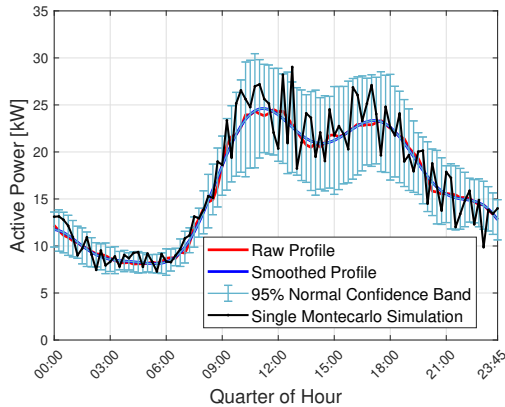


Figure 4.8: Raw mean profile (red), Fourier-smoothed profile (blue), 95% gaussian confidence interval of the smoothed profile (error bars) and example of Monte Carlo draw (black) for Substation 7 in the month of July - working days

the curve of the working days. Similar behaviors can be found for other substations. These statistical behaviors justify the dataset division employed for this work.

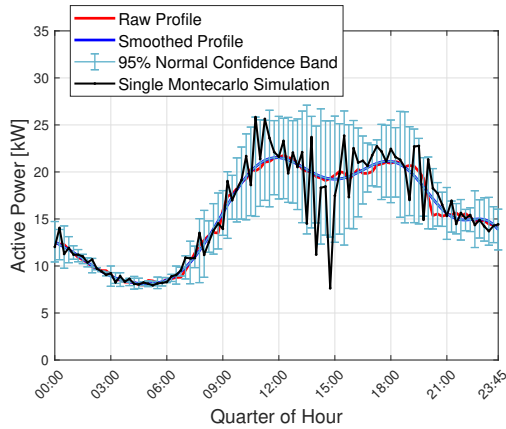


Figure 4.9: Raw mean profile (red), Fourier-smoothed profile (blue), 95% gaussian confidence interval of the smoothed profile (error bars) and example of Monte Carlo draw (black) for Substation 7 in July - pre-holiday days

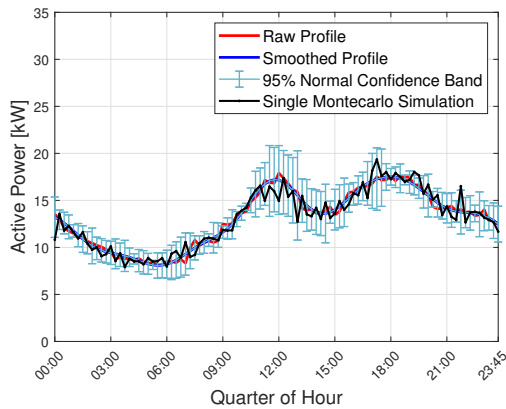


Figure 4.10: Raw mean profile (red), Fourier-smoothed profile (blue), 95% gaussian confidence interval of the smoothed profile (error bars) and example of Monte Carlo draw (black) for Substation 7 in July - holiday days

4.3.5 Simulation Results

Simulations have been performed on the test network with Monte Carlo scenarios, extracting 3000 daily profiles of loads and RESs to represent different operating conditions, taking into account hourly, daily, and seasonal variation. The computational time is approximately 4 secs for each equivalent day (on a dedicated workstation, core i, 64GB RAM), proving the scalability of the solution. The imposed range for voltage magnitudes is (0.95-1.05 p.u.), whereas the congestion limit is considered above 90% of the branch rated power. Available BESS modules are 500 kWh/1000 kW, and all LV nodes are considered as possible locations.

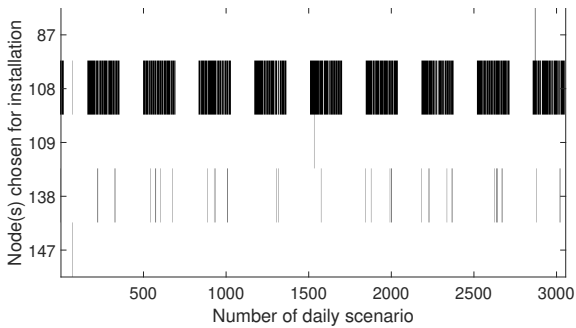


Figure 4.11: Result of BESS allocation

Optimal allocations suggested by the daily simulations performed are always composed of one or two BESS modules, located in a limited group of nodes (which have been selected in Figure 4.11, for readability). Some days do not require any storage when no violations are identified.

Therefore, the four following substations can be identified as suitable for storage installations, marked also in Figure 4.7:

- n° 220 "Nuvoloni" (bus 87)
- n° 47 "Casino" (buses 108 and 109)
- n° 87 "Mercato Ortofrutta" (bus 138)
- n° 97 "Filtri Poggio" (bus 147), same of the hydropower plant.

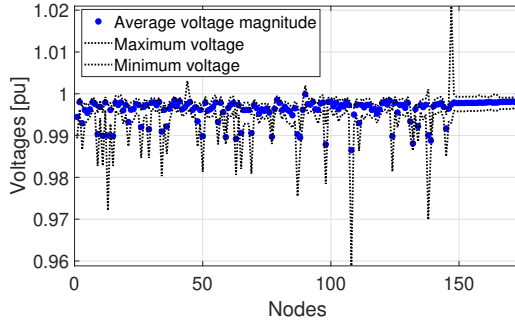


Figure 4.12: Voltage profiles over all MonteCarlo simulation in AMAIE grid

Figure 4.12 shows average voltage magnitudes computed at all nodes (sorted by number, not physical connections) with "gap areas" to represent the variations registered in all simulated scenarios: both at MV and LV level, the imposed range is respected. The highest gaps occur at substations where the load is highly fluctuating or where strong regulations can be performed.

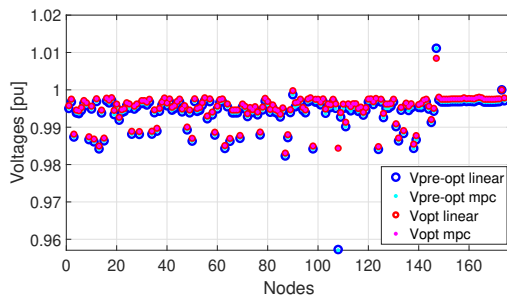


Figure 4.13: Comparison of voltages of all nodes (MV and LV) before and after optimization, one day at 8 PM, with model validation

More specifically, Figure 4.13 shows an example (in a sample scenario, at 8 PM) of comparison of bus voltage magnitudes before and after the optimization (effectively flattening the profile), each with the corresponding validation via non-linear power flow (having circles and dots overlapping testifies the

great match of outputs). As expected, the error on voltages computation due to linearization is negligible, with a mean (over 174 nodes and 72000 time intervals) equal to $3.13 \cdot 10^{-4}\%$ and a maximum of 0.006%.

Also, branch congestion limits can require, in some conditions, BESS contribution in order to respect the imposed limits. In particular, MV lines tend to be far from their maximum flows, while MV/LV transformers may encounter critical conditions: a focus on transformers flows (compared to their limit, the dotted lines) is offered in Figure 4.14. In this situation, energy storage is the only way to ensure problem feasibility.

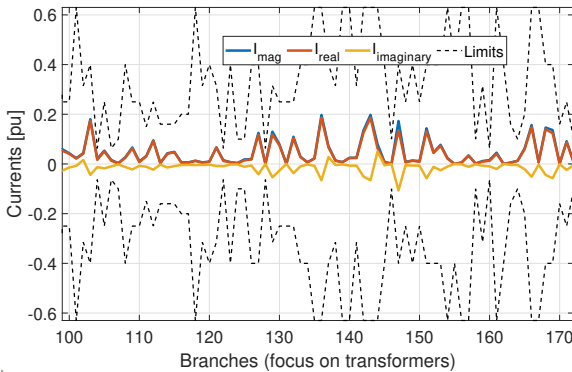


Figure 4.14: Computed values of transformers' currents (magnitude, real and imaginary components) compared to their limits

In the end, Table 4.2 reports the maximum (in absolute value) power injections required to the controllable devices at any simulated time in order to improve voltage profiles and relieve congestions. It is noted that these contributions are required in different moments: depending on the different criticalities, BESS injections can be mainly active (to solve congestion problems) or reactive (when compensation is needed to boost the voltage profile). Reactive regulation supplied by DGs is often close to the maximum allowed by [298].

Table 4.2: Minimum/maximum injections from BESS and DGs at all nodes among all scenarios, compared to corresponding limits

Nodes	Generators		Batteries		
	Q_g [kVAr]	Q_g^{max} [kVAr]	Q_b [kVAr]	P_b [kW]	S_b^{max} [MVA]
39	-47	± 48			
45	-4.7	± 5			
87			790	-347	± 1
108			653	800	± 1
109			429	-341	± 1
138			557	428	± 1
147	-46	± 2.5	130	-193	± 1

4.4 Application: Microgrid/Energy Community Profile Optimization Algorithm

Microgrids and energy communities are emerging as valuable means for the efficient and flexible management of DERs, including storage devices, among others [312]. Crucial to these applications is the capability to guarantee reserve capacity and compensate for RES and demand forecast errors [313]. Hence, control algorithms for day-to-day microgrid operation represent a hot research topic: a recent review can be found in [314]. An industrial microgrid operation solution is presented in [315]; a stochastic optimization framework for microgrids is introduced in [316], while an MPC based approach is given in [317]. Similar algorithms can be employed for aggregators [19, 318].

In this context, in the PODCAST project, a procedure for optimizing the profile of a microgrid or an energy community has been defined. The procedure comprises two parts: a day-ahead and an intra-day stage. The controlled assets are two BESSs, described in Section 4.4.3.1. MATLAB and AMPL [319] were used for modeling and testing the procedure.

The proposed procedure is able to optimize the profile of a microgrid or an energy community. The algorithm consists of two levels: In the day-ahead stage, planning of the optimal profile is carried out with mathematical programming (with a predefined time horizon and granularity). In the intra-day stage, the profile planned in the previous day is realized with a simple rule based on the actual microgrid/energy community absorption (which is updated about every three minutes). Thanks to this method, a microgrid or an

energy community can optimize its profile and participate in the day-ahead market and propose a profile that will be realized in the next day. In the following sections, the algorithm will be introduced in a general way, then the description of their application to the PODCAST project will be given, together with the results of some experiments carried out on a real distribution network.

4.4.1 Day-ahead Stage

The goal of the optimum problem is to minimize the following objective function:

$$\min_{P_{imp}, P_{exp}} \sum_{i=1}^{\frac{L}{\Delta t}} c_{imp}(i) \cdot (P_{imp} \cdot \Delta t) - c_{exp}(i) \cdot (P_{exp}(i) \cdot \Delta t) \quad (4.12)$$

where:

- Δt represents optimized profile granularity [h];
- L is the length of the optimization [h];
- i is the time index, which spans from 1 to $\frac{L}{\Delta t}$;
- $c_{imp}(i)$ is the cost of energy when the distribution grid imports from the microgrid/energy community, relative to time i [€/kWh];
- $c_{exp}(i)$ is the cost of energy when the distribution grid exports to the microgrid/energy community, relative to time i [€/kWh];
- $P_{imp}(i) \geq 0$ is the imported power at time i [kW];
- $P_{exp}(i) \geq 0$ is the exported power at time i [kW].

The decision variables $P_{imp}(i)$ and $P_{exp}(i)$ are defined by Equations (4.13)–(4.15):

$$P_{imp}(i) \leq x_{imp}(i) \cdot P_{imp}^{max}(i) \quad (4.13)$$

$$P_{exp}(i) \leq x_{exp}(i) \cdot P_{exp}^{max}(i) \quad (4.14)$$

$$x_{imp}(i) + x_{exp}(i) \leq 1 \quad (4.15)$$

where:

- $x_{imp}(i)$ is the binary variable equal to 1 if the distribution grid is importing energy from the microgrid at time i (otherwise, it is 0);
- $x_{exp}(i)$ is the binary variable equal to 1 if distribution grid is exporting energy to the microgrid at time i (otherwise, it is 0);
- $P_{imp}^{max}(i)$ is the parameter that represents the maximum power that the distribution grid can import from the microgrid at time i [kW];
- $P_{exp}^{max}(i)$ is the parameter that represents the maximum power that the distribution grid can export to the microgrid at time i [kW];

Hence, also $x_{imp}(i)$ and $x_{exp}(i)$ are decision variables.

Next, The first SoC of the storages is equal to the last measured SoC: for the s -th storage.

$$SoC_s(1) = SoC_s^{meas} \quad (4.16)$$

with S being the number of storages in the microgrid/energy community, $s = 1, \dots, S$ and SoC_s^{meas} is the last measured SoC of the s -th storage .

The dynamics of the SoCis considered in Equations (4.17)–(4.22):

$$SoC_s(i) = SoC_s(i-1) + \frac{\Delta t}{E_s^{nom}} (\eta_s^{ch} \cdot P_s^{ch}(i) - \eta_s^{dsc} \cdot P_s^{dsc}(i)) \quad (4.17)$$

$$P_s^{ch}(i) \leq x_s^{ch}(i) \cdot P_s^{max,ch} \quad (4.18)$$

$$SoC_s(i) \leq SoC_s^{max} \quad (4.19)$$

$$SoC_s(i) \geq SoC_s^{min} \quad (4.20)$$

while the constraints related to the maximum and minimum power are:

$$P_s^{dsc}(i) \leq x_s^{dsc}(i) \cdot P_s^{max,dsc} \quad (4.21)$$

$$x_s^{ch}(i) + x_s^{dsc}(i) \leq 1 \quad (4.22)$$

with $s = 1, \dots, S$ and $i = 1, \dots, \frac{L}{\Delta t}$ and where:

- $SoC_s(i)$ is the state of charge of the storage system at time i [p.u.];
- E_s^{nom} is the nominal capacity of the s -th BESS [kWh];
- η_s^{ch} is the charging efficiency of the s -th BESS [p.u.];
- η_s^{dsc} is the discharging efficiency of the s -th BESS [p.u.];
- $P_s^{ch}(i) \geq 0$ is the variable indicating the charging power of the s -th BESS [kW];
- $P_s^{dsc}(i) \geq 0$ is the variable indicating the discharging power of the s -th BESS [kW];
- $x_s^{ch}(i)$ is the binary variable equal to 1 if the BESS is charging, 0 otherwise;
- $x_s^{dsc}(i)$ is the binary variable equal to 1 if the BESS is discharging, 0 otherwise;
- $P_s^{(max, ch)}$ is the maximum charging power of the s -th BESS;
- $P_s^{(max, dsc)}$ is the maximum discharging power of the s -th BESS;
- SoC_s^{max} is the maximum SoC state for the s -th BESS [p.u.]
- SoC_s^{min} is the minimum SoC for the s -th BESS [p.u.]

Hence, also $P_s^{ch}(i), P_s^{dsc}(i), x_s^{ch}(i), x_s^{dsc}(i)$ are decision variables.

Among the other constraints, a desired property is the continuity of the algorithm, that is, the algorithm should be able to run every day, with little initial SoC condition differences between consecutive days, as indicated by Equation (4.23):

$$\varepsilon \leq SoC_s\left(\frac{L}{\Delta t}\right) - SoC_s(1) \leq \varepsilon \quad (4.23)$$

where:

- ε is the parameter representing the tolerance between initial and final state [p.u.];

For limiting the degradation of the batteries, a limit on an approximation of charging and discharging cycles is enforced:

$$\frac{\eta_s^{ch}}{E_s^{nom}} \cdot \sum_{i=1}^{\frac{T}{\Delta t}} P_s^{ch} \leq l_s^{ch} \quad (4.24)$$

$$\frac{\eta_s^{dsc}}{E_s^{nom}} \cdot \sum_{i=1}^{\frac{T}{\Delta t}} P_s^{dsc} \leq l_s^{dsc} \quad (4.25)$$

where:

- l_s^{ch} is the maximum number of charging cycles;
- l_s^{dsc} is the maximum number of discharging cycles;

Finally, the load balance equation for the microgrid/energy community is:

$$\begin{aligned} Load(i) + P_{exp}(i) + \sum_{s=1}^S \eta_s^{ch} P_s^{ch}(i) &= \\ &= PV(i) + P_{imp}(i) + \sum_{s=1}^S \eta_s^{dsc} P_s^{dsc}(i) \end{aligned} \quad (4.26)$$

where Load(i):

- Load(i) is the load forecasting at time i (see Section 3.4) [kW];
- PV(i) is the generation forecasting at time i (see Section 3.3) [kW];

To summarize, the day-ahead optimization is based on a MILP, the objective is defined in Equation (4.12) and constraints are in Equations (4.13)–(4.26), while the decision variables are:

- $x_s^{ch}(i), x_s^{dsc}(i)$;
- $P_s^{ch}(i), P_s^{dsc}(i)$;

- $P_{imp}(i), P_{exp}(i)$;
- $x_{imp}(i), x_{exp}(i)$.

