Ecole Centrale Lille

Faculté Polytechnique de Mons

PhD thesis

Data-based investigations of Low Voltage Distribution Systems: Machine Learning Applications for the monitoring of the network under ageing and variable atmospheric conditions

by

Egnonnumi Lorraine CODJO

A report submitted in partial fulfilment of the requirements for the degree of

PhD in Electrical Engineering in cooperation with the

Power system and markets research group (PSMR Group), Université de Mons

Laboratoire d'électrotechnique et d'électronique de puissance (L2EP), Centrale Lille

Doctoral Defense Committee

Pr Jacques LOBRY	Université de Mons	Chairman
Pr Emmanuel DE JAEGER	Université Catholique de Louvain	Jury member
Pr Manuela SECHILARIU	Université de Technologie de Compiègne	Reviewer
Pr Corine ALONSO	Université de Toulouse	Jury member
Pr Jean-Paul GAUBERT	Université de Poitiers	Reviewer
Dr Bashir BAKHSHIDEH ZAD	Université de Mons	Secretary
Pr Bruno FRANCOIS	Ecole Centrale de Lille	Thesis supervisor
Pr François VALLEE	Université de Mons	Thesis supervisor

November 2022

Acknowledgment

This work was made possible with the financial and intellectual support of some people.

I would first like to thank the "**Région Hauts-de-France**" (in France) and the "**Fonds de la Recherche Scientifique**" (FNRS in Belgium) for their financial support during this research activities.

I would also like to thank my supervisors Pr. Bruno FRANCOIS and Pr. Francois VALLEE for their patience, their unconditional support, their directives, their supervision and their advice during those years.

Many thanks to Dr. Bashir BAKHSHIDEH ZAD and all my fellow colleagues from both the Power system and markets research group (PSMR Group) and the Electrical engineering and power electronics laboratory (L2EP).

I thank also ORES for kindly giving access to metering data.

Finally, thanks to Fabrice, my mum, my dad and my beloved family for their support and kind words during the best and hardest moments of the last few years.

CODJO Egnonnumi Lorraine

Abstract

This thesis explores potential applications of energy data measured by smart meters in low voltage (LV) electrical networks. From these data, a heuristic algorithm makes it possible to propose an equivalent architecture of the network and to identify its impedances by an optimal minimization of modeling errors. The model obtained is used to study the impact of the variations and increase of the ambient temperature on the operation of the electrical network. Variations from load demands, PV generations, the ambient temperature and working environment conditions increase the leakage current through the cable insulation and so accelerate the ageing of the infrastructure. Some impacts analysis are carried out to characterize and model the cables insulation degradation, based only on the measurement data from smart meters. Since satisfactory mathematical models based on the physics cannot be developed for the application, several supervised machine learning methods are applied to access the condition of the electrical system. Case studies are analyzed to compare the accuracy of learning methods for different degradation scenarios. The proposed frameworks offer promising perspectives for the early identification of LV cable conditions by using SM measurements combined to ML approaches, Load Flow computations and Monte Carlo using scenario simulations to calculate the network voltages.

Keywords : Low voltage distribution networks; supervised Machine Learning methods; cable condition degradation; Load Flow computation; smart meter; Monte Carlos scenarios; cable insulation wear; cables conditions assessment.

Résumé

Cette thèse explore des applications potentielles des données énergétiques mesurées par les compteurs communiquant des réseaux électriques de basse tension (BT). A partir de ces données, un algorithme heuristique permet de proposer une architecture équivalente du réseau électrique et d'identifier ses impédances par une minimisation optimale des erreurs de modélisation. Le modèle obtenu est utilisé pour étudier l'impact de la variation et de l'augmentation de la température ambiante sur le fonctionnement du réseau électrique. Les variations de la demande des charges électriques, des générateurs PV, de la température ambiante augmentent le courant de fuite à travers l'isolation du câble et accélère le vieillissement de l'infrastructure. Des analyses d'impacts sont menées pour caractériser et modéliser la dégradation de l'isolation des câbles en se basant uniquement sur les données issues des compteurs « intelligents ». Etant donné que des modèles mathématiques satisfaisants basés sur la physique ne sont pas développables, plusieurs méthodes d'apprentissage automatique supervisé sont appliquées pour évaluer l'état de dégradation des câbles. Des cas d'études sont analysés pour comparer la précision des méthodes d'apprentissage pour différent scenarii de dégradations. Les outils proposés offrent des perspectives prometteuses pour l'identification précoce des défauts dans les câbles BT en utilisant des mesures issues des compteurs combinés à des approches d'apprentissage et des calculs de flux de puissances utilisant des simulations de scénarios de Monte Carlo pour calculer les tensions.

Mots clés : Réseaux de distribution basse tension; ; Méthodes d'apprentissage automatique supervisées; dégradation de l'état du câble; calcul des flux de puissance; compteurs communicant, scénarios de Monte Carlo; usure du matériau isolant des câbles; évaluation de l'état des câbles.

Contents

List of Figures

List of Table

Acronyms and Abbreviations

AC	Analog Current
ADF	Advanced Driving Functions
AMI	Advanced Metering Infrastructure
AMM	Advanced Meter Management
ANN	Artificial Neural Network
CAPEX	Capital Expenditure
CART	Classification And Regression Trees
CDF	Cumulative Distribution Function
DMS	Distribution Management System
DG	Distributed generation
DGs	Distributed generators
DPVGs	Distributed Photovoltaic Generators
DSO	Distribution System Operator
DT	Decision Tree
EEA	European Environment Agency
EU	European Union
EV	electric vehicles
FP	False Positive
GIEC	Groupe d'experts intergouvernemental sur l'évolution du climat
GIS	Geographic Information System
HV	High Voltage
ICT	Information and Communications Technology
IT	Information Technology
kNN	k-nearest neighbor
LF	Load Flow
LR	Logistic Regression

LSQ	Least Squares Optimization
LV	Low Voltage
MC	Monte Carlo
ML	Machine Learning
MV	Medium Voltage
MTBF	Mean Time Between Failure
NRLF	Newton-Raphson load flow
ONERC	Observatoire National sur les Effets du Réchauffement Climatique
OPEX	Operational Expenditure
PLM	Power Line Modems
PV	Photovoltaics
PVC	Polyvinyl Chloride
RE	renewable energies
RES	Renewable Energy Sources (RES)
RF	Random Forest
RMS	Root Mean Square
RLC load	Resistor (R), Inductor (L) and Capacitor (C) load
ROC	Receiver Operating Characteristic
SCADA	Supervisory Control and Data Acquisition
SDG	Smart Distribution Grid
SGs	Smart Grids
SM	Smart Meters
SVM	Support Vector Machine
TSO	Transport System Operator
TURPE	Tarif d'Utilisation du Réseau Public d'Electricité (Public Transmission User Tariff)
VVC	Volt Var Control
TP	True Positive
TN	True Negative
XLPE	Cross-Linked Poly-Ethylene

Symbols

$Cons(q)$	Power consumption $Cons$ from the network at quarter q
CTi,j	Value assigned to their connection test between node i and node j
Ei,j	Euclidean distance between node i and node j
$Inj(q)$	Power injection Inj from the network at quarter q
it	iteration
$length_i$	length of the cable between node i and node j
$Load(q)$	Load demand at quarter q
N	Total number of nodes for the radial electrical network
ND	Net Demand
$P(q)$	Real power P, at quarter q
Pn	Nominal power (Puissance nominale)
$Prod(q)$	PV generation $Prod$ at quarter q
$Q(q)$	Reactive power Q, at quarter q
q	A quarter of an hour ($q = \{1 .. 96\}$ based on each daily ¼-hourly period)
R_{iso}	Insulation resistance
Ri	per-unit-length serie resistance
R_t	Temperature based resistance
T	squared connection matrix
V_{n_i}	measured voltage of node i
V_{z_i}	calculated voltage of node i
Xi	per-unit-length serie reactance
x_k	input vector of data for any machine learning algorithm
yk	output vector of data obtained from any machine learning algorithm
Z_{br}	Total impedance matrix of the radial network
Zi	Self-impedance between node i and node j

General introduction

Contents

1. European sustainable policies involve massive connections of sources to the LV network

At the origin, electrical distribution systems have been designed for a unidirectional power flow going from centralized large power plants towards small-scale end-users (Figure 1).

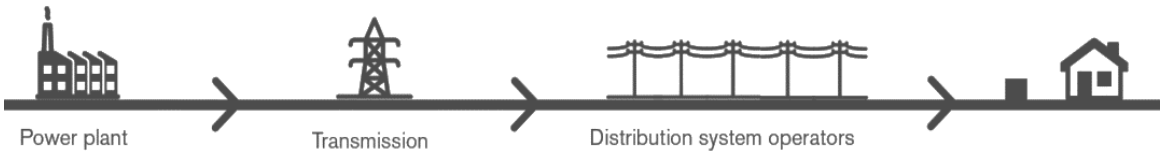


Figure 1: Traditional electrical network with large power plants and unidirectional electricity flow [1].

Since some decades, the European Union has developed particular incentive policy decisions in order to decarbonize the economy and reduce the energy dependence. According to the European Environment Agency (EEA) report "Trends and Projections in Europe 2021" published in October 2021, the EU has achieved its three 2020 climate and energy targets [2]:

- 20% reduction of its greenhouse gas emissions by 20% compared to 1990 levels,
- 20% of the energy consumed from renewable energy sources (RES), and
- 20% improvement of the energy efficiency.

One of the highlights of the report was the need to the continued introduction of renewable sources for electricity generation in order to make more sustainable the electrical energy sector and to achieve the new EU's renewable energy targets. A new 2030 target with a 55% reduction in net greenhouse gas emissions has indeed been set; the latter requiring additional efforts and new policies (such as the faster decline of the energy consumption) compared with the efficiency gains achieved from 2005 to 2020. In France, the targets of the Law « loi d'action et de mobilisation » are to reach a 40% reduction of gas emission in 2030 and to increase by 30% the renewable energy production while decreasing by 50% the nuclear production in 2025.

To suit with this changing energy scenario, since 2007, the design of the electricity system has been re-thought within the integration of new "green and cost friendly" energy sources. These renewable energy sources are mainly characterized by reduced powers and are consequently connected in low voltage or medium voltage networks. Thereby, traditional one-directional systems are migrating towards bidirectional systems with new required sensors, information and technology solutions and active management strategies (Figure 2).

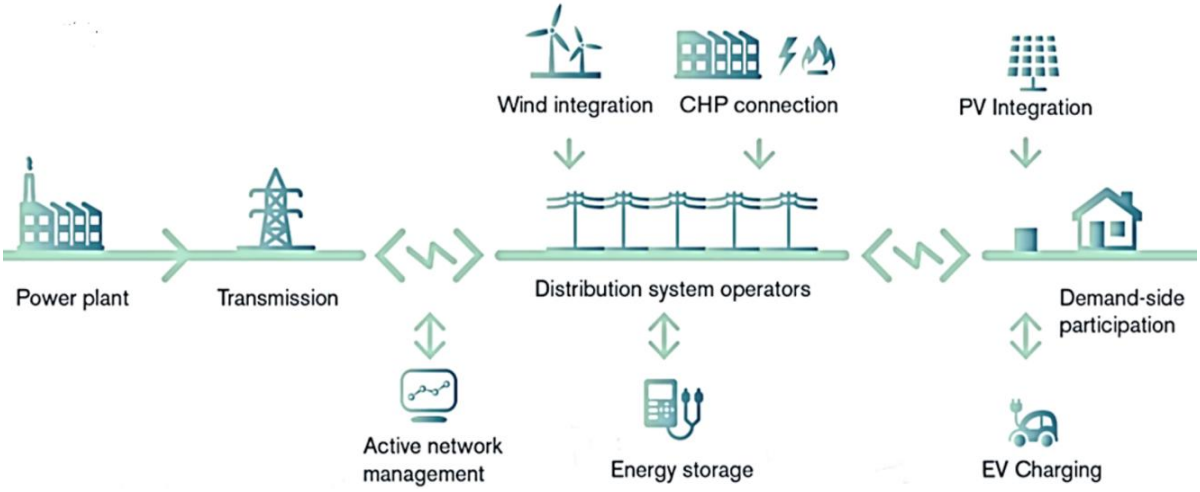


Figure 2: Smart grid with bidirectional electricity flow with decentralized electricity generations [1].

2. Smart Meters and ICT as new available technologies for developing a smarter grid

Moreover, the European Directive of July 13th, 2009 requires to install some intelligent measurement systems in order to enable the active participation of consumers in the electricity supply market. The idea is to enable future business models for a more transactive grid as the provision of dynamic pricing for consumers. The implementation and effective use of individual communicating Smart Meters (SM) is being generalized, as well in France as in Belgium [3]. The goal was that at least 80% of consumers would be equipped with intelligent measurement systems by 2020 [4]. As an example, the energy transition law for green growth of August 18, 2015 is a transposition of that directive in France [5]. In consequence, the French demonstration project Pilot Linky involved the installation of smart meters. In a same way, developments in Information and Communication Technologies (ICT) have been constant in the capacity and speed of data transmission. The result is a growing interest in creating value through a better management and organization of resources.

3. New big data applications

Smart Meters provide to consumers information about their electricity bills and real consumption. They are also sources of temporal and localized information about the electricity flow and so can lead to several applications as shown in (Figure 3).

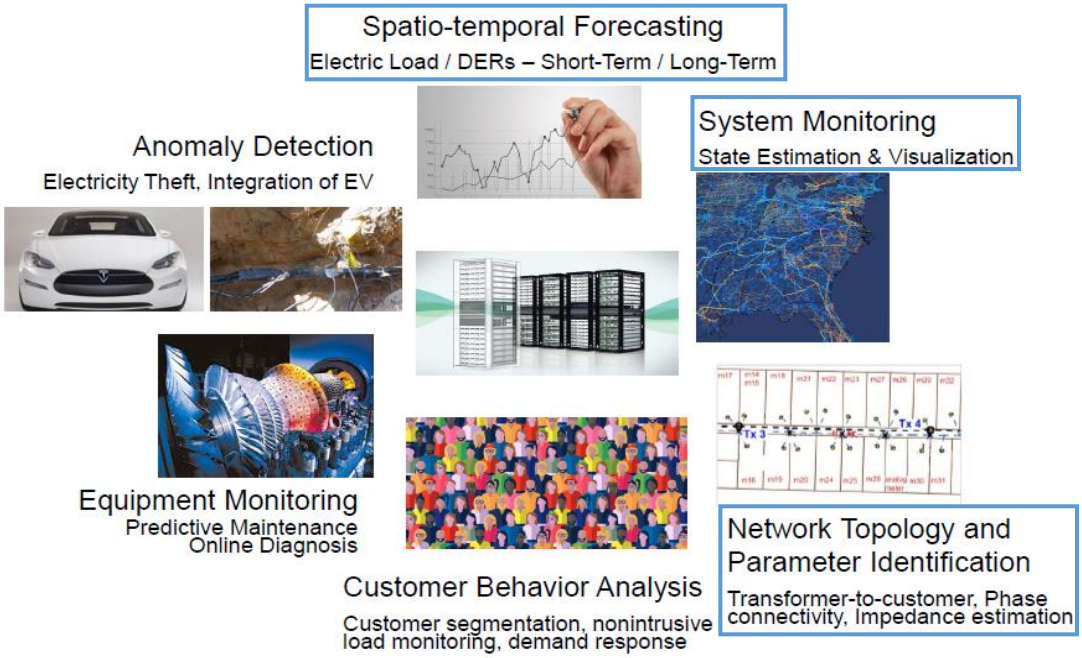


Figure 3: Big data applications in Power distribution systems [6].

Investment in smart meters is currently mainly justified on the basis of the expected reduction of DSO's operational costs, typically resulting from the cost reduction of meter reading, reduction of power theft, remote activation and deactivation of services, faster detection of power outages. With new masses of available data, artificial intelligence techniques can be applied for analysis and decision-making. Some demonstration projects have been experimented and some major ones are listed hereunder.

Regarding the European Research and Innovation Program (Horizon 2020 program), the InterFlex project has emerged from April 2016 to December 2019. The project brings together five European electricity distributors (respectively in Sweden, Netherlands, Czech Republic, Germany, France) around improving the reliability and automation of the electrical network [7] by using industry-scale demonstrators such as the Nice Smart Valley set up in France. One of the project's achievement relied on the remote management of distributed generation by deploying control boxes associated to smart meters that can contribute to relieving grid constraints. As another project, we can quote Atrias, a Belgian project focused on interconnection of all energy players to the network data (either SM data or energy market data). In France, the SESAM Grids project, a partnership between ENGIE (multinational electric company) and academic laboratories, has investigated the reliability and the security of smart grids using meteorological stations and a hundred smart meters deployed on a smart demonstrator at the scale of an economic activity zone [8]. Thanks to SM, the individual measurements along the LV feeder can provide a lot of new information. The task of this research work is to combine these available data with signal processing and data mining methods for improving the knowledge of the LV system through the development of new Machine-Learning based models.

Why creating an upgraded model of the electrical network cable by using a SM database?

To supply electricity, the distribution system operator has to supervise their network in real time in accordance with standards and try to anticipate their maintenance. As exposed in section 1, the arrival of generation in low voltage (LV) networks involves new management challenges. A wiring model (mathematical or physical) of power lines, whether realistic or hypothetical, facilitates the analysis of the electrical network under operation and helps the network management and operation setting. In addition, the growing of stakeholders in the electrical system (renewable energy based producers, electric vehicles, etc.), the lack of control flexibilities over the ageing infrastructures and the increasingly environmental issues are forcing grid operators to adapt their practices and strategy for operating their electricity networks in order to make it more flexible and intelligent. Finally, with the fourth industrial revolution (the numerical one), the energy sector does not stay away from innovation and is consequently witnessing an increase in investments and projects in the area. Using Smart Meter data to better assess the network cable condition (and over a training period) will provide an adaptive approach to the network monitoring and to its maintenance.

In what extent those new models can overcome the existing ones?

This question is clearly a hot topic because it extrapolates the problematic of using conventional models of the electrical network. In fact, impedances vary with the age, environmental phenomena and most of the time with the different reinforcement and maintenance works performed in the past. Hence, the values of the network impedance are not exactly known with accuracy and are, generally, based on the theoretical data originating from the initial deployment of the network. Moreover, Low Voltage networks are still facing problems with the very poorly known technical characteristics and obsolescence of their equipment (e.g. circuit breakers). The impedance matrix resulting from the new models will open the possibility of predicting the consequences of ageing and help to decide for equipment replacement and investments in electrical networks.

What purpose with what kind of extracted information?

With the location of meters, the topology of an equivalent electrical network can be built and the network impedances can be estimated. Hence, this data-driven model can be used to analyze the variations of line parameters according electrical operating conditions but also physical and environmental conditions as material ageing, degradations, seasonal and climate meteorological variations, and so on. For network operators, adaptive impedance based models could be used to perform an appropriate fault tolerance control procedure or to better integrate the uncertainty in network parameters in their (short to long-term) management process. This is an opportunity to have new robust tools that can be integrated into the techno-economic analysis of the network while ensuring a permanent monitoring and diagnosis of its wired infrastructure without interfering with its operation.

Information from SM can be also a new opportunity for improving/securing the electrical network and the quality of service provided to consumers. The targeted applications (through the developed tools) are hoped to significantly improve the reliability, safety and availability of LV distribution systems. In a nutshell, the research targets of this PhD Thesis are therefore oriented towards :

- a better analysis of the LV network sensitivity to temperature or PV generation variations and to the cable condition degradation,
- and the development of an early detection of emerging faults thanks to Machine Learning techniques.

The final interest is to be able to guarantee the maximal availability of the electrical grid although various changes in the infrastructure by reducing the failure times of the electrical network and so maximizing the electrical availability, in a context of the growing demand from numerous emerging users and/or decentralized generation.

4. Thesis objective

Transmission lines are critical infrastructures for the sharing of electricity and have motivated earlier the use of specific meters, monitoring instruments and communication systems. And so, during many years, research projects have been focused on HV and MV transmission lines. This presented research project addresses the upgrading of the Low Voltage distribution network structures modelling and cable condition assessment through the exploitation of information provided by Smart Metering devices installed at homes. Smart meter measurements in the residential sector can drastically enlarge the hosting capacity of distribution network (impact of adding new distributed photovoltaic generators (DPVGs) to the electrical distribution system) and enable cost-effective planning approaches. Proposed contributions aim to set up some cost-effective approaches for LV electrical distribution systems through the opportunity offered by the extended installation of those smart metering devices. The Smart Meter data can then be seen as an opportunity to create new adaptive modelling techniques/tools for those initially poorly metered networks. Potential applications are targeting an advanced monitoring and an early diagnosis of LV distribution network degradations before a main failure.

The main contributions of this research work are:

- the circuit cell based modelling of an electrical network architecture with SM data,
- an impact analysis of temperature variations onto voltage variations,
- the application and comparison of various Machine Learning (ML) techniques for condition assessment of LV networks.

Modelling of an electrical network architecture with SM data

The individual measurements along the LV feeder can provide a lot of new information. Those last ones could be combined with signal processing and data mining methods for developing new models while trying to overcome the lack of knowledge about the cable condition parameters (degradation, seasonal variation appearance) and/or LV network operating conditions.

Impact analysis of temperature variations onto voltage variations

Regarding the climate changes of the last two decades, this contribution brings an original and relevant aspect to the management of distribution networks. Indeed, climatic and environmental phenomena (such as temperature variation and global warming) bring a lot of voltage issues in the LV systems. The extreme temperatures have created some problems of unavailability of the electricity production and a drop in the transmission capacity of LV lines / cables. There is then a need of considering some upgraded values (temperature-based resistance values) of the network parameters in the existing Load Flow (LF) calculation algorithm. This last one will be useful for establishing the impact of the resistance distribution on a Probabilistic Load Flow process.

Application and comparison of various Machine Learning (ML) techniques for the network condition assessment

Assessing the LV network cable condition is part of the network management strategy that allows to keep the critical variables, such as the voltage level, no matter the network operation constraints, in some defined and tolerable margins. To reach this challenge, we need to take advantage of the available SM data in order to detect the soft degradation (at early-stage) of the cable insulation (regardless of the type of fault). A variety of data analysis techniques can be applied to extract a meaningful knowledge from large, noisy databases. Among these techniques, an important feature present in Machine Learning methods is the ability to be adaptable to the local characteristics that are contained into data. To apply these ML techniques and define the input databases, the relationships between the operating conditions of the electrical network, its nodal voltages and thickness variation of cable insulation need to be highlighted.

For the Distribution System Operators (DSO), this research is a step towards a higher reliability and improving cost-effectiveness of LV distribution networks while assessing the grid conditions. The final framework aims to offer promising perspectives for early identification of LV cable conditions by using available SM measurements. An important aspect of this work should therefore be devoted to evaluate how signal processing, data analysis

methods and probabilistic Load Flow computation could contribute to access, even more, to predict the current condition of the LV networks.

5. Outline of the manuscript

The content of the thesis report is organized according to the followed scientific works:

Chapter 1 presents the context of this research and a state of the art dedicated to the new challenges in LV networks. The organization and operation standards of those networks are presented as well as the role of smart meter technologies in the management of distributed generation. Then, the chapter focuses on new challenges for LV systems, from the problems faced (either users or infrastructures / operation problems) to the emerging technologies and opportunities brought by smart meter and smart grid concepts. Finally, a state of the art is proposed; going from the modelling of cable degradation towards Machine Learning applications for monitoring purposes with measured data available in LV networks.

Chapter 2 presents an analysis of the temperature impact on the characterization and modeling of LV cables. A part of a real-life LV network is considered and further is described. In that way, a model of the network is presented in healthy cable condition to better characterize the line resistance variations by taking into account the temperature and assess its impact on the nodal voltages of LV feeders. A Monte Carlo based Probabilistic Load Flow (LF) algorithm for radial Low Voltage networks is then implemented to calculate these voltages and is applied to different seasonal conditions.

Chapter 3 investigates the influence of changing network impedances and variable operating environment on observed voltage variations. The electrical conductance variation of the cable insulation, due to the degradation of the insulating material, is analyzed. Considering the load demand and photovoltaic generation in the network as well as impedance data, in different degradation conditions, load flow calculations are conducted in order to calculate the nodal voltages and to build the network voltage profile. Then, based on a conducted study for various degrees of insulation wear and network working points, interesting information are revealed about the probability of voltage variation appearance in different nodes of the studied network.

Chapter 4 focuses on the cable condition assessment of a single LV distribution cable. The task is to detect a cable degradation with Machine Learning techniques. First, some data analysis approaches are presented and compared. A focus on their usefulness application to LV systems is performed. Then a comprehensive database is generated and is linking nodal voltages variations with different cable degradation conditions and customer Net Demand changes. Finally, a Learning-based process is built for identification of a cable degradation based on (voltage and Net Demand) data from Smart Meter (SM) measurements.

Chapter 5 addresses the implementation of a generic Learning-based identification / classification framework for an entire LV network. Based on the binary single cable condition assessment developed in Chapter 4, the previous work is extended to a complete feeder. Multiple Machine Learning algorithms are set up to do a multi class classification of all electrical lines and are compared in terms of performance.

The least chapter is the general conclusion with recommendations of this work. It concludes this research project by presenting a comparative analysis of the performance of the Learning-

based identification / classification framework for different types of LV networks. Moreover, a discussion is provided on the usage, usefulness and current limitations of the developed framework. Finally, this chapter proposes some guidelines for Low Voltage distribution network operators, focused on the digitalization and reinforcement of LV networks.

6. List of publications

Journals

Chapter 4

E. L. Codjo, B. Bakhshideh Zad, F. J.F. Toubeau, B. Francois and F. Vallée, " Machine Learning-Based Classification of Electrical Low Voltage Cable Degradation," **Energies,** 14(10), N°2852, 10.3390/en14102852, 2021.

International conferences

Chapter 3

E. L. Codjo, B. Bakhshideh Zad, F. Vallee and B. Francois, "Analysis of Low-Voltage Network Sensitivity to Voltage Variations Due to the Insulation Wear," 2020 **55th International Universities Power Engineering Conference (UPEC),** Torino, Italy, 2020, pp. 1-6, doi: 10.1109/UPEC49904.2020.9209782.

Chapter 2

E. L. Codjo, F. Vallee and B. Francois, "Impact of the line resistance statistical distribution on a Probabilistic Load Flow computation," 2020 6th **IEEE International Energy Conference (ENERGYCon), Gammarth, Tunis, Tunisia,** 2020, pp. 637-642, doi: 10.1109/ENERGYCon48941.2020.9236570.

National conference

Chapter 2

E. L. Codjo, F. Vallee and B. Francois, "Impact de la distribution statistique de la résistance de ligne sur un calcul de Load Flow probabiliste," **Symposium de Génie Electrique**, Cité des Congrès de Nantes du 6 au 8 juillet 2021.

Chapter 1: Context and new challenges in LV networks

Contents

1.1. Introduction

The aim of this chapter is first to introduce the current context in LV electrical distribution networks. The general organization of electrical systems is recalled and gives basic information on distribution networks. Historically, LV networks have been developed for supplying electrical energy to consumers. Main features regarding standards for the operation are presented as well as the connection of new users in LV networks. No sensors were used in LV networks but the situation is changing with the introduction of smart meters (SM). An overview of the SM technology is then presented.

Another change is the arrival of more and more distributed generators. Therefore, new challenges for the operation of LV networks arise and the second part of this chapter explains the different issues. Planning and reinforcement decisions in distribution were classically decided by studying critical scenarios. Until now, these practices (so called "fit and forget" approach) enabled DSO to maintain the power quality with a limited observability on the dynamic state of electrical quantities. With the recent connection of new users such as electrical vehicle infrastructure, dispersed generation, distributed storage systems, the limited capacity and dynamic constraints of LV networks require an evolving from a passive towards a more active infrastructure and so a management procedure of the latter. This has led the path to what is commonly named the concept of smart distribution grid. Concretely, new controllable items can now be actively managed so that power quality issues can be mitigated in real-time. Also, planning and reinforcement procedures now need to take account for those Active Network Management opportunities as well as for the increased uncertainty brought into those modern distribution systems.

Additionally, problems are also coming from the network infrastructure such as degradation induced by temperature variations and a more active monitoring of the infrastructure is needed in order to anticipate outages that could prevent those active distribution systems from providing the electricity coming from widespread dispersed generation units.

To cover these problems, new technologies, such as smart metering devices, can be considered to pave the way for smarter LV networks.

In the last part of this chapter, a particular attention is paid to the state of the art in the LV cable condition assessment. The approaches recently proposed to study the modelling of cable degradations are explained. Various measurement data may be used to monitor the correct operation (and their recent applications) of distribution systems and are also exposed. In that way, Machine Learning methods that could be adapted for this purpose are listed and reviewed at the end of this Chapter.

1.2. Context

1.2.1. Organization of electrical distribution systems

The overall structure of a national electrical system is divided into three parts: the High Voltage (HV) transmission network, the Medium Voltage distribution Network and the Low Voltage (LV) distribution network. The HV network is the link between the large production plants, the large industrial sites concentrating a significant portion of electricity demand, the interconnections with other countries and the MV distribution network. The MV distribution systems (within a 20kV voltage level) ensures the electricity delivery from the transmission

network (through different substations) towards the end-users and usually consists of HV (High Voltage) / MV-LV (Medium Voltage – Low Voltage) transformer stations, overhead / underground lines, protection devices and interconnection equipment. The local consumers are connected to the LV systems with a 400V three-phase network or a 230V single-phase network. Moreover, the networks are tree-shaped / not looped with short feeders while there are radial in Belgium [9], [10] and [11]. Figure 4 shows the synoptic diagram of the deployment and operation of passive low voltage networks where the consumer is a fairly passive end stakeholder.

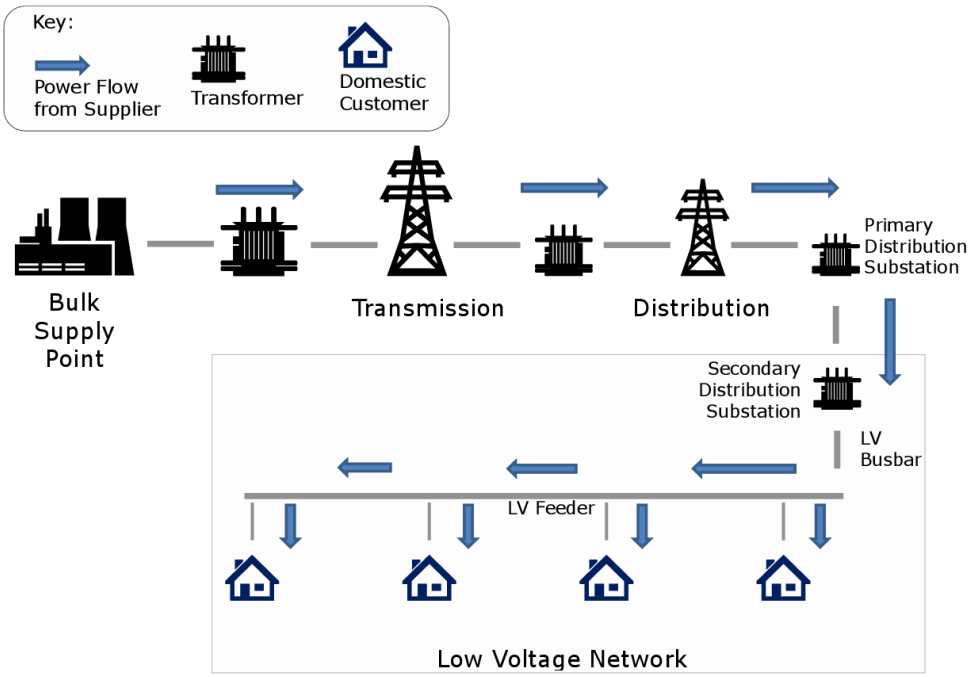


Figure 4: Passive LV network [12].

The different voltage levels are interfaced by transformers in power stations. The functions of these power stations are mainly the voltage transformation, voltage adjustment, distribution of energy flows and network protection. Power stations also contribute to the measurement of energy flows by metering equipment, tariff change using the centralized 175 Hz remote control, safety of the transmission network through the frequency-metric load shedding system and the continuity of the power supply by automatic reclosing systems [13].

According to the amount of delivered power, the different voltage levels offer, to the systems managers, the possibility to reduce voltage drops and losses (from high currents). The general architecture in terms of lengths and voltage levels results from a techno-economic optimization under sometimes compromises. Indeed, the distribution networks constitute usually the larger architecture of the electrical network. In France, over 622 187 km for MV and 701 858 km for LV are existing [14].

A system operator has two missions: the planning and the management of electrical networks. Network planning activities must anticipate changes in electricity consumption and production, in particular through the development of scenarios. These works help for the decision in the investment or capital expenditure (CAPEX) choices of electrical infrastructures. (multi year concession in France and Belgium) For an existing electrical network, normative prescriptions are essential for the proper operating, and must be strictly satisfied despite the arrival of new energy sources or new uses.

1.2.2. Standards in LV networks for operation and connection of new users

The supply of electricity must meet the requirements of standards and guides, which allow users to use electricity in complete safety with a good quality. In Europe, the standard EN 50160 [15] specifies the characteristics of the voltage supplied by public distribution networks within standardized operating ranges. In France, the NFC-13-100 [16] standard defines the installation rules at the interface between public and private networks; NFC-14-100 [17] defines the rules for the installation of public networks; and NFC-15-100 [18] defines the recommendations to be observed for domestic electrical installation. All these standards are detailed in guides, technical specifications or sheets developed by main stakeholders in the electricity sector [19, 20]. As consequence and example, the voltage level for a new user connection depends on his rated power (Table 1). Voltage and current constraints are the main factors involved in the sizing of network infrastructures.

Table 1: Reference connection voltage in France.

	Connection in MV (20kV)	Connection in LV
Consumer (*Pn*)	$Pn \leq$ min (40 MW, 100/*d* MW) *d* is the distance to the power station [km]	$Pn \leq 12$ kVA in single phase (limited to 6 kVA for producers) $Pn \leq 250$ kVA in three phases
Producer (*Pn*)	$Pn \leq 12$ MW Until 17 MW if allowed	

Voltage constraints

The RMS voltage (averaged over 10 minutes) at the connection point of a LV user must remain at all times within a range of ± 10% around the nominal voltage of the LV network (standard EN 50160). Before connecting an user, the Distribution System Operator (DSO) checks that the voltage variation between the MV / LV transformer and any point of the LV network does not exceed preset values according to the voltage at the secondary of the MV / LV transformer. A node is said to be under high voltage constraint (or overvoltage) if its voltage is greater than 10% above the allowable voltage range; in low voltage constraint if its voltage is less than 10%.

Current and apparent power constraints

The current / apparent power constraints are defined to satisfy the rated limits of the equipment and ensure a safe height for the aerial lines. The following definitions are used during the decision-making studies :

- A transformer is said to be overloaded when the apparent power fed through the latter is greater than its maximum apparent power.
- A conductor is said to be in congestion when the current passing through it is greater than its maximum permissible current for more than 10 minutes.

To ensure that the current and voltage constraints are met at any times, the DSO usually studies pessimistic scenarios of production and consumption in steady state (the so-called “fit and forget” approach), namely :

- maximum consumption and zero production in order to detect risks of constraint of low voltage and high current;

- minimum consumption and maximum production to detect stress risks of high voltage and reverse high current.

Also, regular maintenance operations based on (passive) historical procedures are carried out and are the main sources of operating expenses (OPEX). The energy transition leads to rethink these expenses and adaptations of electrical networks, with the emergence of new challenges (e.g. integration of high shares of renewable-based generation) and opportunities (e.g. towards a better knowledge of the system state through active monitoring) offered by the digitalization of the distribution systems. Some details are therefore provided hereafter about the technologies allowing this digitalization, namely: smart metering devices.

1.2.3. Overview on the Smart meter technology

Presentation

As explained in the general introduction, DSOs are pushed by the EU to install Smart Meters (SM). SM are electrical counters built up with an embedded communication device. This electronic device is able of recording electrical measurements (such as voltage levels, active and reactive powers at the coupling point) and to communicate each day the recorded information to a hub (or data concentrator).

Illustrative SM example : Linky

As a good example of the current trend towards digitalization of modern distribution systems, Linky is a smart meter widely implemented in the French distribution network (fig. 5). It records the cumulative electricity consumption on various time "indexes" and returns them daily to ENEDIS via the electrical network (with a Power Line Communication). Technically, Linky measures consumption over 10-minute time slots and stores the corresponding information locally, i.e. 144 metering data per day in a memory. The choice of this recording frequency is based on considerations related to the volume of information to be stored. In Belgium, 15-minute time slots are implemented (to be in accordance with electricity market time slots).

Ten indexes aggregate the consumption over (at most) ten different time ranges; this allows ENEDIS to transmit this information to suppliers on a monthly basis for customer invoicing purposes. The choice of "only" 10 "supplier" indexes aims to reduce the size of the data storage infrastructure at the DSO (15,000 servers at ENEDIS) [21]. Linky is also a suitable tool, within certain limits, to take into account different pricing policies depending on the supplier, and in particular “dynamic” prices.

Four "network" indexes are issued from the national TURPE tariff ranges (Public Electricity Grid Use Tariff).

Four "production" indexes, one to measure the energy injected by the consumer and the three others used by the DSO to monitor and to manage its network.

The Linky meter also receives real-time information sent by the DSO, in particular when switching between tariff ranges. These changes lead to index changes but can also order (control) switches. There are an hardware switch (that may control the water heater or other loads such as the electric vehicle) and seven "virtual" switches. Virtual switches are coded information (0 or 1) available on specific outputs of the meter, which therefore make it possible, via a decoder, to control breakers according to an on / off logic.

For the customer, a local bus (‘tele information”) is provided to know in real time the instantaneous RMS value of the current (rounded to the integer value, calculated during a 1 s period) and the apparent powers in consummation and in injection [22]. The transmission is an asynchronous serial communication according the standard IEC 62056-21 clause 5.1 with a transmission speed (baud rate): 6900 +/- 1%.

(a) (b)

Figure 5: Example of popular smart meters: (a) in France ; (b) in Belgium.

Surrounding ICT

As shown in Figure 6, an advanced metering infrastructure (AMI) comprises smart meters, hubs or data concentrators, communication channels and data management systems.

Hubs collect metering data from a cluster of surrounding electric meters communicating by a Power Line Communication (PLC, wired in AM at 50 kHz) and transmit the collected data by GPRS (3G or 4G networks) to a head-end server that executes an IS (Information System). According to the communication networks and associated infrastructure, it takes several hours to get data from all smart meters to the DSO control center. In the reverse direction, it is able to transfer back remote operation requests from the information system to the meters. It also monitors the voltages of the LV distribution network at the MV/LV power station (because hubs are usually located there) and perform the interface with other network devices (eg: Fault Detectors).

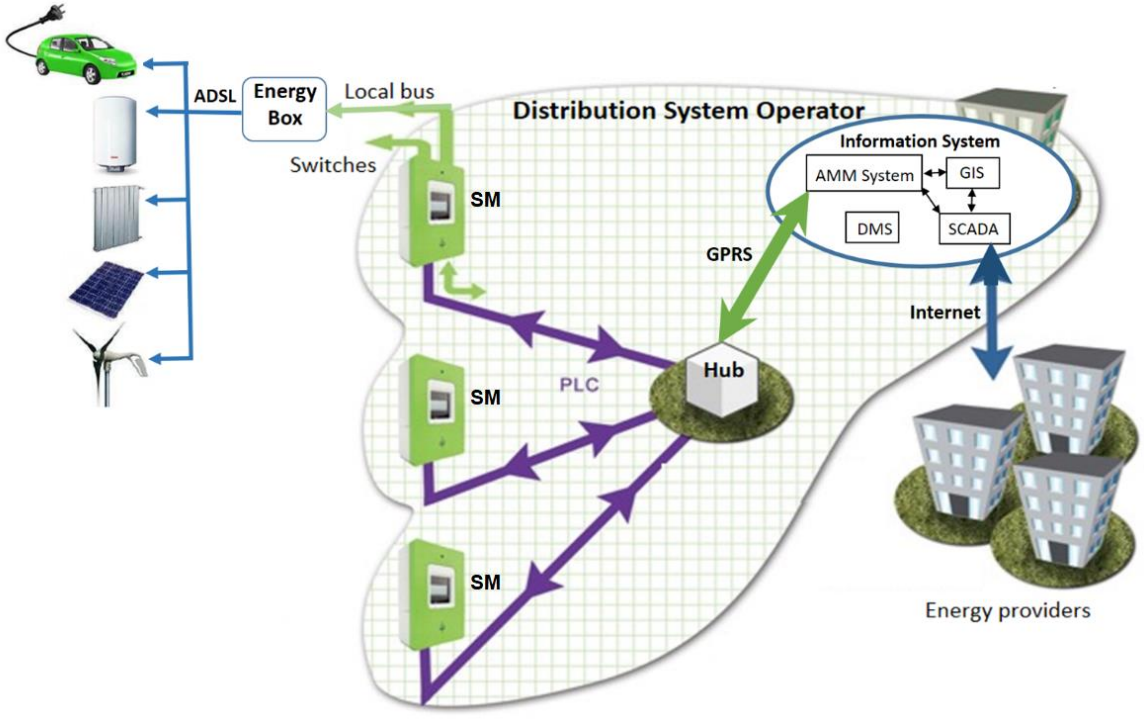


Figure 6: General scheme of a communication infrastructure for metering data transmission.

In the central information system, data coming from the concentrators are received by an Advanced Meter Management (AMM). The AMM performs data acquisition and real-time processing and stores the measurements of the meters, their current status, information relating to the quality of energy, the operational status of the various elements of the system (switches, sensors, etc.), knowledge of the equipment having known failures, and various information about the services of third parties [23]. The Geographic Information System (GIS) stores and processes all types of geographic information via mapping, statistical analysis and database technologies. The GIS provides the SCADA with a topological description of the network and the geolocation of its various equipment.

The Supervisory Control and Data Acquisition (SCADA) makes the link between the physical quantities of the electrical network and all the communication and information functions to make it work. From AMM and GIS data, SCADA processes the information and performs network maintenance operations by sending instructions to the various communicating elements of the distribution network such as protection devices, network equipment or decentralized production. To make decisions, the DSO is helped by a set of functions resulting from traditional methods (load flow, protection plan , etc.) and so-called Advanced Driving Functions (ADF) methods.

These new functions are stored in the Distribution Management System (DMS). These ADF are achievable through new smart sensors, new processing methods of large amounts of information, and new software for real-time or offline applications. These functions enable a better knowledge of the network: state estimate (better confidence despite noise, measurements missing or inaccurate), load distribution calculations, consumption forecast (predictive models, fuzzy logic, neural networks). The network is also more controllable: As example, the Volt Var Control (VVC) supports the conventional adjustment means of the voltage by managing injections / absorption of powers.

As shown, the initial and main application of the SM is for billing purposes but as an important amount of electricity consumption and injection data are collected, they are usually also used for monitoring and /or for a time analysis of the load demand. The large and non-synthetic databases coming from SM measurements bring out question about how to further exploit and valorise those available information. This is especially interesting for Low Voltage distribution networks because until now they do not have effective monitoring devices. In the future active LV networks, smart meters can be used for improving energy consumption forecasting, tracking power generation as well as improving and making more active the diagnosis assessment procedure of the physical infrastructure.

In conclusion, one can note following findings:

- The daily communication of measurement data is well adapted for the implementation of a dynamic tariff with various energy providers but is not suited for fast applications because of significant time delays between smart meters and the information system.

- The rms value of the voltage is not transmitted by smart meters but can be available at the hub location (usually at the MV/LV power station).

- Consumed and injected energy over 15-minute time slots are available from Belgian smart meters while being available on a 10-minute basis in France.

- As a significant time delay exists to get measurements from all smart meters, a real time monitoring (as AC voltage control…) is not feasible with this technology. But smart meter measurements can be used for the monitoring and diagnosis of the state of health of the network.

1.2.4. Distributed generation in existing LV systems

Analysis of the DG growth with installed capacity

In recent years, the generation from renewable energy has experienced a significant development with a global capacity reaching 2799 GW and 609 GW in Europe at the end of 2020 [24]. The objective behind the European Green Deal is to become the world first climate-neutral continent by 2050 (and to cut, by 2030, net emissions by at least 55%). The renewable energy counts for 38% of the energy consumed in the EU in 2021 (compare to 19.7 % in 2019) [25, 26]. Solar and Wind energy have reached respectively 9% and 15% of Europe electricity generation mix. The figure below sums up the annual capacity increase of solar and wind generation in European Union through the recent years.

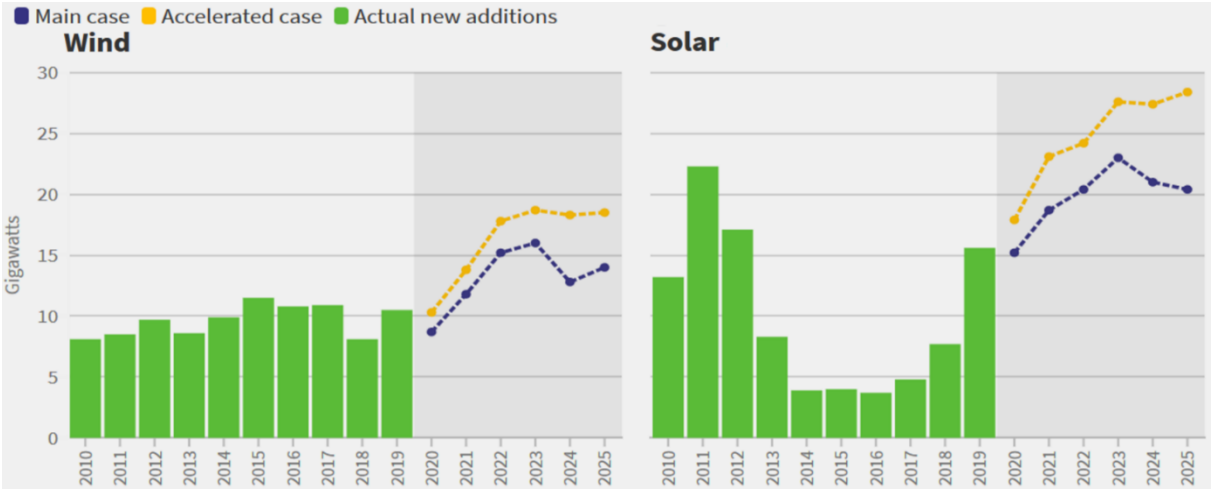


Figure 7: Annual capacity increase of Wind and solar installations in European Union [26].

