Power Losses Estimation in Low Voltage Smart Grids

by

José Ángel Velasco Rodríguez

A dissertation submitted in partial fulfilment of the requirements for the Degree of Doctor of Philosophy in

ELECTRICAL ENGINEERING, ELECTRONICS AND
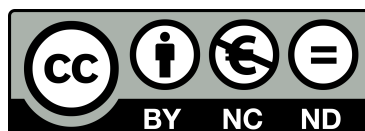
AUTOMATION

Universidad Carlos III de Madrid

Advisor:

Hortensia Amarís Duarte

Tutor:

Hortensia Amarís Duarte

March, 2022

"All we have to do is decide

what to do with the time

that is given to us"

The Fellowship of the Ring, R. R. Tolkien. (1954)

Acknowledgements

First and foremost, I would like to express my gratitude to my thesis advisor Hortensia Amaris, for their wise guidance, high availability, good advices, and patience over all these years. Thank you for giving me the opportunity to undertake this research work.

Thanks to the University Carlos III of Madrid (UC3M) as institution for supporting me with the resources necessaries to carry out this thesis.

I am also very grateful to Andrew Keane, Alireza Soroudi and Valentin Rigoni for hosting me in 2018 at the Energy Institute of the University College Dublin (UCD). My research stay there had greatly helped to improve the quality of this thesis and also was an amazing experience spend that time abroad.

I would also like to thank the rest of members composing the UC3M Electrical Engineering Department. Thanks to all my colleagues at Indra, Capgemini Engineering and Repsol who motivated me in the last stage of this journey.

Most heartfelt appreciation and gratitude to my loved Laura, who have been part of this thesis as much as every single word, supporting me and motivating me specially in the most challenging and arduous times.

Last, but not least, I would like to give special thanks to my family: Jose Antonio, Felisa, Alejandro and Toby, for their endless love and unconditional support over these years, without which I would have never come this far.

Published and Submitted Contents

Journal Papers:

1. J-A. Velasco, H. Amaris and M. Alonso, **"Deep learning-Based Power Losses Estimation Model for Large-Scale Distribution Areas,"** *Electric Power System Research* , vol. 121, Oct. 2020, no. 106054.

   - Published (https://doi.org/10.1016/j.ijepes.2020.106054).

   - Author Contribution: J-A. Velasco wrote the manuscript with input from all authors. He was responsible for conceptualise, analyse and develop the investigation and software of this paper. All authors were responsible for written, editing and review the paper.

   - Totally included in the Thesis, Chapter 4.

   - The material from this source included in this thesis is not singled out with typographic means and references.

Conference Papers:

1. J-A. Velasco, V. Rigoni, A. Soroudi, A. Keane and H. Amaris, **"Optimising Load Flexibility for the Day Ahead in Distribution Networks with Photovoltaics"** in IEEE PES PowerTech 2019, Milano (Italy), July 2019.

   - Published (https://doi.org/10.1109/PTC.2019.8810963).

   - Author Contribution: J-A. Velasco wrote the manuscript with input from all authors. He was responsible for preparing the case study, develop the model, prepare the results and figures.

- Partially included in the Thesis, Chapter 5.

- The material from this source included in this thesis is not singled out with typographic means and references.

2. J-A. Velasco, H. Amaris, M. Alonso and M. Casas, **"Energy Losses Estimation Tool for Low Voltage Smart Grids"** in International Conference and Exhibition on Electricity Distribution (CIRED 2019), Madrid (Spain), June 2019.

- Published (http://hdl.handle.net/10016/29232).

- Author Contribution: J-A. Velasco wrote the manuscript with input from all authors. He was responsible for preparing the case study, develop the model, prepare the results and figures.

3. J-A. Velasco, H. Amaris and M. Alonso and M. Miguelez,**"Stochastic Technical Losses Analysis of Smart Grids under Uncertain Demand"** in International Universities Power Engineering Conference (UPEC2018), Glasgow (Scotland), September 2018.

- Published (https://doi.org/10.1109/UPEC.2018.8542099).

- Author Contribution: J-A. Velasco wrote the manuscript with input from all authors. He was responsible for preparing the case study, develop the model, prepare the results and figures.

- Totally included in this Thesis, Chapter 3.

- The material from this source included in this thesis is not singled out with typographic means and references.

4. J-A. Velasco, H. Amaris and M. Alonso and M. Miguelez, **"Energy Losses Estimation in Low Voltage Smart Grids by using Loss Maps,"** in International Conference on Energy, Environment and Economics (ICEEE2018), Edinburgh (Scotland), August 2018.

- Published (https://doi.org/10.32438/WPE.5218).

- Author Contribution: J-A. Velasco wrote the manuscript with input from all authors. He was responsible for preparing the case study, develop the model, prepare the results and figures.

Abstract

One of the European Union Targets was to replace at least 80% of all traditional energy meters with electronic smart meters by 2020. However, by the end of 2020, the European region (EU 27 including the UK) had installed no more than 150 million smart electricity meters, representing a penetration rate of 50% for smart meters. By 2026, It is expected that there will be more than 227 million smart meters in households due to the updated planning and target numbers, which will affect many European markets, including western and northern Europe. This scenario would contribute to the general purpose of building a more sustainable distribution system for the future.

This thesis contributes to the field of power losses estimation and optimization in low-voltage (LV) smart grids in large-scale distribution areas. To contextualize the importance of the research, it has been necessary to explain the unbalanced nature of low voltage distribution networks where there is a huge deployment of smart meter rollout, and there is also uncertainty related to renewable energy generation. Main results of the thesis have been applied in two smart grid research projects: the national project OSIRIS (*Optimización de la Supervisión Inteligente de la Red de Distribución*) and the European project IDE4L (*Ideal Grid For All*).

Smart metering infrastructure allows distributor system operators (DSOs) to have detailed information about the customers energy consumption or generation. Smart meters measure the active and reactive energy consumption/generation of customers using different discrete time resolutions which range from 15-60 min. A large-scale smart meter rollout allows service providers to gain information about the energy consumed and produced by each customer in near-real time. This knowledge can be used to compute the

aggregated network power losses at any given time. In this case, network power losses are calculated by means of customers' smart meters measurements, in terms of both active and reactive energy consumption, and by the energy measured by the smart meter supervisor located at the secondary substation (SS).

The problem of network losses estimation becomes more challenging as a results of the existence of not-technical losses due to electricity fraud or smart meter measurements anomalous (null or extremely high) or even because there are customers' smart meters that can be out of service.

One of the differential keys of LV smart grids is the presence of single-phase loads and unbalanced operation, which makes it necessary to adopt a complete three-phase model of the LV distribution network to calculate the real value of the power losses. This scenario makes the process of power loss estimation a computationally intensive problem. The challenge is even greater when estimating the power losses of large-scale distribution networks, composed of thousands of SSs.

In recent years, environmental concerns have led to the increasing integration of a considerable number of distributed energy resources (DERs) into LV smart grids. This fact prompts DSOs and regulators to provide the maximum energy efficiency in their networks (i.e., the smallest power loss values) and maximum sustainable energy consumption. Detailed understanding of the network's behavior in terms of power losses and the use of electricity is necessary to achieve this energy efficiency.

However, the above scenario presents some drawbacks. The integration of DERs units, such as photovoltaic (PV) panels, into distribution networks can produce an increment of network power losses if the DERs units are not optimally located, coordinated, or controlled. Additionally, the network can experience technical contingencies such as cable's overloads and nodal over-voltages or can lead to an inefficient system operation due to high energy losses or cables that exceed thermal limits. Moreover, there is a great uncertainty associated with the distributed power generation from PVs because its energy generation depend on weather conditions, including ambient temperature and solar irradiance, which are highly intermittent and fluctuating. Uncertainty is also present in some loads with stochastic behavior, such as plug-in electric vehicles (PEV), which adds an

uncertainty layer and makes their optimal integration more complex.

Therefore, DSOs require advanced methods to estimate power losses in unbalanced large-scale LV smart grids under uncertain situations. Such estimations would facilitate the deployment of policies and practices that lead to a safe and efficient integration of DERs in the form of flexibility mechanisms. In this context, flexibility mechanisms are essential to achieve optimal operation conditions under extreme uncertainty. Flexibility mechanisms can be deployed to tackle the imbalance between generation and demand that results from the uncertainty that is latent in LV smart grids.

These flexibility mechanisms are based on modifying the normal power consumption (for the demand side) or power generation (for the generation side), according to a flexibility scheduling at the request of the network operator.

In summary, DSOs face the challenge of managing network losses over large geographical areas where there are hundreds of secondary substations and thousands of feeders, with multiple customers and an ever-increasing presence of renewable DERs. Power losses estimation is thus paramount to improve network energy efficiency in the context of the European Union energy policies. This situation is complicated by the unbalanced operation of those networks and the presence of uncertainty. To address these challenges, this thesis focuses on the following objectives:

1. Power losses estimation in unbalanced LV smart grids under uncertainty.

2. Power losses estimation in unbalanced LV smart grids in large areas with a presence of DERs.

3. Flexibility scheduling for power losses minimization in unbalanced smart grids under uncertainty.

The mentioned objectives are achieved by taking advantage of smart metering infrastructures, machine and deep learning models and mathematical programming techniques which allows DSOs to reduce their total power losses within the distribution network. This approach entails using flexibility mechanisms to operate the distribution network optimally and enhance the load management and DG expansion planning.

According to the objectives identified earlier, the main contributions of this thesis are the following:

1. Power losses estimation in unbalanced LV smart grids under uncertainty conditions.

   ✓ An optimization-based procedure to estimate load consumption of non-telemetered customers.

   ✓ A Markov chain-based process to estimate intra-hour load demand for data having a low resolution and for non-telemetered customers or customers which smart meters provide incorrect measurements.

2. Power losses estimation in unbalanced LV smart grids in large-scale areas with a presence of DERs.

   ✓ A data mining approach to reduce a high-dimensionality dataset in smart grids to yield a reduced set of relevant features.

   ✓ A clustering process to obtain representative feeders within a large-scale distribution area of smart grids.

   ✓ A deep learning-based power losses estimator for large-scale LV smart grids. The method is formulated as a deep neural network that uses as input features the power load demand and power generation of a set of representative feeders. The model gives, as output, the power losses of the whole area.

3. Flexibility scheduling for power losses minimization in unbalanced smart grids under uncertainty.

   ✓ A robust optimization model for the flexibility scheduling optimization model for unbalanced smart grids with distributed resources, such as PV panels and PEV devices.

Nomenclature

Variables expressed in upper case are expressed in real magnitude (A, kV, kW, kVA or KVAr) meanwhile variables expressed in lower case are in per unit system. Variables with brackets $[U]$ denotes a matrix meanwhile bold letters $\mathbf{u}$ denotes vectors. Magnitudes with a flat bar on top $\bar{u} = Re\{u\} + jIm\{u\}$ denotes a complex number. Note that the flat bar used for mean value is different $\overline{\mathbf{u}}$. Subindex are used to indicate the belonging of a variable to a specific set $u_k, \forall k \in \Omega_k$. Electrical magnitudes of voltage and current as well as other time series has a temporal dependency, but to deliver a more readable notation, that dependency is omitted in certain equations in this thesis in the following way: $u_k(t) \rightarrow u_k$. Latin letters $(\bar{x}, s^2, r_{x,y}, \ldots, etc.)$ are used for sampled variables while greek letters $(\mu, \sigma^2, \rho_{x,y}, \ldots, etc.)$ for the statistics of the whole population.

As reference power S_B is used the rating of the secondary substation transformer (MV/LV), meanwhile the reference voltage U_B it set to the nominal line-to-line voltage $(400V)$. For the rest of principal electrical magnitudes, the reference current is $I_B = S_B/\sqrt{3}U_B$ meanwhile the reference impedance is $Z_B = U_B^2/S_B$ (for star connection).

Acronyms

AI	Artificial Intelligence
AP	Aggregated Patterns
APE	Absolute Percentage Error
AR	Auto-Regressive
AIC	Akaike Information Criteria
ARIMA	Auto-Regressive Integrated Moving Average
ANN	Artificial Neural Network

AMI	Advanced Metering Infrastructure
ADN	Active Distribution Network
BES	Battery Energy System
BIRCH	Balanced Iterative Reducing and Clustering using Hierarchie
BFS	Back Forward Sweep
BR	Bayesian Regularisation
BP	Back Propagation
CCG	Column and Constrained Generation
CNN	Convolutional Neural Network
COUENNE	Convex Over and Under ENvelopes for Nonlinear Estimation
CLARANS	Clustering Large Applications based on RANdom Search
CURE	Clustering Using REpresentative
CDF	Cumulative Distribution Function
CVR	Conservative Voltage Reduction
CWFR	Cluster-Wise Fuzzy Regression
DR	Demand Response
DLC	Daily Load Curve
DSO	Distribution System Operator
DG	Distributed Generation
DER	Distributed Energy Resources
DL	Deep Learning
DT	Decision Trees
DO	Deterministic Optimisation
EM	Expectation Maximisation
EV	Expected Value
ECT	Energy Consumption Tendency
EEV	Expectation of the expected Value Solution
ELFM	Empirical Loss Factor Method
EH	Equivalent Hours
EHL	Equivalent Hours of Losses

FS	Flexibility Scheduling
FNN	Feedforward Neural Networks
FCN	Fuzzy C-Number
GMM	Gaussian Mixture Model
GS	Global Silhouette coefficient
GBDT	Gradient Boosted Decision Trees
GLM	Generalised Linear Models
GAM	Generalised Additive Models
HMM	Hidden Markov Models
IGDT	Information Gap Decision Theory
KNN	K-Nearest Neighbours
KDE	Kernel Density Estimation
LASSO	Least Absolute Shrinkage and Selection Operator
LDA	Linear Discriminant Analysis
LDC	Load Duration Curve
LC	Load Curtailment
LV	Low Voltage
LFM	Loss Factor Method
LSTM	Long Short-term Memory
LF	Load Factor
ML	Machine Learning
MV	Medium Voltage
MA	Moving Average
MAPE	Mean Absolute Percentage Error
MDMS	Meter Data Management System
MINLP	Mixed-Integer Non-Linear Programming
MISE	Mixed-Integrated Squared Error
MAE	Mean Absolute Error
MSE	Mean Squared Error
MLP	Multilayer Perceptron

NTL	Non-Technical Losses
NTM	Non-Telemetered Customers
NLP	Non-Linear Programming
NTL	Non-Technical Losses
NOCT	Nominal Operating Cell Temperature
OPF	Optimal Power Flow
OLS	Ordinary Least Squares
PL	Power Losses
PrL	Percent Loading
PF	Power Flow
PC	Principal Component
PCA	Principal Component Analysis
PLC	Power Line Communications
PLE	Peak Losses Energy (LFM)
PLD	Peak Losses Demand (LFM)
PLM	Power Losses Minimisation
PRIME	PoweRline Intelligent Metering Evolution
PEV	Plug-in Electric Vehicle
PS	Peak Saving
PV	Photovoltaic
PDF	Probabilistic Density Function
RGHM	Renewable Generation Harvesting Maximisation
RO	Robust Optimisation
ROCK	RObust Clustering using linKs
RP	Representative Patterns
R&D	Research & Development
REP	Recourse Problem
RMSE	Root Mean Squared Error
RBF	Radial Basis Function
SARIMA	Seasonal ARIMA

SM	Smart Meter
STC	Standard Test Conditions
SP	Single-Phase
3P	Three-Phase
SS	Secondary Substation
SO	Stochastic Optimisation
SVM	Support Vector Machines
SGD	Stochastic Gradient Descent
TL	Technical Losses
OF	Optimal Function
OCH	Optimal Charging
OSIRIS	Optimización de la Supervision Inteligente de la Red de DIstribución
VUM	Voltage Unbalance Minimisation
VSS	Value of Stochastic Solution
UOPF	Unbalanced Optimal Power Flow
WLS	Weighted Least Squares
WEC	Weakly Energy Consumption

Sets

Ω_C	Set of nodes with low-voltage customers
Ω_{PV}	Set of nodes with PV panels
N_ω	Set of scenarios for stochastic programming
V	Set of network nodes/buses
E	Set of network edges/feeder lines
$\{a, b, c, n\}$	Set of network phases
Ω_d	Set of days in the time horizon
T_d	Set of hours in the time horizon
$\mathcal{N}$	Set of neurons of the DNN
$\mathcal{E}$	Set of directed edges of the DNN
$\Omega_{T,i}$	Input training data set for DNN model

$\Omega_{T,o}$	Output training data set for DNN model
Υ	Set of hyper-parameter combinations

Parameters

U_N	Reference voltage magnitude
S_N	Reference power
S_{inv}	Rate inverter AC/DC
$T_{amb,k}$	Ambient Temperature of the PV panel at node (k)
$T_{cell,k}$	Cell temperature of the PV panel at node (k)
$\cos(\phi_{iv})$	Power factor of the inverter AC/DC of the PV panel
δ_{PV}	PV panel temperature coefficient
$r_{Cu,90°}$	Electrical resistance of Cooper (Ω/m) with $90°C$
$\rho_{Cu,90°}$	Electrical resistivity of Cooper $(\Omega \cdot mm^2/m)$ with $90°C$
S	Cross cable section (mm^2)
α_{Cu}	Temperature coefficient for cooper resistivity $(\%/\circ C)$
P_{IMP}^S	Aggregated imported active power demand registered by the SM supervisor
$P_{k,p}^d$	Active power demand of customer (k) in phase (p)
$E_{D,\tau}^k$	Energy consumption customer (k), during time interval (τ)
$E_{IMP,\tau}$	Energy consumption registered by SM supervisor during time interval (τ)
$\ell_{i,j}^p$	Length of the phase feeder conductor that connects nodes (i) and (j)
r_{ij}^p	Electrical resistance of phase feeder that connects node (i) to node (j)
P_{ctd}	Contractual power
T_{ch}	PEV charging session
$\widehat{P}_D$	Metered aggregated active power demand
P_{PV}	Aggregated active power generation from PV panels
$\mathcal{G}$	Network graph
T_h	Number of hours of a year
LL	Max. power losses rate of a distribution component (LFM)
$\mathbb{X}$	Input of the supervised machine learning model
$\bar{z}_{kj}^{pp}$	Complex self-impedance of phase (p) and nodes (k)-(j)

$\bar{z}_{kj}^{pp'}$	Complex impedance between phases (p)-(p') and nodes (k)-(j)
$e_{k,j}$	Feeder line that connects nodes (k) and (j)
$\widehat{G}_k$	Solar irradiance (W/m^2) at node (k)
PV_k	Peak power output of a PV panel (W/m^2) at node (k)
$\mathbf{p}_{i,j}^{(t)}$	Probability shifting from the state $X_t = i$ to the state $X_{t+1} = j$
X_t	Discrete state of a Markov chain
$\mathcal{P}$	Transitional probability matrix
X	Feeder's features data matrix $X \in \mathbb{R}^{n \times \tilde{p}}$
M	Normalised feeder's features data matrix
$\widetilde{U}$	PCA Projection matrix
$\widetilde{M}$	Reduced normalised feeder's features data matrix
S	Covariance matrix $S \in \mathbb{R}^{\tilde{p} \times \tilde{p}}$
λ_k	k^{th} Eigenvalue associated to the covariance matrix S
$\eta(\%)$	Cumulative variability captured
β_λ	Total number of eigenvalues of the covariance matrix S
ε	Stop criteria for the feeders clustering
$\hat{k}$	Objective number of representative feeder's clusters
$\zeta_{\mathbf{u}}$	Uncertainty budget set for uncertainty variable $\mathbf{u}$
L_{in}	Input layer of the DNN model
L_o	Output layer of the DNN model
$\mathcal{G}_k$	The kth neuron of the DNN model
ζ	Dropout of the DNN model
η	Learning rate of the DNN model
ϑ	Number of epochs of the DNN model
ν	Batch size of the DNN model

Variables

P_D	Aggregated active power demand of all low-voltage customers
Q_D	Aggregated reactive power demand of all low-voltage customers
$\widetilde{P}_D$	Estimated aggregated active power demand

$\widetilde{P}_{PV}$	Estimated aggregated active power generation from PV panels
$\widetilde{P}_{PEV}$	Estimated aggregated PEV charging demand
$\widetilde{P}_{BES}^{+}$	Estimated aggregated BES charging load
$\widetilde{P}_{BES}^{-}$	Estimated aggregated BES discharging load
P_{IMP}	Imported total active power
P_{EXP}	Exported total active power
$E_{D,\tau}$	Aggregated energy consumption of customers during time interval (τ)
x_{eq}	Phase load unbalance equilibrium variable
p_i^p	Total active power injected in phase (p)
p_g^p	Total active power generated in phase (p)
p_d^p	Total active power demand in phase in phase (p)
i_{ij}^p	Current magnitude in phase (p) from node (i) to node (j)
I_T^p	Total secondary substation phase current magnitude in phase (p)
φ_p	Total secondary substation phase current angle in phase (p)
P_{LOSS}	Aggregated active power losses
P_{LOSS}^p	Aggregated active (technical) power losses of phase p
$\mathcal{L}_p$	Relative total active (technical) power losses of phase p
$K_{LOSS}(S_D)$	Relative power losses linear constant
$P_{LOSS,T}$	Aggregated (technical) active power losses
$P_{LOSS,NT}$	Aggregated (non-technical) active power losses
L_f	Loss factor
L_F	Load factor
a_{loss}	Coefficient for the LFM
$\widehat{P}_t$	Average active power demand ratio (kW) (LFM)
$\widehat{P}_{L,t}$	Average active power losses ratio (kW) (LFM)
P_t^{max}	maximum active power demand (peak demand) (kW) (LFM)
P_f	Power flow in a distribution component (LFM)
P_{pl}	System peak load (LFM)
$E_{f,y}$	Yearly energy consumption (LFM)
$\widehat{E}_{loss,t}$	Total energy losses in time period t

$\mathbb{Y}$	Output of the supervised machine learning model
$\mathcal{F}_N$	Represents the mapping function of the supervised ML model
R^2	Determination Coefficient
$\sigma^2_{\hat{y},y}$	Covariance between predicted variable $\hat{y}$ and reference y
$\sigma^2_{\hat{y}}$	Variance of predicted variable $\hat{y}$
σ^2_y	Variance of target variable y
ϵ	Prediction error
ϵ_b	Bias error
ϵ_v	Variance error
ϵ_i	Irreducible error
$\tilde{p}^p_{(k,\bullet,t)}$	Estimated power consumption/generation of customer (k) at time (t)
$\mu^p_{(k,\bullet,t)}$	Expected value of power consumption/generation of customer (k) at time (t)
$\sigma^p_{(k,\bullet,t)}$	Dispersion of power consumption/generation of customer (k) at time (t)
$\phi(\bullet)$	Kernel function for KDE method
π_ω	Probability of scenario ω
χ^*_ω	Optimal solution of scenario ω in stochastic optimization
χ^*_d	Optimal solution of the deterministic optimization problem
$d_{\omega,\omega'}$	Distance between each pair of PDFs of scenarios ω and ω'
$\Phi_{\mathbf{u}}$	Uncertainty set for uncertainty vector parameter $\mathbf{u}$
$\overline{\mathbf{u}}$	Expected value for uncertainty vector parameter $\mathbf{u}$
$\bar{u}^p_k$	Phase voltage (complex) of node (k) and phase (p)
$\bar{I}^{pp}_{kj}$	Phase current (complex) between nodes (k) and (j) phase (p)
$\bar{I}^{pp}_{kj}$	Phase current (complex) between nodes (k) and (j) phase (p)
$[\bar{U}^{abc}_k]$	Column vector of complex phase voltages at node (k)
$[\bar{Z}^{abc}_{k,j}]$	Complex impedances matrix between nodes (k) and (j)
$[\bar{Z}^n_{k,j}]$	Complex impedances matrix (phase-neutral) between nodes (k) and (j)
$[\bar{I}^{abc}_{k,j}]$	Complex current column vector between nodes (k) and (j)
$\bar{I}^{nn}_{k,j}$	Complex neutral phase current between nodes (k) and (j)
$[R^{abc}_{i,j}]$	Resistance matrix between nodes (i) and (j)
$[X^{abc}_{i,j}]$	Reactance matrix between nodes (i) and (j)

$[\bar{Y}_{k,k'}^{abc}]$ Admittance complex matrix between nodes (k) and (k')

$g_{k,k'}^{pp'}$ Conductance between phases (p)-(p') and nodes (k)-(k')

$b_{k,k'}^{pp'}$ Susceptance between phases (p)-(p') and nodes (k)-(k')

$\bar{i}_{i,k}^{p}$ Complex current injection at node (k) and phase (p)

$\bar{i}_{k,j}^{p}$ Complex current flow from node (k) to node (j) in phase (p)

$\bar{s}_{i,k}^{p,cal}$ Calculated complex apparent power injection at node (k) in phase (p)

$\bar{s}_{i,k}^{p,sp}$ Specified complex apparent power injection at node (k) in phase (p)

$\bar{s}_{g,k}^{p}$ Specified complex apparent power generation at node (k) in phase (p)

$\bar{s}_{d,k}^{p}$ Specified complex apparent power demand at node (k) in phase (p)

$p_{i,k}^{p,sp}$ Specified active power injection at node (k) in phase (p)

$p_{g,k}^{p}$ Active power generation at node (k) in phase (p)

$p_{d,k}^{p}$ Active power demand at node (k) in phase (p)

$\bar{s}_{k,j}^{p}$ Complex apparent power flow from node (k) to node (j) in phase (p)

$p_{k,j}^{p}$ Active power flow from node (k) to node (j) in phase (p)

$\mathcal{X}$ State vector variable of the unbalanced power flow

$i_{k,j}^{p,re}$ Real part of the current in phase (p) from node (k) to node (j)

$i_{k,j}^{p,im}$ Imaginary part of the current in phase (p) from node (k) to node (j)

$p_{k,PEV,t}^{p}$ Active power demand of PEV unit in node (k) phase (p) at time (t)

C_r Cluster Centroids

$\mathcal{D}(c_r, c_i)$ Euclidean distance between cluster centroids c_r and c_i

$\mathbb{S}(c_i)$ Silhouette of the representative feeder associated to cluster centroid c_i

α_r Average Euclidean distance factor of representative feeder (r)

h Number of hidden layers

n_k Number of neurons of the hidden layer (k)

$\mathbb{X}_{d,r}$ Scaled daily demand RP vector

$\mathbb{X}_{g,r}$ Scaled daily generation RP vector

$p_{g,r}(t)$ Daily generation RP vector of the feeder (r) at time (t)

$p_{d,r}(t)$ Daily demand RP vector of the feeder (r) at time (t)

z_k Signal of the kth neuron

$w_{j,k}$ Edge weight that connects neuron j^{th} and neuron k^{th}

b_k	Bias factor of the kth neuron
o_k	Output of the kth neuron
$\mathfrak{L}$	Loss function for the target DNN
Λ	Accuracy of the supervised machine learning model

List of Figures

List of Tables

Contents

Chapter 1

Introduction

1.1   Introduction

A low voltage (LV) smart grid is essentially a distribution network, which consists of a secondary substation (SS) with a transformer that steps down the voltage from the medium voltage (MV) side (20 kV) to the LV side (400 V); it also has a set of feeders (individual lines) that provide power supply to residential, commercial or industrial customers. The feeders can be underground or aerial. A LV smart grid is characterized by the existence of an advanced metering infrastructure (AMI), which provides a bidirectional channel of communication among smart meters. The data collected by smart meters can be used to improve distribution network operations such as network balancing, DR and energy flexibility [McKenna et al., 2012].

LV distribution networks have evolved to smart grids due to the incorporation of a communication layer, which is composed by the installation of the smart meters and their associated infrastructure - known as "smart metering infrastructure". The vast rollout of these infrastructures that has taken place in many European countries opens new possibilities to gather precise information about the energy consumed by telemetered customers and the power injected from DG units. This information can help to mitigate the obstacles mentioned in [López et al., 2015].

Nonetheless, this deployment has not been completed in some LV distribution areas. At present, it is common to find a high proportion of non-telemetered customers with

contractual power above 15 kW [Nijhuis et al., 2017].

Smart metering infrastructure has become an essential aspect of the successful operation of policies and practices to improve energy efficiency, such as demand response (DR) programs for load flexibility. These infrastructures represent a bidirectional communication channel between consumers who participate in the DR program and thus offer demand flexibility. With this layer of smart communication, the LV distribution network has been transformed into a smart grid. End users can become active users in the grid and provide flexibility through modifying their energy consumption pattern.

The massive rollout of smart meters carried out by DSOs in distribution networks in recent years has contributed to the digitalization of these networks. The global result has been a better understanding of the customers' load demands and an improved network observability. Nonetheless, considerable uncertainty regarding customers' load demands still remain, adding uncertainty in the network operations. This uncertainty is due to several factors, such as smart meters' incorrect measurements; the presence of non-telemetered customers, especially large ones; and energy fraud, among other issues.

Smart grids and distribution networks have a radial topology, with a unique supply point which corresponds to the secondary substation. From the secondary substation, one or more feeders emerge and supply customers in a small area - less than a few kilometers. The topology usually follows the urban layout of the streets in the case of residential and commercial distribution areas, and in most cases, the feeder cables are underground.

In the case of an industrial or rural area, the feeders can adopt the aerial type, which is either fixed in the walls of buildings or are routed through poles [Short, 2004]. This radial topology has important consequences for the operation of the system since it determines the capacity of the network to inject energy from customers to the network. The fact that most networks are underground means that network reconfiguration is not possible for this type of topology.

Low-voltage distribution networks are characterized by an unbalanced operation because the customers' connections are mainly single-phase. These customers correspond with the category of smart meter type 5, which comprises the largest fraction of most LV distribution networks. Table 1.1 indicates the share of each type of smart meter in
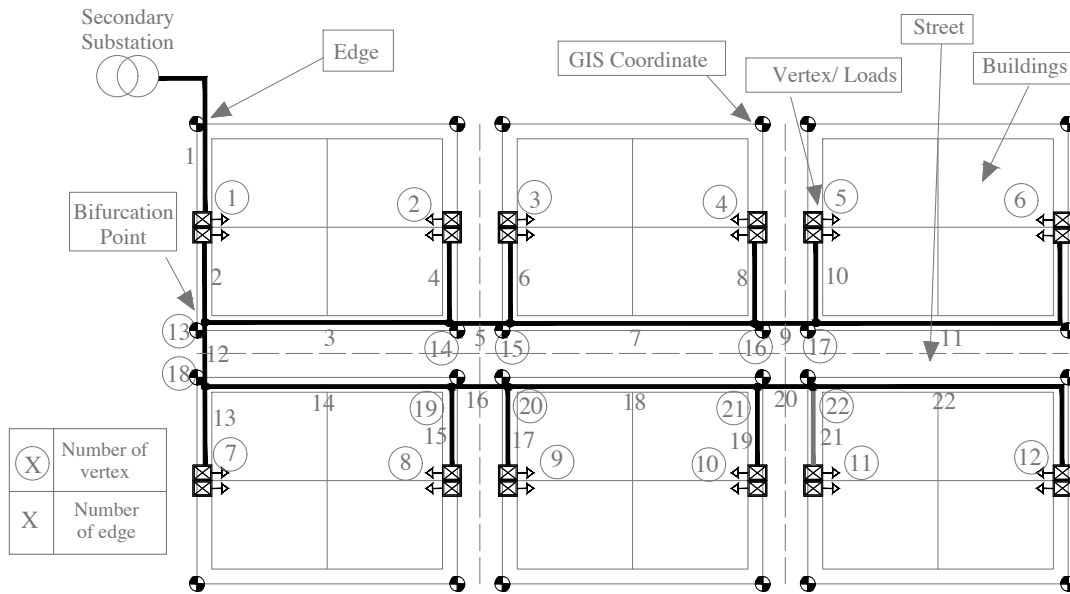
Figure 1.1: Urban layout of an LV smart grid

the networks studied in the OSIRIS project [UC3M et al., 2017] and the classification of smart meters applied in the Spanish LV distribution networks. As customers are mainly single-phase, the power demand is not uniformly spread between the three phases of the network [Short, 2004].

Table 1.1: Smart meter: share by type among networks studied in the OSIRIS project [UC3M et al., 2017]

SM	Power contracted	Share
Type 5	$P_{ctd} \leq 5$ kW	74.21 %
Type 5	$P_{ctd} \leq 15$ kW	95.90 %
Type 4	$P_{ctd} \leq 43.5$ kW	99.35 %

Also the condition is unbalanced since it no possible to balance in real-time the power consumption in each phase neither the power injections coming from distributed energy resources.

Therefore, a total quantity of active power injected in a given network node P can be distributed in each phase, as indicated in (1.5)-(1.7), where x_{eq} indicates the balance between the total power injection between the phases, so that if the balance is $x_{eq} =$

1/3, the total power injected is equally divided among the three phases and the operation results in a balanced condition.

However, if the balance results in a different quantity, $x_{eq} \neq 1/3$, the operation results in an unbalanced condition's and here, so the power total injected is not equally distributed among phases. The unbalanced operation has consequences in terms of the power flow dispatch between phases and the resulting power losses.

We consider a three-wire model for a single-feeder LV distribution network, illustrated in Fig. 1.2. Three different loads are connected, one to each phase of the network. The total power injected in each phase is modified from 0 to 1, in steps of 0.1. For each step, the power flow equations were calculated to obtain the power loss for each phase. A total power injection (consumption) of 100 kW with a unitary power factor was considered for a 1000-m single-feeder line. Nominal line-to-line voltage was fixed at 400 V and used as the reference voltage, and a reference power of 100 kVA was used. An aluminum underground conductor with a cross-section of 240 mm^2 (Table 1.3 and Table 1.2) was considered for the phases, and a conductor of 150 mm^2 was considered for the neutral cable.

Table 1.2:   Underground impedance reference values (cable Al XZ1 0.6/1kV) [Union Fenosa Distribucion, 2011b]

Cross cable section S (mm^2)	**Resistance** r $20°C$ (Ω/km)	**Resistance** r $90°C$ (Ω/km)	**Reactance** x (Ω/km)
50	0.641	0.821	0.093
95	0.320	0.410	0.083
150 (AS)	0.206	0.264	0.099
240	0.125	0.160	0.079
240 (AS)	0.125	0.160	0.093

To consider Cooper (Cu) based conductors, the resistance can be calculated using the electrical resistivity and the cross section as follow:

$$r_{Cu,90°} = \rho_{Cu,90°} \cdot S \qquad (1.1)$$

Table 1.3: Aerial impedance reference values (cable Al RZ 0.6/1kV) [Union Fenosa Distribucion, 2011b]

Cross cable section (S)	Resistance $r^{50°C}$	Reactance x
50 mm^2	0.718 Ω/km	0.1 Ω/km
95 mm^2	0.359 Ω/km	0.1 Ω/km
150 mm^2	0.231 Ω/km	0.1 Ω/km
240 mm^2	-	0.1 Ω/km

Where:

$$\rho_{Cu,90°} = \rho_{Cu,20°} \cdot (1 + \alpha_{Cu} \cdot (90° - 20°)) \tag{1.2}$$

$$\rho_{Cu,20°} = 1/58 \quad \Omega mm^2/m \tag{1.3}$$

$$\alpha_{Cu} = 0.00393 \tag{1.4}$$

The results are indicated in Fig. 1.3, where in the upper plot, is indicated the load fraction (in per unit) assigned to each phase for each balanced faction x_{eq}. Note that for $x_{eq} = 0.5$ the operation of the single-feeder is balanced and the three phases have the same power assignation. In the middle plot is shown the total power losses produced in each phase for each balance fraction considered. Note that the extreme situations $x_{eq} = 0$ and $x_{eq} = 1$ leads to the maximum power losses in phases b and c respectively (since phase a is considered always 1/3 of the total amount of load to be distributed). Finally, in the lower plot is hosted the total power losses of the network which include the power losses associated to the power flow in the three phases as well as the power flowing in the neutral cable. It can be noted that for the equilibrium point $x_{eq} = 0.5$, exists a minimum of the total power losses. This is motivated by the fact that if the system is balanced, the current flowing in the phases results in a lower magnitude compared to unbalanced condition. This change affects dramatically to the power losses values since the calculation is proportional to the square of the current.
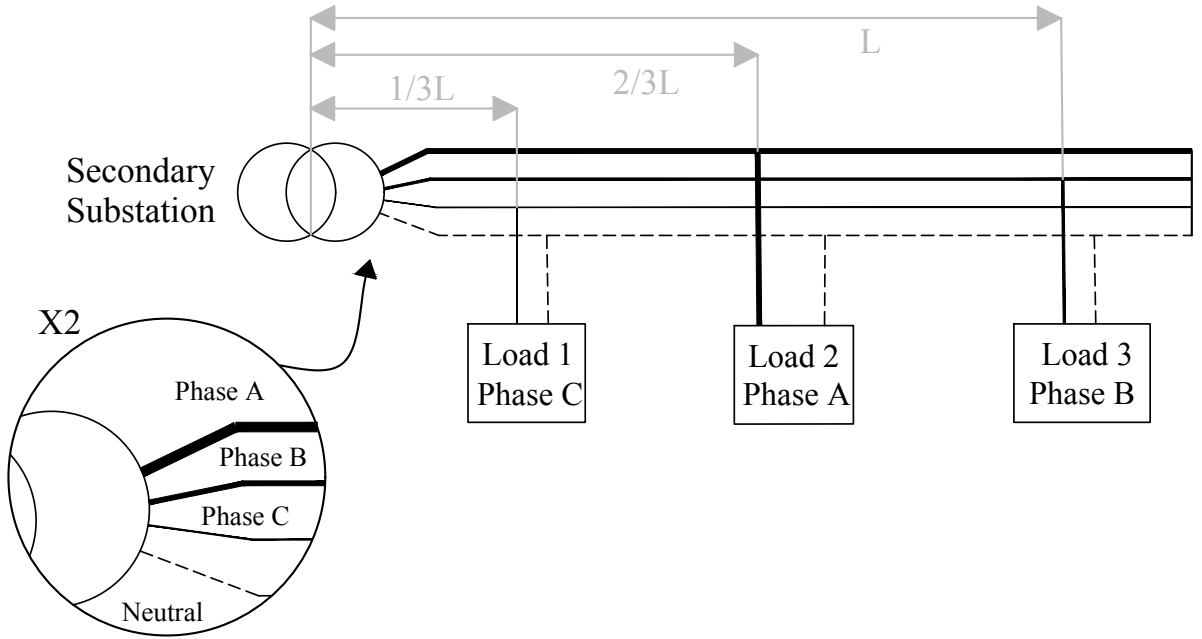
Figure 1.2: Four-wires model in a simple-feeder

$$p_i^a = p_g^a - p_d^a = P \cdot 1/3 \tag{1.5}$$

$$p_i^b = p_g^b - p_d^b = P \cdot 2/3 \cdot (1 - x_{eq}); \quad x_{eq} \in (0,1) \tag{1.6}$$

$$p_i^c = p_g^c - p_d^c = P \cdot 2/3 \cdot (x_{eq}); \quad x_{eq} \in (0,1) \tag{1.7}$$

1.2   Motivation

In recent years, LV distribution systems have experienced a vast rollout and deployment of smart metering systems and increased integration of distributed energy resources (DERs) such as Photovoltaic (PV) panels, plug-in electric vehicles (PEV) and battery energy systems (BES) into the network. These aspects pose a challenge for the operation of the system due to the increase of power losses and the possible technical contingencies, such as overloading and over-voltages, if DERs are not properly allocated and integrated. Moreover, there is much uncertainty associated with load demand variation of the power generation from DERS. Nonetheless, environmental concerns require DSOs and regulators
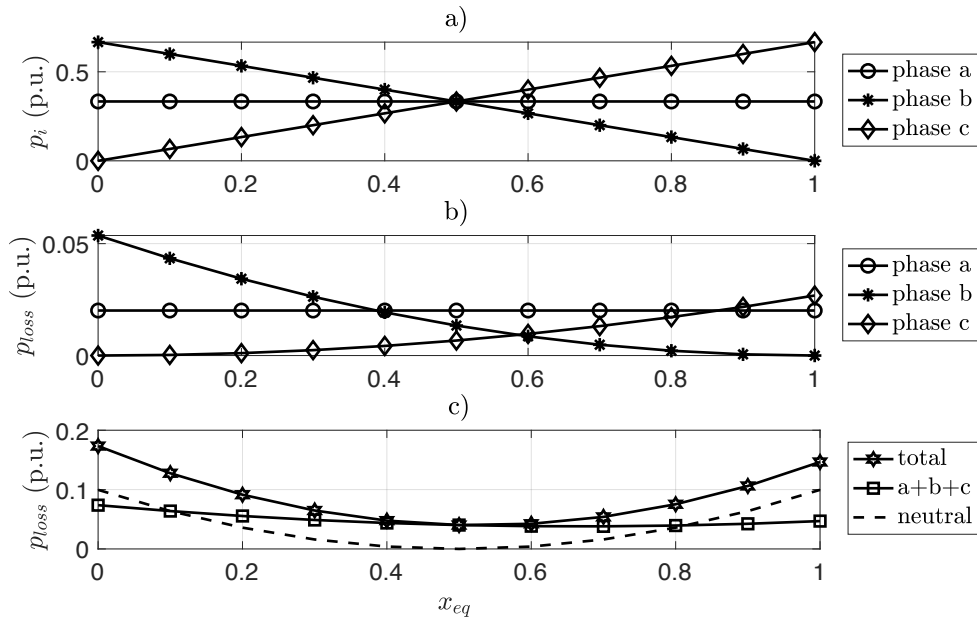
Figure 1.3: Unbalance operation of a four-wires model in a simple-feeder

to provide maximal energy efficiency in the network, with the least possible power loss values, as well as maximal sustainable energy consumption.

This produces a trade-off, because if renewable-based resources are not properly integrated, their operation can increase the power losses. Therefore, DSOs require advanced methods to estimate the power losses levels in large-scale LV smart grids under uncertain and unbalanced situations. Such information can facilitate the deployment of policies and practices that would lead to the safe and efficient integration of DERs in the form of flexibility mechanisms.

1.3 Objectives

This thesis contributes to the field of power loss estimation and optimization for LV smart grids in large-scale distribution areas. The main objectives of the thesis are the following:

1. Power losses estimation in unbalanced LV smart grids under uncertainty.

2. Power losses estimation in unbalanced LV smart grids in large areas with a presence of DERs.

3. Flexibility scheduling for power losses minimization in unbalanced smart grids under uncertainty.

The above objectives can be achieved by taking advantage of smart metering infrastructures, machine learning models and mathematical programming. The fulfillment of such objectives improves the knowledge about the energy lost in distribution networks. This information also provides DSOs with operational tools to manage their power losses in the whole distribution network through flexibility mechanisms.

1.4   Thesis Organisation

This document is organised in ten chapters as follows:

- Chapter 1 correspond with this introductory chapter where a background to the research study is presented as well as thesis motivation, objectives and organisation.

- Chapter 2 presents a succinct literature review of the unbalanced low-voltage smart grids topic tackled, taking particular care of the power losses estimation and the power losses optimisation. The chapter include the state-of-the-art in power losses estimation in unbalanced low-voltage smart grids as well as the optimization techniques used in such networks.

- Chapter 3 presents the power losses estimation for unbalance low-voltage smart grids, including load demand uncertainty characterisation, smart meter data handling (smart metering data infrastructure, issues related to the data accuracy, quality and resolution) as well as non-telemetered customers. Additionally presents an stochastic power losses analysis for non-telemetered customers based on a Markov process model.

- Chapter 4 presents a two-stage deep learning-based power losses estimation model for large-scale unbalanced low-voltage smart grids under uncertainty and with presence of distributed generation. A case study with numerical results is provided, including a clustering process to select the most representative feeders from the distribution area.

- Chapter 5 presents a robust flexibility scheduling model for unbalanced smart grids under uncertainty. The model addressed the load flexibility scheduling considering demand response as a mechanism in the presence of distributed energy resources.

- Chapter 6 brings the thesis to a close summarising the overall conclusions and key findings of the research conducted, to outline the direction for further research work.

Chapter 2

Literature Review

2.1 Introduction

Traditionally, low-voltage distribution system has been operated with radial topologies with a unique supply point (the SS), consequently, the power flow direction is defined from the SS to each customer connection point. This operation mode makes that the energy flow in one direction, in what it is known as the passive mode of a network. In that situation, in a simple low-voltage feeder (Fig. 2.1) with smart meters at the customer's installation as well as at the SS (smart meter defined as supervisor) it could be expected that active losses at certain time period in a low-voltage feeder (P_{LOSS}) can be calculated by adding up all active load demand imported by the customers (P_D) as indicated in (2.1) where (Ω_C) is the set of customers connected to a feeder and connected to the different phases $\{a, b, c\}$, and subtract them from the imported active power measured at the SS smart meter (P_{IMP}^S) as indicated in (2.2).

Although the power losses of a specific feeder represent a low quantity (in the absence of NTLs and massive presence of DERs) it may vary around 3% of the power delivered to customers, when it comes to the power losses of a large-scale distribution area (>1000 feeders) the quantity of energy lost becomes a considerable amount in terms of energy efficiency. Is indeed in this situation when it becomes interesting the power losses estimation, because a better knowledge of how and where the energy is lost, contributes to a better deployment of polices and practices that lead to higher rated energy efficiency

[Usman et al., 2018]. Regarding this, a large distribution area could easily be composed of hundreds of secondary substations and thousands of distribution feeders. Each feeder belonging to that area has different topological characteristics (such as cross section, length, overhead/underground configuration) as well as different levels of DG presence, smart meters coverage, voltage unbalance as well as load demand.

$$P_D = \sum_{k \in \Omega_C} \sum_{p \in \{a,b,c\}} P_d^{k,p} \tag{2.1}$$

$$P_{LOSS} = P_{IMP}^S - P_D \tag{2.2}$$
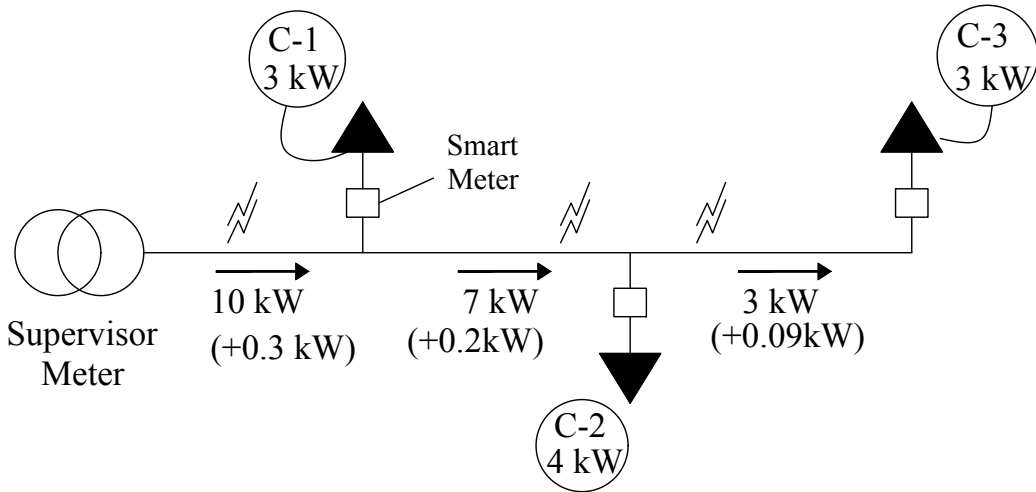
Figure 2.1: Low-voltage feeder illustration

In terms of energy, the energy losses during a time period (τ) can also be calculated as the difference between the energy imported to the network in the period of time $(E_{D,\tau}^S)$ and the aggregation of the energy consumed by all of the customers $k \in \Omega_C$ supplied in the same time $(E_{D,\tau}^k)$.

$$E_{LOSS,\tau} = E_{IMP,\tau} - E_{D,\tau} = E_{IMP,\tau} - \sum_{k \in \Omega_C} E_{D,\tau}^k \tag{2.3}$$

In this context, total system losses indicate how effectively and efficiently a distribution system is delivering power to those customers.

Therefore, at first sight, total system losses are the difference between the delivered power to the system and the power consumed by customers. This approach leads to calculate the balance between the energy delivered by the SS and the total energy registered by customers smart meters.

It can be seen that smart meters data resolution and its mapping to the grid topology constitutes a great influence in the accuracy of the power losses calculation. Feeder monitoring produces high value information for low-voltage grid operation and maintenance providing online updated network topology.