The algorithm (implemented in AMPL [319] environment) defines the optimal microgrid/energy community profile:

$$P_{TOT}(i) = P_{exp}(i) - P_{imp}(i) \quad (4.27)$$

which is positive quantity if the microgrid/energy community is absorbing negative if the microgrid/energy community is exporting. Examples of finalized profile and battery optimization profiles are reported in Figures 4.15 and 4.16.

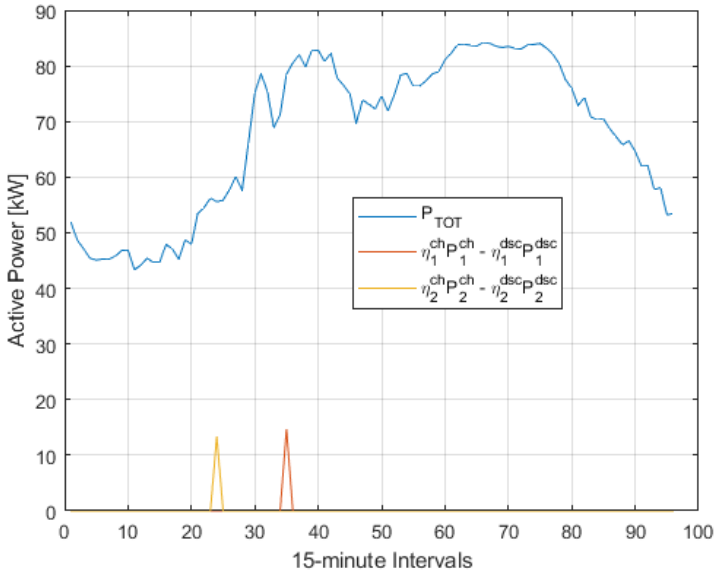


Figure 4.15: Example of day-ahead optimization output: forecasted load and BESS setpoints

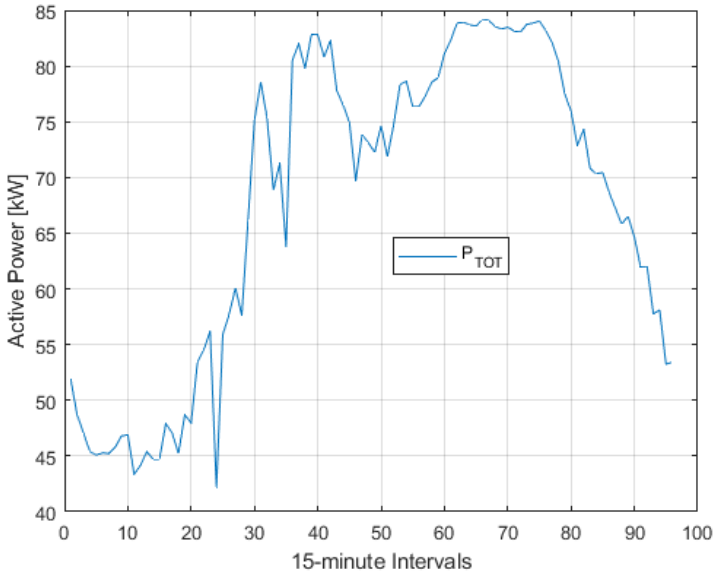


Figure 4.16: Example of day-ahead optimization output: final net load optimized profile

4.4.2 Intra-day Stage

The profile defined by the optimization of the day before must be realized during the following day. Batteries, in this case, are used to compensate for forecasting errors made on load and renewable generation. The intra-day control is performed with a granularity equal to one minute. Its goal is to realize the P_{TOT} profile defined by the optimizer of the day before. For this reason, within each Δt -minute interval of a day (defined by the time index $i = 1, \dots, \frac{L}{\Delta t}$). The intra-day stage carries out the following control logic on the storage devices:

$$P_{storage}(i, m) = P_{tot}(i) - P_{meas}(i, m) \quad m = 1, 2, \dots, 60 \cdot \Delta t \quad (4.28)$$

where:

- $P_{TOT}(i)$ is the day-ahead optimized profile for quarter of interval i ;
- $P_{meas}(i, m)$ [kW] is the microgrid/energy community absorption measured at minute m of interval i ;
- $P_{storage}(i, m)$ [kW] total setpoint that the S BESSs have to provide at minute m of interval i . Each BESS must provide $\frac{P_{storage}(i, m)}{S}$.

In particular, if $P_{storage}(i, m) \leq 0$ BESSs operate in charging mode, while if $P_{storage}(i, m) \geq 0$ batteries go in discharging mode.

Moreover, the $P_{storage}(i, m)$ is set to zero whenever Equations (4.19) and (4.20) (maximum and minimum SoC constraints) are violated.

4.4.3 Experiments

Two experiments aimed at ensuring that the algorithm can derive and follow an optimized profile were carried out the first and the second days of December of 2020 during the Ph.D. on a portion of the distribution grid of the PODCAST project (described in Section 3.4.1 and Section 4.3.3).

The start and the end of the two experiments are reported in Table 4.3.

Table 4.3: Experiments for BESS profile optimization algorithm

Exp. n.	Start	End
1	1st December 2020 - 08:33	1st December 2020 12:32
2	2nd December 2020 15:29	2nd December 2020 19:30

4.4.3.1 Battery Energy Storage System of the PODCAST project

The BESS controlled by the algorithm are two systems of 30 and 70 kWh nominal energy and of 30 and 70 kW nominal power. Each storage is composed of a string of Toshiba SCiBTM modules [320] (see Figure 4.17).

The modules are mounted each on separate but equal racks (Figure 4.18) inside two twin switchboards, together with all the auxiliary systems, listed in the following:

- one Power center for parallel connection with breaker;
- auxiliary distribution switchboard;

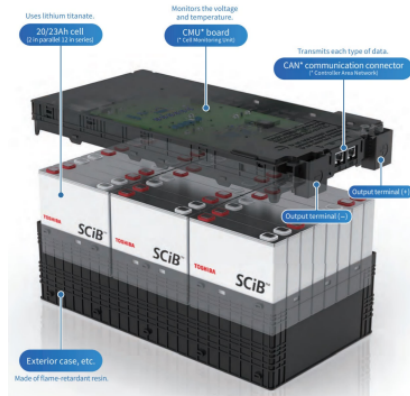


Figure 4.17: Image of the Toshiba SCiB™ Cell [320]

- lighting system;
- cooling system;
- ventilation system.

The two switchboards are placed inside a shelter (see Figures 4.19 and 4.20), which was situated in a rural area that corresponds to a portion of the distribution system which was characterized by a high penetration of PV systems.

The two BESSs are connected with the SCADA-DMS system. There are three ways to control the two systems:

- in person at the shelter (done mainly for maintenance purposes);
- through a remote algorithms server, connected to the DMS via the Open Platform Communications - Unified Architecture (OPC-UA) protocol (see Figure 4.21), wherein the intra-ahead algorithm is implemented.
- via the DMS (see Figures 4.22 and 4.23), wherein the proposed optimization procedure is implemented;

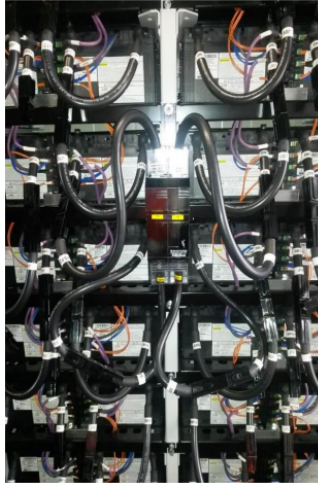


Figure 4.18: Project battery cells mounted on a switchboard



Figure 4.19: Shelter used for hosting the BESSs



Figure 4.20: Internal view of the shelter with the two BESSs

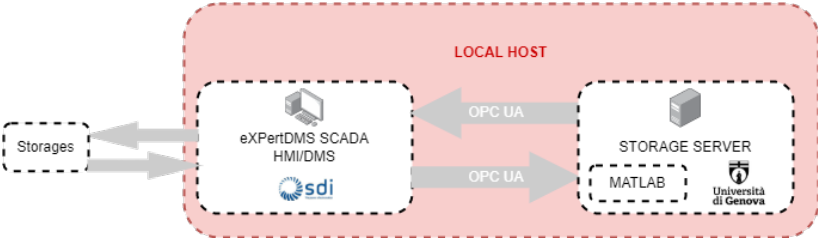


Figure 4.21: Communication scheme for the remote control of the BESSs

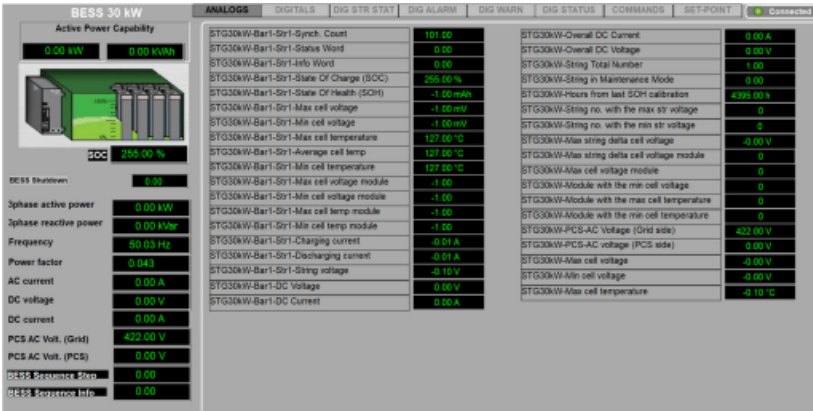


Figure 4.22: DMS interface for monitoring the storage systems

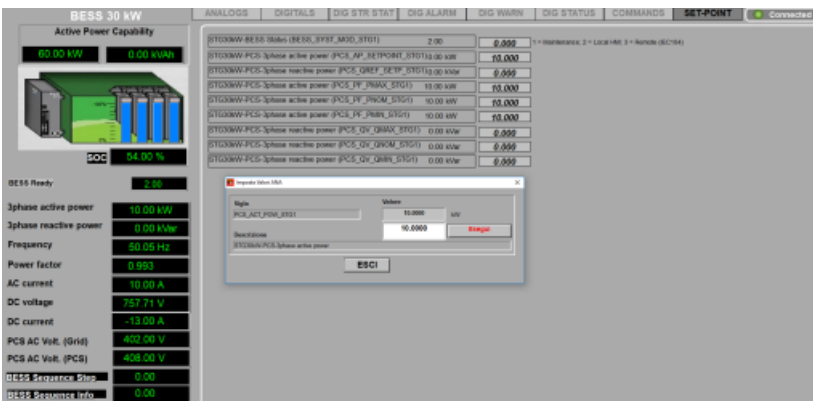


Figure 4.23: Example of controlling remotely the BESSs

4.4.3.2 Chosen Parameters

Several parameters have to be set in order to run the day-ahead stage optimization. $P_s^{max,ch}$, $P_s^{min,ch}$ are set to the respective values of the available BESSs. All the considered parameters are collected in Table 4.4:

Finally, the solver chosen for the optimization is CPLEX [308], called from MATLAB [321] via AMPL environment [319].

4.4.4 Round-Trip Efficiency Estimation

Before presenting the results of the experiments, the round-trip efficiencies of the batteries (namely, η_s^{ch} , η_s^{dsc} for $s = 1, 2$) have been estimated.

They have been computed in several trials, on the 28th of September 2020, where significant setpoints (such as 5, 10, and 15 kW) were set for a single BESS. At the same time, both the given setpoint and the state of charge of the batteries have been noted in order to estimate the relationship between the given setpoint and the actual setpoint realized by the BESSs.

The supposed relationship is linear, with the efficiency being the slope of the linear model:

$$y = \eta \cdot x + \varepsilon \quad (4.29)$$

where:

- y is the effective setpoint (computed from the measured SoC);
- x is the given setpoint;
- η is the battery round-trip efficiency;
- ε is a gaussian error;

The given and the actual setpoints are different for different reasons. Firstly, the given setpoint is a supervisory control setpoint: the battery EMS tries to realize it, and its effective realizations depend on several factors tied to the BESS system state, the grid state, and the EMS settings themselves, which are hard to predict without uncertainties. Hence, round-trip efficiency is modeled by Equation (4.29) in a black-box fashion. Also, two efficiencies are distinguished for each BESS: charging and discharging efficiency. Since it is a ratio between to power setpoints, the efficiencies are dimensionless.

The final values are the ones reported in Table 4.5

Table 4.4: Choice of parameters for the experiments of microgrid/energy community profile optimization

Parameter	Description	Choice	Notes
Δt	Time granularity	0.25 h	15 minutes (standard for the Italian Market)
L	Optimization horizon	24 h	
$\frac{L}{\Delta t}$	Number of Optimization Intervals	96	standard for the Italian Market
S	Number of BESSs	2	
l_{ch}, l_{dsc}	Maximum charging and discharging cycles allowed	1000	
c_{exp}	Costs of buying energy	0.2 €/kWh	Costs where considered constant throughout the day
c_{imp}	Revenue of selling energy	0.1 €/kWh	Costs where considered constant throughout the day
ε	Maximum allowed SoC difference from day-to-day operation	0.1	Corresponding to a maximum SoC difference
$P_{imp}^{max}, P_{exp}^{max}$	Maximum power exchanged by microgrid	70 MW	
SoC_s^{max}, SoC_s^{min}	Anti-degradation limits	10%, 90%	
$\eta_s^{ch}, \eta_n^{dsc}$	Charging and Discharging round-trip efficiencies	Table 4.5	Chosen with algorithm described in Section 4.4.4

Table 4.5: Estimated round-trip efficiencies

Battery	Efficiency	Symbol	Value [p.u.]
30 kW(h)	Charging	η_1^{ch}	0.98
30 kW(h)	Discharging	η_1^{dsc}	0.89
70 kW(h)	Charging	η_2^{ch}	1.04
70 kW(h)	Discharging	η_2^{dsc}	1.00

With the efficiencies, all the parameters have been chosen. Therefore, the experiments could take place.

4.4.4.1 First Experiment

The optimized day-ahead profile for the first experiment is reported in Figure 4.24.

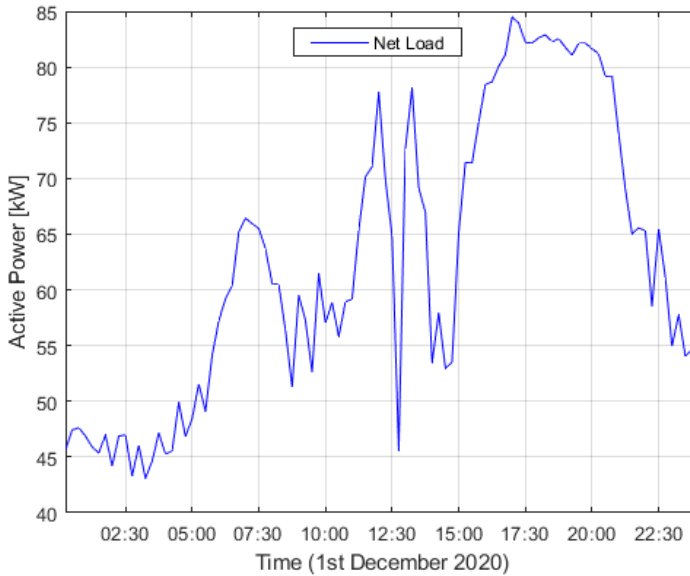


Figure 4.24: Experiment n.1 - Optimized profile in the day-ahead stage load profile

Around noon, the values are particularly volatile because a preliminary experiment was made the day before.

This fact underlines the importance of measuring the uncontrolled load of the aggregate and the controlled devices separately, such as the BESSs, in order to separate the actual demand from the demand modified by actions performed for flexibility, energy efficiency, or other reasons.

In Figure 4.25 it is the output of the two procedures: both the day-ahead stage output (in blue) and the realized intra-day profile (in red).