Moreover, governments are setting up financial rules to help users wishing to own any decentralized production system for personal uses. So, they can inject this electricity into the network and selling it to a buyer at a price fixed by the law (feed in tariff) [27, 28]. Hence, since the last decade, the distribution networks have welcomed more and more producers based on renewable energies (RE) whose installed capacity is between a few kilowatts and about ten megawatts. Usually, they are connected to the nearest electrical substation/feeder after applying an hosting connection process for checking the future current operation.

The main sources of RE developed today are (Figure 8):

- Photovoltaic generation, predominant in terms of number of installations,

- wind power plants, predominant in terms of installed capacity.

In addition to renewable energy sources, there are cogeneration system in the network. Cogeneration is a technique in which two different forms of energy are produced

simultaneously, both electricity and heat, within a single production unit. Some examples are the Stirling engine, the micro-generation boiler, the DHW (domestic hot water), etc.

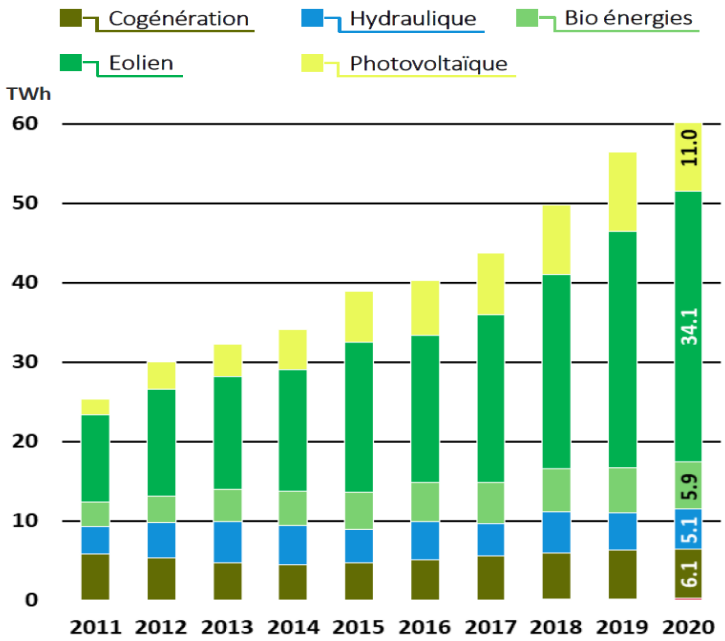


Figure 8 : Energy production in the ENEDIS distribution network [29].

Hosting process for new connections

In conventional LV networks, each local Distribution System Operator (DSO) has the major role of applying standards for the safety and the quality of the energy delivered. Before a LV connection, the DSO checks that the connection of the new user satisfies the transit capacities of LV network structures and voltage ranges at all times allowed on the LV network. When the new user is a producer, the DSO also checks the withstand voltage, direction of transit and maximum power at the distribution station in worst case conditions [30].

If no constraint violation is detected, the user can be connected to the network immediately. Otherwise, the DSO can solve some overvoltage issues by lowering the voltage plug of the MV / LV transformer (manually and under no supply) or reinforce the part of the network under stress (change of transformers, change of lines/cables, new electrical paths), that means modify the network impedances. These solutions are relative to the design and sizing of the distribution system. They are not suited for a large development of DGs because

- the plug of the MV / LV transformers can be only operated manually and exclusively when the power is off, then the modification of the voltage reference point is exceptional,
- reinforcement studies do not take into account the intermittency of most renewable energies connected to distribution networks.

1.3. New challenges for the operation of LV systems and Smart Grid opportunities

1.3.1. Incoming problems from new users

New DGs in LV networks introduce new problems because of their geographical dispersion and their intermittent nature.

The first change is that the connection of small sized Distributed Generation (DG) creates new bidirectional flows of the electricity in LV networks when the local produced energy is exceeding the consumed one. The protection plan should then be revisited with new settings for relays and breakers. Another problem is the possible increase of the voltage beyond the allowed limited. Those issues related to voltage control will therefore be discussed later.

As they are powered by a primary renewable energy, classically, wind and photovoltaic production systems do not regulate their output produced power and so, for the distribution system operator, are not controllable. Two characteristics have to be managed. Firstly, the variability of the power flows requires more dynamical regulation process in distribution networks. Secondly, the imperfect production prediction requires reactivity in case of unexpected situations onto occurring constraints.

In addition, new uses, including electric vehicles and heat pumps, will change the consumption profiles, these unknown characteristics complicate the planning and management for TSOs and DSOs. These uses can indeed create potentially critical consumption peaks for networks if too many loads of this type are connected to them simultaneously. In addition, it can be noted that new ways of consuming electricity are appearing, with the gradual development of collective or individual self-consumption business models, which again modify the consumption curves of users without having sufficient visibility on the latter for the time being.

1.3.2. Problems from the infrastructure point of view

The performances of the electrical network infrastructure is evolving in time and so should be monitored to prevent localized failures that can lead to possible more generalized failures (partial blackouts). Two causes can be identified when talking about physical infrastructure degradation: The ageing of materials and the climate change. In general, climate change has an impact on the environment conditions of electrical networks, since extreme temperatures (hot or cold) are likely to be reached more often [31], and soil characteristics (drying out, for example) are also likely to vary [32]. It then becomes necessary to take them into account in the management and planning procedure of modern power systems.

1.3.3. Technologies for a smarter LV network

In this context, the TSOs and DSOs must succeed in finding a compromise between the necessary adaptations to these new uses and compliance with existing quality standards in terms of supply and distribution of electricity (see part 1.2.2). To this end, new technologies, as the monitoring of renewable energy generation, integration of controllable loads and distributed storage, … are tested and implemented in order to enable a more precise and real-time control of electrical installations. With these hardware actuators, the distribution systems will require more intelligence for the network management and may enable a greater optimization of the investments made in the networks (also with regard to the carbon footprint). Hence, the passive conventional networks are gradually being transformed into a more dynamic, more active and more intelligent one. In that way, smart distribution grids (SDG) play both the role of an electrical feeder and economic exchange support for the production and the consumption of electricity.

This PhD thesis falls within this framework that aims to find opportunities offered by digitalization at the service of electricity networks subject to a major energy transition. To that end, some possible issues that could be dealt with, namely voltage control and active diagnosis assessment of the physical infrastructure, are highlighted hereafter.

1.3.4. Issues on voltage control

The increased penetration of distributed energy production resources within the existing distribution systems leads to technical challenges related to important voltage variations exceeding the imposed standard limits. This situation can lead to damage the consumer's equipment or the network components. Voltage control is becoming then, very critical and crucial to manage. Therefore, voltage control is, so far, crucial for the safety of those networks as well as for minimizing the electrical losses.

Dynamic regulation of the voltage

The quick increasing penetration of intermittent renewable energy resources, such as photovoltaic (PV) and wind generation, brings emerging operational challenges to the control of the voltage. To analyze this phenomena, let us consider a voltage source (the voltage is set at the MV/LV transformer) supplying a customer with a line, which is modelled as a R_2, X_2 series circuit.

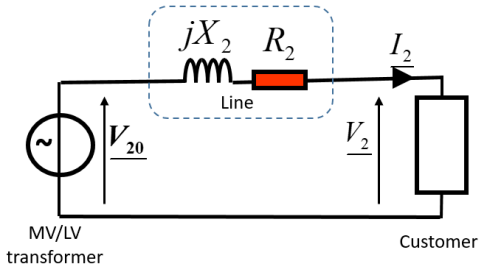


Figure 9 : Electrical scheme of a voltage source supplying a customer.

By considering a lagging angle φ of the customer current within the applied voltage (V_2), the following vector diagram can be drawn on Figure 10.

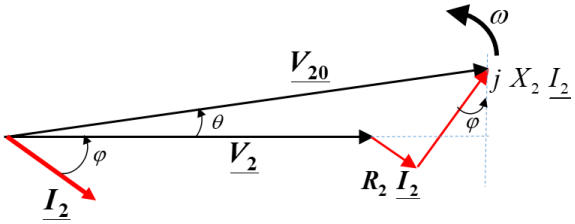


Figure 10 : Vector diagram of the electrical circuit in Figure 9.

Projections of voltage vectors into an orthogonal frame yields to the following scalar equations:

$$V_{20}\cos(\theta) = V_2 + R_2 I_2 \cos(\varphi) + X_2 I_2 \sin(\varphi) \qquad (1.1)$$

$$V_{20}\sin(\theta) = -R_2 I_2 \sin(\varphi) + X_2 I_2 \cos(\varphi) \qquad (1.2)$$

In practice, impedances are small and so does the angle θ :

$$V_{20}\cos(\theta) \approx V_{20} = V_2 + R_2 I_2 \cos(\varphi) + X_2 I_2 \sin(\varphi) \quad (1.3)$$

$$V_{20}\sin(\theta) \approx 0 = -R_2 I_2 \sin(\varphi) + X_2 I_2 \cos(\varphi) \quad (1.4)$$

So the voltage drop can be calculated by :

$$\Delta V = V_{20} - V_2 \approx R_2 I_2 \cos(\varphi) + X_2 I_2 \sin(\varphi) \quad (1.5)$$

Considering that powers (P_2 and Q_2) at the customer connection point are known (measured or calculated), and under a single-phase network assumption, we establish that the voltage drop depends on these powers:

$$\Delta V = V_{20} - V_2 \approx R_2 \frac{P_2}{V_2} + X_2 \frac{Q_2}{V_2} \quad (1.6)$$

Moreover, LV lines have mainly a resistive impedance combined with a limited exchange of reactive power, in consequence, we get:

$$\Delta V \approx R_2 \frac{P_2}{V_2} \quad (1.7)$$

Consequently, when a DG unit injects active power at a network node, the voltage at the point of coupling is rising (with an impact on the neighbouring node voltages). Distributed Generators can then completely modify the network voltage profile among lines, cause over-voltages in the area close to the DGs (and so not detectable at the HV/MV power station or MV/LV transformer) and introduce significant uncertainties in the operating planning of the distribution systems. In addition, one of the most significant changes in the network operation is the need for a fast voltage regulation (with a required fast time response), which can accurately follow the voltage variations associated with the intermittent operation of renewable energy based DG. The traditional voltage setting procedures with existing assets are therefore not suitable to cope with the massive arrival of DGs. Therefore, a new setting is mandatory to ensure a regulated voltage level within the allowed limits. These conclusions are shared in several studies of the scientific literature on the impact of decentralized generation in distribution networks.

On line monitoring of the voltage

Moreover, the modeling approaches known until now for LV networks face a critical lack of adaptability to the structure of Smart Grids (SGs). During the past years, the LV network modeling approaches was based on Load Flow computations using typical load patterns, due to the lack of real-time monitoring. Thereby, to ensure a cost-effectiveness planning management, worst-case situations for the network control are considered, as representative scenarios, by the DSO. The fluctuation of renewable energy sources now increases the volatility and level of uncertainty in the operation of smart grids. To ensure the network stability and the detection of voltage peaks, the previous modelling approaches need to be updated to probabilistic load flow calculations where operational constraints and congestion probability are taken into account. Moreover, considering the constant variability of the LV network state caused by residential load variations and distributed PV generation, the modelling approaches need to be renewed. Thanks to the large deployment of SM, new modeling tools based on multidimensional

approaches are emerging taking into account the impacts of uncontrollable consumption loads and intermittent distributed photovoltaic generators (DPVGs) sources.

1.4. State of the art on LV cable condition assessment

1.4.1. Scope of the literature review

As already stated, Low Voltage (LV) distribution networks are the last stage of the electrical power network, which supplies many dispersed small-scale loads. Those radial networks consist of a set of equipment such as MV-LV (Medium Voltage - Low Voltage) transformer substations, overhead/underground lines, protection systems, etc. LV feeders are designed to feed a limited amount of end-users in order to reduce the influence of a localized interruption. In case of fault in a LV network, a few of customers is, indeed, disconnected. Thereby, either LV interruption problems or degraded operations of LV equipment (such as the cable ageing and deterioration) have received less attention until now. On another hand, due to insulation failure (related to ageing and degradation), the insulation material can badly become more conductive and less effective for preventing the leakage current between the conductors. Moreover, with the increased penetration of DGs, distribution systems become more and more active stakeholders of the energy transition and localized interruptions could have an increased impact on the global security of supply and on the business activity in the coming years. This leads to the necessity of providing a special attention to the diagnosis assessment of the LV network physical infrastructure

Therefore, this section is focused on investigating the literature review related to the LV cable condition assessment as well as on the use of smart metering and intelligent techniques application for an enhanced management of smart distribution grids.

1.4.2. Modelling of cable degradation

The French standard NF C 15-100 (harmonized with European standard HD 384) specifies that the insulating material of LV electrical cables must oppose to the current all along the conductor [33]. In fact, deteriorations of the insulation material can increase the discharge of leakage currents in the cable insulation, which can lead to overcurrent and voltage variation issues, so it can have adverse impacts on the efficient operation and the safety of the network. In addition, LV distribution networks (initially designed for unidirectional power flows) are currently subject to bidirectional power flows and frequent voltage variations arise from the constant growth of decentralized photovoltaic (PV) installed capacity. The voltage variation problems [34-36] tend to accelerate the current leak through the insulation material of LV cables. Monitoring the insulation material degradation becomes then a relevant issue.

Investigations have been carried out for modelling this insulation degradation through its resistivity variations. Also, the impacts of physical and electrical phenomena on the insulation (such as stress in applied voltage) and on the line resistance have been analyzed. By studying the degradation of high-voltage cables having PVC (PolyVinyl Chloride) as the insulating material, the author in [37] has shown the impact of environmental conditions (phenomena such as temperature and humidity variations) on the ageing of this insulation material. Also, in [38], the authors developed a test procedure for electrical cable insulation material.

Authors in [39] have evaluated the insulation resistance characteristics of LV cable under extreme experiments (such as flame contact, over-current, and accelerated degradation). The studies [38] and [39] reveal that the network external operating conditions directly affect the insulating properties of the material. According to their obtained results, a 10°C increase in temperature seems to be doubly inversely proportional to the insulation resistance value. Reference [39] shows that accelerated degradation in the LV cable significantly reduces its equivalent life since it reduces the insulation resistance of the cable.

In the Netherlands, research works [40-42] investigated underground LV cables with a focus on jacket damage. Researchers in [40] worked on how water ingress can progressively degrade the LV cable. To do this, the experimental study was carried out with two different plastic insulated cables (Cross-Linked Poly-Ethylene (XLPE) and PVC) artificially damaged by drilling a 8 mm hole into each cable. The cables have been tested in water exposure conditions and gradual degradation linked to partial discharges has been observed in the insulating material after any water evaporation phenomena. The different experiments showed that at a sufficient degree of degradation, breakdown can occur in the PVC cable (due to leakage currents) while the XLPE cable was still under operation. Indeed, the PVC by decomposing produces hydrogen chloride, which makes the water more conductive. In [41], the same researchers made a comparative study on how polymeric insulation materials can affect degradation growing in LV Underground Cables. That study pointed out the major role of the insulating material decomposition in the cable degradation rate, regarding its chemical composition. From this aspect, PVC material has led to a significant degradation unlike PolyUrethane (PU) and XLPE. In Hungary, references [43-45] have studied the effects of Distributed Generation (DG) on the ageing and degradation of PVC insulation. Firstly, the work was focused on the thermal ageing of LV cables. Reference [43], by implementing a periodic thermal ageing test, has shown an inverse correlation between hardness and the conductive properties of the cable. In [45], the authors have tracked the thermal ageing of the PVC insulation material under various temperatures and different plasticizer contents. The results reveal the best PVC specimens for the monitoring and characterizing thermal ageing in PVC insulated cables. However, these studies only focus on the physicochemical properties of the insulation material of cables.

1.4.3. Monitoring applications with measurement data

Authors in [46] show the benefits of voltage measurements with smart meter by using different literature criteria. In that way, they review smart meter data applications in a context of improving the DSOs network operation and planning.

By investigating the LV system state estimation using smart meters, the study in [47] concludes that the knowledge of power consumption and voltage at all nodes is the best combination for performing a state estimation. The loss of data in real time is also considered by using pseudo-measurement data from the historical database.

The work in [48] addresses the problem of fault detection in underground electrical cables and overhead transmission lines. The work focuses on a 15km – 132 kV / 50Hz HV line (3 km in underground and 12 km in overhead) connecting a source to a RLC load. After modelling and simulation of the power system (due to the lack of Smart Meter data), the temporal values of measured voltages and currents are filtered and digitized to obtain discrete values that are used as inputs to an ANN (Artificial Neural Network) algorithm. However, this fault detection

technique does not take into account the physical behaviour of the conductors. Talking about power cable, the authors in [49] analyse the damage and the life expectancy of electric cables in Low Voltage overhead distribution networks. By studying the Stresses-Strains of the non-isolated driver, they show its incidence on the cable resistivity. That article is useful for understanding how the ageing of the cable affects its operation. However, to improve the quality of estimations, it is essential to know the network architecture under study and the parameters of the lines. This parameter estimation has been the subject of several works including those of [50]. They present an approach for the estimation of line parameters under consideration of measurement tolerances with the network state estimation as an end-purpose. As we will do later (section 2.3 and 4.2), they start with the modelling of the electric lines with a π-line model. Then they add some constant scaling errors to their power flow results to represent the measured signal errors (on their ideal values). After studying the parameter sensitivity to the previous error, they try to calculate compensation factors by solving a LSQ (Least Squares Optimization) minimization problem.

Based on an opposed approach, the study carried out in [51] has explored the physical degradation of the cable through the sensitivity analysis of LV network voltage variation to various degrees of cable insulation wear. That work lies in the uncertain nature of the position and degree of insulation degradation of the cable. In that context, a Monte Carlo (MC) based method was used to characterize the unknown variables. In the literature, the MC technique has been widely used in order to analyze the uncertain and unobservable nature of the distribution systems. The building of scenarios with a MC technique has proven their effectiveness either for the uncertainty impact assessment associated as in [52-53], or for characterizing the Low Voltage distribution systems (in a sensitivity analysis context) as in [54]. Moreover, the uncertainty impacts associated to the electrical network parameters on the nodal voltages have been evaluated in [55, 56]. Furthermore, reference [57] investigated whether and when alternative maintenance strategies, using historical data, would be more profitable than the currently used corrective maintenance.

1.4.4. Machine Learning application in LV networks

Recently, Machine Learning (ML) techniques have been studied for fault detection in [58-65]. The researchers in [58] set up a tool for the transformation of measured data into a model of failure prediction (failure classification, MTBF estimates - Mean Time Between Failure) in order to meet the requirements for preventive maintenance. The problem is approached from a general point of view, starting from the failure prediction at each component of the bus, to the overall prediction of the network failure probability using Machine Learning Ranking methods. After the implementation and the effectively failure prediction, a classification algorithm is developed for a budget-based preventive maintenance planning. The study in [60] addresses the benefit of a Machine Learning framework for fault detection and classification in power systems. By analyzing the mostly used ML techniques (within consideration of fault types and metrics for those techniques evaluation), the authors have shown the benefits of supervised classifiers in the reliable solving of power system problems. In the same way, a part of the research in [61] was dedicated to the fault diagnosis in LV networks by using a deep learning approach. The results of this study allowed the authors to highlight the most influencing parameters in the fault assessment process such as the fault resistance. In a context of grid monitoring, authors in [63] set up a Power Line Modems (PLM) based solution for the diagnosis

of a distribution network cable. By implementing various ML algorithms (combined to several pre-processing methods) the proposed approach ensures to employ the best algorithm for a given diagnostic procedure. The work has been oriented through a two-stage approach from the degradation detection to the ageing and localized degradation assessment of a XLPE insulated cable based. The key point of this approach relies to the access to the PLM database. Authors in [65] investigated the role of ML in integrity analysis of subsea cables. From the design of a Low Frequency (LF) sonar system to the detection of the cable degradation stage through an accelerated life cycle testing, their study provides a library of LF sonar responses depending on the cable types and conditions.

Regarding voltage issues in the distribution network, researchers in [66] have worked on a centralized voltage control framework within consideration of the uncertainties related to the network working conditions and its physical parameters. Those last ones were the dependency between the temperature variation and the line resistance, the internal resistance of the transformer and the consideration of the shunt admittances of power lines by using a PI line model. The author has implemented a fast decision-making method, which is cost-efficient since the deep reinforcement learning-based agent can automatically adapt its behavior under varying operating conditions. The above ML-based studies give acceptable accuracy results with a good speed and a low calculation burden. However, they do not integrate the assessment of the insulating material properties of the LV network cables associated to their growing insulation degradations.

1.5. Synthesis

The above literature review can be considered as the foundation of this research project. The global objectives have always been to promote the use of SM measurements data at an intermediate stage of network management while overcoming the lack of LV systems monitoring information. On a first hand, studies about modelling LV line in degraded conditions only focused on the physicochemical properties of the insulation material of cables. The main difficulty for considering the insulation material electrical properties lies in their uncertain nature since the position and degree of the insulation degradation is not well known. In this context, one contribution of this research lies in characterizing the uncertain and unobservable nature of the distribution systems cable conditions. On another hand, it will be interesting to investigate the integration of those ML tools in the LV cable condition assessment, using the available network data (no need of extra costly IT devices in the network). Hence, the resulting novel contribution will reside in its proposed learning-based framework to identify / classify the cable lines that present an insulation degradation, relying on nodal voltage and net demand variation profiles of the distribution network. Moreover, the below works do not claim to have explored all the aspects related to the LV network cable conditions assessment. Its brings out some relevant methodological / theoretical contributions to solve the above problems as well as some guidelines for proving the interest of monitoring distribution networks and for showing to which network / degree of degradation (qualitative analysis) these measurement database will be useful.

1.6. Conclusion

In this chapter, we have seen that electrical networks are called to respond in next years to new challenges, which can be summed up by the arrival of new sources of energy production, new energy requests and new load demand characteristics, all in a context of climate change. These new uses, usually connected in single-phase, have an impact both on network losses, but also on compliance with the EN 50160 standard, particularly on voltage limits. Moreover, until now, LV networks, unlike other networks, had few sensors and were therefore monitored without observability because the majority of users were consumers whose sizing load models were fairly well estimated. However, new uses, whose consumption and production profiles are still unknown or imprecise, and the development of new self-consumption habits, are making these models questionable. It may become difficult to maintain a quality of supply without oversizing some parts of the electrical networks, which would increase the CAPEX and OPEX of DSOs.

In a very large number of countries, smart meters (SM) provide large amounts of data. In France, in particular, the Linky program initiated since 2015 is an opportunity for the monitoring and management of these networks, which until then had no observability. A review of SM technologies has been written.

Technical performances of the operation and transmission of these data let us imagine applications in the topic of the monitoring of the LV network and LV cable condition assessment. So, a state of the art has been established to emphasize the digitalization of distribution systems for an active and non-invasive assessment of its physical infrastructure. In the next chapter, a deep analysis is driven on the variation of cable characteristics and on the network operating regarding extreme temperature variations.

Chapter 2: Data based analysis of the resistance variability distribution in LV networks

Contents

2.1.Introduction

As Smart Meter (SM) devices are now largely installed, this chapter explores the development of some applications with these available data. The studied LV network is firstly presented with a focus on the considered architecture. Because technical characteristics of LV networks are usually badly known, a first interest is to use data from SM for modelling the electrical network and the next section presents a data-based modelling. It consists in recovering the network architecture or proposing an equivalent network structure that satisfies measured operating points by SM. By considering impedances between nodes, a model of the architecture is deduced. Then, a method is proposed to identify impedances that are considered as unknown model parameters.

The operation and use of the electrical network are involved in the ageing and degradations of electrical hardware. Studies (such as references [34] and [55]) showed that some climatic impacts (such as heat) modify also significantly the network impedances. With the recurrent temperature elevations of the last decade, the electrical networks are facing some increasing temperatures, over nominal values, which have been assumed when sizing the electrical infrastructures in the past. Assumptions on the temperature, adopted a few decades ago when sizing LV electrical networks, do not offer now a sufficient margin of safety with the global warming. These worrying phenomena (such as heatwaves) have also led to numerous power cuts in California, Australia and Brazil. The temperature impact on the modelling of LV cables is addressed in the last part of this chapter as well as the characterization of the resulting operation of the LV network in such situations.

2.2.Presentation of the monitored low-voltage distribution network

In the framework of a pilot project in 2013, the local distribution system operator ORES has installed 600 smart meters in Flobecq, in Hainaut (Belgium). The measurements comprise energy consumption, energy injection and PV generation values that are recorded and integrated over ¼-hourly periods. Data are locally saved and sent, remotely each day, to the DSO Information and Telecommunication systems. For testing and developing data based models and applications, we have used data from a part of this single-phase LV network [51, 52], where each customer is equipped with a Smart Meter (Figure 11).

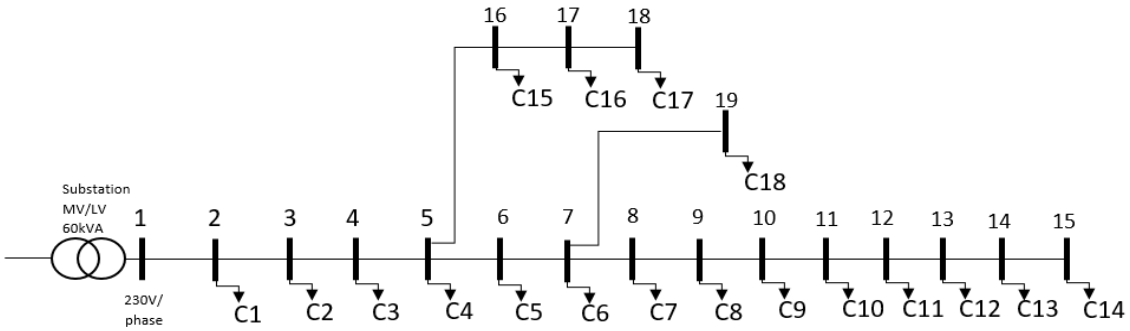


Figure 11: Topology of the monitored LV distribution network.

This radial topology consists of 18 consumers (*Ci*), these residential loads are spaced according to lengths given in Table 2 (see Appendix A for the complete technical parameters).

Table 2: Technical parameter (length) of the above LV distribution network.

Line between :	Length (in m)	Line between :	Length (in m)
Node 1 and Node 2	46	Node 10 and Node 11	44
Node 2 and Node 3	273	Node 11 and Node 12	10
Node 3 and Node 4	62	Node 12 and Node 13	15
Node 4 and Node 5	63	Node 13 and Node 14	14
Node 5 and Node 6	194	Node 14 and Node 15	41
Node 6 and Node 7	26	Node 5 and Node 16	396
Node 7 and Node 8	11	Node 16 and Node 17	93
Node 8 and Node 9	25	Node 17 and Node 18	117
Node 9 and Node 10	21	Node 7 and Node 19	119

In this Thesis, data from SM are used to apply machine learning methods and compare their performances for developing applications such as monitoring of cable ageing, operation, RMS voltage surveillance. With this LV network, different scenario of PV connection at nodes will be also anticipated and considered to estimate the impact on the grid operation.

Working assumptions :

From the above network some assumptions have been made as followed :

- The Low Voltage network is considered balanced during the following studies.
- Due to the short distances (short cable length between nodes; Table 2), the shunt admittances (capacitive phenomenon) can be neglected (see section 2.4.3.1).
- The voltage at node 1 is assumed to be constant and imposed by the substation transformer (see section 2.4.3.2).
- The reactive power is computed by assuming a 0.9 Power Factor value as prescribed in standards (see section 2.5.3.2).
- The cable loses a part of its insulation thickness in case of degradations. Then the cable radius R will be reduced (as well as the resistivity coefficient ρ) while the conductor radius r will remain constant (see section 3.1).
- The insulation radius of the cable is assumed to vary between 0.01 mm and 1 mm while the nominal insulation radius (from the datasheet) of the studied cable is equal to 1.5 mm (see section 3.5.1).

2.3.Model of the power balance at each node

At each quarter of an hour, the SM simultaneously records power consumption (*Cons*) or power injection (*Inj*) and PV generation (*Prod*). Knowing the measurements, the load demand (called *Load*) of the customer at a node q is expressed as in [34] (Figure 12):

$$Load(q) = Cons(q) - Inj(q) + Prod(q) \quad (2.1)$$

where *Cons*, *Inj* and *Prod* are respectively the recorded positive power consumption, positive power injection and positive PV production.

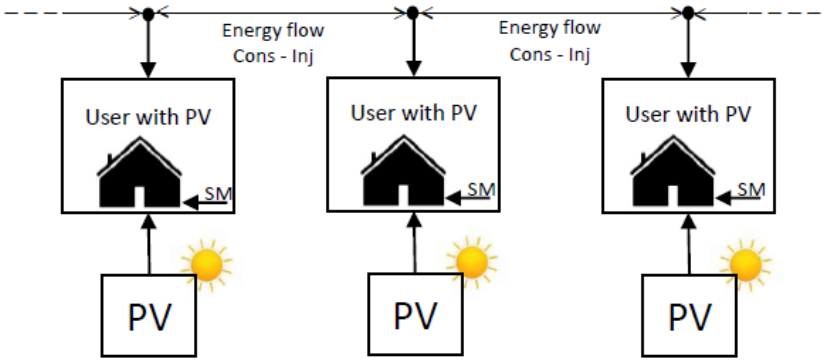


Figure 12: Power flow representation at the customer network node.

The involved network electrical real power *P* is the one at customer node is:

$$P(q) = Cons(q) - Inj(q) = Load(q) - Prod(q) \quad (2.2)$$

The reactive power is computed by assuming a 0.9 Power Factor value as prescribed in standards [67]:

$$Q(q) = P(q) * \tan(\cos^{-1}(0.9)) \quad (2.3)$$

These powers can be positive (if consumption > injection) or negative (if consumption < injection) and will be used later in the Load Flow calculations. Figure 13 shows the time evolution of the exchanged power at node 14.

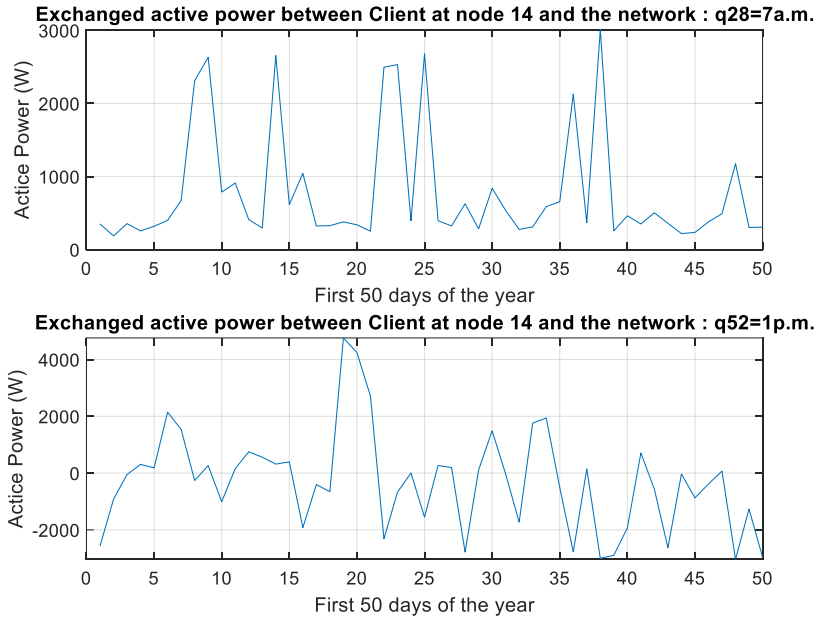


Figure 13: Exchanged active power at node 14.

2.4. Data based modeling of a LV network

2.4.1. General method

Technical characteristics of LV networks are usually poorly known because most of their construction and also evolution (or reinforcement) are performed without any written record. In consequence, the known topology can be wrong, or partially/totally unknown. Based on Smart Meters data, the task of this part is to present a method to automatically:

- identify the topology
- estimate the network impedances

The easiest and more efficient way to recover the topology is to use the knowledge of voltage measurements at each node. The basic strategy is that the more similar two bus voltages are, the more likely they are directly connected. Unfortunately, this test can consider similar voltages on few time steps without being related. So it is very important to consider a large number of samples to reduce these cases among all test results.

In radial networks, each node is fed by only one other node (but can feed several nodes). Therefore, when going from the transformer's node to the end-user's node (figure 1), a new node is connected to one and only one parent node. Hence, the proposed algorithm will start with measurement data from the transformer (node 1) and will find other nodes by analysing the most similar voltages until the last node (the end-user's node).

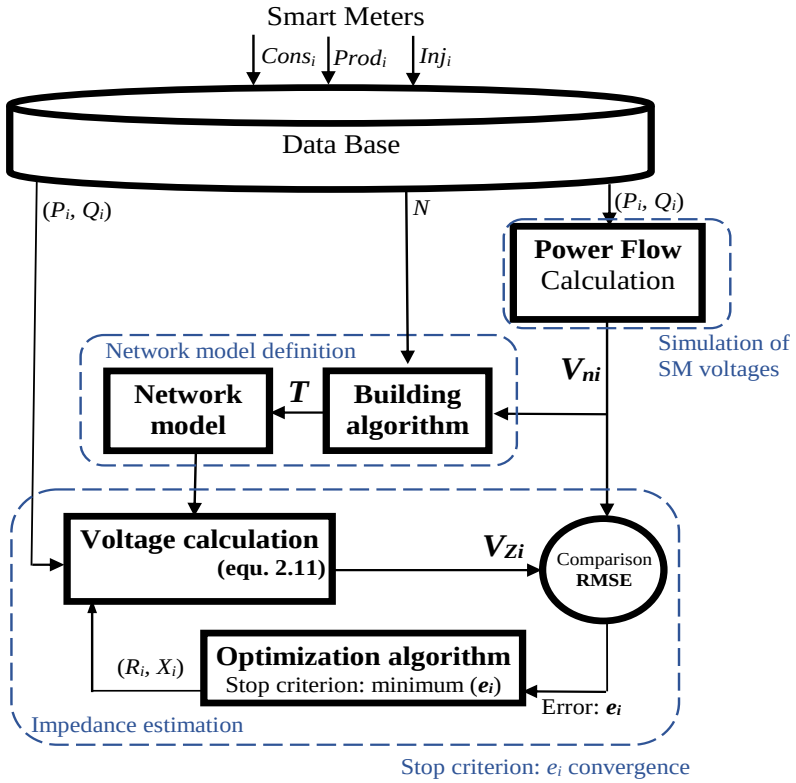


Figure 14: Synoptic diagram of the data based model.

2.4.2. Building Algorithm for topology recovery of the electrical network

2.4.2.1 Matrix formulation of a topology

Recovering the topology of a N nodes radial electrical network consists of defining the connection matrix T, with values that link a parent node with child nodes of the network, as :

$$\begin{cases} T_{i,j} = -1, \quad if\ node\ i\ is\ a\ parent\ of\ node\ j\ ,i\ and\ j\ are\ connected\ by\ a\ line \\ T_{i,j} = 0, \quad otherwise \\ T_{i,i} = 1, no\ link\ to\ consider\ for\ a\ node\ with\ itself \end{cases} \tag{2.4}$$

i and j are respectively the parent nodes and the child nodes i.e. the current are flowing from node i to node j if no generation is connected. $i \in \{1;N\text{-}1\}$ and $j \in \{2;N\}$ (figure 1).

2.4.2.2 Heuristic algorithm

Let us consider for two nodes *i* and *j* of the network the parameters below:

- ($E_{i,j}$) their Euclidean distance,
- ($CT_{i,j}$) a value assigned to their connection test.

For a parent node *i*, the Euclidean distance ($E_{i,j}$) between the measured voltage vector and the measured voltage at other nodes *j* is calculated. A set of 96 measured data (96 quarter-hourly recordings for one day) is available at each node. For each quarter *q*, if ($E_{i,j}$) is the lowest one (which means that ($E_{i,j}$) is less than ($E_{i,1}$) to ($E_{i,j-1}$) and less than ($E_{i,j+1}$) to ($E_{i,N}$)) then it is considered that node *i* and *j* have a similar voltage and so are connected. In consequence, the connection test ($CT_{i,j}$) between both nodes is confirmed, and it is incremented by 1. Otherwise, it is unconfirmed.

Table 3 shows the statistical distribution of the parent node to child node relationship value (based on the last value of ($CT_{i,j}$)) for all the network (with data from the network drawn in section 2.2). Some connection tests were performed by starting from the transformer node (named Parent 1) and by knowing that each child node is supplied with the current from a single node (the parent, which precedes it) but can supply several other nodes (which follows it). For example (table 3), we see in colon 1 (parent node 1), the unique number 95 at the intersection with the second line (child node2). So, 95 Euclidean distances among 96 during the day are minimum between node 1 and 2, and, so these nodes will be considered as connected by a line.

Table 3: Parent node to child node relationship (last value of $CT_{i,j}$).

	Parent	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Child																			
1		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2		95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3		0	94	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4		0	0	95	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5		0	0	1	91	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6		0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0
7		0	0	0	0	0	75	0	0	0	0	0	0	0	0	0	0	0	0
8		0	1	0	0	0	1	12	0	0	0	0	0	0	0	0	0	0	0
9		0	0	0	0	0	0	1	4	0	0	0	0	0	0	0	0	0	0
10		0	0	0	0	0	0	0	0	85	0	0	0	0	0	0	0	0	0
11		0	0	0	0	0	0	0	0	0	85	0	0	0	0	0	0	0	0
12		0	0	0	0	0	0	0	0	0	0	95	0	0	0	0	0	0	0
13		0	0	0	0	0	0	0	0	0	0	0	95	0	0	0	0	0	0
14		0	0	0	0	1	1	0	0	0	0	0	0	95	0	0	0	0	0
15		0	0	0	1	0	0	0	0	0	0	0	0	0	95	0	0	0	0
16		0	0	0	2	85	2	1	1	2	2	0	0	0	0	2	0	0	0
17		1	0	0	1	1	2	0	1	1	0	1	1	1	1	3	91	0	0
18		0	0	0	1	2	7	2	0	0	2	0	0	0	0	2	1	94	0
19		0	0	0	0	0	2	79	16	1	0	0	0	0	0	10	1	1	83

The topology matrix *T* obtained from the complete database corresponds to 95% (i.e. 1 error over 19 nodes) of the parent-child relationship. The error is observed during the reconstruction for node 19; knowing that this last one and the node 8 have both node 7 as parent. These first results allow us to confirm that the algorithm works very well for a radial electrical network without or with little branching.

Then, the topology matrix *T* is filled (following the definition given by equation 2.4) by considering an effective connection of a node (child) to a parent node if 95% of the voltage tests is obtained. These first results (Table 4) allow us to confirm that the algorithm works well for radial electrical network without or with limited branching.

Table 4: Network topology.

Children nodes	Parents nodes																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	-1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	-1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	-1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	-1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	-1	1	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	-1	1	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	-1	1	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	-1	1	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	-1	1	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	-1	1	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	-1	1	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	-1	1	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	-1	1	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	1	0	0	0
16	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	-1	1	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	1	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	1
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1

2.4.3. Identification of the network impedances

Even if the topology is known, it is very rare that impedances of LV distribution networks are exactly known because:

- they are varying with age, temperature, humidity,

- they have been modified by the different reinforcement and maintenance works.

Impedance values are strongly uncertain. In this part, an algorithm is proposed to identify those impedances with collected Smart Metering data.

2.4.3.1.Parametric modelling of the LV network

A lumped pi-model with cable shunt admittances is currently used for the modelling of lines. But, due to the short distances (short cable length between nodes; Table 2), the shunt admittances (capacitive phenomenon) from the general PI model can be neglected as demonstrated in [55]. Therefore, the line impedance can be considered as a combination of per-unit-length series resistance R_i and reactance X_i as (Figure 15) :

$$Z_i = length_i * (R_i + jX_i) \tag{2.5}$$

where Z_i is the self-impedance of the line *i* between the parent node *i-1* and the child node *i*. R_i in (Ohm/km) X_i in (Ohm/km) and $length_i$ (km) represent respectively the line resistance, the line reactance and the length of the line.

Figure 15 shows the series model of the above LV electrical line.

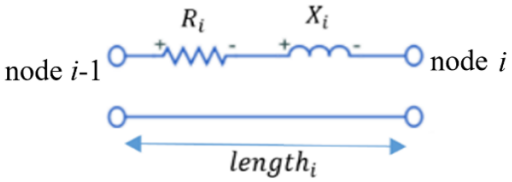


Figure 15*:* Equivalent series model of a healthy line.

Following the topology of the considered distribution system, a z-parametric diagram can be established by using the previous model of lines (Figure 16).

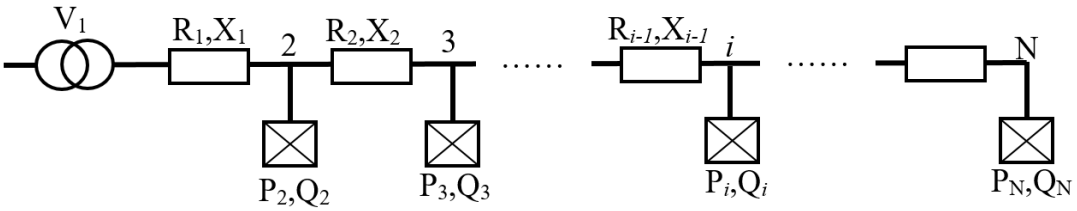


Figure 16: Parametric diagram of a LV network.

By relying on the powers and voltages measured by Smart Meters, a data-based optimal identification of the electrical network impedances Z_i (of a healthy single-phase line) is now presented. As lengths between SM are known, impedances are expressed as:

$$Z_{br_i} = Z_i * length_i \tag{2.6}$$

where $i \in \{2\ ;\ N\}$ and *N* is the total number of nodes in the network.

The impedance matrix is then obtained:

$$Z_{br} = \begin{pmatrix} Z_{br_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & Z_{br_N} \end{pmatrix} * I \tag{2.7}$$

I is the diagonal identity matrix (*N-1 x N-1*).

2.4.3.2.Objective and cost function

To find the best estimation of the *Zi* impedance parameters of this model, an optimization algorithm is used so that calculated voltages with the model are matching with measured ones. The cost function to be minimized is defined as the RMSE (Root Mean Square Error) error between measured node voltages V_{n_i} and approximated voltages V_{z_i} (with *Zi* parameters).

$$e_i = \sqrt{\frac{\sum_{t=1}^{96*Jour} \left(V_{n_i}(t) - V_{Z_i}(t)\right)^2}{96 * jour}} \tag{2.8}$$

where *jour* equal respectively to *1* or *364* depending on daily or annual impedance identification.

As a 15min time step is implemented for measurements, 96 measurements are considered for one day (or *96*364=34944* for one year) and for a node *i*. The task is then to minimize this voltage error:

$$J_i = min(e_i) \tag{2.9}$$

The voltage at node 1 (*i*=1) is assumed to be constant and imposed by the substation transformer. By considering all other nodes, the cost function for the studied network is written as follows:

$$J = \sum_{i=2}^{N} J_i \tag{2.10}$$

Hence, voltages at other nodes can be estimated following the voltage drop expression (derived from equation 1.6):

$$\left|V_{Z_{i+1}}\right| - \left|V_{Z_i}\right| = \frac{R_{i+1}.P_{i+1} + X_{i+1}.Q_{i+1}}{\left|V_{Z_{i+1}}\right|} \tag{2.11}$$

We recall that R_{i+1} and X_{i+1} are unknown, and are found through all the time steps. In installed SM, the measured voltage ($V_{n_i}(t)$) is not communicated to the hub and is not available in the information system. Anyway, with power measurements (P_i, Q_i), these voltages are calculated by a power flow algorithm (figure 14).

As the voltage drops between line sections are different and independent from one to another, this optimisation problem is non-linear with as many objectives functions *Ji* (associated to a single-parameter *Zi*) as truncated lines. Hence, each cost function *Ji* has its own optimum solution *Zi*. However, the best estimation of the network parameters is the best compromise that optimizes the set of all costs. Hence, a bi optimization is developed. Firstly, (*N-1*) mono objective optimizations are implemented individually by finding the best impedance values for each objective functions J_i (equ. 2.9). Then, knowing the (*N-1*) minimum that can be achieved, a multi objective optimization (equ. 2.10) is implemented with these goals and by considering a global objective function, which is based on a weighted sum of all J_i.

Step 1 : Mono objective minimization with fminunc solver

As a first resolution tool, we used the "fminunc" tool from Matlab as a non-linear optimization algorithm. In fact, that solver is regularly used (combined or/not to another algorithm) for multidimensional objective function using Quasi-Newton Methods and quadratic

interpolation. For finding the local optimum, we write a function to evaluate each objective function J_i at an iteration k ($k \in \{1; i\}$). Without doing that (which means solving directly J_i, the solver tries only to adequately minimise the first objective function.

Step 2: Multi objective minimization with fgoalattain solver

In this second step, found impedances are reset. Previous found minimized cost function values are then used as weights to rewrite the unique cost function (equation (2.12) but based on equation (2.10) syntax) and the Matlab goal attainment solver is used.

$$J = \sum_{i=2}^{N} (\text{New}\_J_i - J_i \text{ (found at } Step1) \qquad (2.12)$$

2.4.3.3.Results with a 4 node network

To illustrate how the optimization algorithm works, a simplified four nodes single-phase network is used (Figure 17).