In Fig. 2.1 is described that in normal operating conditions (14:00 PM), customers 1 and 3 (C-1 and C-3, respectively) are demanding 3 kW and customer 2 (C-2) is demanding 4 kW (with unitary power factor all of them). In this situation the power supplied by the SS is equal to the aggregated active power consumption of customers (10 kW) plus the associated technical active power losses (estimated in 0.3 kW). This power losses are associated to the joule effect and are calculated for each feeder section connecting each pair of nodes i-j by means of the square of the current magnitude flowing in that feeder section multiplied by the electrical resistance (Ω/m) and the length of the feeder section (m) as indicates in (2.4). However the calculation of this expression requires to solve the power flow problem which requires to have detailed information about network topology, load demand and DG units.

$$P_{LOSS} = \sum_{\forall i,j, i \neq j} i_{i,j}^2 \cdot r_{i,j} \cdot \ell_{i,j} \tag{2.4}$$

where:

- $i_{i,j}$ (A) is the the current flowing through the feeder conductor that connects nodes i and j.

- $r_{i,j}$ (Ω/m) is the linear resistance of the feeder conductor. Depends on the material ρ (Ω/mm^2) and the cross section area S (mm^2) by the relation $r = \rho/S$.

- $\ell_{i,j}$ (m) is the length of the feeder conductor.

The total active power losses of distribution network can be calculated as the aggregation of the power losses of all the feeders connected to that SS. Similarity, the active power losses of a set of low-voltage distribution network distributed across a large-scale distribution area is the aggregation of the losses of each single SS. However, this intuitive scheme is only valid under the following situations:

1. All the customers are monitored through the smart metering infrastructure.

2. There is no presence of DER devices (such as PV panels).

3. There is no presence of electricity theft (no illegal connections and no manipulated metering devices).

4. There are no loops, i.e.pure radial distribution network.

Power losses in a three-phase LV distribution network can be calculated as the aggregation of the three phases power losses produced in each feeder phase as states (2.5).

$$P_{LOSS} = \sum_{p \in \{a,b,c\}} P_{LOSS}^p = \sum_{p \in \{a,b,c\}} \sum_{\forall i,j, i \neq j} (i_{i,j}^p)^2 \cdot r_{i,j}^p \cdot \ell_{i,j}^p \tag{2.5}$$

where:

- $i_{i,j}^p$ (A) is the phase current magnitude

- $\ell_{i,j}^p$ (m) is the length of the phase feeder section.

- $r_{i,j}^p$ (Ω/m) is the longitudinal phase resistance of the feeder that connects the nodes i and j.

This power losses magnitude of power losses can be expressed as a fraction of demand, S_D (2.6)-(2.9), using the loss factor $\mathcal{L}_p$, which results in expression (2.9).

$$S_D = \sqrt{P_D{}^2 + Q_D{}^2} = \sqrt{P_D{}^2 + \sin(\arccos(\varphi))^2} \tag{2.6}$$

$$P_D = \sum_{p \in \{a,b,c\}} U_N^p I_T^p \cos(\varphi_p) \tag{2.7}$$

$$Q_D = \sum_{p \in \{a,b,c\}} U_N^p I_T^p \sin(\varphi_p) \tag{2.8}$$

$$\mathcal{L}_p = \frac{P_{LOSS}}{S_D} \tag{2.9}$$

Where I_T^p is the total secondary substation phase current, $\cos(\varphi_p)$ is the power factor of the phase p and U_N^p is the secondary nominal phase voltage in the secondary substation. The term $\mathcal{L}_p$ varies with the load demand of the network (affected by the distributed generation), and under some equilibrium circumstances ($x_{eq}^{min} \leq x_{eq} \leq x_{eq}^{max}$) (1.5)-(1.7) behaves linearly with the average phase current magnitude I_T^p, (2.10) where the parameter K_{LOSS} can be inferred for different demand/generation conditions. Note that, even if $\mathcal{L}_p$ behaves linearly, it depends on the variation of the total phase current magnitude, which depends non-linearly of the load demand and generation of the customers and distributed resources.

$$\mathcal{L}_p = \frac{\sum_{p \in \{a,b,c\}} \sum_{\forall i,j, i \neq j} i_{i,j}^{p\ 2} \cdot r_{i,j}^p \cdot \ell_{i,j}^p}{\sqrt{\left(\sum_{p \in \{a,b,c\}} U_N^p I_T^p \cos \varphi_p\right)^2 + \left(\sum_{p \in \{a,b,c\}} U_N^p I_T^p \sin \varphi_p\right)^2}} \propto K_{LOSS}(S_D) \cdot I_T^p \tag{2.10}$$

In actual LV distribution networks, the majority of residential and small-size commercial customers (contractual power less or equal to 15 kW) are telemetered but there are still some commercial and industrial customers (with contractual power from 15 kW to 50 kW) which are only required to provide energy consumption in monthly basis [López et al., 2015]. Consequently, the power demand measured at the secondary substation meter corresponds mainly to the aggregated power demand of commercial and industrial customers and in a minor percentage of the aggregated demand by telemetered residential customers, i.e. $P_{ctd} \leq 15$ kW (Table 2.1).

In this way, it becomes clear that the simple assumption of adding up all demand smart meters measurements and subtract them from the injecting power measured at the secondary substation's meter is unacceptable for losses estimation in low-voltage networks, if there are non-telemetered customers. This assumption is neither valid if there are distributed generation units in the low-voltage customer installations. Due to the existence of the customers that do not have an smart meter that reports power change with

Table 2.1: Smart Meter types in terms of measurement precision and phase connection

SM	Power Contracted	Precision (P)	Precision (Q)	Connection
Type 5	$P_{ctd} \leq 15$ kW	A ($\pm 2\%$)	3 ($\pm 1\%$)	SP
Type 4	15 kW $< P_{ctd} \leq 43{,}5$ kW	B ($\pm 1\%$)	2 ($\pm 2\%$)	3P
Type 3	$P_{ctd} > 43{,}5$ kW	B ($\pm 1\%$)	2 ($\pm 2\%$)	3P

SF: Single-Phase, 3P: Three-Phase, SM: Smart Meter Type [Union Fenosa Distribucion, 2011b, Union Fenosa Distribucion, 2011a].

a high resolution and so it is necessary to split customers demand in demand telemetered ($\widetilde{P}_D$) and estimated demand ($\widehat{P}_D$).

$$P_D = \widehat{P}_D + \widetilde{P}_D \tag{2.11}$$

In presence of DER devices such as PV panels (P_{PV}), as well as other PEVs or BES devices, the direction of the power flow could be either from the SS point to the customers connection point or from the customers to the SS point, and so it can exist power exported to the MV network. Power generation in LV smart grids are mainly associated with PV array panels. Note that power injections from PVs, power exchanges from BES units and power consumption from PEVs are denoted with wide tilde ($\tilde{}$) due to the fact that are uncertain quantities that have to be estimated.

Let's supposes that in the low-voltage feeder showed in Fig. 2.1 all the three customers are telemetered and additionally DG units are installed and connected to the customer connection point as shown in Fig. 2.2.

In this new operating conditions customers C-1 and C-3 are demanding 3 kW and have DG units (PV panels) that are generating 4 kW, so exists a net power injection to the network of 2 kW. Moreover, the customer C-2 is demanding 4 kW and have a PV facility injecting 2 kW. In this situation customer 2 absorb the excess of power of customers 1 and 3, producing losses in the network due to the power flow in the segments of line that connect customers C-1 and C-3 with C-2. However, in those circumstances the supervisor meter, located at the secondary substation, does not measure any power flow at the SS (consumption or generation) and so the losses estimation procedure based
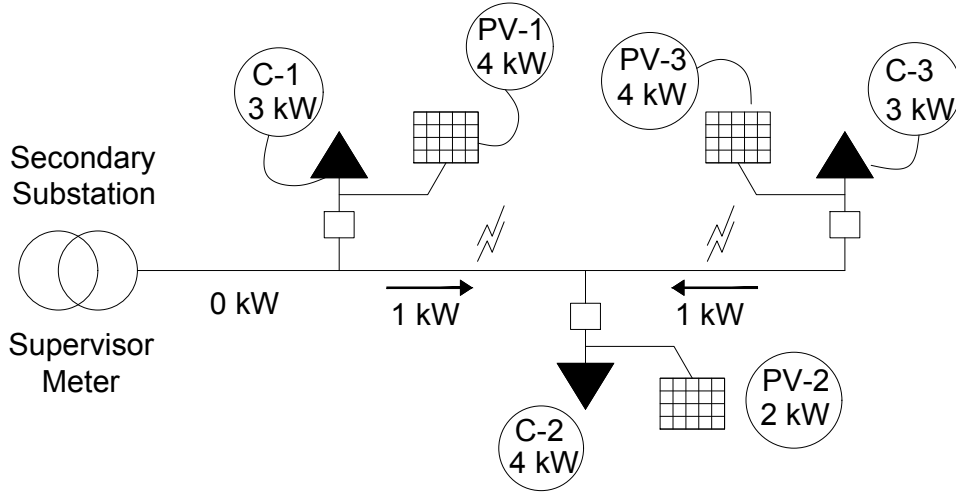
Figure 2.2: Low-voltage simple feeder with DER devices

on energy balance (2.2) cannot be applied straightforward. Moreover, in this situation the DSO is not capable to known in which segments of the network is taking place the loss of power. Another inconvenient associated with massive injection of power generation coming from distributed sources in low-voltage feeders, is that, if the power generated by DG units are higher than the aggregated customers power demand, then, the supervisor meter measures an exported power to the MV network and again the approach (2.2) fails. So, with penetration of DERS, the energy balance to be applied is the indicated in (2.12).

$$P_{IMP} - P_{EXP} = P_{LOSS} + P_D - P_{PV} \tag{2.12}$$

Where the imported power P_{IMP} and exported P_{EXP} are measured by the smart meter supervisor located at the SS. Total power demand by customers P_D is composed by the aggregated power consumption measured by the customers smart meters $\widehat{P}_D$ and by the estimated $\widetilde{P}_D$ power demand (2.11). P_{PV} is the power injected from DG units (usually PV panels). It has to be noted that power generation from PV units depends on solar irradiation and ambient temperature which adds uncertainty in the estimation of PV generation.

Equation (2.2) has to be extended if there are Plug-in Electric Vehicles (PEVs) and/or Battery Energy Storage (BES) units connected to the grid. In this case, power consumption at charging stations $\widetilde{P}_{PEV}$ have to be added to the power balance equation as (2.13).

Similarly, power consumption from PEV behaves in a similar way although demand less power.

$$P_{IMP} - P_{EXP} = P_{LOSS} + P_D - \widetilde{P}_{PV} + \widetilde{P}_{PEV} \tag{2.13}$$

In some cases, there are BES units along with the solar PV units to absorb energy from PV units (charging) and inject energy during discharge. Consequently the power balance equation is (2.14).

$$P_{IMP} - P_{EXP} = P_{LOSS} + (P_D + \widetilde{P}_{BES}^{+}) - (\widetilde{P}_{PV} + \widetilde{P}_{BES}^{-}) + \tilde{P}_{PEV} \tag{2.14}$$

PEV devices normally demand power directly from the grid, but other schemes are possible, as shown in (2.15).

$$P_{IMP} - P_{EXP} = P_{LOSS} + (P_D) - (\widetilde{P}_{PV} + \widetilde{P}_{BES}^{-}) + (\widetilde{P}_{PEV} + \widetilde{P}_{BES}^{+}) \tag{2.15}$$

2.2   Technical and Non-Technical Losses

Total active power losses are composed by Technical Losses (TLs) $P_{LOSS,T}$ which are produced by joule effect, due to the normal operation of power systems and by Non-Technical Losses (NTLs) $p_{LOSS,NT}$ witch are associated with unidentified losses due to the unauthorized customers or undesirable operating condition (electricity theft, wrong measurements, unauthorised use of supply/illegal connections and meter tampering or meter bypass). In Fig. 2.3 it is shown a low-voltage feeder with metered and non-metered customers, technical and non-technical power losses as well as the presence of DG units.
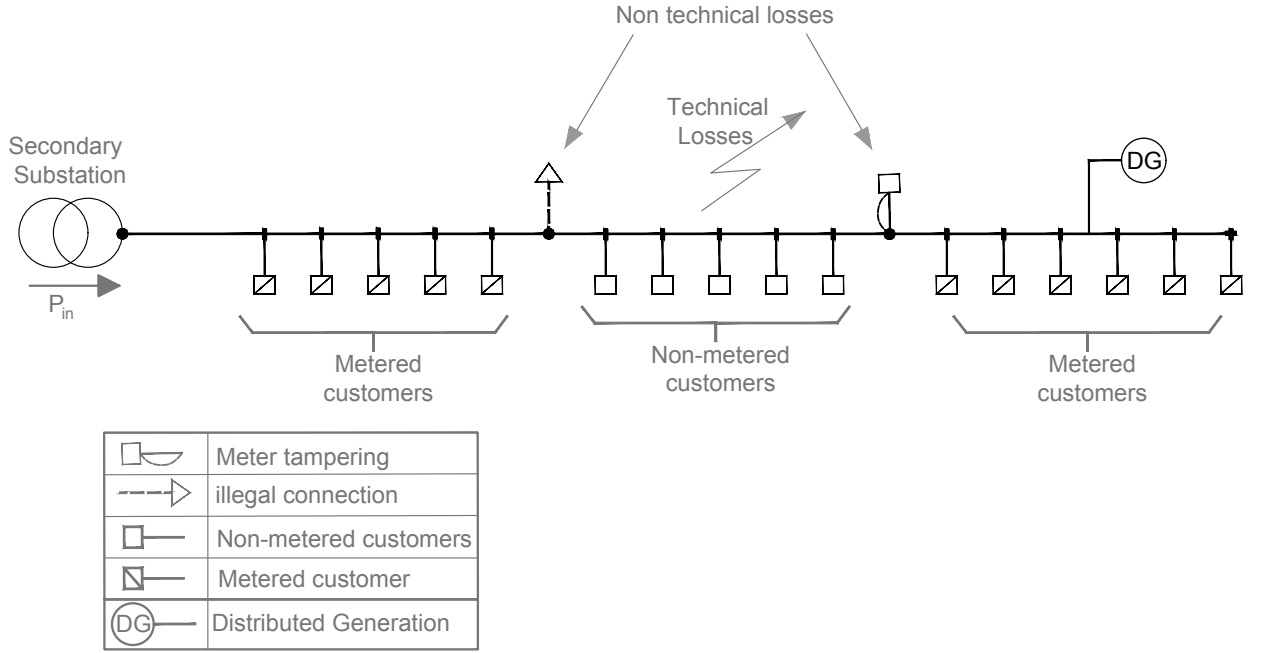
Figure 2.3: Low-voltage feeder with non-technical power losses

$$P_{LOSS} = P_{LOSS,T} + P_{LOSS,NT} \tag{2.16}$$

TLs are produced by the current flow through the cables, and so it depends on the load demand and power injections to the grid; the grid topology (radial or weakly meshed), the feeders physical characteristics (conductors material, cross section) and also of the calculated power flow equation. It depends also on the electrical connection point of customers and DER units to the grid. TLs can be calculated from the power flow calculation [Ljungberg, 2006].

$$P_{LOSS,T} = f(P_D, \widetilde{P}_{PV}, \widetilde{P}_{PEV}, \widetilde{P}_{BES}, \mathcal{G}) \tag{2.17}$$

Where $\mathcal{G}$ is the graph that represents the LV smart grid topology: number of vertices (buses) and edges (lines), spatial distribution of customers and DERs (phase connection), edge weight (conductor material and cross section).

NTLs are extremely difficult to evaluate due to the nature of their unexpected source [Pedro, 2009] consequently they are estimated as the difference between the measured total power losses and the calculated TLs. The accuracy of its estimation depends on the

accuracy of the smart meters measurements, the metering systems state as well as the confidence estimation of the uncertain parameters that affect TLs.

$$P_{LOSS,NT} = \overbrace{P_{LOSS}}^{TOTAL} - \underbrace{P_{LOSS,T}}_{TLs} = \tag{2.18}$$

$$= \underbrace{\overbrace{P_{IMP} - P_{EXP} - \widehat{P}_D}^{TOTAL\ LOSSES} - \underbrace{\widetilde{P}_D + \widetilde{P}_{PV} - \widetilde{P}_{PEV} - \widetilde{P}_{BES}}_{ESTIMATED} - \underbrace{\overbrace{f(\widehat{P}_D, \widetilde{P}_D, \widehat{P}_{PV}, \widehat{P}_{PEV}, \widehat{P}_{BES}, \mathcal{G})}^{TLs}}_{CALCULATED}}_{MEASURED}$$

As it can be seen in (2.18) NTLs calculation requires accurate and reliable information of the energy flow through network lines and entails a challenge for grid operators:

1. It is necessary to estimate the total power losses which are composed by TLs due to the consumption and generation of known customers and DER units, and NTLs due to the non-metered demand and uncertain injection coming from DERs (PV, BES, PEV).

2. The calculation of TLs is not deterministic since it depends on uncertain quantities such as the non-metered demand $\widetilde{P}_D$, the PV injections $\widetilde{P}_{PV}$, contribution of BES $\widetilde{P}_{BES}$ and the PEV consumption $\widetilde{P}_{PEV}$.

$$P_{LOSS} = \underbrace{\overbrace{P_{IMP} - P_{EXP} - \widehat{P}_D}^{BALANCE} - \underbrace{\widetilde{P}_D + \widetilde{P}_{PV} - \widetilde{P}_{PEV} - \widetilde{P}_{BES}}_{ESTIMATED}}_{MEASURED} = \tag{2.19}$$

$$= \underbrace{\overbrace{f(\widehat{P}_D, \widetilde{P}_D, \widehat{P}_{PV}, \widehat{P}_{PEV}, \widehat{P}_{BES}, \mathcal{G})}^{TLs}}_{CALCULATED}$$

Therefore, according to 2.19, power losses in an unbalanced low-voltage smart grid with presence of DERs such as PV, BES or PEV devices can be calculated by two approaches. Firstly, power losses can be computed by the energy balance between the measurements collected by the smart metering infrastructure and the estimated power injections (2.19). The second one, is by solving the unbalanced power flow which also requires to estimate the uncertain power injections mentioned.

Intuitively those characteristics are highly correlated with the power losses and so it is possible to discern groups or cluster of feeders that share common pattern and are statistically similar regarding those characteristics. Moreover, within each cluster of feeders it is possible to find representative feeders of the cluster, from which power losses can be inferred and consequently extrapolated to the remanning feeders belonging to the group which represent.

2.3 Power Losses Estimation Methods

2.3.1 Methods Based on Power Flow & Analytical Equations

The Empirical Loss Factor Method (ELFM) used by most utility engineers to calculate energy loss was developed in [Buller and Woodrow, 1928]. Loss Factor Method (LFM) can be defined as a quantity that, when multiplied by the peak losses, would yield an average loss. This average loss, when multiplied by the hours in the time period being analysed, would yield the energy losses.

Equation (2.20) define the ELFM and is used to determine loss factor (L_f), requiring inputs of load factor (L_F) and a variable coefficient (a_{loss}), which would normally vary between the range $a_{loss} \in (0.15, 0.3)$.

$$L_f = a \cdot L_F + (1 - a_{loss}) \cdot L_F{}^2 \qquad (2.20)$$

There are many inadequacies in using ELFM to determine losses, because this method is intended to be used only during peak losses time, and during other specific time periods it cannot be used. Variable coefficient a_{loss} depends on the load shape. However, it is difficult to know how this coefficient could be determined for any specific type of application. Traditionally this coefficient has been developed by each utility over years, most likely with no documented technical analysis. [Gustafson, 1983] presented a way to calculate the loss factor L_f for the specific load on a system. The results of the method proposed LFM can be used to verify the variable coefficient a_{loss}.

Load factor L_F is defined as the ratio of the average active power demand $\widehat{P}_t$ (kW)

supplied during a designated period t (h) to the peak or maximum active power demand P_t^{max} (kW) occurring in that period. Load factor also may be derived by dividing the active energy demanded (kWh) in the period by the product of maximum active power demand (kW) and the number of hours in the period, equation (2.21).

$$L_F = \frac{\widehat{P}_t}{P_t^{max}} = \frac{\int_0^t p(t)dt}{t \cdot P_t^{max}} \tag{2.21}$$

Taking into account the previous definition, Loss Factor L_f can be defined as the ratio of the average active power losses $\widehat{P}_{L,t}$ (kW) in a period of time t to the peak active power losses $\widehat{P}_{L,t}^{max}$ (kW) in that period. Also, L_f can be defined as the sum of the active power demand squared and divided by the product of the hours in the time period and the maximum active power demand, equation (2.22).

$$L_f = \frac{\widehat{P}_{L,t}}{\widehat{P}_{L,t}^{max}} = \frac{\int_0^t p(t)^2 dt}{t \cdot P_t^{max}} \tag{2.22}$$

However LFM is only suitable to calculate energy losses, while marginal losses and no-load losses can not be calculated with the method. Energy losses can be calculated by using Percent Loading (PrL) on the different components of the distribution system. PrL is obtained by dividing the power flowing through a distribution component (P_f) by the power capacity of that component (S_n), equation (2.23).

$$PrL = \frac{P_f}{S_n} \tag{2.23}$$

This PrL along with demand data is used to find the losses. In PLM a random sampling plan is used. [Flaten, 1988] proposed PLM to estimate losses. Peak Losses Energy (PLE) is defined as yearly energy generation, and demand losses are defined as Peak Losses Demand (PLD). So PrL is defined for each kind of losses in (2.24) and (2.25).

$$PLE = \frac{E_{f,y}/T_h}{S_n} \tag{2.24}$$

$$PLD = \frac{P_{pl}}{S_n} \tag{2.25}$$

where,

- $E_{f,y}$ The yearly energy consumption per type of distribution equipment (MWh).

- P_{pl} The system peak load per type of distribution equipment (MW).

- T_h Number of hours of the year (8760).

Equations (2.24) and (2.25) are calculated for each distribution system component (transformers, lines, etc.). Energy losses in the period of time t in terms of PrL can be written as:

$$\widehat{E}_{loss,t} = (PrL)^2 \cdot LL \tag{2.26}$$

where

- PrL Percent Load, either, PLD or PLE.

- LL Distribution component rated active power loss at maximum current capacity (kW).

Equation (2.26) can be applied to each component of the system in combination with other correction factors such as: power factor, temperature correction or load profile correction. The limitations of the study are in the arduous random sampling conductor loads process, in the accuracy of historical transformer loss data as well as neutral conductor losses.

2.3.2 Machine Learning-based methods

Machine learning algorithms based on supervised learning, traditionally used for regression problems, can be applied for power losses estimation [Geron, 2017]. Supervised machine learning models are based on mathematical algorithms that train a model to learn from a set of training data. As a result they are affected by bias and variance [Harrington, 2012]. The goal of any supervised machine learning algorithm is to best estimate the mapping function ($\mathcal{F}$) for the output variable $\mathbb{Y}$, which in the case of this thesis is the power losses estimation given the input data matrix $\mathbb{X}$, which in this case is the set of features. The mapping function is often called the target function because it is the function that a given

supervised machine learning algorithm aims to approximate [Deisenroth et al., 2020]. Features based on the network operation data can include the shown in Table 2.2.

$$\mathbb{Y} = \mathcal{F}_N(\mathbb{X}) \tag{2.27}$$

Table 2.2: Features based on network operation for power losses estimation

Source	Feature
Smart Meters	Historical power consumption from smart meters
Smart Meters	Smart Meters penetration (ratio of non-telemetered customers)
Smart Meters	Customer consumption pattern
Network	Area of the Network (Residential, Industrial, Commercial)
Network	Network Topology characteristics
Network	Energy tariff's (influence in consumption patterns)
Network	Contractual customer power (influence in consumption magnitude)
Network	DG penetration and DERs presence
External	Calendar days
External	Weather data (temperature) and meteorological special events

Calendar days refer to: day of the week, holidays, non-working days, special days and events; Topology characteristics refers to: feeder length, number of nodes and number of ramifications

Based on a historical data set, related to the network for which the power losses estimation is desired, an appropriate definition of the features to be used can lead to very accurate results without the need of performing traditional power flow calculations. To achieve that is essential to manage the bias-variance trade-off, since the machine learning model will learn from the data related to the operation of the unbalanced smart grid in terms of power consumption, distributed generation, MV/LV power exchanges at secondary substations, voltage unbalanced, etc. If this network operation data set is not correctly defined the power losses predictions performed will lead to unfeasible and unrealistic results.

The following supervised machine learning models [Murphy, 2012] for regression tasks can be considered as well-suited for the task of power losses estimation:

☐ Linear regression models.

- Polynomial models [Chang et al., 2010].

- Ordinary Least Squares (OLS) [Goldberger, 1964].

- Weighted Least Squares (WLS) [Mandel, 1990].

- Generalised Linear Models (GLM) [Nelder and Wedderburn, 1972].

- Generalised Additive Models (GAM) [Hastie, 1990].

- RIDGE regression [Hoerl, 1970].

- LASSO (*Least Absolute Shrinkage and Selection Operator*) [Tibshirani, 1996].

- ELASTIC-NET regression [Zou and Hastie, 2005].

- Linear Discriminant Analysis (LDA) [McLachlan, 2004].

- Bayesian Linear regression [Goldstein and Wooff, 2007].

☐ Statistical-based models.

- Logistic Regression [Hosmer and Lemeshow, 2000].

- Poisson Regression [Greene, 2003].

☐ Support Vector Machines (SVM) Regressor [Cortes and Vapnik, 1995].

☐ K-Nearest Neighbours (KNN) Regressor [Altman, 1992].

☐ Artificial Neural Networks (ANN) [Simon S, 1999].

☐ Decision Trees (DTs) [Quinlan, 1986].

☐ Ensembles methods.

- Random Forest [Breiman, 2001].

- Gradient-Boosted Decision Trees (GBDT) [Hastie and Friedman, 1991].

2.3.2.1 Model Performance

The following performance metrics are widely used in predictive modelling (regression and classification tasks) within machine learning problems [Hastie and Friedman, 1991]:

Determination Coefficient (R^2)

Given by equation (2.28) it provides a measure of how well predicted outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model.

$$R^2 = \frac{\sigma_{\widehat{y},y}^2}{\sigma_{\widehat{y}}^2 \cdot \sigma_y^2} = \frac{n\left(\sum_i y_i \cdot \widehat{y}_i\right) - \sum y_i \cdot \sum_i \widehat{y}_i}{\sqrt{n\sum_i y_i^2 - \left(\sum_i y_i\right)^2} \cdot \sqrt{n\sum_i \widehat{y}_i^2 - \left(\sum_i \widehat{y}_i\right)^2}} \qquad (2.28)$$

Where $\sigma_{\widehat{y},y}^2$ is the covariance between the prediction $\widehat{y} = (\widehat{y}_1, \ldots, \widehat{y}_n)$ and the target variable $y = (y_1, \ldots, y_n)$, σ_y^2 and $\sigma_{\widehat{y}}^2$ are the variance of the target and the prediction variables respectively and finally n is the number of samples predicted.

Root-Mean-Square Error (RMSE)

Given by equation (2.29) which provides the prediction errors by means of the standard deviation (sd) of the residuals. It is a measure of how concentrated the data is around the linear regression model fitted. Mean Square Error (MSE) is also used as a metric.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i \in (1, \ldots, n)} (y_i - \widehat{y}_i)^2} \qquad (2.29)$$

Average Percentage Error (APE)

It is the absolute value of the error prediction in relation to the magnitude of the real variable to be predicted. It is a measure of how close the model is to the target. It can be used only for non-negative target variables. If there exist null values in the target variable, then it is necessary to apply scale factor for each time step i, so the APE metric (2.30) is a time-series.

$$APE_i = \frac{|\widehat{y}_i - y|}{y_i}; \quad \forall i \in (1, \ldots, n) \qquad (2.30)$$

Mean Average Percentage Error (MAPE)

It is the mean value of the Average Percentage Error (APE) (2.31).

$$MAPE = \frac{1}{n} \sum_{i \in (1,\dots,n)} APE_i \tag{2.31}$$

Precision

In classification problems, it is the fraction of relevant samples correctly classified (True Positives) among the aggregation of TP samples and False Positive (FP) (samples labeled by the predictive model as False but actually are True). In statistics, precision is also known as Positive Prediction Value (PPV).

$$Precision = \frac{TP}{TP + FP} \tag{2.32}$$

Recall

In classification problems, is the fraction of relevant samples correctly classified (True Positives) among the aggregation of TP samples and False Negative (FN) (samples labeled by the predictive model as True but actually are False). Recall is also known as sensitivity.

$$Recall = \frac{TP}{TP + FN} \tag{2.33}$$

2.3.2.2 Model Assessment

The prediction error ϵ for any learning model can be discomposed into three components: bias ϵ_b, variance ϵ_v and irreducible error ϵ_i as states (2.34).

$$\epsilon = \epsilon_b + \epsilon_v + \epsilon_i \tag{2.34}$$

Irreducible Error

The irreducible error ϵ_i depends on how the regression problem is framed. It is manifested in factory such as the fact of not including variables that are highly correlated with the target variable and so influence the mapping of the remaining input features to the output

target variable. A well-conducted and rigours data analysis can mitigate the probability of having irreducible errors in the regression problem solved by machine learning models.

Bias Error

Bias error ϵ_b arises from the fact of representing a problem by a simpler model. For example, linear models have a high bias which leads to learn faster than other complex models and are easier to understand but generally they are less flexible and have a lower predictive performance on real-life complex problems. As a result, it can be concluded that:

1. Low-Bias error: The model makes very few and mild assumptions and hypothesis about the nature of the mapping function $\mathcal{F}$. For instance, an ANN adapts their network topology and neurones to learn both the simplest linear mapping function as well as a tough non-linear non-convex function. Machine learning models with low-bias are: ANN, DT, KNN and SVM models.

2. High-Bias error: The model makes many and strong assumptions and hypothesis about the nature of the target function $\mathbb{Y}$. For instance, a linear model assumes that exists a linear relationship between the target and the features. In the side of high-bias machine learning models it can be included: Linear Regression models, LDA and Logistic Regression.

Variance Error

Respect to the variance error ϵ_v is the amount by which the target function will change if different input data is used. The target function is estimated from the used training data, so it can be expected that the model will pose some variance when the training data set change. If the variance of the model is low, the magnitude of the target function will not deviate too much from the original input training dataset and so it will mean that the model is appropriate at detecting the hidden underlying mapping between the input variables and the output target variable. On the contrary, models with high variance are strongly influenced by small changes in the training data set.

Generally, non-linear machine learning models are quite flexible (i.e. they learn quite well the underlying structure of the problem) but suffer from high variance (when the input of the model change, the output change dramatically). For instance, decision trees have high variance, that is even higher if the trees are not pruned before its use. It can be concluded that:

1. <u>Low-Variance error</u>: Examples of low-variance machine learning models include the family of Linear Regression models due to the strong assumption of linearity as well as Linear Discriminant Analysis and Logistic Regression.

2. <u>High-Variance error</u>: Examples of high variance machine learning models include Artificial Neural Networks, Decision Trees, k-Nearest Neighbours and Support Vector Machines.

Variance also is related to the concept of over-fitting, which means that a model is completed adapted to the training data set in such a way that when it comes the time to produce inference with a different dataset, the performance obtained is very poor, getting target output with high deviation from the expected. This can be tackled with cross-validation techniques such a K-fold cross validation [Hastie and Friedman, 1991]. It can be seen that exist a trade-off between Bias-Variance when it comes to decide the model to use. The goal of any supervised machine learning model is to achieve low bias and low variance. In turn the algorithm should achieve good prediction performance. In general, linear machine learning models often have a high-bias but a low-variance meanwhile Non-linear machine learning algorithms often have a low-bias but a high-variance.

2.3.2.3 Bias-Variance trade-off

The parameterisation of machine learning model is often a task related to the balance of bias and variance. For instance:

1. KNN models has low-bias and high-variance, but the trade-off can be changed by increasing the number of neighbours (value of K) that contribute to the prediction and in turn it increases the bias of the model [Murphy, 2012].

2. ANN models has low-bias and high-variance, but the trade-off can be changed with dropout regularisation (cancelling connections between neurones) [Murphy, 2012].

The relationship between bias and variance in machine learning is antagonist (Fig. 2.4) in the sense that increasing the bias decreases the variance, and inversely, increasing the variance decreases the bias. In fact, the real bias and variance error terms can not be determined analytically since the actual underlying target function is not known [Bishop, 2006].
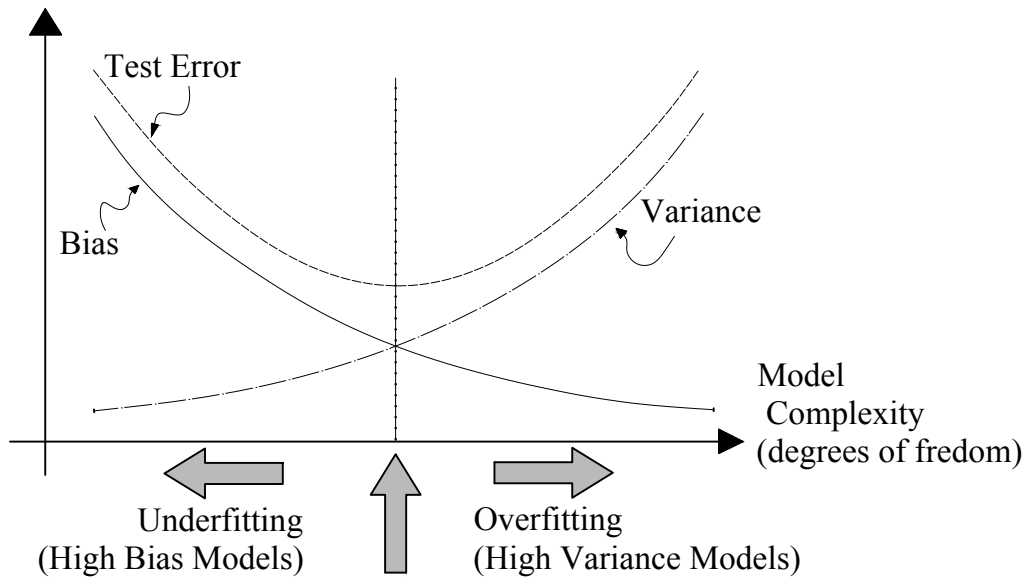
Figure 2.4: Trade-off between bias and variance

The challenge is to find a method that offers a relation low-bias and low-variance, appropriate for the problem to be solved and the requirements. Ensemble methods deals with the bias-variance trade-off. They rely on the machine learning paradigm where multiple models (often called "weak learners") are trained to solve the same problem and their predictions are combined to get better results. The principal hypothesis is that if weak models are properly combined, the resulting model is more accurate and robust. Generally, basics machine learning models (such a linear models or decision trees) perform very poorly either because they have a high bias (i.e low degree of freedom) or because they have too much variance to be robust (i.e the final decision depends, considerably, on the training data set). Then, the idea of ensemble methods is to try reducing the bias

and/or variance of such weak learners by combining several of them (Fig. 2.5) to create a strong learner (or ensemble model) that achieves better performances. There exist two techniques to build ensemble methods: boosting and bagging.
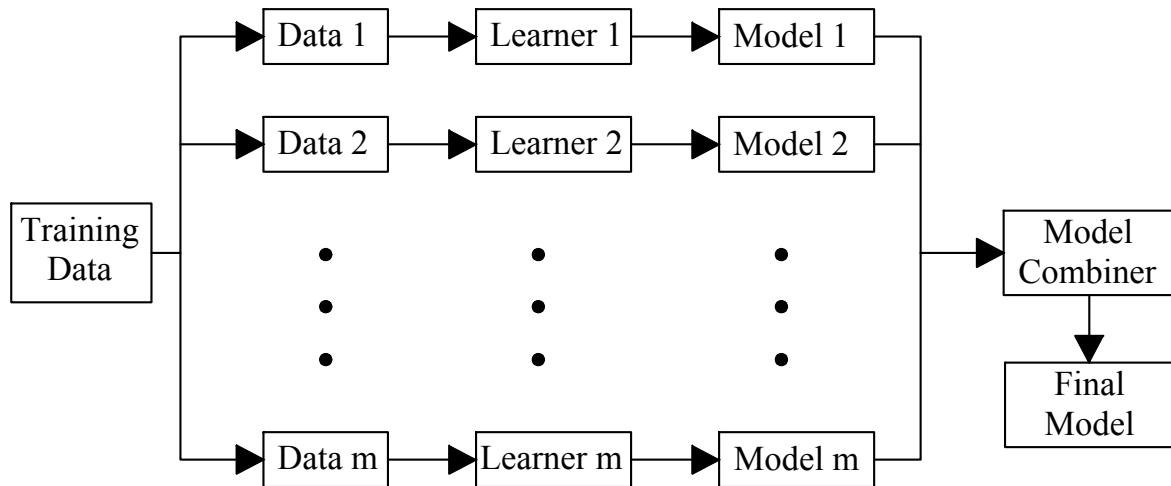
Figure 2.5: Ensemble method procedure illustration

Precision & Accuracy

Precision and accuracy are two terms that frequently are understood as synonyms, but it it not right. Accuracy of a predictive model can be defined as the degree of closeness of the prediction quantity respect to the true value. In contrast, the precision relates to the reproducibility and repeatability of the prediction performed, i.e, it can be defined as the degree to which repeated predictions under different circumstances provides the same quantity. Precision and accuracy are related to bias and variance in the sense that the amount of inaccuracy correspond to the level of bias (high-bias, low accuracy) and the variability corresponds with the quantity of imprecision (the lack of precision, high-variance, low precision).

Boosting Technique

Boosting is a machine learning technique that provides a prediction model based on an ensemble of weak prediction models which usually are decision trees [Hastie and Friedman, 1991].

However, they produce models with low interpretability and require large computational effort. Adaptive boosting (adaboost) and Gradient Boosting are variants of the Boosting technique. If they are applied to decision trees they receive the name of Gradient Boosted Decision Trees (GBDT) which usually outperforms Random Forest ensemble. Finding the best ensemble model is a difficult optimisation problem. Then, instead of trying to solve it in one single-shot (finding directly all the coefficients and weak learners that give the best overall additive model), it can be used an iterative optimisation process that is much more tractable, even if it can lead to a sub-optimal solution. Gradient Boosting casts the problem into a gradient descent one: In each iteration a weak learner is fitted to the negative direction of the gradient of the current fitting error with respect to the current ensemble model [Hastie and Friedman, 1991]. AdaBoost is a meta-estimator that begins by fitting a regressor on the original dataset and then fits additional copies of the regressor on the same dataset but where the weights of instances are adjusted according to the error of the current prediction. The core principle of AdaBoost is to fit a sequence of weak learners (i.e., models that are only slightly better than random guessing, such as small decision trees) on repeatedly modified versions of the data. The predictions from all of them are then combined through a weighted majority vote (or sum) to produce the final prediction. The data modifications at each so-called boosting iteration consist of applying weights to each of the training samples. Initially, those weights are all set to, so that the first step simply trains a weak learner on the original data. For each successive iteration, the sample weights are, individually, modified, and the learning algorithm is reapplied to the reweighed data. At a given step, those training examples that were incorrectly predicted by the boosted model induced at the previous step have their weights increased, whereas the weights are decreased for those that were predicted correctly. As iterations proceed, examples that are difficult to predict receive ever-increasing influence.

Each subsequent weak learner is thereby forced to concentrate on the examples that are missed by the previous ones in the sequence. For non-linear problems, Stochastic Gradient Descent (SGD) technique can be applied to the objective function of the predictive task. SGD is an iterative method for optimising a loss function that poses suitable smoothness properties such as differentiability. It can be understood as a stochastic ap-

proximation of the gradient, since it replaces the actual gradient (calculated from the entire data set) by an estimation (calculated from a random subset of the data). Especially in high-dimensional optimization problems this reduces the computational burden, achieving faster iterations in trade for a lower convergence rate.

Bagging Technique

Bagging is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning model, usually decision tree models. It also mitigates variance and helps to deal with overfitting. The idea of bagging is the same as boosting, which is that the aggregation of many weak learners typically outperforms a single learner over the entire training set and has less overfit. Bagging also can be parallelised, as each single bootstrap can be processed individually before to be combined with the rest. One of the disadvantages of this technique is that for weak learners with high bias, bagging technique will also lead to a model with high bias. In contrast to GBDT ensembles, Random Forest applies the general technique of bagging to the decision tree learners.

Bagging and Boosting are similar techniques in the sense that both train a set of combined weak learners to build a strong learner that obtains better performance than a single learner and both of them can efficiently alleviate variance. Nonetheless, while in the Bagging technique the learners are built independently, in Boosting the method tries to add new models that perform better than the previous ones. Bagging and Boosting generate several training data sets by random sampling and build the final decision by combining the n learners. However, only Boosting specifies a data weighting policy to scale the most difficult cases and Bagging it is the only one that reduces the probability of overfitting problems.

Other ensemble methods

Other ensemble method is the Stacking model which considers heterogeneous weak learners, learns them in parallel and models are combined by training a meta-model to output a prediction based on the different weak models predictions.

Deep Learning-based methods

Deep-learning models are a type of machine learning models which are a subtype of artificial intelligence field in which the feature extraction process is performed automatically in contrast to machine learning models. These deep learning methods such as Multi-Layer Perceptron (MLP) or Long Short-Term Memory (LSTM) models can be used as well as power losses estimators, presenting some advantages and disadvantages [Bishop, 2006]. MLPs are deep artificial neural networks which are inspired in the biological nervous systems. It combines multiple nonlinear processing layers, using simple elements operating in parallel known as neurones. MLPs are composed by an input layer (where is fed the relevant features), several hidden layers (where the learning process takes places) and an output layer (which is the target output). All the layers are interconnected among themselves through the neurons, in which a mathematical operation is performed and an activation function is applied to obtain an output. Each hidden layer uses as input the output of the previous layer [Simon S, 1999].

In regression tasks supervised machine learning models present advantages and disadvantages. Machine learning models are straightforward and easy to implement and interpret [Russel and Norving, 1994]. Additionally, most of them are readily developed, trained, and deployed by using standard hardware, they do not require high-performance computing infrastructure. In this regard, a conventional laptop with common computer power and memory specifications is enough to develop worthwhile models in short periods of times. Moreover, in terms of data requirements, they do not require very large data set to achieve reasonable results. However, they require knowledge about the relevant features to be used as input variables in the regression problem (regressors). This implies that a broad and deep problem-domain knowledge is required to define and select those relevant features. The inappropriate selection of the input features can lead to inaccurate predictions.

In addition, machine learning models have difficulties in learning some non-linear-multivariate relationships. Although deep learning models can extract the relevant features from high-dimensional data sets, they are able to learn complex non-linear relationships achieving in general better performance and accuracy than machine learning

models. Also, deep learning methods can capture very complex patterns with time-sequential dependencies. As disadvantages, deep learning models are considered very computationally expensive algorithms, requiring high-performance computers GPU for training and validation. Furthermore, they require very large data set to extract meaningful features for the regression task. In terms of software, the recent development of powerful libraries based on cutting edge tools as PYTHON has boosted the use of machine learning and deep learning models in scientific production. For instance the library SCIKIT-LEARN [Pedregosa et al., 2011] provides access to a great variety of linear and non-linear machine learning models as well as the framework TENSORFLOW [Abadi et al., 2015] or the library PYTORCH [Paszke et al., 2019] does. Other libraries for specific type of machine learning models are XGBOOST [Chen and Guestrin, 2016] or CATBOOST [Anna Veronika Dorogush, 2018] for GBDT models. Both of them are based on PYTHON programming language.

Lastly, a new class of models is emerging. They combine physical models with machine learning models and they are known as hybrid models. Examples of them are Hidden Markov Models (HMM) [Zucchini et al., 2009] which are general-purpose models for time series data based on a sequence on states and a probability of change (Markov Model). These hybrid methods are useful and efficient in cases of lack of accurate data. However, HMMs rely on strong assumptions and physical knowledge of the process to be predicted.

In the last decade, the use of load profiles and regression analysis has been proposed for losses estimation [Queiroz et al., 2012, Fu et al., 2016, Ibrahim et al., 2017] which are mainly applied to small-size balanced distribution networks. Nonetheless, these techniques require an accurate network data (topology and demand). Consequently, their applicability to large-scale distribution areas is limited because the accurate network data is not always available. Loss Factor-based methods (LFM) consist of computing a quantity defined as losses factor at representative feeders using as input data: the peak demand, the customer's load profile, and the feeder length. LSF methods have been applied in distribution networks using reduced network data [McKenna et al., 2012, Urquhart et al., 2015, Poursharif et al., 2018]. However, they do not consider the connection of DG units, whose presence is ever-increasing in LV distribution networks, mainly as rooftop PV Panels, with

a relevant impact on network losses [C.T.Gaunt et al., 2017]. In this context, distribution systems behave as Active Distribution Networks (ADN) and their optimal operation involves power losses estimation and management [Usman et al., 2018].

Depending on the DG penetration and the spatial-temporal correlation between demand and DG, network losses can be reduced or increased. This fact aggravates the process of network losses estimation using only the data recorded by customers's smart meters and the SS smart meter supervisor. Therefore, any losses estimation method has to consider the presence of DG to provide an accurate losses estimation. To deal with the lack of accurate network data, top-down and bottom-up approaches have been proposed in [Dortolina and Nadira, 2005].

The application of these techniques to large areas is based on the definition of some explanatory variables for each feeder. In the first step (top-down), the different areas are clustered according to the selected explanatory variables, and representative feeders are defined for each cluster group. In the second stage, by means of power flows, loss functions are obtained for each representative feeder's cluster. Losses for each feeder are obtained according to the load that each feeder carries-out. However, the method has been applied only in balanced MV distribution networks, where the network data is more homogeneous and complete. Moreover, the cited method does not consider the unbalance operation of the LV distribution networks so the application of this method to unbalanced low voltage smart grids would provide incorrect results.

Notwithstanding, the concept of representative feeders within a large distribution area seems to be a potential approach to infer power losses level in LV networks, considering their singularities. Network Loss estimation accuracy depends on the closeness between the representative feeders and all those who represent [Usman et al., 2018].

A recent overview of representative feeders works can be found in [Bletterie et al., 2018]. As distribution feeders have different topological structures and different characteristics (e.g., number of clients, line length, energy supply), some authors have proposed the application of a clustering process to obtain losses in a set of representative feeders [Hong and Chao, 2002, Dashtaki and Haghifam, 2013]. In [Hong and Chao, 2002], a Fuzzy C-Number (FCN) clustering technique is used to obtain a losses equation applying

a cluster-wise fuzzy regression analysis. In [Dashtaki and Haghifam, 2013] a clustering-based method is presented to find feeders that have similar characteristics. For each cluster obtained, the mean feeder is selected to estimate the technical losses of the network. The application to large-scale areas is achieved using an iterative process where random feeder parameters are selected to create new network topologies. Although the method has been designed for LV distribution areas, it does not consider the unbalanced operation of these networks, consequently, the losses estimation provided by this method will be useful only for balanced feeders. In [Mateo et al., 2018] a methodology to obtain the large-scale representative networks is presented. However, unbalanced operation in LV distribution feeders is not considered. Machine Learning (ML) algorithms have also been applied to estimate losses [Monteiro et al., 2018, Wang et al., 2017, Hsu et al., 1995, Kang et al., 2006, Leal et al., 2009, Chen et al., 2005, Li et al., 2019]. In [Wang et al., 2017] eXtreme Gradient Boosting (XGBoost) is used for estimate statistical line losses parameters of distribution feeders. The method requires a high number of representative feeders and, to simplify the process, a great amount of historical data is needed to apply the clustering feeder stage. This is an important limitation for expanding its applicability to large-scale areas with reduced network data available. In [Hsu et al., 1995], a feedforward Artificial Neural Networks (ANNs) for feeder loss analysis is proposed. Key factors of feeder losses, such a feeder loading, power factor, feeder length, and transformer capacity, are used as input variables to the network. The proposed method provides accurate results, but it requires a large and detailed training set to build up the neural network. Based on the same idea, [Kang et al., 2006] proposed an ANN-based model to estimate power losses in distribution and transmission networks. From the works cited, it has been proved that ANN models are well suited to provide an estimation of power losses in distribution feeder but only for a specific feeder (topology configuration) under particular operation conditions. However, if the operation conditions of the LV feeder changes, the ANN model has to be trained again to adapt itself to the new circumstances (such a weather conditions, day of the week, etc.). Another aspect to be highlighted is that most of the ANN-based methods analysed in the literature are static and do not consider dynamic changes of the demand, therefore they are not applicable in real time operation. Additionally, following

the process of power losses estimation by the ANN-based model (or with any model), it is necessary to perform a statistical study to quantify the closeness of the power losses estimation to the real losses because the representative feeder's membership rely on the fact that feeders have certain characteristics in common.