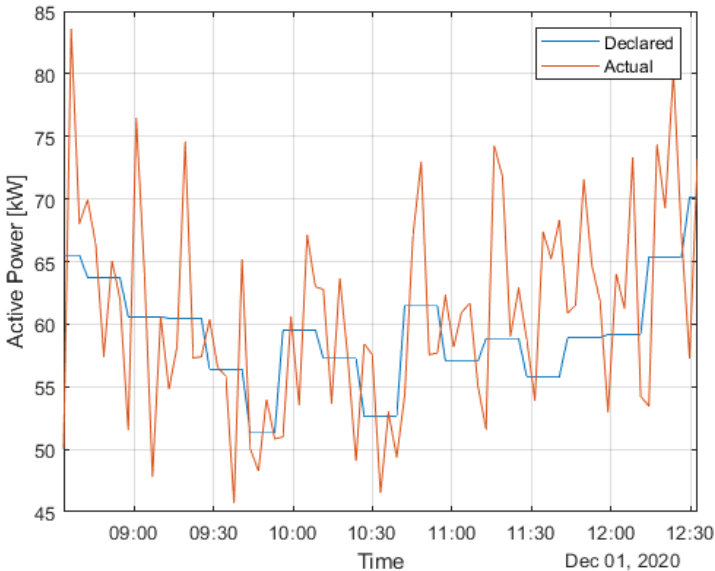


Figure 4.25: Experiment n.1 - Realized load profile

The profile realized by the intra-day procedure is the actual load requested by the microgrid/energy community plus the BESSs actions. The realized profile is sampled every three minutes, while the target profile is constant for each 15-minutes interval. So it is not immediate to see the performance of the profile optimization algorithm: this can be inspected by looking at Figure 4.26, where the equivalent requested and realized energy profile is reported.

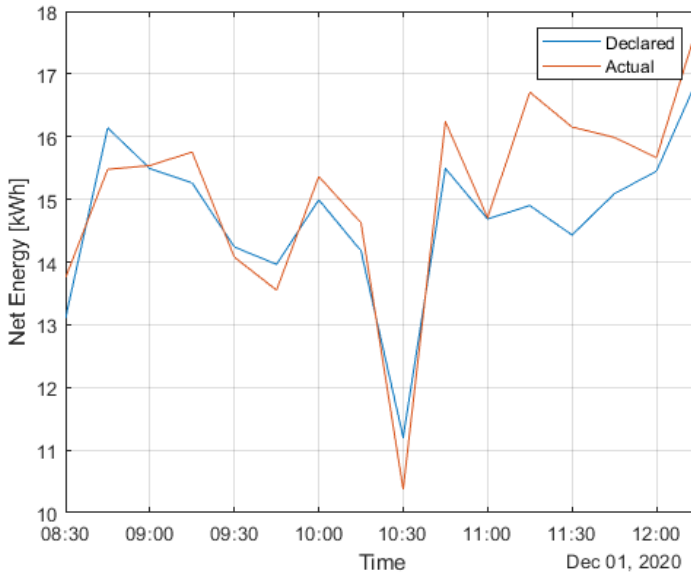


Figure 4.26: Experiment n.1 - Realized energy profile

The overall error committed during the experiment is equal to 0.6 kWh in MAE, a satisfying level, also given the peculiarity of the load around mid-day, as already noted (see Figure 4.24). The error is wider at the end of the experimentation because BESSs could not compensate as well as before the error. Indeed, as Figures 4.27 and 4.28 testify, the BESS had to mostly act as generator toward the grid in order to respect the profile. This caused that around 11:30 AM, the SoCs reached the lower limit set by Equations (4.19) and (4.20). Since the auxiliary systems of the BESS are powered by the batteries and never switched off, this means that the SoC decreases naturally. Hence, the BESSs can only operate in charging mode, until the SoC returns to acceptable values for operating in discharging mode (as it happened around noon, see blue lines in Figures 4.25 and 4.26).

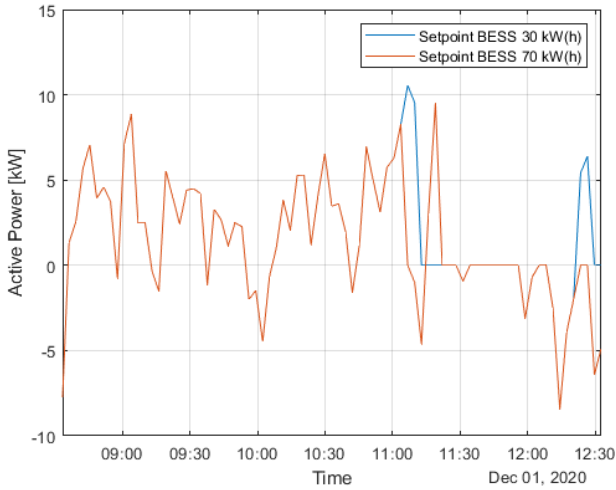


Figure 4.27: Experiment n.2 - Implemented setpoints

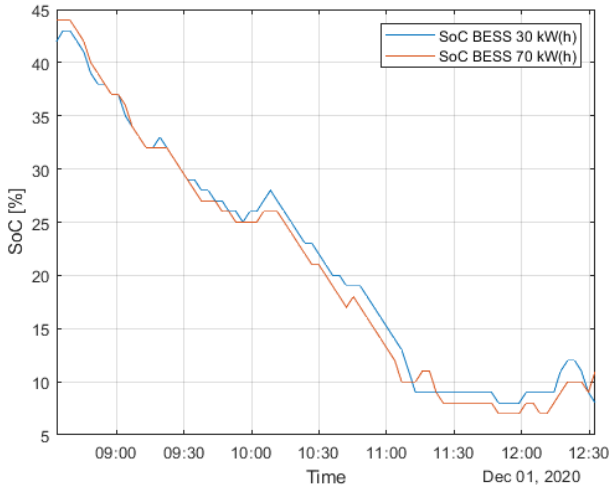


Figure 4.28: Experiment n.2 - Realized energy profile

4.4.4.2 Second Experiment

The day-ahead optimized profile for experiment 2 day is illustrated in Figure 4.29.

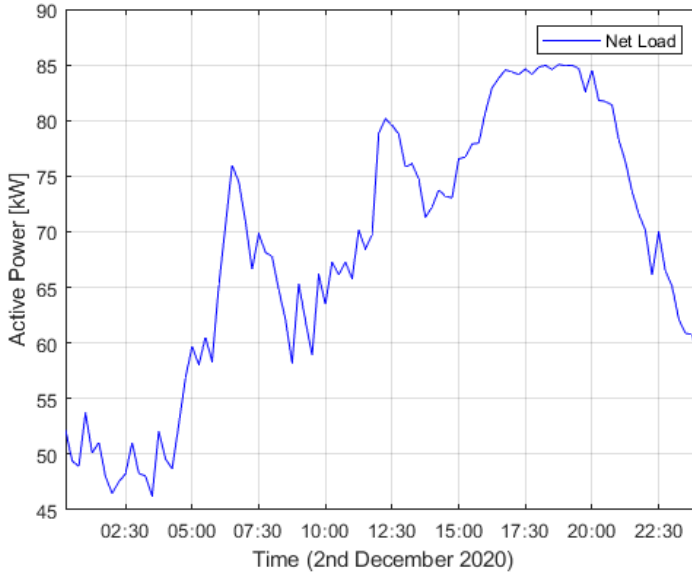


Figure 4.29: Experiment n.2 - Optimized profile in the day-ahead stage

In this case, the experiment was carried out in the afternoon to test a moment of the day with a higher and more constant load. Indeed, in the first experiment (Figure 4.25), the load never exceeded 70 kW, while in the second experiment (Figure 4.30), the load was almost always above 80 kW.

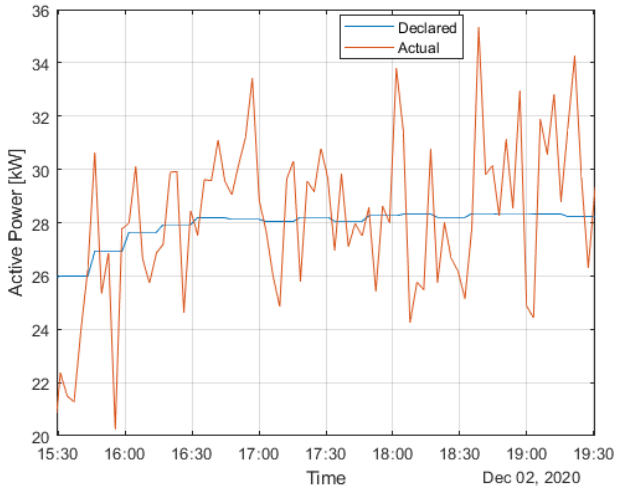


Figure 4.30: Experiment n.2 - Realized load profile

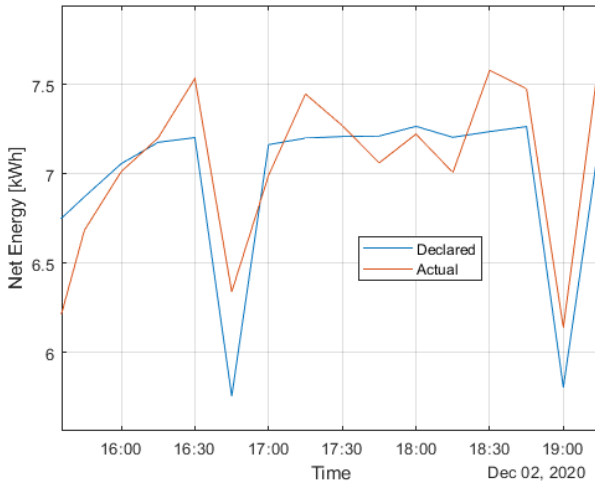


Figure 4.31: Experiment n.2 - Realized energy profile

The declared profile was initially higher than the actual one. This is the reason why the BESS had to raise the total demand by charging, as shown in Figure 4.32

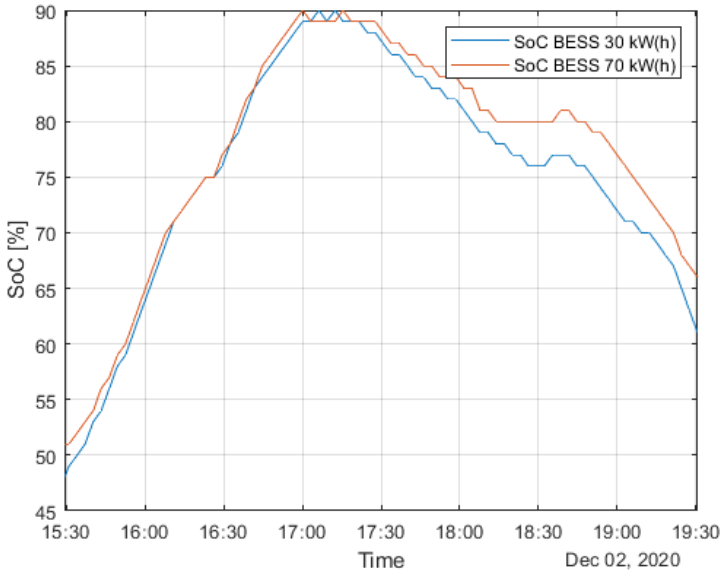


Figure 4.32: Experiment n.2 - Realized energy profile

This has been possible until 5 PM, when the SoC reached the 90% threshold, and the setpoint were set to 0 kW for not violating the Equations (4.19) and (4.20) constraints. Nevertheless, since the BESSs are always switched on, the SoC naturally drops, hence giving the possibility of action in both charge and discharge directions in this case. As a result, the MAE is better in this experiment (0.25 kWh).

After 5:30 PM, BESSs made the reverse regulation by discharging and lowering the demand, ending at 65% of SoC.

In Figure 4.33 the implemented setpoints of both BESS are reported.

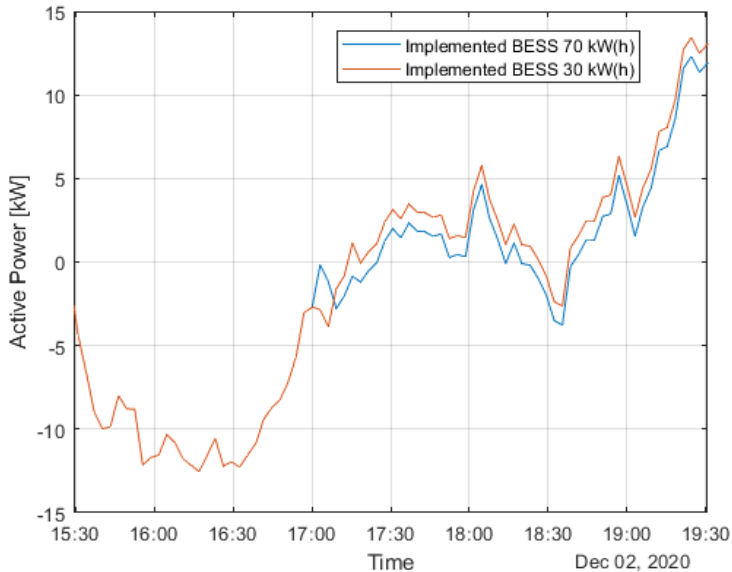


Figure 4.33: Experiment n.2 - Implemented setpoints

They are equal until the 70 kW storage hits the 90% SoC limit, leaving the other BESS to optimize the microgrid/energy community profile. After a brief period of no action or little charges (due to natural discharging of BESSs, the BESSs both contributed to lower the net demand). Note that along all the experiment neither battery reached the 15kW (or -15 kW) of implemented setpoint (see Figures 4.27 and 4.33).

4.5 Chapter Conclusions

In this chapter, an overview of prescriptive analytics was given. It was framed in the context of decision-making process automation as the set of tools that can achieve the highest Levels of Automation (LOAs).

Then, the various practical mathematical techniques to build prescriptive analytics solutions were presented. In particular, deterministic and stochastic

mathematical programming methods were introduced, as well as its intersections with ML tools, and many alternatives (such as data-driven stochastic programming methods, multiparametric programming, genetic algorithms, reinforcement learning, Supervised Learning (SL), simulation, and risk assessment frameworks)

Afterward, the main applications of prescriptive analytics on power systems were reviewed, such as HVAC optimal control, EV smart charging, BESS sizing and siting.

Finally, two applications developed during the Ph.D. are presented. The first [61] is about an algorithm that aims to solve stochastic planning, including DER regulation and sizing and siting of storage devices. Priority in reactive compensation was indeed given to DGs contributions, in accordance with modern network standards, whereas the contributions of BESS were minimized while solving voltage problems and congestions.

The application showed the potential of network planning also considering DER operation, an essential step towards smarter distribution networks. The developed tool can be integrated into the DMS within the PODCAST project in order to exploit the use of historical data coming from smart metering. Probabilistic planning was based on daily Monte Carlo simulations, extracting load and RES profiles based on a modeling technique based on a gaussianity assumption for the quarterly load values.

The presented formulation resulted in being robust and effective, guaranteeing the respect of grid constraints at any time and on all the components. The accuracy of the proposed model has been proven to be far acceptable for the tasks, with negligible percentage errors on voltage magnitude calculation, gaining the advantages of linearity in comparison with other techniques.

Future developments of this work can include reconfiguration, a more detailed cost function of BESS devices (adding, for example, installation costs), and the minimization of network losses, together with the voltage deviations from nominal values. Furthermore, it will be interesting to perform a second-iteration optimization on a unique full-time simulation, considering only the identified substations for BESS installation in order to refine the obtained solution.

Also, looser assumptions about the load profile can be thought (perhaps exploiting the IGSC algorithm, see Section 2.2), and more tools on how to select the BESS location can be added, for example, by integrating the risk for which the BESS is protecting the DSO: this could lead to lower prescribed BESS capacity, with negligible impact on security constraints.

Finally, a deterministic two-stage microgrid or energy community optimization algorithm with two BESS was presented, composed of day-ahead optimization and intra-day realization. The application was developed within PODCAST project with good results on two experiments of 4 hours each.

Some developments are needed for this application to be applied constantly on the distribution grid for the whole day. In particular, separating of BESS contributions to the load from the load itself can enhance the forecasting algorithm performance. Also, investigating stochastic optimization techniques for the day-ahead stage and incorporating very short-term forecasts in the intra-day stage can further help the successful application of the same algorithm in other contexts.

CHAPTER 5

Conclusions

In this thesis, the world of analytics and power systems were presented in conjunction to see how the first can help the second meet its epochal challenges.

In the beginning, a brief history and the current state of power systems and data analytics are presented, and an outline of the subsequent work is given. The power system is under pressure to transition from its fossil-fuel-based generation to one where Renewable Energy Sources (RESs) are the prominent source of energy. This shift is to happen in a relatively short amount of time for mitigating the adverse effects of climate change, for which the energy system is one of the main responsible. In this context, data analytics, which has already revolutionized many aspects of society, science, and technology, is helping modern power grids to be more proactive and sustainable by leveraging all the possible forms of flexibility to address the variability and stochasticity of RESs.

The rest of the thesis breaks down power system analytics into four components: Descriptive, Diagnostic, Predictive, and Prescriptive Analytics.