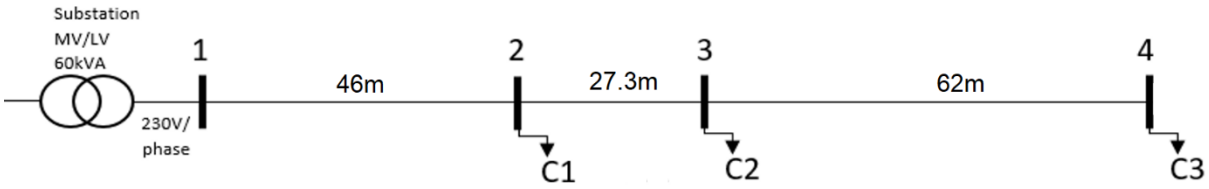


Figure 17: Topology of the simplified four nodes network.

Figure 18 shows daily profiles of the active power values measured for customers at nodes 2, 3 and 4.

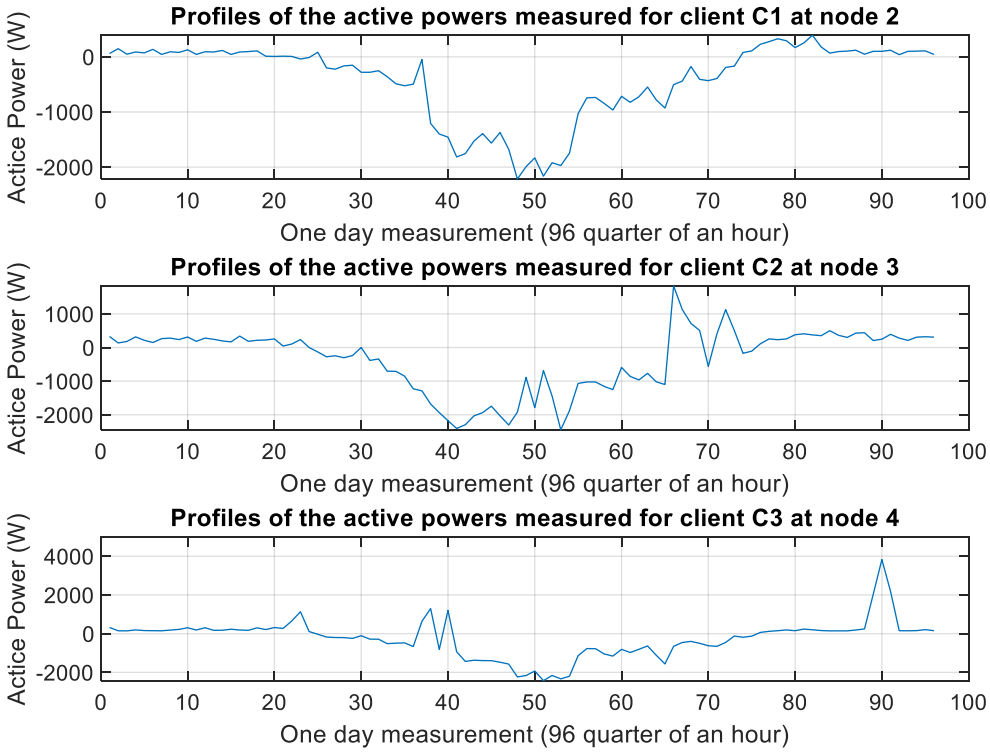


Figure 18: Profiles of the active powers measured in c1, c2 and c3.

Two study cases are considered. In case 1, we consider that lengths of lines are unknown so decision variables are the matrix Z with impedances in ohms of the electrical network (table 5). In case 2, we assume that lengths are known, they can be estimated by measuring the communication speeds between smart meters.

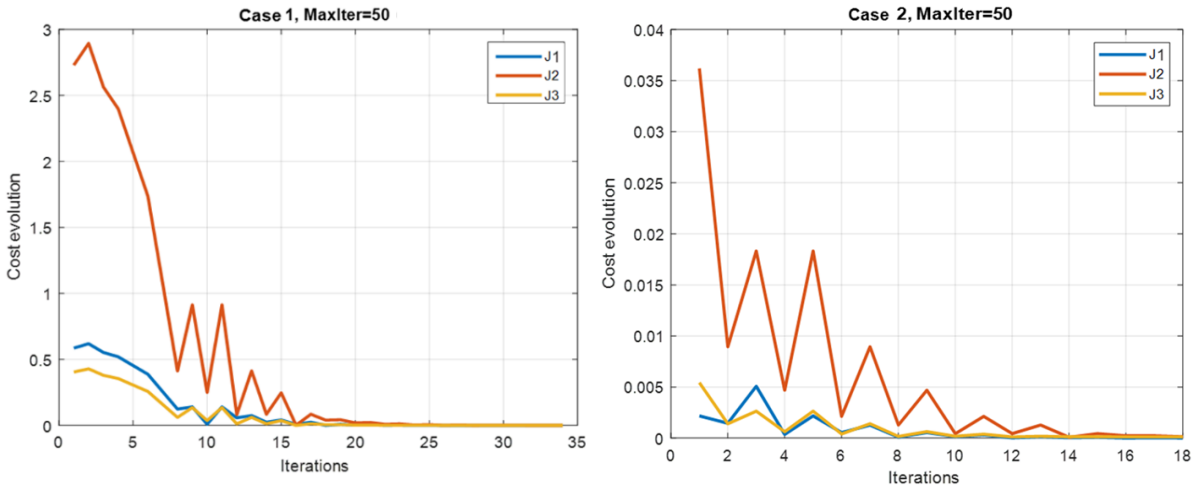
Table 5: Simulation cases specification.

<table>
<tr><td colspan="2">Case 1 : Unknown length of cables</td></tr>
<tr><td>Decision variables :
impedances of the network
$Z = \begin{bmatrix} Z_{l1} & 0 & 0 \\ 0 & Z_{l2} & 0 \\ 0 & 0 & Z_{l3} \end{bmatrix}$</td><td>True values of impedances:
$Z_{l1} = 0.046 * (0.21 + 0.31i) = 0.00966 + 0.01426i$
$Z_{l2} = 0.273 * (0.32 + 0.24i) = 0.08736 + 0.06552i$
$Z_{l3} = 0.062 * (0.32 + 0.24i) = 0.01984 + 0.01488i$
(km) (ohms/km) (ohms)</td></tr>
<tr><td colspan="2">Case 2 : Known length of cables</td></tr>
<tr><td>Decision variables :
impedances of cables
$Z = \begin{bmatrix} Z_1 & 0 & 0 \\ 0 & Z_2 & 0 \\ 0 & 0 & Z_3 \end{bmatrix} * \begin{pmatrix} 0{,}046 \\ 0{,}273 \\ 0{,}062 \end{pmatrix}$</td><td>True values of impedances (per unit length):
$Z_1 = 0.21 + 0.31i$
$Z_2 = 0.32 + 0.24i$
$Z_3 = 0.32 + 0.24i$
(Ohms/km)</td></tr>
</table>

1st Test : Ideal measured voltages are considered

In this test, the calculated voltages (with a power flow), are considered exactly equal to the nodal voltage that could be measured during the monitoring. The results obtained after simulations are shown in Figure 19, for the two cases defined above and different numbers of iterations (where J1, J2 and J3 are respectively the cost function for line 1 to 3).

Case 1 Case 2



$Z_1 = 0.0096 + 0.012i$ $Z_1 = 0.2061 + 0.3088i$

$Z_2 = 0.0871 + 0.059i$ $Z_2 = 0.3200 + 0.2478i$

$Z_3 = 0.0198 + 0.018i$ $Z_3 = 0.3199 + 0.2075i$

Figure 19: Cost function convergence and estimated impedances: (Case 1) impedances in ohms ; (Case 2) impedances in ohms/km.

After 6 iterations, errors at all nodes (functions *J1*, *J2* and *J3*) are significantly reduced. Figure 20 shows the boxplot of the voltage errors at the ending node through iterations.

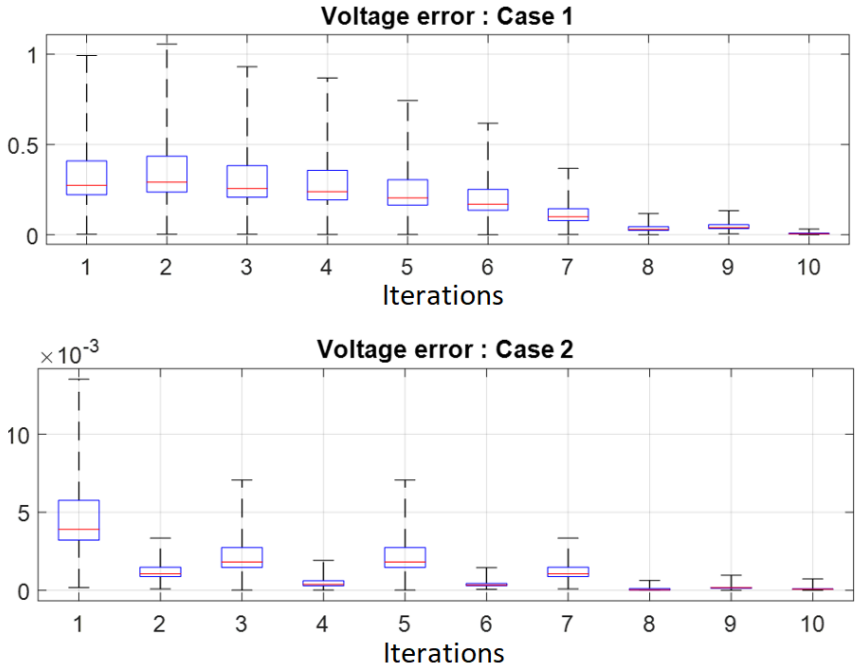
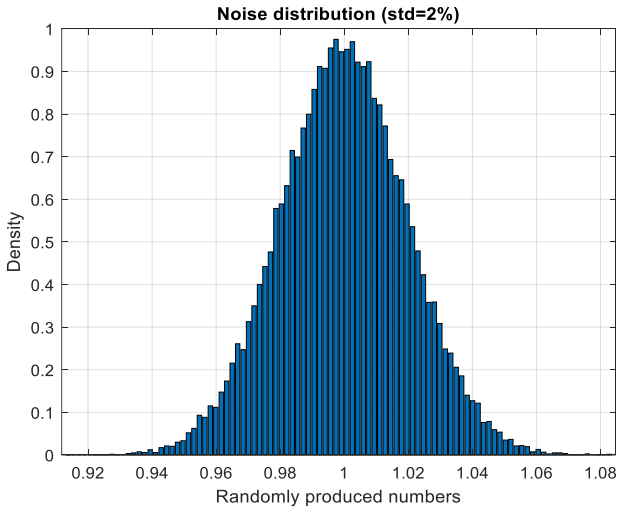


Figure 20: Boxplot of voltage error between measured and estimated impedance (node 4).

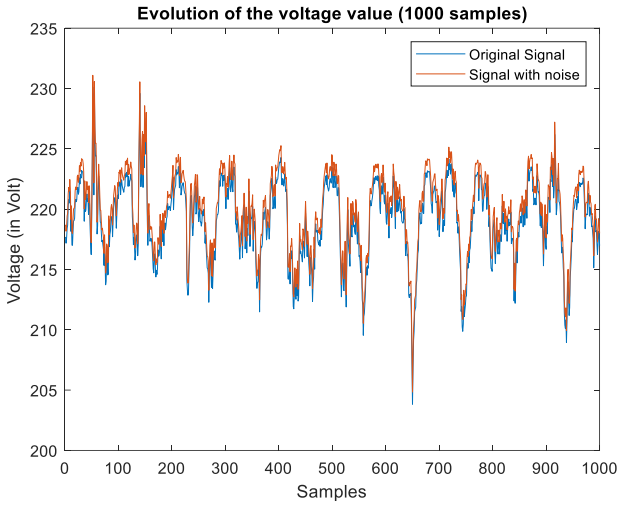
Even if the combination of solvers allows the perfect optimization of all cost functions, it is an optional point to do a goal attainment after the "*fminunc*" minimization (case 1). We see that by taking into account the fact that the line length is known (Case 2) the voltage error is significantly reduced.

2nd Test : Measured voltages with noises

In practice, smart meters have a standard measurement error of 2% for the voltage and 3% for the powers. To assess the impact or measurement errors, a measurement noise is added to the measured voltages (v_{ni}) except at the power station node. A random Gaussian signal is parametrized with a mean equal to 1 and a standard deviation of 2% [68] (Figure 21).



(a)



(b)

Figure 21: Voltages profiles : (a) Added noise ; (b) Noisy measurement of the voltage.

By assuming that the line lengths are known (case 2), the challenge now will be to evaluate the ability of the algorithm to deal with noisy data in order to continue to converge to a best estimation of impedances. The obtained estimated impedance values with respected errors and relative errors are :

$$Z_1 = 0.26 + 0.341i,\ \Delta Z_1 = 0.051 - 0.031i,\ \ \Delta Z_1 / Z_1 = 24\% - 0.1\%i$$

$$Z_2 = 0.44 + 0.28i,\ \Delta Z_2 = 0.122 + 0.041i,\ \Delta Z_2 / Z_2 = 38\% - 17\%i$$

$$Z_3 = 0.43 + 0.24i,\ \Delta Z_3 = 0.112 - 0.13i,\ ,\ \Delta Z_3 / Z_3 = 35\% - 54\%i$$

The figure below shows the boxplot of the voltage errors over the measurement time and through some iterations (at the last node of the network) :

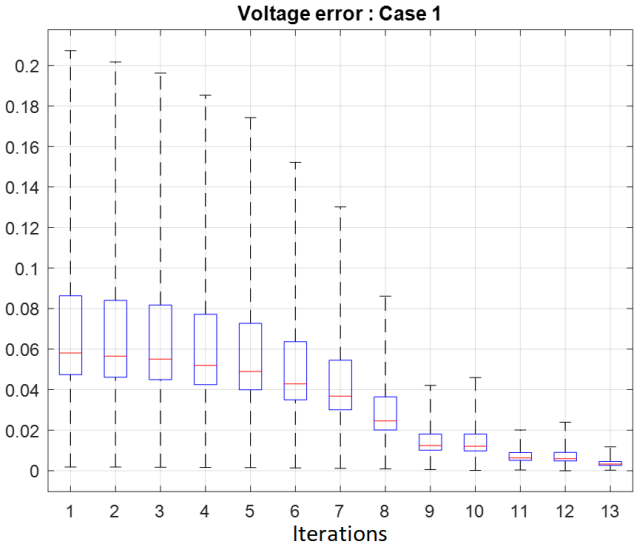


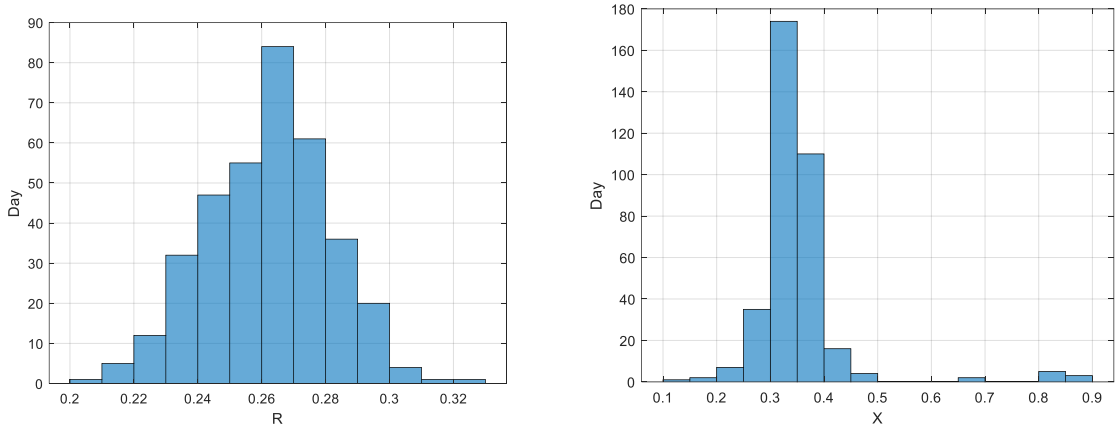
Figure 22: Boxplot of the voltage magnitude error for noisy data.

As we can see in Figure 22, the random noise affects the distribution of the voltage error with a variation from 3% to 41% between the noisy calculated voltage (V_{Zi}) and the perfect measured one.

An interrogation can be emitted on the fact that the new impedances are those close to the parameters of the network during the measurements: assumption to validate by adding in the grid definition, the voltage at the substation.

2.4.4. Statistical distribution of the impedances *Z* during one year

In this part, the impedance of the impedance matrix is re-estimated every day during one year and the noisy measured voltages are considered. A collection of different values are obtained and can be represented has a distribution. As example, the Figure 23 represents the distribution of R_1 and X_1 for 20 days during the year 2013.



(a) (b)

Figure 23: Distribution of the resistance (a) and the reactance (b).

The average values and standard deviations can be calculated for each line. The table below presents the estimation results of the line Z_2 over a cumulative period of 20 days for a single-phase network:

Table 6: Daily variations of the averaged impedance ($Ri + Xi$) values in mOhm

Day	R_1	X_1	**Day**	R_1	X_1
1	0.280	0.317	11	0.228	0.322
2	0.244	0.386	12	0.261	0.379
3	0.239	0.322	13	0.231	0.353
4	0.282	0.360	14	0.253	0.386
5	0.238	0.353	15	0.259	0.325
6	0.257	0.257	16	0.238	0.345
7	0.266	0.370	17	0.249	0.350
8	0.269	0.394	18	0.243	0.394
9	0.240	0.353	19	0.278	0.436
10	0.269	0.383	20	0.276	0.293

For the resistance and the reactance values, we can see that the variations of R_1 or X_1 is respectively around ±25% and ±100% of their typical values during the year.

That observation will be helpful for the study of the environmental impacts over the lines. The assumptions made about the LV network structure allows us to learn its impedance. Using the goal attainment solver, we were able to first validate the solution resulting from the first solver and also to weight our cost function, which can be very helpful when working with an unbalanced network. Regarding voltage drops over the measured samples, we can assume that the LV grid monitored is healthy.

However, using the noisy voltage as measured voltage highlighted the impact of the substation voltage in the optimization result. So, sensor at the substation is an important step for the best monitoring of an electric grid (even if the sensor is imperfect: error, communication time delay (which varies according to the distance with the collection point), partial sensitivity to noise). The impact of the voltage angle is neglected because the lines have short lengths.

2.5.Temperature based model of the resistance distribution

With climate change, we are witnessing an increased the warming. The main goal of this study is to investigate the dependency of voltages and power flows in lines with the temperature/weather. The obtained resistance distributions enable the implementation of a probabilistic Load Flow (LF) calculation and so to estimate voltages and power flows according to temperature variations.

2.5.1. Motivation of the study

The passage of a current in overhead power lines causes a warming due to Joule effect losses. A higher ambient temperature limits the cooling of overhead lines by convection and radiation. Thus, in a hot weather, the temperature within the power line itself increases. In order to avoid reaching too high temperatures, which would damage conductors and insulators, the maximum current that the line can carry must be reduced. For the transport of electricity under high voltage, an automatic procedure is implemented and is called Dynamic Line rating. For low voltage lines, to our knowledge, there is no procedure and the values of the maximum currents of the protections cannot be modified remotely. In case of a permanent warming, the ground will also heat up, which will induce also the heating of underground cables, a power derating may be scheduled. Temperature also has a comparable effect on substations: the maximum capacity of power substations and transformers is reduced by around 0.7% per additional degree over the standard value [69].

In order to follow the evolution of the climate, the ONERC (National Observatory on the effects of global warming) considers as an indicator the difference in temperatures compared to the average over the reference period 1961-1990. The warming of the global average temperature is very clear: strongly negative difference until 1940, then most often negative difference until around 1980, then a net warming. The temperature difference has been almost systematically positive since the beginning of the 1980s (Figure 24).

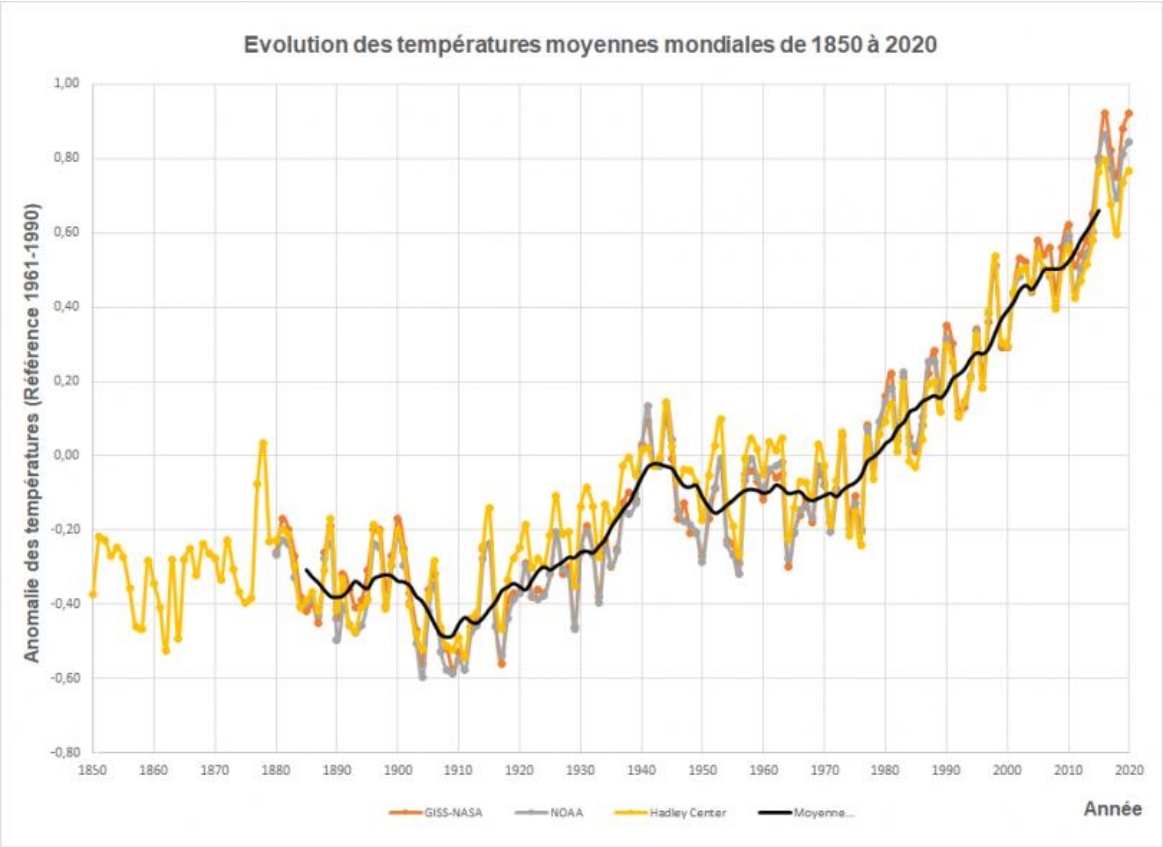


Figure 24: Historic temperature variation in the world from 1850 to 2019 [70].

With this tendency, the evolution of annual average temperatures in Europe shows a clear warming since 1900 as measured in France (Figure 25).

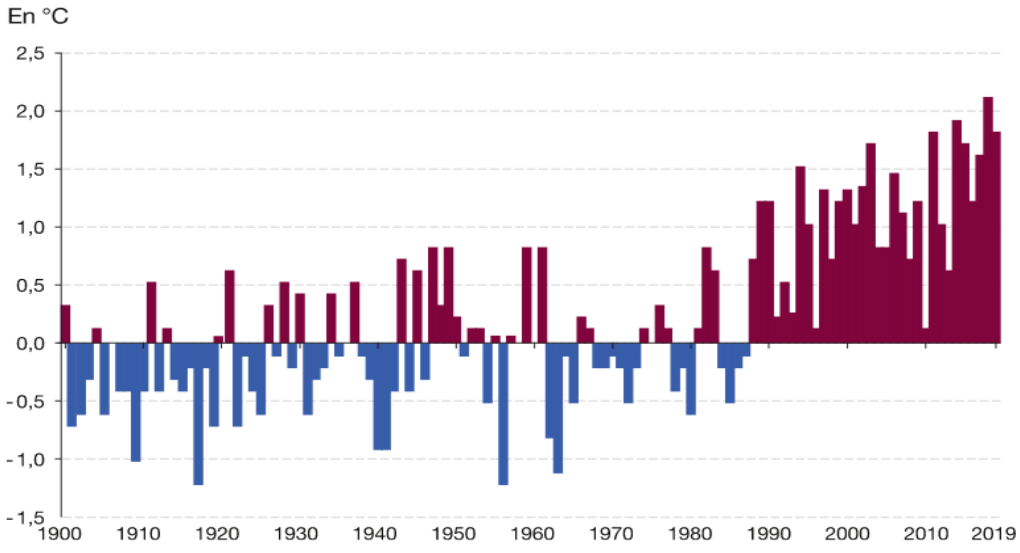


Figure 25: Historic yearly temperature variation in France [70].

For the future, the various considered scenarios by the GIEC are clearly also considering a significant temperature increase whatever the decided actions (Figure 26).

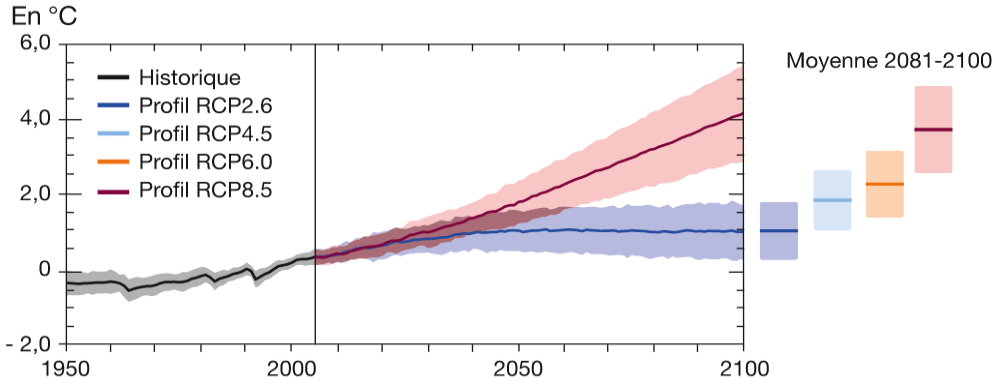


Figure 26: Projection of average temperature variation according to different scenarios [71].

For the operation of the electrical network, the instantaneous value of the temperature is more relevant and the increase of extreme temperature are critical (Figure 27).

The sizing of the electrical infrastructures was calculated with temperature assumptions, which seemed to represent a sufficient safety margin at the time. Now, with global warming this margin is drastically reduced. As example: to avoid power cuts and fire outbreaks, a minimum safety height is needed between a power line and the obstacles below because the conductors lengthen and get closer to the ground depending on the heat. For the MV network, the safety height must be guaranteed up to an outside temperature of 40°C. This is a threshold that was exceeded in many places during the heat waves of June and July 2019 (See Appendix B for July 2022 temperatures). If the network is exposed to maximum temperatures higher than those for which it was designed, it may not be able to carry all the electricity. If this is the case, load shedding may become unavoidable even if the available generation capacities theoretically meet demand.

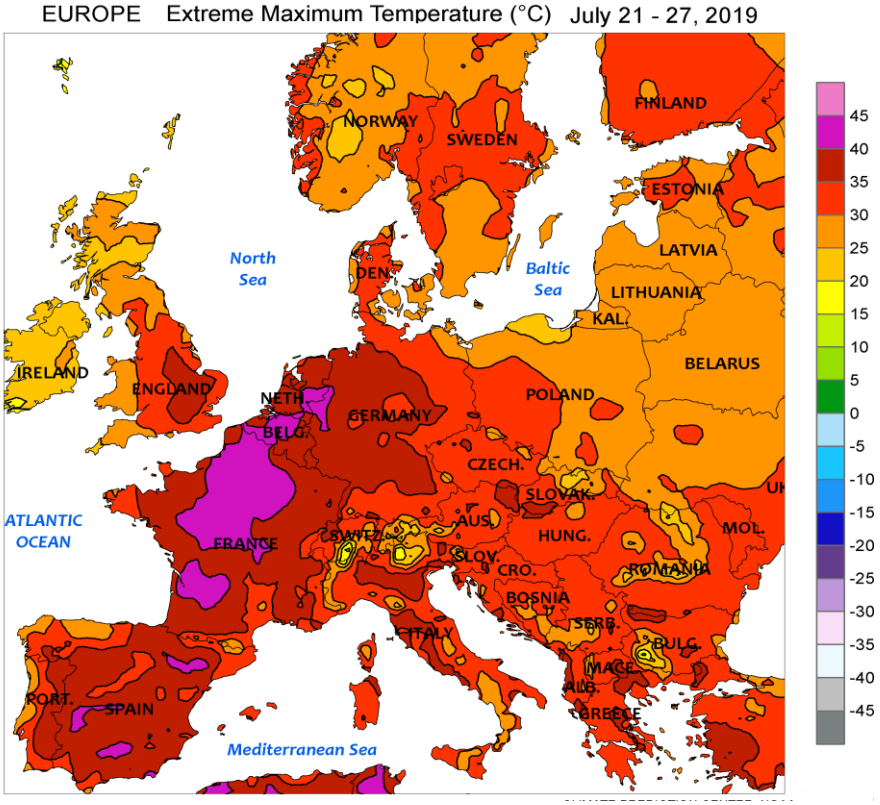


Figure 27: Measured maximal temperatures between 21th and 27 th July, 2019 [72].

Low voltage infrastructures are mainly in overhead. Underground low and medium voltage lines are a solution, while noting that during heatwaves, the ground will also heat up, which will increase the heating of underground lines. Moreover, underground cables have economic limits because they are more expensive than an overhead line.

In this part, the proposed data based model is enhanced to study the effect of the temperature onto the operation of LV networks. The variation of the impedance along the year is modelled and then a probabilistic load flow is developed to quantify the impact onto the voltage variations. To do so, the line resistance distribution is considered as an uncertainty source that needs to be characterised.

A probabilistic LF is implemented in order to estimate voltages and improve the identified impedance values. References [73], [74] and [75] investigated the LV management issue by developing a probabilistic tool for Power Flow calculation considering the fluctuations of PV productions and load demands as an uncertainty regarding the probabilistic evaluation. While the work in [74] used directly smart meter data to establish the PV production profile, the work in [73] based its uncertainty study on a forecasting method by using meteorological data from numerical weather models. Unlike the works of [76] and [55], which defined a confidence interval around the line resistance nominal value, annual and seasonal resistance variation profiles has been defined on the basis of temperature data for this work.

A complete process is developed, for some specific case studies. The objective is to show, to what extent the use of such an upgraded tool can demonstrate the actual impact of the resistance distribution on a Probabilistic Load Flow computation.

2.5.2. Problem Formulation

2.5.2.1.Thermal modelling of a line resistance

By taking into account the temperature variation over one year, a resistance distribution model is built. The inductance is kept fixed because magnetic effects of the lines can, indeed, be considered as independent from the external temperature. The line resistance is expressed with the temperature and the electrical conductor resistivity as:

$$R\ (Ohms) = \ \rho * \frac{l}{S} \tag{2.13}$$

$$R\ (Ohms/m) = \ \rho * \frac{1}{S} \tag{2.14}$$

$$R_t\ (Ohms/km) = \ [\rho_o * (1 + \alpha_o * \Delta T)] * \frac{1}{S * 10^3} \tag{2.15}$$

where l is the length of the line; the parameter ΔT (with $\Delta T = T_t - T_o$) is the temperature variation between the external temperature T_t and the steady state temperature T_o.

The other parameters in equation (2.15) and their value (used in the remaining of chapter 2) are defined in Table 7.

Table 7 : Specifications of the line.

Line conductor parameters	
Parameter	**Value and unit**
T_o : Steady state temperature	20 °C
α_o : temperature coefficient at 20°C	$4x10^{-3}$ K^{-1}
S : conductor section	50 mm^2
ρ_o : Resistivity at 20°C	$17,2x10^{-9}$ Ohms.m

2.5.2.2.Impact of the temperature onto impedances and the network operating

As the external temperature during the year is varying, the line resistance will change also. So a large set of resistance values is obtained. The seasonal or annual average value of the line resistance can be considered.

Table 8: Deterministic values for the fixed resistance value in *mOhms/km.*

CASE 1 : ANNUAL STUDY	CASE 2 : SEASONAL STUDY
0.333	Winter : 0.322 Spring : 0.336 Summer : 0.337 Fall : 0.337

The average value of the resistance distribution in a month can be calculated as well as the average value over the year, but these deterministic data do not model the power variability and the uncertainty of powers.

By considering the resistance as a random variable, the probability and so the cumulative distribution function (CDF) can be calculated over one year horizon (Figure 28).

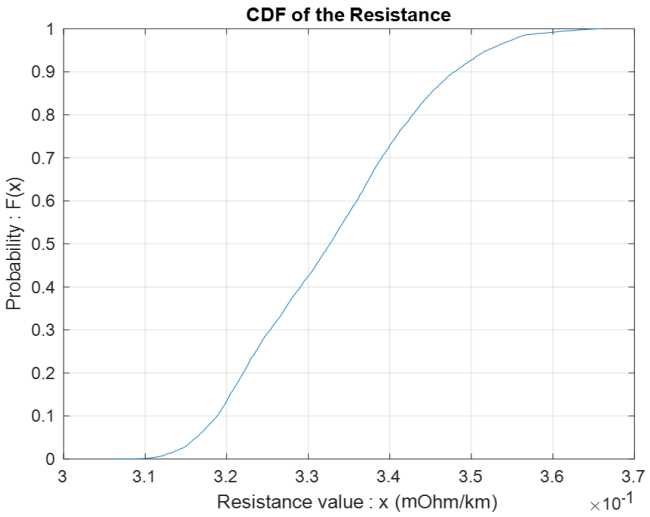


Figure 28: Annual CDF of the line resistance.

Figure 29 shows the CDF (see Appendix C for CDF construction process) of the same line resistance for each season. The CDF profile for the resistance is logically explained by the temperature distribution with higher values during the warmest months.

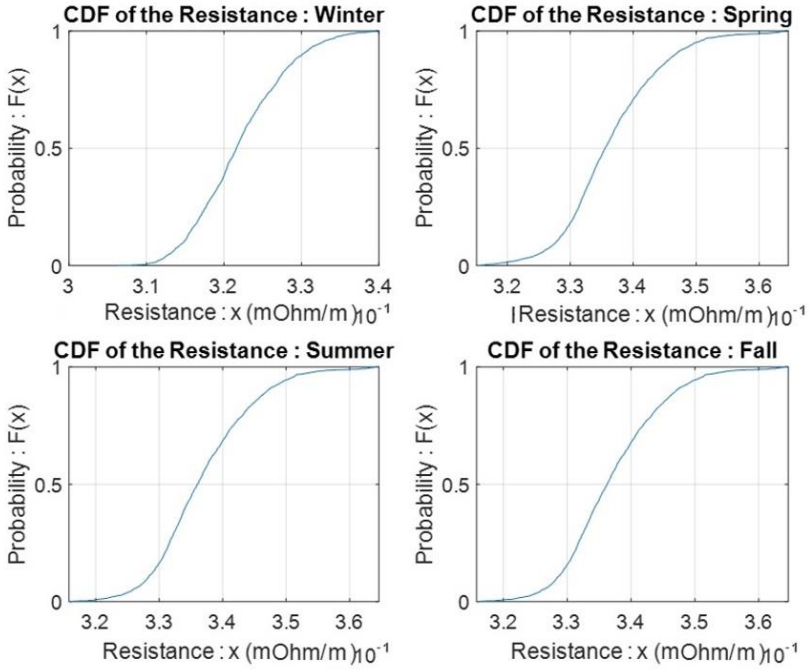


Figure 29: Seasonal CDF of the line resistance.

2.5.3. Probabilistic modelling of variables

Randomly introducing the line resistance distribution in the probabilistic tool is important to show how and when (regarding the season and the time of the day) the effect of the temperature variation influences the power flows and RMS voltages in the electrical network. By computing the voltage magnitude at each node with a power load flow calculation, it is also possible to evaluate the critical nodes (in terms of voltage magnitude constraint) by using a probabilistic criterion that considers an acceptable variation range of ±10% around the nominal magnitude [77].

Table 9: Chosen voltage characteristics.

EN50160 standard for Power Quality	
Tolerable overvoltage percentage	$\leq +10\%$ of V_o
Tolerable voltage dip percentage	$\geq -10\%$ of V_o

where V_o is equal to 230V.

Two cases have been considered to explore the impact of the thermal variations of the line resistance. The first case considers one unique typical daily profile of load demand and PV production for each quarter of an hour (q) over the whole year and will enable us to evaluate if the resistance variation can generally impact or not the voltage variation. The second studied

case will go deeper in the analysis by evaluating the voltage variation for the hottest and the coldest season and so is considering a daily profile related to the each season in the year (table 10).

Table 10: Simulation cases.

CASE 1 : ANNUAL PROFILE	CASE 2 : SEASONAL PROFILE
96 Load and PV production profiles for each customer corresponding to the 96 quarters of one hour ranging from q1 to q96	96 Load and PV production profiles for each customer corresponding to the 96 quarters of one hour and per season, ranging from $q1_1$-$q1_4$ to $q96_1$-$q96_4$ (index 1 for Winter, 2 for Spring, 3 for Summer, 4 for Fall).

2.5.3.1.Probabilistic modelling of the load demand and PV production

As the database of available data from smart meters is limited, it is interesting to use also data from a probabilistic model of these powers. In a previous research work [78], power data have been analysed and, as example, Figure 30 shows recorded data of the energy consumption at a node in April.

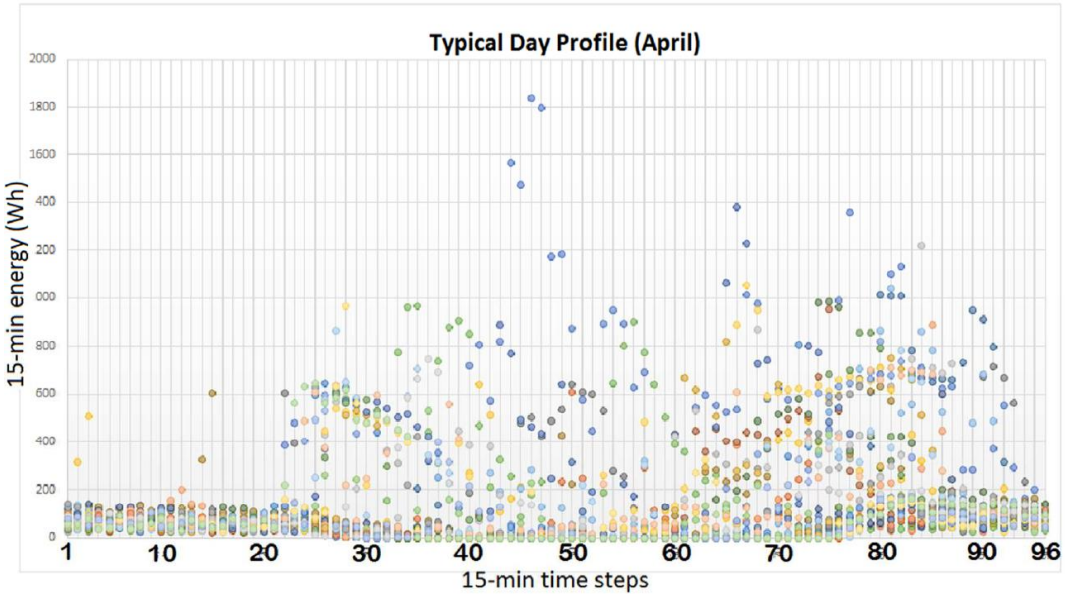


Figure 30: Graphical Representation of the Typical Day Profile of energy consumption at the point of common coupling of one PV user, in April [78].

For each 15 min time step, the load demand and PV production can be considered as random variables. Hence, two Cumulative Distribution Functions (CDFs) of Probability, one for the load demand, one for the PV production can be computed for each 15 min time step q and for each customer. By sampling these characteristics, many daily power profiles can be built through this Monte Carlo process and then used to confirm (or not) tendencies of the studies impacts over a large amount of considered profiles.

To consider powers as time depending data [79], 96 Cumulative Distribution Functions for each one of the 96 quarters of a day are modelled and used for the sampling of the customer load demand ($Load_{q,i}$ for each customer i at each quarter q). The produced PV power (called $Prod_{q,i}$) is built in a same way. The exchanged customer to network power is calculated by using equation (2.2). As example, the CDF for Load and PV production, for customer 14 (of

the network shown in Figure 11) at two different quarters of an hour, are shown in Figure 31. It can logically be observed that more important PV generation values are observed at midday.

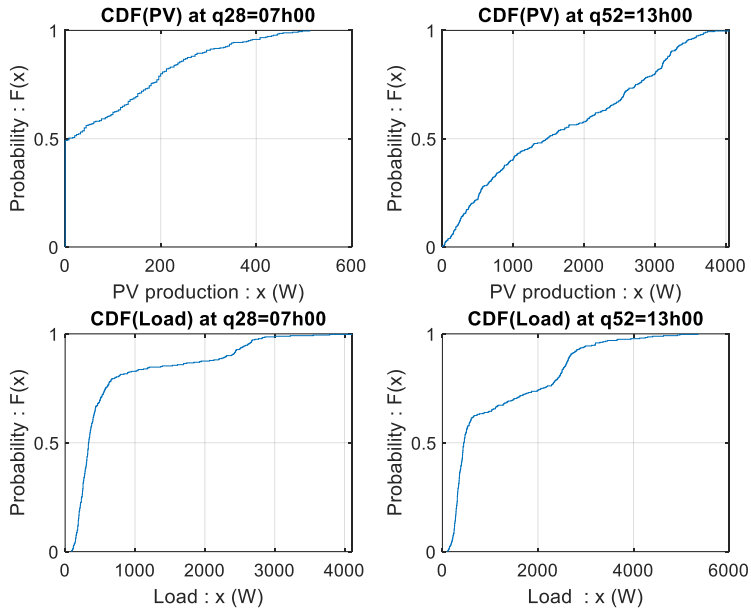


Figure 31: CDF for the PV production and the Load for some quarters of an hour.

In a same way, by considering the resistance value as a random value (see part 2.6.2.2), an annual or seasonal CDF is obtained. Then, for each studied cases, sampled values of the line resistance can be randomly generated.

2.5.3.2.Implementation of a Probabilistic Load Flow and developed algorithms

The issue of this research study is to calculate possible overvoltages and undervoltages in the studied LV network according to the temperature variations and so resistance variations of line cables. A classical probabilistic Load Flow calculation has been derived to estimate all node voltages.

Figure 32 shows the flowchart of the developed and implemented tool (in MATLAB®) of the probabilistic Load Flow with either fixed or sampled resistance values. This algorithm is linked to an upgraded version of the work done in [80] and also by considering the deterministic or probabilistic modelling of the line resistance.

Monte Carlo (MC) simulations are used to build different daily power profiles over the year by randomly sampling the 96 CDF in a day (part 2.6.3.1). Each quarter is indexed by the variable *q*. One profile (or scenario) is indexed by the iterative variable *it*, and a large number N_{it} of profiles are generated (Figure 32). The reactive power value is generated with a 0.9 Power Factor: $Q_{q,it} = P_{q,it}*\tan(\cos^{-1}(0.9))$. According to the study case (table 10), CDF are selected.

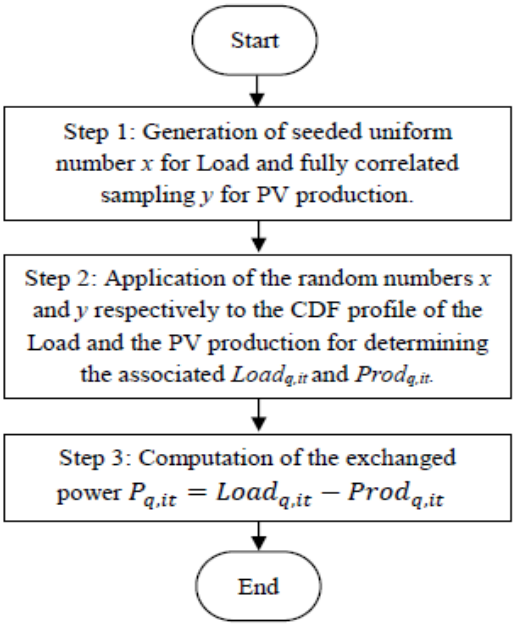


Figure 32: Random power exchange calculation at each quarter of an hour for one customer.

Resistance values will be also generated by another MC method using the annual or seasonal CDF profiles that have been modeled in part 2.6.1 (Figure 33).

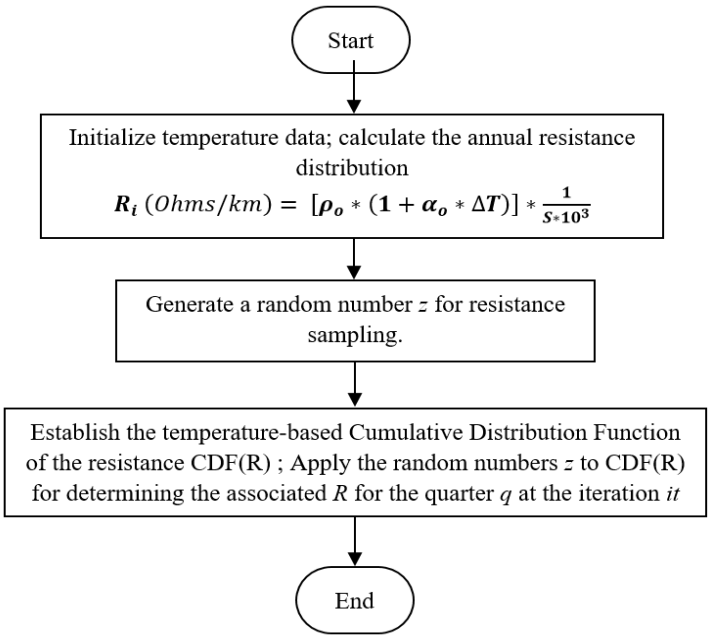


Figure 33 : Flowchart of the resistance probabilistic modelling.