Unbalanced operation of the LV distribution networks plays a key role in the power losses levels [Fang et al., 2019]. LV remains highly unexplored in terms of monitoring, even with the large deployment of smart meters infrastructures [Fang et al., 2019]. For instance, the phase connection of each customer is still a non-trivial question and remain unknown in the majority of the networks [Xu et al., 2018].

Analytical equations for losses calculations in idealised feeders have been proposed in the literature. In [Chang, 1968] idealised feeders were used considering uniformly distributed loads and Losses equations are formulated for lines, transformers, and regulators. To generalise, in [Schultz, 1978] it is proposed the relationships between losses, topology configurations and load level for MV radial networks with uniformly distributed loads. However, the losses expressions provided are not applicable to some specific cases so the applicability to real networks is unviable.

Other methods were formulated based on load profiles to estimate losses such as the Percent Loading (PrL) method [Flaten, 1988]. The method consists of using an empirical equation for losses estimation in each network equipment. However, the use of the proposed equations requires a deep knowledge of the network. Empirical equations that relate losses with demand was developed in [Gustafson and Baylor, 1989]. In this kind of equation-based methods, simplified feeder models were used for losses calculation by using load flows and then a curve fitting process was applied to obtain the coefficients of the quadratic equations for losses estimation. Also, quadratic expression to estimate losses, based on load flows, under step capacitor banks switching and transformer's taps changes were presented in [Baldick and Wu, 1991]. However, these methods do not provide information about marginal losses and no-load losses.

An approximate method based on the use of Daily Load Curves (DLCs) is presented in [Shenkman, 1990] where the square-integrable load is obtained by applying the second statistical moment of the representative DLC for each load. The method was improved

in [Taleski and Rajicic, 1996] considering nodal voltages as variables. Nodal voltages are calculated by load flow techniques for several scenarios of power demand. The use of nodal voltages instead of network rated voltage increases the accuracy of the losses calculation. However, the applicability of these methods to large distribution systems is limited.

Exact and simplified feeder models for power losses calculation are applied in the Taipower distribution network [Chen et al., 1994]. The exact power losses model allows calculating power losses with a high accuracy but a substantial effort is needed to prepare the required input data. On the contrary, the simplified power losses model, based on conventional regression techniques, allows estimating power losses using information of feeder length, conductor size, loading level and power factor. Although the simplified power losses model is quite simple to use, it does not provide an accurate estimation.

An extended approach to estimate losses is based on Loss Factor techniques. This Factor, is multiplied with the peak demand to obtain the average losses of the system. In [Mikić, 2007] energy losses are estimated by means of the calculation of load variance and load mean losses using LFM. In [Queiroz et al., 2012] it is proposed to replace the traditional LFM and the Equivalent Hours (EH) by a Loss Coefficient (LSC) and Equivalent Hours of Losses (EHL), respectively, to improve the accuracy of losses estimation. This is due to the fact that average demand refers to a period of time instead of the maximum demand which is fixed to a time instant. In [Fu et al., 2016], Load Factor (LF) and Minimum Load Factor (MLF) are used to propose two expressions for an improved LSF. In [Ibrahim et al., 2017] reference feeder models for Medium Voltage (MV) networks are used to estimate losses using LFM. The use of the Loss Factor as losses estimator is justified because maximum demand is a quantity usually available. Therefore, LM is a direct way to estimate losses. Despite the lately improvements, LFM still provides losses estimation with low accuracy because the method is only based on the relationship between peak demand and peak losses and between peak losses and mean losses. Moreover, the method heavily depends on load profiles ignoring the topology of the network.

LFM-based methods can be applied when accurate demand data is available. However, the lack of accurate demand data is one of the open unsolved challenges in loss estimation, together with the no availability of accurate topology network data. In

[Gustafson and Baylor, 1988] power losses calculation for the different components of the system, which are useful in economical studies to determine the savings in power losses, are calculated through demand response practices. In [Shenkman, 1990] a series of statistical moments to estimate power losses in distribution networks is proposed by using the daily demand profile for different loads. In [Baldick and Wu, 1991] it is shown the correlation between power losses and different network parameters. Also, it is suggested its use in the studies related to the networks capacitors location to mitigate technical contingencies. In [Chen et al., 1994] it is presented a power losses model based on linear regression techniques using as information the customer's demand. A Monte Carlo simulation is performed in which the network parameters are changed using different load profiles. In [Taleski and Rajicic, 1996] it is proposed a method for power losses calculation in radial distribution networks based upon the power flow equations and using the statistical moments. In [Chiang et al., 1997] it is shown a correlation between the power losses estimation and the network parameters. In [Hong and Chao, 2002] it is presented a method for estimating power losses based on fuzzy logic techniques such us (Fuzzy-C-Number) or CWFR (Cluster-Wise Fuzzy Regression).

If accurate data for the network topology data and demand data is available, technical losses can be calculated using load flow analysis. In [Sun et al., 1980] it is presented a comprehensive loss estimation method using detailed feeder and load models in a load flow program. The proposed three-phase load flow procedure, unlike the existing ones, incorporate a Load Duration Curve (LDC) to calculate energy losses. Principal outcome of the study is that losses estimation accuracy depends on the accuracy of the data and there is a component of losses constant (without load demand) due to the transformers, and other components that vary with demand. Nevertheless, the proposed method requires in-depth knowledge of the system and modelling the distribution feeders, which is a difficult and time-consuming task. A radial power flow method to find peak demand losses and energy losses in distribution feeders in the Columbian Power System was presented in [Cespedes et al., 1983] and simplified feeders' models for power flows analysis to calculate power losses were proposed in [Vempati et al., 1987]. Distribution feeders involve three-phases lateral lines and/or two-phases and single-phase lateral lines. This variety of feeders

makes difficult to generalise the losses estimation process for large distribution areas.

In [Grigoras and Scarlatache, 2013] decision trees were applied for losses estimation where clustering feeders solved the problem of dealing with a variety of feeders improving the accuracy of power losses estimation. In [Hong and Chao, 2002] the variety of feeders is solved by using a Fuzzy C-Number (FCN) clustering technique with a cluster-wise fuzzy regression to obtain a loss formula. In [Dashtaki and Haghifam, 2013] an alternative approach for losses estimation with a great variety of feeders is presented. The method is based on matching the feeder to study with a set of benchmark feeders. Various features are considered such as: main feeder length, lateral branches length, number of branches, number of clients and conductor size. In [Wang et al., 2017] a set of descriptive variables for the feeders is defined where a statistical procedure to select a reduced group of variables is applied. In this work the optimal number of clusters is obtained by the *K-medoids* algorithm and the XGBoost model is applied for the estimation of the statistical line loss of distribution feeders. Additionally, XGBoost model is compared with others traditional machine learning methods such as decision trees, neural networks and random forest and results showed that XGBoost outperforms all of them. A drawback of the proposed method is that it requires the statistical value of losses in each feeder (from historical records) and the load flow power losses values, so depth knowledge of the network is essential which limits its applicability to large distribution areas. Furthermore, the weakness of these methods is the difficulty of finding two feeders to be exactly equal and consequently, the clustering proposed method or matching is not always possible

In [Rao and Deekshit, 2006] two schemes for estimating losses in HV and LV distribution networks are proposed by means of load flow and classical equations using limited data collected from feeders. However, the proposed methodology underestimate power losses and so it results inadequate. In the proposed scheme a load flow model of the feeder is used in combination with energy measured data to estimate the average loss corresponding to each feeder. The method is based on a mean loss curve based on a quadratic function. Firstly, the losses for each feeder during each half hour interval were computed based on the measured data of demand, using a load flow program. Secondly, a mean loss curve is obtained for each feeder using regression approaches which provides the mean of

the measured loss at all the loads, and consequently, energy losses of all the distribution feeders can be estimated. For LV distribution feeders, phases currents' measurements are used to calculate losses in idealised feeders. Both methods exhibit a satisfactory accuracy, but they require also accurate demand data. The time resolution of the load flow analysis for losses estimation, typically, is 15, 30 and 60 min as reported in [Gouveia et al., 2021] [McKenna et al., 2012]. This long-time resolution drives to an underestimation of the losses, as it is pointed out in [Urquhart et al., 2015]. In [Poursharif et al., 2018] the time interval of the consumed energy data is analysed. Different approaches are applied to the losses estimation, Nevertheless, all of them are hardly applicable to large distribution systems because of the need of highly accurate demand data.

Machine learning algorithms such as ANNs also were applied to estimate losses. In [Hsu et al., 1995] a feedforward ANN for feeder loss analysis was proposed where feeder loading, power factor, feeder length, and transformer capacity were used as input variables. the ANN training set included real losses calculated by Load flow techniques. The proposed method had an appropriate performance in comparison with conventional regression techniques and with real losses but an accurate and detailed training set was necessary to set up the ANN.

Similarly, in [Kang et al., 2006] feeder loading, transformer capacity and conductor length were used as input variables for a ANN-based simplified power loss model for distribution feeders. The ANN models were developed both for overhead and underground feeders and ANN models were used to estimate power losses in the Taipower distribution network. However, the accuracy of the estimations was not verified. It has to be noted that the use of ANNs to estimate losses could be useful in those situations in which the connectivity between clients and the distribution feeders are not known. In some cases ANNs provides power losses estimations with an accuracy superior to analytical methods [Leal et al., 2009]. However this is achieved at the expense of an excessive computational time. On the contrary, the simplification of the ANN reduces the computational task but the precision of the estimation is smaller. Although most of the previous methods have good accuracy, they depend excessively on the availability of accurate network demand and topology data.

2.3.3 Methods comparison

Losses in LV distribution systems are calculated by means of load flow when topology network data and load demand data are known. However, when this information is incomplete or unavailable, losses have to be estimated. LV Losses estimation methods that have been proposed to date require a large amount of data (e.g. ANN requires data for training) or make use of simplistic loss models. In this context, it is necessary the development of losses estimation methods that take into account the uncertainty of the load demand and the intermittent behaviour of the DERs. Moreover, the unbalance operation of the LV distribution systems has to be taken into account for a precise losses estimation.

All the losses allocations procedures presented make use of some arbitrariness because of the non-linear relation between losses and power flow which complicate the explicit allocation of losses among the end-users [Usman et al., 2018].

Given the presented literature review, it can be noticed that there are a large scientific production published related to power losses estimation in low-voltage balanced and deterministic distribution networks. However, very few tackle the issue considering the three-phase model network neither account explicitly for the uncertainty present in form of intermittent PV injections or unexpected changed in load demand. Existing solutions are based in solving an equivalent single-phase power flow model of the low-voltage distribution network, assuming constant both demand and generation (if its accounted) and calculating the power losses. That procedure leads to non-realistic power losses values and to non-feasible energy efficiency policies and practices. The most relevant and recent works in power losses estimation methods are summarised in Table 2.3.

Table 2.3: Power losses estimation methods comparison

Reference	DERs	Uncertainty	NTM	Unbalanced	Resolution	Approach	Large-Scale	Voltage Level	Test Network
[Mikić, 2007]	No	Variance	No	No	-	AE-LSF	No	LV	-/ - / 11 buses
[Leal et al., 2009]	No	No	No	No	15 min	ML	Yes	MV	-/ 608 feeders/ -
[Urquhart et al., 2015]	No	No	No	No	1min	AE-LSF	Yes	LV	-/ - / 22 buses
[Fu et al., 2016]	DG	No	No	No	-	AE-LSF	Yes	MV	-/ - / 43 buses
[Ibrahim et al., 2017]	No	No	No	No	-	AE	Yes	LV	44.9 MW / 4 feeders / -
[Wang et al., 2017]	No	No	No	No	-	ML	Yes	MV	- / 762 feeders / -
[Poursharif et al., 2018]	No	No	No	Yes	1-30 min	AE-LSF	No	LV	- / 3 feeders / 140 buses /
[Carvalho et al., 2018]	No	PDF	No	Yes	30 min	PPFS	No	LV	400 kVA / 1 feeder / 67 buses
[Fang et al., 2019]	No	No	No	Yes	-	Clustering	No	MV-LV	- / 800 feeders / -
[Gouveia et al., 2021]	No	PDF	No	Yes	1-15min	PPFS	4No	LV	400 kVA / 1 feeder/ 67 buses

DG: Distributed Generation; PFS: Power-Flow Simulation; NTM: Non-Telemetered Customers; SM: Smart Meters; AE: Analytical Equations; LF: Loss Factor, Lf: Load Factor, TP-BT: Top-down/Botton up approach, PFr: Power flow analysis with regression models, ML: machine learning techniques; AE: Analytical Equations

2.4 Optimization Techniques for Power Losses Minimisation

Power Losses minimisation in LV unbalanced smart grids refers to the optimal decision making in respect to the deployment of policies and practices that minimise certain objectives subject to technical constraints. These policies and practices correspond with Flexibility mechanisms to balance power demand and power generation in such a way that there is an adequate value of power losses keeping the network state in normal and safety conditions.

Unbalanced Optimal Power Flow (UOPF) in smart grids is affected by uncertainties such as: customer consumption, BES charging, PV power injections or energy prices. Usually, aggregated household power demand is forecasted by means of time series models or machine learning models providing an expected value for the customer aggregated demand in each period of time. Moreover, the price of the energy consumed by customers and the potential price of the potential flexibility, are variable and may change. They depend both on the market clearing and the generation technologies involved adding uncertainties in the smart grid optimal operation.

Similarly, PV generation involves uncertainties because solar irradiation is variable affecting the expected power generation. Depending on how uncertainty is modelled the optimisation problem can be categorised into:

- Deterministic Optimisation (DO).

- Non-deterministic Optimisation (NDO).

    - Probabilistic Optimisation.

    - Possibilistic Optimisation.

    - Stochastic Optimisation (SO).

    - Robust Optimisation (RO).

A variable or process is deterministic if it can take only one possible outcome (i.e, if the next event in the sequence can be determined exactly from the current event). In

contrast, a variable process is defined as random if it can take on one or more possible outcomes or events. This means that for a optimisation problem, for instance (2.35), if the parameters $(A, \mathbf{b})$ are known and fixed, the optimal solution $\mathbf{x}^*$ will always be the same (for a given starting point $\mathbf{x}_0$ and using the same solver algorithm). In contrast, if the parameters $(A, \mathbf{b})$ are replaced by random variables $(\widetilde{A}, \widetilde{\mathbf{x}})$ then the optimal solution $\widetilde{\mathbf{x}}^* \neq \mathbf{x}^*$ will not be the same since it is affected by the uncertainty in the input parameters and then the optimisation problem is a non-deterministic problem.

$$
\begin{aligned}
\min_{\mathbf{x}} \quad & \mathbf{c} \cdot \mathbf{x} \\
\text{s.t.} \quad & A \cdot \mathbf{x} \leq \mathbf{b} \\
& \mathbf{x} \geq 0 \\
& \mathbf{x} = [x_1, \ldots, x_n]^T \\
& \mathbf{c} = [c_1, \ldots, c_n] \\
& A \in M^{m \times n} \\
& \mathbf{b} = [b_1, \ldots, b_m]^T
\end{aligned}
\tag{2.35}
$$

Non-deterministic optimisation problems may give different outcomes for the same input, i.e there is uncertainty associated with the input parameters that affect the optimal solution.

An optimisation model with stochastic variables, as parameters, is defined as a stochastic optimisation model, because the model describes a mathematical process that use or harness randomness. Common examples include Brownian motion and or Markov Processes. In general, stochastic is a synonym for probabilistic and for random. A random variable is stochastic if there is uncertainty or randomness involved with its outcomes (i.e if the occurrence of events or outcomes involves randomness or uncertainty).

Moreover, an optimisation problem with probabilistic variables as parameters is defined as a probabilistic optimisation model. A common example is Monte Carlo sampling simulation. A random variable or process is probabilistic if it can be summarised and analysed using the tools of probability (i.e if the occurrence of events or outcomes can be defined with a probability). Most notably, the trajectory of a random variable over a period can be described in terms of a probability distribution.

We may choose to describe a variable as probabilistic over stochastic if we wish to emphasise the probability of the sequence's outcomes and the independence of the events. Typically, probabilistic is used to refer to a lack of dependence between observations in a sequence. For example, the rolls of a fair die are probabilistic, so are the flips of a fair coin. A probabilistic variable or a probabilistic sequence can still be summarised using a probability distribution (for instance a Gaussian distribution). We may describe something as stochastic over probabilistic if we are interested in focusing on the probabilistic nature of the variable, such as a partial dependence between the next and the actual event.

Another technique to account for uncertainty is the possibilistic approach by which the uncertainty variable is described using a membership function [Soroudi, 2011]. Within this approach exist methods such us α-cut method with provides the membership function of $\widetilde{\mathbf{y}} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$. The method is the defuzzification which is a mathematical process for transforming a fuzzy number into a consistent number (without uncertainty).

2.4.1 Probabilistic Optimisation

In probabilistic optimization each variable affected by uncertainty is statistically characterised in terms of finding the Probability Density Function (PDF) that best describes the behaviour of the variables affected by the uncertainty. Compared to the deterministic optimization, instead of producing a forecast for each variable during the period under consideration obtaining (n) discrete expected values, a PDF is fitted to the historical data for each variable. Once the best PDFs are fitted for the uncertain variables, a Monte Carlo Simulation is performed (Fig. 2.6), subtracting in each simulation a sample from the PDF of each variable and solving an instance of the optimization model. However, It is necessary to carry out a considerable number of simulations until the final value of the optimal function converges.

Each variable affected by uncertainty can be described in terms of a PDF that best fit to the cumulated histogram of the historical data. This PDF can be a defined one (gaussian normal, exponential, beta, etc.) or the probability density can be estimated using the Kernel Density Estimation (KDE) method (2.36)-(2.38) with mean $(\mu^p_{k,(\bullet),t})$ and a standard deviation $(\sigma^p_{k,(\bullet),t})$ parameters.
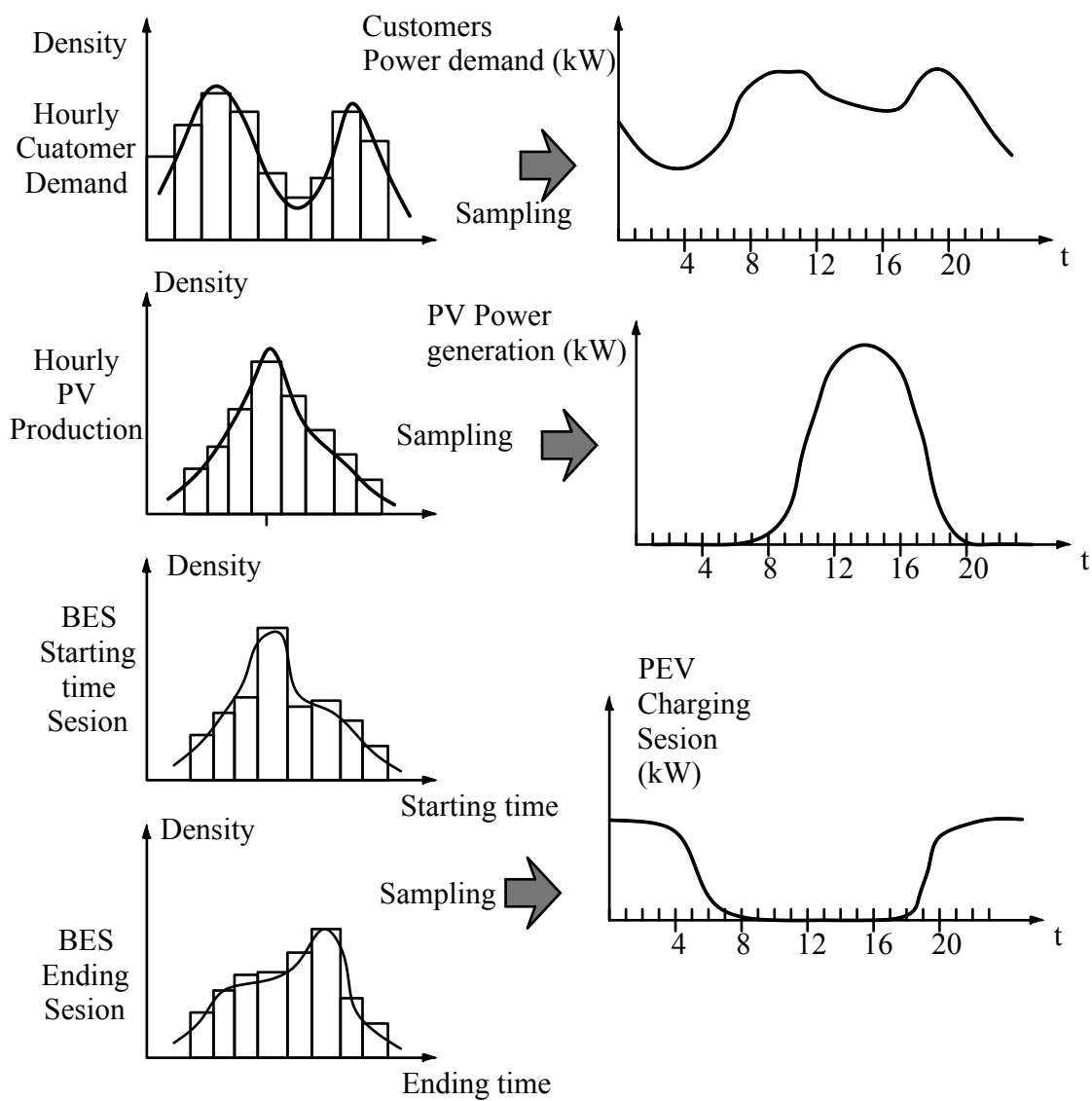
Figure 2.6: Sampling process for Monte Carlo simulation

$$\tilde{p}_{k,(\bullet),t}^{p} \approx \Phi_{(\bullet)}\left(\mu_{k,(\bullet),t}^{p},\ \sigma_{k,(\bullet),t}^{p}\right), t \in T_n, \forall k \in V \tag{2.36}$$

$$\Phi_{(\bullet)}(v) = (nh)^{-1} \sum_{i \in |v|} \phi((v - v_i)/h) \tag{2.37}$$

$$h = \left(\frac{4\ (\sigma_{k,(\bullet),t}^{p})^5}{3n}\right)^{1/5} \tag{2.38}$$

Where $\phi(\bullet)$ is the kernel (window function), v is the set of data points related to the variable, v_i is the ith value of v, h is a smoothing parameter which depends on the standard deviation of the data and the size of the data set.

In contrast with the deterministic optimization, in which a single optimal value is obtained, what is obtained with the probabilistic optimization (Fig. 2.7) is a collection of values for the optimal function (2.39), each one is the result of solving a deterministic optimisation model with a sample extracted from the PDFs. Once the collection of values for the optimal function is large enough to be considered as representative, the PDFs of the objective function value is obtained, which is statistically correlated with the distribution of the uncertain variables.

$$OF_S = \left\{OF_{(1)}, \dots, OF_{(N_s)}\right\} \tag{2.39}$$

2.4.2 Stochastic Optimisation

Also known as Scenario based decision making it is actually a type of probabilistic optimization, since make use of the concept of probability and distribution function. In stochastic optimisation uncertainty modelling is quantified using probability distributions. Variability in DERs generation and customers demand (as well as any other variable in the network affected by uncertainty) can be described in terms of stochastic variables because the outcome (power generation, power demand) involves some randomness and has some uncertainty. It is a mathematical term which is closely related to "randomness" and "probabilistic" and can be considered as "'non-deterministic".
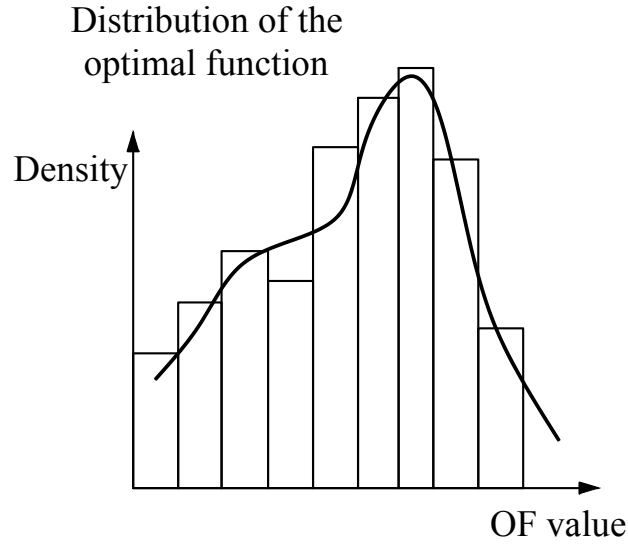
Figure 2.7: Distribution of the optimal function in probabilistic optimization

Having characterised the variables affected by uncertainty, the stochastic optimization seeks to minimise the expected value of the objective function under consideration over a limited set of statistically dependent scenarios (2.40)-(2.41). The expected value of the optimal function is calculated over all plausible scenarios considered as the summation of the products between the probability of the scenario (π_ω) and the objective function. A considerable number of scenarios can be created but in order to keep tractability, scenario reduction techniques need to be applied to reduce a large number of scenarios in a more compact set maintaining the same statistical moments (i.e keeping the same statistical information in less samples). The procedure consists in calculating the statistical moments of the overall scenarios (expected value, standard deviation, skewness and kurtosis) and reducing the number of randomly scenarios.

$$\chi_\omega = p^p_{k,(\bullet),\omega} = \left\{ p^p_{k,(\bullet),\omega,t_0}, \ \ldots, \ p^p_{k,(\bullet),\omega,t_n} \right\} \tag{2.40}$$

$$\min_{\chi_\omega} \sum_{\omega}^{N_\omega = N_\lambda \cdot N_{PV} \cdot N_H} \pi_\omega \cdot OF(\chi_\omega)$$

$$\text{s. t.} \quad f(\chi_\omega) = 0 \tag{2.41}$$

$$g(\chi_\omega) \leq 0$$

$$\chi_\omega \in \mathbb{R}$$

Where π_ω is the probability of the scenario ω. The Value of the Stochastic Solution (VSS) (2.42) is calculated as the difference between average value over scenarios of the objective function evaluated with the stochastic solution of each scenario (Recourse Problem, REP) (2.43) and the average value over scenarios of the objective function evaluated with the deterministic solution (Expectation of the expected Value Solution, EEV) (2.44), random variables substituted by their respective values.

If the number of scenarios is too large, the stochastic optimisation problem can result intractable. To solve this, the number of scenario can be reduced. This is carried out by defining distance metric between each pair of probability distributions scenarios $d(\omega, \omega')$ (such as Fortet-Mourier and Wasserstein [Henrion et al., 2009]) and then using this metric and the probability of each scenario, obtaining a reduced set of scenarios $\Omega'_{N_\omega} \subset \Omega_{N_\omega}$ keeping the statistical moments (expected value and variance) mostly constants. This kind of technique alleviates the computational burden of the stochastic programming problem but at expense of a loss of information. So it should be done in such a way that a trade-off between the two facts is achieved.

$$VSS = EEV - REP \tag{2.42}$$

$$REP = \frac{1}{N_\omega} \sum_{\omega}^{N_\omega} [OF(\chi_\omega^*)]_\omega \tag{2.43}$$

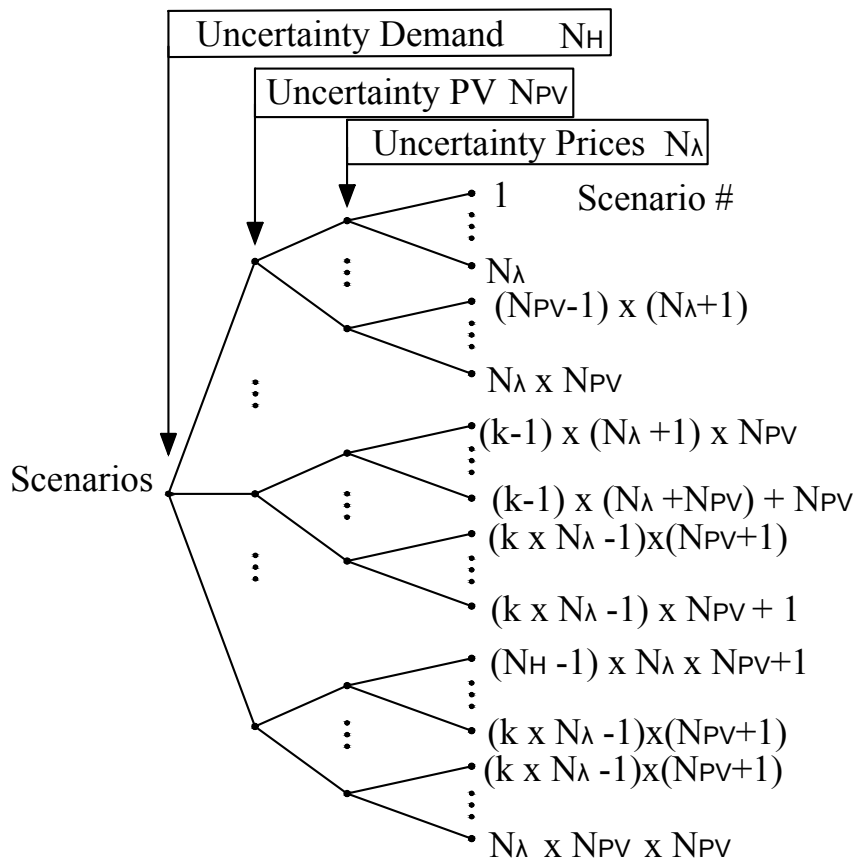$$EEV = \frac{1}{N_\omega} \sum_{\omega}^{N_\omega} [OF(\chi_d^*)]_\omega \tag{2.44}$$

Figure 2.8: Illustration of a scenarios tree in an stochastic programming problem with uncertain variables namely: power demand, PV power generation and energy prices

2.4.3 Robust Optimisation

Robust optimisation (RO) treats the input parameters as random variables considering that they can take values within an uncertainty set and quantifying the total uncertainty by means of an uncertainty budget [Ben-Tal et al., 2009]. In contrast with probabilistic and stochastic optimization, RO does not require accurate probabilistic information nor relies in a computational extensive discretisation of the uncertainty based on scenarios. Instead, RO models the uncertainty by decision variables within a pre-specified uncertainty set in such a way that the optimal solution founded by RO is immune to all possible situations of the uncertainty sources. In Stochastic Programming (SP) uncertainty is described through a probability function but it suffers from the curse of dimensionality. On the contrary, RO uncertainty is not stochastic (e.g., there are no random variables) it is deterministic and based on bounded and convex sets. Moreover, in Stochastic Optimisation (SO) the solution obtained is feasible with some confidence interval and the uncertainty is defined by their Probability Distribution Function (PDF), in RO the solution is feasible for any possible realisation of the uncertainty, which is defined as a given set [Shafie-Khah et al., 2018b].

RO has emerged as a practical alternative framework to stochastic programming for generation scheduling under uncertainty. For decision-making process that is repeated, SO results more favourable due to the lower cost in expectation. On the other hand, RO is more suitable for long term decision problems.

Robust Optimisation (RO) has been applied to properly model DERs and customers uncertainties [Abdi et al., 2017]. In the presence of uncertainty, RO can capture the available flexibility to respond to any combinations of uncertain operations of DERs and power demand (i.e. the system can cope with worst case conditions) and still the optimization model can be computationally tractable.

There exist multiple ways to address and model RO formulations [Ben-Tal et al., 2009, Bertsimas and Thiele, 2006]. The aim of applying RO is to find a solution for the optimization of DERs in LV networks in such a way that the solution obtained is robust enough to DERs uncertainties. This is modelled through the definition of an uncertainty set, that accounts for the uncertainty of each parameter associated with DERs. Usu-

ally, the uncertainty set is represented by the budget- constrained polyhedral approach where boundaries are defined by means of linear inequality constraints. Each vertex of the polyhedron is determined by the intersection of the constraints.

In robust optimisation the uncertainties are modelled using the worst-case scenario in an uncertainty set. Depending on the number of decisions variables and the uncertainty characterisation, RO can be also classified in:

- Single-stage Robust Optimisation.

- Adaptive Robust Optimisation (ARO).

- Adaptive Robust Stochastic Optimisation (ARSO).

- Distributionallly Robust Optimization (DRO).

- Adaptive Distributionallly Robust Optimization (ADRO).

Single-stage Robust Optimisation

Robust optimization is appropriate to identify optimal decisions under uncertainty if these decisions are critical in economic and security terms. Robust optimization allows for robustness control, adapting the decisions to the required level of robustness (protection). In contrast to this, Stochastic Programming is generally computability unstable. Robust optimization problems are generally tractable since their sizes do not depend on the accuracy of the uncertainty description. Contrary to stochastic programming or probabilistic programming where the size of the problem depends on the number of scenarios considered or the number of simulations executed.

Uncertainty sources that can be considered are:

- Load customers demand: Every load demand customer forecast is attached with an uncertainty set Φ_H.

- Energy prices: the energy daily prices depends on the market and so it can fluctuate.

- PH/PEV demand: Every PH/PEV charging pattern demand forecast is bounded with an uncertainty set Φ_{PEV}.

- PV generation: Every PV unit forecast is bounded with an uncertainty set Φ_{PV}.

To deal with the variables affected by uncertainty (PV injections, load demand, fluctuations or unexpected PEV demand) single-state robust UOPF is formulated. The objective is to find a solution that minimises the objective function OF between all the feasible solutions for all situations of the uncertain sources. The uncertainty set is a non-empty compact set, which means that the set is closed (containing all its limits points) and bounded (having all its points lie within some fixed distance of each other). The compactness assumption of the uncertainty set implies that the uncertainty is bounded, which reflects the reality.

Unlike stochastic or probabilistic approaches, robust optimization does not rely on statistical information nor depends on expensive discretisation of uncertainty in scenarios. Instead, RO models uncertainty by decision variables within a pre-specified uncertainty set. In such a way, it provides an optimal solution (χ_r^*) which is immune to all possible realisations of the uncertainty sources within the defined uncertainty set (Φ_u). The worst-case values are found on the extreme uncertainty set intervals.

Consequently, it is considered a budget-constrained polyhedral uncertainty set given by equation (2.45) which provides tractable solutions, and it is based on the deviation of the uncertainty parameters from their expected value.

$$\Phi_{\mathbf{u}} = \left\{ \mathbf{u} \in R^q | u^L \leq u_i \leq u^U : \sum_{i=1}^{q} \frac{u_i - \bar{u}}{u_i} \leq \zeta_{\mathbf{u}} \right\} \tag{2.45}$$

Where: $\mathbf{u}$ is the uncertainty vector parameter under consideration, u_i is the ith entry of the uncertain parameter $\mathbf{u}$, u^L is the lower bound of the uncertainty set given by expression (2.46), u^U is the upper bound, given by expression (2.49), and $\zeta_{\mathbf{u}}$ is a user defined parameter that represents the budget of uncertainty imposed to the total absolute deviation of the uncertainty-related variables u_i from their expected values $(\bar{u}_i)$.

$$u^L = \bar{u} - \Delta u; \forall i = (1 \ldots, q) \tag{2.46}$$

$$u^U = \bar{u} + \Delta u; \forall i = (1 \ldots, q) \tag{2.47}$$

The process of building robust uncertainty sets requires to model the correlation of these sets across time and space. For instance, power demand in a specific location and time can vary in the interval (a; b), and this variation can be correlated with the variation of the energy price in the interval (c; d). Note that if the uncertainty budget is $\zeta_{\mathbf{u}} = 0$, then uncertainty set is a singleton $\Phi_u = \bar{u}$ which corresponds with the case of no-uncertainty. As the uncertainty budget get larger, the robustness of the solution so does. If $\zeta_{\mathbf{u}} = q$, then all entries of the uncertain parameter can simultaneously take their upper and lower bounds. The uncertainty set chosen relies on the consideration of intervals centralised around the expected value i.e., it assumes that the uncertain parameters follow a normal distribution.

Uncertainty sets are similar to the envelope bound model from Information Gain Decision Theory (IGDT), which is a non-probabilistic and non-fuzzy approach for quantify the uncertainty [Haim, 2006]. In this approach uncertainty is defined as the distance between the actual value of the variable and the expected value. In (2.48) it is defined that within the uncertainty budget (a set) $U(\rho, \hat{u})$ are included all the values of the uncertain variable u which deviation from $\hat{u}$ is lower than $\rho \cdot \hat{u}$, where, ρ is the uncertainty horizon of variable u (budget) and $\hat{u}$ is the expected value.

$$u \in U(\rho, \hat{u}) = \left| \frac{u - \hat{u}}{\hat{u}} \right| \leq \rho \tag{2.48}$$

In some cases, a two-stage robust optimisation can be applied. In the First-stage decision variables known as "here-and-now" are decided before uncertainties are realised. In the Second-stage decision variables known as "wait-and-see" are decided after the exact model parameters have been observed.

In that scheme a preventive view is considered, since it is desired to be protected against the worst realisation of the uncertain parameters $u \in U$. (i.e. the values for which the objective function is maximised). For the worst uncertainty realisation, it is desired to minimise the objective function $OF(x, u)$ to get the optimal decision variable $\mathbf{x}^* \in \chi$. By this way, the model provides preventive protection against the worst realisation (from an optimization point of view) within the uncertainty set. For instance, an objective function related to the total flexibility cost would produce an optimal solution $\mathbf{x}^* \in \chi$

because the uncertain parameter related to the energy prices takes the highest value within the defined uncertainty set for that variable. In the same way, in the case of the objective function related to the total power losses it would produce an optimal solution $x^* \in \chi$ for the scenario in which an the PV generation produce the highest value within the uncertainty set for the scenario in which the PV generation produces the highest value within the uncertainty set and the load demand customers takes the lowest value inside its uncertainty set.

So, in this robust optimization model exists two levels: in the first level the decision variable value is fixed and in the second level the objective function is maximised considering the value of the adopted decision variable fixed in the upper level.

$$
\begin{aligned}
&\min_{\mathbf{x} \in X} \max_{\mathbf{u} \in U} \quad OF(\mathbf{x}, \mathbf{u}) \\
&\text{s.\,t.} \qquad f(\mathbf{x}, \mathbf{u}) = 0 \\
&\qquad\qquad g(\mathbf{x}, \mathbf{u}) \leq 0
\end{aligned}
\tag{2.49}
$$

In this problem, the decision variable $\mathbf{x}^* \in \chi$ can be the nodal phase voltage (real and imaginary components). Meanwhile, uncertain parameters $\mathbf{u} \in U$ can be those variables affected by uncertainty namely: customers demand, PV generation, energy prices and PEV charging sessions.

Adaptive Robust Optimisation (ARO)

Usually, once the uncertainty realises it is possible to mitigate the potentially pernicious effects of the uncertainty. This is tackled by the Adaptive Robust Optimization(ARO).

$$
\begin{aligned}
&\min_{\mathbf{x} \in X} \max_{\mathbf{u} \in U} \min_{\mathbf{y} \in Y(\mathbf{x},\mathbf{u})} \quad OF(\mathbf{x}, \mathbf{u}, \mathbf{y}) \\
&\text{s.\,t.} \qquad\qquad f(\mathbf{x}, \mathbf{u}, \mathbf{y}) = 0 \\
&\qquad\qquad\qquad g(\mathbf{x}, \mathbf{u}, \mathbf{y}) \leq 0
\end{aligned}
\tag{2.50}
$$

The overall purpose is to minimise the objective function value under the worst uncertainty realisation within the robust set, but incorporating a corrective view after the uncertainty materialise. This comprises three levels, the first-level represents a planning strategy of the decision variable $\mathbf{x} \in \chi$ prior to the uncertainty realisation $\mathbf{u} \in U$ and

seeks to minimise the objective function value. The second level represents the uncertainty realisation $\mathbf{u} \in U$ within an uncertainty set Φ_u, in the worst possible manner and so it seeks to maximise the objective function value. The third level represents the operation actions to mitigate effects of uncertainty, once it realises. Decision variables at this level are represented by vector $\mathbf{y} \in Y(\mathbf{x}, \mathbf{u})$ which generally depends on both first decision variables x and uncertain parameter $\mathbf{u}$. Therefore, the model provides both ex-ante protection (planning decisions) and ex-post correction (operational decisions). On the computational side, this represents a try-level problem with require a specific solution algorithm that generally relies on merging the second and third level problem using duality theory and solving the resulting bi-level problem through a Column-and-Constraint Generation (CCG) algorithm [Ding et al., 2017].

Chapter 3

Power Losses Estimation in Low Voltage Smart Grids

3.1 Introduction

One of the European Union Targets was to replace at least 80% of all traditional energy meters with electronic-based smart meters by 2020 [European Comission, 2019]. However, at the end 2020, the European region (EU 27 including UK) was capable to replace only up to 150 million smart electricity meters, which corresponds to a penetration rate of 50% of smart meters. It is expected to surpass 227 million of smart meters (Households) in 2026 [Ostling, 2021] due to the updated planning and target numbers that will be carried out in many European markets such as Western and Northern Europe. This large-scale smart meters roll-out would not only allow service providers to gain information about the energy consumed and produced by each customer in real time but it would also allow them to compute network power losses at any given time. Network power losses of smart grids can be computed using customers' measurements (active and reactive energy) and the energy measurement measured by the low-voltage supervisor located at secondary substations. The more accurate the load demand data are, the more precisely power losses can be calculated. However, in some low-voltage networks, there are non-telemetered customers, who do not provide more information about the energy they consume or produce in real time. This complicates power losses calculations because this information is necessary for

estimating the load demand of non-telemetered customers. Moreover, in some instances, smart meter measurements do not have the required accuracy, the data they provide can be anomalous (null or extremely high) or the device can even be out of service. These challenges make it necessary to estimate load demand to perform any real-time power analyses and, especially, power losses calculations.

3.2   Network Modelling

Consider a low-voltage smart grid in terms of a graph [Jungnikel, 2009] composed by a set n nodes (or buses) $V = \{v_1, \ldots, v_n\}$ and a set of m edges (or lines) $E = \{e_{(k,j,p,\ell)} | \forall k, j \in V, k \neq j, \forall p \in \{a, b, c, n\}, \ell \in (1, \ldots, m)\}$. Since low-voltage networks are operated in radial configuration (no loops) the graph considered has a tree topology, normally composed by a mean feeder with laterals (Fig. 3.1).
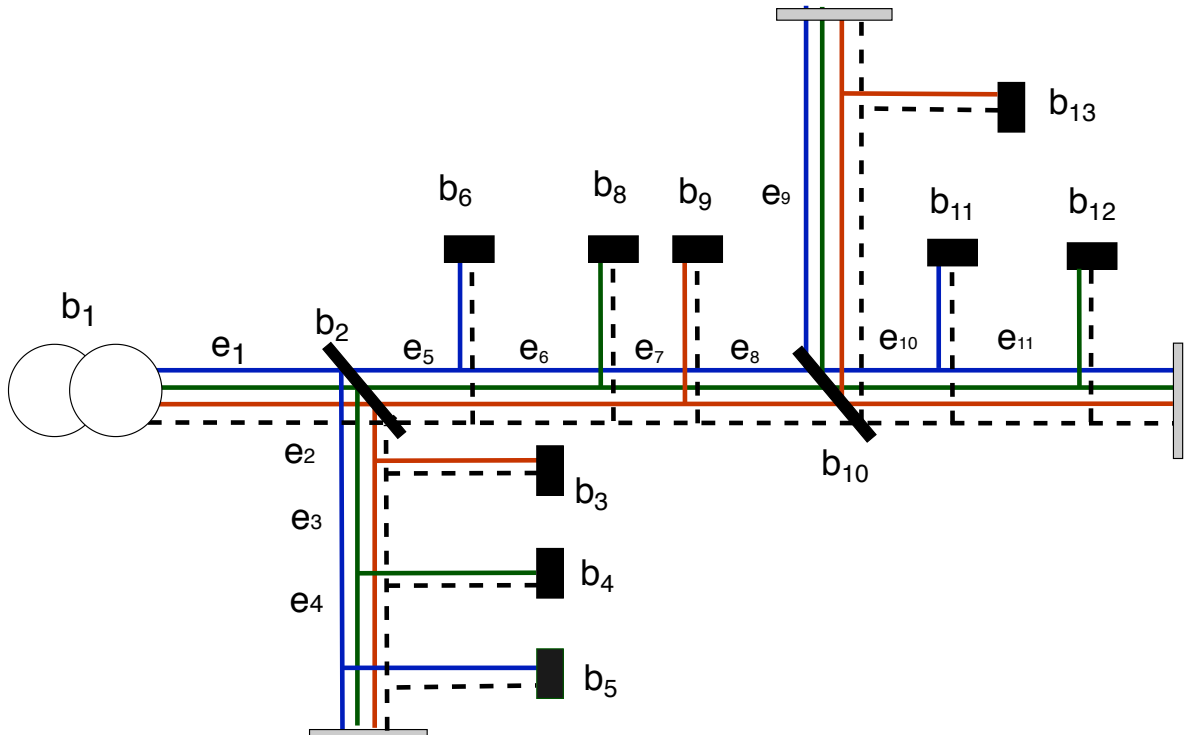
Figure 3.1: Low-voltage smart grid three-phase graph

Moreover, in these networks, the presence of single-phase loads is noteworthy, and so the system operation is performed, usually, in unbalanced conditions. Consequently, each

individual feeder can be defined as a four-wires line section, so the edge $e_{(k,j,p,\ell)}$ results in a four-tuple of the three phases plus the neutral cable $\{a, b, c, n\}$.