The most basic forms of analytics are firstly presented: Descriptive (“What happened to the system?”) and Diagnostic analytics (“Why it happened?”). Visualization, simulation, and explainability were introduced, together with some researches that leveraged these tools to give more insights to power system agents. The Instantaneous Growing Stream Clustering (IGSC) algorithm, developed during the Ph.D., was presented in this context. It consists of a technique for constructing a Discrete Markov Chain (DMC) from sequential (possibly streaming) data, requiring low computing capability and memory, hence capable of being applied to Internet of Things (IoT) devices. The states of the DMC are defined as clusters centroids of the data, hence having the same exact dimension of the data and representing a state in the feature space. Hence, the algorithm can be both used as a simulation tool and as a visualization tool. The technique was presented in [52] and further analyzed in [53].

Then, Predictive analytics (“What will happen to the system?”) was introduced, using the common CRISP-DM framework, analyzing its components, namely: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Particular focus was put on forecasting as a crucial subsegment of Predictive analytics for power systems. Afterward, a review of Predictive analytics applications in power systems was made, with a particular focus on load and Photovoltaic (PV) forecasting. Three applications were presented: PV forecasting and distribution network load forecasting, drawn from work made within the PODCAST project, and non-commercial building energy forecasting, drawn from the PREDICT project. The PV forecasting technique devises an innovative hybrid procedure, integrating neural networks and Clear Sky Models (CSMs) via a decision rule, was published in [56] and a simplified version in [57]. The building energy forecasting technique was published in [60] and leveraged deep transfer learning on thermal images and text mining techniques for occupation estimation and forecasting.

Lastly, Prescriptive analytics (“What should happen to the system?”) was presented. Mathematical programming, its main technique, was overviewed in its deterministic and stochastic forms. Also, other methods, especially the ones leveraging Machine Learning (ML) were reviewed. Two applications drawn from the PODCAST project were presented. The first involved

Volt/Var optimization including new Battery Energy Storage System (BESS) optimal sizing and siting for a Distribution System Operator (DSO), and was published in [61]. The second, a two-stage deterministic programming technique for microgrid/energy community profile optimization, leveraging two BESSs installed in the distribution grid for the project.

The above work shows that neither tools nor capabilities are missing from the analytics side to meet the future power system's uncertain challenges. Nor is it the case that power system professionals are not aware of the great possibilities of analytics. All seems set for analytics revolution to in power systems.

Nevertheless, as noted in Glassman J., Shao B., St Louis R., "*Don't get the cart before the horse: There are no shortcuts to Prescriptive analytics*" [143], where several C-level professionals of prominent fields were interviewed on the topic of Prescriptive analytics, one cannot jump directly into Prescriptive analytics, without addressing several obstacles in its way.

Data quality is one of them (such as right granularity, integration, and accuracy), but also cultural changes are needed.

For example, shared taxonomies are needed: it is challenging to communicate when different names refer to different concepts, which, in turn, may affect data quality.

Support from decision-makers is also crucial. Also, using the correct tools for the correct problems is very important. Not necessarily more complicated tools translate better results (which is sometimes the case for Prescriptive analytics) [143].

While this is true in general, in the case of power systems, at least two aspects are very auspicious in that regard. First, there is all the support one can get from society and policy-makers in realizing the energy transition. Secondly, prescriptive analytics has long been present in this field, as the power system is one of the most automated systems globally (despite being one of the most complex). Moreover, Predictive analytics, which is crucial in unlocking prescriptive analytics potential, is one of the most mature analytics.

This maturity can also overflow into other related energy-intensive sectors like the naval, industrial, and building sectors, among others.

In that respect, this thesis has tried to give some structure to many tools and names that make up analytics, with the help of some concrete applications, with the hope to be useful for other researchers to also contribute to power systems with data analytics.

Bibliography

- [1] G. B. Zorzoli, *I due volti del mercato elettrico. Storia Tecnologie e Liberalizzazione del settore elettrico in Italia*. Cappelli Identity Design srl, 2011.
- [2] K. G. Høyer, “The history of alternative fuels in transportation: The case of electric and hybrid cars,” *Utilities Policy*, vol. 16, no. 2, pp. 63–71, 2008. Sustainable Energy and Transportation Systems.
- [3] K. Johansen, “Blowing in the wind: A brief history of wind energy and wind power technologies in denmark,” *Energy Policy*, vol. 152, p. 112139, 2021.
- [4] C. Binz, T. Tang, and J. Huenteler, “Spatial lifecycles of cleantech industries – the global development history of solar photovoltaics,” *Energy Policy*, vol. 101, pp. 386–402, 2017.
- [5] D. Swift-Hook, “2.03 - history of wind power,” in *Comprehensive Renewable Energy* (A. Sayigh, ed.), pp. 41–72, Oxford: Elsevier, 2012.
- [6] A. S. Bahaj, “Generating electricity from the oceans,” *Renewable and Sustainable Energy Reviews*, vol. 15, no. 7, pp. 3399–3416, 2011.

- [7] H. Le Treut, R. Somerville, U. Cubasch, Y. Ding, C. Mauritzen, T. Mokssit, A. Peterson, M. Prather, Editors: S. Solomon, , D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller, “Historical overview of climate change. in: Climate change 2007: The physical science basis. contribution of working group i to the fourth assessment report of the intergovernmental panel on climate change,” tech. rep., IPCC, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2007.
- [8] IEA, “World energy outlook 2021,” tech. rep., IEA, Paris, 2021.
- [9] IPCC 2021, Editors: V. Masson-Delmotte, P. Zhai, A. Pirani, S. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou, “Summary for policymakers. in: Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate change,” tech. rep., IPCC, Geneva, Switzerland, 2021.
- [10] IPCC, 2014 Core Writing Team, Editors: R. Pachauri, and L. Meyer, “Climate change 2014: Synthesis report. contribution of working groups i, ii and iii to the fifth assessment report of the intergovernmental panel on climate change,” Tech. Rep. 151pp., IPCC, Geneva, Switzerland, 2014.
- [11] V. Eyring, S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, “Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization,” *Geoscientific Model Development*, vol. 9, no. 5, pp. 1937–1958, 2016.
- [12] I. Gunnarsdottir, B. Davidsdottir, E. Worrell, and S. Sigurgeirsdottir, “Sustainable energy development: History of the concept and emerging themes,” *Renewable and Sustainable Energy Reviews*, vol. 141, p. 110770, 2021.
- [13] B. Davidsdottir, “Sustainable energy development: The role of geothermal power,” in *Comprehensive Renewable Energy* (A. Sayigh, ed.), pp. 273–297, Oxford: Elsevier, 2012.

- [14] IRENA, “Renewable capacity statistics 2021,” 2021. The International Renewable Energy Agency, Abu Dhabi.
- [15] IRENA, “Renewable energy highlights,” 2021. The International Renewable Energy Agency, Abu Dhabi.
- [16] IRENA, “Innovation landscape brief: Artificial intelligence and big data,” 2019. The International Renewable Energy Agency, Abu Dhabi.
- [17] K. S. Ratnam, K. Palanisamy, and G. Yang, “Future low-inertia power systems: Requirements, issues, and solutions - a review,” *Renewable and Sustainable Energy Reviews*, vol. 124, p. 109773, 2020.
- [18] D. Chiaroni, F. Frattini, A. Di Lieto, C. Troglio, P. Boccardo, and M. Bagnacavalli, “Renewable energy report,” tech. rep., Energy & Strategy Group, sep 2021.
- [19] S. Bianchi, A. De Filippo, S. Magnani, G. Mosaico, and F. Silvestro, “Virtus project: A scalable aggregation platform for the intelligent virtual management of distributed energy resources,” *Energies*, vol. 14, p. 3663, Jun 2021.
- [20] Y. Zhang, T. Huang, and E. F. Bompard, “Big data analytics in smart grids: a review,” *Energy Informatics*, vol. 1, p. 8, Aug 2018.
- [21] C. Kacfeh Emani, N. Cullot, and C. Nicolle, “Understandable big data: A survey,” *Computer Science Review*, vol. 17, pp. 70–81, 2015.
- [22] W. A. Günther, M. H. Rezazade Mehrizi, M. Huysman, and F. Feldberg, “Debating big data: A literature review on realizing value from big data,” *The Journal of Strategic Information Systems*, vol. 26, no. 3, pp. 191–209, 2017.
- [23] H. Akhavan-Hejazi and H. Mohsenian-Rad, “Power systems big data analytics: An assessment of paradigm shift barriers and prospects,” *Energy Reports*, vol. 4, pp. 91–100, 2018.
- [24] “Apache hadoop.” [Online]. Available: <https://hadoop.apache.org>. Accessed: 2022-01-08.

- [25] S. Sagioglu, R. Terzi, Y. Canbay, and I. Colak, “Big data issues in smart grid systems,” in *2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA)*, pp. 1007–1012, 2016.
- [26] J. Lawrence, J. Renshaw, S. Seachman, M. Vernier, L. Min, M. Schwartz, and K. Baker, “Five artificial intelligence grand challenges for the electric power industry,” tech. rep., Electric Power Research Institute (EPRI), sep 2021.
- [27] C. Zhang and Y. Lu, “Study on artificial intelligence: The state of the art and future prospects,” *Journal of Industrial Information Integration*, vol. 23, p. 100224, 2021.
- [28] P. Marquis, O. Papini, and H. Prade, *Elements for a History of Artificial Intelligence*, pp. 1–43. Cham: Springer International Publishing, 2020.
- [29] H. L. van der Maas, L. Snoek, and C. E. Stevenson, “How much intelligence is there in artificial intelligence? a 2020 update,” *Intelligence*, vol. 87, p. 101548, 2021.
- [30] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, B. C. V. Essen, A. A. S. Awwal, and V. K. Asari, “The history began from alexnet: A comprehensive survey on deep learning approaches,” *CoRR*, vol. abs/1803.01164, 2018.
- [31] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, “Language models are few-shot learners,” in *Advances in Neural Information Processing Systems* (H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, eds.), vol. 33, pp. 1877–1901, Curran Associates, Inc., 2020.
- [32] P. E. Hart, “An artificial intelligence odyssey: From the research lab to the real world,” *IEEE Annals of the History of Computing*, pp. 1–1, 2021.

- [33] A. L. Friedman, *Taylorism*, pp. 13519–13520. London: Palgrave Macmillan UK, 2018.
- [34] N. S. Nise, *Control Systems Engineering - 6th Edition*, ch. 1.2. John Wiley and Sons, 2011.
- [35] Smart Grid Coordination Group, “Smart grid reference architecture,” tech. rep., CEN-CENELEC-ETSI, sep 2012.
- [36] Smart Grids European Technology Platform, “Smartgrids—strategic deployment document for european electricity networks of the future.” 2010.
- [37] J. D. McDonald, “A holistic approach to becoming a data-driven utility,” in *Big Data Application in Power Systems* (R. Arghandeh and Y. Zhou, eds.), pp. 3–27, Elsevier, 2018.
- [38] L. Joppa and C. Herweijer, “How ai can enable a sustainable future,” tech. rep., Microsoft Co., 2015.
- [39] IEA, “Energy efficiency 2019,” tech. rep., IEA, Paris, 2019.
- [40] M. Uslar, M. Specht, C. Dänekas, J. Trefke, G. J. M. Rohjans, Sebastian, C. Rosinger, and R. Bleiker, *Standardization in Smart Grids*. Springer, 2013.
- [41] L. Adua, “Reviewing the complexity of energy behavior: Technologies, analytical traditions, and household energy consumption data in the united states,” *Energy Research Social Science*, vol. 59, p. 101289, 2020.
- [42] Global e-Sustainability Initiative (GeSI) and Accenture Strategy, “Smarter2030 - ict solutions for 21st century challenges,” tech. rep., GeSI, 2015.
- [43] T. Ahmad, H. Zhang, and B. Yan, “A review on renewable energy and electricity requirement forecasting models for smart grid and buildings,” *Sustainable Cities and Society*, vol. 55, p. 102052, 2020.
- [44] I. K. Nti, M. Teimeh, O. Nyarko-Boateng, and A. F. Adekoya, “Electricity load forecasting: a systematic review,” *Journal of Electrical Systems and Information Technology*, vol. 7, p. 13, Sep 2020.

- [45] Y. Wang, Q. Chen, C. Kang, M. Zhang, K. Wang, and Y. Zhao, "Load profiling and its application to demand response: A review," *Tsinghua Science and Technology*, vol. 20, no. 2, pp. 117–129, 2015.
- [46] F. Conte, B. Gabriele, G.-P. Schiapparelli, F. Silvestro, C. Bossi, and M. Cabiati, "Optimal positioning of pmus for fault detection and localization in active distribution networks," in *2021 IEEE Madrid PowerTech*, pp. 1–6, 2021.
- [47] H. Wang and Z. Li, "A review of power system transient stability analysis and assessment," in *2019 Prognostics and System Health Management Conference (PHM-Qingdao)*, pp. 1–6, 2019.
- [48] F. Adinolfi, F. D'Agostino, M. Massucco, A. Morini, M. Saviozzi, and F. Silvestro, "Pseudo-measurement modeling using neural network and fourier decomposition for distribution state estimation," *Innovative Smart Grid Technologies Conference (ISGT), Istanbul, 12–15 October 2014*, pp. 1–6.
- [49] G. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.
- [50] J. Kelly and W. Knottenbelt, "Neural nilm: Deep neural networks applied to energy disaggregation," in *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, BuildSys '15*, (New York, NY, USA), p. 55–64, Association for Computing Machinery, 2015.
- [51] B. H. Chandler, N. and N. G. Herschel, "Gartner's business analytics framework: Gartner report," tech. rep., Gartner Inc., Stamford (US-CT), 2011.
- [52] S. Massucco, G. Mosaico, M. Saviozzi, F. Silvestro, A. Fidigatti, and E. Ragaini, "An instantaneous growing stream clustering algorithm for probabilistic load modeling/profiling," in *2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pp. 1–6, 2020.
- [53] S. Massucco, G. Mosaico, M. Saviozzi, F. Silvestro, A. Fidigatti, and E. Ragaini, "A markov chain load modeling approach through a stream

- clustering algorithm,” in *2020 AEIT International Annual Conference (AEIT)*, pp. 1–6, 2020.
- [54] R. Wirth and J. Hipp, “Crisp-dm: Towards a standard process model for data mining.,” in *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, p. 29–39, 2000.
- [55] C. Schröer, F. Kruse, and J. Marx Gómez, “A systematic literature review on applying crisp-dm process model,” *Procedia Computer Science*, vol. 181, pp. 526–534, 2021.
- [56] S. Massucco, G. Mosaico, M. Saviozzi, and F. Silvestro, “A hybrid technique for day-ahead pv generation forecasting using clear-sky models or ensemble of artificial neural networks according to a decision tree approach,” *Energies*, vol. 12, p. 1298, Apr 2019.
- [57] G. Mosaico and M. Saviozzi, “A hybrid methodology for the day-ahead pv forecasting exploiting a clear sky model or artificial neural networks,” in *IEEE EUROCON 2019 -18th International Conference on Smart Technologies*, pp. 1–6, 2019.
- [58] M. Saviozzi, S. Massucco, and F. Silvestro, “Implementation of advanced functionalities for distribution management systems: Load forecasting and modeling through artificial neural networks ensembles,” *Electric Power Systems Research*, vol. 167, pp. 230–239, 2019.
- [59] M. Crosa, F. D’Agostino, S. Massucco, P. Pongiglione, M. Saviozzi, and F. Silvestro, “The Podcast Project: An Application of Volt/Var Optimization to the Electric Distribution Network of Sanremo (Italy),” in *2019 1st International Conference on Energy Transition in the Mediterranean Area (SyNERGY MED)*, pp. 1–6, May 2019.
- [60] G. Mosaico, M. Saviozzi, F. Silvestro, A. Bagnasco, and A. Vinci, “Simplified state space building energy model and transfer learning based occupancy estimation for hvac optimal control,” in *2019 IEEE 5th International forum on Research and Technology for Society and Industry (RTSI)*, pp. 353–358, 2019.