Figure 34 shows the flowchart of the developed and implemented tool (in MATLAB®) of the probabilistic Load Flow with either fixed or sampled resistance values (for the case 1 : annual study algorithm). This algorithm is linked to an upgraded version of the work done in [80] by considering the deterministic or probabilistic modelling of the line resistance.

At each iteration *it* (*it* ϵ 1…*Nit*), the line resistance value is fixed, or sampled (depending of the simulation case and by using MC method described in Figure 33). Then the exchanged client to network power is generated (based on the algorithm of described by Figure 32) to obtain the active and reactive powers (respectively $P_{q,it}$ and $Q_{q,it}$) for the ongoing iteration. Secondly, both

powers associated to each customer and the reference voltage at the MV/LV post-station V_{nl} is used as input for the Load Flow calculation [77, 78] performed in order to characterize each simulated Monte Carlo state (the Load Flow calculation algorithm will be described below in Figure 35).

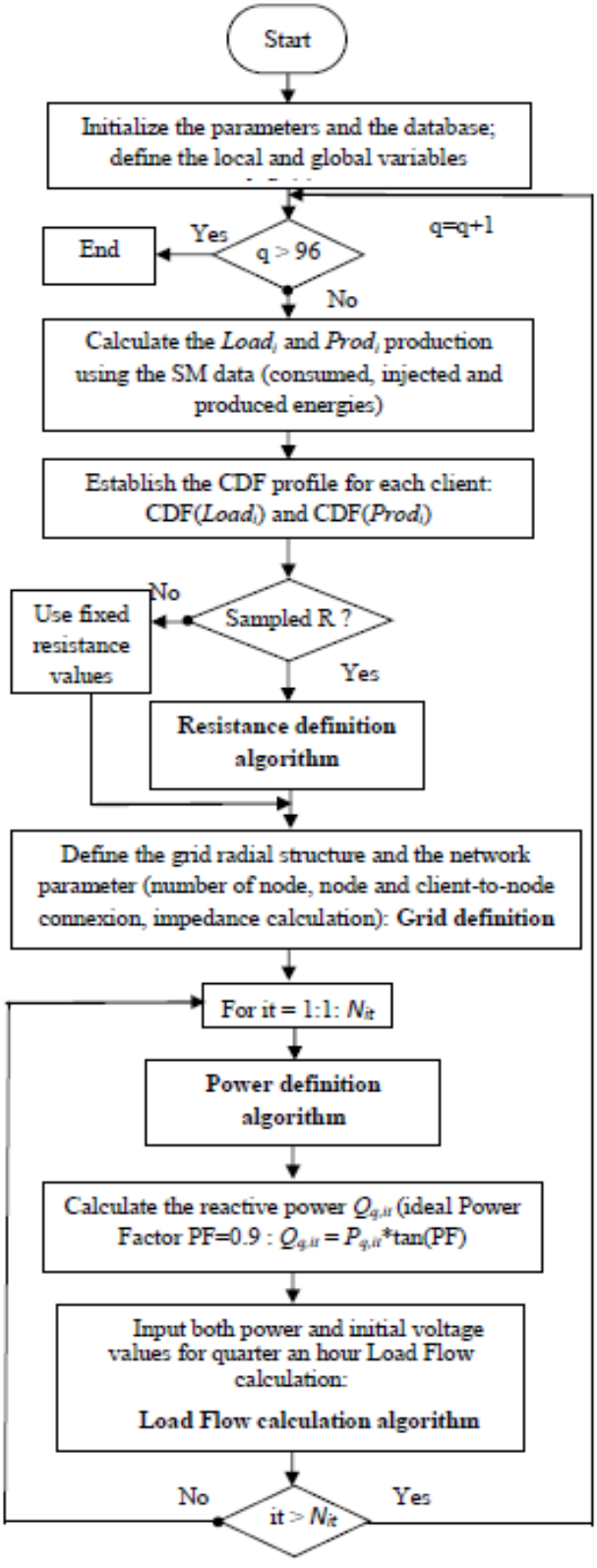


Figure 34 : Flowchart of the developed and implemented algorithm for grid definition and Load Flow calculation in case 1 (Annual study algorithm).

Both powers associated to each customer and the reference value of the voltage at the MV/LV post-station V_{nl} are used as inputs for the Load Flow calculation [81, 82], which is performed in order to characterize each simulated Monte Carlo state (the flowchart of the **Load Flow calculation algorithm** is shown in Figure 35).

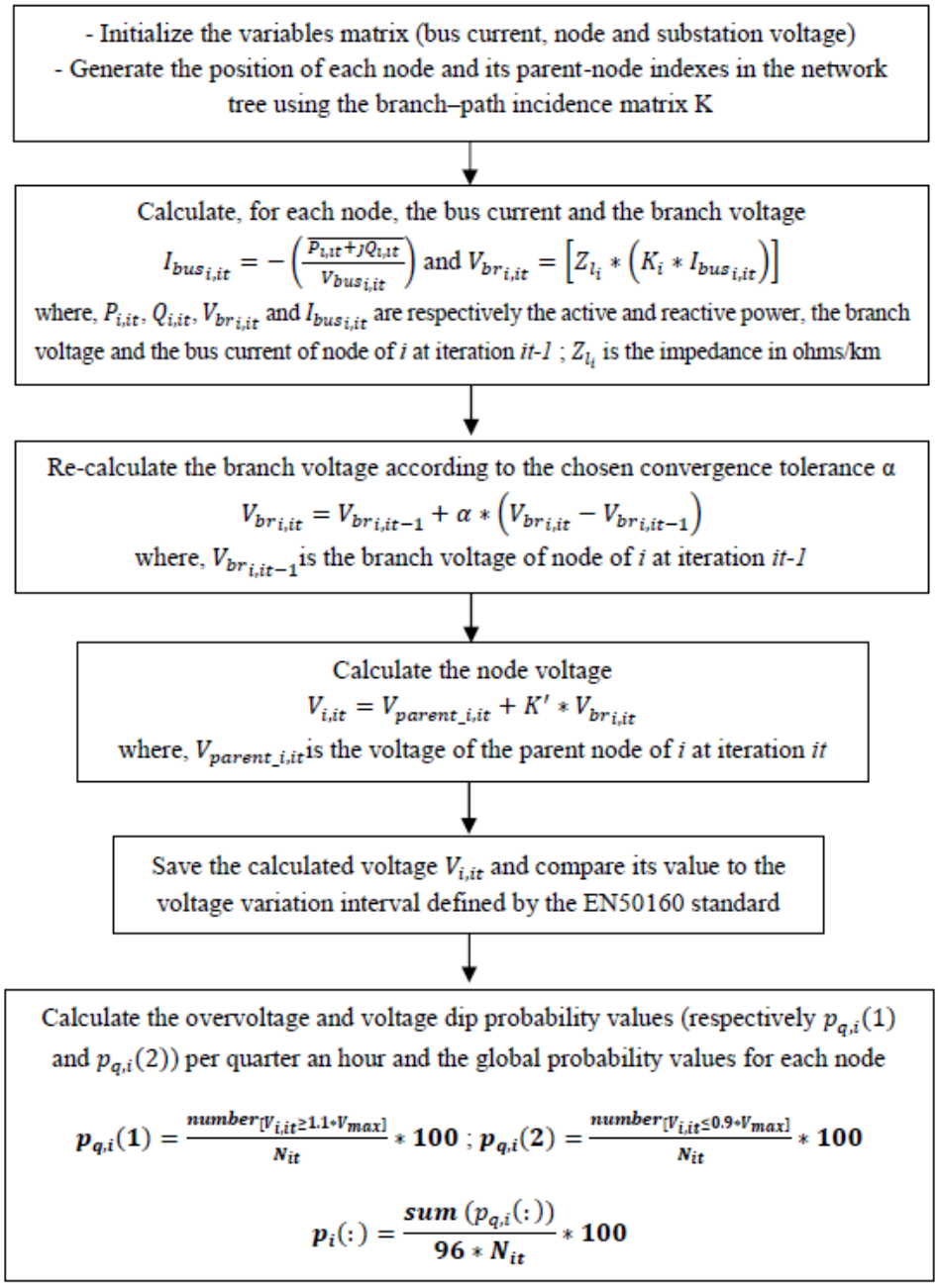


Figure 35: Flowchart of the Load Flow calculation.

The flowchart in Figure 36 summarizes the simulation process of the second simulation case. This process is the same as the first one, except the change in the SM data that has to be considered per season. So for each of the four seasons, the algorithm in Figure 35 is computed.

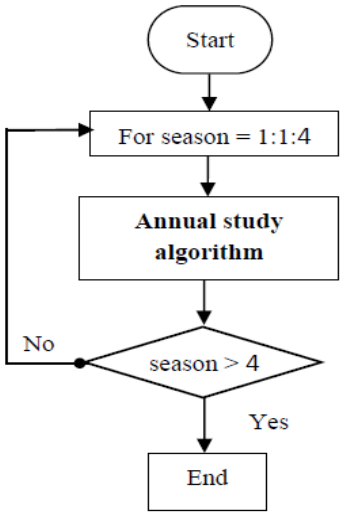


Figure 36: Flowchart of the implemented algorithm for case 2 simulation.

2.5.4. Simulation results and reliability analysis

Figure 37 shows the obtained overvoltage probability at each node of the network for case 1 with sampled value of the resistance. Node 14 is the most critical node of the network while the three first nodes next to the ML/LV station are quite safe (node 2, 3 and 4).

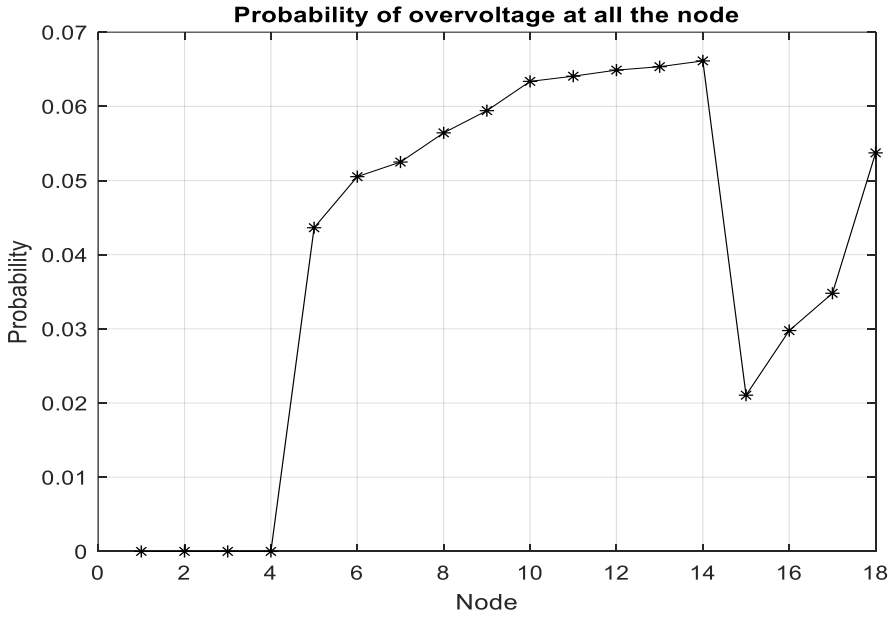


Figure 37: Probability of overvoltage at all nodes for an annual study with sampled value of the resistance.

The voltage dip and overvoltage probability, at node 2 (the one directly linked to the post-station) and at node 14 (the one at the end of the first radial line), resulting of the probabilistic tool simulation are shown on the tables below. The table 11 corresponds to the results for a fixed value of the line resistance and the table 12 for a sampled value on the annual CDF.

Table 11: Global voltage variation probability for two nodes on the same branch : for a fixed value of the resistance.

N_{IT} = 10000	PROBABILITY	
Node	***Overvoltage***	***Dip***
2	0 %	0 %
14	7.101 %	0.684 %

Table 12: Global voltage variation probability for two nodes on the same branch : for sampled values of the resistance.

N_{IT} = 10000	PROBABILITY	
Node	***Overvoltage***	***Dip***
2	0 %	0 %
14	6.613 %	0.666 %

Regarding those values, using a sampled value of the resistance has an impact on the results of the probabilistic Load Flow. From table 11 values to table 12 values, a 6.9% decrease is observed on the calculated probabilities. There is no voltage dip or overvoltage occurrence during the annual simulation for node 2. This can be explained by the direct link of that node to the transformer secondary side (46m length between both nodes); which decreases the variation effect. Table 13 gives the overvoltage probabilities at node 14 for different cases.

Table 13: Global voltage variation probability for the most critical node.

N_{it} = 10000		Probability (occurrence percentage) for node 14	
		Overvoltage	***Dip***
Winter	Fixed R annual	1.781 %	5.656 %
	Fixed R Winter	2.184 %	5.944 %
	Sampled R Winter	2.029 %	6.025 %
Spring	Fixed R annual	9.012 %	0 %
	Fixed R Spring	11.090 %	0.009 %
	Sampled R Spring	10.192 %	0.012 %
Summer	Fixed R annual	10.551 %	0 %
	Fixed R Summer	13.885 %	0.001 %
	Sampled R Summer	12.443 %	0 %
Fall	Fixed R annual	1.469 %	1.633 %
	Fixed R Fall	1.594 %	1.657 %
	Sampled R Fall	1.619 %	1.869 %

Globally, it can be seen that using a same annual mean value of the line resistance, for seasonal study, affects the reliability indices of each season. In other words, it leads to consider the network safer than it actually is. By using a fixed mean value of the resistance based on its specific distribution to each season, the probability percentages found are more realistic and by integrating the sampled resistances values, the probability values are refined to values much more in relation with the weather conditions.

For winter simulations, the apparition of voltage dips is more probable than overvoltage, because during that season the temperatures are very low and then decrease the resistance value. For summer cases, the overvoltage occurrence is higher with a very important increase, compared to the winter values, for all the cases. Furthermore, using sampled values of R_t, the seasonal study shows that there is a 5.83% more risk for an overvoltage to appear in summer in comparison with the annual study result.

Sampled values of R_t in summer allow us to observe a 10.4% decrease of the overvoltage probability regarding results for fixed summer resistance values. The simulation results and the obtained accuracy, especially in the hottest and coldest season, validate our hypothesis about the impact of the weather conditions on the overhead line resistance in a probabilistic LF study. It is therefore very helpful to include temperature-based resistance distribution in the network model within a probabilistic LF study.

2.6.Conclusion

Chapter 2 presents the temperature impact on the characterization and modeling of LV cables. This preliminary study finds all its interest in a logic of familiarization with the measurement data. It is part of a perspective of setting up the first models aimed at validating the completeness of the database. In addition, applying reverse engineering (exploiting instrumentation data to identify the system internal parameter) can make a huge difference for approximating the network theoretical impedance (often used by the DSO for their management process).

It is known that voltage control is an important point in the electrical network management. European standards have specific rules over the voltage magnitude variation. A significant point is then to identify, with the most accuracy possible, the critical nodes especially when there is a large penetration of PV production in the network. All the above cases converge towards a validation of the impact of the line resistance on the probabilistic LF study result especially in a seasonal study. This contribution provides a useful tool for voltage control of a radial LV network with consideration of PV production variation and temperature impact on the network. By identifying the critical node with the higher accuracy possible and regarding the imposed European limitations, some active network management studies can be developed in order to avoid overvoltage or voltage dip risk at an hourly scale and per season.

However, these temperature variation problems and the overvoltage problems they caused, considerably increase the cable ageing. This ageing process typically leads to embrittlement,

cracking, and eventual failure of insulation materials, exposing then the conductor and risking potential short circuit and leakage current. The conclusions of the above study therefore lead us to wonder about the impact of the degradation of said insulation on the evolution of the electrical parameters of the cable within a low-voltage electrical network. The above investigation results depend on the specific parameters of the considered cable; hence, there are much needed information for each type of cable used in order to have a precise structure of the network. Given the importance of those parameters values, of the each deployed cables, that we do not know, this opens up the view to integration of AI methods and shows their importance in this research.

Chapter 3 : Impact analysis of line degradations on LV network voltages

Contents

3.1.Introduction

In this chapter, a framework is proposed to evaluate the impacts of the cable insulation degradation on the nodal voltages in a LV distribution system. The insulation degradations can create a flow of leakage currents that are distributed in the overall electrical network and thus can cause serious problems on the correct operation of the latter.

For the study, as the criticality of the cable insulation degradation is not known, an uncertain variable is considered and can change within a predefined range. The presented analysis method is organized in two stages. Extensive Monte Carlo simulations are carried out in order to characterize the uncertain nature of the insulation material degradation. Then, load flow calculations estimate the nodal voltages in each scenario created by MC simulations.

3.2.Characterization of the Cables Insulation Degradation

Electrical cables are daily subject to mechanical damages, excessive heat, ageing of material, and electrical stress. These operating conditions cause degradations of the cable insulation material and in extreme cases, the cable can totally or partially loose its insulation. In that situation, the insulation impedance decreases and a leakage current flows between the cable and the ground. The impedance associated to that leakage current includes the resistance of the degraded cable insulation as well as the ground resistance. The remaining of this section focuses on the calculation of the resistance associated to the degraded insulation.

In a cable with a degraded insulation, the leakage current flows radially outwards from the center towards the surface of the cable along its length. A cylindrical cable is considered with a total radius *R*, a length *L* and a conductor radius equal to *r* (Figure 38).

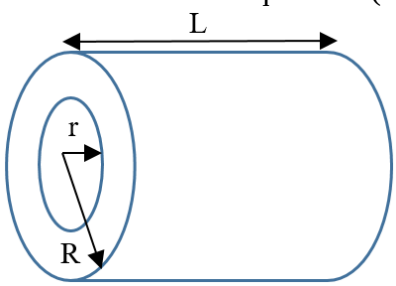


Figure 38: Geometric representation of a portion of a cylindrical cable.

The radius corresponding to the insulating material is equal to *R-r*. An elementary section with a radius *x* and an insulation material thickness *dx* is considered (infinitesimally small layer of insulation) [83]. The area of this elementary section (the cylinder area) is equal to $2\pi Lx$. The insulation resistance of this elementary cylindrical area is given by:

$$R_{iso-dx} = \frac{\rho dx}{2\pi Lx} \tag{3.1}$$

where $R_{iso\text{-}dx}$ and ρ are respectively the resistance and the resistivity coefficient of the insulation material.

From equation (3.1), the insulation resistance of the cable is calculated by integrating the thickness value *dx* over the radius corresponding to the insulating material:

$$R_{iso} = \frac{\rho}{2\pi L}\int_{r}^{R}\frac{dx}{x} = \frac{\rho}{2\pi L}\ln\frac{R}{r} \tag{3.2}$$

The above equation gives a general formulation of an electrical cable insulation resistance. By assuming that the cable is losing a part of its insulation thickness in case of degradations, the cable radius R will be reduced (as well as the resistivity coefficient ρ) while the conductor radius r will remain constant. That radius variation will tend to decrease the insulation resistance value (equation (3.2)). Figure 39 shows the CDF profile of the résistance with scenarios for insulation material degradation.

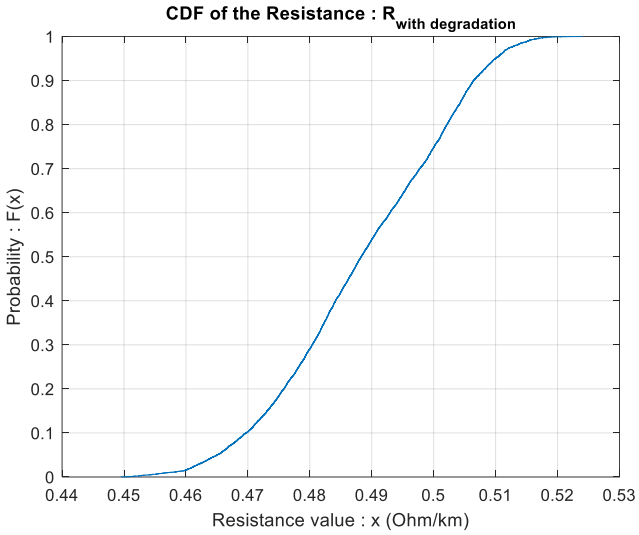


Figure 39: CDF profile of the résistance with degradation.

3.3.Modelling of the LV line in degraded conditions

To model the electrical line, in damaged insulation conditions, the resistance variation (R_{iso}) due to the insulation degradation, established in equation 3.1, is incorporated in the R, L model of the line. Indeed, a shunt variable resistance, between the leakage point (named *d* in Figure 40) and the ground, models the current discharge over an electrical insulation material. Figure 40 shows the representation of this new electric path (series combination of insulation resistance R_{iso} and ground resistance R_g) in the line model.

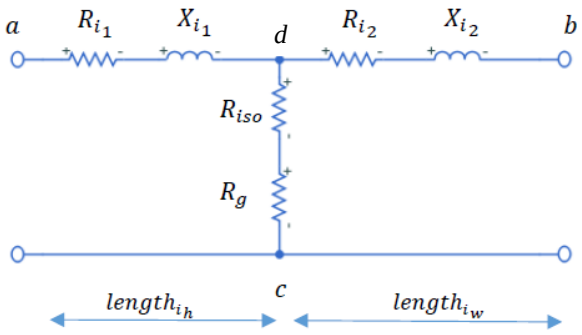


Figure 40: Equivalent electrical circuit of a line having a damaged cable insulation.

Considering that $length_i$ is the total length of the line *i*, $length_{i_h}$represents the length of the healthy part of this line and $length_{i_w}$ is the length of the section starting from the leakage point to the next node.

$$length_{i_w} + length_{i_h} = length_i \qquad (3.3)$$

From the circuit in Figure 40 , three impedances are defined according to the different parts of the model:

$$Z_{ad} = R_{i_1} + jX_{i_1} \tag{3.4}$$

$$Z_{bd} = R_{i_2} + jX_{i_2} \tag{3.5}$$

$$Z_{cd} = R_{iso} + R_g \tag{3.6}$$

To apply a load flow calculation method of voltages, these three impedances Z_{ad}, Z_{bd} and Z_{cd} (represented by equation (3.4) to (3.6)) are converted to an equivalent delta connection circuit, which is represented in Figure 41.

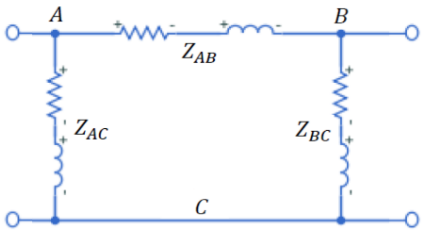


Figure 41: Equivalent electrical circuit model of the damaged line.

This new PI model has resistive parallel branches with different resistance values depending on the position of the leakage point. The delta model parameters are based on the following equations:

$$Z_{AB} = \frac{Z_{ad}\,Z_{bd} + Z_{bd}\,Z_{cd} + Z_{cd}\,Z_{ad}}{Z_{cd}} \tag{3.7}$$

$$Z_{BC} = \frac{Z_{ad}\,Z_{bd} + Z_{bd}\,Z_{cd} + Z_{cd}\,Z_{ad}}{Z_{ad}} \tag{3.8}$$

$$Z_{AC} = \frac{Z_{ad}\,Z_{bd} + Z_{bd}\,Z_{cd} + Z_{cd}\,Z_{ad}}{Z_{bd}} \tag{3.9}$$

3.4. Proposed Method to analyze uncertain degradations on network voltages

3.4.1. Principle

The impact of the insulation degradation of the line on the nodal voltages is now analyzed. Monte Carlo simulations are used to create wide series of scenarios that cover the possible values of the unknown insulation resistance of the line. The impacts on the nodal voltages are then evaluated by load flow calculations. The overall procedure of the proposed method is depicted in Figure 42.

3.4.2. Monte Carlo (MC) algorithm

In the proposed MC tool, firstly, the possible variation range of the uncertain variable (insulation resistance of the line) is defined. Then, a proper number of points is created between the lower and upper bounds of the defined range according to the desired accuracy. Afterwards, the probability density function corresponding to those points is obtained by calculating their

standard deviation and mean value. Then, the obtained probability density function is transformed into a cumulative distribution function (CDF).

In order to create *N* scenarios for the studied uncertain parameter (insulation resistance of the cable), a sampling procedure is applied to the obtained CDF. A uniformly distributed random value between 0 and 1 is chosen. It is assigned to the CDF on the vertical axis and its corresponding value on the horizontal axis gives the variation that the uncertain parameter under study can have in one scenario. The sampling procedure is repeated *N* times in order to create *N* scenarios.

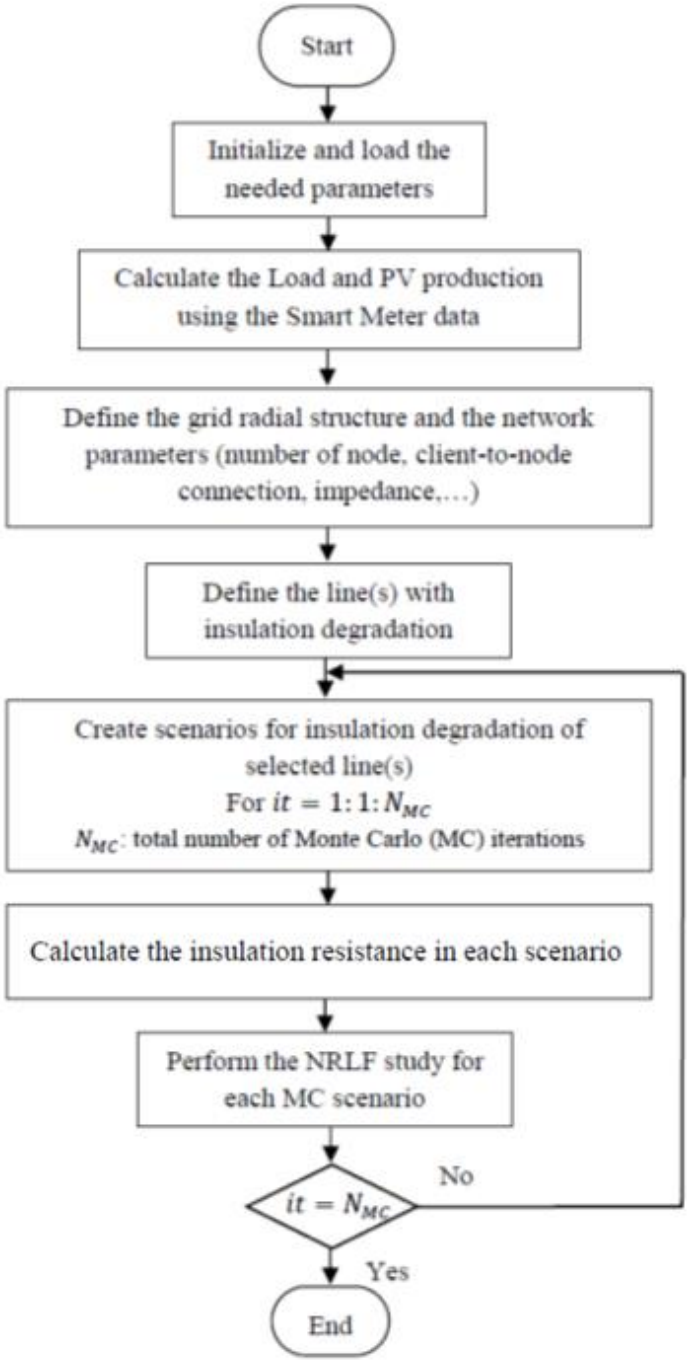


Figure 42: Overall procedure of the proposed method (NRLF : Newton-Raphson load flow)

3.4.3. Newton-Raphson load flow (NRLF) study

The Newton-Raphson load flow (NRLF) study is used to calculate the network voltages for all scenarios created by the MC simulations. In the NRLF, the nonlinear algebraic equations of the nodal powers are linearized by expanding them through Taylor series. They constitute the so-called Jacobian matrix, which gives the linearized relationships between the small changes in the real and reactive powers with respect to the small changes in the nodal voltage angles and magnitudes as below:

$$\begin{bmatrix}\Delta P\\ \Delta Q\end{bmatrix} = \begin{bmatrix}\frac{\partial P}{\partial \theta} & \frac{\partial P}{\partial V}\\ \frac{\partial Q}{\partial \theta} & \frac{\partial Q}{\partial V}\end{bmatrix}\begin{bmatrix}\Delta \theta\\ \Delta V\end{bmatrix} = \begin{bmatrix}J_1 & J_2\\ J_3 & J_4\end{bmatrix}\begin{bmatrix}\Delta \theta\\ \Delta V\end{bmatrix} \quad (3.10)$$

$\Delta\theta$ and ΔV denote the vectors of small variations in the voltage angles and magnitudes at the P-Q buses, respectively. Also, ΔP and ΔQ are the vectors of errors between the scheduled and calculated powers at the P-Q (load) buses. The mathematical relations to obtain the elements of the Jacobian matrix have been given in [84]. The document [85] proposed a structured process to build the power flows equations in NRLF for an AC system. The flowchart to calculate the Jacobian matrix from equations (3.7) to (3.9) are explained by the figure below:

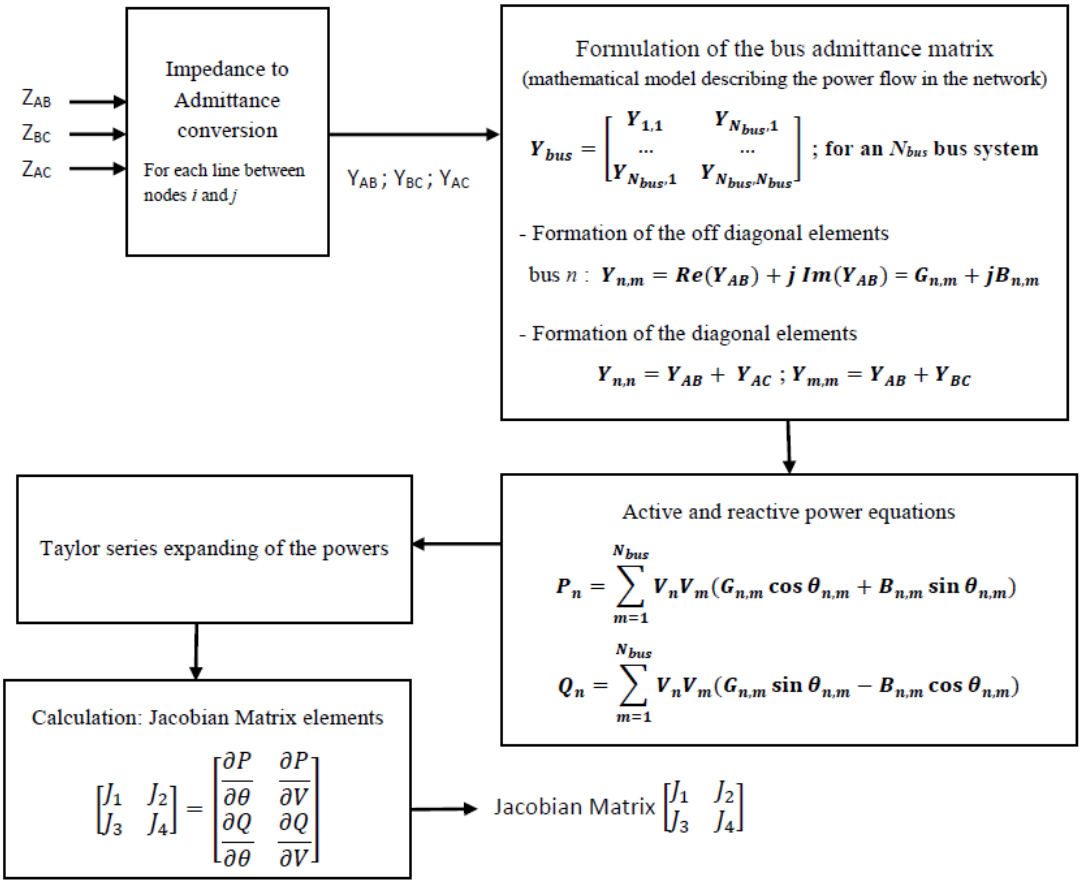


Figure 43: Flowchart to calculate the Jacobian matrix in NRLF.

Once the content of Jacobian matrix is determined, using ΔP and ΔQ, which are errors between the sensed powers and the computed ones, system voltages are updated. Then, the new

voltages are used to update ΔP and ΔQ vectors as well. In the next iteration of the NRLF, the Jacobian matrix elements will be updated in order to obtain the new voltages and eventually new ΔP and ΔQ. The NRLF in this iterative-based procedure minimizes the errors (ΔP and ΔQ). The iterative procedure stops when a predefined error threshold is met.

3.5. Investigations on impacts of cable degradations

3.5.1. Followed scientific method and definition of study cases

The proposed method based on the MC simulations and NRLF calculations is used here in order to evaluate impacts of insulation degradations on the nodal voltages. The cable insulation degradation is taken into account according to the formulation developed in section 3.2. It is supposed that due to cable degradations, the thickness (radius) of insulating material is reduced. The radius of the insulating material is considered as an unknown (uncertain) variable, which will be characterized through the created Monte Carlo scenarios. 3000 scenarios are created during the MC study. The NRLF study is then performed in each created scenario to calculate the nodal voltages.

For the scenario building, the insulation radius of the cable is assumed to vary between 0.01 mm and 1 mm. The former value corresponds to an extreme insulation degradation and 1 mm to moderate degradation situations. The nominal insulation radius of the studied cable is equal to 1.5 mm [86].

The conductor cross section of the studied cable is equal to 50 mm2, from which the radius of the conductor is equal to 3.9 mm. The insulation resistance (R_iso) is calculated by using equation (3.2) considering the created scenarios for the insulation radius variations. It is supposed that the insulation degradation occurs in a section of cable with 1 meter length (L=1 meter). In addition, the resistivity of insulating material (ρ) is considered equal to 3.9 MΩ cm [87]. It should be noted that, theoretically, the insulating materials of cables (e.g. PVC) have a volume resistivity bigger than 10^9 Ω cm. However, in this work, in order to consider the deterioration of the insulating characteristic (besides the reduction of thickness of insulating materials), a lower value of resistivity is taken into account. In reality, the resistivity of the cable insulation decreases with the ageing, moisture, voltage and temperature increase.

In order to evaluate impacts of insulation degradations, three study cases (as presented in Table 15) are defined on the studied LV network presented in Figure 11 of section 2.2.

Table 14: Considered study cases

STUDIED CASE	DESCRIPTION
Case 1	Insulation wear at the beginning of the feeder (line between nodes 2 and 3)
Case 2	Insulation wear at the end of the feeder (line between nodes 13 and 14)
Case 3	Simultaneous insulation wear in the selcted lines i.e. lines between nodes 2 and 3, 8 and 9, 14 and 15, as well as, 17 and 18

The nominal voltage at bus 1 is equal to 230 volts and considered as a constant value (i.e. bus 1 is the slack node).

In order to study the impacts onto the operating conditions of the LV network, the three different locations of degradations will be tested in two situations: heavy load demand with a light PV generation and high PV generation with a low load demand. The former operating point studies the impacts of the insulation degradation on the network having initial voltage drops, while the latter one investigates it in the network with initial voltage rises.

3.5.2. Operation in high load and low PV generation conditions

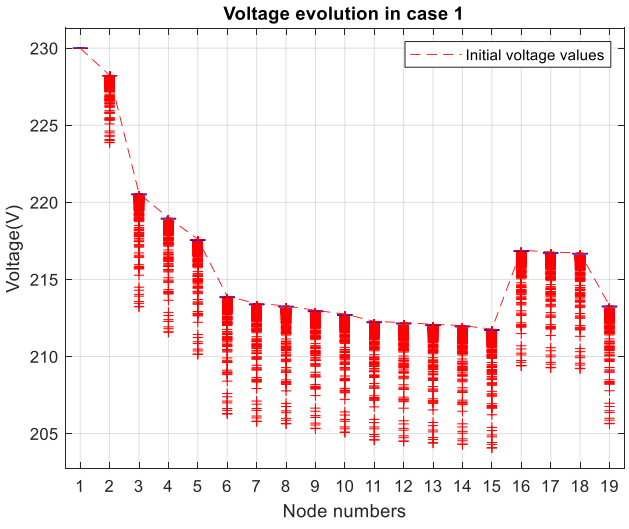
In the first studied operation conditions, load demands are assumed at 85% of their nominal values with no PV productions. This situation with a high load demand should induce voltage drops. Figure 44 shows the boxplots of obtained nodal voltages in considered scenarios for the three studied cases. The box includes the 75th to the 25th percentiles of the voltage profile.

The dashed curve shows initial voltages with healthy cables (no insulation degradation).

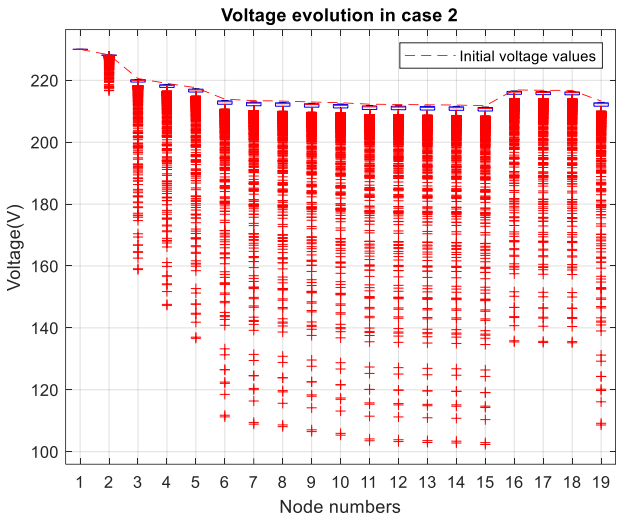
As it can be seen in Figure 44(a), when the cable insulation degradation happens near the bus 1 (case 1), the boxplot of voltage variations has a small range. However, in the extreme insulation degradation scenarios, voltage drops to -10 volts are observed.

In Figure 44 (b), in case of a degradation far from the slack bus (case 2), the boxplot of nodal voltage variations becomes larger. This is explained by the fact that the leakage current in the branch between nodes 13 and 14 passes through whole lines in the upward direction of node 13, which causes more voltage drops compared to case 1. In the extreme insulation degradation scenarios of the case 2, the nodal voltages reduce to around 100 volts, which indicates more than 115 volts of voltage reduction compared to the voltages of healthy lines.

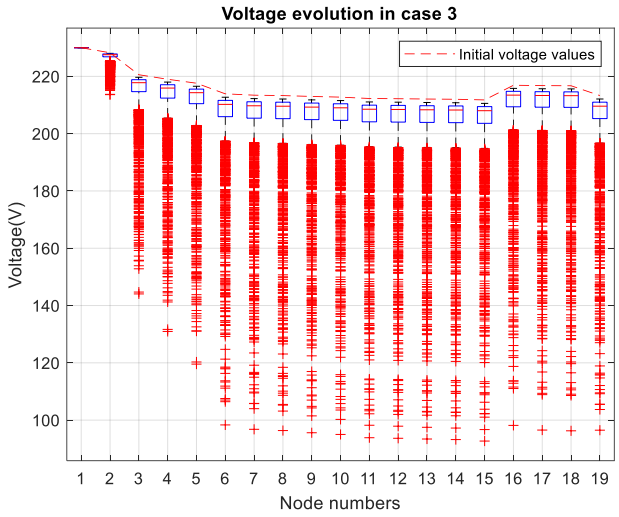
Finally, in Figure 44(c), for simultaneous insulation degradations in four lines, the boxplots of nodal voltages are noticeably enlarged with respect to the ones in cases 1 and 2. In the extreme insulation degradation scenarios of case 3, the voltages can go down to 90 volts.



(a)



(b)



(c)

Figure 44: Evolution of voltages in heavy load - light PV generation situation: (a) case 1 (damage between nodes 2 and 3); (b) case 2 (damage between nodes 13 and 14); (c) case 3.

For all studied cases, most of voltage samplings (in the boxplot) are over the minimum voltage level (90% of 230V) and the standard is satisfied. The red points in the above figures highlight the voltages values outside the boxplot in the created scenarios.

As it can be seen, the extreme voltage drops appear in the voltage results as outliers. The outliers in the voltage results are justified by the fact that the insulation conductance ($1/R_{iso}$) applied in the NRLF study is a nonlinear equation. Consequently, for specific low values of insulation resistance, the insulation conductance noticeably increases, which leads to extreme voltage drops shown by the outliers.

Moreover, the difference in voltage threshold between nodes 15 and 16 is linked to the fact that node 15 is at the end of the network and therefore has the lowest position, while node 16 is a branch of node 5 in the middle of the network.

3.5.3. Operation in low load demand and high PV generation conditions

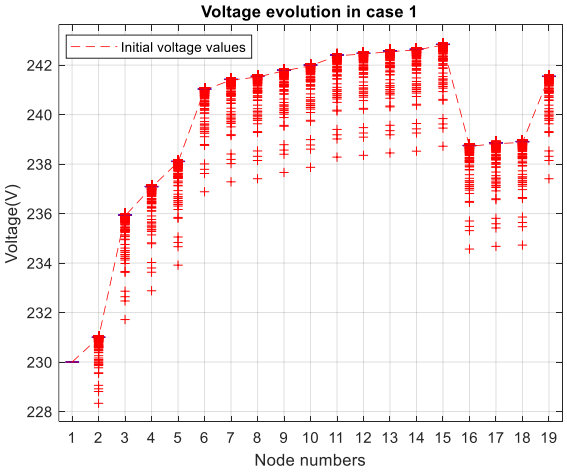
In the second studied operation conditions, the load demands are considered low (equal to 13% of their nominal values) while PV generations are at 80% of their rated values (on each node). As it can be seen in Figure 45, voltage rises appear and are caused by the PV generations when there is no cable insulation degradation.

In Figure 45(a), when the insulation degradation happens at the line between nodes 2 and 3 (i.e. case 1), the node voltages decrease by around 4-5 volts. In addition, the boxplots in Figure 45(a) are very small because the voltage variations are very small among the different scenarios.

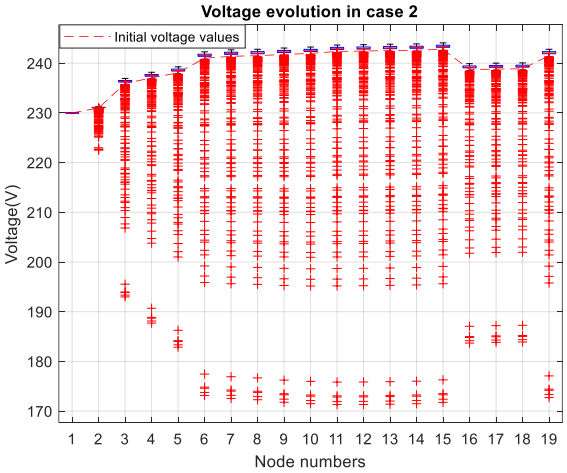
If degradations are located at the end of the feeder (case 2), the nodal voltage variations are considerably reduced and the heights of boxplots are very small (Figure 45(b)). In the most extreme degradation scenarios, the nodal voltages can reach 170 volts.

In case 3, with simultaneous insulation degradations along the line, the boxplots of nodal variations become larger, and the voltages drop to around 140 volts (figure 44.c).

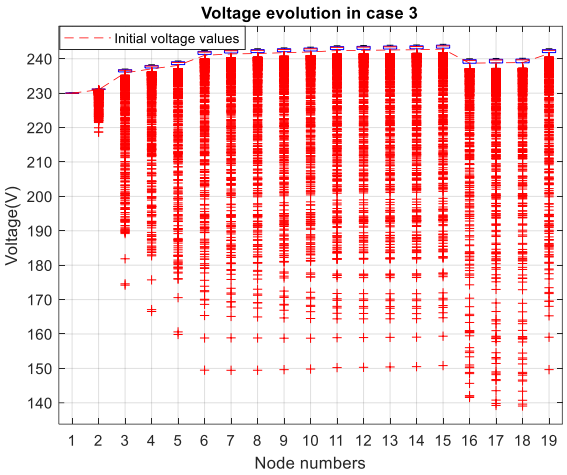
For all studied cases, most of voltage samplings (in the boxplot) are under the maximum voltage level (110% of 230V) and the standard is satisfied. Moreover, we can note that most of voltage samplings are concentrated and does not vary a lot.



(a)



(b)



(c)

Figure 45: Evolution of voltages when the PV generations are high and load demands are less important : (a) in case 1; (b) in case 2; (c) in case 3.

3.6. Conclusion

In this chapter data from smart meters are used to investigate the influence of the degradation of the insulating line material on the voltage variations.

Due to the uncertain nature of the cable insulation degradation, a probabilistic framework based on the Monte Carlo simulations is developed to generate different uncertain degrees of the insulation degradation. The load flow calculations are then carried out in order to obtain the nodal voltages associated to each created scenarios. Regarding the extensive scenarios, the insulation wear of cable lines can cause important voltage drops of more than 50% depending on the position and degree of the cable insulation degradation. The exposed method provides the distribution of nodal voltages variations at different nodes of the studied network. With these results, the probability of voltage variations can also be established.

However in this study, the various generated states of the insulation wear have been generated as probability values because the ageing process of cable insulation materials is not known in real time. This chapter, then, points out the interest of implementing an automated framework for the classification of electrical Low Voltage cables degradation as long as the degree of degradation is not a self-updating quantitative variable. To do so, chapter 4 is oriented towards the development of a machine learning-based classification tool for this purpose.