3.3 Network Equations

Consider the four-wires line section showed in Fig. 3.2 between sending node k and receiving node j. The relationship between the electrical magnitudes phase voltage, and phase currents and the self-impedance and mutual impedances is defined by the network equations. These equations are obtained by applying Ohm's law to all the phases as it is indicated in expression (3.1).
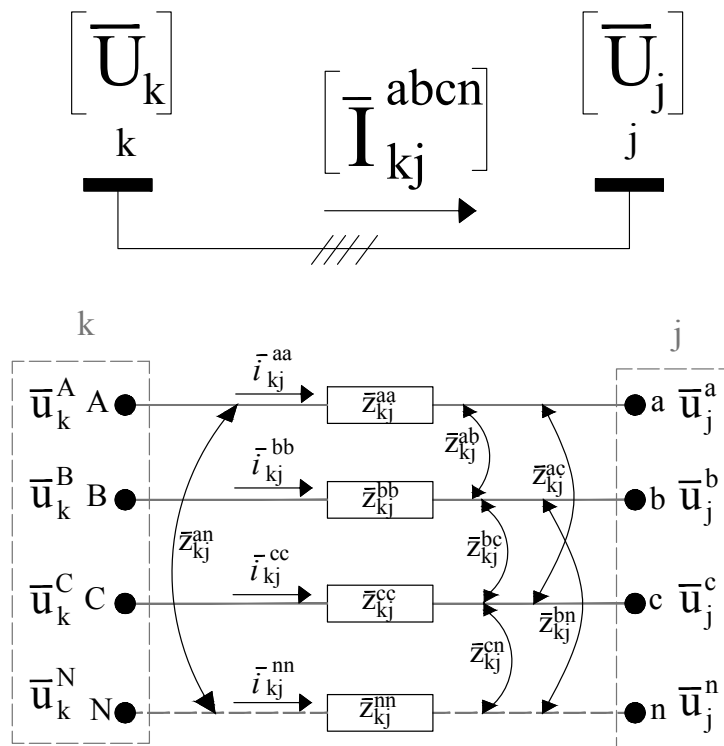
Figure 3.2: Section of four-wires feeder section model

$$
\begin{bmatrix} \bar{u}_k^A \\ \bar{u}_k^B \\ \bar{u}_k^C \\ \bar{u}_k^N \end{bmatrix} - \begin{bmatrix} \bar{u}_j^a \\ \bar{u}_j^b \\ \bar{u}_j^c \\ \bar{u}_j^n \end{bmatrix} = \begin{bmatrix} \bar{z}_{kj}^{aa} & \bar{z}_{kj}^{ab} & \bar{z}_{kj}^{ac} & \bar{z}_{kj}^{an} \\ \bar{z}_{kj}^{ba} & \bar{z}_{kj}^{bb} & \bar{z}_{kj}^{bc} & \bar{z}_{kj}^{bn} \\ \bar{z}_{kj}^{ca} & \bar{z}_{kj}^{cb} & \bar{z}_{kj}^{cc} & \bar{z}_{kj}^{cn} \\ \bar{z}_{kj}^{na} & \bar{z}_{kj}^{nb} & \bar{z}_{kj}^{nc} & \bar{z}_{kj}^{nn} \end{bmatrix} \begin{bmatrix} \bar{i}_{kj}^{aa} \\ \bar{i}_{kj}^{bb} \\ \bar{i}_{kj}^{cc} \\ \bar{i}_{kj}^{nn} \end{bmatrix}
\tag{3.1}
$$

The above expressions could be expressed in a more compact way by gathering the phase components, (3.2). By this way, it is obtained the phase impedance matrix (3.3) isolated from the neutral cable components (3.4) as well as the neutral cable impedance $\bar{z}_{kj}^{nn}$.

$$
\begin{bmatrix} [\bar{U}_k^{abc}] \\ \bar{u}_k^N \end{bmatrix} - \begin{bmatrix} [\bar{U}_j^{abc}] \\ \bar{u}_j^N \end{bmatrix} = \begin{bmatrix} [\bar{Z}_{kj}^{abc}] & [\bar{Z}_{kj}^n] \\ [\bar{Z}_{kj}^n]^T & \bar{z}_{kj}^{nn} \end{bmatrix} \begin{bmatrix} [\bar{I}_{kj}^{abc}] \\ \bar{i}_{kj}^{nn} \end{bmatrix}
\tag{3.2}
$$

Where:

$$
[\bar{Z}_{ij}^{abc}] = \begin{bmatrix} \bar{z}_{kj}^{aa} & \bar{z}_{kj}^{ab} & \bar{z}_{kj}^{ac} \\ \bar{z}_{kj}^{ba} & \bar{z}_{kj}^{bb} & \bar{z}_{kj}^{bc} \\ \bar{z}_{kj}^{ca} & \bar{z}_{kj}^{cb} & \bar{z}_{kj}^{cc} \end{bmatrix}
\tag{3.3}
$$

$$
[\bar{Z}_{kj}^n] = \begin{bmatrix} \bar{z}_{kj}^{an} & \bar{z}_{kj}^{bn} & \bar{z}_{kj}^{cn} \end{bmatrix}^T
\tag{3.4}
$$

$$
[\bar{U}_k^{abc}] = \begin{bmatrix} \bar{u}_k^A & \bar{u}_k^B & \bar{u}_k^C \end{bmatrix}^T
\tag{3.5}
$$

$$
[\bar{U}_j^{abc}] = \begin{bmatrix} \bar{u}_j^a & \bar{u}_j^c & \bar{u}_j^c \end{bmatrix}^T
\tag{3.6}
$$

$$
[\bar{I}_{kj}^{abc}] = \begin{bmatrix} \bar{i}_{kj}^{aa} & \bar{i}_{kj}^{bb} & \bar{i}_{kj}^{cc} \end{bmatrix}^T
\tag{3.7}
$$

Note that the phase impedance matrix (3.3) is complex and could be separated in real and imaginary part corresponding with the resistance R and reactance X matrix (3.8).

$$\left[\bar{Z}_{kj}^{abc}\right] = \left[R_{kj}^{abc}\right] + j\left[X_{kj}^{abc}\right] =$$

$$\begin{bmatrix} r_{kj}^{aa} & r_{kj}^{ab} & r_{kj}^{ac} \\ r_{kj}^{ba} & r_{kj}^{bb} & r_{kj}^{bc} \\ r_{kj}^{ca} & r_{kj}^{cb} & r_{kj}^{cc} \end{bmatrix} + j \begin{bmatrix} x_{kj}^{aa} & x_{kj}^{ab} & x_{kj}^{ac} \\ x_{kj}^{ba} & x_{kj}^{bb} & x_{kj}^{bc} \\ x_{kj}^{ca} & x_{kj}^{cb} & x_{kj}^{cc} \end{bmatrix} \tag{3.8}$$

It is assumed that all of the diagonal terms of the impedance matrix (3.3) are equal as well as all of the mutual impedances.

This is a good enough approximation for distribution grids which makes the problem more tractable [Short, 2004]. Therefore, the elements of the diagonal are replaced by the self-impedance term (3.9) and the elements out of the diagonal are equal and could be replaced by the mutual term (3.10).

$$\bar{z}_{kj}^{aa} = \bar{z}_{kj}^{bb} = \bar{z}_{kj}^{cc} = \bar{z}_{kj}^{s} \tag{3.9}$$

$$\bar{z}_{kj}^{ab} = \bar{z}_{kj}^{ac} = \bar{z}_{kj}^{ba} = \bar{z}_{kj}^{bc} = \bar{z}_{kj}^{ca} = \bar{z}_{kj}^{cb} = \bar{z}_{kj}^{m} \tag{3.10}$$

Since the majority of the low-voltage distribution models have expressed the network parameters in sequence terms [Short, 2004] the relationship between the phase impedance terms $\bar{z}_{kj}^{s}$ and $\bar{z}_{kj}^{m}$ and the sequence terms $\bar{z}_{kj}^{o}$ and $\bar{z}_{kj}^{1}$ are indicated in (3.11).

$$\bar{z}_{kj}^{s} = (\bar{z}_{kj}^{o} + 2\bar{z}_{kj}^{1})/3$$

$$\bar{z}_{kj}^{m} = (\bar{z}_{kj}^{o} - \bar{z}_{kj}^{1})/3 \tag{3.11}$$

Assuming that the network distribution network is well-grounded (that is, null voltage in neutral cable: $\bar{u}_{k}^{n} = \bar{u}_{j}^{N} = 0$) it is possible to apply Kron's reduction technique and working only with the phase components: {a,b,c}. Therefore, imposing the zero voltage condition to the neutral cable in equation (3.2), allows to obtain the expression (3.12) for the currents, which allows the clearing of the neutral cable current on function of the phase currents, expression (3.13).

$$\left[\bar{Z}_{ij}^{n}\right]^{T}\left[\bar{I}_{ij}^{abc}\right] + \bar{I}_{ij}^{nn}\bar{z}_{ij}^{nn} = 0 \tag{3.12}$$

$$\bar{I}_{kj}^{nn} = -\frac{1}{\bar{z}_{kj}^{nn}} \left[\bar{Z}_{kj}^{n} \right]^{T} \left[\bar{I}_{kj}^{abc} \right] \tag{3.13}$$

Therefore, the three phase components are used to calculate the network (3.14), and then, the current in the neutral cable is calculated by means of (3.14).

$$\begin{bmatrix} \bar{u}_k^A \\ \bar{u}_k^B \\ \bar{u}_k^C \end{bmatrix} - \begin{bmatrix} \bar{u}_j^a \\ \bar{u}_j^b \\ \bar{u}_j^c \end{bmatrix} = \left[\bar{Z}_{kj}^{abc} \right] \begin{bmatrix} \bar{i}_{kj}^{aa} \\ \bar{i}_{kj}^{bb} \\ \bar{i}_{kj}^{cc} \end{bmatrix} \tag{3.14}$$

Consider now the power injection at bus k depicted in Fig. 3.3 where $\left[\bar{I}_{i,k}^{abc} \right]$ is the column vector of net phase currents injections at node k (3.15), $\left[\bar{U}_{k'}^{abc} \right]$ is the column vector of phase voltages (3.16) of the node k' which belongs to the set of nodes connected to node k: $k' \in \Omega_k = \{1, \ldots, m\}$, and $\left[\bar{Y}_{kk}^{abc} \right]$ is the admittance matrix (3.17) of the three-phase line between nodes k and $k' \in V$.
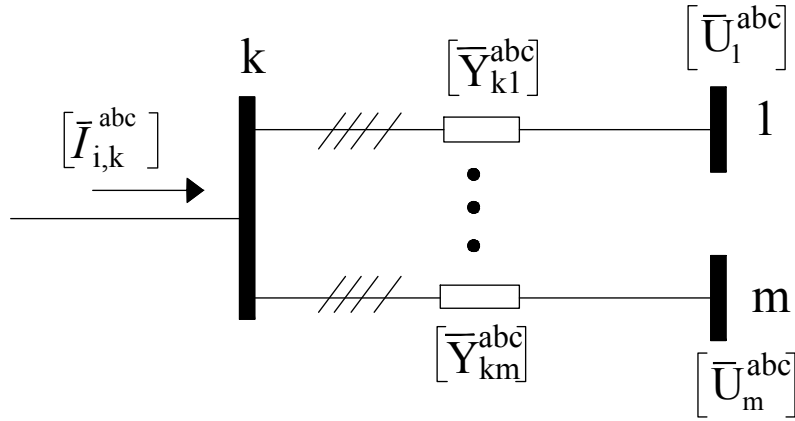
Figure 3.3: Network model injection

$$\left[\bar{U}_{k'}^{abc} \right] = \left[\bar{u}_{k'}^a \ \bar{u}_{k'}^b \ \bar{u}_{k'}^c \right]^{T} \tag{3.15}$$

$$\left[\bar{I}_{i,k}^{abc} \right] = \left[\bar{i}_{i,k}^a \ \bar{i}_{i,k}^b \ \bar{i}_{i,k}^c \right]^{T} \tag{3.16}$$

$$\left[\bar{Y}_{kk'}^{abc}\right] = \left[\bar{Z}_{kk'}^{abc}\right]^{-1} = \begin{bmatrix} \bar{y}_{kk'}^{aa} & \bar{y}_{kk'}^{ab} & \bar{y}_{kk'}^{ac} \\ \bar{y}_{kk'}^{ba} & \bar{y}_{kk'}^{bb} & \bar{y}_{kk'}^{bc} \\ \bar{y}_{kk'}^{ca} & \bar{y}_{kk'}^{cb} & \bar{y}_{kk'}^{cc} \end{bmatrix} \tag{3.17}$$

The net current injection at node k for each phase $p \in \{a, b, c\}$ could be calculated by applying Kirchhoff's current law [Sereeter et al., 2017] by means of the expression (3.18).

$$\bar{i}_{i,k}^{p} = \sum_{k' \in \Omega_k} \sum_{p' \in \{a,b,c\}} \bar{y}_{kk'}^{pp'} \bar{u}_{k'}^{p'}, \quad k' \neq k \tag{3.18}$$

By extending the above expression to all nodes that constitute the network $n = |V|$ and all phases $\forall p \in \{a, b, c\}$, the vector of net currents injections is obtained by means of the expression (3.19).

$$\begin{bmatrix} \left[\bar{I}_1^{abc}\right] \\ \vdots \\ \left[\bar{I}_b^{abc}\right] \end{bmatrix} = \begin{bmatrix} \left[\bar{Y}_{11}^{abc}\right] & \cdots & \left[\bar{Y}_{1b}^{abc}\right] \\ \vdots & \ddots & \vdots \\ \left[\bar{Y}_{b1}^{abc}\right] & \cdots & \left[\bar{Y}_{bb}^{abc}\right] \end{bmatrix} \begin{bmatrix} \left[\bar{U}_1^{abc}\right] \\ \vdots \\ \left[\bar{U}_b^{abc}\right] \end{bmatrix} \tag{3.19}$$

Where the complex admittance $\bar{y}_{kk'}^{pp'}$ could be separated in real and imaginary part, as it is done in expression (3.20) obtaining by this way the conductance $g_{kk'}^{pp'}$ and the susceptance $b_{kk'}^{bb'}$. These magnitudes could be calculated from the resistance and reactance $r_{kk'}^{pp'}$ and $x_{kk'}^{pp'} \ \forall p \in \{a, b, c\}$ as it is indicated in (3.20) and (3.21).

$$\bar{y}_{kk'}^{pp'} = g_{kk'}^{pp'} + j \cdot b_{kk'}^{pp'}, \quad k' \neq k \tag{3.20}$$

$$g_{kk'}^{pp'} = \frac{r_{kk'}^{pp'}}{(r_{kk'}^{pp'})^2 + (x_{kk'}^{pp'})^2}, \quad k' \neq k \tag{3.21}$$

$$b_{kk'}^{pp'} = \frac{-x_{kk'}^{pp'}}{(r_{kk'}^{pp'})^2 + (x_{kk'}^{pp'})^2}, \quad k' \neq k \tag{3.22}$$

Complex equations are decomposed in their real and imaginary parts such as phase voltage (3.23), current injection (3.24) and current flow (3.25).

$$\bar{u}_k^p = u_k^{p,re} + j \cdot u_k^{p,im} \tag{3.23}$$

$$\bar{i}_k^p = i_k^{p,re} + j \cdot i_k^{p,im} \tag{3.24}$$

$$\bar{i}_{kj}^p = i_{kj}^{p,re} + j \cdot i_{kj}^{p,im} \tag{3.25}$$

The network state is determined by solving the non-linear set of equations that constitute the unbalanced power flow $F(\mathcal{X}) = 0 : \mathbb{R}^n \to \mathbb{R}^n$ (3.26) that constitutes the power mismatch between the specified power injections (sp) and the calculated power injections (cal). This non-linear and complex equation is divided into real (active power) and imaginary (reactive power) components.

$$F(\mathcal{X}) = 0 = \begin{bmatrix} F_p(\mathcal{X}) \\ F_q(\mathcal{X}) \end{bmatrix} = \begin{bmatrix} p_{i,k}^{p,sp} + p_{i,k}^{p,cal} \\ q_{i,k}^{p,sp} + q_{i,k}^{q,cal} \end{bmatrix} \tag{3.26}$$

where $\mathcal{X}$ is the state variable, indicated in expression (3.27) that corresponds with the real and imaginary phase voltages $\{u_k^{p,re}, u_k^{p,im}\}$, $\forall k \in V$ $\forall p \in \{a, b, c\}$ indicated in expressions (3.28) and (3.29) where b is the number of buses of the distribution grid.

$$\mathcal{X} = \begin{bmatrix} [X_{re}] \\ [X_{im}] \end{bmatrix} \tag{3.27}$$

$$[X_{re}] = \begin{bmatrix} u_1^{a,re} u_1^{b,re} u_1^{c,re} \dots u_b^{a,re} u_b^{b,re} u_b^{c,re} \end{bmatrix}^T \tag{3.28}$$

$$[X_{im}] = \begin{bmatrix} u_1^{a,im} u_1^{b,re} u_1^{c,im} \dots u_b^{a,im} u_b^{b,im} u_b^{c,im} \end{bmatrix}^T \tag{3.29}$$

It can be noted that the calculated (cal) complex power injection in each node k and phase p is given by expression (3.30).

$$\bar{s}_{i,k}^{p,cal} = \bar{u}_k^p \left(\bar{i}_{i,k}^p \right)^* \tag{3.30}$$

Working in the above expression and introducing the equation of current injection (3.18) is possible to find an expression for the injected power in function of the lines' parameters and nodal voltages as shown in (3.31).

$$\bar{s}_{i,k}^{p,cal} = \bar{u}_k^p \sum_{j\in\Omega_k} \sum_{p'\in\{a,b,c\}} \left(\bar{y}_{kj}^{pp'}\right)\left(\bar{u}_j^{p'}\right)^* \tag{3.31}$$

Now, active and reactive power injections are obtained by extracting the complex power injection in real and imaginary part as indicates expression (3.32).

$$\bar{s}_{i,k}^{p,cal} = p_{i,k}^{p,cal} + j \cdot q_{i,k}^{p,cal} = Re\{\bar{s}_{i,k}^{p,cal}\} + j \cdot Im\{\bar{s}_{i,k}^{p,cal}\} \tag{3.32}$$

Active and reactive power injections are expressed in terms of the real and imaginary components of voltage and current injection as it is shown in expressions (3.33) and (3.34).

$$p_{i,k}^{p,cal} = u_k^{p,re} i_{i,k}^{p,re} + u_k^{p,im} i_{i,k}^{p,im} \tag{3.33}$$

$$q_{i,k}^{p,cal} = u_k^{p,im} i_{i,k}^{p,im} - u_k^{p,re} i_{i,k}^{p,im} \tag{3.34}$$

Where active and imaginary current injections $i_{i,k}^{p,re}$, $i_{i,k}^{p,im}$ are obtained from (3.24)-(3.25) and expressed in (3.35) and (3.36).

$$i_{i,k}^{p,re} = \sum_{j\in\Omega_k} \sum_{p'\in\{a,b,c\}} g_{kj}^{pp'} u_j^{p',re} - b_{kj}^{pp'} u_j^{p',im} \tag{3.35}$$

$$i_{i,k}^{p,im} = \sum_{j\in\Omega_k} \sum_{p'\in\{a,b,c\}} b_{kj}^{pp'} u_j^{p',im} + g_{kj}^{pp'} u_j^{p',im} \tag{3.36}$$

Finally, expression (3.37) and (3.38) are reorganised introducing (3.36) and (3.41) obtaining (3.39) and (3.40).

$$p_{i,k}^{p,cal} = u_k^p \sum_{j\in\Omega_k} \sum_{p'\in\{a,b,c\}} u_j^{p'} \left[g_{kj}^{pp'} \cos(\theta_{kj}^{p'}) + b_{kj}^{pp'} \sin(\theta_{kj}^{p'})\right] \tag{3.37}$$

$$q_{i,k}^{p,cal} = u_k^p \sum_{j\in\Omega_k} \sum_{p'\in\{a,b,c\}} u_j^{p'} \left[g_{kj}^{pp'} \sin(\theta_{kj}^{p'}) - b_{kj}^{pp'} \cos(\theta_{kj}^{p'})\right] \tag{3.38}$$

where $\theta_{kj}^{p'} = \theta_k^{p'} - \theta_j^{p'}$ is the difference between the voltage angles, being $\bar{u}_k^{p'} = u_k^{p'} \angle\theta_k^{p'}$ and $\bar{u}_j^{p'} = u_j^{p'} \angle\theta_j^{p'}$ the phasors of the phase voltage in the nodes of the line $k - j$.

On the other hand, the specified (sp) power injections are determined by the power balance between the positive power injected (generation) due to DER units, and the negative power injection (consumption) due to the customers demand, as indicates expression (3.39).

$$\bar{s}_{i,k}^{p,sp} = \bar{s}_{g,k}^{p} - \bar{s}_{d,k}^{p} \qquad \forall p \in \{a,b,c\}, \forall k \in V \tag{3.39}$$

Besides, the specified complex power injection $\bar{s}_{i,k}^{p,sp}$ is separated in active and reactive power injections as indicated in expressions (3.40) and (3.41).

$$p_{i,k}^{p,sp} = p_{g,k}^{p,sp} - p_{d,k}^{p,sp} \qquad \forall p \in \{a,b,c\}, \forall k \in V \tag{3.40}$$

$$q_{i,k}^{p,sp} = q_{g,k}^{p,sp} - q_{d,k}^{p,sp} \qquad \forall p \in \{a,b,c\}, \forall k \in V \tag{3.41}$$

The active and reactive power specified in each bus depends on the type of bus considered. To solve the power flow problem, one of the buses is chosen to be the $slack-node$ (denoted with k^*) and its power injections are equal to the net demand of the system including the total power losses. This slack node corresponds to the low-voltage side of the Secondary Substation, where the reference voltage is fixed to the rated value, which is the same of the reference voltage so $\bar{u}_k^p = 1\angle 0$ p.u. The remaining network buses $\forall k \in V/k^*$ are considered as PQ nodes where phase voltage $\bar{u}_k^p$ is variable and nodal power injections are due to the DER power injection. These DER units can inject power as shown in Table 3.1. Note that active power generation coming from DERs involve uncertainty since it depends on the weather conditions and so it requires to be estimated or forecasted. The same happen to the load demand of non-telemetered customers.

The network state (3.42) can be obtained by solving the non-linear systems equations indicated in (3.43) and (3.44).

$$\bar{s}_{kj}^{p} = \bar{u}_k^p \bar{i}_{kj}^p \qquad \forall p \in \{a,b,c\}, \forall k,j \in V, j \neq k \tag{3.42}$$

$$p_{kj}^p = Re\{\bar{s}_{kj}^p\} = u_k^{p,re} i_{kj}^{p,re} + u_k^{p,im} i_{kj}^{p,im} \qquad \forall p \in \{a,b,c\}, \forall k,j \in V, j \neq k \tag{3.43}$$

Table 3.1: Type of buses for power flow analysis

Bus Type	Voltage	Generation	Demand
Slack	Fixed	Variable	Variable
PQ: Demand & Gen.	Variable	Parameter	Parameter
PQ: Gen.	Variable	Parameter	0
PQ: Demand	Variable	0	Parameter

Note that demand and generation parameters are input data that can be forecasted based on smart meters measurement and weather measurements, or it can be considered as uncertain variables (i.e. parameters which value is not certain quantity).

$$q_{kj}^p = Im\{\bar{s}_{kj}^p\} = u_k^{p,im} i_{kj}^{p,im} - u_k^{p,re} i_{kj}^{p,re} \qquad \forall p \in \{a,b,c\}, \forall k,j \in V, j \neq k \qquad (3.44)$$

Where current flows $i_{kj}^{p,re}$ and $i_{kj}^{p,im}$ are calculated using Kirchhoff's current law (3.44) decomposed in active and imaginary parts (3.45)-(3.46).

$$i_{kj}^{p,re} = \sum_{p' \in \{a,b,c\}} g_{kj}^{pp'} \left(u_j^{p,re} - u_k^{p,re}\right) - b_{kj}^{pp'} \left(u_j^{p,im} - u_k^{p,im}\right) \qquad (3.45)$$

$$i_{kj}^{p,im} = \sum_{p' \in \{a,b,c\}} b_{kj}^{pp'} \left(u_j^{p,re} - u_k^{p,re}\right) + g_{kj}^{pp'} \left(u_j^{p,im} - u_k^{p,im}\right) \qquad (3.46)$$

Meanwhile the power flows from $j \to k$ are obtained by (3.47) and (3.48).

$$p_{jk}^p = (-1) \left(u_j^{p,re} i_{kj}^{p,re} + u_j^{p,im} i_{kj}^{p,im}\right) \qquad (3.47)$$

$$q_{jk}^p = (-1) \left(u_j^{p,im} i_{kj}^{p,im} + u_j^{p,re} i_{kj}^{p,re}\right) \qquad (3.48)$$

3.3.1   Power Losses Equations

Active (technical) power losses of the low-voltage smart grid can be calculated as the absolute value of the aggregation between the sending and receiving power in the other extreme of the feeder section as indicates (3.49).

$$P_{LOSS,T} = \sum_{e_{(k,j,p,\ell)} \in E} |p_{kj}^p + p_{jk}^p| \tag{3.49}$$

Alternatively, active power losses can also be calculated by means of current flow in each feeder and the feeder impedance: resistance r_{kj} (Ω/m) as indicated in (3.50).

$$P_{LOSS,T} = \sum_{e_{(k,j,p,\ell)} \in E} r_{k,j} \cdot \ell_{k,j} \cdot (i_{k,j}^p)^2 \tag{3.50}$$

3.3.2   Customers Load Demand

Customer's load demand have a dependency with the local bus voltage to which is connected [Stephen et al., 2014], [Onen, 2016], so customer's power demand is modelled by means of a ZIP model (3.51).

$$p_{d,k}^{p,sp} = \tilde{p}_{d,k}^p \left[c_{p,k}^{p,1} (u_k^p)^2 + c_{p,k}^{p,2} (u_k^p) + c_{p,k}^{p,3} \right] \quad , \forall k \in \Omega_C, \forall p \in \{a,b,c\} \tag{3.51}$$

where:

- u_k^p is the phase voltage magnitude in p.u.

- $\tilde{p}_{d,k}^p = \widetilde{P}_{d,k}^p / S_B$ is the active power demand of the customer in p.u.

- $c_{p,k}^{p,1}$, $c_{p,k}^{p,2}$, $c_{p,k}^{p,3} \in (0,1)$ Are the active power demand sensibility coefficients that determine the voltage dependency of the customer k connected to phase p. These coefficients can be fitted by running multiple power flows with different voltage levels.

Customers load demand $\tilde{p}_{d,k}^p$ is represented as time series based on the smart meters's measurements with a τ sampling time (ranging from 15 minutes to 1 hour). In (3.52) it is shown the power load demand from a customer $k \in \Omega_C$ in the period of time from $t = 0$ to $t = T$.

$$\tilde{p}_{d,k}^p = [p_{d,k,t=0} + p_{d,k,t=\tau} + \ldots + p_{d,k,t=T-\tau} + p_{d,k,t=T}] \tag{3.52}$$

If smart meters measurements are sampled at each hour, it is required to estimate the load demand at inferiors instants of time (for instance every 15 min).

Table 3.2: Example of time series smart meter measurements

Period	Timestamp	Demand (W) $p_{d,k}$
$t = 0$	2020-12-03 00:00:00	2.755
$t = \tau$	2020-12-03 00:15:00	2.876
$\vdots$	$\vdots$	$\vdots$
$t = T - \tau$	2020-12-03 13:10:00	2.948
$t = T$	2020-12-03 13:25:00	2.876

$\tau = 15\text{min}; T = 24\text{ h} \cdot 60\text{ min/h} = 1440\text{ min}$

For demand response actions it is interesting to know the power demand of individual household appliances. Total customer demand $(P_{D,i})$ (3.53) can be separated in power demand dependent on temperature $p_{d,k,T}^{p}$ (such a Air Conditioner (AC), Electrical heaters) and the remaining household appliances's demands (such as oven, computers, TV, fridge, hot water, lightning, etc.) $p_{d,k,NT}^{p}$, which do not depend on the temperature.

$$p_{d,k}^{p} = p_{d,k,NT}^{p} + p_{d,k,T}^{p} \tag{3.53}$$

$p_{d,k,T}^{p}$ depends on the outdoor household temperature T_{ext} and the indoor temperature T_{int} as indicated in Fig. 3.4 where $T_{int} \in (T_{max}, T_{min})$.

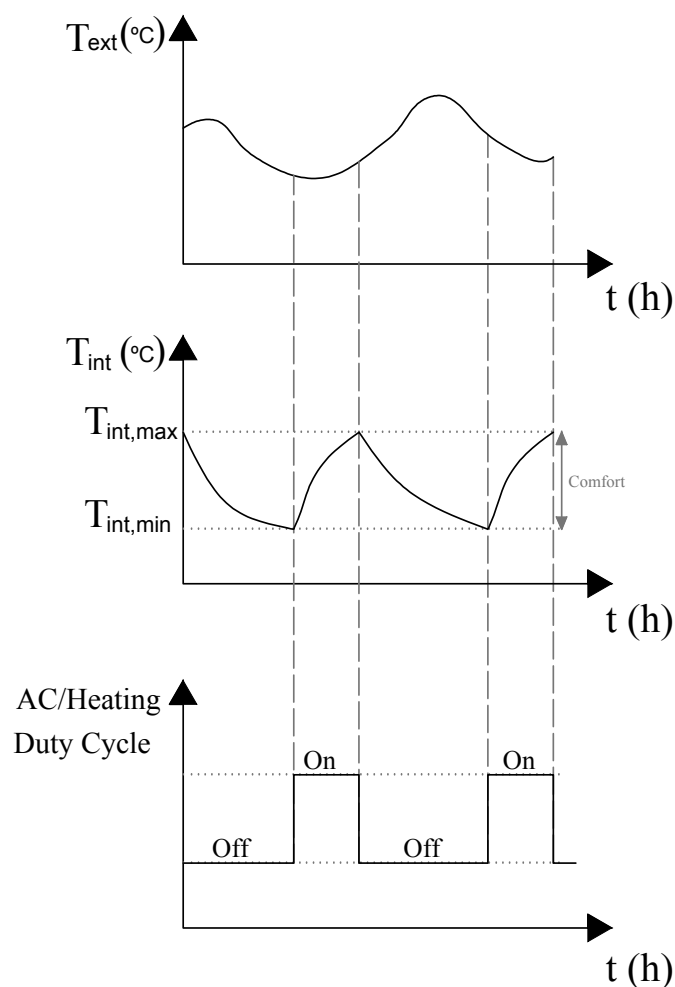$$p_{d,k,T}^{p} = f_c(T_{ext}, T_{int}) \tag{3.54}$$

Figure 3.4: Relationship between the heating and cooling devices operation with the outdoor temperature and indoor temperature

3.3.3   Distributed Energy Resources (DERs)

DERs include:

1. Photovoltaic (PV) Panels.

2. Plug-in Electric Vehicles (PEV).

3. Battery Energy System (BES).

Photovoltaic Panels

PV panels generate power depending on the solar irradiance G (W/m^2) in the inclined plane, ambient temperature T_{amb} as well as the orientation and inclination angle. The variable power output of each PV facility is indicated in expression (3.55) [Lorenzo et al., 1994].

$$p_{g,k,t}^{p,sp} = \tilde{p}_{pv,k}^p = \frac{P_{PV,k}}{S_B} \cdot \frac{\widehat{G}_k}{1000} \left\{ 1 + \delta_{PV}(T_{cell} + 25) \right\}, \quad \forall k \in \Omega_{PV} \qquad (3.55)$$

where:

- P_{PV_k} is the peak power output of the PV facility (number of PV panels multiplied by the their peak power) in Standard Test Conditions (STC) which means under a irradiation of $G = 1000 \ W/m^2$ and ambient temperature of $T_{amb} = 25°C$.

- $\widehat{G}_k$ is the solar irradiance (W/m^2) that receives the PV panel of the customer k on the inclined plane. Notice that this component has a hat which indicates that this variable is forecasted or estimated.

- δ_{PV} is the temperature coefficient of power $(\%/°C)$.

- $\widehat{T}_{cell,k}$ is the temperature $(°C)$ of the panel surface witch depends on the ambient temperature $\widehat{T}_{amb,k}$ and the solar irradiance $\widehat{G}_k$ and the Nominal Operating Cell Temperature (NOCT) $(°C)$ (3.56)

$$\widehat{T}_{cell,k} = \widehat{T}_{amb,k} + \widehat{G}_k \cdot \left(\frac{NOCT - 20}{800} \right), \quad \forall k \in \Omega_{PV} \qquad (3.56)$$

Note that solar irradiation $\widehat{G}_k$ and ambient temperature $\widehat{T}_{amb,k}$ use hat notation $\widehat{}$ because are two weather variables that have to be previously forecasted.

PV panels are connected to the grid through a PV inverter DC/AC (with rated apparent power S_{iv}) and they can be used to provide flexibility by means of PV energy curtailment. This represent one of the mechanism to provide load flexibility in low-voltage distribution networks with high penetration of distributed resources, and is known as PV production curtailment. This is based in modify the power factor $\cos(\phi_{iv})$ of the inverter devices AC/DC that connects the PV panel arrays with the network in such a way that

the total active power injected in the network is lower than the rated capacity of the inverter (Fig. 3.5) and the active power injected to the grid can be reduced (3.57) at the expense of the injection of reactive power (3.57)-(3.58). In any case, the operating points of PV inverters have to guarantee (3.59) as illustrated in Fig. 3.5.

$$\tilde{p}^p_{pv,k} = \cos(\phi_{iv}) \cdot S_{iv}, \quad \forall k \in \Omega_{PV} \tag{3.57}$$

$$\tilde{q}^p_{pv,k} = \sin(\phi_{iv}) \cdot S_{iv}, \quad \forall k \in \Omega_{PV} \tag{3.58}$$

$$(\tilde{p}^p_{pv,k})^2 + (\tilde{q}^p_{pv,k})^2 \leq (S_{iv})^2 \ , \forall k \in \Omega_{PV} \tag{3.59}$$

Where ϕ_{iv} is the power factor of the inverter which usually vary between $0.95 \leq \cos(\phi_{iv}) \leq 1$.
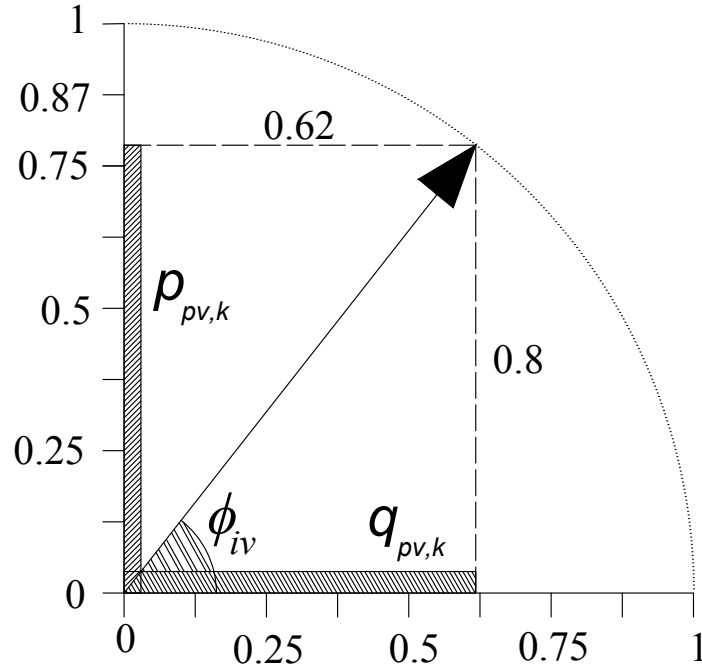

Figure 3.5: PV Inverter limits

Plug-in Electric Vehicles (PEV)

Plug-in Electric Vehicles (PEV) are vehicles that incorporate a Battery Energy System (BES) with a capacity that range between 50 and 100 kWh. Each PEV unit can be connected at the household's connection point through a single-phase or three-phase charging station (depending on the charging power). Additionally, PEVs can be connected to public charging stations, however in this thesis, only household' connection points are considered. PEV units are modelled in such a way that arrivals and charging sessions duration have an stochastic component and depends on the type of the day (weekdays, weekends, holidays or special events) and the time of the day, according to [Flammini et al., 2019]. In Fig. 3.6 it is shown the time evolution of the PEV charging session and the uncertainty associated.

The Fig. 3.6 illustrates the expected value of the charging power demanded by the PEV in real magnitude and the extreme bounds delimited by one standard deviation (light grey) and two standard deviations (darker grey), which means that 68.27% of the trajectories will fall within the light grey delimited area, and 99.73% of the charging pattern will fall within the aggregation of the two areas.

To model the PEV charging pattern, the expression indicated in (3.60) is formulated, which includes an expected value term ($\mu_{k,\ PEV,t}^{p}$) and a deviation term ($\sigma_{k,\ PEV,t}^{p}$). Both terms are characterised by a Probability Density Function (PDF) ($\widehat{f_h}$) defined in (3.61) according to the Kernel Density Estimation (KDE) method.

$$p_{k,PEV,t}^{p} = \mu_{k,PEV,t}^{p} + \sigma_{k,PEV,t}^{p}, \quad \forall t \in T_{ch}, \forall k \in \Omega_{PEV} \tag{3.60}$$

$$p_{k,PEV,t}^{p} \approx \widehat{f}_{PEV}(v) = (nh)^{-1} \sum_{i \in |v|} \phi((v - v_i)/h) \tag{3.61}$$

where: v is the variable to be represented (in this case the PEV charging pattern, which is a set of data points of charging sessions), $\phi(\bullet)$ is the gaussian function, v_i is the ith value of v, h is the smoothing parameter indicated in expression (3.62).

$$h = \left(\frac{4\ \sigma^5}{3\ |v|} \right)^{1/5} \tag{3.62}$$

where: σ is the standard deviation of the PEV charging patterns. The expected value component of the PEV demand correspond with the nominal value of the PEV charging point during the charging session T_{ch}.
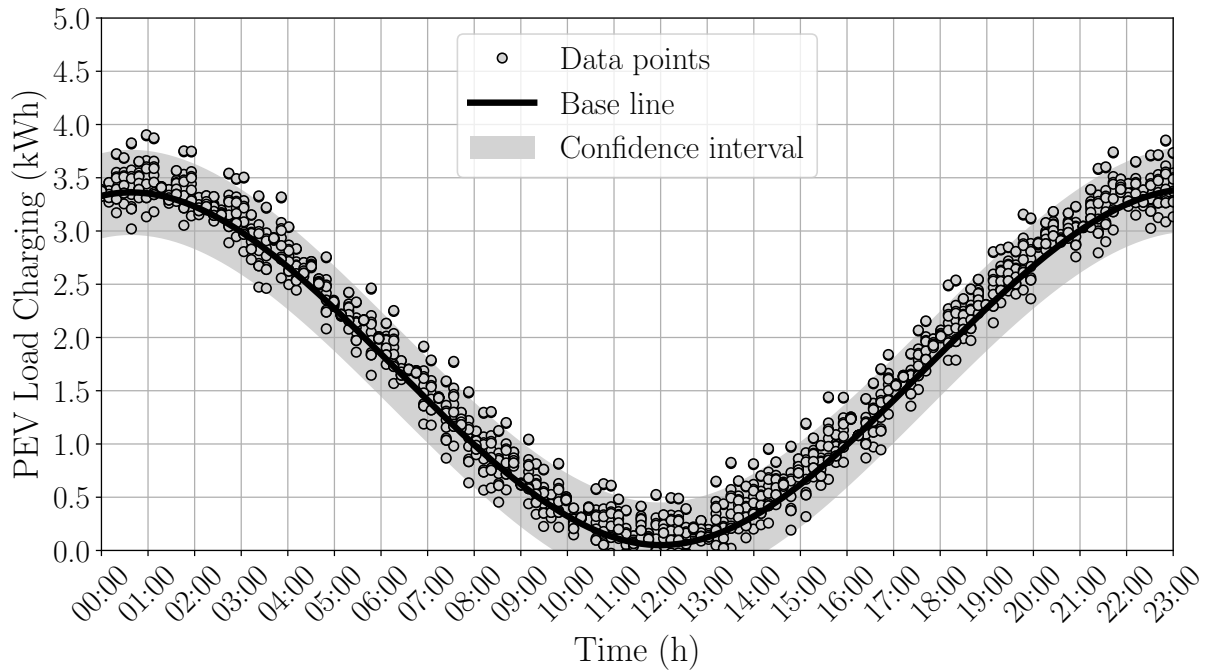
Figure 3.6: PEV charging load pattern

3.3.4   Load Demand Modelling: Non-Telemetered Customers

To calculate technical power losses in low-voltage smart grids is required the customers load demand, which is obtained from the customer's smart meter measurements (active and reactive energy) and the energy measurement registered by the low-voltage supervisor's smart meter deployed at secondary substations. However, the resolution of those measurements not always is as lower as it would be desired (normally 1 hour) and in some low-voltage networks, some customers meter report energy consumption in monthly basis. This fact makes more challenging the process of calculation of the technical power losses since an inaccurate load demand will lead to wrong power losses estimation. To tackle this issue, in this section is presented a stochastic approach for the modelling of the load demand under uncertainty (e.g., non-telemetered customers and uncertain smart meters

readings). Load demand of non-metered customers is modelled by means of a top-down approach. Intra-hour load demand profiles of customers are synthetically generated by applying a Markov process. Finally, a case study is conducted to provide numerical results of the framework proposed. The data and network used in the case study corresponds to the roll-out deployed by the Spanish R&D (Research and Development) demonstration project OSIRIS [UC3M et al., 2017].

In low-voltage smart grids there are telemetered customers for whom smart meters provide hourly measurements. There are also non-telemetered customers, who have monthly energy meters. Moreover, in some situations, smart meters from telemetered customers fails (not measurement) or they can provide false readings. Consequently, the hourly load demand from non-telemetered and telemetered customers with missing measurements has to be estimated. Non-telemetered customers have energy meters that provide measurements of the energy consumed during the most recent months. For each non-telemetered customer, historical monthly energy measurements are used to infer the hourly energy consumption by applying a top-down approach using three levels of resolution:

- Upper Level: Monthly Energy Consumption.

- Middle Level: Daily Energy Consumption.

- Lower Level: Hourly Energy Consumption.

Upper-level: Monthly Energy Consumption

Available data from non-telemetered customers consist of historical monthly energy values measured during the most recent months. This information for energy consumption during the current month is typically lacking because it is collected once the month ends. To estimate the current month's energy consumption, an Energy Consumption Tendency curve (ECT) curve is deduced using the historical monthly measurements. The monthly energy consumed from the previous month (last available measurement) to the current month (still unknown) is inferred through an interpolation process based on polynomial functions (splines) [Massopust, 2010]. The ECT for the energy consumption of a non-telemetered customer is shown Fig. 3.7.
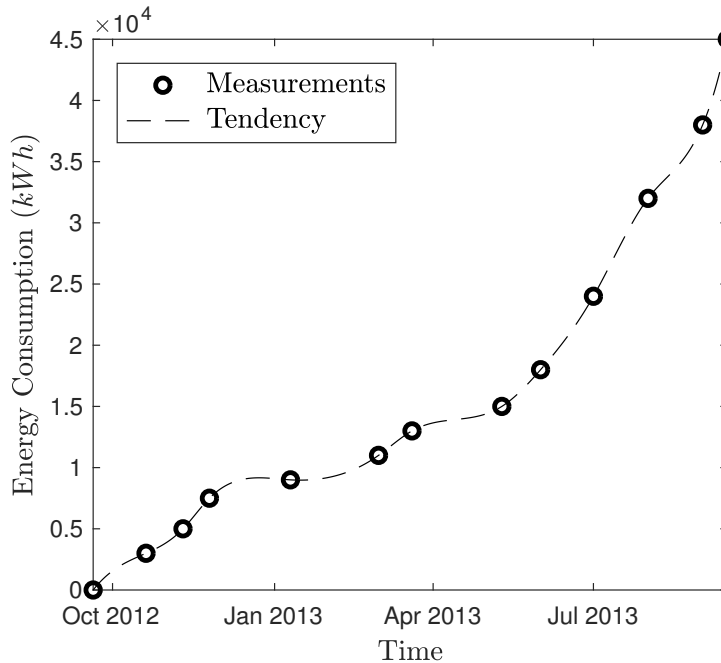
Figure 3.7: ECT curve for the historical measurements of a non-telemetered customer from September 2012 to September 2013.

Middle-level: Daily Energy Consumption

Once the energy estimation for the current month is obtained, it is possible to estimate each week's consumption. This weekly estimation process is necessary in order to differentiate the energy consumption of the working days from holidays. The weekly profile is deduced from the weekly consumption registered by the smart meter located at the secondary substation. This secondary substation weekly profile is considered the reference Weekly Energy Consumption (WEC) profile for the low-voltage network. In Fig. 3.8, the estimated WEC profile for a non-telemetered customer is shown. The first three points correspond to holidays, the next five points correspond to the work days and the last two points correspond to Saturday and Sunday. The WEC therefore provides the daily energy consumption for each day of the week.

Lower Level: Hourly Energy Consumption

At this level, the complete set of hourly values that are missing is estimated using an optimisation process. The procedure proposed is formulated as a Non-Linear Programming
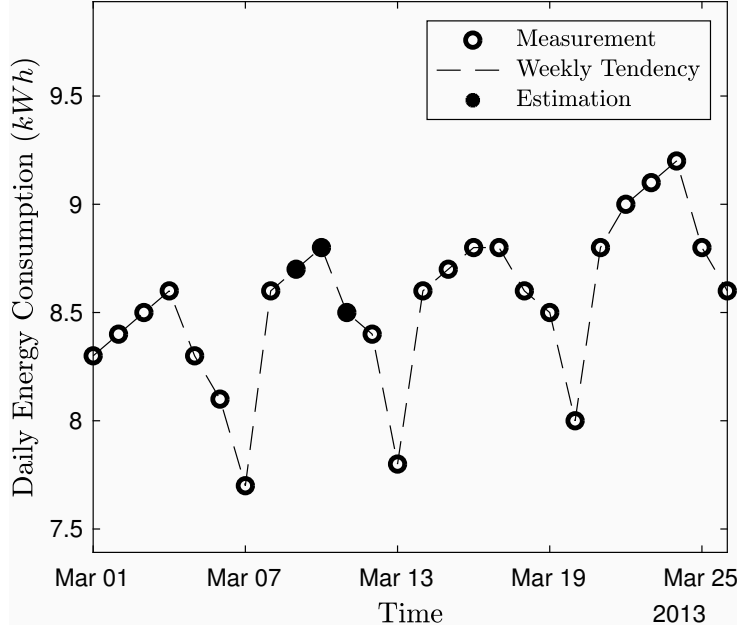
Figure 3.8: WEC profile of a non-telemetered customer in May 2013.

(NLP) optimization problem as described below in (3.63)-(3.66).

Minimise the objective function:

$$Min \quad z = \left\{ \sum_{i \in \Omega_C} \sum_{d \in \Omega_d} \sum_{t \in T_d} \left(P_{i,t,d} - \widehat{P}_{i,t,d} \right)^2 \right.$$
$$+ \sum_{i \in \Omega_{c'}} \sum_{d \in \Omega_d} \left(E_{i,d} - \widehat{E}_{i,d} \right)^2$$
$$\left. + \sum_{d \in \Omega_d} \sum_{t \in T_d} \left(\sum_{i \in \Omega_c} P_{i,t,d} - \widehat{P}^S_{t,d} \right)^2 \right\} \tag{3.63}$$

Subject to:

$$\sum_{i \in \Omega_C} \sum_{t \in T_d} P_{i,t,d} = \sum_{i \in \Omega_c} E_{i,d}; \quad \forall d \in \Omega_d \tag{3.64}$$

$$\sum_{i \in \Omega_C} \sum_{t \in T_d} P_{i,t,d} \leq \sum_{h=1}^{24} P_{s,t,d}; \quad \forall d \in \Omega_d \tag{3.65}$$

$$\sum_{i \in \Omega_c} E_{i,d} \leq \widehat{E}^S_d; \quad \forall d \in \Omega_d \tag{3.66}$$

Where $T_d = (t_1, \ldots, t_n)$ is the daily time horizon. The objective function (3.63) is composed of three terms. The first term represents the total square error between

the real hourly demand $(P_{i,t,d})$ and the hourly measurement $(\widehat{P}_{i,t,d})$ for the 24 hours of the day d. The second term refers to the square error between the real daily energy consumption $(E_{i,d})$ and the measured daily value $(\widehat{E}_{i,d})$. The third term in the equation refers to the square error related to the daily energy demanded from the complete set of network customers $(\sum_{i=\Omega_C} P_{i,t,d})$ and the daily energy registered at the secondary substation $(\widehat{P}^S_{t,d})$. Constraint (3.64) equalises the 24-hour power consumption with the daily energy consumption by each customer. Constraint (3.65) establishes that the 24-hour power consumption registered at the secondary substation has to be greater than the power demanded by the network customers. Finally, constraint (3.66) establishes the energy constraint for consumption throughout the day.

Intra-hour load demand modelling

The modelling of intra-hour load demand for network customers is modelled as a stochastic Markov process [Howard and Karlin, 1984]. Therefore, high-resolution load demand customer profiles are modelled as a discrete time Markov chain (X_t), where t is the parameter running over an index set (T) that corresponds to discrete units of time $T = (t_0, t_n)$ (where t_0 starts the hour and t_n finishes the hour), and the value that takes (X_t) is the state of the Markov process [Pinsky and Karlin, 2011]. The Markov chains are characterised by the property (3.67), which defines the probability of shifting from the state $X_t = i$ to the state $X_{t+1} = j$ in t discrete units of time, which only depends on the previous state.

$$
\begin{aligned}
\mathbf{p}^{(t)}_{i,j} &= Prob\Big(X_{t+1} = j | X_t = i\Big) \\
&= Prob\Big(X_{j+1} = j | X_t = i, X_{t-1} = i_{t-1}, \ldots, X_0 = i_0\Big)
\end{aligned}
\tag{3.67}
$$

The probability function described in (3.68) is also called a one-step transition probability. To model a complete Markov process of k states, a transitional probability matrix $\mathcal{P}$ had to be built, which constitutes a first-order Markov chain. This considers the current state and that the probability of the next state. Since discrete time steps are considered (e.g., 1 minute), the Markov chain is homogenous so that the entry $\mathbf{p}^{(t)}_{i,j}$ of the matrix $\mathcal{P}$ provides the probability of shifting from state i to j.

$$\mathcal{P} = \begin{pmatrix} \mathbf{p}_{1,1} & \cdots & \mathbf{p}_{i,k} \\ \vdots & \ddots & \vdots \\ \mathbf{p}_{k,1} & \cdots & \mathbf{p}_{k,k} \end{pmatrix} \in \mathbb{R}^{k \times k} \tag{3.68}$$

The ith row is the probability distribution of the values of X_{t+1} under the condition $X_t = i$. The number of rows of matrix $\mathcal{P}$ determines the number of states in the Markov process. Note that the sum of all values in a single row (a group of states) is equal to the unit.

A first-order Markov process is set up using the hourly load demand probability functions for the hourly measurement, and the data are estimated from the low-voltage network customers's measurements. A transitional probability matrix $\mathcal{P} \in \mathbb{R}^{60 \times 60}$ (with 1-min. resolution) is used to model intra-day and high-resolution load demand profiles. The first state of the Markov chain sequence is chosen by means of a random variable that selects values in the interval (0,1) [McLoughlin et al., 2010]. Next, the matrix $\mathcal{P}$ is applied to determine the next states.