- [61] S. Massucco, G. Mosaico, P. Pongiglione, M. Saviozzi, and F. Silvestro, “Probabilistic planning for distribution networks including optimal der regulation and storage allocation,” in *2020 IEEE Power Energy Society General Meeting (PESGM)*, pp. 1–5, 2020.
- [62] Y. Lu, R. Garcia, B. Hansen, M. Gleicher, and R. Maciejewski, “The state-of-the-art in predictive visual analytics,” *Computer Graphics Forum*, vol. 36, no. 3, pp. 539–562, 2017.
- [63] M. Friendly, “A brief history of data visualization,” in *Handbook of Computational Statistics: Data Visualization*, vol. III, Heidelberg: Springer-Verlag, 2006.
- [64] J. W. Tukey, “The Future of Data Analysis,” *The Annals of Mathematical Statistics*, vol. 33, no. 1, pp. 1 – 67, 1962.
- [65] Y. Dodge, *Exploratory Data Analysis*, pp. 192–194. New York, NY: Springer New York, 2008.
- [66] M. Angelini, G. Santucci, H. Schumann, and H.-J. Schulz, “A review and characterization of progressive visual analytics,” *Informatics*, vol. 5, no. 3, 2018.
- [67] A. D. Nimbarte, N. Smith, and B. Gopalakrishnan, “Human factors evaluation of energy visualization dashboards,” *Ergonomics in Design*, 2021.
- [68] Mathworks, Inc., “Statistics and machine learning toolbox,” 2021.
- [69] J. L. Cremer, I. Konstantelos, and G. Strbac, “From optimization-based machine learning to interpretable security rules for operation,” *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 3826–3836, 2019.
- [70] T. Miller, “Explanation in artificial intelligence: Insights from the social sciences,” *Artificial Intelligence*, vol. 267, pp. 1–38, 2019.
- [71] C. Molnar, *Interpretable Machine Learning*. 2019.
- [72] J. Pearl, “Causal inference in statistics: An overview,” *Statistics Surveys*, vol. 3, no. none, pp. 96 – 146, 2009.

- [73] H. Jia and A. Chong, “eplusr: A framework for integrating building energy simulation and data-driven analytics,” *Energy and Buildings*, vol. 237, p. 110757, 2021.
- [74] European Technology Innovation Platforms – Smart Networks for Energy Transition (ETIP SNET) WG4 and ETIP SNET TF3, “Digitalization of the energy system and customer participation,” tech. rep., ETIP SNET, 2018.
- [75] C.-C. Sun, A. Hahn, and C.-C. Liu, “Cyber security of a power grid: State-of-the-art,” *International Journal of Electrical Power Energy Systems*, vol. 99, pp. 45–56, 2018.
- [76] T. Cioara, I. Anghel, M. Antal, I. Salomie, C. Antal, and A. G. Ioan, “An overview of digital twins application domains in smart energy grid,” 2021. arXiv preprint 2104.07904.
- [77] G. Shao, S.-J. Shin, and S. Jain, “Data analytics using simulation for smart manufacturing,” in *Proceedings of the Winter Simulation Conference 2014*, pp. 2192–2203, 2014.
- [78] J. Lan, Q. Guo, and H. Sun, “Demand side data generating based on conditional generative adversarial networks,” *Energy Procedia*, vol. 152, pp. 1188–1193, 2018.
- [79] Y. Gu, Q. Chen, K. Liu, L. Xie, and C. Kang, “Gan-based model for residential load generation considering typical consumption patterns,” in *2019 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5, 2019.
- [80] Z. Pan, J. Wang, W. Liao, H. Chen, D. Yuan, W. Zhu, X. Fang, and Z. Zhu, “Data-driven ev load profiles generation using a variational auto-encoder,” *Energies*, vol. 12, no. 5, 2019.
- [81] D. B. Crawley, L. K. Lawrie, F. C. Winkelmann, W. Buhl, Y. Huang, C. O. Pedersen, R. K. Strand, R. J. Liesen, D. E. Fisher, M. J. Witte, and J. Glazer, “Energyplus: creating a new-generation building energy simulation program,” *Energy and Buildings*, vol. 33, no. 4, pp. 319–331, 2001. Special Issue: BUILDING SIMULATION’99.

- [82] B.-M. Hodge, D. Lew, and M. Milligan, “Short-term load forecast error distributions and implications for renewable integration studies,” in *2013 IEEE Green Technologies Conference (GreenTech)*, pp. 435–442, 2013.
- [83] J. de Hoog and K. Abdulla, “Data visualization and forecast combination for probabilistic load forecasting in gefcom2017 final match,” *International Journal of Forecasting*, vol. 35, no. 4, pp. 1451–1459, 2019.
- [84] T. Hong, J. Xie, and J. Black, “Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting,” *International Journal of Forecasting*, vol. 35, no. 4, pp. 1389–1399, 2019.
- [85] M. G. De Giorgi, P. M. Congedo, M. Malvoni, and D. Laforgia, “Error analysis of hybrid photovoltaic power forecasting models: A case study of mediterranean climate,” *Energy Conversion and Management*, vol. 100, pp. 117–130, 2015.
- [86] Y. Wang, Q. Chen, and C. Kang, *Overview of Smart Meter Data Analytics*, pp. 1–35. Springer Singapore, 2020.
- [87] R. A. Sevlian, J. Yu, Y. Liao, X. Chen, Y. Weng, E. C. Kara, M. Tabone, S. Badri, C.-W. Tan, D. Chassin, S. Kiliccote, and R. Rajagopal, “Vader: Visualization and analytics for distributed energy resources,” 2017.
- [88] R. J. Hyndman, X. Liu, and P. Pinson, “Visualizing big energy data: Solutions for this crucial component of data analysis,” *IEEE Power and Energy Magazine*, vol. 16, no. 3, pp. 18–25, 2018.
- [89] Commission for Energy Regulation (CER), “Cer smart metering project - electricity customer behaviour trial, 2009-2010 [dataset].” 1st Edition. Irish Social Science Data Archive. SN: 0012-00. <https://www.ucd.ie/issda/data/commissionforenergyregulationcer/>.
- [90] N. Smith, A. D. Nimbarte, B. Gopalakrishnan, and T. Wuest, “Real-time energy visualization system for light commercial businesses,” *Sustainable Energy Technologies and Assessments*, vol. 34, pp. 68–76, 2019.

- [91] “Temporal knowledge discovery in big bas data for building energy management,” *Energy and Buildings*, vol. 109, pp. 75–89, 2015.
- [92] B. Diong, G. Zheng, and M. Ginn, “Establishing the foundation for energy management on university campuses via data analytics,” in *South-eastCon 2015*, pp. 1–7, 2015.
- [93] A. Cunningham, J. D. Hart, U. Engelke, M. Adcock, and B. H. Thomas, “Towards embodied interaction for geospatial energy sector analytics in immersive environments,” in *Proceedings of the Twelfth ACM International Conference on Future Energy Systems, e-Energy ’21*, (New York, NY, USA), p. 396–400, Association for Computing Machinery, 2021.
- [94] M. Cordeil, A. Cunningham, T. Dwyer, B. H. Thomas, and K. Marriott, “Imaxes: Immersive axes as embodied affordances for interactive multivariate data visualisation,” in *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology, UIST ’17*, (New York, NY, USA), p. 71–83, Association for Computing Machinery, 2017.
- [95] D. Aggarwal, E. Curry, and K. C. Davis, “Employing virtual power analytics and linked data for enterprise it energy informatics,” in *Proceedings of Semantic Web Information Management on Semantic Web Information Management, SWIM’14*, (New York, NY, USA), p. 1–4, Association for Computing Machinery, 2014.
- [96] R. Chiosa, M. S. Piscitelli, and A. Capozzoli, “A data analytics-based energy information system (eis) tool to perform meter-level anomaly detection and diagnosis in buildings,” *Energies*, vol. 14, no. 1, 2021.
- [97] Y. Wang, Q. Chen, T. Hong, and C. Kang, “Review of smart meter data analytics: Applications, methodologies, and challenges,” *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3125–3148, 2019.
- [98] J. Hu and A. V. Vasilakos, “Energy big data analytics and security: Challenges and opportunities,” *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2423–2436, 2016.

- [99] Yi Wang, Q. Chen, C. Kang, M. Zhang, K. Wang, and Y. Zhao, “Load profiling and its application to demand response: A review,” *Tsinghua Science and Technology*, vol. 20, no. 2, pp. 117–129, 2015.
- [100] A. Arif, Z. Wang, J. Wang, B. Mather, H. Bashualdo, and D. Zhao, “Load modeling—a review,” *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 5986–5999, 2018.
- [101] G. L. Ray and P. Pinson, “Online adaptive clustering algorithm for load profiling,” *Sustainable Energy, Grids and Networks*, vol. 17, 2019.
- [102] S. S. Hosseini, S. Kelouwani, K. Agbossou, A. Cardenas, and N. Henao, “Adaptive on-line unsupervised appliance modeling for autonomous household database construction,” *International Journal of Electrical Power & Energy Systems*, vol. 112, pp. 156 – 168, 2019.
- [103] I. Benítez, A. Quijano, J.-L. Díez, and I. Delgado, “Dynamic clustering segmentation applied to load profiles of energy consumption from spanish customers,” *International Journal of Electrical Power & Energy Systems*, vol. 55, pp. 437 – 448, 2014.
- [104] I. Benítez, J.-L. Díez, A. Quijano, and I. Delgado, “Dynamic clustering of residential electricity consumption time series data based on hausdorff distance,” *Electric Power Systems Research*, vol. 140, pp. 517 – 526, 2016.
- [105] H. Izakian, W. Pedrycz, and I. Jamal, “Clustering spatiotemporal data: An augmented fuzzy c-means,” *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 5, pp. 855–868, 2013.
- [106] H. Izakian and W. Pedrycz, “Agreement-based fuzzy c-means for clustering data with blocks of features,” *Neurocomputing*, vol. 127, pp. 266 – 280, 2014.
- [107] F. McLoughlin, A. Duffy, and M. Conlon, “A clustering approach to domestic electricity load profile characterisation using smart metering data,” *Applied Energy*, vol. 141, pp. 190 – 199, 2015.
- [108] A. Rouhani and A. Abur, “Real-time dynamic parameter estimation for an exponential dynamic load model,” *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1530–1536, 2016.

- [109] G. M. Gilbert, S. Naiman, H. Kimaro, and B. Bagile, “A critical review of edge and fog computing for smart grid applications,” in *Information and Communication Technologies for Development. Strengthening Southern-Driven Cooperation as a Catalyst for ICT4D* (P. Nielsen and H. C. Kimaro, eds.), pp. 763–775, Springer International Publishing, 2019.
- [110] ABB Group, “Sace emax 2 lv breaker.” [Online] <https://new.abb.com/low-voltage/products/circuit-breakers/emax2>.
- [111] D. Gamerman and H. F. Lopes., *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Champan & Hall/CRC, 2006.
- [112] J. A. Silva, E. R. Faria, R. C. Barros, E. R. Hruschka, A. C. P. L. F. de Carvalho, and J. Gama, “Data stream clustering: A survey,” *ACM Comput. Surv.*, vol. 46, no. 1, 2013.
- [113] J. Jockusch and H. Ritter, “An instantaneous topological mapping model for correlated stimuli,” in *IJCNN’99. International Joint Conference on Neural Networks. Proceedings (Cat. No.99CH36339)*, vol. 1, pp. 529–534 vol.1, July 1999.
- [114] T. Kohonen, “Self-organized formation of topologically correct feature maps,” *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, 1982.
- [115] B. Fritzke, “A growing neural gas network learns topologies,” in *Proceedings of the 7th International Conference on Neural Information Processing Systems*, (Cambridge, MA, USA), p. 625–632, MIT Press, 1994.
- [116] D. Vazquez, T. Freichard, and C. Laugier, “Growing hidden markov models: An incremental tool for learning and predicting human and vehicle motion,” *The International Journal of Robotics Research*, vol. 28(11–12), p. 1486–1506, January 2009.
- [117] P. B. de Mello Martins, V. B. Nascimento, A. R. de Freitas, P. B. e Silva, and R. G. D. Pinto, “Industrial machines dataset for electrical load disaggregation,” *IEEE Dataport*, 2018.
- [118] T. Caliński and J. Harabasz, “A dendrite method for cluster analysis,” *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, 1974.

- [119] D. Arthur and S. Vassilvitskii, “K-means++: The advantages of careful seeding,” *SODA '07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1027–1035, 2007.
- [120] P. Bholowalia and A. Kumar, “Article: Ebk-means: A clustering technique based on elbow method and k-means in wsn,” *International Journal of Computer Applications*, vol. 105, no. 9, pp. 17–24, 2014.
- [121] Y. Zhang, S. Tang, H. Sun, and H. Neumann, “Human motion parsing by hierarchical dynamic clustering,” in *British Machine Vision Conference*, Sept. 2018.
- [122] “Dynamical clustering github repository.” [Online]. Available: https://github.com/yz-cnsdqz/dynamic_clustering. Accessed: 2020-02-06.
- [123] J. T. Prince, “A paradigm for assessing the scope and performance of predictive analytics,” *Information Economics and Policy*, vol. 47, pp. 7–13, 2019.
- [124] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, “Energy forecasting: A review and outlook,” *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 376–388, 2020.
- [125] S. C. Hoi, D. Sahoo, J. Lu, and P. Zhao, “Online learning: A comprehensive survey,” *Neurocomputing*, vol. 459, pp. 249–289, 2021.
- [126] Y. Li, S. Zhang, R. Hu, and N. Lu, “A meta-learning based distribution system load forecasting model selection framework,” *Applied Energy*, vol. 294, p. 116991, 2021.
- [127] T. Hong and S. Fan, “Probabilistic electric load forecasting: A tutorial review,” *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, 2016.
- [128] R. Weron, “Electricity price forecasting: A review of the state-of-the-art with a look into the future,” *International Journal of Forecasting*, vol. 30, no. 4, pp. 1030–1081, 2014.
- [129] T. Hong, “Crystal ball lessons in predictive analytics,” *EnergyBiz*, vol. 12, no. 2, pp. 35–37, 2015.

- [130] M. H. Ramos, S. J. van Andel, and F. Pappenberger, “Do probabilistic forecasts lead to better decisions?,” *Hydrology and Earth System Sciences*, vol. 17, no. 6, pp. 2219–2232, 2013.
- [131] C. Sweeney, R. J. Bessa, J. Browell, and P. Pinson, “The future of forecasting for renewable energy,” *WIREs Energy and Environment*, vol. 9, no. 2, p. e365, 2020.
- [132] N. Amjady and M. Hemmati, “Energy price forecasting - problems and proposals for such predictions,” *IEEE Power and Energy Magazine*, vol. 4, no. 2, pp. 20–29, 2006.
- [133] C. Kath and F. Ziel, “Conformal prediction interval estimation and applications to day-ahead and intraday power markets,” *International Journal of Forecasting*, vol. 37, no. 2, pp. 777–799, 2021.
- [134] I. Kanda and J. Q. Veguillas, “Data preprocessing and quantile regression for probabilistic load forecasting in the gefcom2017 final match,” *International Journal of Forecasting*, vol. 35, no. 4, pp. 1460–1468, 2019.
- [135] J. Zhang, Y. Wang, M. Sun, and N. Zhang, “Two-stage bootstrap sampling for probabilistic load forecasting,” *IEEE Transactions on Engineering Management*, pp. 1–9, 2020.
- [136] Z. Wang, W. Wang, C. Liu, B. Wang, and S. Feng, “Short-term probabilistic forecasting for regional wind power using distance-weighted kernel density estimation,” *IET Renewable Power Generation*, vol. 12, no. 15, pp. 1725–1732, 2018.
- [137] Q. T. Phan, Y. K. Wu, and Q. D. Phan, “A hybrid wind power forecasting model with xgboost, data preprocessing considering different nwps,” *Applied Sciences*, vol. 11, no. 3, 2021.
- [138] H. Panamtaash, Q. Zhou, T. Hong, Z. Qu, and K. O. Davis, “A copula-based bayesian method for probabilistic solar power forecasting,” *Solar Energy*, vol. 196, pp. 336–345, 2020.
- [139] L. Han, Y. Qiao, M. Li, and L. Shi, “Wind power ramp event forecasting based on feature extraction and deep learning,” *Energies*, vol. 13, no. 23, 2020.

- [140] I. Dimoukias, M. Amelin, and M. R. Hesamzadeh, “Forecasting balancing market prices using hidden markov models,” in *2016 13th International Conference on the European Energy Market (EEM)*, pp. 1–5, 2016.
- [141] W. A. Yousef and S. Kundu, “Learning algorithms may perform worse with increasing training set size: Algorithm-data incompatibility,” *Comput. Stat. Data Anal.*, vol. 74, p. 181–197, jun 2014.
- [142] R. D. D. Veaux and D. J. Hand, “How to Lie with Bad Data,” *Statistical Science*, vol. 20, no. 3, pp. 231 – 238, 2005.
- [143] R. S. L. Jeremy Glassman, Benjamin Shao, “Don’t get the cart before the horse: There are no shortcuts to prescriptive analytics,” in *Proceedings of the Annual Hawaii International Conference on System Sciences*, pp. 1055–1064, jan 2019.
- [144] R. Tawn, J. Browell, and I. Dinwoodie, “Missing data in wind farm time series: Properties and effect on forecasts,” *Electric Power Systems Research*, vol. 189, p. 106640, 2020.
- [145] J. Barker, “Machine learning in m4: What makes a good unstructured model?,” *International Journal of Forecasting*, vol. 36, no. 1, pp. 150–155, 2020.
- [146] I. Guyon and A. Elisseeff, *An Introduction to Feature Extraction*, pp. 1–25. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.
- [147] M. Christ, N. Braun, J. Neuffer, and A. W. Kempa-Liehr, “Time series feature extraction on basis of scalable hypothesis tests (tsfresh – a python package),” *Neurocomputing*, vol. 307, pp. 72–77, 2018.
- [148] V. Bolón-Canedo and A. Alonso-Betanzos, “Ensembles for feature selection: A review and future trends,” *Information Fusion*, vol. 52, pp. 1–12, 2019.
- [149] T. Januschowski, J. Gasthaus, Y. Wang, D. Salinas, V. Flunkert, M. Bohlke-Schneider, and L. Callot, “Criteria for classifying forecasting methods,” *International Journal of Forecasting*, vol. 36, no. 1, pp. 167–177, 2020.