Chapter 4: Design of a Machine Learning-based classification tool for detecting electrical Low Voltage cables degradation

Contents

4.1. Introduction

Currently, the monitoring of HV transmission lines with specific meters and communication systems is a cost-effective approach. In LV distribution systems, it would be very expensive to deploy sensors and dedicated information and communication technologies for monitoring cables and lines because the entire electrical network is large with many points of common coupling. Moreover, as explained in the introduction, part 4, the architecture, sizing and parameters of existing lines in LV networks are poorly known. Without satisfactory mathematical models (which are based on the physics), a data based modeling can be an advantageous alternative to monitor the correct operating of LV systems.

A variety of data analysis techniques can be applied to extract a meaningful knowledge from large databases. Among these techniques, an important feature present in Machine Learning methods is their ability to be adaptable and so parametrized with the data. The evolution of learning theory in recent years can be explained in particular through the development of data servers for storing information and the rise of Big Data methods. To tackle the challenge of monitoring LV networks, the research work in this chapter aims to take advantage of available data from smart meters and test Machine Learning (ML) capabilities in order to detect the soft degradation of cable insulation at early-stage and regardless of the type of fault.

So, in the second part of this chapter, we are going to present the general scheme of used Machine Learning algorithms using smart meter measurements for the monitoring of low-voltage distribution grids.

Before the development for the LV network application with ML algorithms, data bases have to be generated, e.g. useful data that contain operating points of the LV network systems in normal operation conditions and degraded operation conditions. In the third part, a framework is proposed to generate and organize data that will be used by algorithms. Thanks to the previously presented line model (chapter 2, part 2.2), impedance values are generated for different conditions related to the thickness. Then, within sensed data by smart meters, node voltages are estimated with Low Flow calculations and a “knowledge” database is built. The “knowledge” data base is split into two subsets: a “training subset” and a “test subset”. The “training subset” will be used to implement and parameter the considered ML techniques for classifying the state of lines. The “testing subset” is containing data not used previously during the learning stage. It will be used to evaluate achieved performances and compare the considered ML techniques.

In the fourth part, graphical tools are presented to analyze the results and interpret them. The best ML technique is the one, which gives minimum classification errors.

In part five, various ML are experimented and assessed within the same “knowledge” database (and so conditions). The purpose is to classify the cable insulation wear and so automatically detects faults.

4.2.State of the art on implemented Machine Learning methods

4.2.1. Objective of the investigation

Thanks to measurements from smart meters (part 2.2 and 2.3), two types of input data are available: the net demand and the node voltage at each network node and every quarter over a

time horizon. Relying on these data, we are going to explore, apply and compare different machine learning techniques to detect the operating condition of lines and cables of the electrical network. This problem is a classification, also known as pattern recognition, discrimination, or supervised learning.

In a classification task the response variable is discrete and is known as the class variable, and its values as the classes. Here, a pattern k is defined by a vector of input data x_k and an output y_k. In a simple and first approach, two states have to be decided for each line (between two nodes of the network): no insulation wear or insulation wear of the line. So, the considered output data y_k is a discrete value (that may correspond to +/- 1 or a label). This separation of objects into two categories, or "classes", corresponds to a binary classification problem. Learning then consists of determining a function f such that $f(x_k) = y_k$.

Indeed, we are going to consider supervised learning approaches to design and parameter ML techniques with a dataset of input data corresponding to known output data (labeled data). The building of this learning database will be detailed in part 4.3.1. Figure 46 presents a visual representation of the supervised ML process. The objective is to generate automatically knowledge rules (so-called the **Model**) from a database containing "samples" of inputs (so-called the **Data**) with the corresponding outputs. With new input data (represented by the **circle** symbol in Figure 46), a classification of the observed operating mode can be predicted into two classes (represented by the **star** and **square** groups in figure 45).

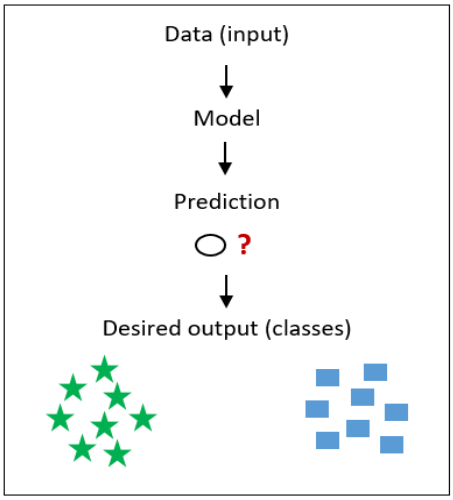


Figure 46: Visualization of supervised Machine Learning approach.

The objective of a classification task is to derive a rule or set of rules, which determine the class each input data belongs, or is most likely to, belong to. These rules must be determined from another set of data (the training data set), whose class values are known. These supervised learning approaches can be divided into two categories [88]:

- Classification methods dispatch the input observations in categorical groups and lead to the construction of predictive models with discrete responses (y_k.).
- Regression methods describe the relationship between inputs variables (so called predictors) and the outputs (through a mathematical function) and lead to the construction of predictive models with continuous responses.

There are many methods used for classification. In following parts, we are going to present the general scheme of the supervised machine learning methods that have been used in this

research work. As the corresponding algorithms are well known and are available as “open source” codes, we will not present implementation details.

4.2.2. k-nearest neighbors algorithm

The k-nearest neighbor (kNN) is a supervised Machine Learning (ML) algorithm that can be used for classification and regression models. A kNN algorithm classes new unlabeled data according to their similarity to *k* existing “reference” data, which are called neighbors, in a training set (kNN representation [89].Figure 47).

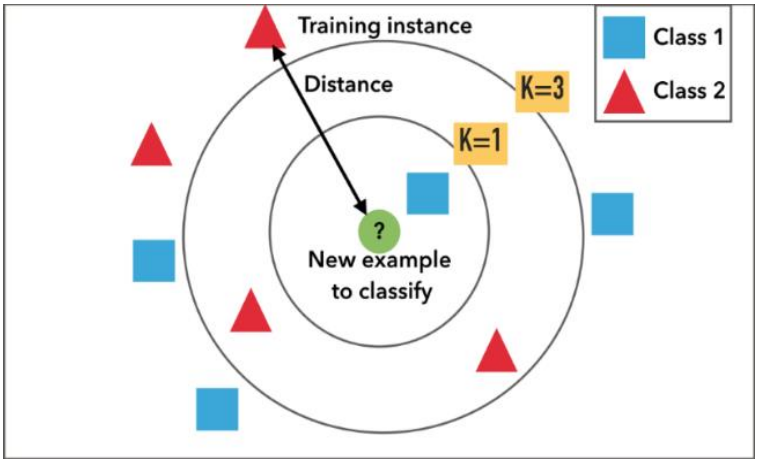


Figure 47: kNN representation [89].

Each new observation *x* is compared to those that already exist by using a distance calculation (such as the Euclidean distance, cosine of the angle formed by the two observations, etc.). Then, the algorithm assigns to the observation *x*, the class, which is the most frequent among the *k* examples of the training set closest to *x*. Hence, the class with the *k* smallest distances is assigned to *x*. Hence, KNN algorithm assumes the similarity between the new observation and available cases and put the new observation into the class that is most similar to the available categories.

The algorithm therefore requires knowing *k*, the number of neighbors to consider. For choosing the right *k*, the kNN algorithm can be run several times with different values of *k*. Then the right number of neighbors *k* will be the one that has led to the best performance (i.e., the lowest error and the best prediction accuracy).

For classification applications, kNN have no parameters we can tune to improve the performance. Strictly speaking, kNN does not have any learning involved and no objective function to optimize. kNN supports non-linear separations and can only provide a class (label) as an output. Studies have proved that kNN is a simple but highly efficient and effective algorithm for solving real life classification problems (such as the recommendation of movies on NETFLIX) [90, 91]. In electrical engineering applications, kNN is mostly used for fault detection and classification but also for power quality classification. The kNN algorithm has also the advantage to be a versatile and easy to understand and to implement method with no need of initial assumptions. However, when the volume of samples in the dataset (so called predictors) increases, the kNN algorithm tends to become slower. Even if there are more precise classification algorithms, kNN remains a first and simple algorithm to model a classification

problem and can achieve a high classification accuracy in problems with unknown distributions, while familiarizing with the available database.

For this study, the kNN algorithm has been implemented with the Euclidean distance as the distance measure because of the ease of calculations and possible manual checking of obtained results. Also, a limited number of neighbors (k=5) has been applied.

4.2.3. Decision tree

Decision Tree (DT) is a supervised Machine Learning (ML) algorithm used in both regression and classification problems (usually called CART : Classification And Regression Trees). For classification purposes, DT is a largely used non-parametric method.

A decision tree is based on a hierarchical representation of the data structure in the form of sequences of decisions (tests) with the goal of predicting a class. The root represents the whole domain. Each level of the tree represents a partition, with nodes at that level representing partition cells (domain subsets). So, the end-nodes are the classification and the intermediate nodes are the tests on the properties of the observations (each circle of Figure 48). In other words, building a decision tree is a recursive process, going from the properties (drawn by branches) to the conclusions about an observation (drawn by leaves). In this data modeling, each observation is described by a set of intermediary variables, which are tested in the nodes of the tree. Testing is done in internal nodes and decisions are made in leaf nodes.

Let us consider a binary classification problem where we have to classify a student in two classes, the obtained results from the ML technique can be correct (true) or false in each class. This classification will be made based on his result to three different exams ($\boldsymbol{X_1}$, $\boldsymbol{X_2}$ and $\boldsymbol{X_3}$ called in scientific jargon **the features**). If the student has at least a certain value (so-called here the **median**) in each of the three exams, he will **success** the semester. However if he has less than the **median** value in one of the exams, he will **fail** the semester.

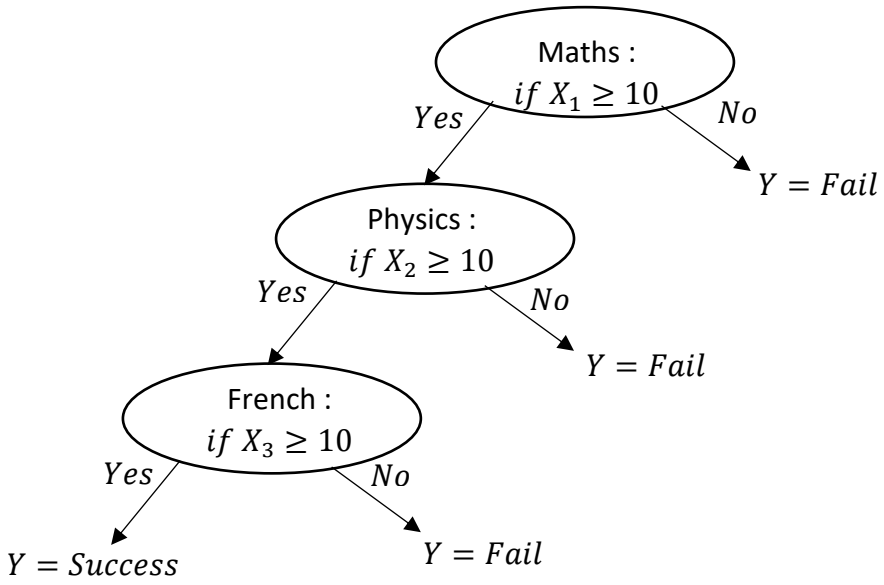


Figure 48: Decision tree representation.

The decision tree starts with a root node (property of X_1 in Figure 48) and branches toward possible outcomes. Each of those outcomes leads to additional nodes (property of X_2 and X_3), which also branch toward other outcomes. In other words, it is a visual representation of the decision-making directly related to the problem to be solved.

A Decision Tree is a commonly used and highly understandable Machine Learning method. It is a reliable algorithm for separating a dataset (predictor variables set) into several given classes by providing some clear indications about the most relevant predictors. For classification problems, a DT algorithm does not need much computation and does not rely on functional assumptions (i.e., it is not affected by any non-linearity) while it can build very complex trees and encounter overfitting problem. Also, the creation of optimal decision trees can be obstructed by the presence of dominating classes. DT accuracy reduces however when the number of training examples to the number of classes is low. Decision Tree is a widely used algorithm that gives high-quality results with the data, which mostly depends on conditions [92, 93]. In electric power system applications, DT is used in load consumption prediction and load forecasting, preventive and corrective control, power systems security assessment, etc. [94]. The DT algorithm, in this study, is an adjusted binary classification decision tree.

4.2.4. Logistic regression

The Logistic Regression (LR) is a parametric model, which supports linear solutions and can derive to a high confidence level (regarding its prediction). LR is a powerful algorithm for finding boundaries between two classes. Mathematically, a LR algorithm uses regression to predict the probability (between 0 and 1) of a new observation *x* to be classified into *y*, a given class (see Figure 49).

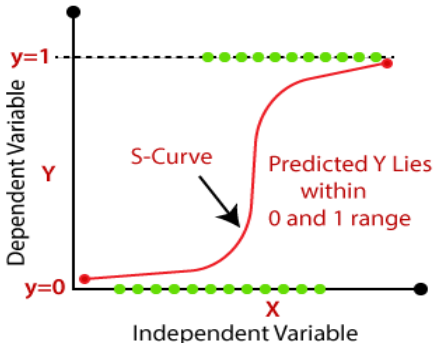


Figure 49: Logistic regression representation [95].

A mathematical representation of LR is now developed. Considering a two-class classification problem, an analogy is made between the labels and the output classes as shown in Table 15.

Table 15: Analogy between labels and classes.

State of the line	Labels/Class	Class in LR (variable *y*)
No insulation wear	H	1 for the positive class
Insulation wear	M	0 for the negative class

The output $h(x)$ of a Logistic Regression model is the probability of a new observation *x* to be classified into a class *y* and will be bounded by a function *h*, as below:

$$0 \leq h(x) \leq 1 \tag{4.2}$$

$$\begin{cases} If\ h(x) \geq 0{,}5 : predict\ y = 1 \\ If\ h(x) < 0{,}5 : predict\ y = 0 \end{cases} \tag{4.3}$$

For this classification problem, a sigmoid function *h* can be used to map predictions to probabilities:

$$h(u) = \frac{1}{1 + e^{-u}} \tag{4.4}$$

The input of the sigmoid function (u) is the weighted sum (weights: θ) of the observation vector inputs (x). Then $h_\theta(x)$ can be written as below:

$$h(u) = h(\theta^T x) = h_\theta(x) \tag{4.5}$$

θ is a vector of weights with the same size as the observation input vector *x*. The key point is then to find the right values for parameters θ by solving a minimization problem of a cost function (J):

$$\min_\theta [\, J(\theta)\,] \tag{4.6}$$

The cost function is a measurement of the prediction error over *M* observations (indexed by the variable *m*).

$$J(\theta) = \frac{1}{M} \sum_{m=1}^{M} cost\big(h_\theta\big(x^{(m)}\big), y^{(m)}\big) \tag{4.7}$$

m is the index of observations and *M* is the total number of observations in the training subset. The function *cost* is the quadratic classification error that is expressed as follows [96]:

$$cost(h_\theta(x), y) = \frac{1}{2}\Big(\frac{1}{1 + e^{-(\theta^T x)}} - y\Big)^2 \tag{4.8}$$

$$cost(h_\theta(x), y) = \begin{cases} -\log\big(h_\theta(x)\big)\ if\ y = 1 \\ -\log\big(1 - h_\theta(x)\big)\ if\ y = 0 \end{cases} \tag{4.9}$$

The cost function to be minimized will be equal to:

$$J(\theta) = -\frac{1}{M} \sum_{m=1}^{M} \Big[y^{(m)}\, log\Big(h_\theta\big(x^{(m)}\big)\Big) + \big(1 - y^{(m)}\big)\, log\Big(1 - h_\theta\big(x^{(m)}\big)\Big)\Big] \tag{4.10}$$

Logistic regression method is the go-to method for binary classification problems (problems with two class values). LR is easy to implement, fast and very efficient to train. LR algorithm gives a good accuracy for simple datasets and the provided model coefficients can be interpreted as indicators of predictor importance. LR has the advantages to be less likely to lead to over-fitting except in high dimensional datasets. Logistic Regression methods are used, in electrical

engineering, for electricity monitoring, visualization and prediction but also for fault detection in renewable energy production [97].

4.2.5. Random Forest

Random Forest (RF) are supervised Machine Learning (ML) algorithms used to solve both regression and classification problems. The Random Forest is a flexible ensemble Learning method, which is based on combination of the results from a multitude of decision trees (constructed at the RF training time) – the forests. For classification problems, RF is an effective method because it corrects the overfitting problem of DT (as presented in section 4.2.3). The output of the RF is a more accurate and stable prediction resulting from the merges of the built multiple decision trees (see Figure 50).

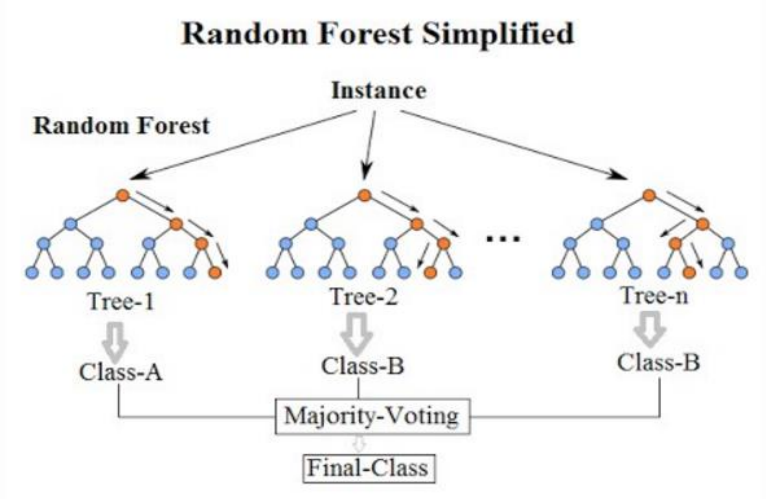


Figure 50: Random Forest representation [98].

There are two main ways for combining the outputs of multiple decision trees into a random forest :

- Bagging (general idea of the bagging method is that a combination of learning models increases the overall result.) ;
- Boosting (similar idea as the previous one except that the samples are weighted for sampling so that samples, which were predicted incorrectly get a higher weight and are therefore sampled more often).

Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes the model's prediction (see figure below). The key effect is that the trees protect each other from their individual errors (as long as they do not constantly all err in the same direction). While some trees may be wrong, many other trees will be right, so as a group the trees are able to move in the correct direction. So, for random forest to perform well, the predictions need to have low correlations with each other. Random forest adds additional randomness to the model, while growing the trees. Instead of searching for the most important feature while splitting a node, it searches for the best feature among a random subset of features. This results in a wide diversity that generally results in a better model. Consequently, in random forest, only a random subset of the features is taken into consideration by the algorithm for splitting a node.

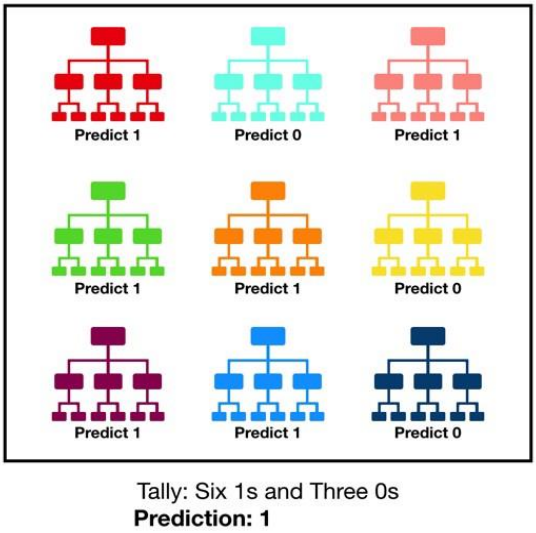


Figure 51: Visualization of a Random Forest Model Making a Prediction [99].

In electrical engineering, the random forest algorithm is used to evaluate power sustainability and cost optimization in Smart Grid as well as to classify faults for active power system networks.

Random Forest has the advantage to be versatile and make it easy to view the relative importance it assigns to the input features. RF produces, even without hyper-parameter tuning, a great result most of the time. It is also one of the most used algorithms, because of its simplicity and diversity.

However, the input data characteristics can affect the RF performance. Also, a large number of trees can make the algorithm too slow and ineffective for real-time predictions. In general, these algorithms are fast to train, but quite slow to create predictions once they are trained. A more accurate prediction requires more trees, which results in a slower model. In most real-world applications, the random forest algorithm is fast enough but in some situations where run-time performance is important, other approaches would be preferred. Our performed work will focus on bagging decision trees for random forest classification in chapter 5.

4.2.6. Support Vector Machine

Support Vector Machine is a supervised Machine Learning (ML) algorithm used in both regression and classification problems. SVM is a robust prediction method that is based on statistical learning frameworks.

The purpose of a SVM is to learn a linear function in the feature space of input data that deviates from the learning outputs by at most a prescribed distance (maximum margin). For a classification problem into N classes, the algorithm finds a hyperplane in an N-dimensional space that distinctly separates the data points. The optimal hyperplane divides a plane into parts where each class has it side. As example, for a binary classification problem, Figure 53 shows hyperplanes that divide the plane into two parts. There are many possible hyperplanes that could be chosen to separate the two classes of data points. The objective of a SVM algorithm will be to find a plane that has the maximum margin, i.e the maximum distance between data points of both classes, in order to correctly separate these two classes. Maximizing the margin distance

provides some reinforcement so that future data points can be classified with more confidence. The data points that are sufficient to determine the maximum margin (optimal separating hyperplane) are the support vectors.

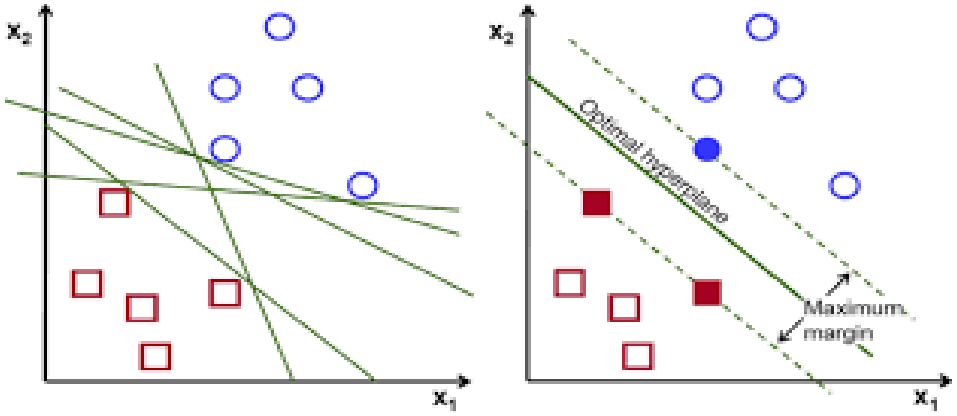


Figure 52: Support Vector Machine representation for two classes [100].

The dimension of the hyperplane depends upon the number of classes. If the number of classes is 2, then the hyperplane is just a line. If the number of input classes is 3, then the hyperplane becomes a two-dimensional plane. For more classes, SVM algorithms build themselves the above hyperplane and so are very interesting for multiclass problems.

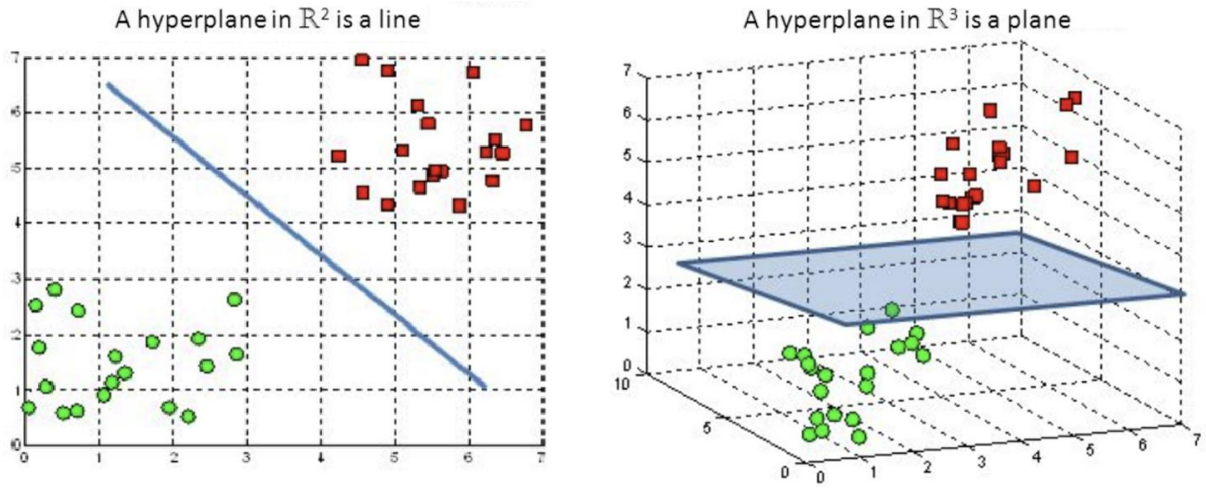


Figure 53: Visualization of a SVM hyperplane for two and three classes problems [100].

The One-vs-One strategy splits a multi-class classification into one binary classification problem per each pair of classes. For the *N*-class instances dataset, *N**(*N* -1)/2 binary classifier models have to be generated. Using this classification approach, the primary dataset are split into one dataset for each class opposite to every other class. The number of class labels present in the dataset and the number of generated binary classifiers must be the same.

The One-vs-All strategy splits a multi-class classification into one binary classification problem per each pair of classes. For the N-class instances dataset, N-binary classifier models have to be generated. The number of class labels present in the dataset and the number of generated binary classifiers must be the same.

The support vector machine is a powerful tool for binary classification, capable of generating very fast classifier functions following a training period. For the purpose of this

study, the SVM algorithm is a full trained, multiclass model, which used the One-vs-All method.

4.3.Proposed approach for the development of a ML based classifier

4.3.1. Building of a knowledge database

In a first stage, a working database is created with the voltage values and Net Demand values at each node of the considered LV distribution network. The Net Demand (ND) data are obtained with SM measurements at each quarter of an hour (q), the PV production PV_i and the Load demand $Load_i$ at each node i, as expressed:

$$ND_i(q) = Load_i\,(q) - PV_i(q) \tag{4.1}$$

To generate the database in healthy cable conditions, only the line model of section 2.4.3.1 is used to calculate the nodal voltage with these SM measurements. To generate the database in degraded cable conditions, thickness scenarios are used to represent the radius variations (according environmental conditions and among seasons, a year,…). They are generated by a random sampling of cdf (chapter 3, section 3.4). Then, the variable values of the cable impedances are calculated. Impedance values are used by a Newton-Raphson Load Flow (NRLF) based algorithm (as explained in section 3.4.3) that calculates the AC node voltages with measured data (each quarter q of each day). Figure 54 shows the flowchart of the synthetic creation of the knowledge database containing nodal voltages (the global flowchart of the proposed approach including the classification process is presented in Figure 61).

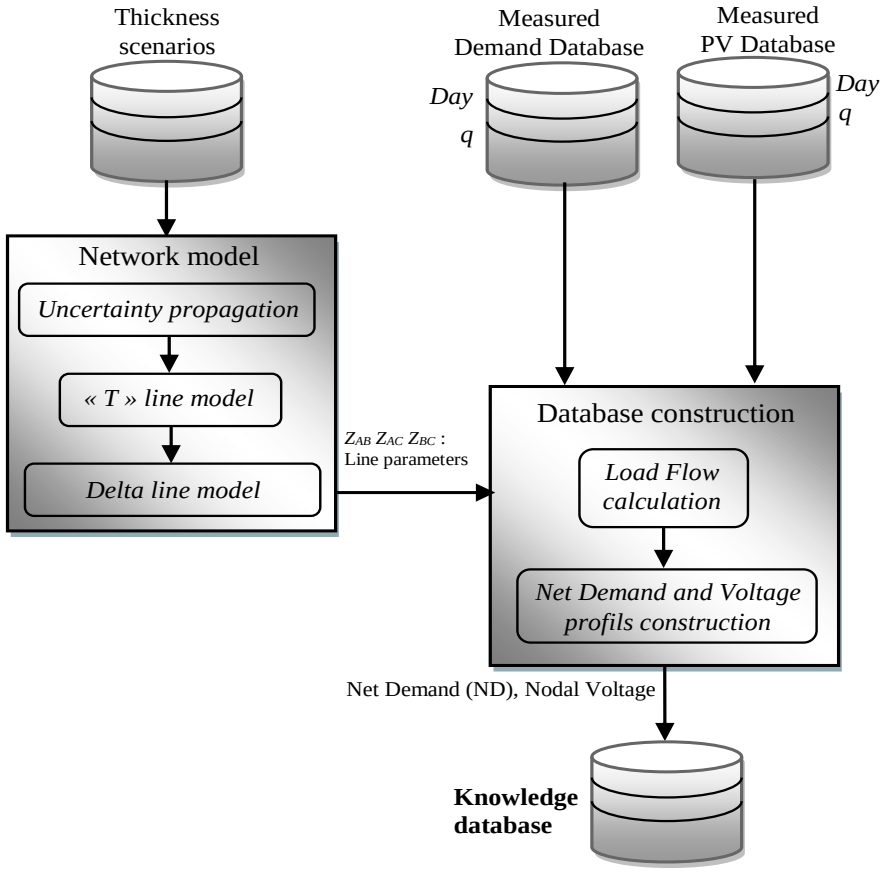


Figure 54: Flowchart of the synthetic creation of the database.

4.3.2. Labelling data

For the evaluation of the cable state, two classes are defined for each observation in the database (see Table 16). The class H is associated with the cables without insulation wear while the class M is used to label the cables that present a certain degree of insulation wear.

Table 16: Label table for the observations.

Cable state	Labels	Description
No insulation wear	H	H for Healthy
Insulation wear	M	M for Medium

For the sake of simplicity, the presented analysis is carried out on a part of the network, which is in the backward direction of node 3 (Figure 11). The input node (i.e., node 2) is connected to customer C1 while the output node (i.e., node 3) is connected to customer C2. The first node (i.e., node 1) is connected to the secondary side of the transformer and is supposed to be at the 230V reference value. In this study, one month of SM data is used to build the knowledge dataset. For each day, 96 measurements are collected. The total number of observations is thus equal to 2880 measurements (i.e., 30×96). Those 2880 observations are created while ensuring to cover a uniformity of both classes (healthy state and medium degraded state) in the synthetic dataset. Table 17 shows how the cable states are distributed in the knowledge database.

Table 17: Observations in the knowledge database.

State	Number of observations
H	1441
M	1439
Total	2880

4.3.3. Building of a training and validation subset

As a first step, a “knowledge” database is created with two types of input data: the net demand and the node voltages every quarter of an hour over a given time horizon. Supervised Machine Learning algorithms consist of two stages namely the learning and the testing stage.

A part of input data with the corresponding output data (class, labels) from the knowledge database are stored in a subset, which is called training subset. A learning algorithm uses this set to adapt and set parameters of the applied ML technique in order to predict the right output (predicted states).

The quality of the ML based classifier is assessed by its ability to generalize, i.e. its ability to correctly classify new examples that have not been used for learning. Then, the remaining observations from the original knowledge database are stored in another subset : the testing subset. During the testing stage, a prediction is made with those input data, the prediction error is calculated with all of these samples and then the accuracy of the parametrized ML algorithm can be assessed with these non-learned data (figure 55).

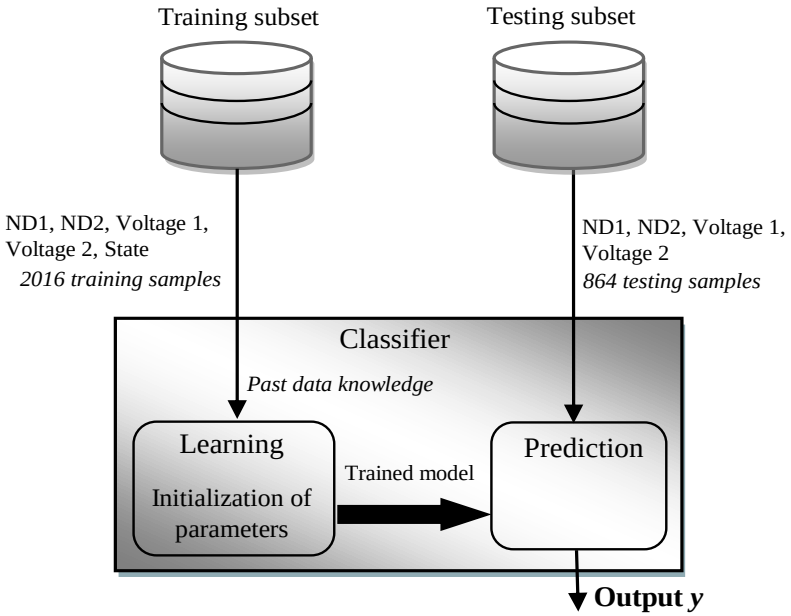


Figure 55: Flowchart of the classification process: from the model training to the implemented prediction.

In practice, for selecting the observations in each data subset, a random logical selection is applied. Table 18 summarize the repartition of the data, which are used in each subset.

Table 18: Partition of the data.

Database	**Percentage**	**Observations**
Knowledge database	100%	2880
Training subset	70%	2016
Test subset	30%	864

Selected data in the training subset are very important because it is directly related to the value and quality of the presented knowledge that will be learned. The training subset has been built by using about half of samplings from each class (or state or label) in order to construct a balanced representation of the targeted classification function through these data (Table 19).

Table 19: Distribution of the observations in each subset.

Observations state	**Knowledge database**	**Training subset**	**Testing subset**
H	1441	1000	441
M	1439	1016	423
Total	2880	2016	864

Figures 56 and 57 show the specified classification process for each implemented algorithm.

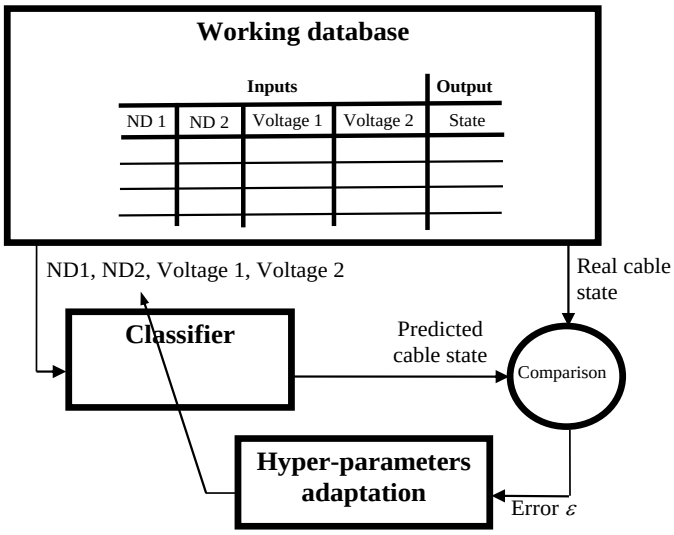


Figure 56: Classification process specified to Decision Tree (DT) and k-nearest neighbor (kNN) algorithms.

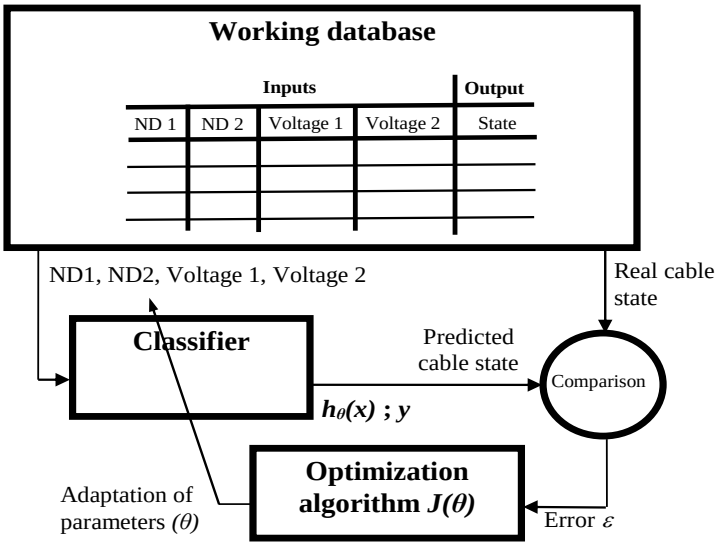


Figure 57: Classification process specified to Logistic regression (LR) algorithm.

4.4. Graphical tools for result analysis and interpretation

In order to analyze the obtained classification results from each ML technique, confusion matrix and Receiver Operating Characteristic (ROC) are proposed to represent and assess their performances.

4.4.1. Confusion matrix: 2D representation

The confusion matrix is an error/accuracy table that helps representing and visualizing the performance of any supervised learning algorithm. For a classification problem in two classes, the obtained results from the ML technique can be correct (true) or false in each class, the confusion matrix is then a 2D representation of results. As an example of a binary classification problem, we consider the prediction, for twenty persons, of pregnancy or not. The result of the test for each person will take one of this two values: **positive** (if the person is pregnant) and

negative (if the person is not pregnant). As for each test, the algorithm can give a correct or a wrong classification. The 2-D confusion matrix is obtained with the results of four evaluations (TP, FP, FN, TN) as presented in the table 20.

TP means True Positive and that the algorithm predicted a positive result and it is true (correct). It also means that we predicted that twelve women are pregnant and she actually is.

TN means True Negative and that the algorithm predicted a negative result and it is true (correct). It also means that we predicted that five men are not pregnant and he actually is not.

FP means False Positive and that the algorithm predicted a positive result and it is false (wrong). It also means that we predicted that a man is pregnant but he actually is not.

FN means False Negative and the algorithm predicted a negative test and it is false (wrong). It also means that we predicted that two women are not pregnant but they actually are.

Table 20: Confusion matrix for two classes prediction.

Output classes			
	positive	TP 12 	FP 1 
	negative	FN 2 	TN 5 
		positive	**negative**

Target classes

For a classification problem in more classes, the size of the confusion matrix is increasing. As example, the 2D confusion matrix of a classification in four classes is shown in Figure 58. 3453 observations are considered in the test database with a balanced sharing among class. Hence 864 observations from each class are representing the Target Class or Correct Class with true values.

For the 3453 observations, the classification technique must retrieve the correct class. Each colon of the confusion matrix shows the obtained performance according to the class. As example, the 864 observations from the H class has been perfectly classed in class H (first colon). Among the 864 observations from the C class, only 476 has been perfectly classed in class C (colon 4), hence 55,1% are well classed (green percentage in the last line). By adding all observations in any colon, 864 classed observations are retrieved.

When observations are badly classed, it is interesting to study and understand the miss-classification. Each line of the confusion matrix analyzes and quantifies the confusion. As example (Figure 58), 1626 classed observations are in class A (second line) but only 864 are

correctly classed and corresponds to 53.1% of observations classed in class A. Bad classed observations in class A are shared in the two other classes B (10,9%) and C (11,2%).

The diagonal cells correspond to observations that are correctly classified. The off-diagonal cells correspond to incorrectly classified observations. Both the number of classed observations and the percentage of the total number of classed observations are shown in each cell. As example (Figure 58), 489 observations from class B have been classed among the 3453 observations of the test database. So, it represents 14,1% of well classed observations. Finally, from the whole test database, 77,9% of observations has been correctly classified and the source of well classified observations among classes is shown in the diagonal.

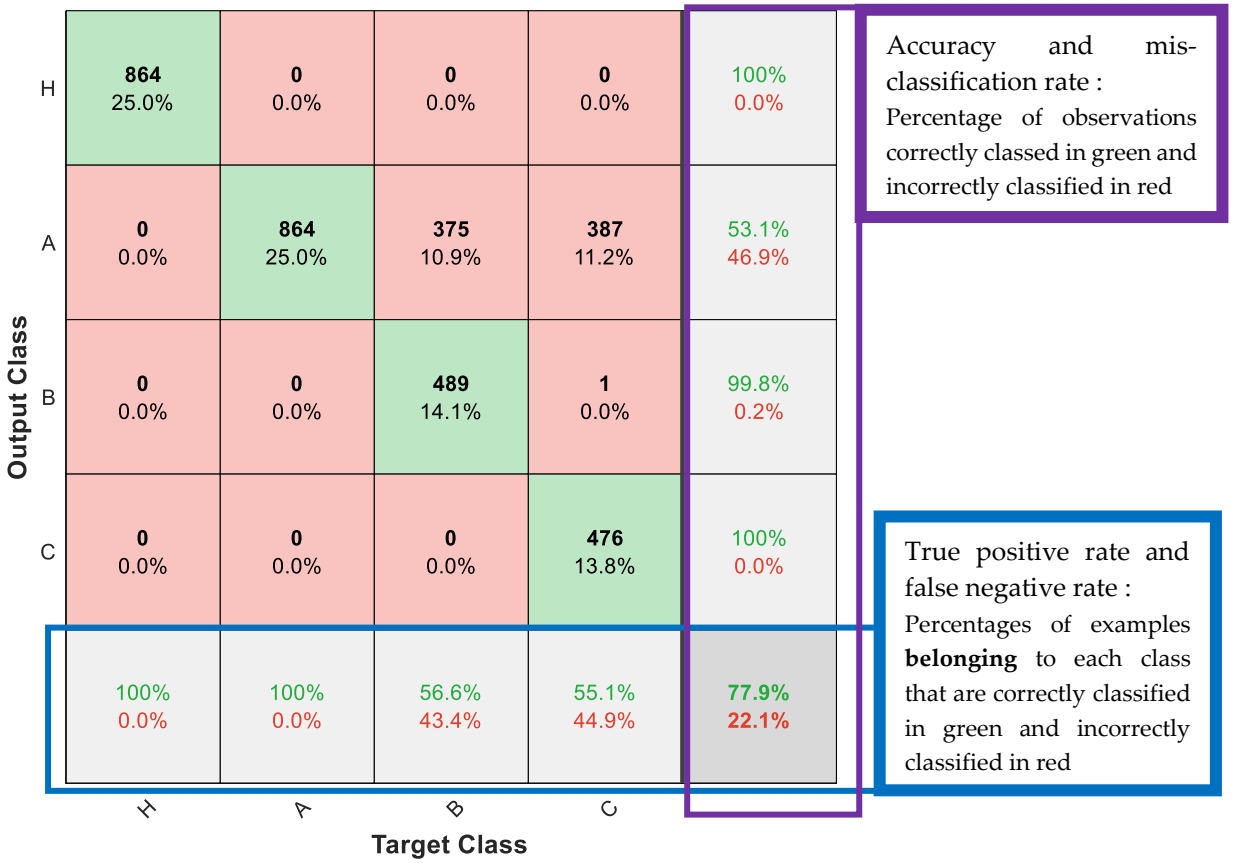


Figure 58: Example of a confusion matrix for four-classes prediction.

4.4.2. Confusion matrix: 3D representation

The 3D confusion matrix is the representation of the previous 2D error matrix with the number of observations in the third axis (see Figure 59).

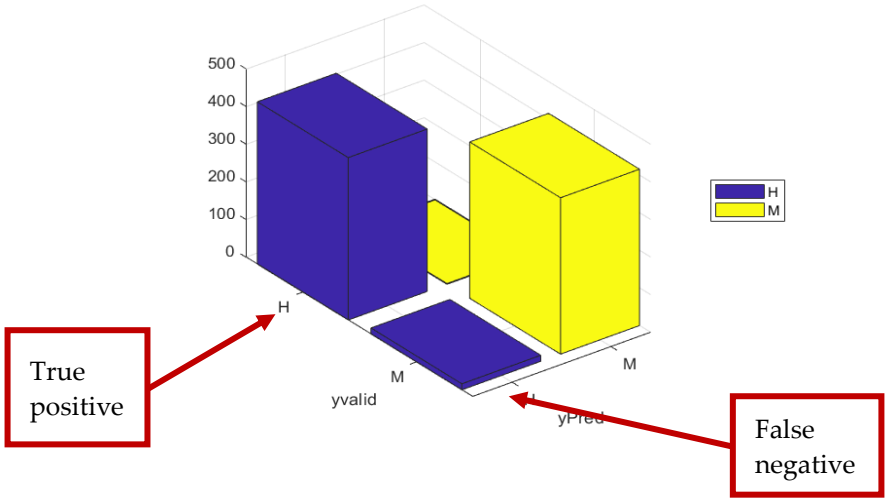


Figure 59: Example of 3D confusion matrix.

4.4.3. ROC diagram

The ROC (Receiver Operating Characteristic) curve is the representation of the diagnosis ability of the classifier. Indeed, the true positive rate (TP ; the sensitivity) is plotted against the false positive (FP ; the probability of false alarm) for each algorithm. As shown in Figure 60, the more the classifier curve is above the diagonal line, the more its performance are good. On the other hand, more the classifier curve is under the diagonal line, more its performance are bad.

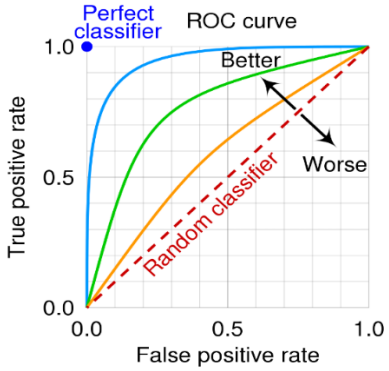


Figure 60: Characteristics of the ROC diagram [101].

4.5. Application for a cable diagnosis

4.5.1. Flowchart of the implemented ML approach

The proposed framework includes each process in Figure 54 and the classification process (Figure 55). Therefore, we obtain the global flowchart of figure 61 (from the SM measurement to the prediction results).