Load demand probability density functions (PDFs)

To determinate the entries $\mathbf{p}_{i,j}$ for the transitional probability matrix $\mathcal{P}$, the Probability Density Functions (PDFs) and the Cumulative Distribution Functions (CDFs) have to be determined. PDF and CDF are denoted with $f(x)$ and $F(x)$, respectively, and their definitions are provided in (3.69) and (3.70) (where X denotes the Markov chain). This describes the relationship between them [D. C. Montgomery, 1999].

$$Prob\left(i \leq X \leq j\right) = \int_i^j f(x)dx = F(i) - F(j) \tag{3.69}$$

$$Prob\left(X \leq j\right) = \int_{-\infty}^j f(x)dx = F(j) \tag{3.70}$$

Due to the difficulty associated with selecting a theoretical parametric distribution (normal, log-normal, etc.), the Kernel Density Estimation (KDE) [D. C. Montgomery, 1999] is chosen to carry out the fitting procedure. Recently, in the scientific literature, other authors have used KDE to fit PDFs and CDFs to the continuous operational variables

of the network, such as load demand in [Hyndman and Fan, 2010] or wind speed in [Xu et al., 2015] and [Miao et al., 2016]. Therefore, to avoid making assumptions about the statistical distribution of the load demand data, the KDE method is applied. The non-parametric representation of the PDF of the load demand variable S_d is defined by means of a smoothing function (3.71) as follows:

$$\hat{f}(\tau) = \frac{1}{n\hat{h}} \int_1^{24} \varphi(\tau) \left(\frac{\tau - S_d(t)dt}{\hat{h}} \right) \tag{3.71}$$

Where $\hat{f}(\tau)$ is the estimated density function, τ is the time in hours, $\varphi(\tau)$ is the Gaussian kernel (a standard normal probability density function [Hastie and Friedman, 1991]) and $\hat{h}$ is the bandwidth (a positive smoothing parameter). One of the criteria used to obtain the value for $\hat{h}$ is the Mean Integrated Squared Error (MISE) (3.72).

$$MISE(\hat{h}) = E \left[\int \left(y(\tau) - y^{'}(\tau) \right)^2 d\tau \right] \tag{3.72}$$

Additionally, $y^{'}(\tau)$ is the unknown real density function. Due to that fact, the MISE formula cannot be applied in a straightforward fashion. Therefore, the most popular solution for using an appropriate value for $\hat{h}$ is to apply the Silverman's rule of thumb [Silverman, 1986] by means of (3.73).

$$\hat{h} = \left(\frac{4\hat{\sigma}_{S_d}^5}{3n} \right)^{1/5} \approx \frac{\hat{\sigma}_{S_d}1.06}{n^{1/5}} \tag{3.73}$$

Where $\hat{\sigma}_{S_d}$ is the standard deviation of the variable S_d.

3.4   Case study

3.4.1   Description

The data and network used in process correspond to the roll-out deployed for the Spanish R&D demonstration project OSIRIS (*Optimizacion de la Supervision Inteligente de la Red de DIstirbuticion*), which is an innovative project to develop knowledge, tools and new equipment for optimising the supervision of low-voltage Smart Grids [UC3M et al., 2017].

The project scenario concerns a primary substation that supplies power to $25,849$ customers in a region located in the south of Madrid (Spain). In Fig. 3.9 one of the low-voltage distribution networks in the OSIRIS project is shown. The 680 kVA network has eight low-voltage residential and commercial feeders with 32 load consumption points. There are 22 telemetered customers (individual contractual power $\leq$ 15 kW) and 10 non-telemetered customers with a total contractual power of 442 kW. Data related to cables electrical properties (resistance and reactance) considered are indicated in Table A.1.
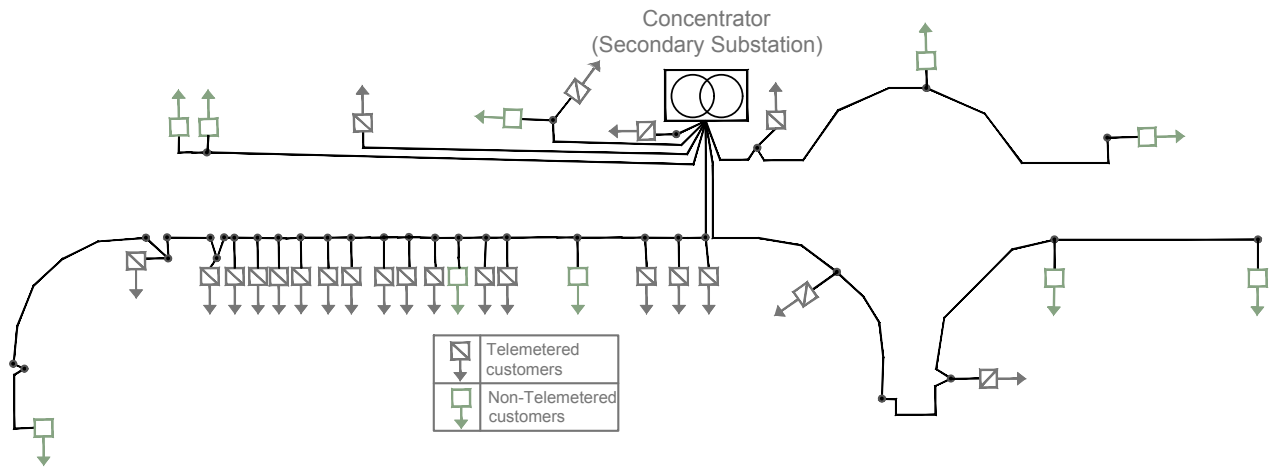
Figure 3.9: OSIRIS real low-voltage smart grid under study

3.4.2 Data

Smart Metering Infrastructure

In the OSIRIS demonstration project, the communication infrastructure between smart meters and the Meter Data Management System (MDMS) has been designed specifically for low-voltage networks. It is based on PoweRline Intelligent Metering Evolution (PRIME) [López et al., 2015], (Fig. 3.10). Smart meters are associated with telemetered customers, whereas the concentrators are located at the secondary substations. Concentrators send their data up to the so-called gateway, which manages communications with the MDMS using the medium voltage infrastructure as a communication medium. A much more detailed explanation of the communication architecture deployed in the OSIRIS project can be found in [López et al., 2015].
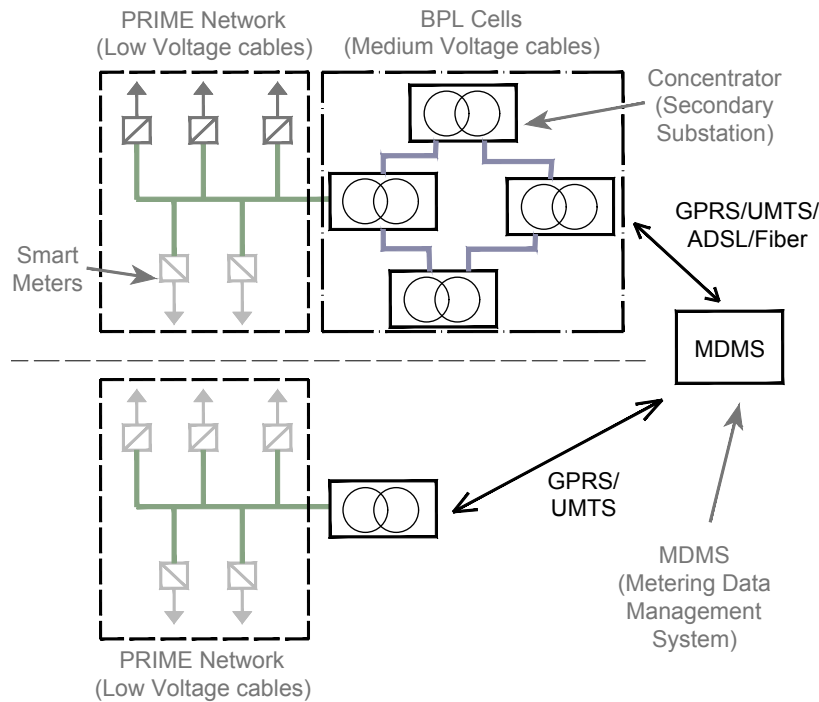
Figure 3.10: Communications layer illustration

Smart Meters Data Acquisition

Each telemetered customer has an individual smart meter that registers the active load energy (imported and exported), in addition to the reactive load demand (four quadrants). This provides hourly and daily measurements. A smart meter supervisor is located at the secondary substation. In this study, smart meter data from all the metered customers have been collected by the concentrator of one of the seven networks operated by the OSIRIS project from September 2013 to September 2014.

Statistical study of load demand data

A statistical study of the hourly demand consumption of the OSIRIS network is shown in Fig. 3.10. The histograms describe power demand across different daily time-slots. During the first hours of the day (01:00 and 06:00) and the last hours (20:00 and 22:00), load demand is concentrated around a certain value and it displays a clear skewness. Meanwhile, for the rest of the hours of the day (10:00 to 18:00), the load demand is spread in a more heterogeneous fashion. Fig. 3.11 shows the PDFs (left) and CDF (right)

distributions obtained by the KDE method, respectively.
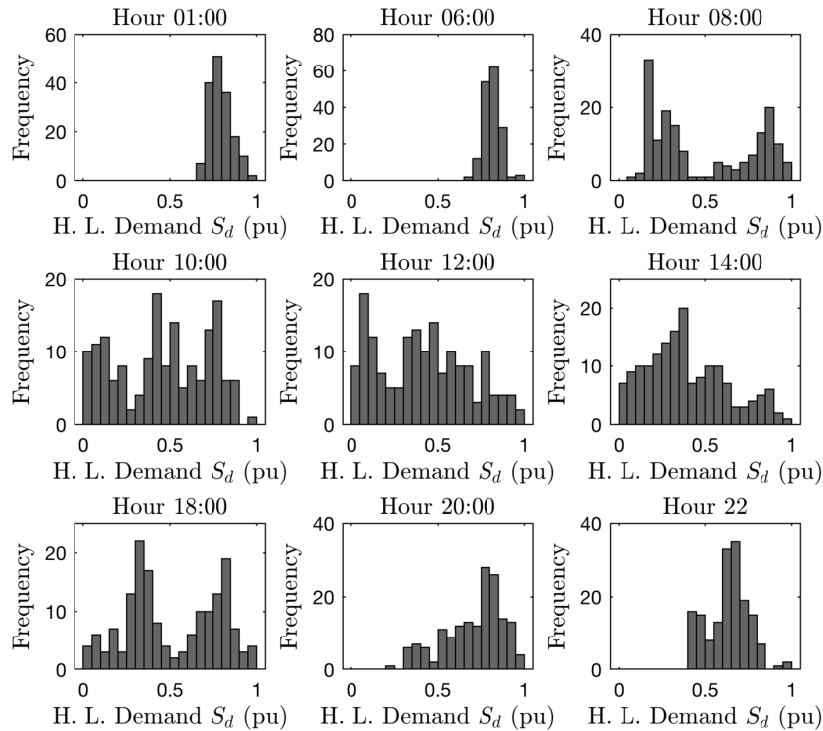
Figure 3.11: Hourly load demand histograms

In Fig. 3.12-left is shown the daily energy consumption estimated for a particular customer and it can be notice the points where energy demand estimation has been carried out with the model (3.63)-(3.66). In Fig. 3.12-right is shown the hourly load demand for the above customer and also it can be seen the estimated load demand points as well as the working load pattern used.

For each customer (telemetered and non-telemetered), a stochastic high-resolution load demand has been synthetically generated (Fig. 3.13-left). The initial point corresponds to 9:00 AM, and the final point corresponds to 10:00 AM.

These points are known using the hourly load demand. Between these two hours, five Markov realisations are presented as indicates Fig. 3.14-left. Fig. 3.14-right shows the linear regression calculated for relative TLs. Points of the relative TLs for every stochastic realisation are concentrated in the middle of the cloud. Moreover, the bandwidth error
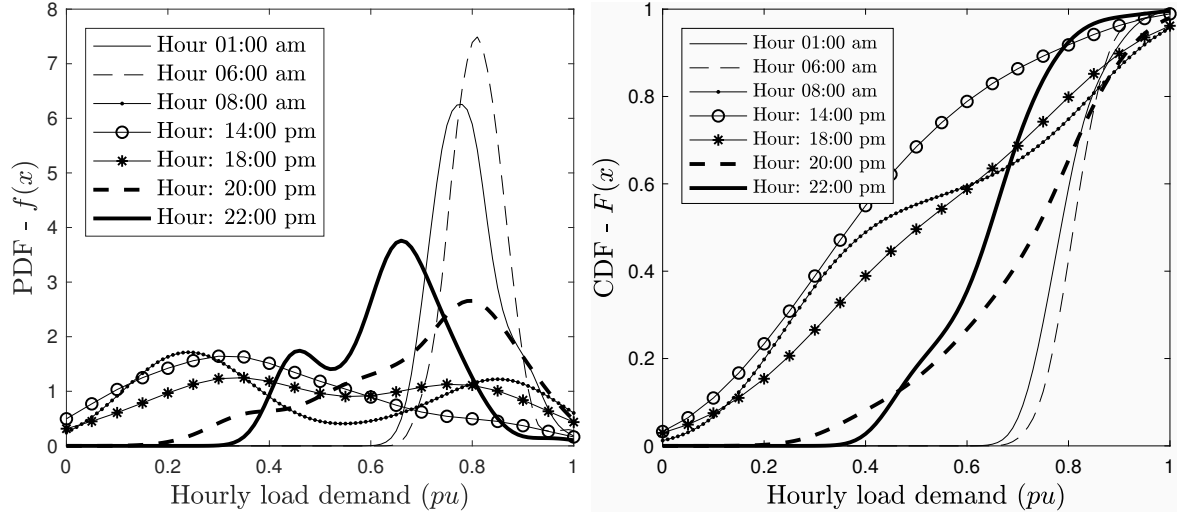
Figure 3.12: Left: PDFs obtained for hourly load demand; Right: CDFs obtained for hourly load demand.

represents the variability with respect to the regression model prediction.

3.4.3   Results

Power Losses Estimation

Technical power posses have been calculated by the three-phase unbalanced power flow with the full three-phase network model (explained in section 3.3 of this chapter) by solving the state equation (3.26) for every stochastic high-resolution load demand profile synthetically generated. Technical active power losses $P_{LOSS,T}$ (3.50) are expressed as a fraction of the load demand S_D (2.6) obtaining the relative active power losses $\mathcal{L}_p$ defined previously in (2.9).

Fig.3.15-left shows the box-plot representation of the stochastic TLs for each percentage level of loading capacity at the secondary substation. Fig. 3.15-right shows the PDF and CDF functions through KDE for the stochastic TLs obtained by means of the Markov process. For this specific OSIRIS network, hourly TLs do not exceed 2% with a high density in the interval 0-1.5 %.

The median values (middle line of the box) are distributed according to a linear evolution [D. C. Montgomery, 1999]. Thus, linear regression is proposed as a method to fit
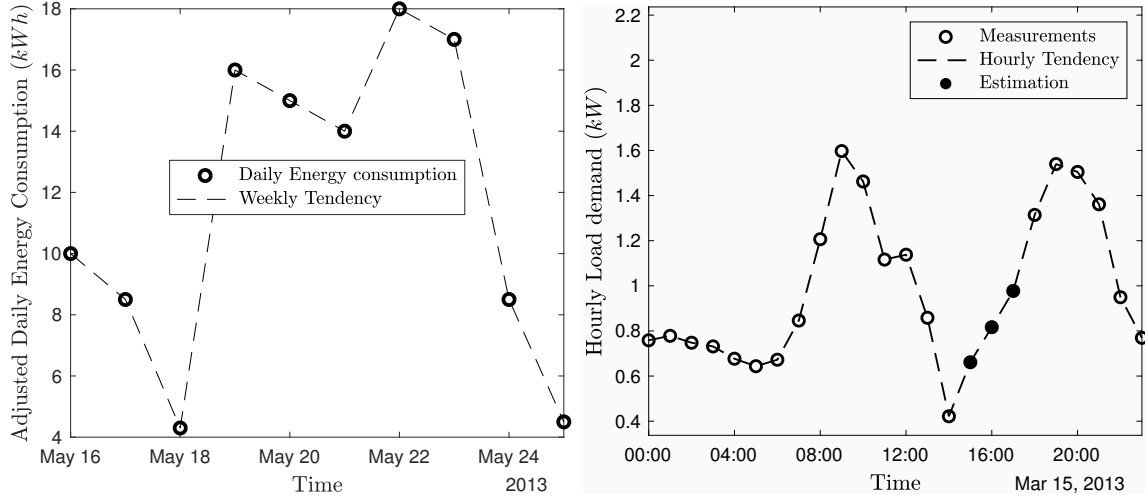
Figure 3.13: Left:Daily energy consumption tendency, Right: Hourly load demand modelling

stochastic TLs by means of the ordinary least squares model (3.74), where $\mathcal{L}'_p$ and $\tilde{S}_d$ are the mean values of the relative active TLs and load demand, respectively, $cov(S_d, \mathcal{L}_p)$ is the covariance of active TLs ($\mathcal{L}_p$) and load demand S_d; and $var(S_d)$ is the variance of load demand S_d.

$$\mathcal{L}'_p = \beta S_d + \alpha \tag{3.74}$$

Where:

$$\beta = \frac{cov(S_d, \mathcal{L}'_p)}{var(S_d)} \tag{3.75}$$

$$\alpha = \widetilde{\mathcal{L}}_p - \beta \widetilde{S}_d \tag{3.76}$$

Considering the PDFs obtained, a Monte Carlo Simulation is performed solving the unbalanced three-phase power flow in each iteration of the network presented. For each iteration of the simulation, the total power losses (active and reactive) are calculated and the results of the simulation are indicated in 3.16 where it can be seen that the distribution of the power losses (both active and reactive) are exponential, being the highest value achieved (active power losses) around 3% of the total power demand of the network.
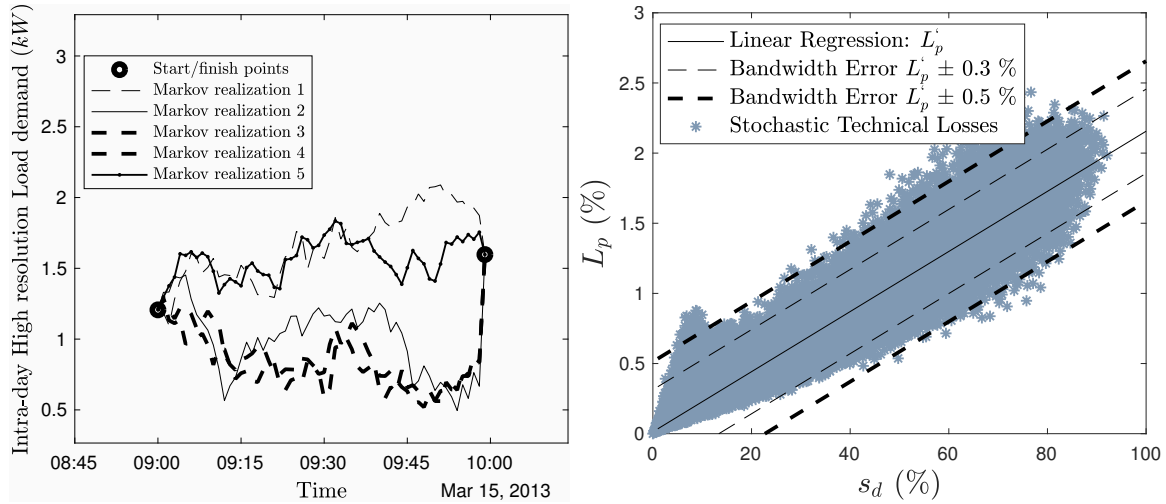
Figure 3.14: Left: Fifteen intra-hour high-resolution load demand realisations of the Markov process. Right: Stochastic Power Losses obtained

3.5 Conclusions

In this chapter, power losses of a real smart grid have been stochastically analysed. Load demand from the customers smart meters has been gathered through the communications infrastructure. The existence of non-telemetered customers as well as the missing load demands data have been taken into account in order to generate an estimate them by means of a top-down approach. Missing load demand data have been estimated through an NLP optimisation process. Additionally, intra-hour high-resolution load demand profiles have been synthetically generated by means of Markov chains. For every stochastic realisation of the Markov process, the technical power losses of the network has been calculated by means of unbalanced power flow. Then, the evolution of technical power losses with the demand have been statistically modelled throng a linear model, obtaining a losses model.
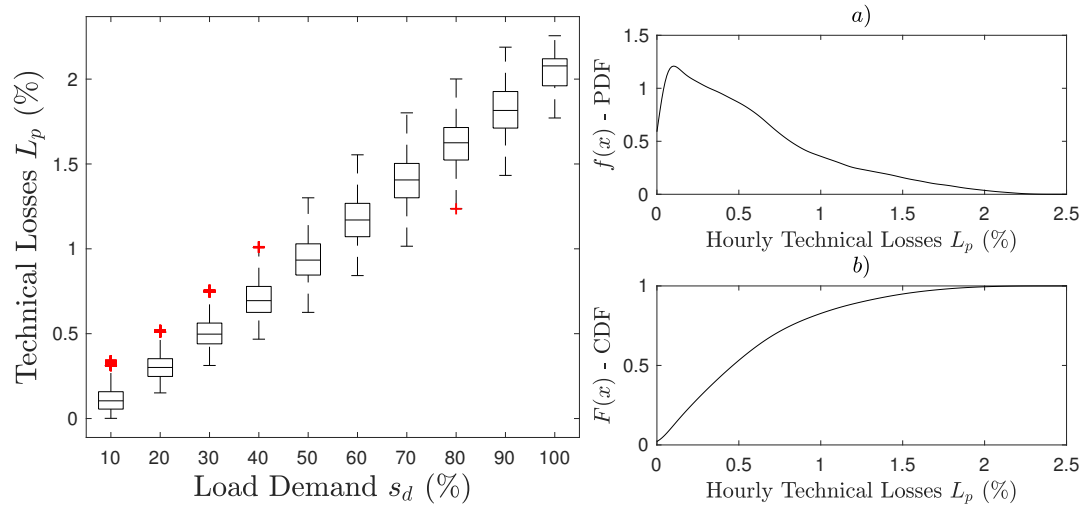
Figure 3.15: Left:Box-plots for stochastic TLs, right: PDF and CDF obtained.
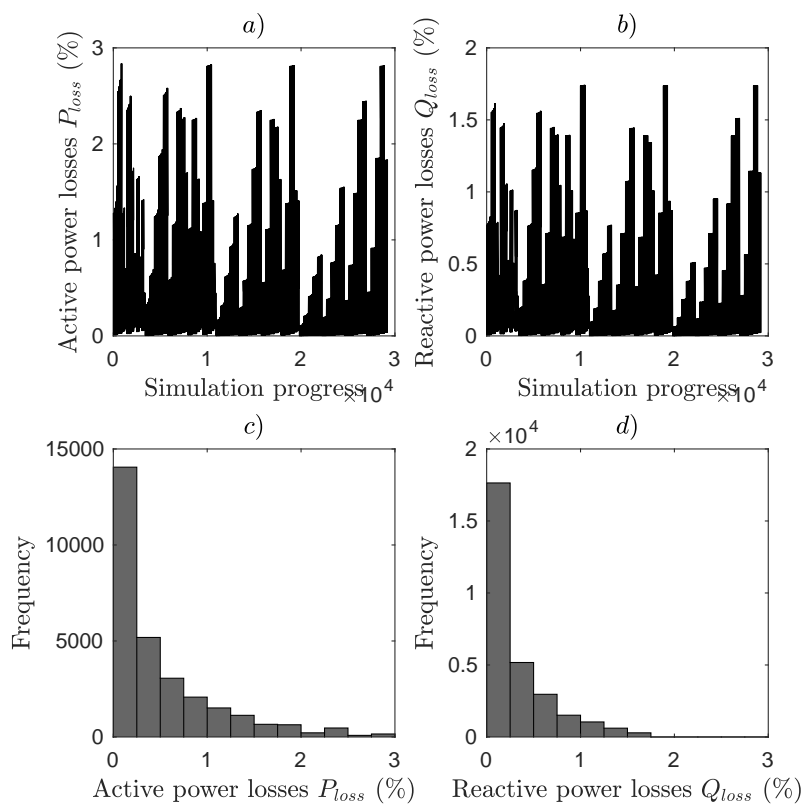
Figure 3.16: Power losses Montecarlo simulation

Chapter 4

Power Losses Estimation in Large-Scale Distribution Areas

4.1 Introduction

Energy regulators incentivise utility companies to reduce power losses (low voltage and medium voltage) to increase energy efficiency in distribution networks [CNMC, 2019]. In this chapter, a deep learning-based power losses predictor model is built for power losses estimation in large-scale low-voltage distribution areas.

Recently, a new group of machine learning models has been applied in power systems analysis [Monteiro et al., 2018]. The application of these techniques has emerged due to the increasing amount of data available and the increasing number of requirements related to network operation. Moreover, these models can reveal hidden insights in the data generated in the power network, something that traditional approaches cannot provide.

Power losses estimation can be considered as a regression learning problem, in which a tradeoff between bias and variance has to be handled. Ensembles models and Deep Neural Networks (DNN) are machine learning models well suited for that purpose [Grus, 2015]. In the the first group, the most widely used techniques are Random Forest and Gradient Boosted Decision Trees (GBDT) models, which provide accurate results but they are computationally expensive. In the second group, appear the DNN models among other models [Goodfellow et al., 2016b]. DNNs are artificial intelligence systems inspired by the

biological neural networks that constitute human brains and are categorised as supervised machine learning. Without any prior knowledge, they automatically identify the input to provide some specific output in terms of the training received [Schmidhuber, 2015]. DNNs are capable to capture complex non-linear relationships (such as power loss estimation) and they represent a compromise between accuracy and performance. Within this type of models, it can be found Long-Short Term Memory (LSTM) models or Convolution Neural Networks (CNN) if the data flows from the input layer to the output layer and back again, or Feedforward Neural Networks (FNN) where the data only flows in one direction, from the input layer to the output layer. For simplicity the latest type is chosen, without loss on performance or accuracy. FNNs are multi-purpose learning systems offering high performance and accuracy, for that reason they are selected for the power losses estimation.

The principal contributions and novelties of the presented work can be summarised as follows:

- A novel deep learning methodology for loss estimation in large-scale low-voltage smart grids with a high penetration of DG.

- The proposed model considers the unbalanced operation of low-voltage networks. This aspect is not often considered in other literature.

- The proposed method can cope with uncertainty in the network, energy consumption, and distribution generation production.

To the best of the authors's knowledge, the estimation of power losses in unbalanced large-scale low-voltage smart grids using deep learning techniques remains highly unexplored in the literature.

4.2 Methodology

This section provides the formulation of the proposed technical power losses estimation methodology for large-scale low-voltage smart grids which could easily be composed of

hundreds of SS and thousands of distribution feeders. Each feeder belonging to that area can have different topological characteristics (such as cross-section, length, and overhead-/underground configuration) and different levels of DG penetration, smart meter penetration, voltage unbalance, and load demand. The methodology comprises the following five steps:

1. **Data Collection**: low-voltage feeders' data and smart meter measurements are collected from the large-scale low-voltage distribution area under investigation. From this step, a feeder's data matrix containing the customised feeder's characteristics is obtained.

2. **Data Normalisation**: Due to the fact that feeder's characteristics vary considerably, a normalisation data process is performed to obtain the normalised feeder's data matrix.

3. **Features Extraction**: To deal with the curse dimensionality, a new feeder's data matrix with reduced dimensions is obtained, which contains the most discriminatory information providing a meaningful system representation.

4. **Feeder's Clustering**: The entire set of low-voltage feeders belonging to the large-scale low-voltage distribution area is classified into feeder's clusters where the representative feeder is selected.

5. **Deep Neural Network Power Losses Model**: A novel deep learning model is formulated to estimate technical power losses in the entire low-voltage area. To this end, it is necessary to select the optimum hyper-parameters, to perform the training and validation test sets as well as to select the network raw data.

4.3 Data Collection

Actual low-voltage distribution networks are characterised by a massive roll-out of smart metering deployment among low-voltage residential customers [López et al., 2015].

Large low-voltage distribution areas are characterised by high levels of variability such as: feeder properties, number and spatial location of customers and number and spatial location of DG units. Therefore, the network data input set to be applied to the deep learning model has to be representative enough to describe the nature and behaviour of the feeders, but it also has to be reduced to prevent overfitting the deep learning model. The more relevant feeder's characteristics are summarised in Table 4.1, which include information regarding the level of unbalance, the presence of DG units, the ratio of customers with smart meters as well as information regarding the physical configuration of the feeders (length, section, configuration, etc.). Note that in total, there are 14 customised feeder characteristics, because characteristics with subindex p are defined for each phase $p \in \{a, b, c\}$.

$X_{1,p}^f$ repressed the power ratio between power contracted by customers connected to phase $p \in \{a, b, c\}$ of feeder f and power contracted by all the customers connected to the same feeder. X_2^f represents the ratio between power contracted by the customers connected to the feeder f and power rating of the transformer located in the SS from which the feeder belongs. X_3^f represents the ratio between the number of telemetered customers and the total number of customers connected to the feeder. X_4^f represents the ratio between the DG power and the power demand of customers connected to the feeder. X_5^f represents the the spatial-location of the DG units along the feeder using the projected Euclidean distance respect to the location of the SS. X_6^f represents the spatial-location of the customers along the feeder using the projected Euclidean distance respect to the location of the SS. X_7^f represents the ratio between the average energy produced by the DG unit of a customer and its average energy consumption. X_8^f represents the the total feeder impedance. $X_{9,p}^f$ represents the the expected value (or mean Value, MV) of the power demand of phase $p \in \{a, b, c\}$ of the feeder f in Working days (WD). $X_{10,p}^f$ represents the dispersion (or Standard Deviation SD) of the power demand of phase $p \in \{a, b, c\}$ of the feeder f in WD. $X_{11,p}^f$ represents the MV of the power demand of phase $p \in \{a, b, c\}$ of the feeder f in Non-Working days (NWD). $X_{12,p}^f$ represents the SD of the power demand of phase $p \in \{a, b, c\}$ of the feeder f in NWD. X_{13}^f represents the ratio between number of customers connected to a feeder f and the total number of customers connected to the

Table 4.1: Low-voltage network properties for representative feeder's clustering

Property	Definition	Data source
$X_{1,p}^f$	Unbalance power level phase p	SM
X_2^f	Feeder loading level	N
X_3^f	Smart meters deployment	N
X_4^f	DG penetration level	N
X_5^f	DG spatial location	N
X_6^f	Customers spatial location	N
X_7^f	Self-consumption ratio	N
X_8^f	Impedance feeder path	N
$X_{9,p}^f$	Demand MV WD phase p	SM
$X_{10,p}^f$	Demand SD WD phase p	SM
$X_{11,p}^f$	Demand MV NWD phase p	SM
$X_{12,p}^f$	Demand SD NWD phase p	SM
X_{13}^f	Ratio customers per feeder	N
$X_{14,p}^f$	Ratio customers per phase	N

MV: Mean Value, SD: Standard Deviation, WD: Working days, NWD: Non-Working days, SM: Smart Meters Data, N: Network Data. p is the index for phases: $\{a, b, c\}$

.

same SS. Finally $X_{14,p}^f$ represents the ratio between the customers connected to the phase $p \in \{a, b, c\}$ of feeder f and the total number of customers connected to the same feeder

As the majority of low-voltage feeders have radial topology and the presence of loops (weakly meshed topologies) is scarce, the location of the DG units along the feeder plays a crucial role regarding the levels of power loss [Ochoa and Harrison, 2011].

Hence, the spatial location of the DG units (X_5^f) is determined based on the Euclidean distance (d_e^*) between the centroid of the DG unit coordinates and the SS coordinates. The characteristic X_5^f takes values according to expression (4.1). Depending on the value of that characteristic, there are three possible situations: DG located at the main feeder head, located at the feeder centre or located at the end of the feeder as illustrated in Fig. 4.1.

$$X_5^f = \frac{d_e^*}{L} - \frac{1}{2} = \begin{cases} \frac{-1}{2} & \text{if } d_e^* = 0 \ \rightarrow \text{Feeder Head} \\ 0 & \text{if } d_e^* = \frac{L}{2} \ \rightarrow \text{Feeder Centre} \\ \frac{1}{2} & \text{if } d_e^* = L \ \rightarrow \text{Feeder Tail} \end{cases} \quad (4.1)$$

Where L denotes the length of the feeder and d_e^* is the Euclidean distance obtained with (4.2).

$$d_e^* = \sqrt{(X_{ss} - X^*)^2 + (Y_{ss} - Y^*)^2} \quad (4.2)$$

Where (X_{ss}, Y_{ss}) are the GPS coordinates of the SS and (X^*, Y^*) are the GPS coordinates of the DG unit calculated with (4.3).

$$X^* = \frac{\sum_{i=1}^{N_{DG,f}} X_i}{N_{DG,f}}, \quad Y^* = \frac{\sum_{i=1}^{N_{DG,f}} Y_i}{N_{DG,f}} \quad (4.3)$$

And (X_i, Y_i) are the coordinates of the ith DG units belonging to the feeder f, and $N_{DG,f}$ is the number of DG units connected to feeder f.

The collection of feeder's characteristics can be described in mathematical terms as a feeder's data matrix $X \in \mathbb{R}^{n \times \tilde{p}}$ where the columns are the feeder's characteristics and the rows are the feeders samples , i.e. each one of the feeders belonging to the distribution area from where the data was collected (n feeders) as indicated in (4.4).

$$X = \begin{pmatrix} \mathcal{X}_{1,1} & \mathcal{X}_{1,2} & \dots & \mathcal{X}_{1,\tilde{p}} \\ \mathcal{X}_{2,1} & \mathcal{X}_{2,2} & \dots & \mathcal{X}_{2,\tilde{p}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{X}_{n,1} & \mathcal{X}_{n,2} & \dots & \mathcal{X}_{n,\tilde{p}} \end{pmatrix} \quad (4.4)$$

4.4 Data Normalisation

To obtain a more interpretable description of the feeder's characteristics collected and to give the same weight to all feeder properties, a normalisation process is carried out. This is necessary due to the diverse nature of the properties [de Souto et al., 2008]. For instance, the typical length of low-voltage feeders may vary between 50 and 500 meters,
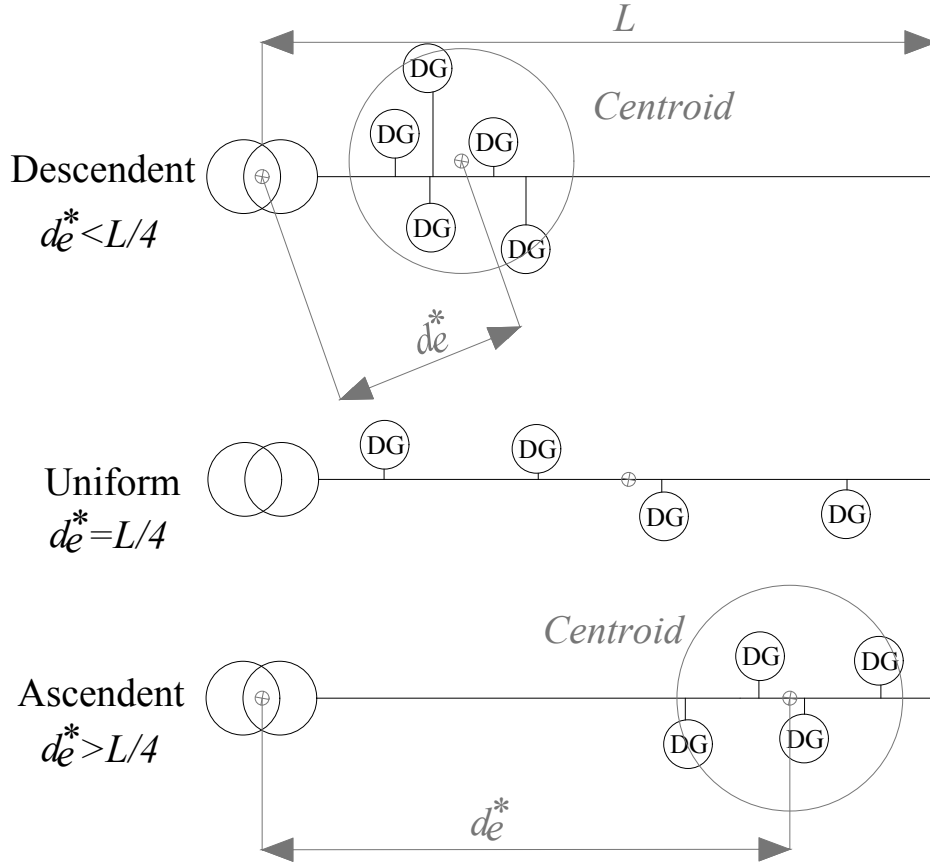
Figure 4.1: Characterisation of the spatial location of the DG units along the feeder

meanwhile, the typical phase unbalance may vary between 0.25 and 0.35. If there is no data normalization stage, the phase unbalance variation will be negligible compared to the feeder length.

The original feeder's data matrix X is transformed into the normalised feeder's data matrix M by applying to each entry $\mathcal{X}_{i,j}$ of the matrix X a normalising function f : $\mathbb{R} \to \mathbb{R}$ defined in expression (4.5) [Andrew Watters and Boslaugh, 2018]. This way the maximum absolute value of the variable is scaled to unit size and vary from zero to one.

$$x_{i,j} = \frac{\mathcal{X}_{i,j} - \mathtt{min}\{\mathcal{X}_j\}}{\mathtt{max}\{\mathcal{X}_j\} - \mathtt{min}\{\mathcal{X}_j\}}, \ \ \forall i \in (1, \ldots, n), \forall j \in (1, \ldots, \tilde{p}) \tag{4.5}$$

Where:

- $x_{i,j}$ is the normalised entry $\mathcal{X}_{i,j}$ of the matrix X.

- $\min\{\mathcal{X}_j\}$ is the minimum value of the characteristic j, (i.e. the minimum value of the column j of the matrix X).

- $\max\{\mathcal{X}_j\}$ is the maximum value of the of the characteristic j.

The resulting normalised feeder's data matrix M is indicated in (4.6).

$$
M = \begin{pmatrix} x_{1,1} & & x_{1,i} & & x_{1,\tilde{p}} \\ x_{2,1} & & x_{2,i} & & x_{2,\tilde{p}} \\ \vdots & \dots & \vdots & \dots & \vdots \\ x_{n,1} & & x_{n,i} & & x_{n,\tilde{p}} \end{pmatrix} \tag{4.6}
$$

Note that each column of the matrix M corresponds with a normalised feeder characteristic M_j (4.7) of dimensions $(1 \times n)$, that is, the M_j vector contains the jth characteristic of the n feeders of the low-voltage smart grid.

$$
M_j = (x_{1,j}, \dots, x_{n,j})^T, \quad \forall j \in (1, \dots, \tilde{p}) \tag{4.7}
$$

4.5   Features Extraction

One of the main aspects in which special attention have to be placed is in the input features (predictors) used for training the power losses predictive model. Large-scale low-voltage smart grids can be comprised of hundreds of substations and thousands of feeders, consequently dealing with such a large dataset could lead to a high computational burden and overfitting problems. In those situations, it becomes necessary to reduce the dimensions of the problem retaining, and at the same time, the maximum original data variability. One of the most well-known dimension reduction techniques is the Principal Component Analysis (PCA) [Jolliffe, 2002], which has been used in different applications such as finance or biology [Lu et al., 2013].

PCA is an unsupervised learning method that reduces the dimensionality of a correlated set of variables into a set of linearly independent uncorrelated variables (which are known as principal components), maintaining the majority of the variability existing in the initial data set [Lu et al., 2013]. This transformation is done in the sense that the first

component has the largest variance (e.g., it captures the maximum possible variability of the initial data set) and the second component has the highest variance subject to the constraint that it is orthogonal to the first component.

The objective is to find a linear mapping given by the projection matrix $\widetilde{U} \in \mathbb{R}^{\tilde{p} \times s}$ which transforms each row of the matrix M (i.e each normalised feeder's observation) expressed as the vector $M_i = [x_{i,1}, \ldots, x_{i,\tilde{p}}] \in \mathbb{R}^{1 \times \tilde{p}}$, $\forall i \in (1, \ldots, n)$, into a lower dimension representation, $\widetilde{M} \in \mathbb{R}^{\tilde{p} \times s}$ (4.8) with dimension $s < \tilde{p}$. Each column of the matrix $\widetilde{M}$ is called a Principal Component (PC). One advantage of using PCA is that the projected feeder's observations M_i are uncorrelated, in the sense that, the components of that column vector are mutually independent [Lu et al., 2013].

$$\widetilde{M} = \widetilde{U}^T M^T \tag{4.8}$$

The number of PCs is usually chosen by defining the total variability to be captured. The number of PCs chosen to capture at least η % of the cumulative variability captured has to fulfil (4.9) where λ_k is the kth eigenvalue associated and β_λ is the total number of eigenvalues of the covariance matrix S.

$$\underset{s}{\operatorname{argmin}} \left\{ \frac{\sum_{k=1}^{s} \lambda_k}{\sum_{k=1}^{\beta_\lambda} \lambda_k} > \eta(\%) \right\} \tag{4.9}$$

The covariance matrix $S \in \mathbb{R}^{\tilde{p} \times \tilde{p}}$ (4.10) can be obtained by computing (in the diagonal terms) the variances s_j^2 of each variable (column) of the matrix M and, in the off diagonal elements, the covariances $s_{j,j'}$, $\forall j, j' \in (1, \ldots, \tilde{p})$, $j' \neq j$, among each pair of variables (columns) of the matrix M. Note that the covariance matrix S is a square matrix that measures the variability of the normalised feeder's data matrix M.

$$S = \begin{pmatrix} s_1^2 & s_{1,2} & \ldots & s_{1,\tilde{p}} \\ s_{2,1} & s_2^2 & \ldots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ s_{\tilde{p},1} & \ldots & \ldots & s_{\tilde{p}}^2 \end{pmatrix} \tag{4.10}$$

4.5.1 First Principal Component ($k = 1$)

The first projection vector $\widetilde{U}_1$ has to be obtained by solving the optimisation problem defined in (4.11) which includes the maximisation of the quantity $\widetilde{U}_1^T S \widetilde{U}_1$ and the normalisation constraint $\widetilde{U}_1^T \widetilde{U}_1 = 1$.

$$\underset{\widetilde{U}_1}{\mathrm{argmax}} \left\{ \widetilde{U}_1^T S \widetilde{U}_1 \right\} \texttt{ subject to } \widetilde{U}_1^T \widetilde{U}_1 = 1 \tag{4.11}$$

This can be solved by using the technique of the Lagrarian multipliers, from which the expression (4.11) is converted in (4.12) where λ_1 is the Langrarian multiplier.

$$\psi_1 = \widetilde{U}_1^T S \widetilde{U}_1 - \lambda_1 \left(\widetilde{U}_1^T \widetilde{U}_1 - 1 \right) \tag{4.12}$$

The expression (4.12) can be differentiated with respect to $\widetilde{U}_1$ obtaining (4.13).

$$\partial \psi_1 / \partial \widetilde{U}_1 = 0 \Rightarrow (S - \lambda_1 I) \widetilde{U}_1 = 0 \tag{4.13}$$

Where $I \in \mathbb{R}^{\tilde{p} \times \tilde{p}}$ is the identity matrix. By the definition of eigenvalues and eigenvectors, from (4.13), it is clear that λ_1 is the eigenvalue of the covariance matrix S, and $\widetilde{U}_1$ the corresponding eigenvector. With the result obtained in (4.13) it can be concluded that the first projection vector $\widetilde{U}_1$ will be the eigenvector associated with the first large eigenvalue λ_1 of the covariance matrix S.

4.5.2 Second Principal Component ($k = 2$)

To find the second PC (i.e the second column of the reduced feeder's data matrix $\widetilde{M}$) the eigenvector $\widetilde{U}_k$ with $k = 2$ has to be determined, which also implies solving the optimisation problem (4.11) but also including the constraint that the projection by $\widetilde{U}_2$ are uncorrelated with $\widetilde{U}_1$ as indicated in (4.14).

$$\underset{\widetilde{U}_1}{\mathrm{argmax}} \left\{ \widetilde{U}_1^T S \widetilde{U}_1 \right\} \texttt{ subject to }$$
$$\widetilde{U}_1^T \widetilde{U}_1 = 1 \quad \widetilde{U}_2^T \widetilde{U}_1 = 0 \tag{4.14}$$

Next, the Langrarian multipliers procedure is followed to obtain $\widetilde{U}_2$, but in this case there are two constraints to consider, thus, two multipliers has to be defined as indicated in (4.15).

$$\psi_2 = \widetilde{U}_2^T S \, \widetilde{U}_2 - \lambda_2 \left(\widetilde{U}_2^T \widetilde{U}_2 - 1 \right) - \mu_2 \left(\widetilde{U}_2^T \widetilde{U}_1 \right) \tag{4.15}$$

Differentiating (4.15) with respect to $\widetilde{U}_2$ it is obtained (4.16).

$$\partial \psi_2 / \partial \widetilde{U}_2 = 0 \Rightarrow S\widetilde{U}_2 - \lambda_2 \widetilde{U}_2 - \mu_2 \widetilde{U}_1 = 0 \tag{4.16}$$

Multiplying both sides of equation (4.16) by $\tilde{U}_1^T$ result in (4.17).

$$\widetilde{U}_1^T S \, \widetilde{U}_2 - \lambda_2 \left(\widetilde{U}_1^T \widetilde{U}_2 \right) - \mu_2 \left(\widetilde{U}_1^T \widetilde{U}_1 \right) = 0 \rightarrow \mu_2 = 0 \tag{4.17}$$

The first term of equation (4.17) as well as the second one are zero due to the zero-correlation constraint. The third term is the unity due to the normalisation constraint. Therefore, this results in $\mu_2 = 0$, which allows us to rewrite the expression (4.16) in the reduced equation (4.18) where, it is concluded that λ_2 is the second projection vector and $\tilde{U}_2$ is the eigenvector associated with the second large eigenvalue of the covariance matrix S.

$$\partial \psi_2 / \partial \widetilde{U}_2 = 0 \Rightarrow (S - \lambda_2 I) \, \widetilde{U}_2 = 0 \tag{4.18}$$

4.5.3 Subsequent Principal Components ($k > 2$)

As a result, the subsequent projection vectors $\widetilde{U}_k$ ($k = 3, \ldots, s$) could be obtained as each eigenvector associated with the kth large eigenvalue λ_k of the covariance matrix S.

4.6 Feeder's Clustering

In this stage, the n feeders from the target large-scale low-voltage smart grids are partitioned into a reduced number of feeder's groups (i.e., clusters), where feeders belonging

to the same cluster are similar [Ke-Lin and M. N. S., 2014]. Then, from each feeder's cluster, a representative feeder is chosen as the closest to the centroid of that cluster.

Clustering techniques can be broadly divided in hierarchical-based, partitional-based and distribution-based clustering techniques [Ke-Lin and M. N. S., 2014]:

- **Hierarchical clustering** produces, as output, a binary tree in which the root node represents the whole data, the leaf nodes represent the clusters obtained, and the intermediate nodes represent the distance between clusters. Some algorithms in this category are BIRCH (*Balanced Iterative Reducing and Clustering using Hierarchies*) [Zhang et al., 1966], CURE (*Clustering Using REpresentative*) [Sudipto et al., 1998], and ROCK (*RObust Clustering using linKs*) [Guha et al., 1999].

- **Partitional clustering** is based on building an initial set of clusters by partitioning the original data set and then redistributing them iteratively by means of calculating the centroid (cluster centre) of each cluster. This kind of technique presumes low time complexity and high efficiency [Ke-Lin and M. N. S., 2014]. Some algorithms in this category are `K-means++` [Hartigan and Wong, 1979] `K-medoids++` [Leonard and Peter, 1990] and CLARANS (*Clustering Large Applications based on RANdom Search*) [Ng and Han, 1998] being the first one the most popular squared error-based clustering algorithm.