- [150] L. Breiman, “Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author),” *Statistical Science*, vol. 16, no. 3, pp. 199–231, 2001.
- [151] M. Rana and I. Koprinska, “Forecasting electricity load with advanced wavelet neural networks,” *Neurocomputing*, vol. 182, pp. 118–132, 2016.
- [152] A. Mellit, A. Massi Pavan, E. Ogliaari, S. Leva, and V. Lughi, “Advanced methods for photovoltaic output power forecasting: A review,” *Applied Sciences*, vol. 10, no. 2, 2020.
- [153] M. Feurer and F. Hutter, *Hyperparameter Optimization*, pp. 3–33. Cham: Springer International Publishing, 2019.
- [154] A. F. Atiya, “Why does forecast combination work so well?,” *International Journal of Forecasting*, vol. 36, no. 1, pp. 197–200, 2020.
- [155] M. Hibon and T. Evgeniou, “To combine or not to combine: selecting among forecasts and their combinations,” *International Journal of Forecasting*, vol. 21, no. 1, pp. 15–24, 2005.
- [156] C. Bergmeir and J. M. Benítez, “On the use of cross-validation for time series predictor evaluation,” *Information Sciences*, vol. 191, pp. 192–213, 2012.
- [157] P. C. Emiliano, M. J. Vivanco, and F. S. de Menezes, “Information criteria: How do they behave in different models?,” *Computational Statistics Data Analysis*, vol. 69, pp. 141–153, 2014.
- [158] S. Kolassa, “Why the “best” point forecast depends on the error or accuracy measure,” *International Journal of Forecasting*, vol. 36, no. 1, pp. 208–211, 2020.
- [159] R. S. Mariano and D. Preve, “Statistical tests for multiple forecast comparison,” *Journal of Econometrics*, vol. 169, no. 1, pp. 123–130, 2012.
- [160] M. Barque, S. Martin, J. E. N. Vianin, D. Genoud, and D. Wannier, “Improving wind power prediction with retraining machine learning algorithms,” in *2018 International Workshop on Big Data and Information Security (IWBIS)*, pp. 43–48, 2018.

- [161] Z. Chen and B. Liu, *Lifelong Machine Learning*, ch. Continual Learning and Catastrophic Forgetting. Morgan & Claypool Publishers, 2018.
- [162] J. Maldonado-Correa, J. Solano, and M. Rojas-Moncayo, “Wind power forecasting: A systematic literature review,” *Wind Engineering*, vol. 45, no. 2, pp. 413–426, 2021.
- [163] F. Adinolfi, F. Baccino, F. D’Agostino, S. Massucco, M. Saviozzi, and F. Silvestro, “An architecture for implementing state estimation application in distribution management system (dms),” in *In Proceedings of the Innovative Smart Grid Technologies (ISGT 2013) Conference*, (Copenhagen), 2013.
- [164] B. Ming, P. Liu, S. Guo, L. Cheng, Y. Zhou, S. Gao, and H. Li, “Robust hydroelectric unit commitment considering integration of large-scale photovoltaic power: A case study in china,” *Applied Energy*, vol. 228, pp. 1341–1352, 2018.
- [165] E. Namor, F. Sossan, R. Cherkaoui, and M. Paolone, “Control of battery storage systems for the simultaneous provision of multiple services,” *IEEE Transactions on Smart Grid*, vol. DOI: 10.1109/TSG.2018.2810781, 2018.
- [166] D. Van Der Meer, G. Chandra Mouli, G. Morales, L. Elizondo, and P. Bauer, “Energy management system with pv power forecast to optimally charge evs at the workplace,” *IEEE Transactions on Industrial Informatics*, vol. 14, pp. 311–320, 2018.
- [167] U. K. Das, K. S. Tey, M. Seyedmahmoudian, S. Mekhilef, M. Y. I. Idris, W. V. Deventer, B. Horan, and A. Stojcevski, “Forecasting of photovoltaic power generation and model optimization: A review,” *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 912–928, 2018.
- [168] M. Kudo, A. Takeuchi, Y. Nozaki, H. Endo, and J. Sumita, “Forecasting electric power generation in a photovoltaic power system for an energy network,” *Electrical Engineering In Japan*, vol. 167, pp. 16–23, 2009.
- [169] S. Pelland, J. Remund, J. Kleissl, T. Oozeki, and K. De Brabandere, “Photovoltaic and solar forecasting: State of the art,” tech. rep., IEA, 2013. Report.

- [170] A. Adinolfi, F. D'Agostino, S. Massucco, M. Saviozzi, and F. Silvestro, "Advanced operational functionalities for a low voltage microgrid test site," in *IEEE Power & Energy Society General Meeting*, (Denver, CO (USA)), 2015.
- [171] F. Conte, S. Massucco, M. Saviozzi, and F. Silvestro, "A stochastic optimization method for planning and real-time control of integrated pv-storage systems: Design and experimental validation," *IEEE Transactions of Sustainable Energy*, vol. 9, pp. 1188–1197, 2018.
- [172] "Arrêté du 8 mars 2013 fixant les conditions d'achat de l'électricité produite par les installations utilisant l'énergie mécanique du vent situées dans des zones particulièrement exposées au risque cyclonique et disposant d'un dispositif de prévision et de lissage de la production." France Electricity Regulation, [Online]. Available: <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000027262791/>.
- [173] E. Ogliari, A. Dolara, G. Manzolini, and S. Leva, "Physical and hybrid methods comparison for the day ahead pv output power forecast," *Renewable Energy*, vol. 113, pp. 11–21, 2017.
- [174] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. Martinez-de Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Journal of Solar energy*, vol. 136, pp. 78–111, 2016.
- [175] M. G. De Giorgi, P. M. Congedo, and M. Malvoni, "Photovoltaic power forecasting using statistical methods: Impact of weather data," *IET Science, Measurement and Technology*, vol. 8, pp. 90–97, 2014.
- [176] L. A. Dao, L. Piroddi, and L. Ferrarini, "Ensemble methods for pv power production prediction," in *International Conference on Clean Electrical Power*, (Santa Margherita Ligure (Italy)), 2017.
- [177] T. Cai, D. Shanxu, and C. Chen, "Forecasting power output for grid-connected photovoltaic power system without using solar radiation measurement," in *International Symposium on Power Electronics for Distributed Generation Systems*, (Hefei (China)), 2010.
- [178] E. Ogliari, A. Gandelli, F. Grimaccia, S. Leva, and M. Mussetta, "Neural forecasting of the day-ahead hourly power curve of a photovoltaic power plant," in *In Proceedings of the International Joint Conference*

- on Neural Networks (IJCNN 2016), Vancouver, BC, Canada, 24–29 July 2016*, pp. 654–659.
- [179] L. Cristaldi, G. Leone, and R. Ottoboni, “A hybrid approach for solar radiation and photovoltaic power short-term forecast,” in *In Proceedings of the IEEE International Instrumentation and Measurement Technology Conference (I2MTC 2017), Turin, Italy, 22–25 May 2017*, pp. 1252–1257.
- [180] A. Asrari, T. X. Wu, and B. Ramos, “A hybrid algorithm for short-term solar power prediction – sunshine state case study,” *IEEE Transactions on Sustainable Energy*, vol. 8, pp. 582–591, 2017.
- [181] J. L. Sánchez-García, E. Espinosa-Juárez, and J. J. Flores, “Short term photovoltaic power production using a hybrid of nearest neighbor and artificial neural networks,” in *IEEE PES Transmission & Distribution Conference and Exposition - Latin America (PES T&D-LA), Morelia (Michoacán, Mexico, 20–24 September 2016*, pp. 1–6.
- [182] T. Hong, “Energy Forecasting: Past, Present, and Future,” *Foresight: The International Journal of Applied Forecasting*, pp. 43–48, Winter 2014.
- [183] Y. K. Semero, J. Zhang, and D. Zheng, “Emd–pso–anfis-based hybrid approach for short-term load forecasting in microgrids,” *IET Generation, Transmission Distribution*, vol. 14, pp. 470–475(5), February 2020.
- [184] C. Kuster, Y. Rezgui, and M. Mourshed, “Electrical load forecasting models: A critical systematic review,” *Sustainable Cities and Society*, vol. 35, pp. 257–270, 2017.
- [185] T. Hong, P. Wang, and L. White, “Weather station selection for electric load forecasting,” *International Journal of Forecasting*, vol. 31, no. 2, pp. 286–295, 2015.
- [186] N. B. Vanting, Z. Ma, and B. N. Jørgensen, “A scoping review of deep neural networks for electric load forecasting,” *Energy Informatics*, vol. 4, p. 49, Sep 2021.

- [187] J. Shi, Y. Liu, and Y. Yang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," *IEEE Transactions on Industry Applications*, vol. 48, pp. 1064–1069, 2012.
- [188] C. M. Huang, S. J. Chen, S. P. Yang, and C. J. Kuo, "One-day-ahead hourly forecasting for photovoltaic power generation using an intelligent method with weather-based forecasting models," *IET Generation Transmission and Distribution*, vol. 9, pp. 1874–1882, 2015.
- [189] M. Rana, I. Koprinska, and V. G. Agelidis, "Solar power forecasting using weather type clustering and ensembles of neural networks," in *In Proceedings of the International Joint Conference on Neural Networks (IJCNN 2016), Vancouver, BC, Canada, 24–29 July 2016*, pp. 4962–4969.
- [190] Z. Li, Z. Lu, Y. Qiao, N. Wang, and K. Ding, "Weather type partition method considering sequential features in photovoltaic forecasting," *The Journal of Engineering*, vol. 13, pp. 1259–1263, 2017.
- [191] Y. Wang, W. Liao, and Y. Chang, "Gated recurrent unit network-based short-term photovoltaic forecasting," *Energies*, vol. 11, no. 2163, 2018.
- [192] F. Wang, Z. Zhen, B. Wang, and Z. Mi, "Comparative study on knn and svm based weather classification models for day ahead short term solar pv power forecasting," *Applied Sciences*, vol. 8, p. 28, 2017.
- [193] C. Coimbra, J. Kleissl, and R. Marquez, *Solar Resource Assessment and Forecasting*, ch. Overview of solar forecasting methods and a metric for accuracy evaluation. Elsevier, 2013. pp. 171–194.
- [194] M. Q. Raza, M. Nadarajah, and C. Ekanayake, "On recent advances in pv output power forecast," *Solar Energy*, vol. 136, pp. 125 – 144, 2016.
- [195] A. Wagner, *Photovoltaik Engineering*. Springer, 2009.
- [196] D. Jordan, S. Kurtz, K. VanSant, and J. Newmiller, "Compendium of photovoltaic degradation rates," *Progress in Photovoltaics: Research and Applications*, vol. 24, pp. 978–989, 2016.

- [197] D. Quansah, M. Adaramola, G. Takyi, and I. Edwin, “Reliability and degradation of solar pv modules—case study of 19-year-old polycrystalline modules in ghana,” *Technologies*, vol. 5, p. 22, 2017.
- [198] V. Quaschnig and R. Hanitsch, “Irradiance calculation on shaded surfaces,” *Solar Energy*, vol. 62, pp. 369 – 375, 1998.
- [199] B. Y. H. Liu and R. C. Jordan, “The interrelationship and characteristic distribution of direct, diffuse and total solar radiation,” *Solar Energy*, vol. 4, pp. 1–9, 1960.
- [200] H. T. Yang, C. M. Huang, Y. C. Huang, and S. P. Yi, “A weather-based hybrid method for 1-day ahead hourly forecasting of pv power output,” *IEEE Transactions on Sustainable Energy*, vol. 5, pp. 917–926, 2014.
- [201] D. King, “Photovoltaic module and array performance characterization methods for all system operating conditions,” tech. rep., Sandia National Laboratories, 1996.
- [202] D. King, J. Kratochvil, W. Boyson, and W. Bower, “Field experience with a new performance characterization procedure for photovoltaic arrays,” tech. rep., Sandia National Laboratories, 1998.
- [203] D. King, W. Boyson, and J. Kratochvil, “Photovoltaic array performance model,” tech. rep., Sandia National Laboratories, 2004.
- [204] E. Skoplaki, A. G. Boudouvis, and J. A. Palyvos, “A simple correlation for the operating temperature of photovoltaic modules of arbitrary mounting,” *Solar Energy Materials & Solar Cells*, vol. 92, pp. 1393–1402, 2008.
- [205] A. M. Muzathik, “Photovoltaic modules operating temperature estimation using a simple correlation,” *International Journal of Energy Engineering*, vol. 4, pp. 151–158, 2014.
- [206] M. Koehl, M. Heck, S. Wiesmeier, and J. Wirth, “Modeling of the nominal operating cell temperature based on outdoor weathering,” *Solar Energy Materials & Solar Cells*, vol. 95, pp. 1638–1646, 2011.
- [207] N. R. Draper and H. Smith, *Applied Regression Analysis*. Wiley–Interscience, 1998.

- [208] M. Ceci, R. Corizzo, F. Fumarola, D. Malerba, and A. Rashkovska, "Predictive modeling of pv energy production: How to set up the learning task for a better prediction?," *IEEE Transactions on Industrial Informatics*, vol. 13, pp. 956–966, 2017.
- [209] S. Mekhilef, R. Saidur, and M. Kamalisarvestani, "Effect of dust, humidity and air velocity on efficiency of photovoltaic cells," *Renewable and Sustainable Energy Reviews*, vol. 16, pp. 2920–2925, 2012.
- [210] X. Zhang, B. Jiang, X. Zhang, F. Fang, Z. Gao, and T. Feng, "Solar photovoltaic power prediction based on similar day approach," in *In Proceedings of the 36th Chinese Control Conference (CCC 2017), Dalian, China, 26–28 July 2017*, pp. 10634–10639.
- [211] A. Bagnasco, M. Saviozzi, F. Silvestro, A. Vinci, S. Grillo, and E. Zenaro, "Artificial neural network application to load forecasting in a large hospital facility," in *In Proceedings of the International Probabilistic Methods Applied to Power Systems Conference (PMAPS 2014), Durham, UK, 7–10 July 2014*, pp. 1–6.
- [212] M. P. Perrone and L. N. Cooper, "When networks disagree: ensemble methods for hybrid neural networks," in *Neural Networks for Speech and Image Processing* (R. J. Mammone, ed.), Chapman-Hall, 1993.
- [213] L. Breiman, J. H. Freidman, R. A. Olsen, and C. J. Stone, *Classification and regression trees*. Wadsworth, 1984.
- [214] X. Wu, V. Kumar, J. Ross Quinlan, J. Gosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, pp. 1–37, 2008.
- [215] "Pvsyst 6.6.4 software." <http://www.pvsyst.com/en/>.
- [216] "Openweathermap inc." [Online]. <https://openweathermap.org/>.
- [217] A. Afram, F. Janabi-Sharifi, A. Fung, and K. Raahemifar, "System identification and data fusion for on-line adaptive energy forecasting in virtual and real commercial buildings," *Energy and Buildings*, vol. 129, pp. 227–237, 2016.