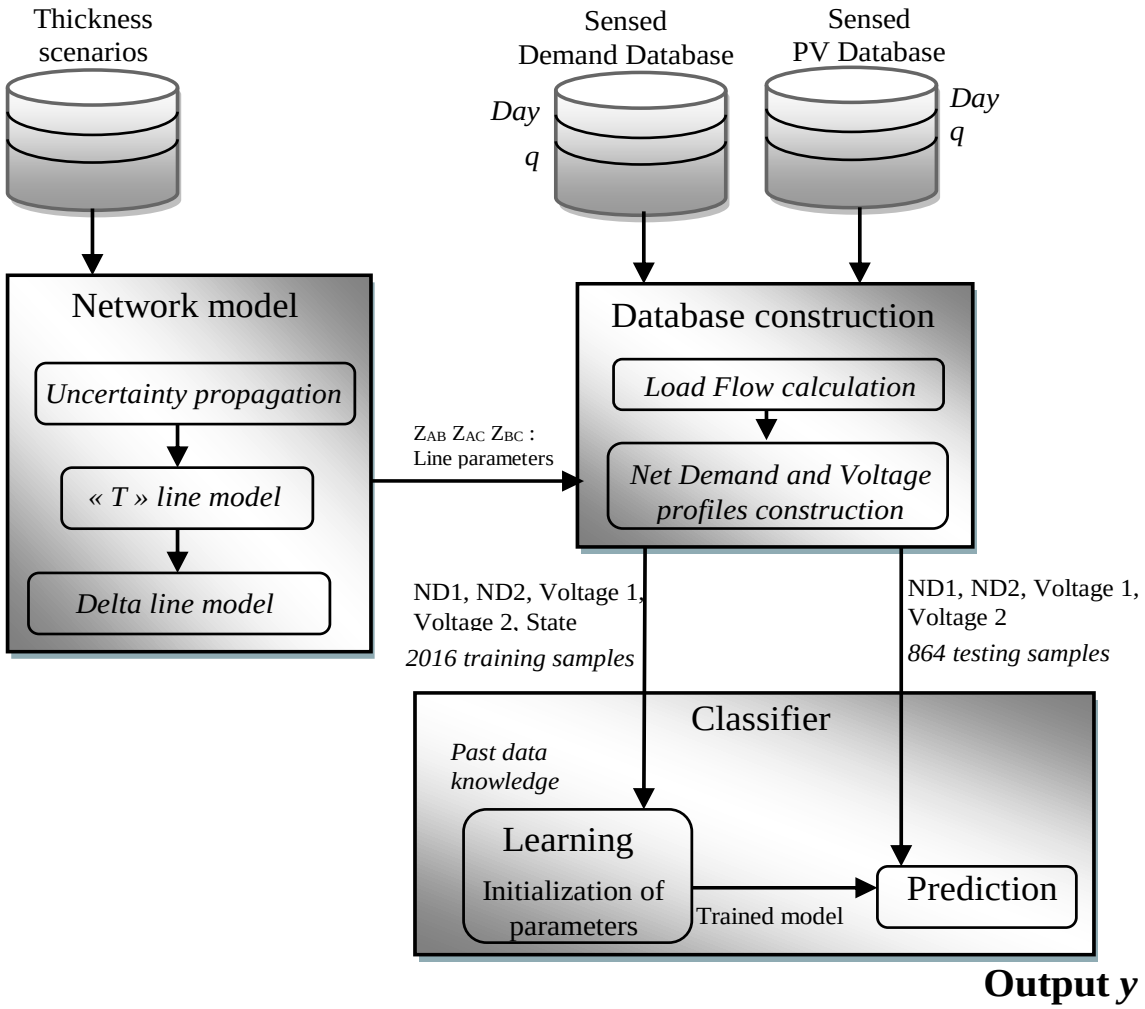


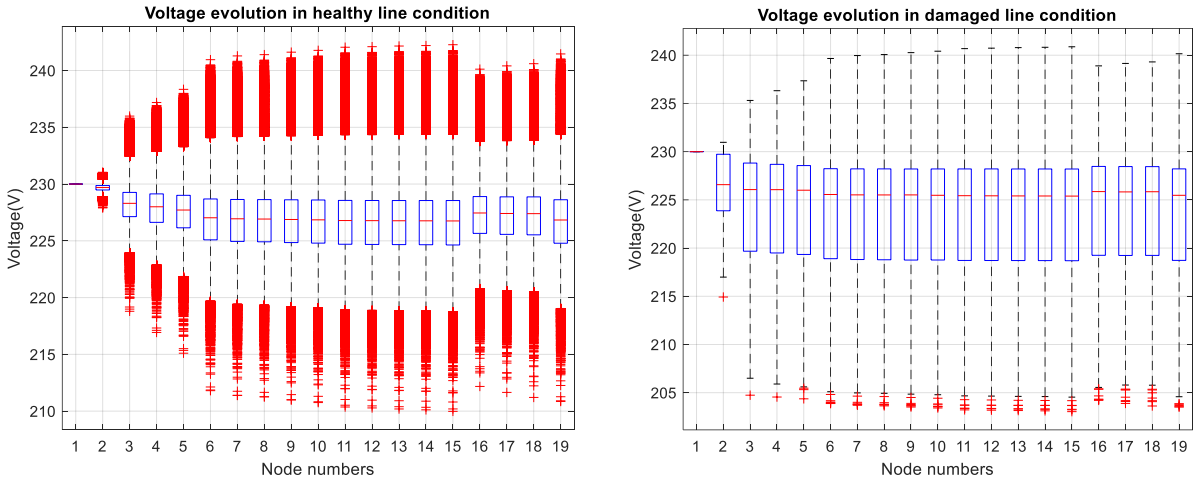
Figure 61: Global flowchart of the implemented approach for the Machine Learning based ELV cable condition classifier.

4.5.2. Analysis of the sensibility of node voltages against damage

A first investigation is carried out in a healthy cable condition to calculate the RMS voltage at each node of the feeder (Figure 11). In this situation, the only cause of voltage variations are the Net Demand variations. The obtained values are limited to the [210 V, 242V] range as shown in Figure 62(a).

Figure 62(b) presents the nodal voltages with the same ND but with a moderated degraded cable located in the line between nodes 2 and 3. It should be noted that the extreme degradation scenarios as studied in [36] have not been considered because these severe faults are easier to observe and to detect as voltage variations are usually outside the standard (+-10% 230V). **The interest in our study is to detect the beginning of degradations very soon**. This process will be useful in managing cable maintenance and in anticipating severe faults or outage. Hence, the moderately degraded cable condition is linked to a soft fault degradation, which will not induce necessarily breakage conditions but just introduces significant variations in the voltage profile. ML techniques will be explored to identify if the cause of voltage variations is due to the Net Demand or to a soft damage of the cable.

In the boxplots of nodal voltage profiles shown in Figure 62, the red positive signs demonstrate the outliers of the voltages in the created scenarios (the box includes the 75th to the 25th percentiles of the voltage profile). The outliers in Figure 62(a) are related to the prosumers ND demand variations while those in Figure 62(b) are due to the variation of the insulation conductance (1/*Riso*). As it can be understood, the increase in insulation conductance (1/*Riso*) leads to the voltage drops shown by the outliers.



(a) (b)

Figure 62: Boxplots of nodal voltages obtained by the Load Flow calculations: (a) for feeder in a healthy cable condition; (b) for feeder with a moderately damaged cable located in the line between nodes 2 and 3.

4.5.3. Diagnosis of a single line

4.5.3.1. Presentation of study cases

The main purpose of this work is to identify if the monitored cable section (i.e., the one between nodes 2 and 3) is either in a healthy working condition (class H) or present any insulation wear (class M). This classification will be made by various ML methods by using an input dataset built from provided Smart Meter data and computed nodal voltages. The degradation of the cable is modeled by a thickness variation. The tested ML algorithms are expected to understand if any variation in the data is related to a cable degradation (based on the Net Demand / voltage level compromise) or to the customer Net Demand. In order to evaluate the importance of chosen data inputs, two cases will be considered and compared.

In **the study case 1**, the Net Demand (*ND*) and the nodal Voltage (*V*) of the node 2 (named ND_1 and V_1) and node 3 (named ND_2 and V_2) are given as input data to the classification algorithm.

In **the study case 2**, only the nodal Voltage (*V*) of node 2 (V_1) and node 3 (V_2) are given as inputs to the classifier. The idea is to evaluate if the tested ML algorithm can really distinguish the effects of thickness variation without the knowledge of the Net Demand variations.

4.5.3.2.Obtained results

Table 21 and

Table *22* show the prediction results obtained by the studied classification techniques for cases 1 and 2.

Table 21: Prediction results for study case 1 with different ML techniques (kNN: k-Nearest Neighbors, DT : Decision tree, LR: Logistic Regression)

Observations state	Real state	kNN	DT	LR
H	433	468	434	447
M	431	396	430	417
Total	864	864	864	864

Table 22: Prediction results for study case 2 with different ML techniques (kNN: k-Nearest Neighbors, DT : Decision tree, LR: Logistic Regression)

Observations state	Real state	kNN	DT	LR
H	433	299	454	524
M	431	565	410	340
Total	864	864	864	864

For the study case 2 (table 22) in which the Net Demand (DT) is not given in the inputs, it can be seen that the implemented algorithms are losing in performance. Therefore, it can be concluded that ND is an essential input for increasing the performances of the classifier by helping to distinguish overloading from cable ageing situations. This is due to its "not to be neglected" impact on the nodal voltage variation range [36].

The constructed tree for DT method is shown in Figure 63 (corresponding to case 1). It reveals that normalized input ND_1 profiles (named x1) will affect the prediction process as well as the normalized output voltage profile V_2 (named x4). As a result, LR and DT lead to predictions with high accuracies in case 1.

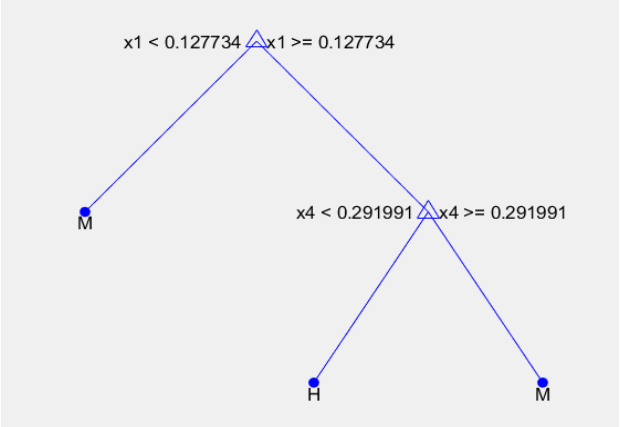


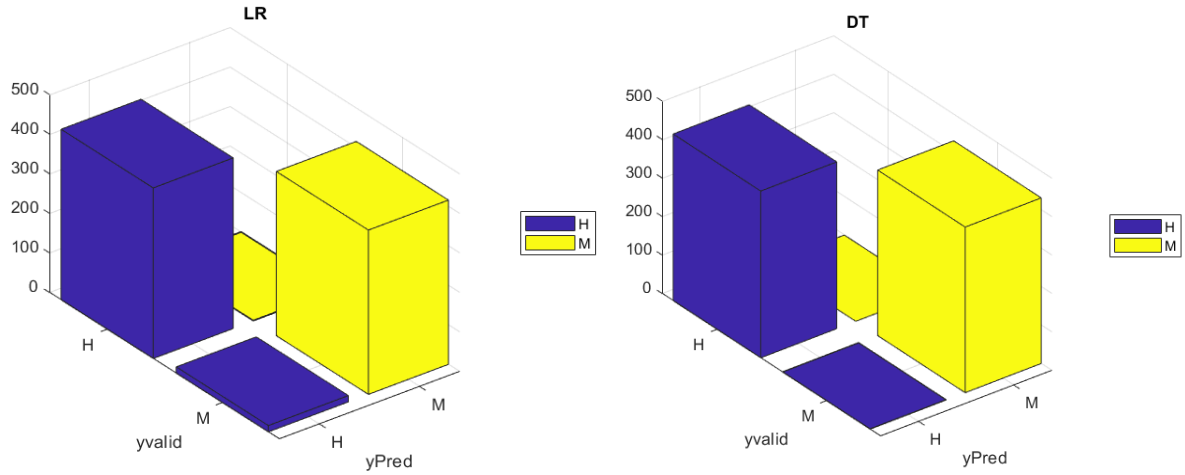
Figure 63: Constructed decision tree in case 1.

Table 23 gives the related training and prediction accuracy for each studied classification method. By comparing these results, it can be concluded that the LR and Decision Tree are great binary classification tools while the kNN method leads to less accurate predictions.

Table 23: Training and prediction results accuracy for application case 1.

Algorithms	Training error	Prediction error (%)	Accuracy (%)
kNN	0.15427	23.264	76.736
DT	0	0.116	99.884
LR	0	2.083	97.917

Figure 64 represents the 3-D confusion matrix of LR and DT methods for the first study case in order to visualize the quality of the classifiers output. The axis yPred and yvalid correspond respectively to the outputs of the classifier (the predictions) and the known cable conditions (real classes from the original dataset). Only few damaged cable conditions could not be predicted with either LR or DT algorithms (small blue block corresponding to 30 observations in Figure 64(a) and 4 observations in Figure 64(b)).



(a) (b)

Figure 64: Confusion matrix of the prediction result: logistic regression method (a) and decision tree method (b).

Figure 65 shows the ROC (Receiver Operating Characteristic) diagram representation of the prediction, which shows the ratio between the true positive (sensitivity) and the false positive (specificity) outputs of the classifier. Knowing that the closer the curve is to 45-degree diagonal of the ROC space, the less accurate would be the prediction result, it can be conclude that kNN is clearly the least efficient algorithm in the studied application case.

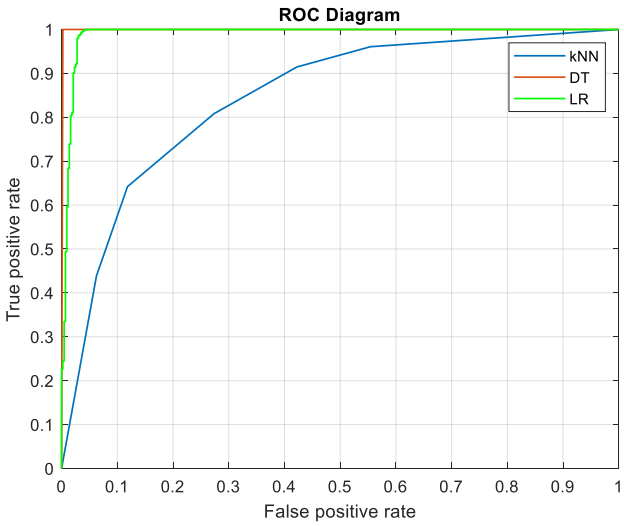


Figure 65: ROC diagram of the prediction result for k-nearest neighbors (kNN), Decision Tree (DT) and Logistic regression (LR) methods.

The obtained simulations with various degrees of insulation wear reveal an interesting information about the added value of data-driven approaches for cable condition assessment. Particularly, this work demonstrates the ability of different classification algorithms to identify, on the basis of only ND and voltage variation, the LV network cable condition assessment. However, this presented work should not be directly extended for other practical applications or be generalized for two reasons. Firstly, the resistance of the insulation material is calculated (in section 3) within consideration of some LV cable electrical properties specific to each manufacturer. Secondly, Machine Learning techniques have been developed here for the degradation detection in operating domains where the causes of observed variations are difficult to interpret. Hence to avoid a direct median separation in the observations, the input database has been built (in section 4.3.1) by excluding the cases of extreme degradations scenarios (severe faults) because they are easily detected without any advanced techniques.

4.6.Conclusion

ML based classification algorithms implemented in this research project are presented. Then, in the second part of the chapter, a Machine Learning-based framework to generate and organize data. A limited number of classification algorithms is then applied to identify a low voltage cable degradation due to the insulation material. From the probabilistic scenarios generation (for the uncertain nature and degree of the cable insulation degradation) to the probabilistic load flow calculations ; the comparisons between the implemented classifiers show that logistic regression and decision tree approaches are powerful binary classification tools with 97.917% and 99.884% accuracy performance, respectively, while the k-nearest neighbors method could not provide accurate predictions.

Despite the existing literature relating to fault detection in electrical networks, the main contribution of the presented study lies in its proposed methodology. Indeed, the proposed work is a novel approach, which lies in the use of data from already largely deployed Smart Meters.

This chapter reveals then the added value of data-driven approach for LV cable condition assessment. The two main contributions are:

- The application of ML algorithms for monitoring of LV grids by using smart meter measurements and voltages calculated by a Power Flow algorithm

- A proposal for practical deployment and utilization of ML algorithms

This chapter paves the way to an effective and timely predictive maintenance of the LV distribution network avoiding expensive solutions for the Distribution System Operators (DSOs) as well as the customers. To extend the study, the next chapter will be oriented towards the classification of all cables in a complete LV network, on the basis of a generalized data-based early identification of electrical low voltage cable degradation due to insulation wear.

Chapter 5: Cable condition assessment and prediction strategies

Contents

5.1.Introduction

Electrical low voltage (LV) networks are the utilities of the electrical networks. Their voltage of 230V or 400V is used every day to power household appliances. This allows the distribution of electrical energy to households. Moreover, with the energy transition the LV grids are becoming more and more important beyond their public utility. The structure and voltage magnitude of LV grids are favoring to the connection of low power electrical equipment - such as electric vehicles (EV) and photovoltaic systems (PV). However, this does not guarantee their reliability of service.

In fact, each low LV feeder is connected to a relatively low amount of customers and there exists as many LV grids as needed to serve all the local customers. Due to this low amount of customers served per feeder, the impact of an electricity interruption is low. Therefore, interruptions on the low LV level have not received much attention. However, all those low impact interruptions add up to high maintenance costs. It is then worth to investigate how to reduce these costs.

Researches have focused on possibilities for visual diagnostics of LV grids based on physical degradation phenomena. However, diagnostics are not available yet on a large scale and visual inspection is impossible as LV components are generally located underground. Also with the ongoing energy revolution, historical network data are available on interruptions, assets and environmental factors.

This chapter has the objective to investigate the use of smart metering devices data in order to make an assessment of the condition of LV distribution feeders. The focus is made on the problem of classification of all the LV cables - at the same computing time - in a complete LV network. This process is made on the basis of a generalized data-based identification of electrical low voltage cable degradations due to insulation wear.

While the previous chapter was related to a binary classification problem (two classes), this chapter addresses a multi-classes classification model to solve. After presenting the studied electrical networks and the framework for data analysis, various ML techniques are applied. To explore performances of ML techniques, results are compared according to the severity of line degradations. An impact analysis of the production connection is also explored on obtained performances. Finally, general conclusions and recommendations are drawn.

5.2. Machine Learning techniques for LV Cable condition classification

5.2.1. Problem and task

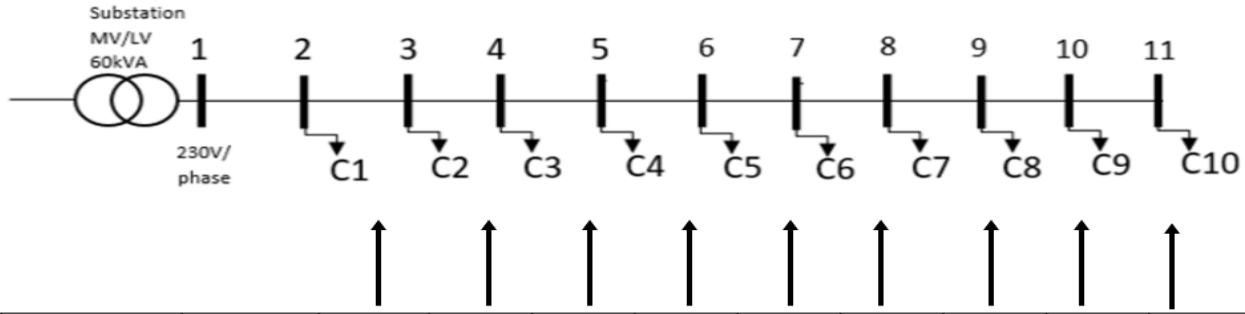
The objective of this work is to study the scalability of the developed Machine Learning techniques for an entire distribution network. The issue is to explore and develop a ML method as generic as possible what else the topology and size of Low Voltage distribution networks.

The study starts by considering a sending point with several customers ; ten customers in this application case. Then, the output vector is extended in order to be able to obtain for each cable, a binary output of the type “healthy/degraded”. The state of each cable, which connects two clients in the feeder, is considered in this chapter as a binary variable to be identified. Therefore, we will end up with as many output variables as there are sections of cable in the network.

The main difficulty is that each client node variations has a significant impact on the other nodes (in the same network). In consequence, a damping of voltage anomalies is usually observed in all the nodes that are located before the node where the anomaly occurs. The idea is to see if those impacts affect the decision accuracy and the calculation speed and how. This is a main issue to a possible extension toward an online application to monitor in real time and/or trigger protections. It would be then interesting to compare the studied classifiers in chapter 4 with as inputs the measurements available only from the meters (the powers consumed, injected, produced).

5.2.2. Electrical network overview and new resulting labels

The Figure 66 shows the generic ten nodes distribution network with the associated lengths (between each node), that are considered from figure 11. 6kVA customers are connected at each node. This network will serve as a basic study case to evaluate and compare the constructed classifiers and the computation results.



Lines classes	H	A	B	C	D	E	F	G	I	J
Lines length (m)		46	273	62	63	194	26	11	25	21

Figure 66: Generic topology of the considered 10-nodes LV distribution network.

For the evaluation of each cable state in a feeder, two classes are defined and are applied to each observation in the database (see Table 16). For the evaluation of a network (as represented in Figure 66), as many classes as lines in the feeder are defined to consider all cable states. Each label from A to J is associated to the presence of a degradation state in the associated line (table 24). The class H is associated with a complete healthy feeder; no insulation wear in none of the cables.

Table 24: Label table for the observations.

Cable state	**Labels**
No insulation wear	H
Insulation wear between node 2 and 3	A
Insulation wear between node 3 and 4	B
Insulation wear between node 4 and 5	C
Insulation wear between node 5 and 6	D
Insulation wear between node 6 and 7	E
Insulation wear between node 7 and 8	F
Insulation wear between node 8 and 9	G
Insulation wear between node 9 and 10	I
Insulation wear between node 10 and 11	J

5.2.3. Method of analysis

The method of chapter four is extended for a network. It consists in generating a data base of measurements representing the operation modes of the electrical network, then, these data are learned to parametrize each considered machine learning technique. Finally, ML techniques are applied to the unknown database and obtained classifications are studied.

The working database (so-called "knowledge" database) in this chapter, is created with the voltage values and Net Demand values at each node of the considered LV distribution network. The Net Demand (ND) data are obtained with SM measurements at each quarter of an hour (q), the PV production PV_i and the Load demand $Load_i$ at each node i, as expressed in equation (4.1). As stated previously, the voltage values at each node is depending of the state of lines's health. So, a database must be created to handle various damage conditions in the network in order to test the efficiency of various ML techniques for their detection. To generate this new database, each damage condition for each cable line, is considered separately by a radius variation. Thickness scenarios are used to represent these radius variations (according environmental conditions and among seasons). Each scenario is associated to a random sampling of cdf (as presented in chapter 3, section 3.4). Then, the variable values of the cable impedances are calculated. This impedance uncertainty and variability are propagated onto node AC voltages that are calculated by using the Newton-Raphson load flow (NRLF) technique (as explained in section 3.4.3) and measured data (each quarter q of each day) as presented in Figure 53.

Supervised Machine Learning algorithms consist of two stages namely the learning and the testing stage with data.

During the training, input data with the correct output classes are required. A learning algorithm uses this set to adapt and set parameters of the applied ML technique, and to create the right output (predicted states). A part of input data with the corresponding known output data (class, labels) from the knowledge database are stored in a subset, which is called training subset. For our presented results, the training dataset comprises 8064 observations.

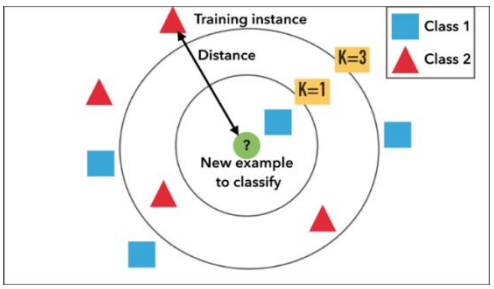
The quality of the ML based classifier is assessed by its ability to generalize, i.e. its ability to correctly classify new examples that have not been used for learning. Then, the remaining observations from the original knowledge database are stored in another subset: the testing subset. For our presented results, the testing dataset comprises 3456 observations. During the testing stage, a prediction is made with those new input data, the prediction error is calculated with all of these samples and then the accuracy of the parametrized ML algorithm can be assessed with these non-learned data (Figure 55).

5.2.4. Summary of the considered Machine Learning methods

For the comparative study, all the ML algorithms presented in section 4.2 will be experimented with the application network. To do so, the below section is a brief summary of each algorithm specification.

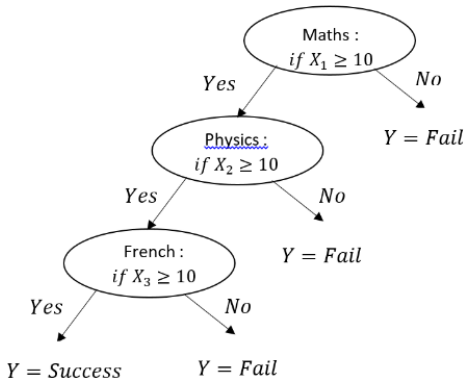
- **k-nearest neighbors (kNN)**

Each new observation is compared to existed ones by using a distance calculation and a number of nearest neighbors. This method is mostly used for fault detection and classification but also for power quality classification.



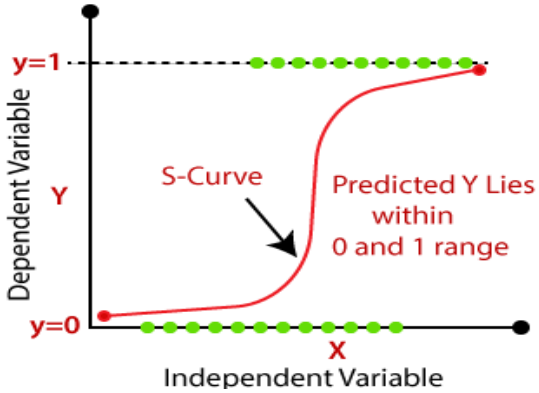
- **Decision Tree (DT)**

Recursive process, going from the properties (as seen in the branches) to the conclusions about an observation (as seen in the leaves). This method is used for Preventive and corrective control, for power systems security assessment.



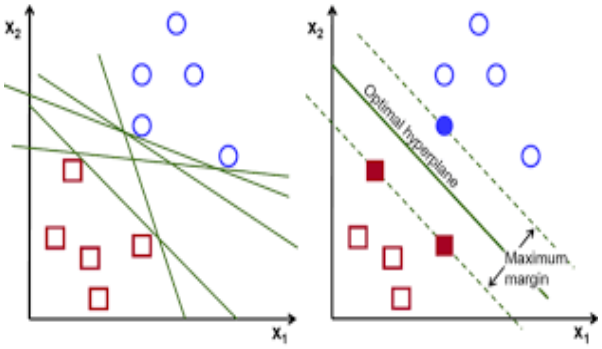
- **Logistic regression (LR)**

LR predict the probability $h_\theta(x)$ of a new observation x to be classified into a given class y. This method is used for electricity monitoring, visualization and prediction, fault detection in renewable energy production.



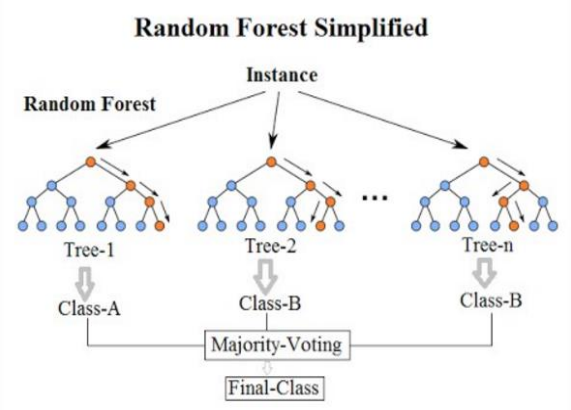
- **Multiclass support vector machine (SVM)**

SVM is a full, trained, multiclass model using binary one-versus-all support vector machine (SVM) method. Robust prediction methods being based on statistical learning frameworks. This method is mostly used for smart grid condition monitoring, identification of miscellaneous electric loads, micro-grids protection.



- **Random Forest (RF)**

Very powerful ensembling machine learning algorithm which works by creating multiple decision trees and then combining the output generated by each of the decision trees. This method is used for smart pricing, classification of electrical grid stability, prediction of power consumption.



5.3. Performance analysis of ML techniques according the line degradation severity

5.3.1. Followed exploring method with only consumers

A sending point with several consumers is considered and so different power demands according the time. These observations are obtained by sampling the CDF of the load demand (as modelled from measurements in part 4.5.2.3). As example, the figure below presents the considered load demand at node 6 and 11 during one day.

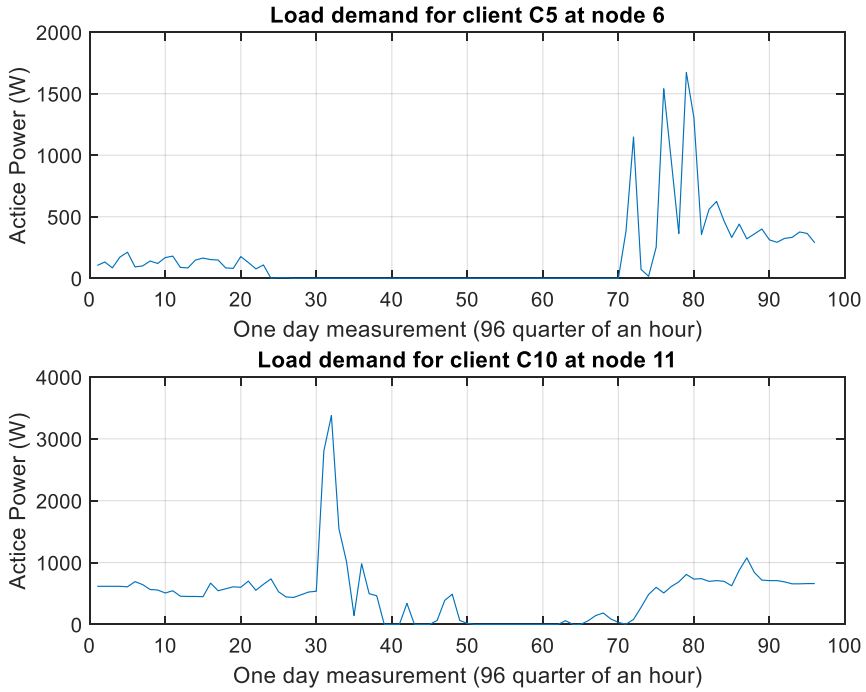
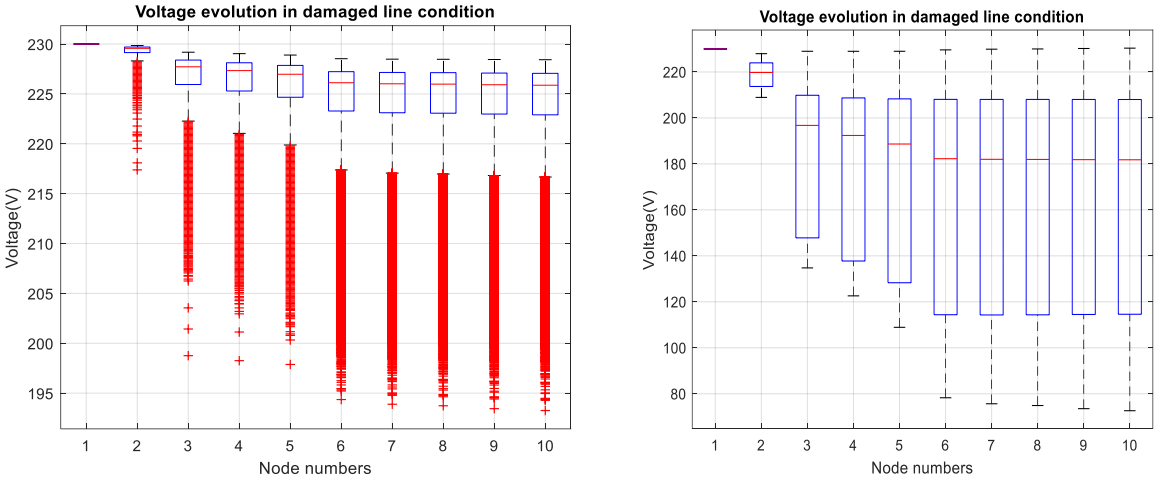


Figure 67: Load demand a) at node 6 and b) at node 11.

Under various degradations in various lines, the purpose is to implement and compare different ML techniques to identify the state of each cable. It should be noted that the extreme degradation scenarios as studied in [36] have not been considered because the computations are associated to light degradation conditions. Two situations will be studied. First, a light degradation of lines is implemented by considering a high domain variation of the random sampling [0.7 ; 1] onto the CDF of the line resistance (Figure 28) to create light variations of the impedance. Then, medium degradations will be considered by another sampling domain [0.7 ; 0.4].

5.3.2. Comparative study of results under light and medium degradations

The presented results show the boxplots of nodal voltages obtained by the Load Flow calculations with consumers. Figure 68(a) is related to the feeder under light degradations while Figure 68(a) is related to the feeder under medium degradations. As expected, the increase in the insulation conductance (1/*Riso*) leads to voltage drops. For each figure, the blue box includes the 75th to the 25th percentiles of the voltage profile. The red positive signs highlight the outliers of the voltages in the created scenarios (outside the percentile box). Under light degradations, average values of the different node voltages are limited to the [226 V, 230 V] range where as under medium degradations, the variation domain is increasing to [182 V, 230 V] out of the standard minimum magnitude.



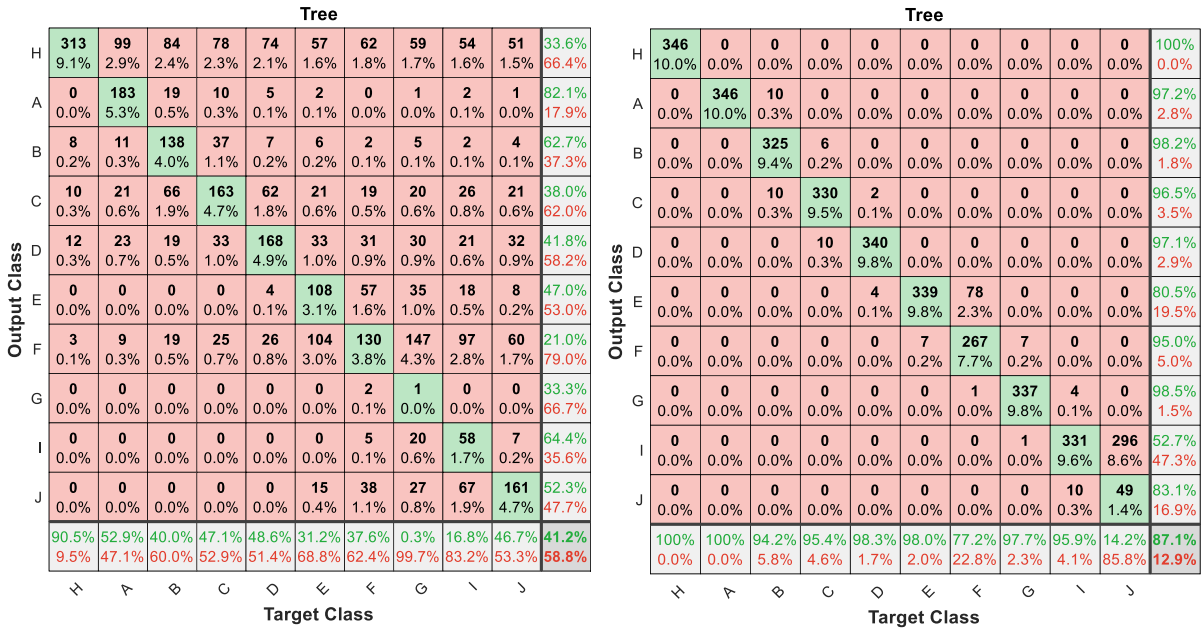
(a) (b)

Figure 68: Boxplots of nodal voltages under: (a) light degradations; (b) medium degradations.

To compare performances of the different machine learning technologies, kNN, DT, SVM and LR methods, all obtained 2-D confusion matrixes are given in the following figures.

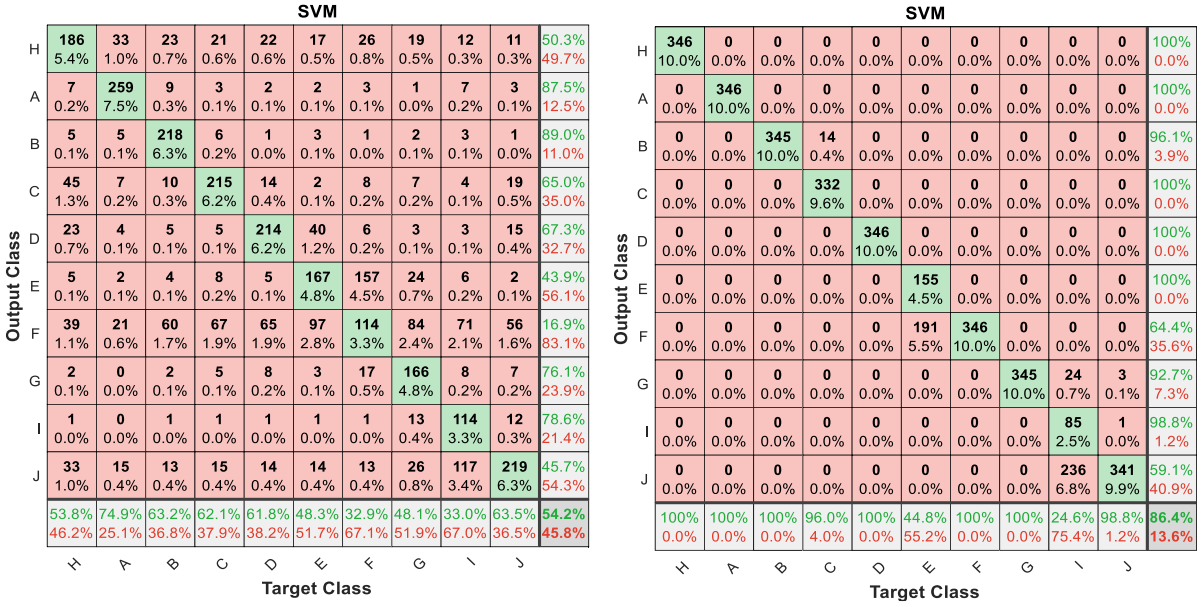
(a) (b)

Figure 69: kNN classification results under: (a) light degradations; (b) medium degradations.



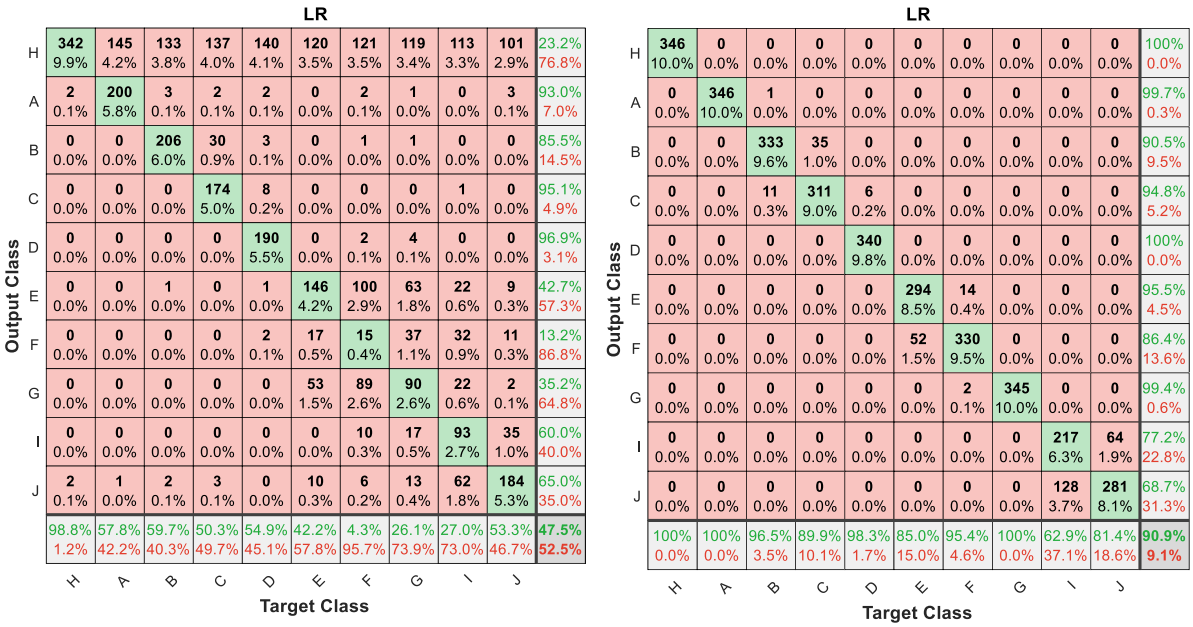
(a) (b)

Figure 70: DT classification results under: (a) light degradations; (b) medium degradations.



(a) (b)

Figure 71: SVM classification results under: (a) light degradations; (b) medium degradations.



(a) (b)

Figure 72: LR classification results under: (a) light degradations; (b) medium degradations.

The testing dataset handles 3456 observations comprising only 346 observations in a healthy state. So, a first challenge of data mining is to detect this healthy state (10% of the observations). A second one will be to identify and locate the degraded line (9 possible states).

Comment 1: For detecting a H state (first colon), LR, Tree and kNN have good performances (>90% as shown in first column of last line for the case H) as shown on figure 69(a), 70(a) and 72(a) under light degradations. When the degradation are stronger, all ML techniques detects very well this state (100%).

Comment 2: For identifying the location of the degraded electrical line, it is difficult to identify a better ML technique. As example, under light degradations, the SVM technique (figure 71) has the best performances with 54,2% (last colon). But when medium degradations are considered, performances are decreased because of bad results regarding states E (44,8%) and I (24.6%). Under medium degradations, the found best technique is LR with a global 90,9% performance.

Comment 3: What else the ML technique, some degraded lines are harder to detect and decrease the obtained general performance. Hence, it seems that more observations should be provided regarding these conditions during the training phase. We recall that equal number of observations for each state have been used in the learning phase for presented results here. In fact, performances of each ML technique could be improved by enriching the learning data base with more data corresponding to operation conditions that are difficult to detect after this first test. This enhancement has not been applied in our work because it is scientifically questionable to compare ML techniques using different learning databases for each one. Anyway this specialization learning can easily been applied for an engineering application.

It can be conclude that medium degradation are easier to detect by ML techniques. Light degradation conditions are difficult to assess due to the low gap between the voltage profile while they are remaining no critical for the operation of the electrical network.

5.4.Impact of PV generation onto performances of ML techniques

5.4.1. Followed exploring method with a PV production

As previously, a departure with several consumers is considered and so different power demands according the time. But now, we are going to consider a PV production at a node whose time profile is obtained by sampling the cdf of the PV production (as modelled from measurements in part 4.5.2.3). As consumption is also present, the Net Load Demand is calculated (as explained in part 2.3). As example, the figure below presents the obtained Net Load Demand at node 6 and 11 during one day.

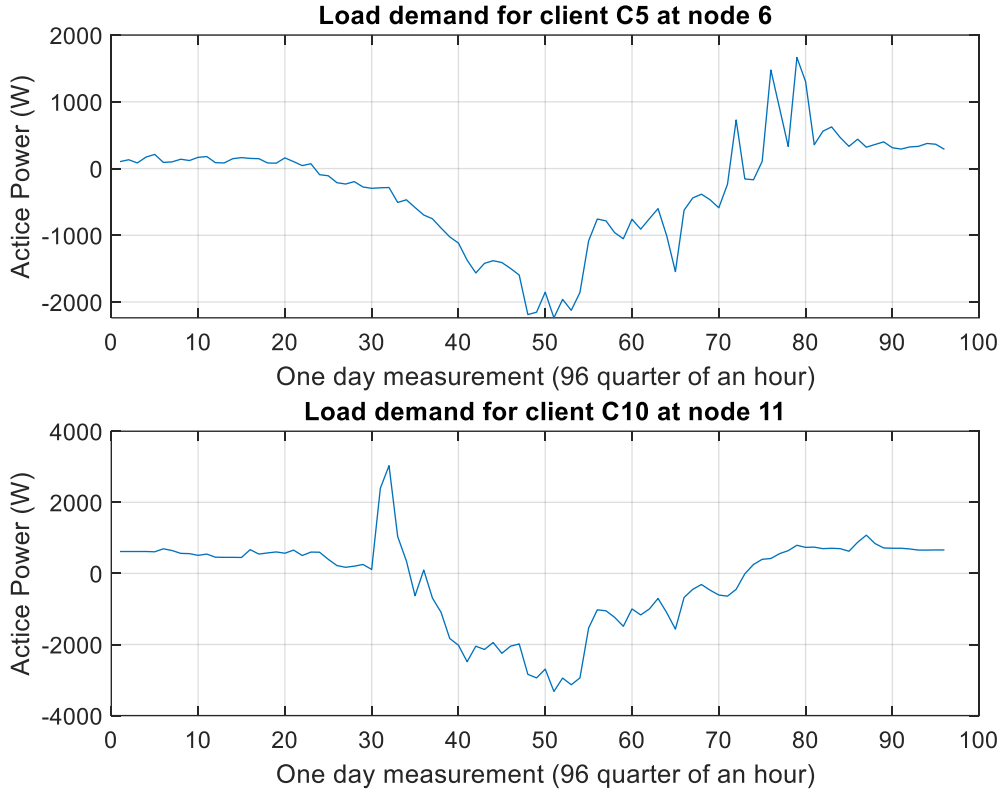
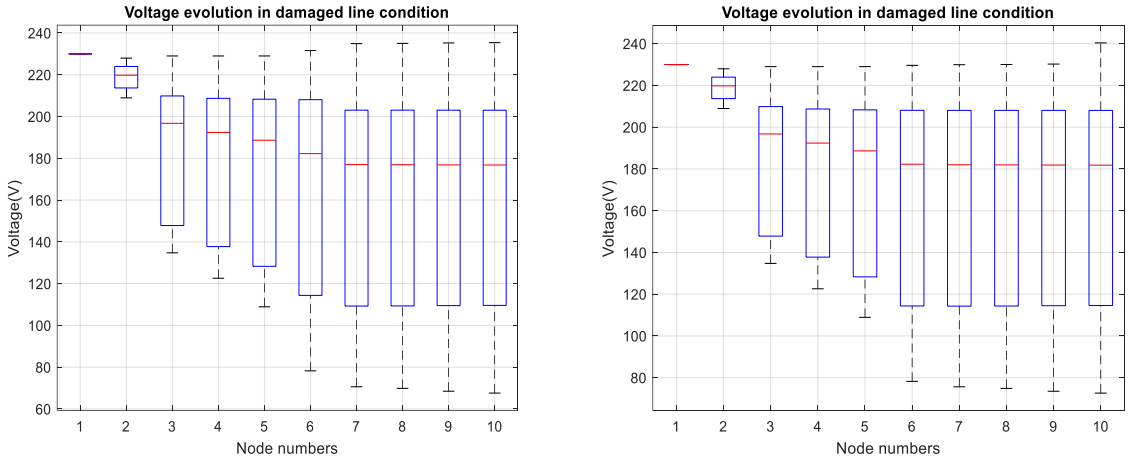


Figure 73: Load demand at a) node 6 and (b) node 11.