- **Distribution-based clustering** techniques are based on probability distribution functions (PDF). By this, the clusters are found as the groups of elements that offer similar probability distribution functions. One algorithm of this kind of technique is known as the Gaussian Mixture Model (GMM) [Reynolds, 2009], which makes use of the Expectation-Maximisation (EM) algorithm [P. et al., 1977]. The initial data set is modelled with a fixed number of Gaussian distributions randomly initialised. The parameters of those distribution functions are then optimised to fit the data set achieving a local optimum. This algorithm produces complex models capturing high correlation and dependence between attributes but depends on the statistical model chosen [Theodoridis and Koutroumbas, 2008].

Table 4.2: Clustering techniques and performance comparison

Technique	Algorithm	Time Complexity	Scalability	Large data sets	Outliers	Noise
H	BIRCH	$O\left(n\right)$	High	Yes	Low	Low
H	CURE	$O\left(\tilde{n}^2 log\left(\tilde{n}\right)\right)$	High	Yes	Low	Low
H	ROCK	$O\left(n^2\right)$	Medium	No	Low	Low
PB	K-means++	$O\left(log(\hat{k})\right)$	Medium	Yes	High	High
PB	K-medoids++	$O\left(\hat{k}\left(n-\hat{k}\right)^2\right)$	Low	No	Low	Low
PB	CLARANS	$O\left(n^2\right)$	Medium	Yes	Low	Low
DB	GMM	$O\left(\epsilon s^3\right)$	Medium	Yes	-	-

H: Hierarchical, PB: Partion-Based, DB: Distribution-Based, CLARANS: Clustering Large Applications based on RANdom Search, CURE (Clustering Using REpresentative, ROCK: RObust Clustering using linKs, Sens.: Sensibility, GMM: Gaussian Mixture Model, BIRCH: Balanced Iterative Reducing and Clustering using Hierarchies

Table 4.2 summarises a comparison between the most extended algorithms of each clustering technique where $\hat{k}$ is the number of clusters, s is the dimension of the reduced observation vector (columns of the matrix $\widetilde{M}$), n is the number of samples (rows of the matrix $\tilde{M}$) and ϵ is the number of iterations of the EM algorithm. It is also indicated the sensibility of the technique to outliers and to the presence of noise.

According to [Rigoni et al., 2016] GMM and improved K-means++ are well suited candidates for clustering feeders with many feeder's characteristics. Nevertheless, GMM is characterized by a higher computational complexity in comparison with K-means++. Therefore, in this thesis the n feeders gathered in the reduced feeder's data matrix $\tilde{M}$ are clustered by means of the K-means++ algorithm, using the uncorrelated s Principal Components obtained in the PCA analysis.

4.6.1 Clustering Algorithm

The K-means++ algorithm partitions the dimension-reduced feeder's data matrix $\widetilde{M} \in \mathbb{R}^{n \times s}$ of feeder's characteristics (which have (n) feeders' observations with (s) Principal

Components) into $\hat{k}$ separated subsets (i.e. feeders' clusters) $C_r = \{c_1, \ldots, c_{n_r}\}$, $\forall r \in (1, \ldots, \hat{k})$. `K-means++` has two steps: the initialization of the centroids and the assignment step.

1. **Centroid Initialization**: The location of the $\hat{k}$ centroids are chosen as follow:

   a.1) Random selection of the rth cluster's centroid, $c_r^{*(0)} = c_i$ where $c_i \in \tilde{M}$ is selected from the feeders data points that compound the feeder's PCs.

   a.2) Computation of the Euclidean distance $\mathcal{D}$ between each feeder data point and the centroid $\mathcal{D}(c_i, c_r^{*(0)}) = ||c_i - c_r^{*(0)}||^2$.

   a.3) Updating of the centroid $c_{r+1}^{*(0)}$ as the one that offers the highest probability according to (4.19).

$$\mathcal{P}\left(c_{r+1}^{*(0)} = c_j\right) = \mathcal{D}(c_j, c_r^{*(0)})^2 / \sum_{k,j \neq k} \mathcal{D}(c_k, c_r^{*(0)}) \qquad (4.19)$$

   a.4) Repeat a.2) and a.3) until $\hat{k}$ centroids has been chosen.

2. **Assignment step**: After initialisation of the centroids, each feeder data point is assigned to the closest centroid cluster by minimising the Mean Square Error (MSE) according to (4.20).

$$MSE\left(C_1, \ldots, C_{\hat{k}}\right) = \sum_{r=1}^{\hat{k}} \frac{1}{n_r} \sum_{i=1}^{n_r} ||c_i - c_r^{*(0)}||^2,$$
$$r \in \left(1, \ldots, \hat{k}\right) \qquad (4.20)$$

Then, the initial cluster set is updated by recalculating the centroids as the mean of the samples of the clusters using (4.21).

$$c_r^{*(0)} = \frac{1}{n_r} \sum_{c_i=1}^{n_r} c_i, \quad \forall r \in (1, \ldots, \hat{k}) \qquad (4.21)$$

The Assignment step (4.20)-(4.21) is repeated q times until the change in all $c_r^{*(q)}$ is sufficiently small ε as indicated in (4.22).

$$c_r^{*(q)} - c_r^{*(q-1)} \leq \varepsilon, \quad \forall r \in (1, \ldots, \hat{k}) \tag{4.22}$$

4.6.2 Clustering Evaluation

To evaluate whether a feeder's sample has been appropriately associated with the right cluster two index can be used: the Silhouette and the Global Silhouette (GS). The Silhouette index is a coefficient that quantifies the similarity of the object (in this case, a feeder sample) with the rest of the elements of the group belonging to its own cluster (in this case, the feeders' group) [Rousseeuw, 1987]. The values that could adopt the Silhouette range from -1 (low relation) to $+1$ (high relation).

The Silhouette of the feeder sample c_i is defined as $\mathbb{S}(c_i)$ and is calculated according to (4.23).

$$\mathbb{S}(c_i) = \frac{b(c_i) - a(c_i)}{max\{a(c_i), b(c_i)\}} \approx \underbrace{1 - \frac{a(c_i)}{b(c_i)}}_{b(c_i) > a(c_i)}, \quad \forall i \in (1, \ldots, n) \tag{4.23}$$

Where:

- $a(c_i)$ is the average distance between the feeder sample c_i and all the other feeders belonging to the same cluster.

- $b(c_i)$ is the smallest average distance between the feeder sample c_i and all the others feeders in all clusters. Notice that for $b(c_i) > a(c_i)$ the expression (4.23) could be reduced.

If a feeder sample c_i have a Silhouette index value $\mathbb{S}(c_i) \sim 1$, is because $b(c_i) \gg a(c_i)$ which implies that the feeder sample is poorly related with its neighbouring clusters, so the clustering configuration is appropriated. On the contrary, if the Silhouette index value is $\mathbb{S}(c_i) \sim -1$, then the feeder sample is likely to belong to a neighbour feeder cluster and, so, the clustering has to be revised. Finally, $\mathbb{S}(c_i) \sim 0$ indicates that the feeder sample c_i is on the border of two neighbouring clusters.

To evaluate the quality of the feeders' cluster process, the Global Silhouette coefficient is used (4.24). The GS coefficient is based on the Silhouette index and provides a general

sense of the quality of the clustering process. The higher the GS coefficient value, the higher the quality of the clustering.

$$GS = \frac{1}{\hat{k}} \sum_{r=1}^{\hat{k}} \frac{1}{n_r} \sum_{i=1}^{n_r} \mathbb{S}(c_i) \tag{4.24}$$

Where n_r is the number of feeders belonging to the cluster C_r, $\forall r \in (1, \ldots, \hat{k})$.

4.6.3   Representative Feeder's Selection

Once the appropriate clusters have been obtained, the representative feeders, denoted with $\tilde{c}_r$ are chosen as the feeder sample closest to the centroid within each feeders' cluster by means of the minimum Euclidean Distance to the centroid c_r^*, which is described in (4.25).

$$\tilde{c}_r = \arg\min_{c_i \in C_r} \left\{ ||c_i - c_r^*||^2 \right\}, \ \ \forall r \in \left(1, \ldots, \hat{k}\right) \tag{4.25}$$

It must be noted that the representative feeder $\tilde{c}_r$ chosen from the cluster C_r evaluates the power losses from the set of feeders belonging to that cluster.

To deal with the differences between the representative feeder $\tilde{c}_r$ and the different feeders belonging to that cluster $c_i \in C_r$ (as it is illustrated in Fig.4.2) an average Euclidean distance factor α_r is used as a weighting factor to extrapolate the representative feeder to the rest of the feeders inside the cluster C_r. This factor is based on the Euclidean distance from each feeder to the reference feeder, in the cluster, according to (4.26).

$$\alpha_r = \frac{1}{n_r} \sum_{i=1}^{n_r} ||c_i - \tilde{c}_r||^2, \ \ r \in \left(1, \ldots, \hat{k}\right) \tag{4.26}$$

4.7   Model Formulation

4.7.1   Model Architecture

The DNN-based power losses model proposed in this thesis is illustrated in Fig.4.3. The model can be described physically as a collection of nodes called artificial neurons, which
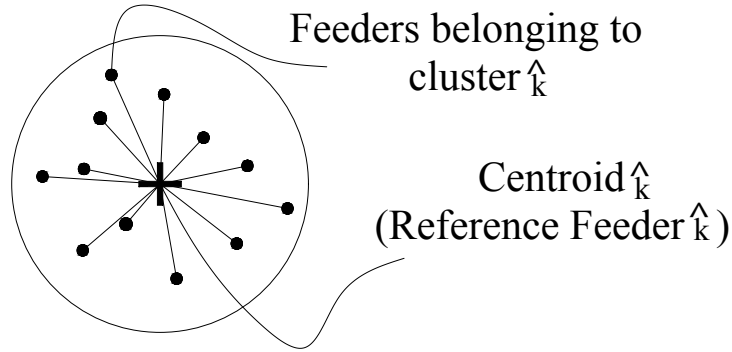
Figure 4.2: Feeders' cluster representation

are the basic units of computation. These neurons are interconnected by weighted links called edges. The input layer L_{in} receives the data input vector denoted by $\mathbb{X}$, which comprises the representative demand and generation conditions of the whole low-voltage distribution area by means of the scaled demand and generation patterns of the representative feeders $\tilde{c}_r$. The output layer L_o provides the model output $\mathbb{Y}$ which represents the total technical losses of the large-scale low-voltage distribution area. Between the input and the output layers are multiple hidden layers, $k \in (1, \ldots, h)$, with h being the number of hidden layers, each one containing n_k number of neurons.

The configuration of the hidden layer's architecture through the connections between the neurons determines the final behaviour of the model, as, in each hidden layer, a non-linear transformation of the aggregate activation of the DNN takes place [Simon S, 1999]. It must be noted that as the size of the model increases, the greater the amount of flexibility obtained. However, this requires a higher computational complexity.

Moreover, to avoid overfitting (when the model performs so closely to the training data but fails with new data), some of the neurons of the model must be cancelled (i.e.: no output). This is achieved by means of the "dropout" technique.

Figure 4.3: Deep neural network-based power losses model proposed
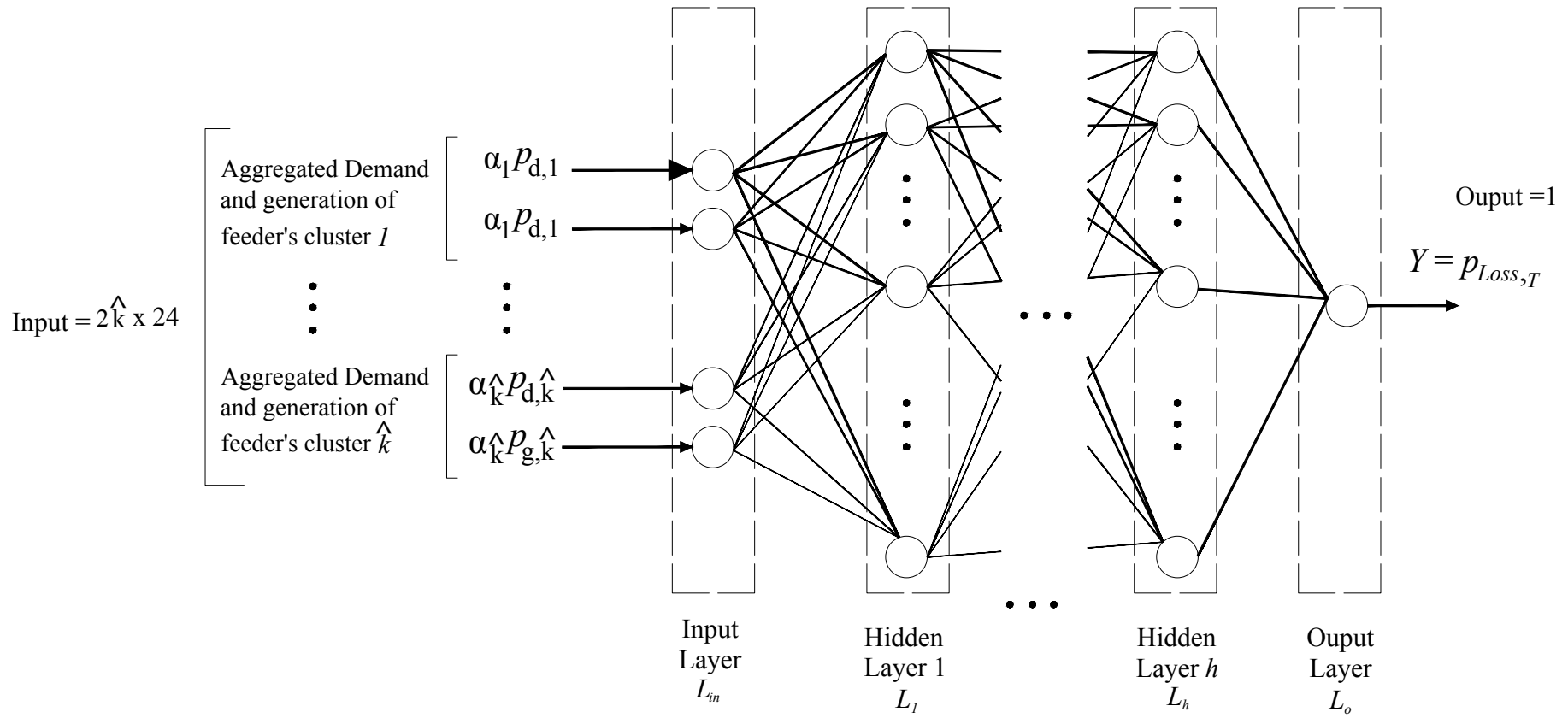
4.7.2 Model Input

As the entire set of feeders belonging to the large-scale low-voltage distribution area has been separated into feeders' clusters, the daily Representative Patterns (RPs) for the power demand and generation (DG) of each feeders' cluster has to be defined and used as input to the deep learning losses model. Fig. 4.3 illustrates the process of scaling each daily APs of each representative feeder to obtain the daily RPs of each feeder cluster in which the entire set feeders has been divided as indicates (4.27). Therefore, the model input vector $\mathbb{X}$ defined in (4.27) describes the representative demand and generation conditions of the entire low-voltage distribution area at one instant of time t. Specifically, $\mathbb{X}$ contains the daily RPs of the demand and generation for each of the feeder's clusters in which the entire set of feeders has been separated.

$$\mathbb{X} = \begin{bmatrix} \mathbb{X}_{d,1} & \mathbb{X}_{g,1} \dots \mathbb{X}_{d,\hat{k}} & \mathbb{X}_{g,\hat{k}} \end{bmatrix}^T \quad (pu) \tag{4.27}$$

Each demand and generation RP is obtained by means of the daily Aggregated Pattern (AP) of the representative feeder, denoted as $p_{d,r}(t)$ (4.29) for demand AP and $p_{g,r}(t)$ (4.31) for generation AP, scaled to the entire feeder's cluster.

In the case of the daily demand RP denoted with $\mathbb{X}_{d,r}$ of each feeder cluster r, is composed by the daily demand AP of the feeder r in the instant time t, which is denoted with $p_{d,r}(t)$, scaled to the entire feeder's cluster r by multiplying by the coefficient α_r as indicates (4.28).

$$\mathbb{X}_{d,r} = \alpha_r \cdot p_{d,r}(t), \quad \forall r \in (1, \dots, \hat{k}) \tag{4.28}$$

Where α_r, defined previously in (4.26), is the extrapolation coefficient, that takes into account the number of feeder's in the cluster r as well as the distance from each one to the reference feeder (centroid), $p_{d,r}(t)$ is the aggregation of the net active power demand of the reference feeder r (i.e the sum of the net active power demand of each customer connected to the reference feeder r) as indicates (4.29) where $\Re$ is the real part.

$$p_{d,r}(t) = \frac{1}{S_{B_r}} \sum_{i \in \Omega_{c,r}} P_{d,r}^{(i)}(t) = \frac{1}{S_{B_r}} \sum_{i \in \Omega_{c,r}} \Re\{S_{T,d,r}(t)\}, \tag{4.29}$$

Where S_{B_r} is the power reference of the system of feeder r and $Re\{S_{T,d,r}(t)\}$ is the net total active power demand of the feeder r in instant time t.

In the case of the daily generation RP denoted with $\mathbb{X}_{g,r}$ of each feeder cluster r, is composed by the daily generation AP of the feeder r in the instant time t, which is denoted with $p_{g,r}(t)$, scaled to the entire feeder's cluster r by multiplying by the coefficient α_r as indicates (4.30).

$$\mathbb{X}_{g,r} = \alpha_r \cdot p_{g,r}(t), \quad \forall r \in (1, \ldots, \hat{k}) \tag{4.30}$$

Where the daily generation AP $p_{g,r}(t)$ of the feeder r in the instant time t is the aggregation of the net active power generation of the reference feeder r (i.e the sum of the active power generation of each DG unit connected to the reference feeder r) as indicates (4.31),

$$p_{g,r}(t) = \frac{1}{S_{B_r}} \sum_{i \in \Omega_{c,r}} P_{g,r}^{(i)}(t) = \frac{1}{S_b} Re\{S_{T,g,r}(t)\} \quad (pu) \tag{4.31}$$

Where $Re\{S_{T,g,r}(t)\}$ is the net total active power generation (by the DG units) of the feeder r in instant time t.

Therefore the input model $\mathbb{X} \in \mathbb{R}^{2\hat{k} \times 1}$ represents the demand and generation conditions at certain time instant t of the whole low-voltage distribution organised in $2\hat{k}$ References Patterns as indicates (4.32).

$$\mathbb{X} = [\alpha_1 p_{d,1}(t) \quad \alpha_1 p_{g,1}(t) \ldots \alpha_{\hat{k}} p_{d,\hat{k}}(t) \quad \alpha_{\hat{k}} p_{g,\hat{k}}(t)]^T \quad (pu) \tag{4.32}$$

In Fig. 4.4 is illustrated the process of scaling each daily APs of each representative feeder to obtain the daily RPs of each feeder cluster.

4.7.3   Model Output

On the other hand the model output $\mathbb{Y} \in \mathbb{R}$ corresponds with the total active power losses of the whole low-voltage distribution area $(p_{loss,T})$ as indicates (4.33) and constitutes the output layer L_o. It is obtained as the result of a mapping function $\mathcal{F}_N$ of the model configuration $\mathbb{W}$ (edge weights matrix) with the model input vector $\mathbb{X}$.
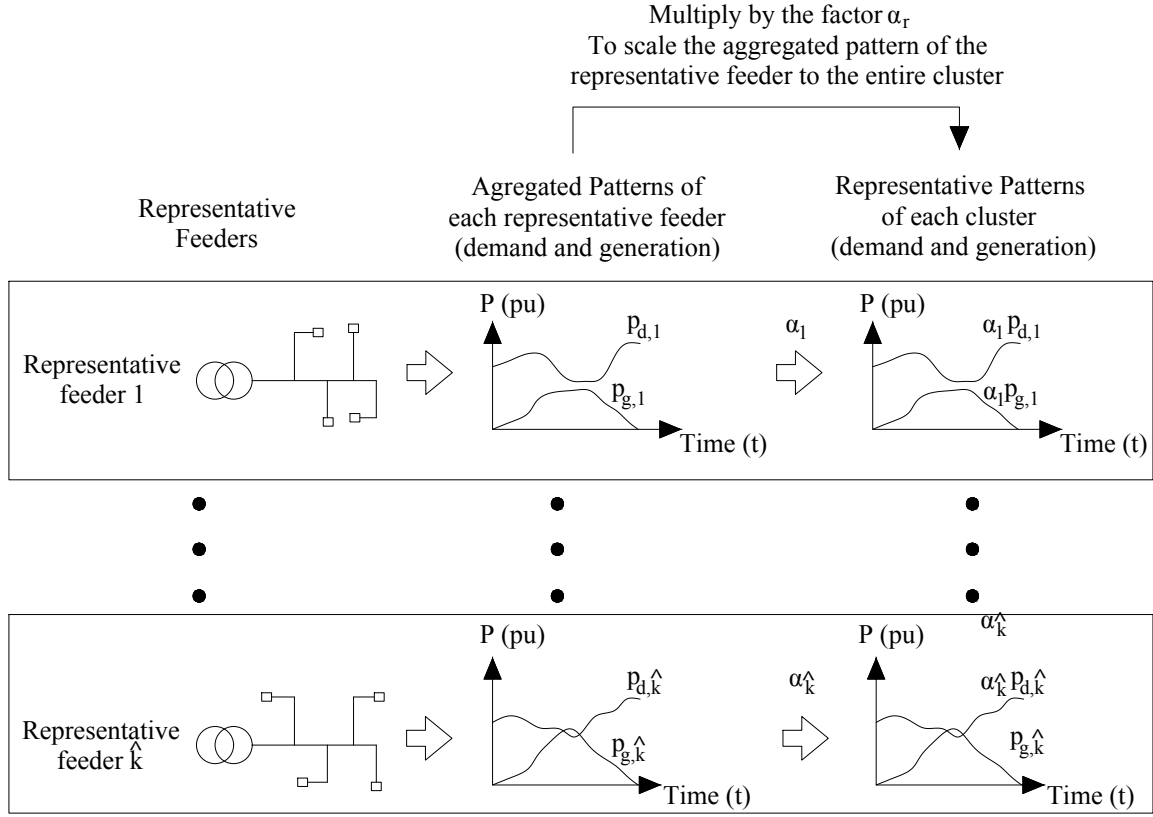
Figure 4.4: Schematic process for obtaining the daily representative patterns

$$\mathbb{Y} = \mathcal{F}_N(\mathbb{W}, \mathbb{X}) = p_{loss,T} \tag{4.33}$$

where $\mathbb{X}$ is the input vector, which is the demand and generation conditions at instant t of the whole low-voltage distribution due to expression (4.32), and is presented to the model in the input layer L_{in}, and $\mathbb{W}$ is the edge weights matrix in such a way that the element (j, k), denoted with $w_{i,k}$, represents the weight associated to the edge that connects neuron j^{th}, denoted with $\mathcal{G}_j$ with the neuron k^{th} of the, denoted with $\mathcal{G}_k$. Note that neurons of the same layer can not be connected. The predecessors neurons of the kth are defined as those who are connected to the kth neuron but are placed in the previous layer. The set of predecessor neurons of the kth neuron is denoted as $\Omega_{n,k}$, being $n_{e,k}$ the number of predecessor neurons of the kth neuron. Each neuron of the model receives real-valued signals coming from predecessor neurons, and produces a new real-value activation signal which is sent again through the output edges of each neuron as illustrated in Fig.4.5.

Therefore, given a matrix of edge weights $W \in \mathbb{R}^{h+2\hat{k} \times max\{n_k\}}$ the function $\mathcal{F}_N$ maps an input vector $\mathbb{X} \in \mathbb{R}^{2\hat{k} \times 1}$ obtaining the output $\mathbb{Y} \in \mathbb{R}$.

The mapping function $\mathcal{F}_N$ is defined over a finite set of neurons $\mathcal{N}$ given by (4.34) and a finite set of directed edges $\mathcal{E}$ given by (4.35) which constitutes the connections among neurons.

$$\mathcal{N} = \{\mathcal{G}_{k,i} | \forall k \in (in, 1, \ldots, h, o), \forall i \in (1, \ldots, n_k)\} \tag{4.34}$$

$$\mathcal{E} \subseteq \mathcal{N} \times \mathcal{N} = \{e_{i,k} | \forall i, j \in \mathcal{N}\} \tag{4.35}$$

Each neuron of the model receives a compound of real-valued signals coming from predecessor neurons, and produce a new real-value activation signal which is sent again through the output edges of each neuron as illustrates Fig. 4.5.
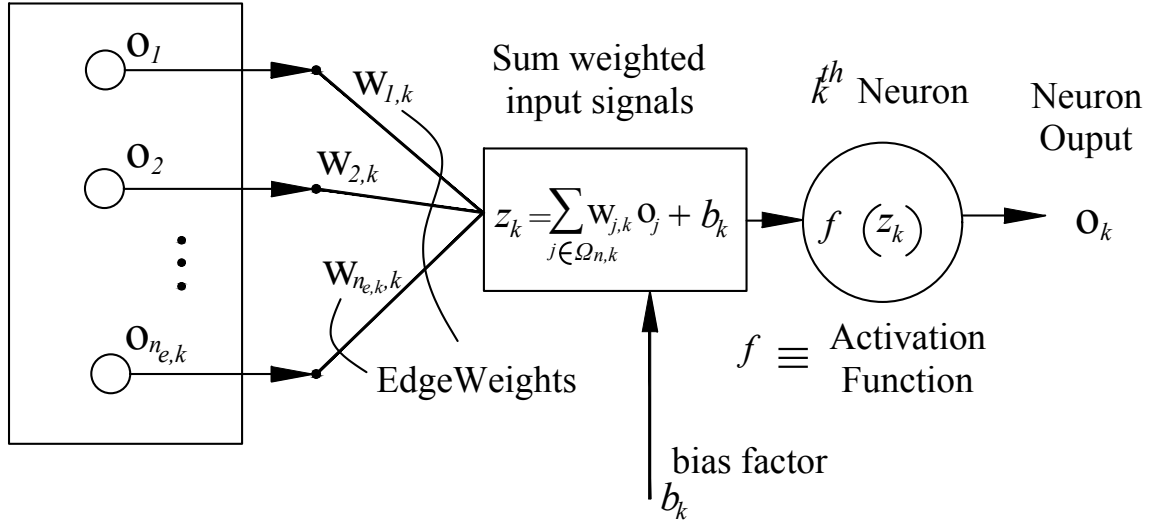

Figure 4.5: Data communication between sequential layers

The kth neuron $(\mathcal{G}_k)$ processes the signal (z_k) defined in (4.36), which is the sum of the outputs of the $n_{e,k}$ predecessor neurons (o_j) multiplied by the corresponding edge weights $(w_{j,k})$ plus a bias factor (b_k), which controls the activation process.

$$z_k = \sum_{j \in \Omega_{n,k}} w_{j,k} \, o_j + b_k \tag{4.36}$$

Then, the output of the kth neuron, defined as (o_k), is calculated by means of the activation function $f(z_k) \in (0, 1)$ as indicated in (4.37).

$$o_k = f(z_k) = \frac{1}{1 + e^{-z_k}} \in (0, 1) \tag{4.37}$$

$$\mathbb{Y} = p_{loss,T} \quad (pu) \tag{4.38}$$

For convenience, each neuron is identified within the layer which belongs, by the number of layer L_i, $i \in (0, 1, \ldots, h)$ and the number of neuron that represents with in the layer N_k, $k \in (1, \ldots, s_g)$, therefore $\mathcal{G}_i : N_k L_g, \forall i \in (1, \ldots, q)$. For instance, $N_3 L_2$ indicates the neuron 3 of hidden layer 2.

The set of edges $\mathcal{E}$ have an associated set of weights $\mathcal{W} \subset \mathcal{N} \times \mathcal{N} = \{w_{i,j}, \forall i, j \in \mathcal{N}\}$ where the value of each edge weight represents its relative importance in comparison with other edges. The set of weights $\mathcal{W}$ is arranged in a squared matrix W in such a way that the element (i, j) of the matrix W will be the edge weight of the edge that connects i^{th} neuron $\mathcal{G}^{(i)} \in \mathcal{N}$ with the jth neuron $\mathcal{G}^{(j)} \in \mathcal{N}$.

There exist a wide collection of activation functions in the machine learning literature: linear, piecewise-linear, threshold, hyperbolic tangent or sigmoid [Simon S, 1999]. The last one is the most used in the field of machine learning because is a smooth s-shape and bounded function (Fig. 4.6).
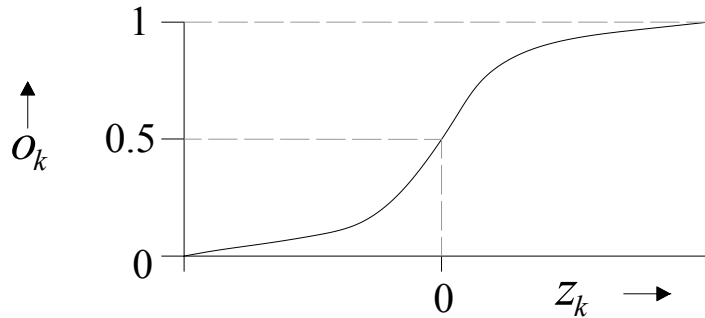
Figure 4.6: Sigmoid function

For instance, consider a network with four hidden layers (i.e $L_h = 4$) with three neurons per layer (i.e $s_1 = s_2 = s_3 = s_4 = 3$) showed in Fig. 4.7 the output of the each neurone of the hidden layer k will be o_k given by (4.37).
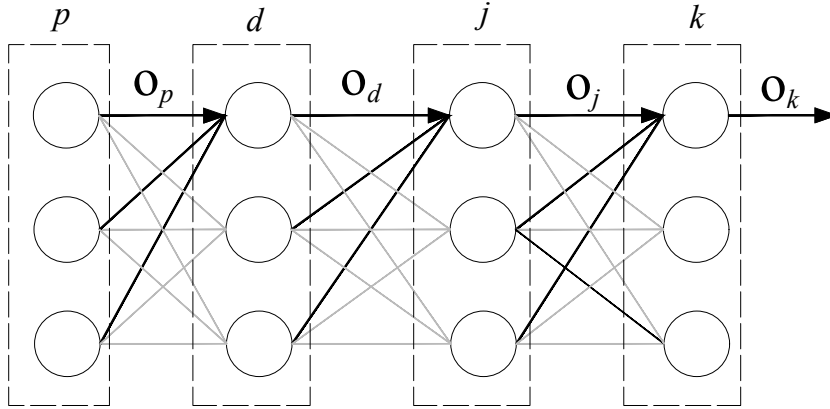
Figure 4.7: DNN computation example

The output of each neuron of the layer j is o_j given by (4.39).

$$o_j = \frac{1}{1 + e^{-z_j}} \tag{4.39}$$

Where z_j is the weighted sum of inputs that arrives at each neurone of layer j given by (4.40).

$$z_j = \sum_{d=1}^{3} w_{d,j} \cdot o_d \tag{4.40}$$

The output of each neuron of the layer d is o_d given by (4.41).

$$o_d = \frac{1}{1 + e^{-z_d}} \tag{4.41}$$

Where z_d is the weighted sum of inputs that arrives at each neurone of layer d given by (4.42).

$$z_d = \sum_{p=1}^{3} w_{p,d} \cdot o_p \tag{4.42}$$

Therefore the output of each neurone in the last hidden layer will we obtained by replacing from expression (4.42) to (4.36), obtaining with that the expression (4.43).

$$o_k = \cfrac{1}{1 + e^{-\left(\sum_{j=1}^{3} w_{j,k} \cdot \left(\cfrac{1}{1+e^{-\left(\sum_{d=1}^{3} w_{d,j} \cdot \left(\frac{1}{1+e^{-\left(\sum_{p=1}^{3} w_{p,d} \cdot o_p\right)}}\right)\right)}}\right)\right)}} \tag{4.43}$$

4.7.4 Model Training

The proposed power losses model is an artificial intelligence model with adjustable parameters that use the error between the actual output and target output to refine those model parameters. Therefore, model training consist of adjusting the model parameters (i.e. the edge weights of the matrix $\mathbb{W}$) by using a training data set in such a way that the model trained performs satisfactory if given a batch (π) of input demand and generation conditions gathered in the input vector $\mathbb{X}^\pi$, the model provides a power losses output ($\widehat{p}_{loss,T}^{(\pi)}$) close enough to the target power losses output ($p_{loss,T}^{(\pi)}$). DNN models can describe rare dependencies in the training data but this also could lead to a very common problem in machine learning, overfitting. This problem is founded in the high level of abstraction due to the number of the hidden layers. To minimise the impact of overfitting Regularisation methods are used to select the network architecture [Bengio et al., 2013]. For instance, Bayesian Regularisation (BR) is one method, which restricts the magnitude of the weights [Monteiro et al., 2018] or structural stabilisation which restricts the number of hidden nodes and/or weights. Different transference functions are tested and compared such as linear, Log-Sigmoid and Tangent-Sigmoid transfer functions, being the last one, the one with the best performance. The adjustment is done by means of minimising the mismatch between the model output when is feed with the known input (i.e the power losses that the model estimate given input demand and generation conditions) and the known output $\widehat{\mathbb{Y}}$ (the known power losses). This paradigm used to train the model is called supervised learning [Bishop, 2006]. By doing this, the active power losses output mapping obtained from the input demand and generation conditions will implicitly contains the knowledge about the problem domain, in this case the power flow equations [Kumar et al., 2017].

To update the edge weights of the model, the most widely used technique is Back-

Propagation (BP), a gradient descent-based algorithm that allows the training of the model to perform a specific task through the calculation of the edge weights values [Schmidhuber, 2015]. Gradient descent is a practical method to find the minimum of difficult functions. BP allows to avoid overshooting by taking smaller steps if the gradient of the function gets shallower. In this case BP minimises a loss function denoted with $\mathfrak{L}$. The selection of the loss function depends on the learning paradigm (supervised, unsupervised or reinforcement) and the activation function chosen [Simon S, 1999].

A widely-used loss function in supervised learning is the Mean-Squared Error (MSE), indicated in (4.44) which consists of minimising the mean squared error between the output power losses, and the target value of power losses for the π scenario.

$$\mathfrak{L} = \frac{1}{\nu} \sum_{\pi=1}^{\nu} \left| \hat{p}_{loss,T}^{(\pi)} - p_{loss,T}^{(\pi)} \right|^2 = \mathfrak{L}(w_{i,j}) \tag{4.44}$$

Where ν is the batch size, i.e the number of samples that are presented to the model until the weight matrix is updated. Th edge weights are updated $\Delta w_{i,j}$ by using the gradient of the loss function with respect to the weights as indicates expression (4.45).

In the training stage, some parameters have to be considered regarding the size of the DNN (by means of the number of hidden layers N_h and number of neurons per layer s) and regarding the learning process by means of the learning rate η [Bishop, 2006].

The weight updates of BP can be done via stochastic gradient descent.

$$\Delta w_{i,j} = w_{i,j}(\hat{t}+1) - w_{i,j}(\hat{t}) = \eta \frac{\partial \mathfrak{L}}{\partial w_{i,j}}, \ \ \forall w_{i,j} \in \mathbb{W} \tag{4.45}$$

Where η is the leaning rate and $\hat{t}$ is a time discrete parameter that indicates the iteration training step. The learning rate η influences the speed and quality of learning and its selection is fundamental, since a high value can boost the learning process but the risk of obtaining suboptimal solutions increases, while a too low value of the learning rate could lead to a more accurate training at an expense of longer training operations.

From expression (4.45) it can be seen that the sign of the gradient $(\partial \mathfrak{L}/\partial w_{i,j})$ indicates how the error varies (whether directly with the weight or inversely). Therefore, the weight must be updated in the opposite direction, "descending" the gradient. By regulating the

learning process through the value of η ensures that the model learn from the data reducing the impact of outliers and/or noise.

The model training algorithm comprise the following steps:

1. **Weight matrix initialisation**: At first $(\hat{t} = 0)$ all the edge weights $w_{i,j}(\hat{t})$ of the edge weights matrix $\mathbb{W}(\hat{t})$ are randomly assigned as indicates (4.46).

$$\mathbb{W}(\hat{t}) \sim U(\underline{w_{i,j}^{min}(0)}, \overline{w_{i,j}^{max}(0)}) \tag{4.46}$$

Where $U(\bullet)$ is the uniform distribution, $\underline{w_{i,j}^{min}(0)}$ is the lower bound for the initialisation of the weights and $\overline{w_{i,j}^{max}(0)}$ is the upper bound.

2. **Propagation of errors**: The first batch $(n_\nu = 1)$ of demand and generation conditions input samples $\mathbb{X}^{(\pi)} = \{\mathbb{X}^{(1)}, \dots, \mathbb{X}^{(\nu)}\}$, belonging to the training data set, is presented to the model and it's propagated forward through the model, layer by layer, until it reaches the output layer, obtaining the corresponding output power losses $\hat{p}_{loss,T}^{(\pi)}(\hat{t}) = \{\hat{p}_{loss,T}^{(1)}(\hat{t}), \dots, \hat{p}_{loss,T}^{(\nu)}(\hat{t})\}$. This calculated sequence of output power losses is compared with the target power losses $p_{loss,T}^{(\pi)}$ calculating the mean-squared output error $\epsilon_o^{(\pi,\hat{t})}$ as indicates (4.47).

$$\epsilon_o^{(\pi,\hat{t})} = \frac{1}{\nu} \sum_{i=1}^{\nu} |\hat{p}_{loss,T}^{(i)}(\hat{t}) - p_{loss,T}^{(i)}|^2 \tag{4.47}$$

The output error is back propagated from the output layer to the previous layers obtaining the error of the hidden layer $L_k, \forall k \in \{1, \dots, h\}$ defined as $\epsilon_k^{(\pi,\hat{t})}$ which reflects the contribution of that layer to the output error. This is calculated by the product between the output error and the column $W_k(\hat{t}) \subset \mathbb{W}(\hat{t}) = [w_{1,k}(\hat{t}), \dots, w_{n_k,k}(\hat{t})]^T$ (n_k is the number of neurones of the layer L_k) of the edge weight matrix $\mathbb{W}(\hat{t})$ corresponding to the hidden layer r as indicates (4.48).

$$\epsilon_k^{(\pi,\hat{t})} = W_k(\hat{t}) \ \epsilon_o^{(\pi,\hat{t})}, \ \ \forall k \in (1, \dots, h) \tag{4.48}$$

The error of the hidden layer $\epsilon_k^{(\pi,\hat{t})}$ is a column vector $(n_k \times 1)$ and contains the error contribution of each neuron of the hidden layer.

3. **Gradient descent**: The gradient of the loss function (which is the slope of the loss function respect to the edge weights value) is obtained by applying the chain rule to the expression (4.49) obtaining the expression (4.49) [Rashid, 2017].

$$\frac{\partial \mathfrak{L}}{\partial W_k(\hat{t})} = -\epsilon_k^{(\pi,\hat{t})} \cdot O_k(I_{n_k} - O_k) \cdot O_{k-1}^T, \quad \forall k \in (1, \dots, h) \tag{4.49}$$

Where $O_k = [o_1, \dots, o_{n_k}]$ is the neuron's outputs of the layer L_k, I_{n_k} is a unit vector vector of size $(1 \times n_k)$ and $O_{k-1} = [o_1, \dots, o_{n_{k-1}}]$ is the neuron's outputs of the hidden layer L_{k-1}.

4. **Weights updating**: The weights matrix is updated then column by column with the corresponding loss function gradient calculated considering a learning rate η . The updated edge weight column $W_k(\hat{t}+1)$ of the layer L_k is calculated as the old one $W_k(\hat{t})$ moving in the opposite direction to the gradient as is indicated in (4.50).

$$W_k(\hat{t}+1) = W_k(\hat{t}) - \eta \frac{\partial \mathfrak{L}}{\partial W_k(\hat{t})}, \quad \forall k \in (1, \dots, h) \tag{4.50}$$

5. **Recalculate output**: One the entire weight matrix $\mathbb{W}(\hat{t}+1)$ has been updated for the first batch, the next batch of samples $(N_v = 2)$ is presented to the model and propagated throw it, obtaining a new sequence of output power losses values. The output error is then recalculated and if the difference between the actual error and the previous one is below a maximum limit (ϵ^{max}), the process is stopped and the model is trained, if not the process go back to step 3.

$$\begin{aligned} \epsilon_o^{(\pi,\hat{t}+1)} - \epsilon_o^{(\pi,\hat{t})} < \epsilon^{max} &\Rightarrow \texttt{Ok} \\ \epsilon_o^{(\pi,\hat{t}+1)} - \epsilon_o^{(\pi,\hat{t})} > \epsilon^{max} &\Rightarrow \texttt{Go to 3} \end{aligned} \tag{4.51}$$

Training data

The training data set is composed by a sequence of known inputs, i.e known demand and generation conditions for each representative feeder, and a sequence of known outputs , i.e the corresponding power losses associated with those demand and generation conditions, which are calculated throw unbalanced power flows.

One requisite of the DNN models is the need of a large training data set. That becomes in a limitation if only available smart meters measurements are used. To tackle with this limitation, demand-generation scenarios are synthetically created for the $\hat{k}$ representative feeders obtained in the clustering process. For each reference feeder a collection of APs for the demand as well as for the generation are synthetically created based on the smart meters measurements, which is gathered from the AMI systems of the whole distribution area. With this, the need for a of large data set is fulfilled. These demand and generation APs created for each representative feeder represents the aggregation of all the customers connected to the feeder (in the case of demand), and all the DG units connected to the feeder (in the case of generation) units connected to that feeder and consist of mean expected value of each customer/DG unit plus a uncertainty component modelled with the standard deviation.

Moreover, demand and generation are time dependent variables in the sense that their values change over the season of the year and even over the different days (working days, weekends and holidays). To consider this variability, synthetically created APs are created for winter and summer season as well as for working-days and non-working days (holidays and weekends).

Demand APs:

1. Demand AP for Winter & Working days $p_{d,r}^{wi+wd}(t)$.

2. Demand AP for Winter & Non-Working days $p_{d,r}^{wi+nwd}(t)$.

3. Demand AP for Summer & Working days $p_{d,r}^{su+wd}(t)$.

4. Demand AP for Summer & Non-Working days $p_{d,r}^{su+nwd}(t)$.

Generation APs:

1. Generation AP for Winter $p_{g,r}^{wi}(t)$.

2. Generation AP for Summer $p_{g,r}^{su}(t)$.

Those APs for demand and generation are obtained from the smart meter measurement over a year. Then based on that, a large number of scenarios is synthetically

produced using the mean expected value $\mu_{(\bullet)}$ of the original APs and adding a component of uncertainty by means of the standard deviation $\sigma_{(\bullet)}$ as indicates expression (4.52).

$$
\begin{aligned}
D_1 &: p_{d,r}^{wi+wd(\omega)}(t) = \mu_{p_{d,r}^{wi+wd}(t)} \pm \sigma_{p_{d,r}^{wi+wd}(t)} \\
D_2 &: p_{d,r}^{wi+nwd(\omega)}(t) = \mu_{p_{d,r}^{wi+nwd}(t)} \pm \sigma_{p_{d,r}^{wi+nwd}(t)} \\
D_3 &: p_{d,r}^{su+wd(\omega)}(t) = \mu_{p_{d,r}^{su+wd}(t)} \pm \sigma_{p_{d,r}^{su+wd}(t)} \\
D_4 &: p_{d,r}^{su+nwd(\omega)}(t) = \mu_{p_{d,r}^{su+nwd}(t)} \pm \sigma_{p_{d,r}^{su+nwd}(t)} \\
G_1 &: p_{g,r}^{wi(\omega)}(t) = \mu_{p_{g,r}^{wi}(t)} \pm \sigma_{p_{g,r}^{wi}(t)} \\
G_2 &: p_{g,r}^{su(\omega)}(t) = \mu_{p_{g,r}^{su}(t)} \pm \sigma_{p_{g,r}^{su}(t)}
\end{aligned}
\tag{4.52}
$$

Where $\omega \in (1, \ldots, \kappa)$ is the index for demand and generation scenarios (which is a daily AP) and κ is the total number of daily AP of each type produces ($\kappa > 1000$).

As the input vector contains both demand and generation tuples of each representative feeder and has been defined 4 demand APs and 2 generation APs to take into account the variability throw the year, its necessary to combine each demand AP (D_1, D_2, D_3 and D_4) with the 2 generation APs (G_1, G_2), that is 8 pairs $D_1 - G_1, D_1 - G_2, \ldots, D_4 - G_1, D_4 - G_2$, where each block represents the κ APs produced. Therefore, for each representative feeder a total number of scenarios $\xi = 8\,\kappa$ is synthetically produced as is illustrated in Fig. 4.8.

The input $\mathbb{X}^{(\omega,1)}$ includes the κ demand and generation daily APs of the combination $D_1 - G_1$ for each representative feeder (i.e $2\hat{k}\kappa$ patterns). As there are 8 combinations, the total input training data (denoted with $\Omega_{T,i}$) set will be composed by the $8(2\hat{k}\kappa)$ demand and generation daily APs as indicates (4.53) where ω denotes the index of APs scenarios created $\forall \omega \in (1, \ldots, \kappa)$.

$$
\Omega_{T,i} = \{\mathbb{X}^{(\omega,1)}, \mathbb{X}^{(\omega,2)}, \ldots, \mathbb{X}^{(\omega,7)}, \mathbb{X}^{(\omega,8)}\}
\tag{4.53}
$$

Moreover, the output power losses training set (denoted with $\Omega_{T,o}$) will be composed by the correspond active power losses daily patterns obtained by unbalanced power flow algorithm based on the formulation presented in [Garcia et al., 2000]. As is illustrated in Fig. 4.8 for each combination of demand and generation AP, it is calculated the active power losses corresponding with the κ AP for each combination.

$$\Omega_{T,o} = \{p_{loss,T}^{(\omega,1)}, p_{loss,T}^{(\omega,2)}, \dots, p_{loss,T}^{(\omega,7)}, p_{loss,T}^{(\omega,8)}\} \tag{4.54}$$
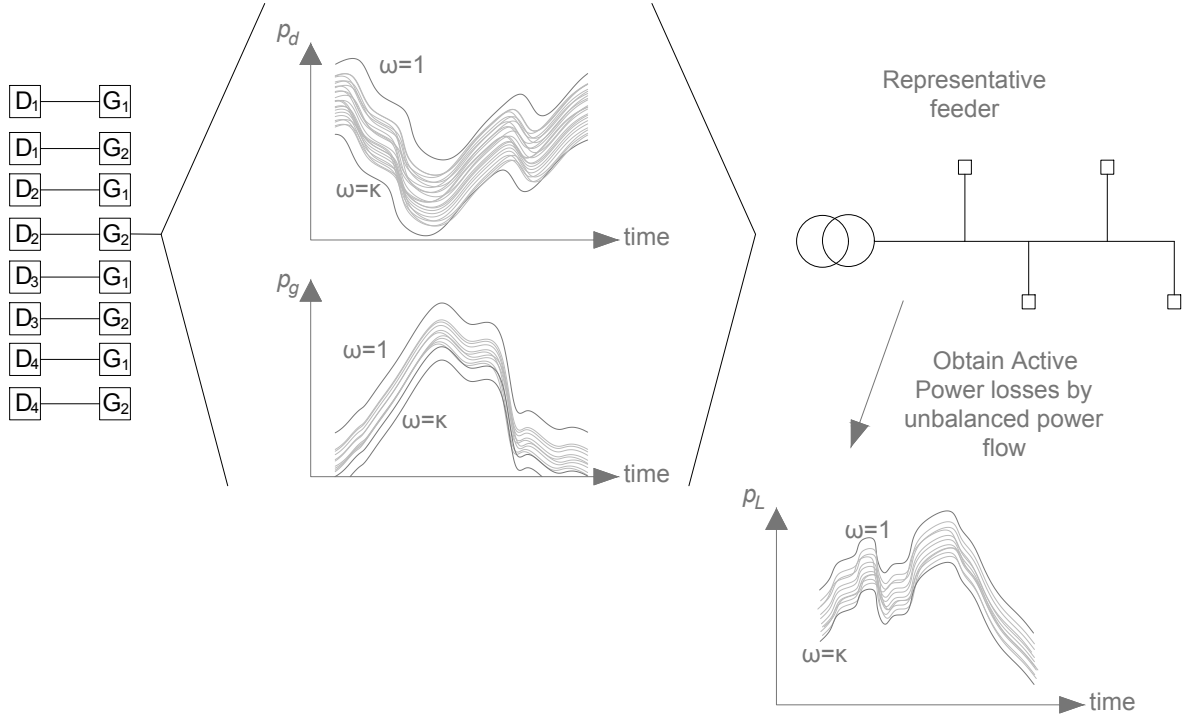
Figure 4.8: Output power losses set calculation throw unbalance power flow

Both input and output training sets together compose the training data set Ω_T as indicates (4.55).

$$\Omega_{T,i} \cup \Omega_{T,o} := \Omega_T \tag{4.55}$$

Model hyper-parameters

One of the more critical parts of the modelling of the DNN is the selection of the DNN control parameters to characterise the model structure and performance. These control parameters are known as hyper-parameters and their configuration plays a crucial role in the performance of the deep learning model. The configuration of the hidden layers architecture through the connections between the neurons determines the final

behaviour of the model since in each hidden layer takes place a non-linear transformation of the aggregate activation of the DNN [Simon S, 1999]. Regarding this, as the size of the model increase, the more flexibility it is obtained to perform accurately however this is at expense of a higher computationally expense. Moreover to avoid overfitting (when the model perform so close to the training data but fails with new data) some of the neurons of the model has to be cancelled (no output), this is achieved throw the technique known as "Dropout". Finding the best configuration for hyper-parameters is not trivial [Bergstra and Bengio, 2012]. In the field on machine learning, there are two main strategies for optimising hyper-parameters: Grid search and Random search [Goodfellow et al., 2016a]. Grid search consists of searching exhaustively through a previously specified set of hyper-parameter candidate values which leads to the most accurate model. On the other hand, random search consists of taking values randomly for each hyper-parameter within an interval defined by an upper bound and a lower bound, and retaining, consequently, the best combination found. It can be said that Random search is very fast but does not guarantee the optimal combination of hyper-parameters. On the contrary, grid search guarantees that the optimal combination of hyper-parameters will be reached although the cost of the computational time increases. In this thesis, the grid search strategy is adopted to tune the hyper-parameters of the deep learning model because it is the method that provides the optimal combination.