- [218] Y. Zhang, M. Beaudin, R. Taheri, H. Zareipour, and D. Wood, "Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators," *IEEE Transactions on Smart Grid*, vol. 6, pp. 2253–2263, 2015.
- [219] L. Gigoni, A. Betti, E. Crisostomi, A. Franco, M. Tucci, F. Bizzarri, and D. Mucci, "Day-ahead hourly forecasting of power generation from photovoltaic plants," *IEEE Transaction on Sustainable Energy*, vol. 9, pp. 831–842, 2018.
- [220] R. Sifontes, M. Marcano, A. Rojas, J. Rengifo, F. Ochoa, P. De Oliveira, and Corpoelec, "Dms/scada data filtering using neural network tool to mid-term load forecasting," in *2014 IEEE ANDESCON*, pp. 1–1, 2014.
- [221] V. Donde, Z. Wang, F. Yang, and J. Stoupis, "Short-term load forecasting based capacity check for automated power restoration of electric distribution networks," in *IEEE PES T D 2010*, pp. 1–8, 2010.
- [222] E. Bessone, C.A. Nucci, S. Bianchi, M. Pentolini, S. Massucco and F. Raco, "The PODCAST project - optimizing distribution networks with renewable energy sources," *AEIT International Annual Conference, Bari*, Oct 2018.
- [223] "Smartgen project." [Online]. Available: <http://www.smartgen.it/>.
- [224] D. Pohoryles, C. Maduta, D. Bournas, and L. Kouris, "Energy performance of existing residential buildings in europe: A novel approach combining energy with seismic retrofitting," *Energy and Buildings*, vol. 223, p. 110024, 2020.
- [225] S. Wang and Z. Ma, "Supervisory and optimal control of building HVAC systems: A review," *HVAC&R Research*, vol. 14, no. 1, pp. 3–32, 2018.
- [226] M. Killian and M. Kozek, "Ten questions concerning model predictive control for energy efficient buildings," *Building and Environment*, vol. 105, pp. 403 – 412, 2016.
- [227] M. Knudsen, R. Hedegaard, T. Pedersen, and S. Petersen, "System identification of thermal building models for demand response – a

- practical approach,” in *CISBAT International Conference*, pp. 936–939, Lausanne, Switzerland, 6–8 September 2017.
- [228] J. R. Dobbs and B. M. Hency, “Model predictive HVAC control with online occupancy model,” *Energy and Buildings*, vol. 82, pp. 675–684, 2014.
- [229] Z. Liu, J. Zhang, and L. Geng, “An intelligent building occupancy detection system based on sparse auto-encoder,” in *IEEE Winter Applications of Computer Vision Workshops (WACVW)*, pp. 17–22, 2017.
- [230] D. Acharya, W. Yan, and K. Khoshelham, “Real-time image-based parking occupancy detection using deep learning,” in *Research@Locate*, 2018.
- [231] A. Bagnasco, S. Massucco, M. Saviozzi, F. Silvestro, and A. Vinci, “Design and validation of a detailed building thermal model considering occupancy and temperature sensors,” in *IEEE International Forum on Research and Technology for Society and Industry (RTSI)*, Palermo, Italy, 10–13 September 2018.
- [232] “Adaptive energy efficiency platform for consumption reduction in non-residential buildings, official website.” [Online]. <https://http://www.research.softeco.it/predict.aspx/>.
- [233] A. Krizhevsky, I. Sutskever, and G. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Neural Information Processing Systems Conference*, pp. 1097–1105, Lake Tahoe, Nevada, USA, 3–8 December 2012.
- [234] H. Drucker, C. Burges, L. Kaufman, A. Smola, J. Alexander, and V. Vapnik, “Support vector regression machines,” in *Advances in Neural Information Processing Systems Conference*, pp. 155–161, Denver, Colorado, USA, 2–7 December 1996.
- [235] J. Karnowski, “Alexnet + SVM,” 2015. [Online]. Available: <https://jeremykarnowski.files.wordpress.com/2015/07/alexnet2.png>.
- [236] S. Prívvara, Z. Váňa, J. Cigler, and L. Ferkl, “Predictive control oriented subspace identification based on building energy simulation tools,” in *Mediterranean Conference on Control and Automation*, pp. 1290–1295, Barcelona, 3–6 July 2012.

- [237] P. V. Overschee and B. de Moor, "N4sid: Numerical algorithms for state space subspace system identification," *IFAC Proceedings Volumes*, vol. 26, no. 2, Part 5, pp. 55 – 58, 1993.
- [238] "Solcast inc.," [Online]. <https://solcast.com.au/>.
- [239] R. A. Verzijlbergh, P. W. Heijnen, S. R. de Roode, A. Los, and H. J. Jonker, "Improved model output statistics of numerical weather prediction based irradiance forecasts for solar power applications," *Solar Energy*, vol. 118, pp. 634–645, 2015.
- [240] N. Engerer and F. Mills, "Kpv: A clear-sky index for photovoltaics," *Solar Energy*, vol. 105, pp. 679–693, 2014.
- [241] R. Parasuraman, T. Sheridan, and C. Wickens, "A model for types and levels of human interaction with automation," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 30, no. 3, pp. 286–297, 2000.
- [242] M. R. Endsley, "Level of automation: Integrating humans and automated systems," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 41, no. 1, pp. 200–204, 1997.
- [243] A.-S. Ulfert, C. H. Antoni, and T. Ellwart, "The role of agent autonomy in using decision support systems at work," *Computers in Human Behavior*, vol. 126, p. 106987, 2022.
- [244] T. B. Sheridan and W. L. Verplanck, "Human and computer control of undersea teleoperators," tech. rep., MIT Man-Machine Laboratory, Cambridge, MA, 1978.
- [245] L. Save and B. Feuerberg, "Designing human-automation interaction : a new level of automation taxonomy," 2013.
- [246] S. Altavilla and E. Blanco, "Are ai tools going to be the new designers? a taxonomy for measuring the level of automation of design activities," *Proceedings of the Design Society: DESIGN Conference*, vol. 1, p. 81–90, 2020.
- [247] B. C. Zanchin, R. Adamshuk, M. M. Santos, and K. S. Collazos, "On the instrumentation and classification of autonomous cars," in *2017*

- IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2631–2636, 2017.
- [248] K. Lepenioti, A. Bousdekis, D. Apostolou, and G. Mentzas, “Prescriptive analytics: Literature review and research challenges,” *International Journal of Information Management*, vol. 50, pp. 57–70, 2020.
- [249] T. Kjeldsen, *Operations Research and Mathematical Programming: From War to Academia - A Joint Venture*. Eolss Publishers, 2008.
- [250] Y. Bengio, A. Lodi, and A. Prouvost, “Machine learning for combinatorial optimization: A methodological tour d’horizon,” *European Journal of Operational Research*, vol. 290, no. 2, pp. 405–421, 2021.
- [251] C. Ning and F. You, “Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming,” *Computers Chemical Engineering*, vol. 125, pp. 434–448, 2019.
- [252] W. Wiesemann, D. Kuhn, and M. Sim, “Distributionally Robust Convex Optimization,” *Operations Research*, vol. 62, pp. 1358–1376, December 2014.
- [253] G. Calafiore and M. Campi, “The scenario approach to robust control design,” *IEEE Transactions on Automatic Control*, vol. 51, no. 5, pp. 742–753, 2006.
- [254] I. Pappas, D. Kenefake, B. Burnak, S. Avraamidou, H. S. Ganesh, J. Katz, N. A. Diangelakis, and E. N. Pistikopoulos, “Multiparametric programming in process systems engineering: Recent developments and path forward,” *Frontiers in Chemical Engineering*, vol. 2, 2021.
- [255] E. O. Arwa and K. A. Folly, “Reinforcement learning techniques for optimal power control in grid-connected microgrids: A comprehensive review,” *IEEE Access*, vol. 8, pp. 208992–209007, 2020.
- [256] DIgSILENT GmbH, “Digsilent power system solutions.” [Online]. Available: www.digsilent.de.

- [257] Y. Himeur, A. Alsalemi, A. Al-Kababji, F. Bensaali, A. Amira, C. Sardinios, G. Dimitrakopoulos, and I. Varlamis, "A survey of recommender systems for energy efficiency in buildings: Principles, challenges and prospects," *Information Fusion*, vol. 72, pp. 1–21, 2021.
- [258] V. Osadská, "Stochastic methods in risk analysis," in *Transactions of the VSB - Technical University of Ostrava, Safety Engineering Series*, vol. 12 of *Issue 1*, pp. 61–67, 2017.
- [259] R. Azmi, W. Tibben, and K. T. Win, "Review of cybersecurity frameworks: context and shared concepts," *Journal of Cyber Policy*, vol. 3, no. 2, pp. 258–283, 2018.
- [260] Y. S. Afridi, K. Ahmad, and L. Hassan, "Artificial intelligence based prognostic maintenance of renewable energy systems: A review of techniques, challenges, and future research directions," *International Journal of Energy Research*.
- [261] J. Ferrero Bermejo, J. F. Gómez Fernández, R. Pino, A. Crespo Márquez, and A. J. Guillén López, "Review and comparison of intelligent optimization modelling techniques for energy forecasting and condition-based maintenance in pv plants," *Energies*, vol. 12, no. 21, 2019.
- [262] J. Drgoňa, D. Picard, M. Kvasnica, and L. Helsen, "Approximate model predictive building control via machine learning," *Applied Energy*, vol. 218, pp. 199–216, 2018.
- [263] D. Deltetto, D. Coraci, G. Pinto, M. S. Piscitelli, and A. Capozzoli, "Exploring the potentialities of deep reinforcement learning for incentive-based demand response in a cluster of small commercial buildings," *Energies*, vol. 14, no. 10, 2021.
- [264] D. Azuatalam, W.-L. Lee, F. de Nijs, and A. Liebman, "Reinforcement learning for whole-building hvac control and demand response," *Energy and AI*, vol. 2, p. 100020, 2020.
- [265] M. Corra2019, E. Fusari, A. Ferrari, and D. Macii, "A system based on iot platforms and occupancy monitoring for energy-efficient hvac management," in *2019 IEEE 5th International forum on Research and Technology for Society and Industry (RTSI)*, pp. 347–352, 2019.

- [266] Z. J. Lee, D. Johansson, and S. H. Low, “Acn-sim: An open-source simulator for data-driven electric vehicle charging research,” in *2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pp. 1–6, 2019.
- [267] Z. J. Lee, G. Lee, T. Lee, C. Jin, R. Lee, Z. Low, D. Chang, C. Ortega, and S. H. Low, “Adaptive charging networks: A framework for smart electric vehicle charging,” 2020.
- [268] T. Ding, Z. Zeng, J. Bai, B. Qin, Y. Yang, and M. Shahidehpour, “Optimal electric vehicle charging strategy with markov decision process and reinforcement learning technique,” *IEEE Transactions on Industry Applications*, vol. 56, no. 5, pp. 5811–5823, 2020.
- [269] H. M. Abdullah, A. Gastli, and L. Ben-Brahim, “Reinforcement learning based ev charging management systems—a review,” *IEEE Access*, vol. 9, pp. 41506–41531, 2021.
- [270] Y. Nakahira, N. Chen, L. Chen, and S. H. Low, “Smoothed least-laxity-first algorithm for ev charging,” in *Proceedings of the Eighth International Conference on Future Energy Systems, e-Energy '17*, (New York, NY, USA), p. 242–251, Association for Computing Machinery, 2017.
- [271] W. Chen, P. Zhuang, and H. Liang, “Reinforcement learning for smart charging of electric buses in smart grid,” in *2019 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2019.
- [272] B. V. Mbuwir, F. Ruelens, F. Spiessens, and G. Deconinck, “Battery energy management in a microgrid using batch reinforcement learning,” *Energies*, vol. 10, no. 11, 2017.
- [273] T. A. Nakabi and P. Toivanen, “Deep reinforcement learning for energy management in a microgrid with flexible demand,” *Sustainable Energy, Grids and Networks*, vol. 25, p. 100413, 2021.
- [274] R. Henry and D. Ernst, “Gym-anm: Reinforcement learning environments for active network management tasks in electricity distribution systems,” *Energy and AI*, vol. 5, p. 100092, 2021.

- [275] L. Ferrari, F. Esposito, M. Becciani, G. Ferrara, S. Magnani, M. Andreini, A. Bellissima, M. Cantù, G. Petretto, and M. Pentolini, “Development of an optimization algorithm for the energy management of an industrial smart user,” *Applied Energy*, vol. 208, pp. 1468 – 1486, 2017.
- [276] GAMS Development Corp., “General algebraic modeling system.” [Online]. Available: <https://www.gams.com/>.
- [277] F. Conte, F. D’Agostino, P. Pongiglione, M. Saviozzi, and F. Silvestro, “Mixed-integer algorithm for optimal dispatch of integrated pv-storage systems,” *IEEE Transactions on Industry Applications*, vol. 55, no. 1, pp. 238–247, 2019.
- [278] F. Conte, S. Massucco, M. Saviozzi, and F. Silvestro, “A stochastic optimization method for planning and real-time control of integrated pv-storage systems: Design and experimental validation,” *IEEE Transactions on Sustainable Energy*, vol. 9, no. 3, pp. 1188–1197, 2018.
- [279] Z. Hameed, S. Hashemi, and C. Træholt, “Site selection criteria for battery energy storage in power systems,” in *2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pp. 1–7, 2020.
- [280] Z. Tang, Y. Liu, J. Liu, R. Li, L. Wen, and G. Zhang, “Multi-stage sizing approach for development of utility-scale bess considering dynamic growth of distributed photovoltaic connection,” *Journal of Modern Power Systems and Clean Energy*, vol. 4, no. 4, pp. 554–565, 2016.
- [281] M. Motalleb, E. Reihani, and R. Ghorbani, “Optimal placement and sizing of the storage supporting transmission and distribution networks,” *Renewable Energy*, vol. 94, pp. 651–659, 2016.
- [282] S. P. Ntomalis, G. A. Nakas, P. A. Gkaidatzis, D. P. Labridis, A. S. Bouhouras, and G. C. Christoforidis, “Optimal siting of bess in distribution networks under high wind power penetration,” in *2018 IEEE International Energy Conference (ENERGYCON)*, pp. 1–6, 2018.
- [283] R. Chedid and A. Sawwas, “Optimal placement and sizing of photovoltaics and battery storage in distribution networks,” *Energy Storage*, vol. 1, no. 4, p. e46, 2019.

- [284] P. Boonluk, A. Siritaratiwat, P. Fuangfoo, and S. Khunkitti, “Optimal siting and sizing of battery energy storage systems for distribution network of distribution system operators,” *Batteries*, vol. 6, no. 4, 2020.
- [285] O. Babacan, W. Torre, and J. Kleissl, “Siting and sizing of distributed energy storage to mitigate voltage impact by solar pv in distribution systems,” *Solar Energy*, vol. 146, pp. 199–208, 2017.
- [286] N.-A. Masood, M. N. H. Shazon, H. M. Ahmed, and S. R. Deeba, “Mitigation of over-frequency through optimal allocation of bess in a low-inertia power system,” *Energies*, vol. 13, no. 17, 2020.
- [287] S. Afxentis, M. Florides, V. Machamint, C. Yianni, P. Norgaard, H. Bindner, J. Kathan, H. Brunner, C. Mayr, C. Anastassiou, V. Efthymiou, and G. Georghiou, “Energy class dependent residential battery storage sizing for pv systems in cyprus,” vol. 2019, pp. 4770–4774, TheInstitution of Engineering and Technology, 2019.
- [288] L. Zhou, Y. Zhang, X. Lin, C. Li, Z. Cai, and P. Yang, “Optimal sizing of pv and bess for a smart household considering different price mechanisms,” *IEEE Access*, vol. 6, pp. 41050–41059, 2018.
- [289] D. Mejía-Giraldo, G. Velásquez-Gomez, N. Muñoz-Galeano, J. B. Cano-Quintero, and S. Lemos-Cano, “A bess sizing strategy for primary frequency regulation support of solar photovoltaic plants,” *Energies*, vol. 12, no. 2, 2019.
- [290] K. Lai, Y. Wang, D. Shi, M. S. Illindala, Y. Jin, and Z. Wang, “Sizing battery storage for islanded microgrid systems to enhance robustness against attacks on energy sources,” *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 5, pp. 1177–1188, 2019.
- [291] M. B. Sanjareh, M. H. Nazari, G. B. Gharehpetian, R. Ahmadihangar, and A. Rosin, “Optimal scheduling of hvacs in islanded residential microgrids to reduce bess size considering effect of discharge duration on voltage and capacity of battery cells,” *Sustainable Energy, Grids and Networks*, vol. 25, p. 100424, 2021.
- [292] A. Boveri, F. Silvestro, M. Molinas, and E. Skjong, “Optimal sizing of energy storage systems for shipboard applications,” *IEEE Transactions on Energy Conversion*, vol. 34, no. 2, pp. 801–811, 2019.