The purpose of this study is firstly to highlight the impact of a PV generation onto ML performances. The presented results are related to medium degradations of lines. Secondly, the impact of the location of this PV production into the network is explored.

5.4.2. Comparative study of results for PV connection at the terminal node and node 6

The boxplots of nodal voltages obtained by the Load Flow calculations are presented on. Figure 74(a) with PV penetration on the node 6 while Figure 74(b) is related to the feeder with PV penetration at the last node, named node 11. For each figure, the blue box includes the 75th to the 25th percentiles of the voltage profile (see Appendix D for the comparative study of the performances without and with PV penetration under medium degradation).

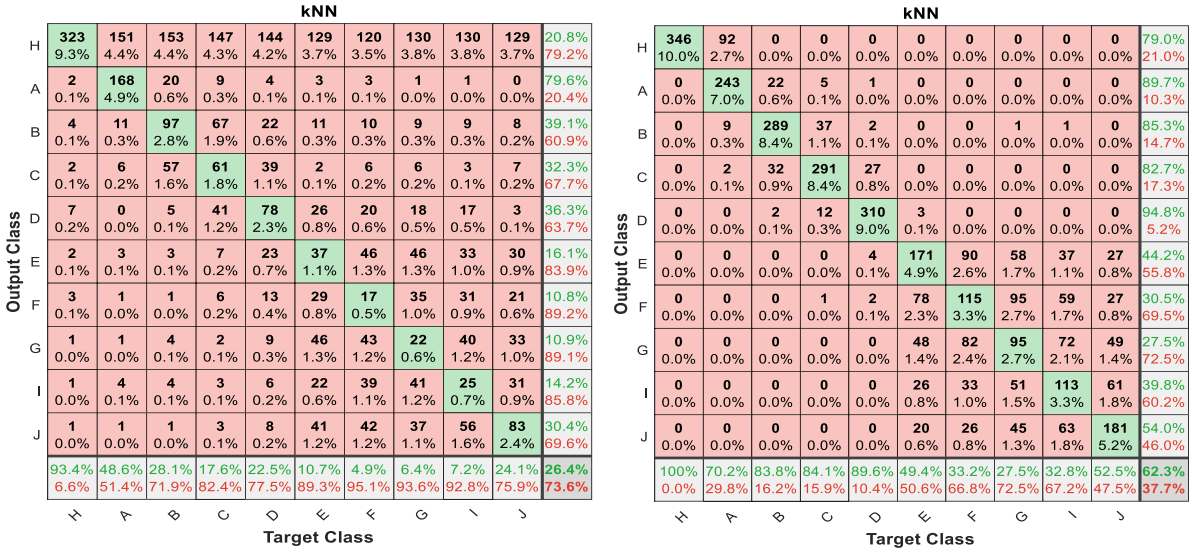


(a) (b)

Figure 74: Boxplots of nodal voltages with a PV production at: (a) the node 6 ; (b) the last node named node 11.

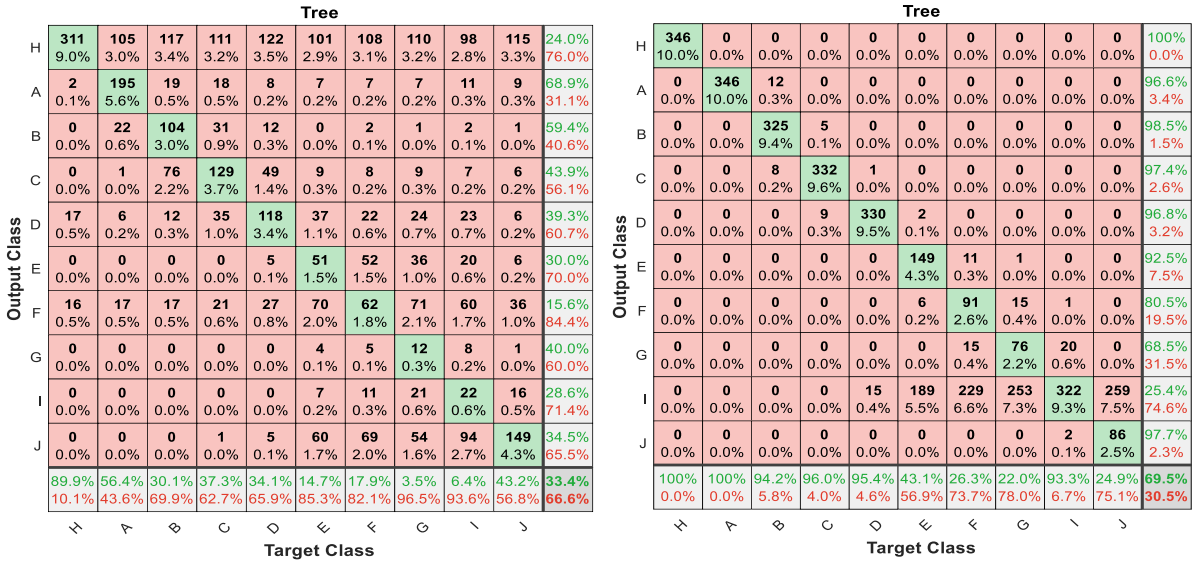
For either consumers or prosumers, the red positive signs highlight the outliers of the voltages in the created scenarios (outside the percentile box). Average values of the different node voltages (red bars) are in the domain limited to the [178 V, 230 V] range with a PV production connected at the last node; whereas when connected at the last node, the variation domain is increasing to [182 V, 230 V]. We retrieved a well-known result as the connection of a production at the end of a network improve the RMS voltage over the all network.

In order to compare performances of the different machine learning technologies, kNN, DT, SVM and LR methods, all obtained 2-D confusion matrixes are given in the following figures.



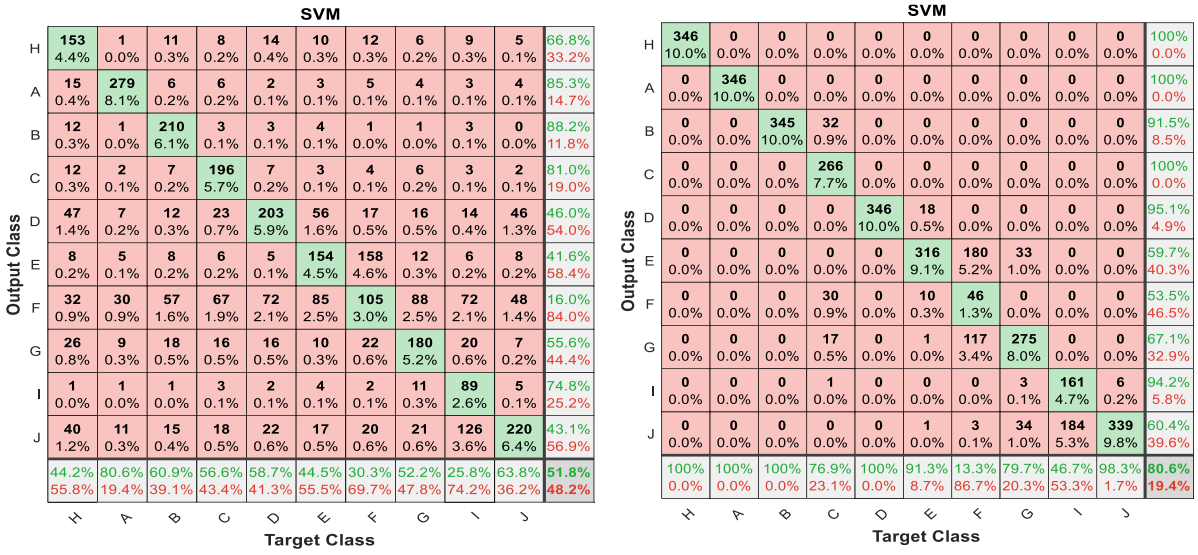
(a) (b)

Figure 75: kNN classification results with a PV production at: (a) the node 6 ; (b) the last node named node 11.



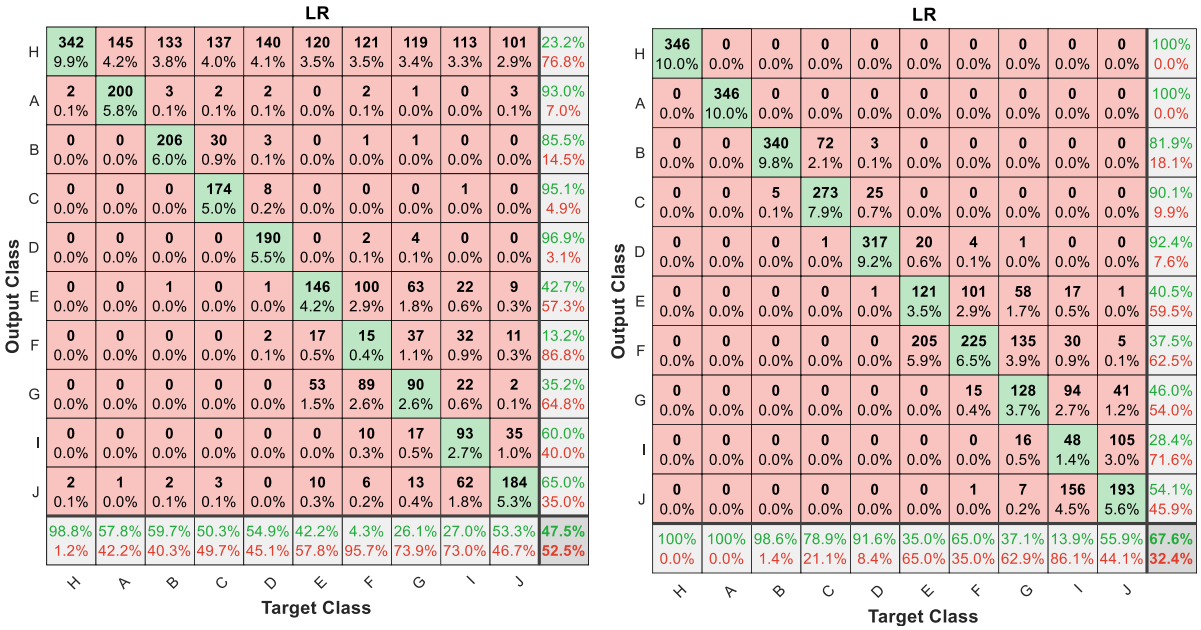
(a) (b)

Figure 76: DT classification results with a PV production at: (a) the node 6 ; (b) the last node named node 11.



(a) (b)

Figure 77: SVM classification results with a PV production at: (a) the node 6 ; (b) the last node named node 11.



(a) (b)

Figure 78: LR classification results with a PV production at: (a) the node 6 ; (b) the last node named node 11.

Comment 1: A first observed tendency by comparing the same situation but without production (figure 69(b), 70(b), 71(b), 72(b)) is that when PV generation is connected, the performances of ML techniques are less. It can be explained because the electricity production has a tendency to increase the voltage and so to hide the voltage droop due of the degraded line (increase of the impedance).

Comment 2: Performances are better for a connection location at the end of the electrical network

Comment 3: Excepted for the SVM, again the healthy state is easily detected and classed (first colon)

Comment 4: For the detection of damaged lines, the SVM technique has the best results. But, it must be noted that the state F is badly detected (13,3%). Again, it seems that more observation should be provided regarded these conditions during the training phase (we recall that equal number of observations for each state have been used in the learning phase for presented results here).

5.5. Discussion and Conclusion

The chapter 5 of this manuscript focuses on how to develop a Machine Learning technique as generic as possible in order to make it applicable to a full Low Voltage distribution feeder. We therefore start by considering a feeder with several customers (a ten nodes network with different nodal length). Then, the output space is extend to englobe as many label as existing cables (line between two nodes). So where as from the classification process of chapter 4 was binary (two classes), now this chapter five addresses a multi-classes classification model to solve. Classification algorithms implemented in this research project are presented. Then, in the

second part of the chapter, a Machine Learning-based framework to generate and organize data. A limited number of classification algorithms is then applied to identify a low voltage cable degradation due to the insulation material.

From the probabilistic scenarios generation (for the uncertain nature and degree of the cable insulation degradation) to the probabilistic load flow calculations ; the comparisons between the implemented classifiers show that the output accuracy can increase form 15.6% to 67.6% ; depending on the degree of line's degradation or the PV impacts.

The obtained results shown that the feeder with only consumers give the better classification accuracy. This can be explained by the structure of the knowledge database of this application, which makes easier the separation between the observations associated with the different classes. More the PV plants are far from the first node, more the nodal voltage profile is variable ; so the PV impact can clearly be analysed.

From previous comments about obtained performances in detection, we extract some conclusions.

- Tree and LR techniques are recommended to detect an H state (with or without PV production).

- The k-nearest neighbours method (k=5) still remaining the technique providing less accurate predictions.

- As when PV generation is connected, the performances of ML techniques are less, a recommendation would be then to add more input data to the ML techniques as the load demand with the Net Demand at prosumer nodes.

- To improve the learning for badly classed states, it could be interesting to consider more observations coming from degraded lines that are related to these states during a second learning stage in order to improve performances of classifiers thanks to more supplied knowledge from these difficult states.

- The final recommendation would be to use at least two ML techniques: one will be specialized (learned with adapted data) for detecting the H state as Tree or LR and, in case of degraded lines, the SVM technique to identify the degraded line.

The above approach shows the added value of data-driven with ML techniques for LV feeder condition assessment as well as for PV penetration impacts on existing LV feeder.

Recommendations and General Conclusion

1. Concluding comments

In this research project, statistical data from SM and the daily power flow measurement have been used for investigating adaptive models of the LV network under variations of temperature and line degradations.

Firstly, a probabilistic Load Flow algorithm has been developed for radial LV network within consideration of the system resistance distribution as an uncertain parameter depending on the temperature. The Load demand and the PV production generally used in classical Load Flow (LF) calculations are computed by using Smart Meter data with a quarter of an hour resolution time. To take into account uncertainties into the network operation, either the resistance value or the network to customer exchanged power are randomly selected at each iteration by using Monte Carlo (MC) methods. Both annual and seasonal dependencies of the line resistance have been implemented. The simulation results have shown that integrating the resistance distribution in a seasonal probabilistic tool can affect the collected reliability indices up to 10.4% depending on the season.

Secondly, the conductance variations of the LV cable insulation, due to the insulating material degradation, have been analysed within its impact on the nodal voltages. To this end, a probabilistic framework has been proposed based on the MC simulations and LF calculations. The MC simulations are used here to characterize the insulation degradation through scenarios and to model the resistance variation. LF calculations are conducted in order to calculate the nodal voltages. Simulation results reveal some voltage drops over 50% depending on the location and level of the cable insulation degradation as well as on the network working conditions.

Finally, a Machine Learning-based framework is proposed for the identification of cable degradation due to the insulation material wear. To this end, a probabilistic tool is firstly developed to generate scenarios including the uncertain nature and level of the cable insulation degradation. Those scenarios are then associated with the load demand and PV generation variations and used to build the nodal voltage database by performing probabilistic load flow calculations. Different supervised learning methods are finally applied to the generated database. In the first (training) stage, the studied ML based classification methods learn from the given inputs, its associated cable condition status in order to be able to predict it, in the second (test) phase with (new) given network operating points. The comparisons between the implemented classifiers show that Logistic Regression and Decision Tree approaches are powerful binary classification tools with respectively a 97.92% and 99.88% accuracy performance while k-nearest neighbors method could not provide accurate predictions. The conducted study reveals the benefit of such a data-driven approach for the cable condition assessment.

2. Overview of performed research works

This PhD manuscript has explored scientific approaches for the modeling of low voltage electrical networks with an interest for identifying their state of physical degradation. The main objective was to address the upgrading of LV network modelling and the cable condition assessment through the exploitation of data provided by Smart Metering devices (SM) installed at homes. The large-scale deployment of SM helps to increase the observability of the electrical

network state of health. This large database is, for electrical engineering, an opportunity to create new adaptive modelling techniques/tools for those initially poorly metered networks. The designed and implemented algorithms enable a various field of applications ranging from temperature based modelling and the impact analysis of line impedance variations. In this work, some frameworks have been set up by applying Monte Carlo (MC) simulations and ML techniques onto energy data at end-users scale. Thereby, the main contributions are:

- In chapter 2, the modelling of an unknown electrical network architecture with SM data, and the impact analysis of temperature variations onto voltage variations and so operation security,
- In chapter 3, the characterization and impact analysis of the cable insulation degradations on LV network voltages,
- In chapter 4, the application and comparison of various Machine Learning (ML) techniques for the condition assessment of a single LV cable with smart meter data,
- In chapter 5, the extension of above techniques for the condition assessment of a LV electrical network with an exploring study about the PV production impact onto the accuracy of degradation detection.

For low voltage distribution networks, issues of these works can help the maintaining, drastically enlarge the hosting capacity and enable the cost-effective planning of reinforcement upgrades by identifying weak parts of the network.

3. Guidelines for the LV distribution system operators

This section presents an open letter to the DSO to guide them in the applied development of the presented above prospective ML techniques.

Need to measure the RMS voltage variations at transformer and SM

The various studies investigated in this project have highlighted the need to measure voltage variations along the feeder. Communication of voltage measurements was not mandatory in installed smart meters. As presented in this manuscript, voltages can be estimated by executing a Power Flow algorithm with power measurements. However, voltage measurements will make possible the reduction of calculation errors. Moreover, the accuracy of voltage measurements is essential, as it is an essential input data of ML techniques for degradation detection and location.

Integration of the insulation resistance in network models for a load flow calculation

The chapter 3 of this document has focused on how to characterize the cable insulation degradations as a variable resistance named R_{iso}. For upgrading the model of the LV distribution network, the above resistance model can be incorporated in the model of the line. Therefore, this variable resistance (more precisely the combination of R_{iso} and a ground resistance R_g) should be considered as a shunt variable resistance, between the leakage point and the ground. The Load Flow calculation could then be more realistic and closed to the real cable conditions of the LV feeder. The modelling of the insulation degradation makes the preventive maintenance

more effective and more accurate to prevent sudden breakdowns on the network and the enormous associated financial losses.

Use of thermal models based power flow calculation for prognostic and on line monitoring

Climate change is expected to raise the global temperature by 2.6°C, on average and in a rather optimistic scenario. As detailed in chapter 2, the thermal exchange with the environment and hence by weather conditions is influencing the temperature of the electrical line. Physical parameters should not be considered as static and an extension of their variation range should be considered. Dynamic thermal modelling of the power system infrastructure, SM data and power flow calculation can be exploited for a prognostic of operating conditions and for controlling the varying capability of connected distributed energy resources (PV curtailment, load shedding, …) in order to reduce power transfer if necessary.

4. Limitations and future works

The final interest of these works is to set up a tool, which can assist the Distribution System Operators (DSOs) in an effective and timely predictive maintenance of the LV distribution network avoiding the costly solutions. Indeed, the obtained result offer promising perspectives for the early detection of cable degradations by combining ML approaches, Load demands profiles and Smart Meter (SM) measurements.

The current study is an exploratory study of SM database and a first step towards a global and generalized data-based early detection of electrical Low Voltage cable degradations due to the insulation wear, by using Machine Learning tools.

The implemented frameworks, during this project, aim to serve as tools for a future monitoring, diagnostic and prognostic of Low Voltage networks with the advantage of better managing the latter (e.g. by detecting failures before they become costly, by improving probabilistic load flows used for techno-economic analysis, by increasing the robustness of the voltage control algorithms…). The targeted applications is hoped to significantly improve the reliability, safety and availability of LV distribution systems.

The obtained results offer promising perspectives for the early detection of cable degradations by combining ML approaches, Load demands profiles and Smart Meter (SM) measurements. However, the work done in this manuscript has some limitations like any research works. Beyond the main contributions of this thesis, some relevant further developments could be explored for more and for better performance and generalization of the implemented tools. The following lines present some perspectives to this work.

The first one is about the condition assessment for any LV feeder. Indeed, this research could be extended to large LV feeders with multiple ramification based on a Cross Nodal Learning that will enable the learning between the models of each line section or cables in the network. Moreover, an Ensemble learning process using multiple learning algorithms to get better predictions can be set up for this global application; as example bagging, boosting and random forests method belong to the same learning method family.

The second one is about the effectiveness and relevant of the SM data by using voltage measurements rather than estimations from power flow calculations. The database can be

complete with **transformer voltage and nodal voltage variation measurement** in order to improve their availability, quality and accuracy. The above results have shown that quality data lead to a better knowledge database and moreover to a better degradation detection.

The third one can be the addition of new inputs as the type of cable, the PV production, … It is a research track to improve the accuracy of ML classifiers by providing adequate and complementary knowledge through more pertinent data inputs.

The fourth one can be about overcoming the limits of Machine Learning applications on detection of cable degradation. The conclusion of chapter 5 shows that in case of light degradations, it is difficult to distinguish between the class H and other classes. So, some new complementary inputs features can be added to the knowledge database.

References

References

[1] European Distribution System Operators (E.DSO). Why smart grids? Available online: https://www.edsoforsmartgrids.eu/home/why-smart-grids/ (accessed on 8 April 2021).

[2] European Environment Agency, "Trends and Projections in Europe 2021", Report No 13/2021, ref : TH-AL-21-012-EN-N, 2021, ISBN 978-92-9480-392-4, doi: 10.2800/80374.

[3] Christophe Magdelaine (2019). Doit-on avoir peur des nouveaux compteurs communicants Linky? https://www.notre-planete.info/actualites/3703-danger-compteur-communicant-Linky (accessed on October 2019).

[4] DIRECTIVE 2009/72/CE DU PARLEMENT EUROPÉEN ET DU CONSEIL du 13 juillet 2009. *Journal officiel de l'Union européenne*. L211/55 – L211/93. https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:211:0055:0093:fr:PDF (accessed on October 2018).

[5] LOI n°2015-992 du 17 août 2015 relative à la transition énergétique pour la croissance verte https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000031044385&categorieLien=id (accessed on October 2018).

[6] Nanpeng Yu. Big Data Analytics in Electric Power Distribution Systems. *The International Academy Research and Industry Association (IARIA)*, 2017.

[7] Enedis. Interflex. Les démonstrateurs smart grids https://interflex-h2020.com/

[8] SESAM-GRIDS. http://www.sesam-grids.org/Pages/Description.aspx (accessed on October 2018).

[9] Thierry Van Cutsem. The electrical system: the very fundamentals. Le blog de l'Université de Liège. 2015.

[10] Heloïse Dutrieux. Méthodes pour la planification pluriannuelle des réseaux de distribution. Application à l'analyse technico-économique des solutions d'intégration des énergies renouvelables intermittentes. *Ecole Centrale de Lille*. Français. NNT:2015ECLI0021, tel-01491053. 2015.

[11] Clémentine Benoit. Models for investigation of flexibility benefits in unbalanced low voltage smart grids. *Electric power*. *Université Grenoble Alpes*. English. NNT:2015GREAT056. tel-01223369. 2015.

[12] Procopiou, A. Active Management of PV-Rich Low Voltage Networks. Degree of Doctor of Philosophy in the Faculty of Science and Engineering – University of Manchester, United Kingdom, 2017.

[13] ERDF, « Description physique du réseau public », DTR d'ERDF, référence ERDF-NOIRES_07E, version 2, avril 2008.

[14] Groupe EDF. L'acheminement de l'électricité - Le réseau de distribution. Available online: https://www.edf.fr/groupe-edf/espaces-dedies/l-energie-de-a-a-z/tout-sur-l-energie/l-acheminement-de-l-electricite/le-reseau-de-distribution (accessed on August 2021).

[15] AFNOR, « NF EN 50160 (C 02-160) Caractéristiques de la tension fournie par les réseaux publics de distribution », Mai 2020, http://dfv-technologie.com/doc_technique_dfv/normes_afnor/EN50160.pdf.

[16] AFNOR, « NFC 13-100 Postes de livraison alimentés par un réseau public de distribution HTA (jusqu'à 33 kV) », April 2015.

[17] AFNOR, « NFC 14-100 Installations de branchement à basse tension », Febuary 2008.

[18] AFNOR, « NFC 15-100 Installations électriques à basse tension », December 2002.

[19] « Fiches et guides SéQuélec | Enedis ». https://www.enedis.fr/fiches-et-guides-sequelec, May 10, 2020.

[20] Promotelec, « Raccordement d'une installation électrique », oct. 1997.

[21] Mireille CAMPANA, François DEMARCQ, Didier PILLET, "Flexibilite du systeme electrique : contribution du pilotage de la demande des batiments et des vehicules electriques", 2020, ref : N° 2019/01/CGE/SG/ICM.

[22] Enedis, Sorties de télé-information client des appareils de comptage Linky utilisés en généralisation par Enedis, Ref: Enedis-NOI-CPT_54E, Vers. 3, 2018.

[23] Victor Gouin. Évaluation de l'impact du Smart Grid sur les pratiques de planification en cas d'insertion de production décentralisée et de charges flexibles. Energie électrique. Université Grenoble Alpes, 2015.

[24] International Renewable Energy Agency (IRENA). Renewable capacity highlights. renewable power generation capacity statistics for the past decade (2011-2020). *31 March 2021*. Available online : https://www.irena.org/publications/2021/March/Renewable-Capacity-Statistics-2021 (accessed on September 2021).

[25] European commission. State of the Energy Union 2021 – Contributing to the European Green Deal and the Union's recovery. *26 October 2021*. Available online : https://ec.europa.eu/energy/sites/default/files/state_of_the_energy_union_report_2021.pdf (accessed on October 2021).

[26] EMBER. Landmark moment as EU renewables overtake fossil fuels. Available online : https://ember-climate.org/project/eu-power-sector-2020/ (accessed on October 2021).

[27] Arrêté du 13 avril 2021 pris en application du décret n° 2020-1561 du 10 décembre 2020 relatif aux aides pour l'électrification rurale. *JORF n°0098 du 25 avril 2021*. Available online: https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000043416408 (accessed on September 2021).

[28] Economie.gouv.fr. Aides et crédits d'impôts. Installation de panneaux solaires : vous avez droit à des aides. Available online: https://www.economie.gouv.fr/particuliers/aides-installation-photovoltaiques# (accessed on September 2021).

[29] Bilan électrique ENEDIS 2020, https://www.enedis.fr/media/2691/download, May 2021

[30] Gérédis, Principes d'étude et de développement du réseau pour le raccordement des clients consommateurs et producteurs BT, https://www.geredis.fr/IMG/pdf/d-r1-rta-14-d_principes_d_etude_et_de_developpement_du_reseau_pour_le_raccordement_des_clients_consommateurs_et_producteurs_bt, 2020.

[31] Intergovernmental Panel on Climate Change, « Summary for Policymakers », http://www.ipcc.ch/report/sr15/, 2018.

[32] Intergovernmental Panel on Climate Change, « Summary for Policymakers », https://www.ipcc.ch/site/, 2019

[33] Norme française NF C 15-100 (Décembre 2002). Installations électriques à basse tension. Available online: http://alternatif33.free.fr/NFC15100_2002.pdf (accessed on 23 September 2020).

[34] Codjo, E. L.; Vallée, F.; Francois, B. Impact of the line resistance statistical distribution on a probabilistic load flow computation. In Proceedings of the *2020 6th IEEE International Energy Conference (ENERGYCon)*, Gammarth, Tunisia, September 28 - 1 October 2020, 637-642.

[35] Bakhshideh Zad, B.; Hasanvand, H.; Lobry, J.; Vallée, F. Optimal reactive power control of DGs for voltage regulation of MV distribution systems using sensitivity analysis method and PSO algorithm. *Int. J. Electr. Power Energy Syst.* **2015**, 68, 52-60.

[36] Bakhshideh Zad, B.; Lobry, J.; Vallée, F. A centralized approach for voltage control of MV distribution systems using DGs power control and a direct sensitivity analysis method. In Proceedings of 2016 IEEE International Energy Conference, Leuven, Belgium, 4-8 April 2016.

[37] Salivon, T. Vieillissement thermique d'isolants en PVC et PELX de câbles électriques en environnement automobile, Ecole nationale supérieure d'arts et métiers – ENSAM, France, 2017.

[38] Techni-Tool Company, The Complete Guide to Electrical Insulation Testing, Ensto Finland OyEnsio.

[39] Kang, S.-D.; Kim, J.-H. Investigation on the insulation resistance characteristics of Low Voltage cable. *Energies* **2020**, *13*(14), 3611.

[40] Kruizinga, B.; Wouters, P.A.A.F.; Steennis, E.F. Fault development on water ingress in damaged underground Low Voltage cables with plastic insulation. In Proceedings of the 2015 IEEE Electrical Insulation Conference (EIC), Seattle, WA, USA, 2015, pp. 309-312.

[41] Kruizinga, B.; Wouters, P.A.A.F.; Steennis, E.F. Comparison of polymeric insulation materials on failure development in low-voltage underground power cables. In Proceedings

of the 2016 IEEE Electrical Insulation Conference (EIC), Montreal, QC, Canada, 2016, pp. 444-447.

[42] Helmholt, K.A.; Groote Schaarsberg, M.; Broersma, T.; Morren, J.; Kruizinga, B.; Wouters, P.A.A.F.; Steennis, E.F.; Baldinger, F. A structured approach to increase situational awareness in low voltage distribution grids. In Proceedings of the 2015 IEEE Eindhoven PowerTech, Eindhoven, Netherlands, 2015, pp. 1-6.

[43] Csányi G.M., Tamus Z.Á., Varga Á. Impact of Distributed Generation on the Thermal Ageing of Low Voltage Distribution Cables. In Proceedings of the 8th Doctoral Conference on Computing, Electrical and Industrial Systems (DoCEIS), May 2017, Costa de Caparica, Portugal, pp.251-258.

[44] Csányi G.M., Tamus Z.Á., Kordás P. (2018) Effect of Enhancing Distribution Grid Resilience on Low Voltage Cable Ageing. In Technological Innovation for Resilient Systems, Proceedings of the Doctoral Conference on Computing, Electrical and Industrial Systems (DoCEIS), 29 March 2018, pp.300-307.

[45] Csányi, G.M.; Bal, S.; Tamus, Z.Á. Dielectric Measurement Based Deducted Quantities to Track Repetitive, Short-Term Thermal Aging of Polyvinyl Chloride (PVC) Cable Insulation. *Polymers* **2020**, 12, 2809.

[46] Helge SELJESETH & al. Benefits of voltage measurements with smart meters. *23rd International Conference on Electricity Distribution, CIRED*. Lyon, 15-18 June 2015.

[47] Abdel-Majeed Ahmad & Braun Martin. (2012). Low voltage system state estimation using smart meters. *Universities Power Engineering Conference (UPEC), 47th International*. 1-6. 10.1109/UPEC.2012.6398598. 2012.

[48] A. Nag & A. Yadav. Artificial Neural Network for Detection and Location of Faults in Mixed Underground Cable and Overhead Transmission Line. *The Sixth International Conference on Computational Intelligence and Information Technology – CIIT*, May 2016.

[49] A. CHOUAIRI & al. Analyse de l'endommagement et prédiction de la durée de vie des câbles électriques dans les réseaux de distribution aériens basse tension: Application aux conducteurs en cuivre et aux câbles en aluminium torsade. *22ème Congrès Français de Mécanique*. Lyon, 24 au 28 Août 2015.

[50] Alexander M. Prostejovsky & al. Distribution Line Parameter Estimation Under Consideration of Measurement Tolerances. IEEE transactions on Industrial Informatics, VOL. 12, NO.2, APRIL 2016.

[51] Codjo, E. L.; Bakhshideh Zad, B.; Vallée, F.; François, B. Analysis of Low-Voltage Network Sensitivity to Voltage Variations Due to the Insulation Wear. In Proceedings of the 55th International Universities Power Engineering Conference (UPEC), Turin, Italy (Virtual Event), 1-4 September 2020.

[52] Klonari, V.; Toubeau, J.-F.; De Grève, Z.; Durieux, O.; Lobry, J.; L.; Vallée, F. Probabilistic simulation framework for balances and unbalanced low voltage networks. *Int. J. Electr. Power Energy Syst* **2016**, 82, 439-451.

[53] Conti, S.; Raiti, S. Probabilistic load flow using Monte Carlo techniques for distribution networks with photovoltaic generators. *Solar Energy* **2007**, 81, 1473-1481.

[54] Klonari, V.; Bakhshideh Zad, B.; Lobry, J.; Vallée, F. Application of voltage sensitivity analysis in a probabilistic context for characterizing low voltage network operation. In Proceedings of the 2016 International Conference on Probabilistic methods Applied to Power Systems, Beijing, China, 2016.

[55] Bakhshideh Zad, B.; Lobry, J.; Vallée, F. Impacts of the model uncertainty on the voltage regulation problem of medium-voltage distribution systems. *IET Gener. Transm. Distrib.* **2018**, 12, 2359-2368.

[56] Bakhshideh Zad, B.; Toubeau, J.-F.; Lobry, J.; Vallée, F. Robust voltage control algorithm incorporating model uncertainty impacts. *IET Gener. Transm. Distrib.* **2019** ; Volume 13. 3921-3931.

[57] Klerx, M. H. P. Condition assessment of low voltage distribution grids. Technische Universiteit Eindhoven, Netherlands, 2020.

[58] Cynthia Rudin & al. (2012). Machine Learning for the New York City Power Grid. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34.2, 17.

[59] Asadi Majd, A.; Samet, H.; Ghanbari, T. k-NN based fault detection and classification methods for power transmission systems. *Prot Control Mod Power Syst*, **2017,** 2.

[60] Baskar, D.; Selvam, P. Machine Learning Framework for Power System Fault Detection and Classification. *International Journal of Scientific & Technology Research* **2019**, 9.

[61] Sapountzoglou, N. Fault detection and isolation for low voltage distribution grids with distributed generation, Université Grenoble Alpes, Grenoble, France, 2016.

[62] Ndeye, L.; Flaus, J.-M.; Adrot, O. Review of Machine Learning Approaches in Fault Diagnosis applied to IoT System. In Proceedings of the 2019 International Conference on Control, Automation and Diagnosis ICCAD'19, Grenoble, France, 2-4 July 2019.

[63] Huo, Y.; Prasad, G.; Lampe, L.; Leung, C. V. (2019, April). Smart-grid monitoring: Enhanced machine learning for cable diagnostics. In Proceedings of the 2019 IEEE International Symposium on Power Line Communications and its Applications (ISPLC), pp. 1-6.

[64] Khan MA, Kim J. Toward Developing Efficient Conv-AE-Based Intrusion Detection System Using Heterogeneous Dataset. *Electronics* **2020**, 9(11), 1771.

[65] Wenshuo, T.; Flynn, D.; Brown, K.; Valentin, R; and Zhao, X. The Application of Machine Learning and Low Frequency Sonar for Subsea Power Cable Integrity Evaluation. In Proceedings of the OCEANS 2019 MTS/IEEE SEATTLE, 2019, pp. 1-6.

[66] Toubeau, J.-F.; Bakhshideh Zad, B.; Hupez, M.; De Grève, Z.; Vallée, F. Deep reinforcement learning-based voltage control to deal with model uncertainties in distribution networks *Energies* **2020**, *13*(15), 3928.

[67] Association négaWatt Electrique - Manuel de travaux pratiques : 6. Calcul d'une liaison aérienne," Université de Liège - Faculté des Sciences Appliquées, http://www.tdee.ulg.ac.be/userfiles/file/6.pdf.

[68] Abdel-Majeed Ahmad & Braun Martin. (2012). Low voltage system state estimation using smart meters. *Universities Power Engineering Conference (UPEC), 47th International*. 1-6. 10.1109/UPEC.2012.6398598. 2012.

[69] Le monde de l'énergie. Réseaux électriques et changement climatique : une menace inévitable. Available online: https://www.lemondedelenergie.com/reseaux-electriques-changement-climatique/2019/09/12/ (accessed on December 2021).

[70] Ministère de la Transition écologique. Impacts du changement climatique : Atmosphère, Températures et Précipitations. Available online: https://www.ecologie.gouv.fr/impacts-du-changement-climatique-atmosphere-temperatures-et-precipitations (accessed on December 2021).

[71] Contribution du Groupe de travail I au cinquième Rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du climat. Available online: https://www.ipcc.ch/site/assets/uploads/2018/03/WG1AR5_SPM_brochure_fr.pdf (accessed on December 2021).

[72] National Centers for Environmental Information. Global Climate Report - Annual 2019. Available online: https://www.ncdc.noaa.gov/sotc/global/201913 (accessed on December 2021).

[73] H. Liu, C. Huang, Y. Chen and Y. Hou, " Probabilistic Power Flow Calculation Method for Low voltage Microgrid," *2013 IEEE Grenoble Conference*, Grenoble, 2013, pp. 1-5.

[74] V. Klonari, F. Vallée, O. Durieux, Z. De Grève and J. Lobry, "Probabilistic modeling of short term fluctuations of photovoltaic power injection for the evaluation of overvoltage risk in low voltage grids," *2014 IEEE International Energy Conference (ENERGYCON)*, Cavtat, 2014, pp. 897-903.

[75] G. Gruosso, P. Maffezzoni, Z. Zhang and L. Daniel, "Probabilistic load flow methodology for distribution networks including loads uncertainty," Electrical Power and Energy Systems 106, 2019, pp. 392–400.

[76] K. Christakou, M. Paolone and A. Abur, "Voltage Control in Active Distribution Networks Under Uncertainty in the System Model: A Robust Optimization Approach," *IEEE Transactions on Smart Grid*, Nov. 2018, pp. 5631-5642.

[77] H. Markiewicz and A. Klajn "Voltage DisturbancesStandard EN 50160 Voltage Characteristics inPublic Distribution System," Wroclaw University of Technology, July 2004.

[78] Vasiliki Klonari, "Probabilistic Analysis of Low Voltage Distribution Networks with Distributed Generation", Ph.D. Thesis, 12/09/2016, University of Mons

[79] D. Wright, "Basics of Monte Carlo Simulation," Geant 4 Tutorial at Lund University, 2018.

[80] F. Vallée, V. Klonari, J. Lobry and O. Durieux, "Study of the combined impact of auto-consumption behaviour and correlation level between prosumers on overvoltage probabilities in low voltage distribution grids," *2014 IEEE PES T&D Conference and Exposition*, Chicago, IL, 2014, pp. 1-6.

[81] M. Hupez, Z. De Grève and F. Vallée, "Simulating Time Dependent Technical Solutions in Distribution Networks Using Sequential Stochastic Analyses," 24th International Conference & Exhibition on Electricity Distribution CIRED, Glasgow, Scotland, June 2017.

[82] T.-H. Chen and N.-C. Yang, "Three-phase power-flow by direct ZBR method for unbalanced radial distribution systems," IET Generation Transmission & Distribution, 2009.

[83] BICC Cables Ltd. Electric cables handbook, 3rd ed.; Wiley-Blackwell, 1997.

[84] H. Saadat, ''Power system analysis,'' Mcgraw Hill, 1999.

[85] Ferhat Emre Kaya , " Newton-Raphson based Load Flow analysis of AC/DC distribution systems with distributed generation", Master of Science in Electrical and Electronics Engineering Department, 09/2019, Middle East Technical University

[86] Nexans Olex Company, "Low Voltage Aerial Bundled Cables," pp. 7-10, www.olex.com.au, October 2010.

[87] J. Sun, H. Gokturk and D. Kalyon, "Volume and surface resistivity of low-density polyethylene filled with stainless steel fibres," Journal of Materials Science, vol. 28, pp. 364-366, 10.1007/BF00357809, 1993

[88] Shalev-Shwartz, S.; Ben-David, S. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, 2014.

[89] Noteworthy the Journal Blog. A Quick Introduction to K-Nearest Neighbors Algorithm. 2017. Available online: https://blog.usejournal.com/a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7 (accessed on 19 November 2020).

[90] Chen, G.; Shah, D. Explaining the Success of Nearest Neighbor Methods in Prediction. In *Foundations and Trends in Machine Learning*, 2018; Publisher: Now Publishers Inc, pp. 105–199.

[91] Imandoust, S. B.; Bolandraftar, M. Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background. *International Journal of Engineering Research and Applications* **2013**, *3*, 605-610.

[92] Carnegie Mellon University – School of Computer Science. Artificial Intelligence: Representation and Problem Solving. Introduction to Learning & Decision Trees. 2007. Available online: https://www.cs.cmu.edu/afs/cs/academic/class/15381-s07/www/slides/041007decisionTrees1.pdf (accessed on 23 November 2020).

[93] Great Learning Blog. Decision Tree Algorithm Explained with Examples. 2020. Available online: https://www.mygreatlearning.com/blog/decision-tree-algorithm/ (accessed on 23 November 2020).

[94] Liu, C.; Rather, Z.; Chen, Z.; Bak, C.L. An overview of decision tree applied to power systems. *International Journal of Smart Grid and Clean Energy* **2013**, 413-419.

[95] JavaTpoint. Linear Regression vs Logistic Regression. 2018. Available online: https://www.javatpoint.com/linear-regression-vs-logistic-regression-in-machine-learning (accessed on 25 November 2020).

[96] Andrew, N. Machine Learning. Available online: https://fr.coursera.org/learn/machine-learning (accessed from September 2020 to Mars 2021).

[97] Ansari, M.; Srivastava, K.; Kaluri, R. Electricity Monitoring, Visualization and Prediction using Logistic Regression, 2017. 10.13140/RG.2.2.20915.94240.

[98] TIBCO Community. Random Forest Template for TIBCO Spotfire. 2020. Available online: https://community.tibco.com/wiki/random-forest-template-tibco-spotfire (accessed on November 2021).

[99] Tony Yiu. Understanding Random Forest - How the Algorithm Works and Why it Is So Effective. Available online: https://towardsdatascience.com/understanding-random-forest-58381e0602d2 (accessed on November 2021).

[100] Rohith Gandhi. Support Vector Machine - Introduction to Machine Learning Algorithms. Available online: https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47 (accessed on November 2021).

[101] Receiver Operating Characteristic (ROC) curve with False Positive Rate and True Positive Rate. Available online: https://en.wikipedia.org/wiki/File:Roc_curve.svg (accessed on November 2021)

Appendixes

Appendix A :

Technical Parameters of the Low Voltage (LV) distribution network

Table A: Technical parameter of the LV distribution network.

Line between :	Length (in m)	R (Ohms/m)	X (Ohms/m)
Node 1 and Node 2	46	206	310
Node 2 and Node 3	273	310	243
Node 3 and Node 4	62	310	243
Node 4 and Node 5	63	310	243
Node 5 and Node 6	194	310	243
Node 6 and Node 7	26	310	243
Node 7 and Node 8	11	310	243
Node 8 and Node 9	25	310	243
Node 9 and Node 10	21	310	243
Node 10 and Node 11	44	310	243
Node 11 and Node 12	10	310	243
Node 12 and Node 13	15	310	243
Node 13 and Node 14	14	310	243
Node 14 and Node 15	41	310	243
Node 5 and Node 16	396	310	243
Node 16 and Node 17	93	310	243
Node 17 and Node 18	117	310	243
Node 7 and Node 19	119	310	243

Appendix B :

Measured extreme temperatures during the 2022 heat waves

In 2022, high temperatures have been measured earlier in the year (before the summer) and in unexpected regions; as example on 18th June, not in the South of France (as usually until now) but on the West side (figure B1).