Consequently, previously to the model training, a set of control parameters have to be defined to characterise the model structure and performance which are the following:

- **No. of hidden layers** h: Larger deep neural networks are likely to obtain better results since the model have more opportunities to learn independent representations.

- **No. of neurons per hidden layer** n_k: A large number of neurons per hidden layer increase accuracy but, a large number may lead the model to overfitting. In contrast, to few neurons per hidden layer may lead the model to under-fitting.

- **Dropout** ζ: Is a regularisation technique to avoid overfitting which consist of cancelling some neurons of the hidden layers [N.Srivastava et al., 2014]. A widely-use

good starting point is to cancel between 20% of the hidden neurons. Cancelling a low number of neurons produces minimal effect and cancelling a high number of neurons may lead the model to under-learning. Neurons are randomly dropped out during training. This means that their contribution to the activation of downstream neurons is temporally removed on the forward pass and any weight updates are not applied to the neuron during the backward pass. Neurons Weights are tuned for specific features providing some specialisation. Neighbouring neurons rely on this specialisation, which if taken too far can result in a fragile model too specialised to the training data. The effect is that the network is capable of better generalisation and is less likely to overfitting the training data.

- **Learning rate** η: is a coefficient to control how the model change depending on the error response, since the weights of the deep learning model are updated based on that error.

- **Weights Initialisation** $\{\underline{w_{i,j}^{min}(0)}, \overline{w_{i,j}^{max}(0)}\}$: The weight initialisation scheme depends on the activation function chosen. In this chapter sigmoid function is chosen so in general $w_{i,j} \in (0,1)$. The lower bound of the minimum weight could be defined randomly as $\underline{w_{i,j}^{min}(0)} \in (0, 0.5)$ means while the upper bound of the maximum weight could be defined randomly $\overline{w_{i,j}^{max}(0)} \in (0.5, 1)$.

- **Number of epochs** ϑ: Is the number of times that the complete training data is presented to the model in the training stage. The number of epochs have to be increased until the validation accuracy (i.e the output error obtained in the validation stage) decreases (not in the case of the training accuracy because that means overfitting).

- **Batch size** ν: Is the the number of training samples which are presented to the model, before the edge weights are updated. The training data set can be divided in one or more batches. Commonly used values for the batch sizes are 32, 64, and 128 samples. If the dataset cannot be divided evenly then some samples could be ignored.

The weights of a DNN have to be calculated via an empirical optimisation procedure known as Stochastic Gradient Descent (SGD) [Robbins and Monro, 1951]. This procedure is an optimisation algorithm that estimates the error gradient for the current state of the model using samples from the training dataset, then updates the weights of the model using the back-propagation algorithm. The optimisation problem addressed by SGD for DNN is challenging and the space of solutions (sets of weights) may be comprised both by global optimum solutions and also by many local optimum solutions [Brownlee, 2016].

One of the most important hyper-parameter to tune for the DNN is the learning rate to achieve good performance. Learning rate controls how quickly or slowly the DNN learns. The amount that the weights are updated during training is determined by the step size or the learning rate. Given a perfectly configured learning rate, the model will learn how to approximate the function, the best, given the available resources (the number of layers and the number of nodes per layer) and the considered training epochs numbers (passes through the training data). Generally, a large learning rate allows the model to learn faster, at the cost of arriving to sub-optimal weights. A small learning rate value allows the model to reach an optimal or even globally optimal set of weights but may take significantly longer to train. Moreover, a learning rate that is too large will result in weight updates that will be too large and the performance of the model (such as its loss on the training dataset) will oscillate over training epochs. A learning rate that is too small may never converge or may get stuck on a suboptimal solution. The range of values to consider for the learning rate is from 0.01 to 10^6. During training, the back- propagation algorithm estimates the error for which the node weights are involved. Instead of updating the weight with the full amount, it is scaled by the learning rate. This means that a learning rate of 0.1 would mean that weights are updated a 10%. Configuring the learning rate is challenging and time-consuming. The learning rate hyper-parameter controls the rate or speed at which the model learns. Specifically, it controls the weights updates during the training and also the end of the training batch examples. The learning rate will interact with many other aspects of the optimisation process, and the interactions may be nonlinear. In general, low learning rates require more training epochs. Conversely, large learning rates require fewer training epochs. Further, low batch sizes are better

suited to low learning rates given the noisy estimate of the error gradient.

Diagnostic plots can be used to investigate how the learning rate impacts the rate of learning and the learning dynamics of the model. Line plots can provide information about the rate of learning over training epochs (fast/slow); if the model has learned too quickly (sharp rise and plateau) or is learning too slowly (little or no change); If the learning rate might be too large via oscillations in loss.

4.7.5 Model Validation

The model hyper-parameters tuning process involves finding the combination of model hyper-parameter values that produces the most accurate model, i.e. the combination that maximise the accuracy Λ of the model defined throw the APE (Absolute Percentage Error) index [Goodfellow et al., 2016a] in expression (4.56)

$$\Lambda = 1 - APE = 1 - \frac{|p_{loss,T} - \widehat{p}_{loss,T}|}{p_{loss,T}} \tag{4.56}$$

Where $(\widehat{p}_{Loss,T})$ is the value of active power losses estimated by the model and $(p_{loss,T})$ is the known active power losses value.

To obtain the combination of model hyper parameter values that produce the most accurate model, for each model hyper parameter, a collection of n different candidate values are defined as indicated in (4.57) and in Table (4.3). For instance, for the particular case of the hyper-parameter "number of hidden layers h", its candidate values are uniformly distributed between a minimum of two hidden layers and a maximum of 20 hidden layers.

It is well-known that validation is usually performed by dividing the data into training, test, and validation sets [Kuhn and Johnson, 2013]. However, in some cases, this strategy leads to suboptimal hyper-parameters [Russel and Norving, 1994]. In this thesis, is proposed the $\mathcal{K}$-Fold technique to validate the power losses model, which has been found to produce a more accurate model because the variability of the data set is completely considered. It must be noted that the selection of the number of $\mathcal{K}$ folds is not obvious. The aim is that the number of $\mathcal{K}$-Folds must be chosen in such an order as to be statistically representative of the problem to solve. Large $\mathcal{K}$ values provide higher accuracy at the expense of a higher computational effort. Meanwhile, small $\mathcal{K}$ values will be faster

Table 4.3: Candidate values of the hyper-parameters' model for $\mathcal{K}$-fold process.

Hyper-parameters	Candidate values				
No. Hidden Layers (h)	2	3	5	10	20
No. Hidden neurons (n_h)	4	6	8	12	16
Dropout (ζ)	0.05	0.10	0.20	0.25	0.5
Learning rate (η)	1e-6	1e-4	1e-3	1e-2	1e-1
Batch size (ν)	0.1ξ	0.25ξ	0.5ξ	1.0ξ	2ξ
Upper edge-weight $\overline{w_{i,j}(0)}$	0.6	0.7	0.8	0.9	1.0
Lower edge-weight $\underline{w_{i,j}(0)}$	0.1	0.2	0.3	0.4	0.5

but less accurate The number of $\mathcal{K}$ folds has to be chosen such that each train/test group of data samples is large enough to be statistically representative. Values $\mathcal{K} = \in (4, 10)$ are very common in the field of applied machine learning [Hastie et al., 2017].

$$
\begin{aligned}
h &= \{h_1, h_2, \ldots, h_{n-1}, h_n\} \\
n_k &= \{n_{k_1}, n_{k_2}, \ldots, n_{k_{n-1}}, n_{k_{n-1}}\}, \quad \forall k \in (1, \ldots, h) \\
\zeta &= \{\zeta_1, \zeta_2, \ldots, \zeta_{n-1}, \zeta_n\} \\
\eta &= \{\eta_1, \eta_2, \ldots, \eta_{n-1}, \eta_n\} \\
\nu &= \{\nu_1, \nu_2, \ldots, \nu_{n-1}, \nu_n\} \\
w_{i,j}^{min} &= \{w_{i,j}^{min}{}_1, w_{i,j}^{min}{}_2, \ldots, w_{i,j}^{min}{}_{n-1}, w_{i,j}^{min}{}_n\} \\
w_{i,j}^{max} &= \{w_{i,j}^{max}{}_1, w_{i,j}^{max}{}_2, \ldots, w_{i,j}^{max}{}_{n-1}, w_{i,j}^{max}{}_n\}
\end{aligned}
\tag{4.57}
$$

The optimal number of epochs ϑ is chosen by applying the "Early Stopping" technique which consists of increasing the number of iterations for training the model as long as the accuracy (Λ) keeps increasing.

$\mathcal{K}$-fold technique consists of splitting the training data set into $\mathcal{K}$ subsets as follow:

$$
\Omega_T := \underbrace{\Omega_{T,x}^{(1)} \cup \ldots \cup \Omega_{T,x}^{(\kappa)}}_{\text{Input training data } \mathbb{X}} \cup \underbrace{\Omega_{T,p}^{(1)} \cup \ldots \cup \Omega_{T,p}^{(\kappa)}}_{\text{Output training data } \mathbb{Y}}
\tag{4.58}
$$

- Initially, the `split 1` is formed by one test fold (Fold 1), subsets $(K = 1)$ and the remaining subsets are used as training data set $K = (2, \ldots, \mathcal{K})$. For the training
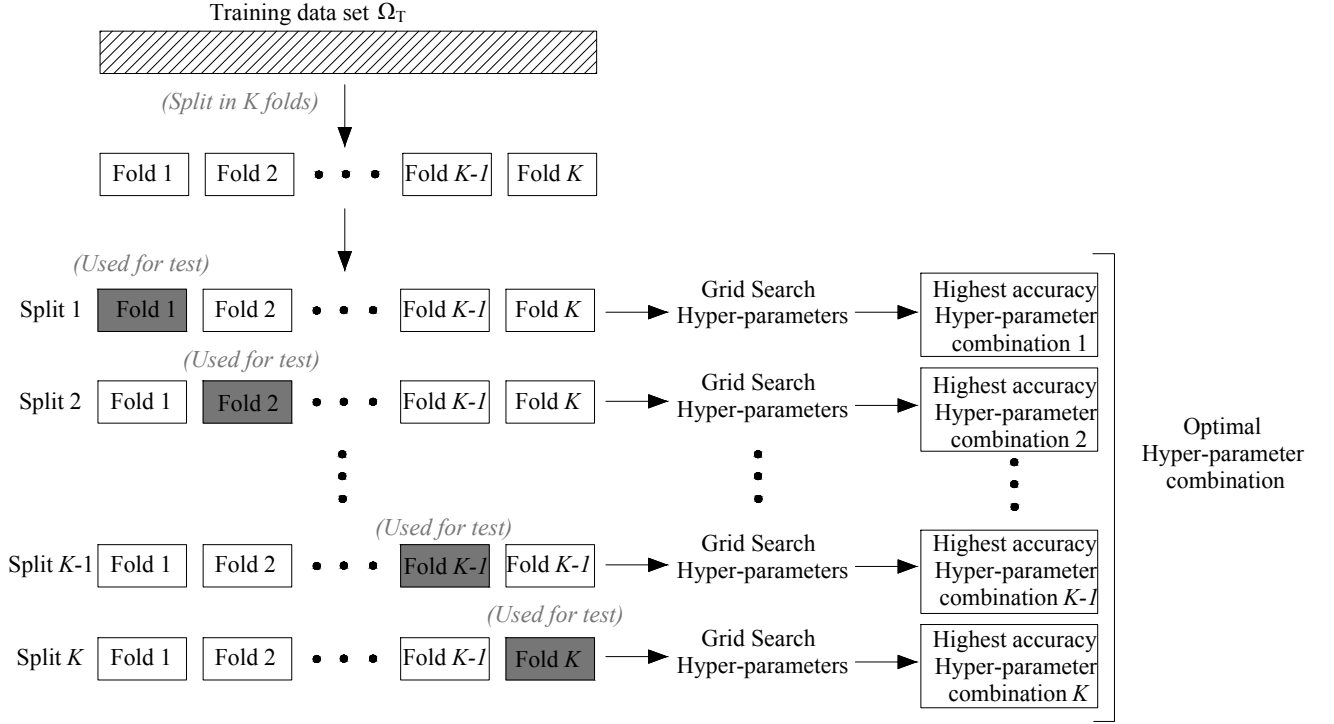
Figure 4.9: $\mathcal{K}$-Fold Cross validation and hyper-parameter tunning procedure

data set, the $n_{hp}^{n_{hp}}$ hyper-parameter parameters combinations are explored to find the most accurate model. This is done by means of building a model with each one of these hyper-parameter's combinations, and then training the model with the training data set of the `split 1`. Each model is tested with the test data set (Fold 1), obtaining the power losses accuracy given by (4.56). From all the hyper-parameter combinations, the one that is chosen produces the highest accuracy. The optimal hyper-parameter combination values, denoted by $\Upsilon^{(1)}$, are retained and the process continues with the next split `split 2`.

$$\Upsilon^{(1)} = \{h^{(1)*}, n_k^{(1)*}, \zeta^{(1)*}, \eta^{(1)*}, \nu^{(1)*}, w_{i,j}^{max(1)*}, w_{i,j}^{min(1)*}\} \qquad (4.59)$$

- In the `split 2` the next fold is taking as test data $K = 2$ and the first fold and the remaining ones are taking as training data $K = (1, 3, \ldots, \mathcal{K})$. For this split the operation is the same to obtain the second hyper-parameter combination values $\Upsilon^{(2)}$ that produces the most accurate model, i.e with he highest accuracy $\Lambda^{(2)}$.

- The process is repeated with the upcoming splits until the `split` $\mathcal{K}$, obtaining $\mathcal{K}$ hyper-parameter combination values associated with $\mathcal{K}$ accuracy values. From those accuracy scores, the hyper-parameter combination associated with the highest is chosen for the finally trained model. The process is illustrated in Fig. 4.9

4.8 Case Study

In this section, the proposed methodology for power loss estimation has been applied to an existing large-scale LV smart grids. The low-voltage distribution area under investigation is located in Madrid (Spain) and the principal characteristics are indicated in Table 4.4. The distribution area is composed of 147 SS with 1,256 feeders (so $n = 1265$). There are 30,429 residential and commercial customers with a total contractual power of 546 MW. In this area, the contractual DG penetration level is 55% and they are based mainly on rooftop PV panels, which are the typical DG source in Madrid region. The average smart meter deployment is 88%, which provides hourly measurements from telemetered customers. To deal with the connection of non-telemetered customers, the hourly power demand is estimated as detailed in section 3.3.4 of this thesis. From the utility company, it is known that the average power phase unbalance due to the unevenly distributed single-phase customers has been found to be 13.8%. The geographical area corresponds to a whole metropolitan ZIP code which covers 605 ha. Data related to cables electrical properties (resistance and reactance) considered are indicated in Table 1.2 and 1.3.

4.8.1 Data

It can be noted that the whole area offers great variability in many parameters, such as:

- Feeder properties: length, impedance, type of connection (aerial, underground).

- Customers: phase connection (three-phase, single-phase), type of demand (residential, commercial, industrial), connection point.

- DG: phase connection (three-phase, single-phase), power rate, connection point.

- Smart meter deployment: telemetered and non-telemetered customers.

- Phase unbalanced.

Table 4.4: Case study data for large-scale power losses estimation

Case study property	Value
Geographical Area Covered (ha.)	605
No. Secondary Substations (No. SS)	147
No. Customers (three phase, single-phase)*	30429
No. Feeders (n)	1256
Power Contracted (MW)	546
Accumulated feeder Length (km)	273
Max. (PV-based) DG-presence level (%)	55
Aver. Smart Meter penetration level (%)	88
Aver. Power Phase Unbalance (%)	13.8
Aver. Ratio Customers/SS	207
Aver. Ratio Customers/Feeder	24
Network Type	100 % Urban
Cables Material	100 %Aluminium
Cables impedance	(0.1 , 3) Ω/km
Network Configuration	80 %Underground
	20 %Overhead

*residential, commercial and industrial

For that reason, the data set must be selected to represent the high network diversity but focusing on the more relevant ones from the point of view of network power losses. Fig. 4.10 shows the histograms of the $\tilde{p}$ feeders's characteristics (according to Table 4.4) and the Probability Density Function (PDF) with a red line. It can be seen that the feeder's characteristics follow three different PDFs:

- The characteristics related to power consumption ($X_{1,a}$, $X_{1,b}$, $X_{1,c}$, $X_{9,a}$, $X_{9,b}$, $X_{9,c}$,

$X_{10,a}$, $X_{10,b}$, $X_{10,c}$, $X_{11,a}$, $X_{11,b}$, $X_{11,c}$, $X_{12,a}$, $X_{12,b}$, $X_{12,c}$) presents a Gaussian distribution.

- Meanwhile, the characteristics related to feeder properties (X_2, X_4, X_7, X_8 and X_{13}) are left-skewed.

- The rest of the characteristics such as smart meter penetration (X_3), DG spatial distribution (X_5), and customers per phase ($X_{14,a}, X_{14,b}, X_{14,c}$) are right-skewed.
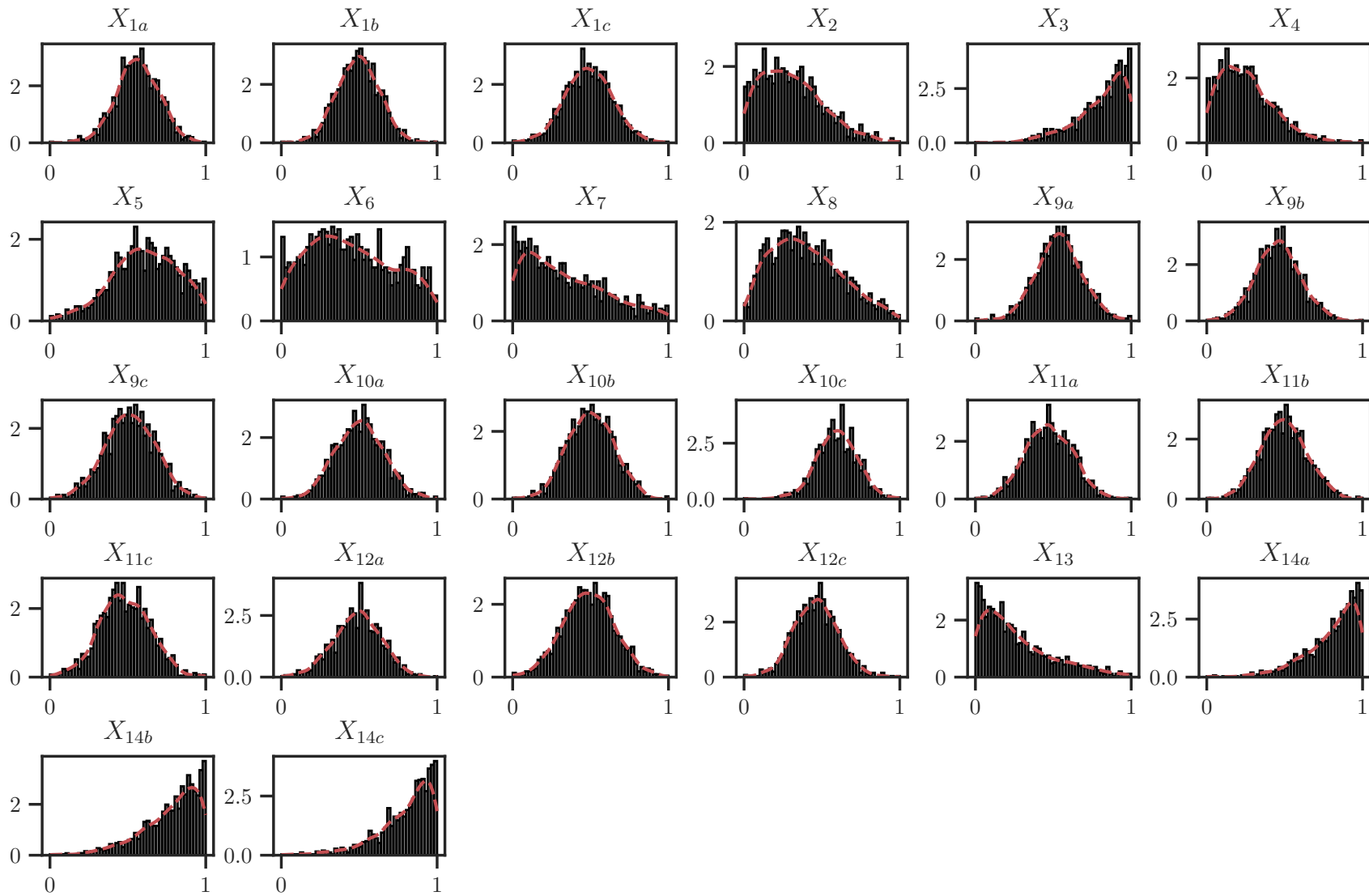
Figure 4.10: PDFs and histograms of the feeders' characteristics

From Fig.4.10, the following conclusions can be deduced:

- It can be seen that phase A is more loaded than the other phases. Consequently, the large low-voltage distribution area is characterised by an unbalanced operation $(X_{1,a}, X_{1,b}, X_{1,c})$.

- The feeder loading level characteristic X_2 varies from 5% to 75%, which means that there are some heavily loaded feeders (75% of the power rating of the transformer) but also that there are others with very reduced loading (5% of the power rating of the SS transformer). It can be observed, through the histogram, that the PDF is tilted to the left side. Then the power contracted by the customers is located primarily between 10% and 50% of the power transformer rating (since the feeder load level is defined as the ratio between power contracted and the power rating of the SS transformer).

- The smart meter deployment (characteristic X_3) varies from a minimum of 55% to the maximum, which is 100%.

- The DG penetration level, which corresponds to the characteristic X_4, is tilted to the left in such a way that the majority of the feeder has a DG penetration of 50% (ratio between power peak installed and power contracted). Related to this, the self-consumption ratio (X_7) shows that over half of the feeders have a self-consumption ratio less than or equal to 50%.

- The spatial distribution of DG units, X_5 units, and customer's X_6 are very close because, usually, the DG facilities are allocated at the same customer's connection point (PV rooftop facilities).

Fig. 4.11 shows the correlation matrix of the feeder's characteristics. As can be seen the feeder's characteristics related to load unbalance are correlated positively $(X_{1,a} - X_{1,b})$ and negatively $(X_{1,b} - X_{1,c})$, which means that when more loading is added to one phase, it comes at the expense of reducing the loading at other phases. Otherwise, the phases will be balanced.

Also positively correlated are the feeders' characteristics $(X_2 - X_3)$ and $(X_7 - X_{13})$. Due to the fact that in Spain smart meter installations are mandatory for customers with power contracted equal to or less than 15 kW, it is expected that highly loaded feeders are prone to have customers with higher power contracted. In addition, it can be noted that highly loaded feeders presented high correlation self-consumption ratios.
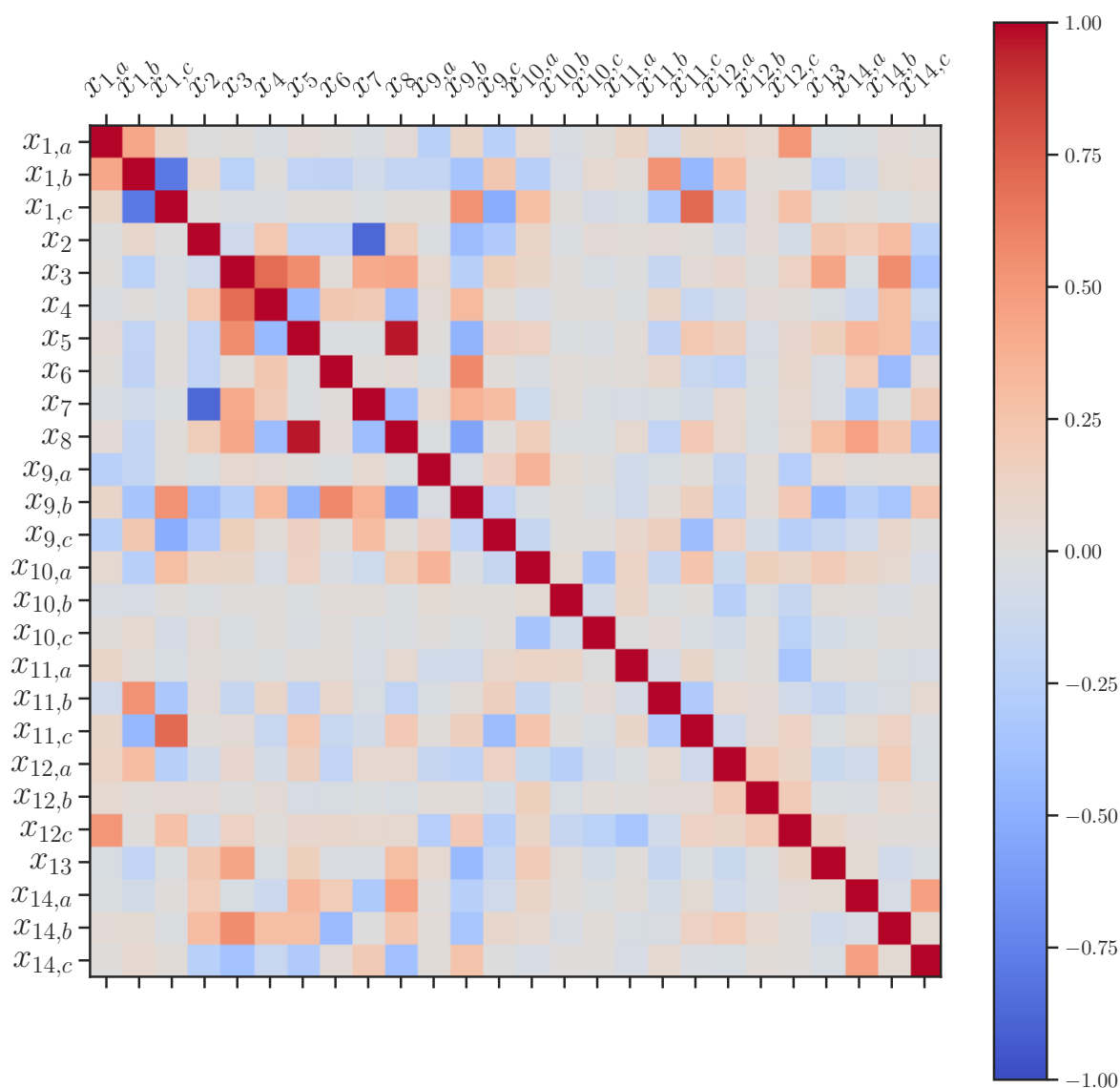

Figure 4.11: Correlation matrix of the feeder's characteristics

4.8.2   Results

Principal Component Analysis

From Fig. 4.10 , it can be noted that there are some feeder characteristics that provide very high values (X_{12}) and others that present small values (X_6). To give the same weight to all feeder characteristics, a normalisation process is performed. In this case, the feeder's characteristics are normalised to vary in the range from 0 to 1. The box plot of each normalised feeder characteristic is obtained and is shown in Fig. 4.12. Each box plot indicates the median value (red medium vertical line in the box) and the inter-quartile range defined by the upper limit, indicating the 75% percentile (Q_3) and the lower limit, indicating the 25% percentile (Q_1). Also, extremes in values are shown which are considered outliers (those feeder samples with a value of up to 50% of the Q_1 and Q_3 limits). It can be noted that the normalisation process maintains the statistical behaviour of the feeder's characteristics, allowing the comparison between them. The feeder's properties that exhibit a clear normal distribution (i.e., $X_{1,a}$, $X_{1,b}$, and $X_{1,c}$) have the inter-quantile range around the median zero. The feeder's characteristics that have skewed statistical behaviour keep maintaining the majority of the feeder's samples in the extremes (for instance, DG penetration level X_4).

Because some of the feeder's characteristics are correlated (Fig. 4.11), a feature extraction process is carried out to obtain a reduced set of uncorrelated feeder's features. This is achieved through PCA analysis, as was explained in section 4.5. The objective of the PCA is to capture the maximum data variability using a reduced number of uncorrelated variables. The selection of the PC projections is performed by (4.9). Fig. 4.13 shows the cumulated variability captured by each Principal Component. It can be seen that the first PC $(k = 1)$ is able to capture 53% of the data variability in the original feeder's characteristics. However, this percentage increases if the second projection PC $(k = 2)$ is added. The first and two PCs together capture up to 79% of the data variability of the original feeder's characteristics. It can also be seen that increasing the number of PC projections does not notably increase the capture of data variability ($\approx 10\%$ each additional PC projection added).
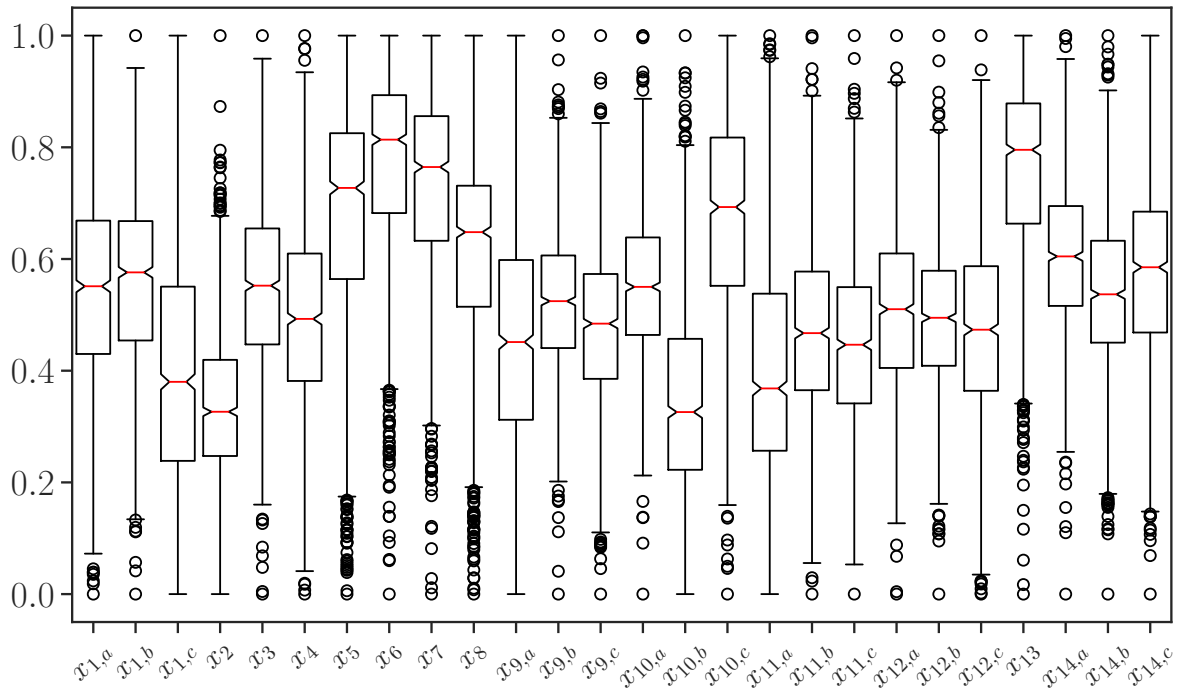
Figure 4.12: Boxplots of the normalised feeder's characteristics

Consequently, for this case, the PCA analysis allows for the reduction of the normalised feeder's data matrix to the two-dimensional feeder's data matrix, considerably reducing the computational burden but maintaining up to 79% of the original data variability.

Clustering and representative feeders selection

The $n = 1256$ feeders of the two-dimensional feeder's data matrix $\tilde{M}^{n \times s}$ are classified into feeder's clusters by means of the `K-means++` procedure explained in section 4.6. The selection of the optimal number of feeder's clusters $\hat{k}$ is not always straightforward but comes from a compromise between a reduced number of clusters and a large value of the Global Silhouette Coefficient (GS). The `K-means++` algorithm has been executed 1,000 times using different initial centroid seeds.

Fig.4.14 shows the evolution of the GS coefficient as the number of clusters is increased (red line), as well as the percentage of variance explained (blue line), which is calculated using the statistic F-test [Hayter, 2012]. It can be appreciated that for $\hat{k} = 2$ clusters, a maximum value of $GS_{\hat{k}=2} = 0.48$ is achieved. However, the percentage of variance ex-
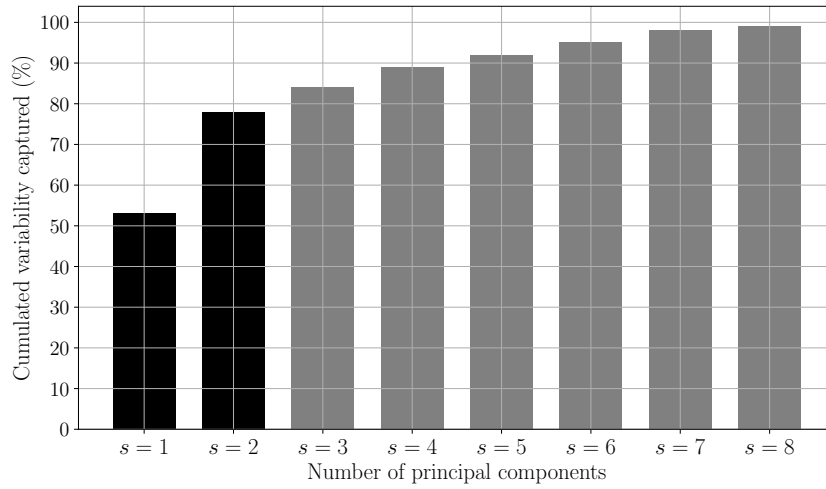
Figure 4.13: Cumulated variability captured by Principal Component of the PCA analysis

plained by two clusters is very low (40%). Consequently, a compromise value of $\hat{k} = 4$ feeder's clusters provides good percentage of variance explained [Hayter, 2012] and a good GS value [Mateo et al., 2018, Rigoni et al., 2016, Bracale et al., 2012, Strunz et al., 2009].

Fig. 4.15.a) shows the scatter plot of the feeder's data for each of the $\hat{k} = 4$ feeder's clusters. Within each feeder's cluster, a representative feeder $\tilde{c}_r$ is chosen as the feeder closest to the centroid's cluster. Finally, Fig. 4.15.b) shows the silhouette index value (4.23) for the chosen clustering configuration (i.e., $\hat{k} = 4$ feeder's clusters). The vertical red line indicates the Silhouette Index value for the chosen clustering configuration. As can be seen, only a few feeder's samples are below the zero limit.

Table 4.5 shows the Euclidean distance factor α_r (used to scale each representative feeder input in the model) as well as the number of feeder's clusters assigned to each cluster.

Table 4.5: Euclidean distance factor per feeder cluster and number of feeder's per cluster

Representative Feeder $\tilde{c}_r$	Euclidean distance factor α_r	No. feeders per cluster n_r
$r = 1$	1.43	314
$r = 2$	1.64	287
$r = 3$	1.75	345
$r = 4$	1.82	310
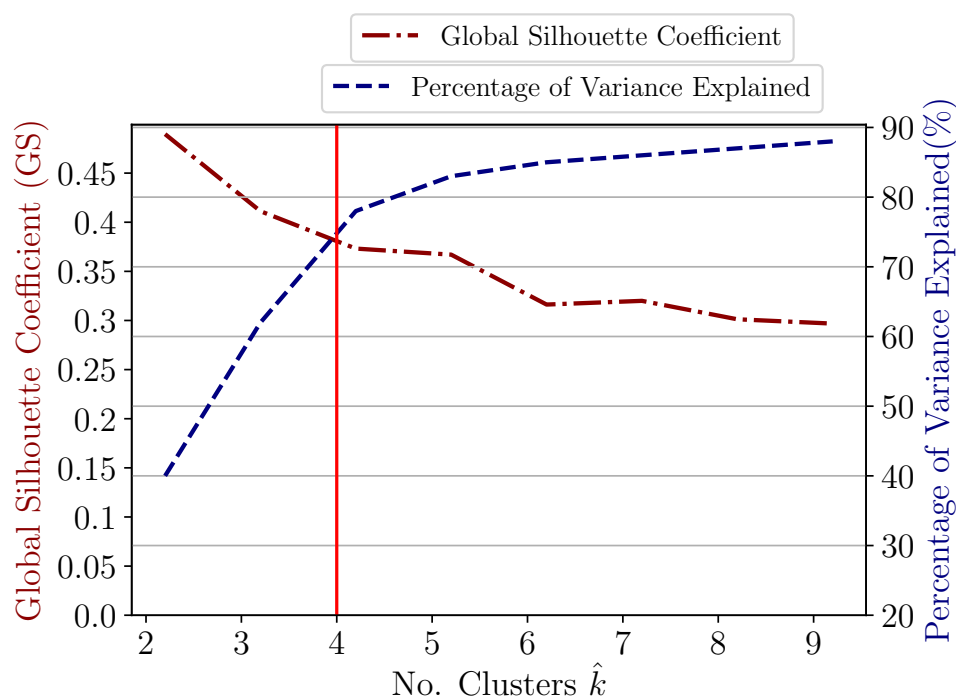
Figure 4.14: Selection of the optimal number of clusters
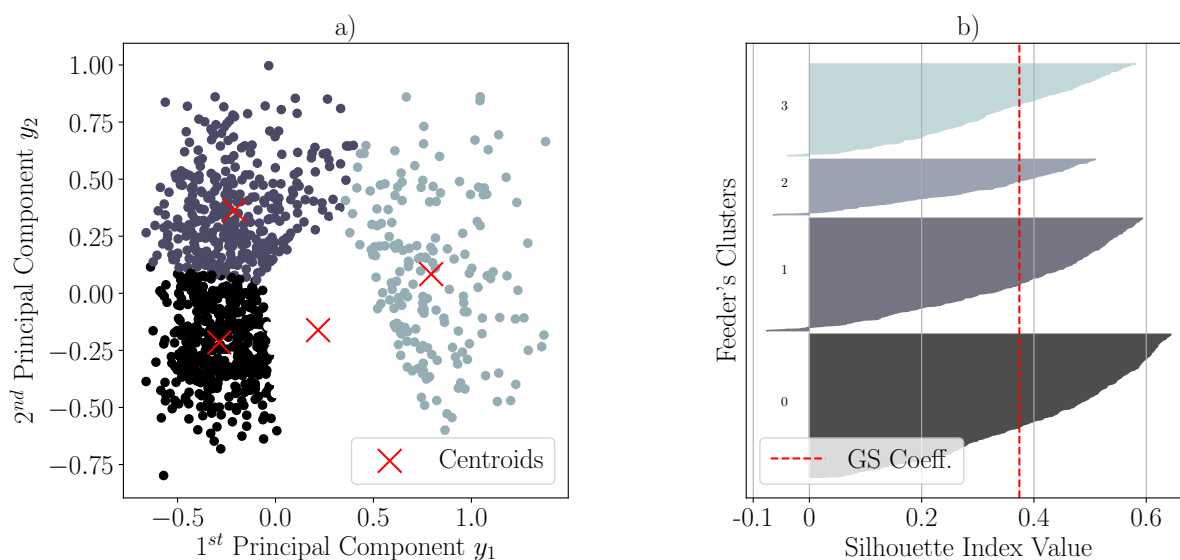
Figure 4.15: Feature extraction and clustering results: a) Scatter plot of the feeder's clusters obtained through first and second PCs, b) Silhouette Index Value plot for $\hat{k} = 4$ feeder's clusters.

Model Training

The training data set Ω_T has been synthetically produced using the smart meter data of the Spanish large-scale low-voltage distribution area gathered over the period from 2014-

2018 with a time resolution of 10 min. Moreover, the generation data from each of the DG units for the same period of time is also available.

For each reference feeder, a collection of APs for the demand and generation has been synthetically created based on yearly smart meter measurement data. These demand and generation APs represent the aggregation of all the customers connected to the reference feeder (in the case of demand APs), and the aggregation of all the DG units connected to the reference feeder (in the case of generation APs).

Demand and generation APs are formed by means of the mean expected value $\mu_{(\bullet)}$ of the time series of demand/generation of the representative feeders plus a uncertainty component modelled through the standard deviation $\sigma_{(\bullet)}$ of the same demand/generation time series. With that mean expected value and standard deviation, a large number of synthetic APs $\kappa > 1000$ are produced.

Moreover, to consider the variability of demand and generation throughout the year (seasons, working days, holidays, etc.) the synthetically produced APs are discriminated for the winter and summer seasons as well as for Working-Days (WD) and Non-Working Ways (NWD) (holidays and weekends) as indicated in Table 4.16. Each scenario has been simulated in every representative feeder, and the corresponding active power losses have been obtained by solving an unbalanced power flow problem (Fig. 4.16) [Velasco et al., 2019]. Fig. 4.17 shows a daily sample of the training data produced.

Table 4.6: Synthetic demand and generation APs formulas

		Demand APs		
Winter	WD	$D_1 : p_{d,r}^{wi+wd(\omega)}(t) = \mu_{p_{d,r}^{wi+wd}(t)} \pm \sigma_{p_{d,r}^{wi+wd}(t)}$		
	NWD	$D_2 : p_{d,r}^{wi+nwd(\omega)}(t) = \mu_{p_{d,r}^{wi+nwd}(t)} \pm \sigma_{p_{d,r}^{wi+nwd}(t)}$		
Summer	WD	$D_3 : p_{d,r}^{su+wd(\omega)}(t) = \mu_{p_{d,r}^{su+wd}(t)} \pm \sigma_{p_{d,r}^{wi+wd}(t)}$		
	NWD	$D_4 : p_{d,r}^{su+nwd(\omega)}(t) = \mu_{p_{d,r}^{su+nwd}(t)} \pm \sigma_{p_{d,r}^{su+nwd}(t)}$		
		Generation APs		
Winter		$G_1 : p_{g,r}^{wi(\omega)}(t) = \mu_{p_{g,r}^{wi}(t)} \pm \sigma_{p_{g,r}^{wi}(t)}$		
Summer		$G_2 : p_{g,r}^{wi(\omega)}(t) = \mu_{p_{g,r}^{su}(t)} \pm \sigma_{p_{g,r}^{si}(t)}$		

$\omega \in (1, \dots, \kappa)$ is the index for demand and generation scenarios (which is a daily AP) and κ is the total number of daily APs of each type synthetically generated.
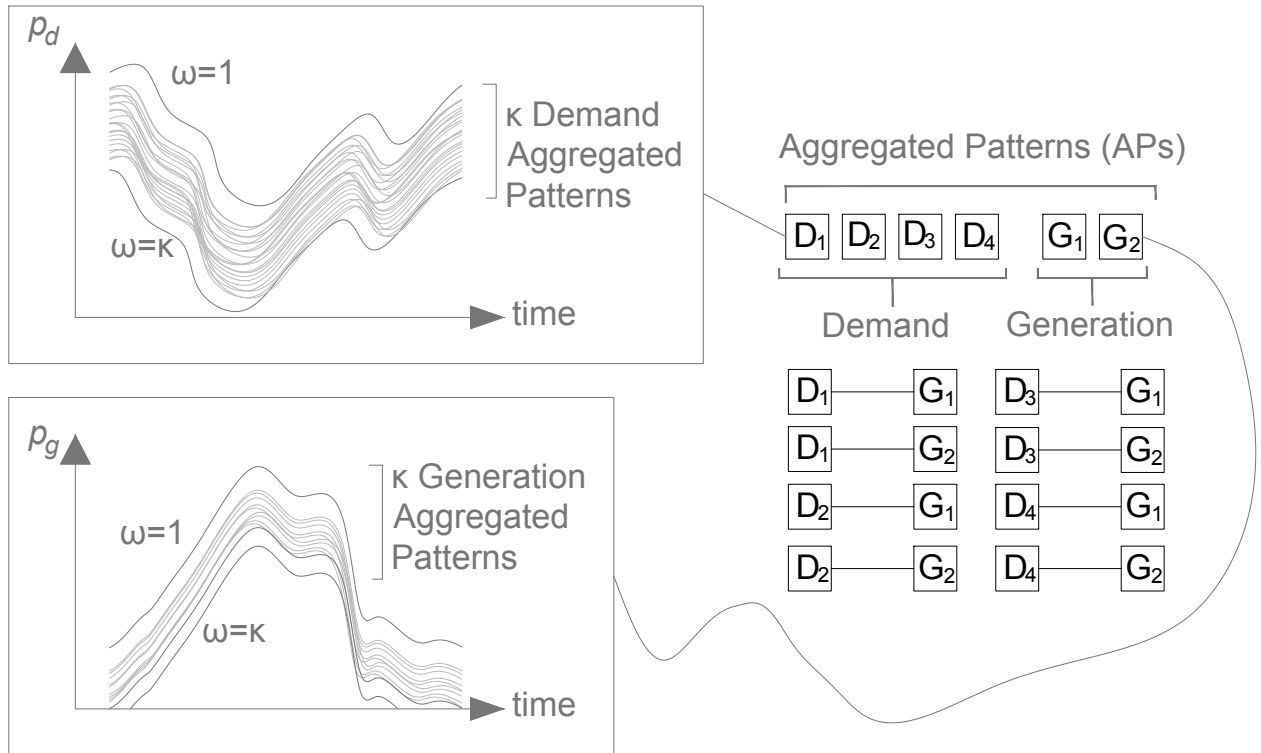
Figure 4.16: Combination of demand APs and generation APs for one representative feeder
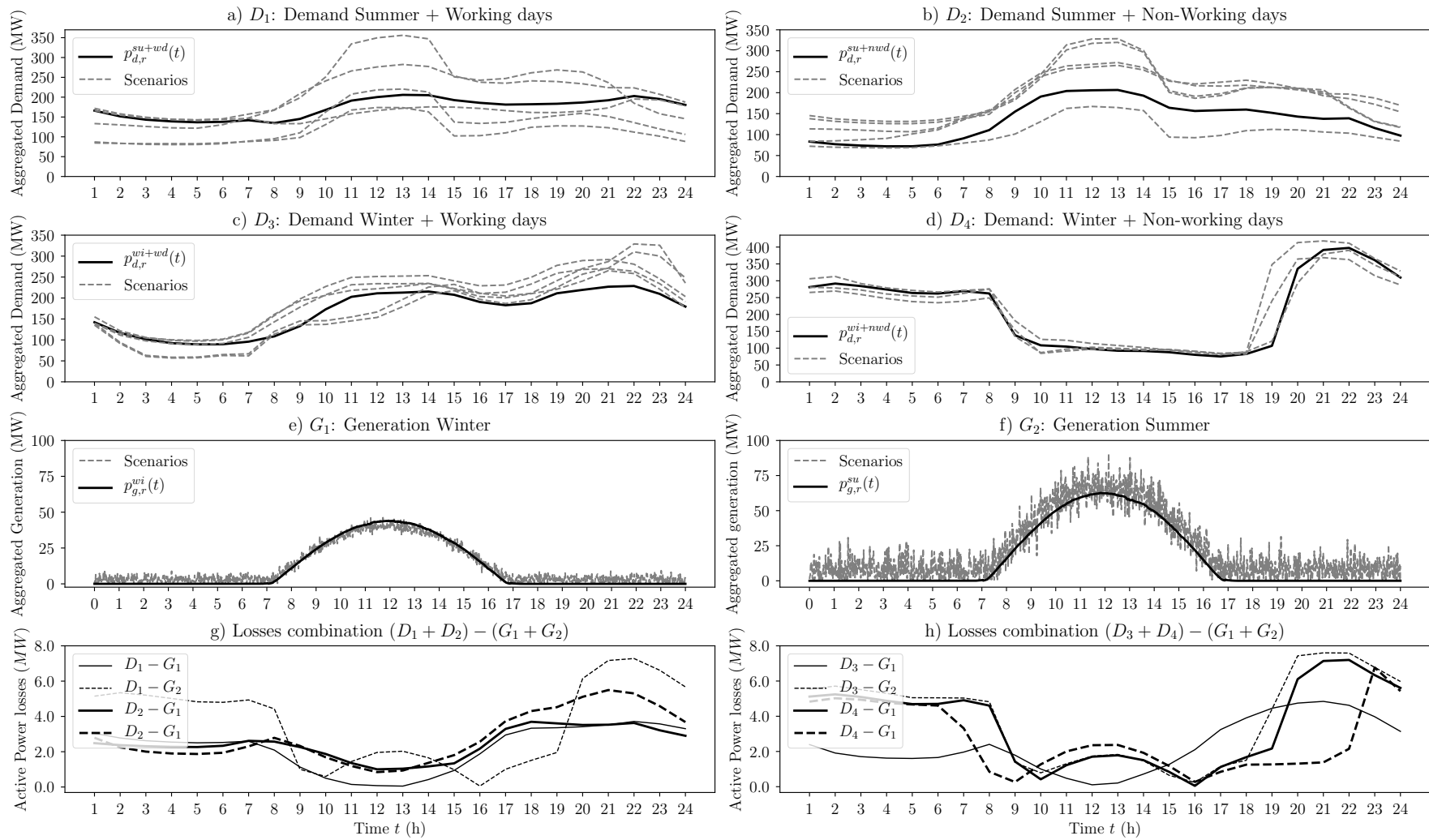
Figure 4.17: Training data for the deep learning losses model

Model Validation

For the $\mathcal{K}$ Cross-Validation procedure, a number of $\mathcal{K} = 4$ folds have been chosen to split the training data set [Hastie et al., 2017]. Table 4.7 shows the hyper-parameter values associated with the highest accuracy score trajectories obtained. The model with the higher accuracy corresponds to $\mathcal{K}$-Fold 1, where the architecture of the model consists of four layers: the input layer has eight input signals (aggregated demand and aggregated generation for each cluster); the first hidden layer has six neurons; the second hidden layer has four neurons; and, finally, the output layer generated the total power losses of the large-scale area. It can be seen that a predominant batch size of 240 provides the model with the highest accuracy. The final architecture is shown in Fig.4.18.