- [293] D. Wu and X. Ma, “Modeling and optimization methods for controlling and sizing grid-connected energy storage: A review,” *Current Sustainable/Renewable Energy Reports*, vol. 8, pp. 123–130, Jun 2021.
- [294] A. Jawad, Nahid-Al-Masood, and S. Munim, “Optimal sizing of bess for attaining frequency stability under high pv penetration,” in *2021 International Conference on Technology and Policy in Energy and Electric Power (ICT-PEP)*, pp. 348–353, 2021.
- [295] B. O. Alawode, U. T. Salman, and M. Khalid, “A flexible operation and sizing of battery energy storage system based on butterfly optimization algorithm,” *Electronics*, vol. 11, no. 1, 2022.
- [296] J. Kelly and P. Leahy, “Optimal investment timing and sizing for battery energy storage systems,” *Journal of Energy Storage*, vol. 28, p. 101272, 2020.
- [297] “Clean energy for all Europeans,” eu directive, European Union, Luxembourg: Publications Office of the European Union, 2019.
- [298] “CEI 0-16: Reference technical rules for the connection of active and passive consumers to the HV and MV electrical networks of distribution Company,” tech. rep., Italian Electrotechnical Committee, Apr 2014.
- [299] “CEI 0-21: Reference technical rules for the connection of active and passive users to the LV electrical Utilities,” tech. rep., Italian Electrotechnical Committee, Apr 2017.
- [300] A. Keane, L. F. Ochoa, C. L. T. Borges, G. W. Ault, A. D. Alarcon-Rodriguez, R. A. F. Currie, F. Pilo, C. Dent, and G. P. Harrison, “State-of-the-Art Techniques and Challenges Ahead for Distributed Generation Planning and Optimization,” *IEEE Transactions on Power Systems*, vol. 28, pp. 1493–1502, May 2013.
- [301] A. Borghetti, M. Bosetti, S. Grillo, A. Morini, M. Paolone, and F. Silvestro, “A Two-Stage Scheduler of Distributed Energy Resources,” in *2007 IEEE Lausanne Power Tech*, pp. 2168–2173, July 2007.

- [302] R. Anilkumar, G. Devriese, and A. K. Srivastava, "Voltage and Reactive Power Control to Maximize the Energy Savings in Power Distribution System With Wind Energy," *IEEE Transactions on Industry Applications*, vol. 54, pp. 656–664, Jan 2018.
- [303] A. Nassaj and S. M. Shahrtash, "An Accelerated Preventive Agent Based Scheme for Postdisturbance Voltage Control and Loss Reduction," *IEEE Transactions on Power Systems*, vol. 33, pp. 4508–4518, July 2018.
- [304] W. Zhang, O. Gandhi, H. Quan, C. D. Rodriguez-Gallegos, and D. Srinivasan, "A multi-agent based integrated volt-var optimization engine for fast vehicle-to-grid reactive power dispatch and electric vehicle coordination," *Applied Energy*, vol. 229, pp. 96–110, 2018.
- [305] I. Afandi, P. Ciufu, A. Agalgaonkar, and S. Perera, "A holistic approach for integrated Volt/Var control in MV and LV networks," *Electric Power Systems Research*, vol. 165, pp. 9–17, 2018.
- [306] I. Kim, "Optimal distributed generation allocation for reactive power control," *IET Generation, Transmission Distribution*, vol. 11, no. 6, pp. 1549–1556, 2017.
- [307] M. Nick, R. Cherkaoui, and M. Paolone, "Optimal Planning of Distributed Energy Storage Systems in Active Distribution Networks Embedding Grid Reconfiguration," *IEEE Transactions on Power Systems*, vol. 33, pp. 1577–1590, March 2018.
- [308] IBM, "Ibm cplex optimizer." [Online]. Available: <https://www.ibm.com/it-it/analytics/cplex-optimizer>.
- [309] F. D'Agostino, S. Massucco, P. Pongiglione, M. Saviozzi, and F. Silvestro, "Optimal DER Regulation and Storage Allocation in Distribution Networks: Volt/Var optimization and Congestion Relief," in *2019 IEEE Milan PowerTech*, pp. 1–6, June 2019.
- [310] A. Garces, "A Linear Three-Phase Load Flow for Power Distribution Systems," *IEEE Transactions on Power Systems*, vol. 31, pp. 827–828, Jan 2016.

- [311] R. P. O' Neill, A. Castillo, and M. Cain, "The IV Formulation and Linear Approximations of the AC Optimal Power Flow Problem," *FERC staff technical paper*, Dec 2012.
- [312] F. Sossan, E. Namor, R. Cherkaoui, and M. Paolone, "Achieving the dispatchability of distribution feeders through prosumers data driven forecasting and model predictive control of electrochemical storage," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 4, pp. 1762–1777, 2016.
- [313] S. Massucco, P. Pongiglione, M. Saviozzi, F. Silvestro, P. Almaleck, and P. Serra, "Algorithm for optimal microgrid operation and control with adaptable constraints and flexible objective function," in *2019 IEEE 5th International forum on Research and Technology for Society and Industry (RTSI)*, pp. 97–102, 2019.
- [314] K. Gao, T. Wang, C. Han, J. Xie, Y. Ma, and R. Peng, "A review of optimization of microgrid operation," *Energies*, vol. 14, no. 10, 2021.
- [315] S. Y. Derakhshandeh, M. E. Hamedani Golshan, and M. A. Masoum, "Profit-based unit commitment with security constraints and fair allocation of cost saving in industrial microgrids," *IET Science, Measurement & Technology*, vol. 7, no. 6, pp. 315–325, 2013.
- [316] H. Farzin, M. Fotuhi-Firuzabad, and M. Moeini-Aghtaie, "A stochastic multi-objective framework for optimal scheduling of energy storage systems in microgrids," *IEEE Transactions on Smart Grid*, vol. 8, no. 1, pp. 117–127, 2017.
- [317] A. Parisio, E. Rikos, and L. Glielmo, "A model predictive control approach to microgrid operation optimization," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 5, pp. 1813–1827, 2014.
- [318] F. Conte, M. Saviozzi, and S. Grillo, "An optimization problem for day-ahead planning of electrical energy aggregators," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 12213–12220, 2020. 21st IFAC World Congress.
- [319] AMPL Optimization inc., "A mathematical programming language." [Online]. Available: <https://ampl.com/>.

- [320] Toshiba Corp., “Scib™ cells.” [Online]. Available: <https://www.global.toshiba/ww/products-solutions/battery/scib/>.
- [321] The MathWorks, Inc., “Matlab software.” [Online]. Available: <https://www.mathworks.com/products/matlab.html>.

List of publications originated during the Ph.D. work

- [1] S. Massucco, G. Mosaico, M. Saviozzi, and F. Silvestro, "A hybrid technique for day-ahead pv generation forecasting using clear-sky models or ensemble of artificial neural networks according to a decision tree approach," *Energies*, vol. 12, p. 1298, Apr 2019.
- [2] S. Bianchi, A. De Filippo, S. Magnani, G. Mosaico, and F. Silvestro, "Virtus project: A scalable aggregation platform for the intelligent virtual management of distributed energy resources," *Energies*, vol. 14, p. 3663, Jun 2021.
- [3] G. Mosaico and M. Saviozzi, "A hybrid methodology for the day-ahead pv forecasting exploiting a clear sky model or artificial neural networks," in *IEEE EUROCON 2019 -18th International Conference on Smart Technologies*, pp. 1–6, 2019. (Presenting author, 3rd Prize at the IEEE R8 Student Paper Contest 2019)
- [4] G. Mosaico, M. Saviozzi, F. Silvestro, A. Bagnasco, and A. Vinci, "Simplified state space building energy model and transfer learning based occupancy estimation for hvac optimal control," in *2019 IEEE 5th International forum on Research and Technology for Society and Industry (RTSI)*, pp. 353–358, 2019.
- [5] S. Massucco, G. Mosaico, P. Pongiglione, M. Saviozzi, and F. Silvestro, "Probabilistic planning for distribution networks including optimal der

- regulation and storage allocation,” in 2020 IEEE Power Energy Society General Meeting (PESGM), pp. 1–5, 2020.
- [6] S. Massucco, G. Mosaico, M. Saviozzi, F. Silvestro, A. Fidigatti, and E. Ragaini, ”An instantaneous growing stream clustering algorithm for probabilistic load modeling/profiling,” in 2020 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), pp. 1–6, 2020. (Presenting author)
- [7] S. Massucco, G. Mosaico, M. Saviozzi, F. Silvestro, A. Fidigatti, and E. Ragaini, ”A markov chain load modeling approach through a stream clustering algorithm,” in 2020 AEIT International Annual Conference (AEIT), pp. 1–6, 2020.
- [8] Bagnasco, A.; Vinci, A.; Silvestro F.; Mosaico, G; ”Soluzioni innovative per il risparmio energetico negli edifici del terziario” in ”Gestione Energia” N. 1/2020. [in Italian ”innovative solutions for energy saving in tertiary sector buildings”].
- [9] F. Conte, G. Mosaico, G. Natrella, M. Saviozzi, F. R. Bianchi, ”Optimal Management of Renewable Generation and Uncertain Demand with Reverse Fuel Cells by Stochastic Model Predictive Control” (submitted to 2022 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)).
- [10] S. Bianchi, A. De Filippo, S. Magnani, M. Milano, G. Mosaico, F. Silvestro, ”Il ruolo dei Virtual Power Plants nella gestione energetica sostenibile” (accepted to Conference of National Laboratory “Artificial Intelligence ed Intelligent Systems” (AIIS)). [in Italian ”The role of Virtual Power Plants in sustainable energy management”].
- [11] P. Almaleck, G. Mosaico, M. Saviozzi, P. Serra, P., ”Electrical Consumption Forecasting in Sports Venues: A proposed approach based on artificial neural networks and arimax models” (Submitted to Energy and Buildings journal).

Projects into which I have been involved during my Ph.D. work

1. **Analysis and Development of probabilistic security assessment techniques in realistic electrical networks**, with "Ricerca sul Sistema Energetico" – R.S.E. S.p.A., year: 2021.

Short description: Development of stochastic optimization algorithms to manage security problems in electrical networks with relevant dimensions (e.g., security-constrained optimal power flow and optimal redispatching in "N-1" analysis). The proposed procedure is based on Benders Decomposition to be suitable for realistic networks.

2. **e-SCALE (Energy and services connected to the aggregate management)**, with "Università di Cassino e del Lazio Meridionale", year: 2020.

Short description: The e-SCALE project has developed and validated a complete service system for aggregator/energy community management. In particular, the focus of this project has been:

- the analysis of models for the evaluation and optimization of ancillary services;
- the definition and realization of use cases for different types of users.

3. **Techniques for developing applications for the probabilistic security assessment**, with "Ricerca sul Sistema Energetico" – R.S.E. S.p.A., year: 2021.

Short Description: This activity has realized a stochastic optimization algorithm in Matlab-GAMS environment to solve a security-constrained optimal power flow problem. The proposed procedure is based on a chance-constraint approach to managing the uncertainties due to RES and load.

4. **PODCAST** (Platform for optimizing distribution through smart meter data and distributed power storage), with Ricerca di Sistema (CSEA – "Cassa per i Servizi Energetici e Ambientali"), years: 2017-2020.

Short description: The project has developed innovative tools for providing services for transmission system operators, distribution system operators, and production and load operators exploiting smart-meter data. In addition, the project has implemented a distribution management system with advanced functionalities (state estimation, load and PV forecasting, load modeling, flexibility management, profile optimization, fault location, volt/var control, and losses minimization).

5. **PREDICT** (Adaptive energy efficiency platform for reducing consumption in non-residential buildings), with "Regione Liguria", years: 2017-2019.

Short description: This project aimed to realize an adaptive platform for energy efficiency and consumption reduction in non-residential buildings. The proposed platform can dynamically elaborate the optimal temperature setpoints for each building zone, considering actual energy consumption, weather data (external temperature, solar irradiation, and wind speed), and building occupancy.

6. **Optimal management and control of conventional, renewable generation and storage systems in the context of energy system evolution**, with Hitachi ABB Power Grids, years: 2019-2022.

Short description: This research project has been focused on:

- Tools and algorithms for load forecasting;
- E-mobility modeling;

- Optimal sizing of storage systems;
- Reliability methodologies for microgrid control systems.

7. **Load modeling project**, with ABB S.p.A., years: 2019-2022.

Short description: The main activity of this project has been the development of a load modeling algorithm. This project implemented an adaptive streaming algorithm to model any load through a Markov Chain. The proposed procedure can cluster with minimal computational effort, allowing real-time modeling.

8. **VIRTUS - Virtual management of distributed energy resources**, with Ricerca di Sistema (CSEA – "Cassa per i Servizi Energetici e Ambientali"), years: 2019-2022.

Short description: The project aims to carry out the prototypical realization of a virtual power plant in an industrial tertiary context. This has to allow the integration of conventional electricity and thermal energy generation with network infrastructures and supply systems along the entire value chain: the organization of the energy portfolio, the use of forecasting systems, the management of customer relationships, the adoption of geographical location systems and the measurement and billing activities.

9. **Energy consumption data analysis in ships project**, with Cetena S.p.A., years: 2021-2022.

Short description: The project aims to develop an IT tool for simple, flexible and effective creation, subdivision, management, and analysis of the information database deriving from a ship, thanks to the innovative techniques based on artificial intelligence in order to identify energetic correlations between the users and identification of possible actions to improve its energy performance (both technological and managerial).

Attended post-graduation courses and foreign experiences during the Ph.D.

1. "Elsevier on Campus" seminar, Genoa - October 10th, 2018 (4 hours).
2. Participation in the doctoral course "Advanced Programming In Matlab And Simulink" - Prof. Alberto Oliveri, DITEN, October 2018 (20 hours).
3. Participation in the "Electrical Systems" course - Prof. Marco Invernizzi - Prof. Stefano Bracco - University of Genoa, DITEN, 2018 and 2019 (60 hours).
4. "Dynamic Optimization with Convergence Guarantees" seminar, Dr. Eric Kerrigan (Imperial College London) - DITEN November, 13th 2018, (2 hours).
5. "The renewable energy industry and its landscape today and tomorrow" seminar - Dr. Andrés Isaza (CMO of GE Renewable Energy) - DITEN - November, 29th 2018 (2 hours).
6. "The transformation of the electricity system, the new management needs and the tools to cope with them" seminar - AEIT-AEE, Milan, December 4th, 2018 (8 hours).

7. Participation in the doctoral course "An Introduction to optimization over time and its application to online machine learning and reinforcement learning" - Prof. Giorgio Gnecco (IMT Lucca) - DIBRIS, February 2019 (18 hours).
8. Participation in the doctoral course "Model and applications of predictive control" - Mauro Gaggero - DIBRIS, March 2019 (18 hours)
9. "Digitization in the electricity and water sector" seminar organized by AEIT - May 28th, 2019, organized by ABB Power Grid at Genova Sestri Ponente (4 hours). x
10. Participation in the Summer School "Data-Driven Analytics and Optimization for Energy Systems", organized by the Danish Technical University (DTU) in Lyngby (Denmark) - June 16-21st, 2019 (44 hours).
11. "Digitization in the electricity and water sector" seminar organized by AEIT - May, 28th 2019, organized by ABB Power Grid at Genova Sestri Ponente (4 hours).
12. Participation in the course "Management and Control of Electrical Systems" - Prof. Stefano Massucco - University of Genoa, DITEN, 2019 (60 hours).
13. Participation in the doctoral course "Introduction to the quantification of uncertainty and the analysis of stochastic sensitivity" held by Prof. Ing. Alessandro Mariotti of the University of Pisa. December 17th, 2019 to December 19th, 2019 (20 hours).
14. Participation in the course "RAM & PHM 4.0: Advanced methods for reliability, availability, maintenance, diagnostics, and prognostics of industrial equipment" held by the Politecnico di Milano, February-March 2021 (32 hours). Also, recipient of one of the scholarships reserved for two Ph.D. students attending the course.
15. Attendant at the "Advanced Machine Learning" Course", University of Liège (online for covid): workshops on Gaussian processes, statistical learning theory, probabilistic prediction, deep learning, simulation-based inference, causality, reinforcement learning, responsible AI, February-May 2021 (30 hours).

-
16. Attendant at the "Design and management of electrical and energy systems", University of Liège (online for covid): 4 modules on power-sharing in DC microgrids, planning, and management of distribution networks, reliability management, sustainability, and integration of electrical networks, February-May 2021 (30 hours).
 17. "Microgrids: Swedish and International Perspectives" seminar organized by the IEEE Power and Energy Society & Power Electronics Society Sweden Chapter, May 26th, 2021, (4 hours).
 18. Deep Learning Specialization Certificate, earned after completing the courses offered by deeplearning.ai via Coursera (2021):
 - "Neural Network and Deep Learning" (27 hours);
 - "Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization" (26 hours);
 - "Structuring Machine Learning Projects" (8 hours);
 - "Convolutional Neural Network" (40 hours);
 - "Sequence Models" (41 hours).
 19. Audited the Reinforcement Learning Specialization, offered by University of Alberta and Alberta Machine Intelligence Institute via Coursera, 2021 (16 hours).