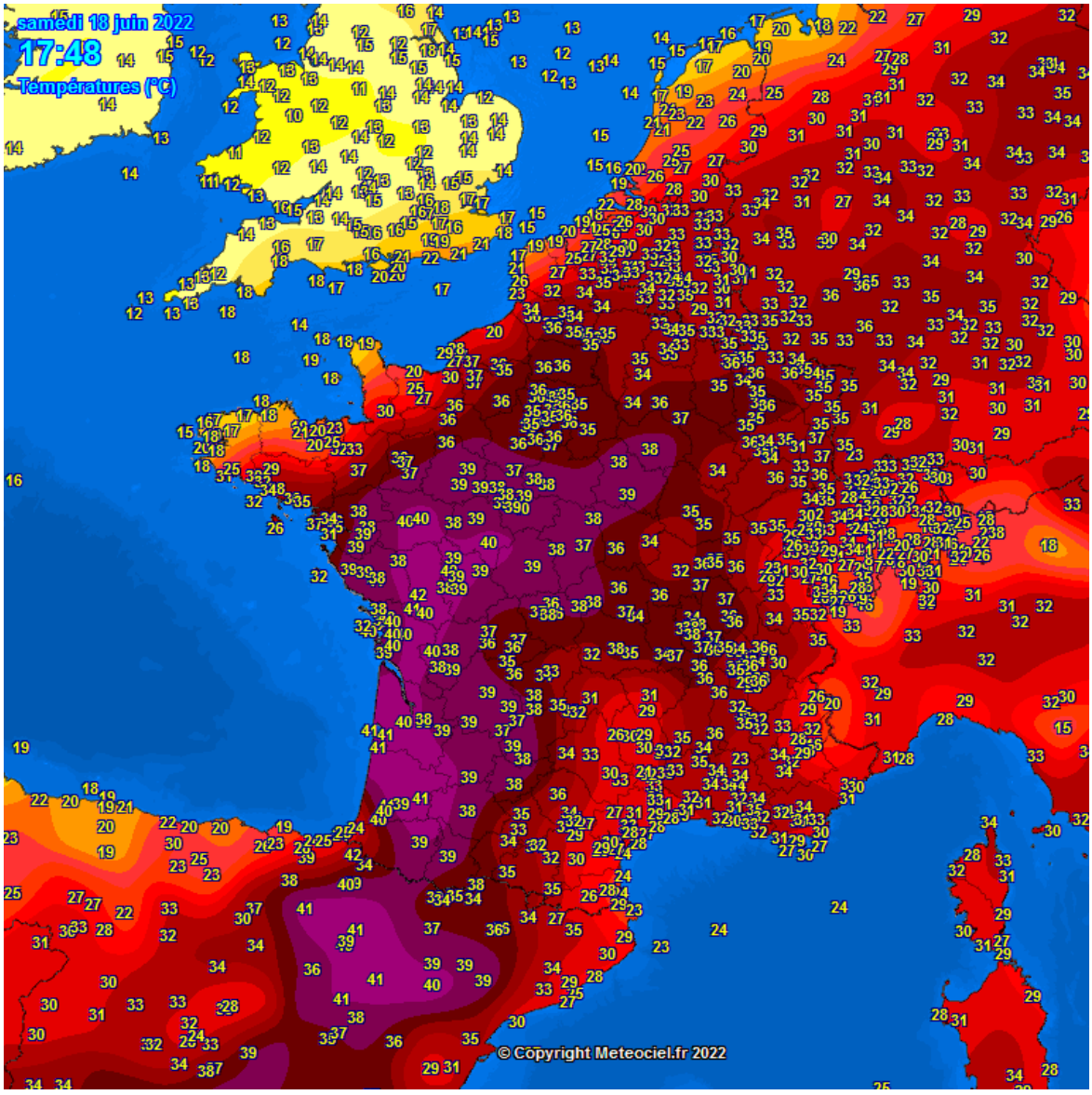


Figure B1: Measured maximal temperatures on June 18th, 2022 (© *https://www.meteociel.fr/observations-meteo/tmaxi.php*)

One month after, high temperatures reappear and again in the North of Europe; as on 19th July (figure B2).

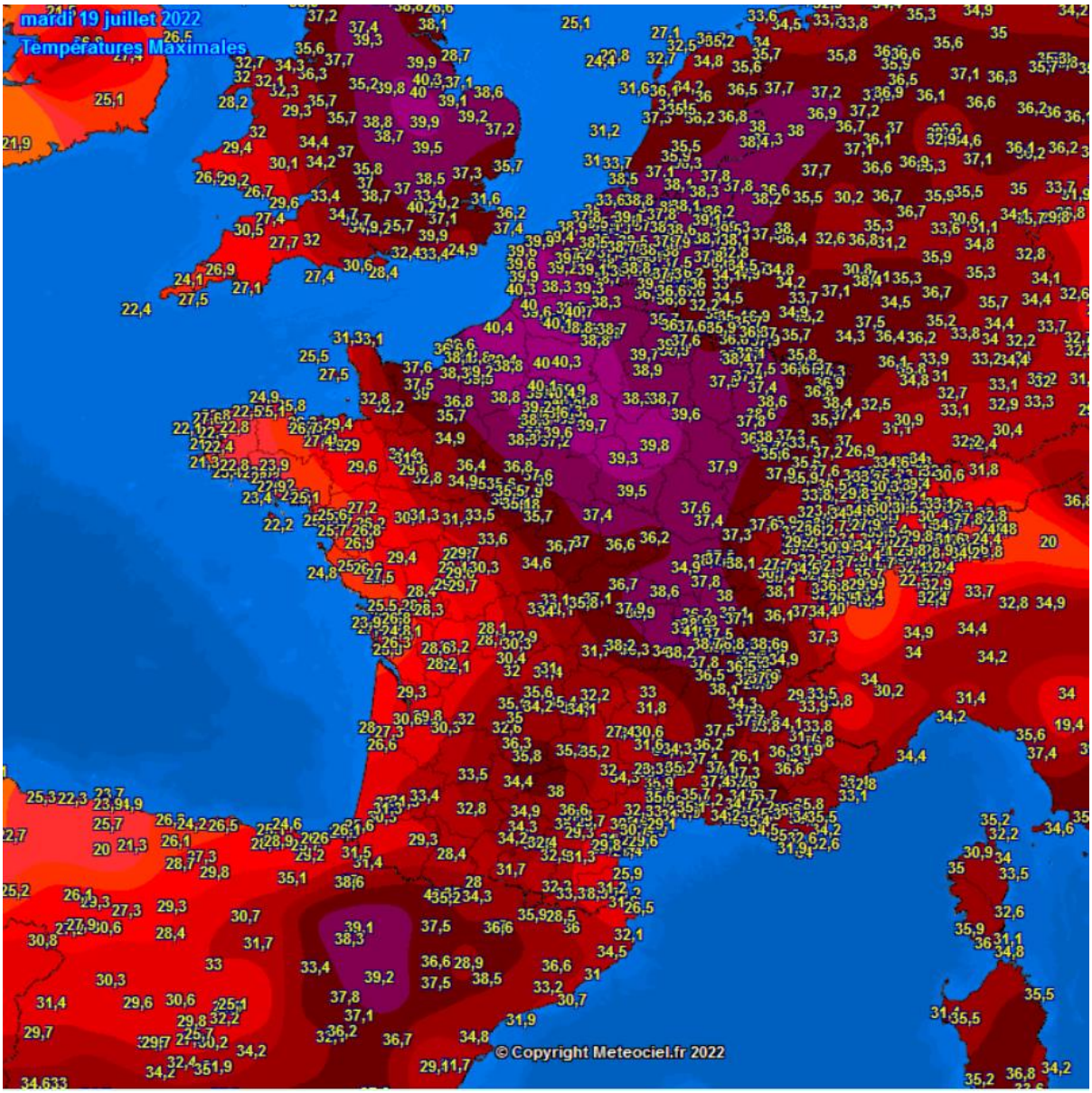


Figure B2: Measured maximal temperatures on July 19th, 2022 (©*https://www.meteociel.fr/observations-meteo/tmaxi.php*)

Appendix C :

Building and plotting a Cumulative Distribution Function (CDF)

This appendix explains how to construct a CDF profile from a known database. By definition, a Cumulative Distribution Function (CDF) is function *F* that represents the probability that a value *X* randomly sampled from a population, and evaluated at *x*, is equal to or less than *x* (mathematically $P[X \leq x]$).

Let remember first, the properties inherent to the CDF :

- Every CDF is non-decreasing
- Considering that the CDF maximum value is at *x*, $F(x) = 1$.
- The CDF ranges is from 0 to 1.
- $P[X = x]) = 0$.

On other hand, CDF is used to evaluate the accumulated probability of the known database (also assimilated to the area under the curve to the left from a point of interest).

Example 1 : Uniform distribution

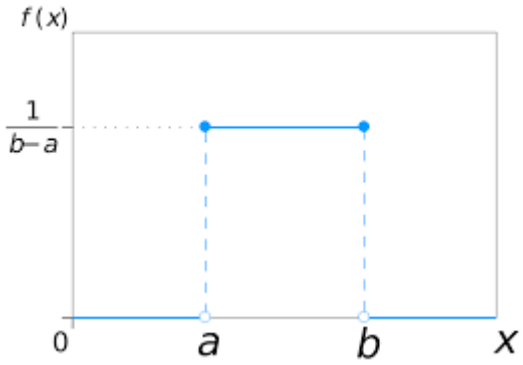


Figure C1: Representation of uniform distribution.

The shape of the distribution is described by the function *f(x)* below :

$$f(x) = \frac{1}{b-a} \quad \text{for } a<x<b$$

Let consider a point of interest *x* as show in Figure C2. We are trying to figure out the CDF i.e. the area of the surface in dark.

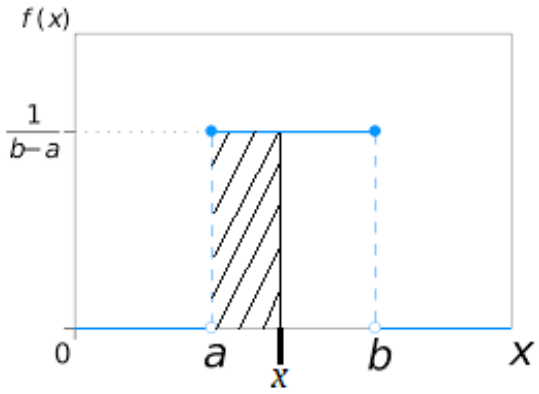


Figure C2: Representation of the point of interest x.

Knowing that the area of a rectangle is :

$$A_{rectangle} = Base * height$$

$$A_{dark} = (x - a) * \big(f(x)\big)$$

So, for a random value *X*, the Cumulative Distribution Function (CDF) will be mathematically expressed as :

$$P[X \leq x] = \frac{x - a}{b - a}$$

Finally, $P[X \leq x]$ calculated for the entire interval [a; b] will be represented as Figure C3.

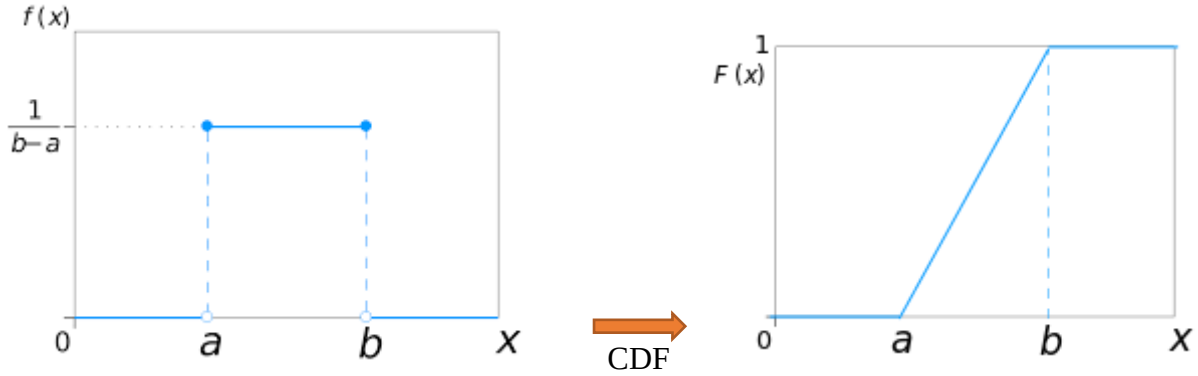


Figure C3: Representation of the CDF of a uniform distribution.

Example 2 : Exponential distribution

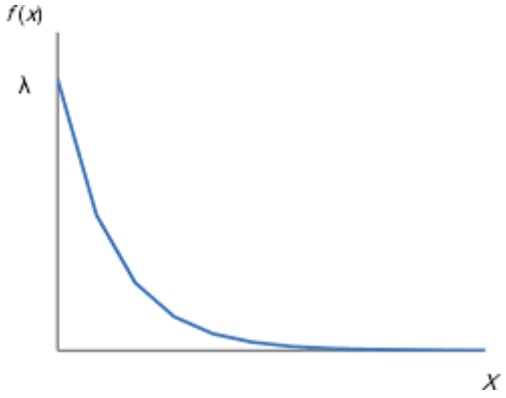


Figure C4: Representation of exponential distribution.

The shape of the distribution is described by the function *f(x)* below :

$$f(x) = \lambda\, e^{-\lambda x}$$

Let consider a point of interest *x* as show in Figure C4. We are trying to figure out the CDF i.e. the area in dark.

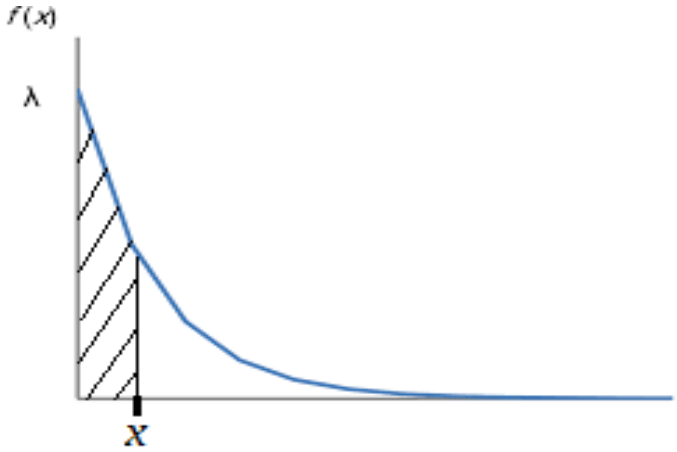


Figure C5: Representation of the point of interest x.

For a random value *X*, the Cumulative Distribution Function (CDF) will be mathematically expressed as :

$$P[X \leq x] = 1 - e^{-\lambda x}$$

Finally, $P[X \leq x]$ calculated for the entire interval [0 ; 10] will be represented as Figure C3.

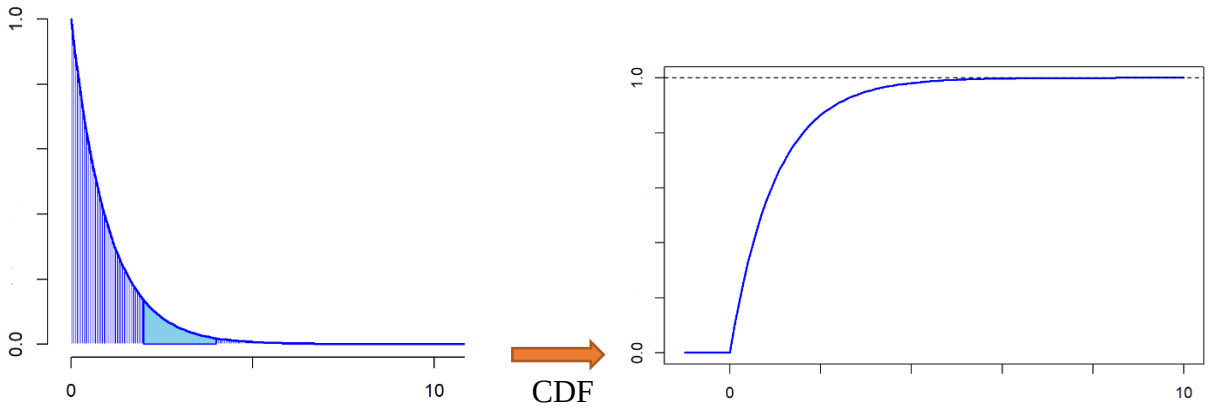


Figure C6: Representation of the CDF of an exponential distribution (© *https://www.alphacodingskills.com/scipy/scipy-exponential-distribution.php*).

Example 3 : Normal distribution

Let consider the normal distribution of Figure C7 represented by its Probability Distribution function (PDF) and the associated CDF calculated based on the above examples.

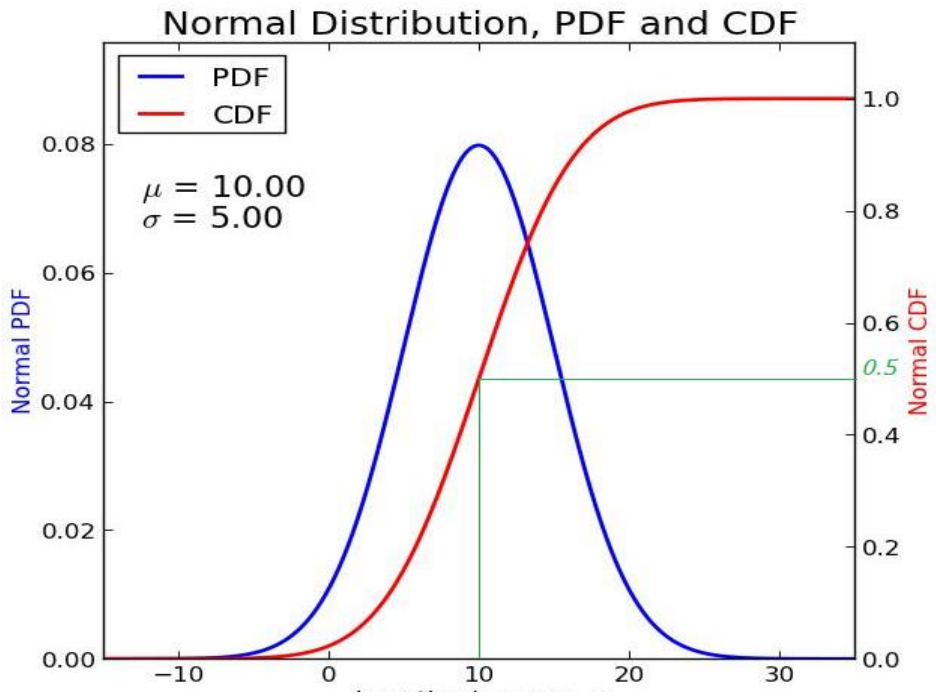


Figure C7: Representation of the CDF of a normal distribution (© *https://dwaincsql.com/2015/05/14/excel-in-t-sql-part-2-the-normal-distribution-norm-dist-density-functions/*).

So, the shape of the distribution (i.e. the PDF) is described by the function f(x) below :

$$f(x) = \frac{1}{\sqrt{2\pi * \sigma^2}} * e^{-\frac{1}{2}*\left(\frac{x-\mu}{\sigma}\right)^2}$$

where μ and σ are respectively the mean and the standard deviation of the dataset.

$$\mu = \frac{1}{N_{total}} \sum_{it=1}^{N_{total}} x_{it} \quad and \quad \sigma = \sqrt{\frac{1}{N_{total}} \sum_{it=1}^{N_{total}} (x_{it} - \mu)^2}$$

Then the CDF of a point of interest x is mathematically expressed by :

$$P[X \leq x] = \int_{-\infty}^{x} \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right)$$

Finally, $P[X \leq x]$ calculated for the entire interval, has been built by the red curve of Figure C7.

The Figure C7 also show that half of the distribution is accumulated by the time that we get to point B=10.

Application in studied cases

Regarding the database of this study (see section 2.5.2.2), the Figure C9 below has been constructed, from the database of Figure C8, to show the CDF of the line resistance for each season in one year.

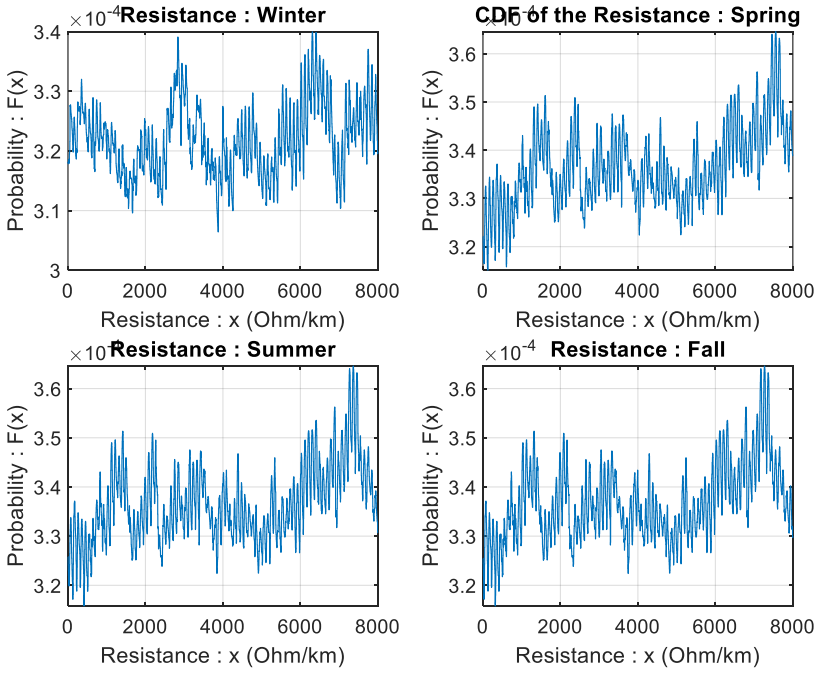


Figure C8: Seasonal profile of the line resistance.

By taking into account the temperature variation over one year, the Figure C8 shows the seasonal resistance distribution model expressed with the temperature and the electrical conductor resistivity.

So, the shape of each graph (named above f(x)) is built as:

$$f(x) = \left[\rho_o * \left(1 + \alpha_o * (x - To)\right)\right] * \frac{1}{S * 10^3}$$

where the parameter (x - T_o) is the temperature variation between the external temperature x (i.e. T_t in equation (2.15)) and the steady state temperature T_o.

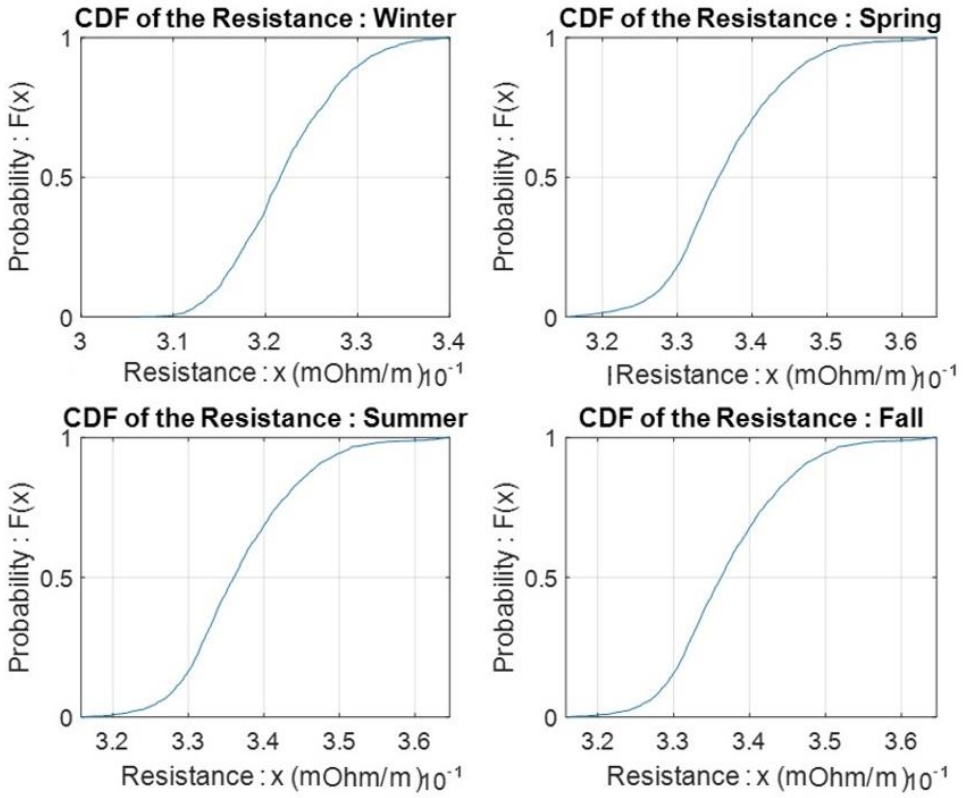
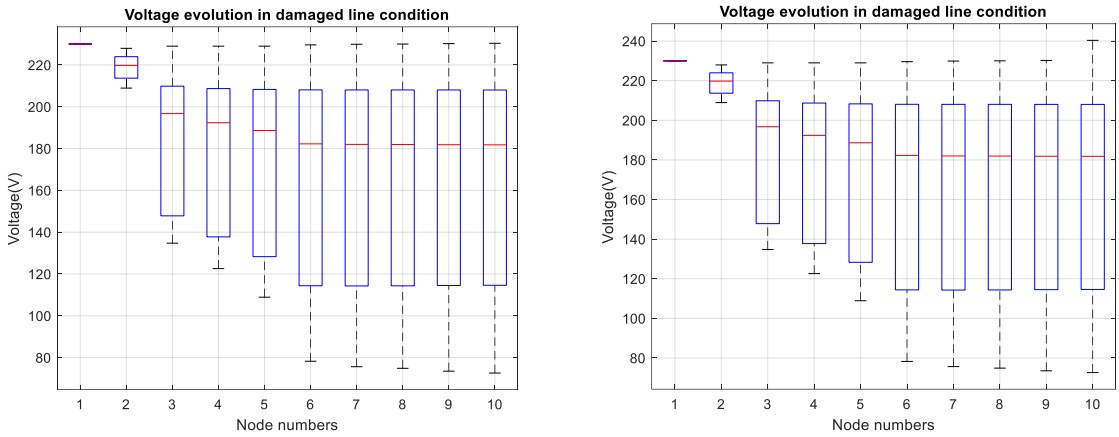


Figure C9: Seasonal CDF of the line resistance during each season.

Appendix D :

Comparison of performances without and with PV penetration under medium degradation

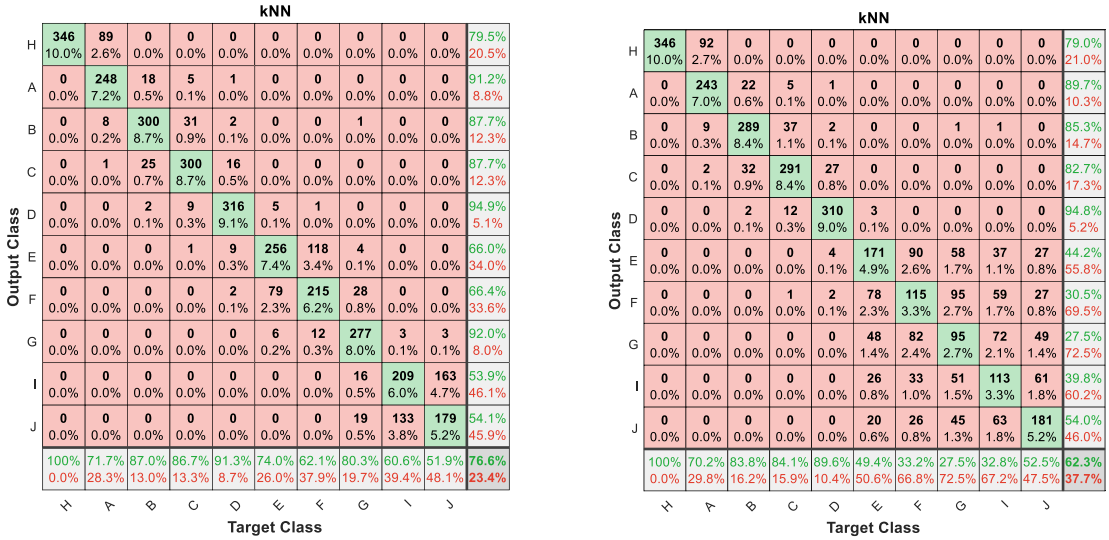
1. Study of the terminal node



(a) (b)

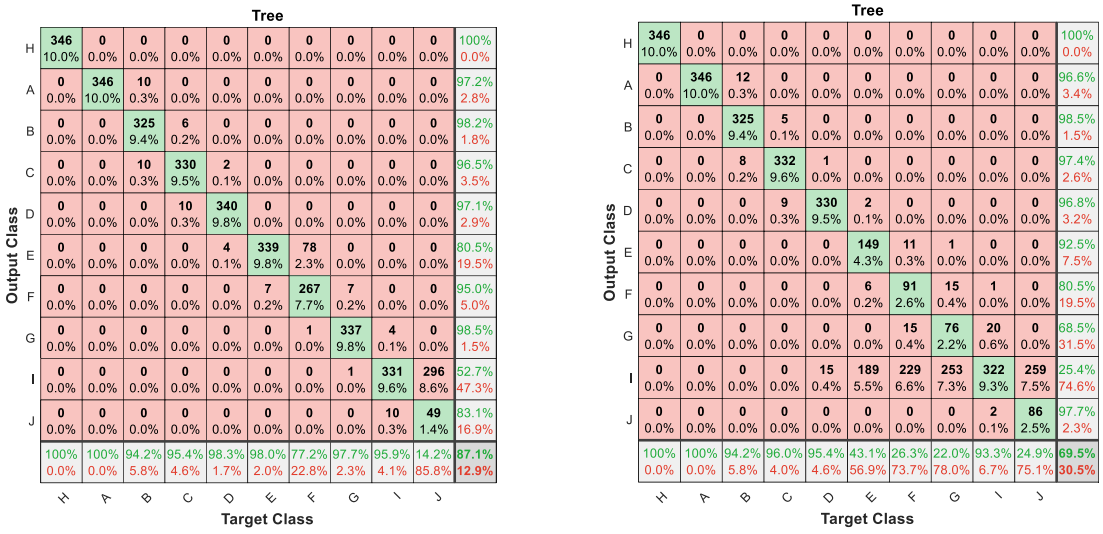
Figure D1: Boxplots of nodal voltages obtained by the Load Flow calculations: (a) for feeder with consumers only ; (b) for feeder with PV penetration on last node named node 10.

The consecutive figures below represents the 2-D confusion matrix, for both application cases, of kNN, DT, SVM and LR methods. The axis TargetClass and OutputClass correspond respectively to the known cable conditions (real classes from the original dataset) and to the outputs of the classifier (the predictions).



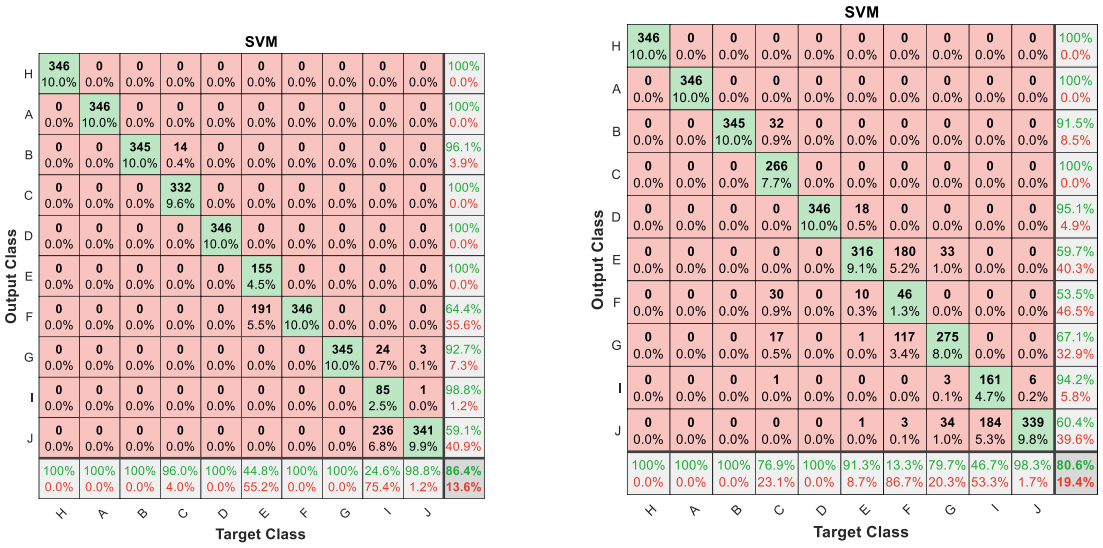
(a) (b)

Figure D2: kNN classification results : (a) for feeder with consumers only ; (b) for feeder with PV penetration on last node named node 10.



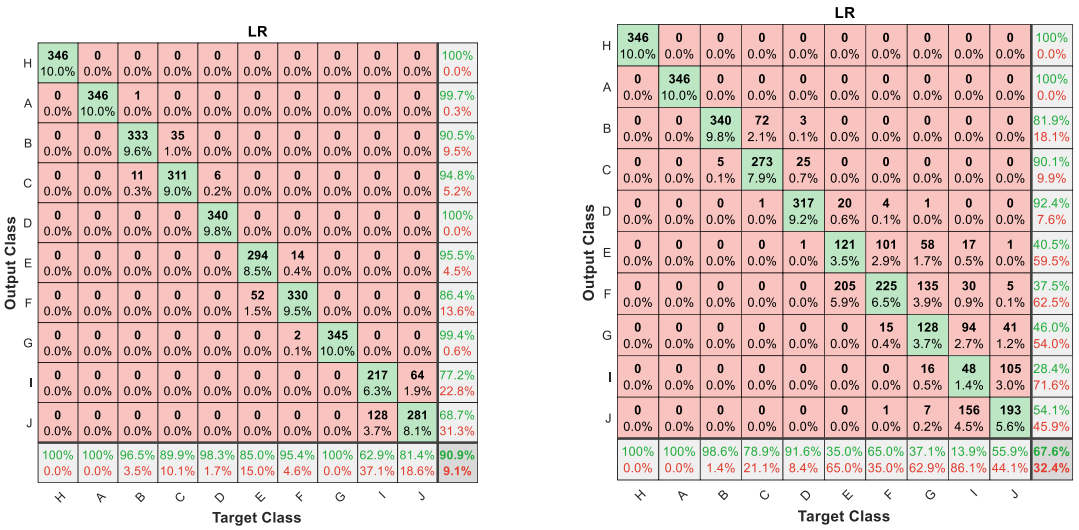
(a) (b)

Figure D3: DT classification results : (a) for feeder with consumers only ; (b) for feeder with PV penetration on last node named node 10.



(a) (b)

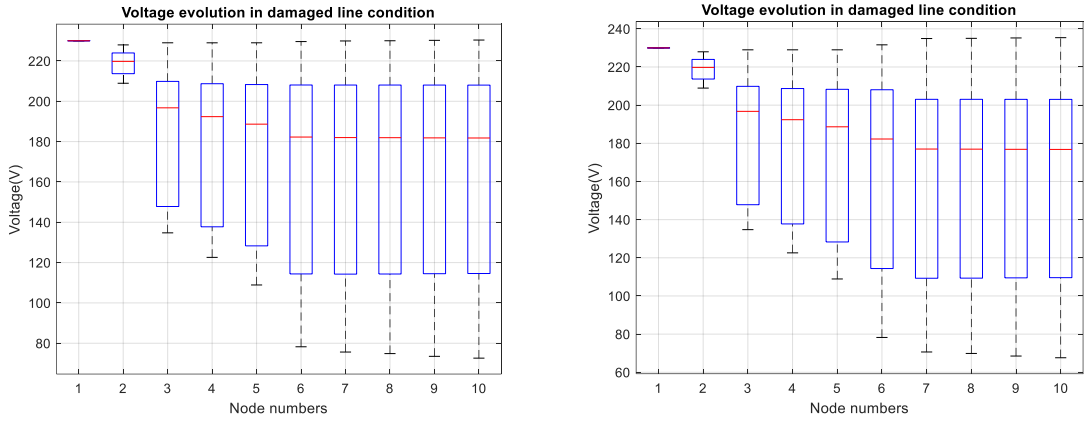
Figure D4: SVM classification results : (a) for feeder with consumers only ; (b) for feeder with PV penetration on last node named node 10.



(a) (b)

Figure D5: SVM classification results : (a) for feeder with consumers only ; (b) for feeder with PV penetration on last node named node

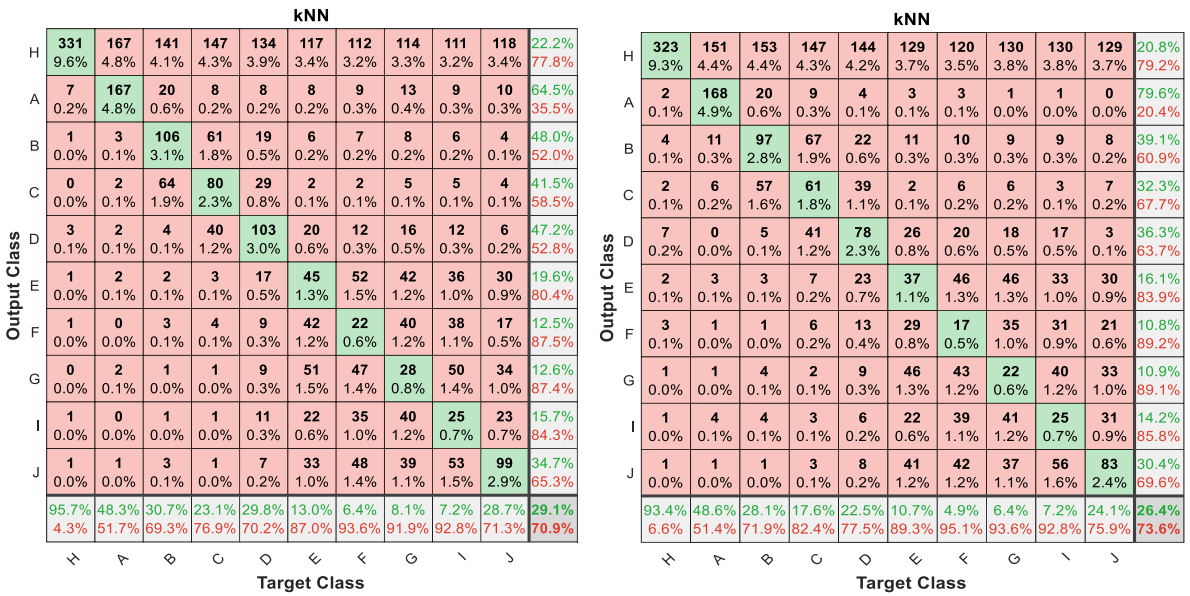
2. Study of the node 6



(a) (b)

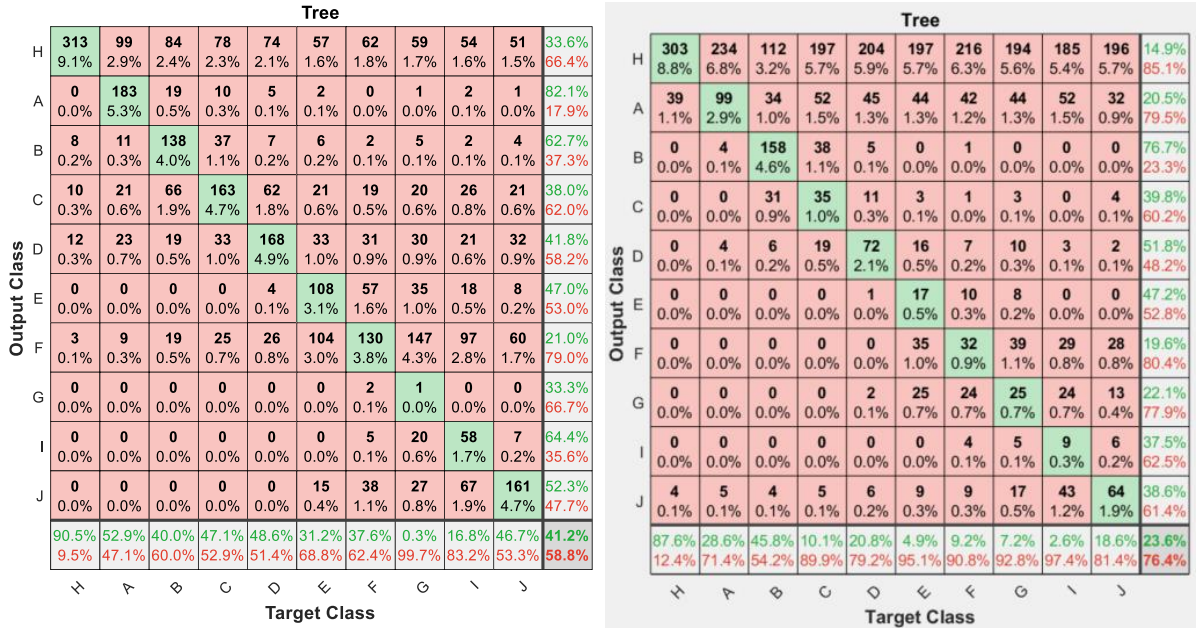
Figure D6: Boxplots of nodal voltages obtained by the Load Flow calculations: (a) for feeder with consumers only ; (b) for feeder with PV penetration on middle node named node 6.

The consecutive figures below represents the 2-D confusion matrix, for both application cases, of kNN, DT, SVM and LR methods. The axis TargetClass and OutputClass correspond respectively to the known cable conditions (real classes from the original dataset) and to the outputs of the classifier (the predictions).



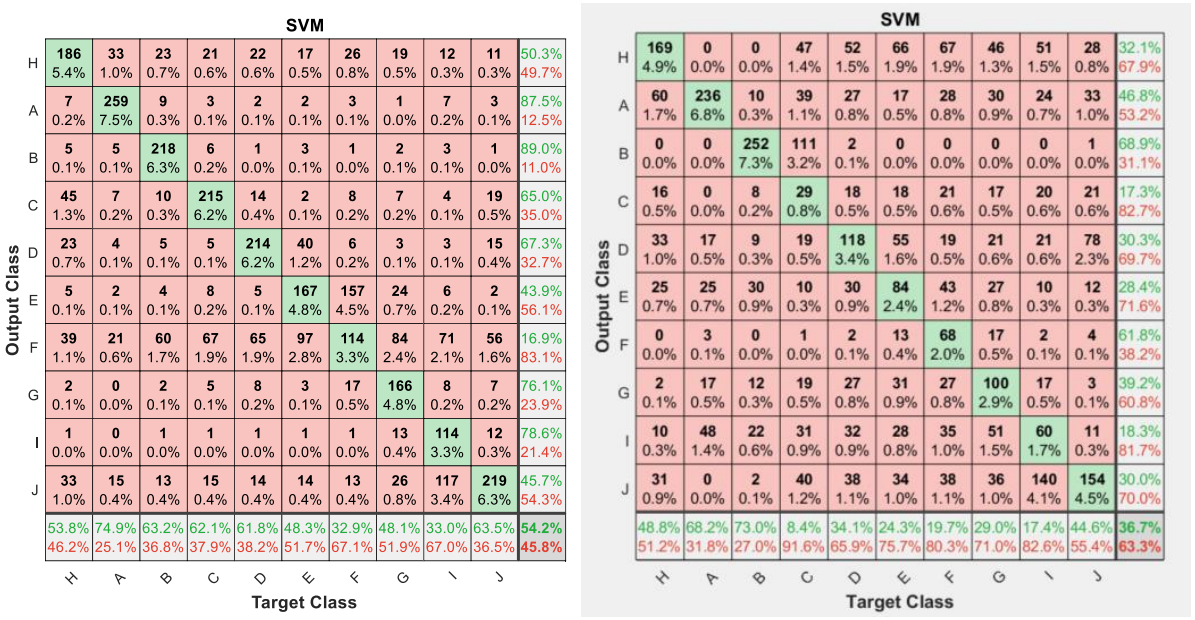
(a) (b)

Figure D7: kNN classification results : (a) for feeder with consumers only ; (b) for feeder with PV penetration on middle node named node 6.



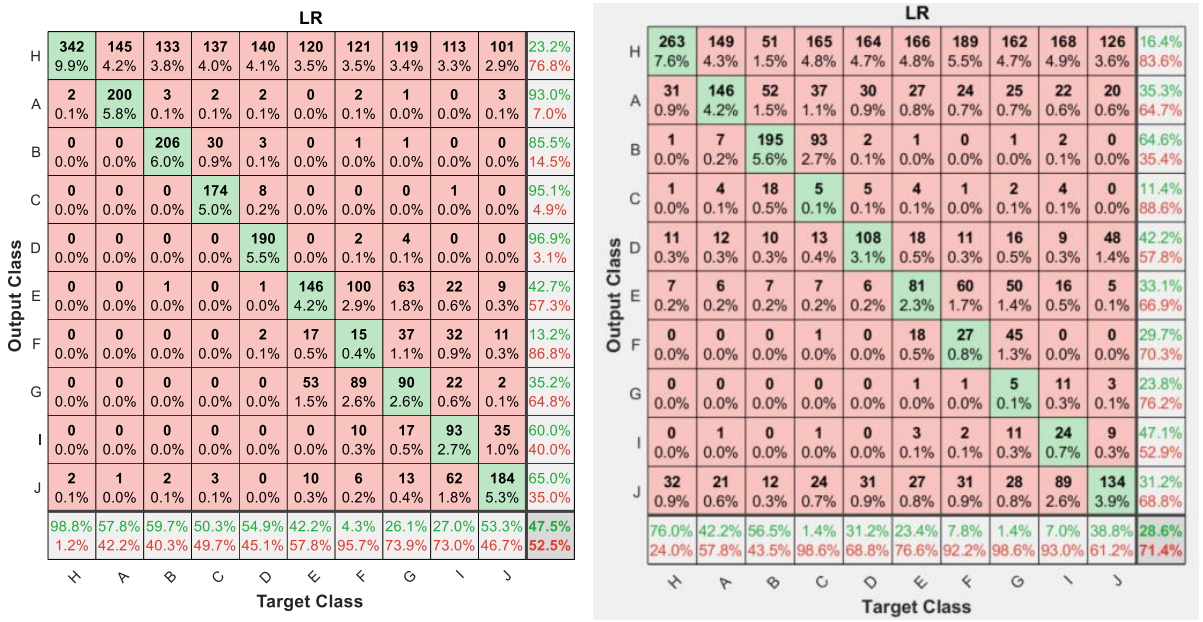
(a) (b)

Figure D8: DT classification results : (a) for feeder with consumers only ; (b) for feeder with PV penetration on middle node named node 6.



(a) (b)

Figure D9: SVM classification results : (a) for feeder with consumers only ; (b) for feeder with PV penetration on middle node named node 6.



(a) (b)

Figure D10: LR classification results : (a) for feeder with consumers only ; (b) for feeder with PV penetration on middle node named node 6.