Table 4.7: Hyper-parameter combinations with the highest accuracy for each split

Accuracy Λ	$\mathcal{K}$ Fold	Architecture $(n_1 : \ldots : n_h)$	Dropout ζ	Learning rate η	Weight initialisation $(w_{i,j}^{min}, w_{i,j}^{max})$	Batch size ν
0.9365	1	(8:6:4:1)	0.24	1e-2	(0.06,0.75)	240
0.9271	2	(8:6:4:1)	0.32	1e-3	(0.27,0.52)	240
0.9205	3	(8:4:8:1)	0.41	1e-2	(0.35,0.60)	240
0.9200	4	(8:12:12:1)	0.38	1e-4	(0.02,0.72)	48
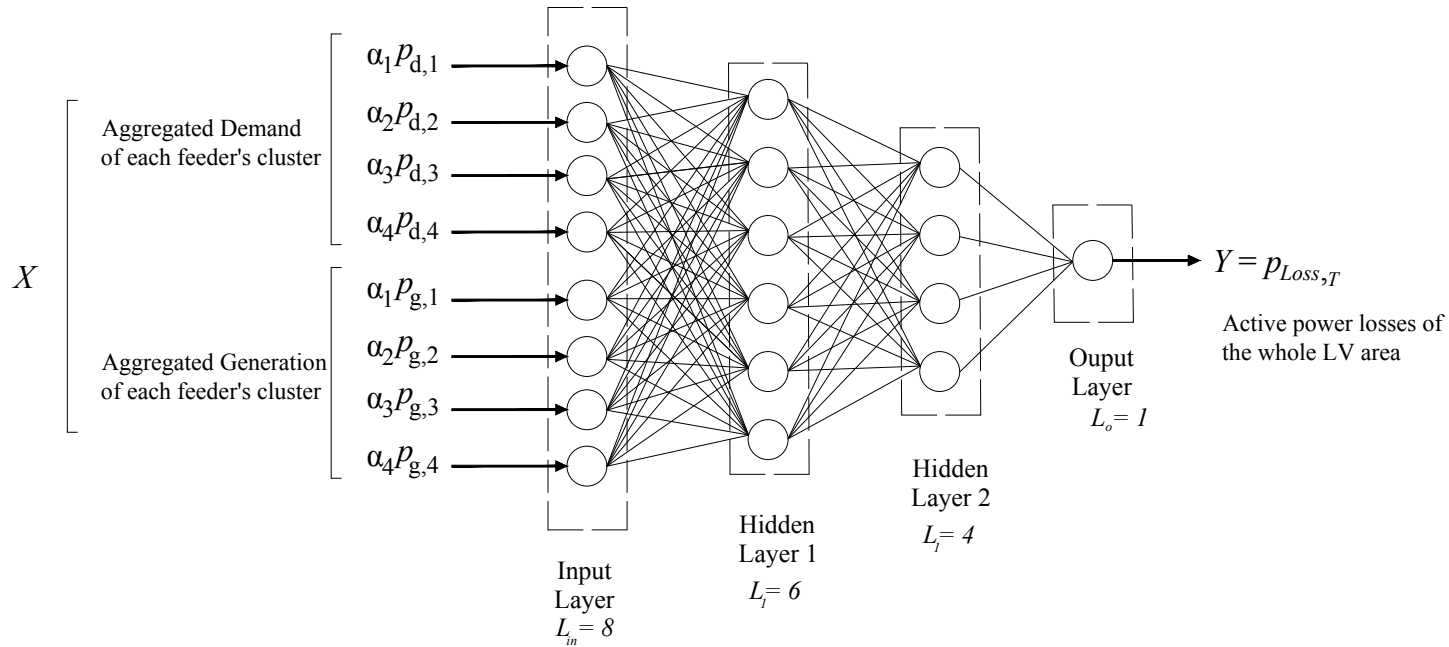
Figure 4.18: Final architecture of the DNN model

DNN Power losses estimation in unbalanced networks

The power losses model presented in this chapter has been trained and validated using Python programming language and the libraries SCIKIT-LEARN [Pedregosa et al., 2011] and TENSORFLOW [Abadi et al., 2015].

Once the deep learning losses model has been trained and validated, it can be applied for the power losses estimation at large-scale for the whole set of low-voltage smart grid networks for a 24-hour period with a time resolution of 10 min. (i.e., the model is fed with the demand and generation data with a time resolution of 10 min.). Fig. 4.19 shows the technical losses estimation on a daily basis for the whole large area. Fig. 4.19.a) shows the aggregated power demand, Fig. 4.19.b) shows the aggregated power generation coming from the PV-based units of each representative feeder. Fig. 4.19.c) shows the losses estimation using the deep learning losses model for the large-scale low-voltage distribution area (solid line); the dashed line shows the losses obtained by an unbalanced power flow. Finally, Fig. 4.19.d) shows the accuracy of the model over the day, which results in an average value of 88% outperforming the results obtained in [Leal et al., 2009, Wang et al., 2017].

Fig. 4.20 shows the power losses estimation for the unbalanced large area by applying the deep learning model proposed in the paper for a period of four weeks (solid line), and the exact power losses obtained with the unbalanced power flow solution (dashed line), it has to be highlighted that the power flow requires the exact network topology and customer/generation data. It has to be noted that the overall accuracy of the deep learning model over a four-week period is 87% compared to the exact power flow solution. Moreover, the computational speed of the deep learning model is almost 10 times inferior to that of the traditional approach (power flow), for this case study the power flow takes 1 hour and the DNN 7 minutes.

Comparative

As is pointed out in the introduction to this chapter, not many methodologies have been proposed, in the literature, to evaluate the power losses in low-voltage large-scale smart grids, considering both unbalanced operation as well as DG. However, to pro-
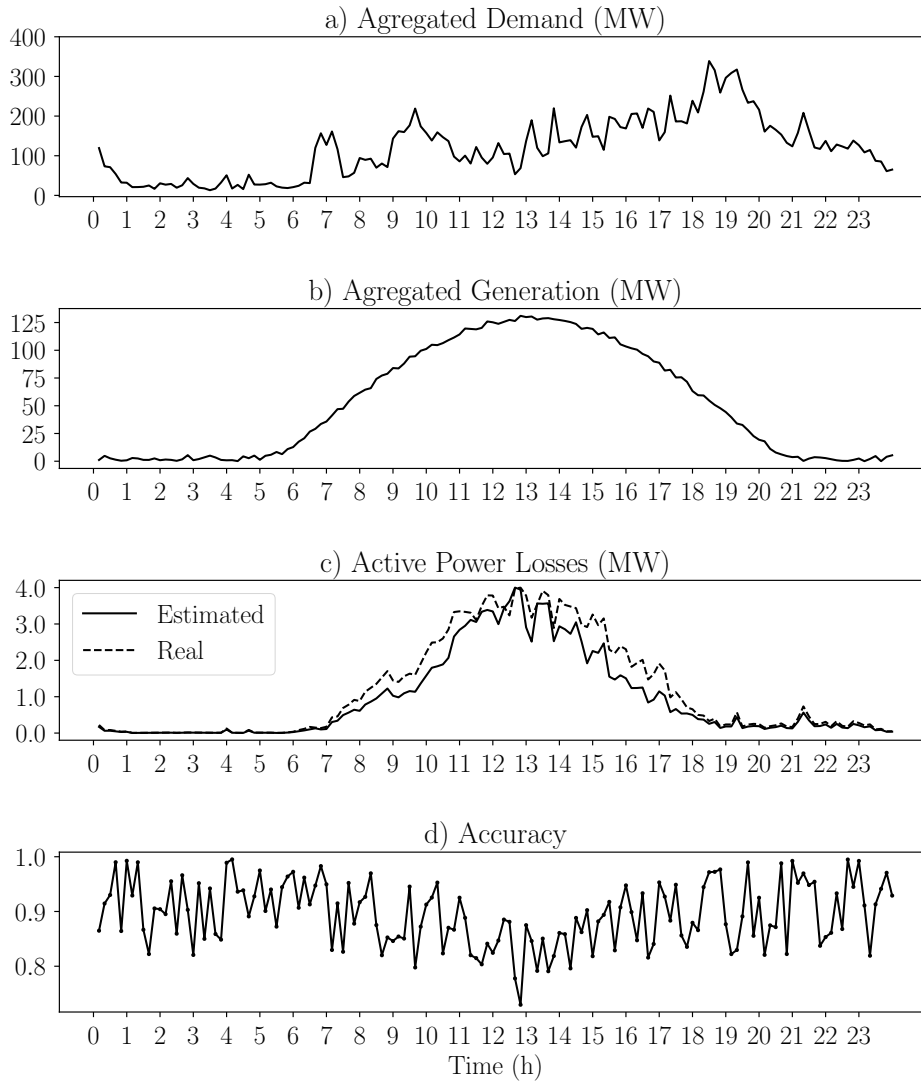
Figure 4.19: Active power losses estimation results on one daily basis

vide a comparative result, the proposed deep learning loss model using out-of-the-box demand and generation data is compared to the losses estimation method proposed by [Heckmann et al., 2013]. It must be noted that [Heckmann et al., 2013] cannot be applied to unbalanced networks.

Fig.4.20 shows the comparison among losses estimation for large low-voltage distribution areas using the proposed deep learning losses model (solid line), the real losses estimation using determinist power flow (dark dashed line), and the losses estimation method proposed by [Heckmann et al., 2013] (grey dashed line). As can be seen in the upper plot, the deep learning method outperforms the results of [Heckmann et al., 2013],
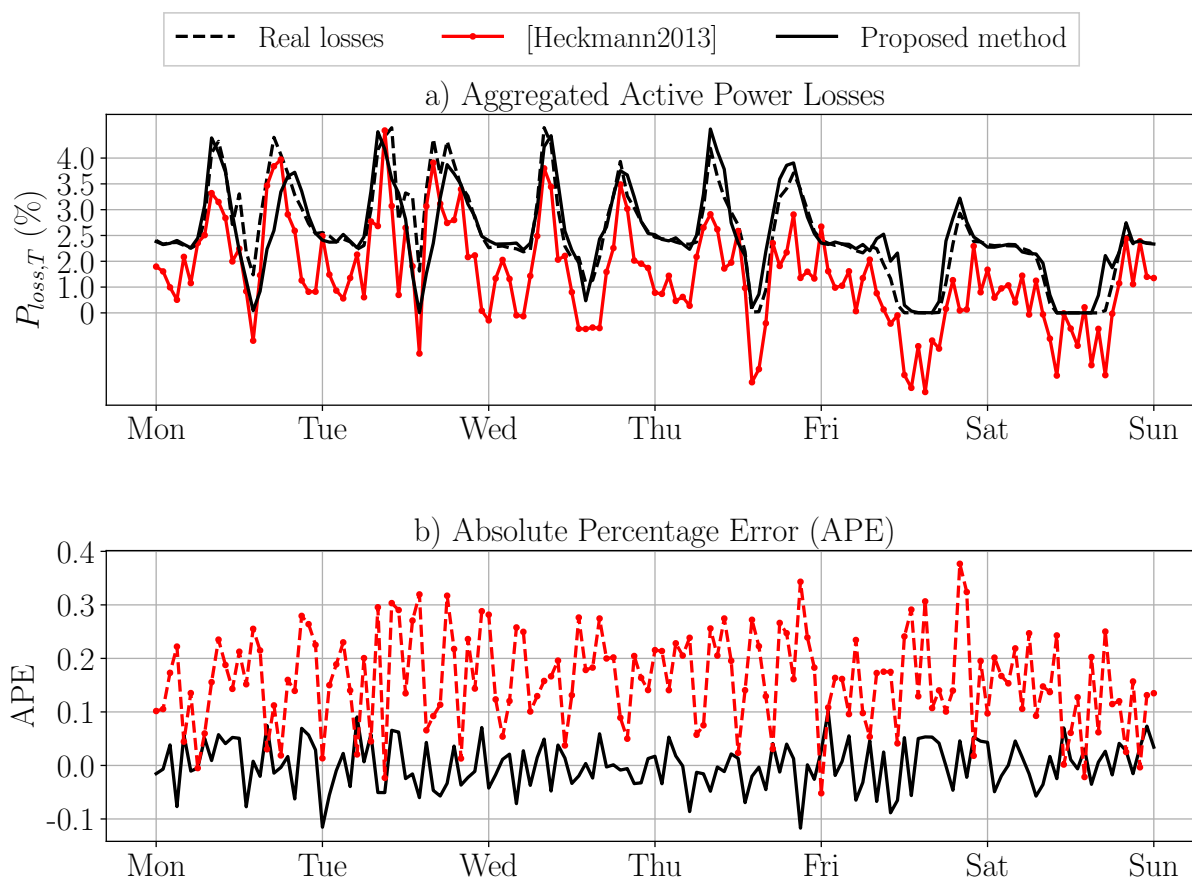
Figure 4.20: Active power losses estimation comparison between the proposed method and the method proposed in [Heckmann et al., 2013] under balanced situations (7-13 October, 2019)

which, due to the lack of an accurate consideration of DG generation, provides negative values of power losses at certain times of the week. The lower plot of the same figure shows the APE results of both methods (the proposed in this paper and the applied in [Heckmann et al., 2013]) compared to the exact power flow results. It can be appreciated that the deep learning method provides very small error values, showing its superior accuracy.

4.9 Conclusions

This chapter of the thesis presented a deep learning-based model for power losses estimation in large-scale low-voltage distribution areas with presence of distributed resources as well as unbalanced conditions. First, the architecture and the operational behaviour of the low-voltage distribution networks have been described. Then, a feature extraction process based on Principal Components Analysis is carried out obtaining a reduced set of features that captures the majority of the variability of the feeder's data. Following that a `k-means++` clustering process is applied to obtain the representative feeders as the closest feeder's to the centroids of the clusters obtained. After that, a supervised power losses model based on a Deep Neural Network is formulated for the whole low-voltage distribution area under study. To train the model, a set of demand-generation scenarios has been synthetically created for those representative feeders, and their corresponding power losses in unbalanced conditions has been calculated using the three-phase power flow model depicted in this thesis. The learning process of the model is conducted following the $\mathcal{K}$-fold procedure obtaining at the same time the optimal combination of model's hyper-parameters. The model customised with the optimal hyper-parameters is then feed with the demand and generation data coming from the metering infrastructure of the representative feeders, scaled up to the number of feeder belonging to the feeder's cluster which represent. The power losses model has been applied to a real utility low-voltage distribution area in Madrid (Spain) exhibiting a good performance as power losses estimator. The proposed methodology has been proved to be a potential real-time operation tool to improve energy efficiency level in large distribution areas with high penetration of distributed resources.

Chapter 5

Power Losses Minimisation in Unbalanced LV Smart Grids Under Uncertainty

5.1 Introduction

Environmental concerns have led to an ever-increasing incorporation of renewable-based Distributed Energy Resources (DERs) in actual distribution systems, in particularly in the low-voltage levels.

Power losses growth and technical contingencies depends on the penetration level, location, and DERs type as [Soroudi, 2012]. For instance, the integration of PEVs poses potential contingencies in distribution networks since massive uncontrollable charging sessions can produce technical problems on residential LV networks (e.g., thermal problems on transformers/cables and/or significant voltage drops).

The appropriate amount and location of DERs into power networks, which is known as Hosting Capacity (HC), can contribute to the reduction of Greenhouse emissions Levels through a more sustainable energy. However, If the DER generation exceeds the network hosting capacity limit (i.e. the resources are misplaced and wrongly sized), it could lead to network's contingencies, such as: over-voltages, overloads and power losses increments.

In the upcoming years, increasing penetration levels of DERs will require robust plan-

ning and operation tools to assess the capabilities and requirements of the networks [Keane et al., 2013]. Ideally, the objective would be to maximise the energy efficiency level of the system (i.e., minimise the energy lost) and at the same time maximise the renewable energy penetration.

Moreover, due to the intermittent behaviour of those renewable-based resources, uncertainty has to be properly modelled in such a way that the operational conditions obtained will be robust enough from changes in solar irradiation, electric vehicles charging schedules or unexpected customers behaviour changes [Olivier et al., 2016]. Those circumstances make challenging the efficient and secure integration of DERs in smart grids without producing technical contingencies such as over-loads and over-voltages.

For the aforementioned reasons, flexibility mechanisms are required to mitigate those contingencies and at the same time to increase the integration of DERs into power networks. These mechanisms will be activated before the contingencies take place to keep the system within normal operational conditions being robust against any uncertainty sources (PV production or customers demand) [Soroudi and Amraee, 2013].

5.2   Flexibility Services for Power Losses Minimisation

Energy flexibility can be defined in terms of a schedule power sequence during certain horizon time (hourly, day a-head or a week in advance) with certain time resolution (from 10 minutes to 1 hour). When the flexibility is provided by a residential customer, the mechanism is known as Demand Response (DR) and consists in the modification of the customer pattern consumption (increasing or decreasing the load demand) at certain time steps. When the flexibility is provided by distributed generators (for instance PV panels) the mechanism is known as load curtailment and consists in limiting the quantity of power injected to the network [Soroudi, 2012]. Other forms of flexibility are peak shaving, which refers to the use of electricity stored in BES during peak hours, reducing the consumption of power from the grid at peak times. The use of BES system for load levelling is also known as "energy arbitrage" because it may be possible to earn a profit by

storing inexpensive electricity when demand is low and using it when the price is higher [Navarro-Espinosa et al., 2014].

When designing flexibility provider mechanisms, it has to be considered aspects related to their functional and technical feasibility:

- Customer's potential flexibility is related with the final energy use: space cooling/heating, water heating, indoor lighting, and other home appliances such as washing/dryer machine.

- DERs potential flexibility is related with their normal operation: PEV charging session, BES charging/discharging events and PV injections.

Additionally, flexibility mechanisms are required to have the following characteristics:

- Sheddability, which refers to the technical potential to load decrease. For instance if a prosumer/DER is supposed to decrease its power consumption/generation at DSO requests, the flexibility action has to be proceeded in safety conditions both for the customer/DER as well as for the network itself.

- Controllability, which refers to the capacity of the customer/DER to increase the level of power consumption/generation to provide flexibility.

- Acceptability, which refers to the availability of the customers/DERs to provide flexibility at a given time when requested.

5.2.1 Prosumer Flexibility

Nowadays LV distribution networks have been transformed into smart grids in which the end-users become active users by modifying their energy consumption and providing flexibility services to the grid. Examples of recent research papers using DR as flexibility providers can be found in [Shafie-Khah et al., 2018a, Shafie-khah et al., 2018, Heydarian-Forushani et al., 2018b, Heydarian-Forushani et al., 2018a]. Although there are works [Baran and Fernandes, 2016] which prevent DR to be a unique solution for flexibility provision to the system.

DR mechanisms encourage the reduction of customers load demand (incentive-based programs) based on economic incentives [Carreiro et al., 2017]. DR mechanisms can be implemented in different timescales, being the day-ahead the time period when Demand bidding and load curtailment are scheduled.

To cope with the mentioned requirements, in this thesis customers are considered prosumers which provide flexibility services. A prosumer is formed by a load consumption, the PEV charging point, BES device (energy storage) and distributed energy generation (by means of PV panels). With this approach, prosumers not only can demand energy from the LV network (normal customer behaviour) but also can inject energy to the grid. Within the proposed prosumer scheme, it is contemplated that each prosumer will have a triplet BES-PEV-PV as indicated in Fig. 5.1 where energy flows are shown to provide flexibility upon an operator request. The PV block (which represent an array of PV panels), injects certain quantity of energy along the day following a daily pattern close to a gaussian bell-shape which median value is around midday. It is considered that the power generation from the PV block can follow three different paths depending on: a) flexibility requirements, b) household appliance demand c) uncertainty levels or d) power losses levels in the network.
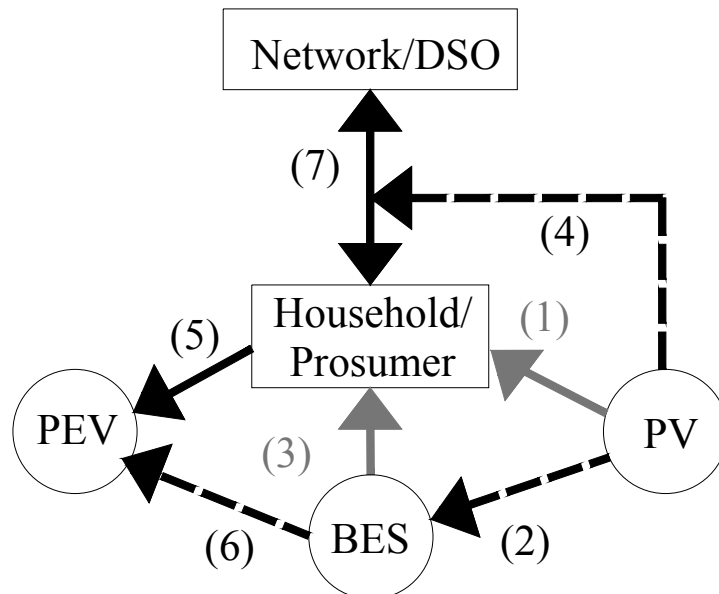
Figure 5.1: Prosumer with triplet BES-PV-PEV and energy flows

In the case a), if the DSO requires to the prosumer to provide flexibility, for instance by reducing the power consumption from the network (load curtailment), the prosumer will reduce the power consumption from the network by demanding the required power from the PV block, path (1). This is conditioned to the uncertainty related to the PV block operation, since it is characterised by an intermittent behaviour of the weather conditions (specifically solar irradiation and ambient temperature) which are intermittent and fluctuating, as well as they depend on calendar season and hours of the day (Fig. 5.2).
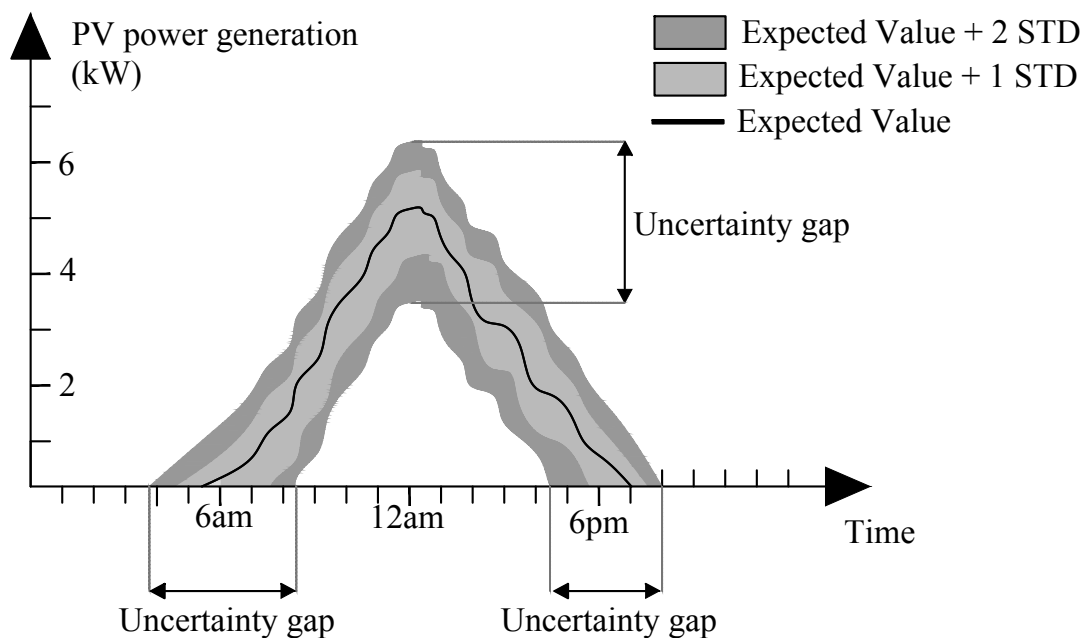
Figure 5.2: PV power production uncertainty

This PV block uncertainty implies a risk for the flexibility provision. Therefore, to tackle with the PV uncertainty, path (2) is considered in which a certain fraction of the PV power generation is used for storing the BES block (which is composed by an array of BES units) and so the stored energy can be used as a power supply to the household, path (3), under PV generation uncertainties. The charging of the BES block is subject to some technical constraints related to its maximum storage capacity and maximum/minimum energy discharge. When these constraints are close to be violated, the BES unit is allowed to discharge energy to the prosumer or the PV unit directly injects the energy to the grid,

path (4). In this case, this injection from the PV block to the network is also constrained by some network constraints. Consequently, if the PV injection does not fulfil the technical networks constraints, PV curtailment is performed to guarantee network constraints.

The prosumer uncertainty is associated to the customer demand, the PV generation, the PEV connection times (arrival and departure), PEV arrival capacity and PEV required departure capacity. The PEV block (which consist of an PEV/PHEV charger installed at the household) is modelled as a household appliance, which demand pattern is characterised by a statistical distribution with an Expected Value (EV) and Standard Deviation (STD). In this case, the PEV uncertainty is spread around the specific hour time when the vehicle is connected to the charging station, and it finishes when the vehicle leaves the household. Also, the PEV charging demand profile during the connection period is considered (Fig. 5.3).
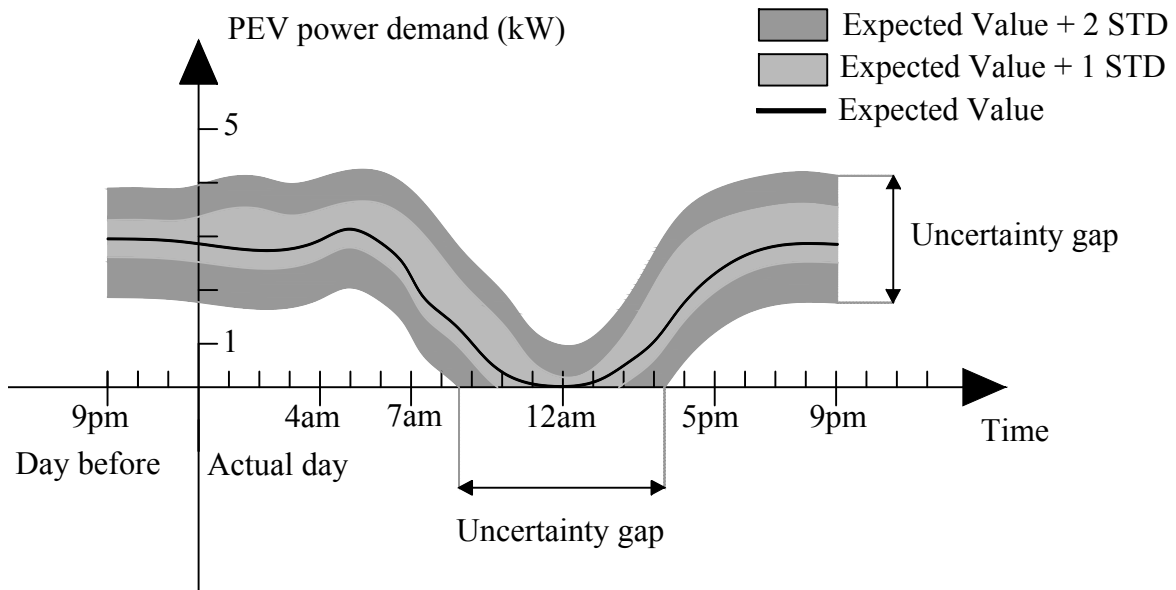
Figure 5.3: Aggregated PEV charging uncertainty model

The power consumption of the PHEV/PEV block is part of the aggregated power consumption of the prosumer, path (5). However, under flexibility requests from the operator, the PHEV/PEV power consumption can be affected. Consequently, to provide the flexibility to the grid without interrupting the PHEV/PEV charging demand, the stored energy at the BESS block will be used to charge the PHEV/PEV by path (6).

Finally, path (7) indicates that the prosumer can import energy from the network and behaves as a customer or it can export energy to the network acting as a generator. Therefore, the coordinated control of the energy flow between the BES block and PV block is used as a flexibility mechanism in the sense that when the customers are requested to provide flexibility from the DSO, they provide that flexibility by the coordinated control of the PV block energy stored in the BES unit as shown in Fig. 5.1. However, it has to be noted that the proper operation of the triplet BES-PEV-PV is affected by uncertainty.

5.2.2   DERs Flexibility

DERs-side flexibility-mechanisms are found in the literature as effective flexibility mechanisms to avoid technical contingencies. PV curtailment is proposed in [Franco et al., 2018] were the optimal PV operation point (volt/var) is determined by solving a Three-phase Optimal Power Flow (TOPF) considering a wide range of demand and PV generation scenarios. The widespread presence of PV units in distribution systems leads to overvoltages as is pointed out in [Gutierrez-Lagos et al., 2019] where an OPF-based CVR (Conservative Voltage Reduction) scheme for unbalance distribution networks (MV and LV) with high penetration of PVs is applied maintaining nodal voltages within statutory limits and minimising the energy imports from the grid [Soroudi et al., 2017].

To address DERs integration issues controlled PEV charging schemes have been proposed. The implementation of controlled PEV/PHEVs charging schemes, in particular centralised schemes, have been proposed in [Quiros-Tortos et al., 2016] where PEV charging operation is optimised to solve current/voltage network's contingencies. PEV Local control schemes allow a greater penetration of PEV than uncontrolled charging operations, meanwhile PEV centralised control achieve a better utilisation of the network capacity at the expense of implementing a significant network communications infrastructure [Richardson et al., 2013]. Regarding the impact of PHEV/PEVs in distribution networks, in [Guo et al., 2019] is studied the coordination between Autonomous PEVs charging/discharging with Network Reconfiguration (NR) to minimise power losses in Medium Voltage (MV) grids. However, It has to be highlighted that PEV optimal charging control could affect the battery lifetime of these devices since these are designed and

manufactured to be charged uninterruptedly [Hoke et al., 2011].

In contrast, dedicated BES systems can be used as flexibility energy providers. In fact, Multi-stage approaches based on scenarios' uncertainties provide accurate-enough representation of the uncertainty and the flexibility. However, they can become computationally infeasible easily, because the required number of variables increases exponentially with the number of stages considered. This inconvenient can be mitigated by selecting representative scenarios using clustering techniques [Li et al., 2016].

5.2.3   Optimization Techniques

Optimal operation of DERs is usually modelled as deterministic problems but, there is an uncertainty involved in the model due to the nature of DERs [Soroudi et al., 2016]. Thus, it is required that some optimization methodologies have to be developed to overcome the problem of DER integration into smart grids. In general, these approaches are based on Optimal Power Flows (OPF) in which a certain Objective Function (OF) is minimised subject to some technical constrains (power capacity, voltage and current limits, etc.). The OF considered in the literature can be Power Losses Minimisation (PLM), Supply Cost Minimisation (SCM), maximisation of renewable-based DG capacity in what is known as Renewable Generation Harvesting Maximisation (RGHM). The control variables are based on flexibility mechanisms. By solving the OPF, the optimal set points of the controlled variables are obtained. Usually, these OPFs approaches have been formulated in a multi-period fashion (typically for the day ahead basis, 24h) to cater with the daily variability of power customers demand and DERs generation.

To achieve the optimal flexibility required to reduce power losses of the system as well as solving network technical contingencies, it is necessary to model a complete three-phase optimal power flow of the LV smart grid network. This requires modelling the phase connection of each device in the network as well as determining the phase voltage and phase current magnitudes, which increases the computational complexity. The optimisation problem consists in a mathematical programming problem in which an objective function is minimised (or maximised) and it is formulated together with a set of equality constrains related to the full three-phase network equations, as well as inequality constraints related

to technical reference values.

Additionally, the optimal operation of distribution networks has been investigated considering DERs controllability, such as DG curtailment to deal with current and voltage constraints. Moreover, recently PLM have also been studied considering volt-var control [Ricciardi et al., 2019] and reconfiguration [Guo et al., 2019]. As demonstrated in [Ochoa and Harrison, 2011] PLM considering one single load condition hardly leads to an overall power losses minimisation. Moreover, the sole objective of PLM tends to reduce the potential DERs capacity that could be connected to distribution networks, hence a trade-off among different objective functions must be found. The concept of rolling optimisation was presented in [O'Connell et al., 2014] in which the OF in solved in time blocks, updating the demand and generation forecast for the upcoming blocks. Thus, more accurate results are obtained due considering time-varying characteristics.

OPF formulations in the literature range from linear models to non-linear models. The first have a straightforward implementation but at expense of a high computational cost as well as complex scalability [Franco et al., 2018]. For that reason, convex formulations [Borghetti, 2013] have been proposed as semi-definite programming [DallAnese et al., 2013] and quadratically constrained [Jabr et al., 2012]. These OPF-based programs could help the DSOs to carry out extended analysis for congestion management (over voltages and over-loadings), impact of Conservative Voltage Reduction (CVR), and the effects of regulatory policies.

One potential solution to counteract the over voltages due to PV injections is the adoption of On-Load Tap Changers (OLTP) transformers. However, the efficient and flexible nodal voltage control requires remote observability [Chen et al., 2005]. Also, to account for the DERs weather uncertainty, probabilistic Monte Carlo simulations have been applied [Li et al., 2019]. Nonetheless, these approaches are computational expensive and time consuming.

Given the presented literature review, it can be noticed that a large scientific production has been published related to DER optimization in LV distribution networks. However, very few considers the three-phase model neither accounts, explicitly, the intermittent PV uncertainty or the unexpected load demand variations. Existing solutions are

based on solving an equivalent single-phase LV power flow assuming constant both demand and generation. However, that procedure leads to non-realistic power losses values.

5.3 Formulation

The robust flexibility scheduling (FS) optimization model proposed for unbalanced LV smart grids correspond with a mathematical Mixed-Integer Non-Linear Programming (MINLP) problem stated in (5.1)-(5.20). It corresponds with a min-max optimization problem in which the outer-level problem it's a minimisation problem for the decision vector variables $\mathbf{x}$ and $\mathbf{y}$ and the inner-level problem is a maximisation problem for the uncertain vector variables $\mathbf{u}$ and $\mathbf{z}$. Uncertainty variables $\mathbf{u}$ and $\mathbf{z}$ are allowed to take values within a specified uncertainty sets. The value that the decision variables $\mathbf{x}$ and $\mathbf{y}$ can take depends on the maximisation of the objective function given $\mathbf{u}$ and $\mathbf{z}$ (worst-case). By this way the optimal value for the decision variables $(\mathbf{x}^*, \mathbf{y}^*)$ is obtained for the worst-case form the uncertainty variables $(\mathbf{u}^*, \mathbf{z}^*)$ point of view.

$$\min_{\mathbf{x},\mathbf{y}} \max_{\mathbf{u},\mathbf{z}} \quad \mathcal{F} : \left\{ \sum_{t \in T} \sum_{k \in V} x_{k,t}^p \cdot \xi_t + y_{k,t}^{p,+} \cdot \lambda_t + y_{k,t}^{p,-} \cdot \lambda_t \right\} \tag{5.1}$$

$$\text{s.t.} \quad \mathbf{x} = \left[x_{k,t}^p \right]^T \in (0,1), \forall k \in V^* \subseteq V, \forall t \in T, \forall p \in (a,b,c) \tag{5.2}$$

$$\mathbf{y}^+ = \left[y_{k,t}^{p,+} \right]^T \in (y_k^{min}, y_k^{max}), \forall k \in V^* \subseteq V, \forall t \in T, \forall p \in (a,b,c) \tag{5.3}$$

$$\mathbf{y}^- = \left[y_{k,t}^{p,-} \right]^T \in (y_k^{min}, y_k^{max}), \forall k \in V^* \subseteq V, \forall t \in T, \forall p \in (a,b,c) \tag{5.4}$$

$$\mathbf{y}^+ + \mathbf{y}^- = 0 \tag{5.5}$$

$$\mathbf{u} = [u_{k,t}^p]^T \in \Phi_{u,t} \tag{5.6}$$

$$\mathbf{z} = [z_{k,t}^p]^T \in \Phi_{z,t} \tag{5.7}$$

$$\vartheta = [v_{k,t}^{p,re} \quad v_{k,t}^{p,im}]^T \in \mathbb{R}, \forall k \in V, \forall t \in T, \forall p \in (a,b,c) \tag{5.8}$$

$$p_{i,k,t}^{p,sp}(\mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{z}) + p_{i,k,t}^{p,cal}(\vartheta) = 0 \tag{5.9}$$

$$p_{i,k,t}^{p,sp}(\mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{z}) = u_{k,t}^p - (p_{d,k,t}^{p,sp} + y_{k,t}^p + z_{k,t}^p) \tag{5.10}$$

$$i_{i,k,t}^{p,re} = \sum_{k' \in \Omega_k} \sum_{q \in (a,b,c)} g_{kk'}^{pq} \, v_{k',t}^{q,re} - b_{kk'}^{pq} \, v_{k',t}^{q,im} \tag{5.11}$$

$$i_{i,k}^{p,im} = \sum_{k' \in \Omega_k} \sum_{q \in (a,b,c)} b_{kk'}^{pq} \, u_{k'}^{q,im} + g_{kk'}^{pq} \, u_{k'}^{q,im} \tag{5.12}$$

$$x_{k,t}^p \cdot y_k^{min} \le y_{k,t}^p \le x_{k,t}^p \cdot y_k^{max} \tag{5.13}$$

$$\overbrace{\sum_{p \in (a,b,c)} \sum_{k,j \in V, \ i \neq j} \left[\left(i_{kj,t}^{p,re} \right)^2 + \left(i_{kj,t}^{p,im} \right)^2 \right] r_{k,j}^p \cdot e_{k,j}^p}^{\mathcal{L}_p} \le L_t^M \tag{5.14}$$

$$i_{kj,t}^{p,re} = g_{k,j}^p \left(v_k^{p,re} - v_j^{p,re} \right) + b_{k,j}^p \left(v_j^{p,im} - v_k^{p,im} \right) \tag{5.15}$$

$$i_{kj,t}^{p,im} = g_{k,j}^p \left(v_k^{p,im} - v_j^{p,im} \right) + b_{k,j}^p \left(v_k^{p,re} - v_j^{p,im} \right) \tag{5.16}$$

$$\Phi_{u,t}(\mu_{\mathbf{u},\mathbf{t}}, \sigma_{\mathbf{u},\mathbf{t}}) = \left\{ u_{k,t}^p \in \mathbb{R} : \sum_{\forall k \in N_u} \frac{|u_{k,t}^p - \mu_{u,t,k}|}{\sigma_{u,t,k}} \le \varphi_{u,t} \right\} \tag{5.17}$$

$$\Phi_{z,t}(\mu_{\mathbf{z},\mathbf{t}}, \sigma_{\mathbf{z},\mathbf{t}}) = \left\{ z_{k,t}^p \in \mathbb{R} : \sum_{\forall k \in N_z} \frac{|z_{k,t}^p - \mu_{z,t,k}|}{\sigma_{z,t,k}} \le \varphi_{z,t} \right\} \tag{5.18}$$

$$\left(i_{kj,t}^{p,re} \right)^2 + \left(i_{kj,t}^{p,im} \right)^2 \le \left(i_{kj}^{max} \right)^2 \tag{5.19}$$

$$\left(v^{min} \right)^2 \le \left(v_{k,t}^{p,re} \right)^2 + \left(v_{k,t}^{p,im} \right)^2 \le \left(v^{max} \right)^2 \tag{5.20}$$

The problem can be formulated in compact form:

$$\min_{\mathcal{X}} \max_{\mathcal{X}} \quad \mathcal{F}(\mathcal{X}) \tag{5.21}$$

$$\text{s.t.} \quad g_i(\mathcal{X}) \leq 0, \forall i \in (1, \dots, m) \tag{5.22}$$

$$h_j(\mathcal{X}) = 0, \forall j \in (1, \dots, s) \tag{5.23}$$

$$\mathcal{X} = \{\mathbf{x}, \mathbf{y}^+, \mathbf{y}^-, \mathbf{u}, \mathbf{z}, \vartheta\} \tag{5.24}$$

Which turns in to the following problem:

$$\min \quad \max \quad \mathcal{F}(\mathcal{X}) \tag{5.25}$$

$$\text{s.t.} \quad \mathcal{L}_p - L_t^M \leq 0 \tag{5.26}$$

$$\mathbf{y}^+ + \mathbf{y}^- = 0 \tag{5.27}$$

$$\left(v_{k,t}^{p,re}\right)^2 + \left(v_{k,t}^{p,im}\right)^2 - \left(v^{max}\right)^2 \leq 0 \tag{5.28}$$

$$\left(v^{min}\right)^2 - \left(v_{k,t}^{p,re}\right)^2 - \left(v_{k,t}^{p,im}\right)^2 \leq 0 \tag{5.29}$$

$$p_{i,k,t}^{p,sp}(\mathcal{X}) + p_{i,k,t}^{p,cal}(\mathcal{X}) = 0 \tag{5.30}$$

5.3.1   Objective Function

The objective function (5.1) is related with the total cost of the flexibility provided by a set of customer participating in a demand response program. Decision variables $\mathbf{x}$ and $\mathbf{y}$ are related with commitment decision (binary decision) of customers to provide flexibility and the dispatch of the amount of flexibility required from them, respectively. The first term of the objective function is related with the flexibility commitment cost ξ_t ($\$$) and the second term with the flexibility dispatch cost λ_t ($c\$/kWh$) is the marginal cost of the flexibility provided, i.e. is the cost associated to the flexibility provided over the planning horizon which depends on the flexibility providers committed and the prices of the flexibility. Decision variables of the problem are the following: $\mathbf{y} = [y_{t,p}^k]^T$ is a real-continuous vector decision variable related to the amount of amount of load flexibility dispatched in time instant $t \in T$ to each customer $k \in V^*$ connected to phase $p \in (a, b, c)$ participating in the demand response program for flexibility. Note that $\mathbf{x} = \left[x_{t,p}^k\right]^T$ is

a binary decision vector of flexibility decisions (i.e. which eligible customer k connected to phase p will participate in the flexibility program) for each time interval $t \in T$ of the flexibility scheduling period, usually 24 h. The objective function (5.1) involves two decision vector variables of the flexibility cost, which is composed by the cost associated with the commitment of flexibility and the cost associated with the dispatch of flexibility. Note that the dispatch decision variable is a real number which can take positive and negative values, but the flexibility cost is a positive magnitude, so the absolute value is used. This entails some computational issues since it is a non-linear function. Therefore, the dispatch decision variable is splitted in positive and negative $\mathbf{y} = \mathbf{y}^+$, $\mathbf{y}^-$ in such a way that if decision variable $\mathbf{y}$ is positive, then $\mathbf{y}^- = 0$ and $\mathbf{y}^+ > 0$, and if $\mathbf{y}$ is negative, then $\mathbf{y}^- > 0$ and $\mathbf{y}^+ = 0$. The objective function is subject to a set of equality and inequality constraints which correspond as follow: network non-linear unbalanced power flow equations (5.9)-(5.13), power net injection per phase in each node (5.15)-(5.16), maximum power losses limit (5.13)-(5.14), uncertainty constraints (5.17)-(5.18) technical limits for current magnitude en each phase feeder lines and phase voltage magnitude in (5.19)-(5.20) respectively.

5.3.2   Power Network Equations

Equality constraint (5.9) establishes that the specified (sp) power injections (which depends on the decision variables) in each node (k), phase (p) and time instant (t) has to be equal to the calculated (cal) power injections, which depends on vector state variable ϑ defined in expression (3). Note that solving this power mismatch is equivalent to finding the solution of $\mathcal{F}(\vartheta, \mathbf{x}, \mathbf{y}) = 0$ which entails a non-linear equation and provides the network state. By this, the network state is constituted by the power mismatch between the specified power injections (sp) at each node, phase and time-step, and the calculated power injections (cal). The state variable is defined by the nodal phase voltage magnitude (in each instant of time) separated in real and imaginary components. Calculated power injections are defined by the three-phase network equation (5.10), in each node (k), phase (p) and time instant (t). The calculated power injection is formed by taking the real part of the complex power injection ($\bar{S} = \bar{U}\bar{I}^*$) and decomposed in real and imaginary

parts. The result is a linear combination of the product between the real phase voltage magnitude and phase current on the one hand, and on the other hand, the imaginary phase voltage and phase current. Phase current magnitude can be expressed in terms of the phase voltage magnitude (state variable) and the network admittance according to the network equation $(\overline{I} = \overline{U} \cdot \overline{Y})$, derived from Kirchhoff's current law, as indicate expressions (5.9)-(5.10).

This gives as result the real and imaginary components of the phase current in function of the real and imaginary components of the phase voltage and the admittance matrix. Note that the conductance $g_{kk'}^{pq}$ and the susceptance $b_{kk'}^{pq}$ correspond with the mutual elements (of the diagonal) of the admittance matrix. In this regard, it is assumed that all the diagonal terms of the admittance matrix are equal as well as all the mutual impedances. This is a good approximation for the network modelling and makes more tractable the problem computation [Short, 2004]. Therefore, the power balance equation (5.9) and power injection (5.11)-(5.12) define the network state of the unbalanced smart grid.

In this regard, network nodes are modelled as PQ loads, where phase voltage magnitude is variable, and the net power injection is fixed by the problem as a result of the balance between specified and calculated power injections. In each node (k), phase (p) and instant of time the power injection is defined as the difference between the power generation (positive) and the power demand (negative) as shown in equation (5.10). Generation (PV injections) is considered as an uncertain variable and is defined as $\mathbf{u}$. This uncertain variable is specified at each node, phase and time instant.

The second component of the injection equation (5.10) is the total demand, composed as follow: first term corresponds with the non-interruptible customer demand $(p_{d,k,t}^{p,sp})$, which is not affected by flexibility and it is forecasted from historic real data availability; second component is the decision variable related to the flexibility dispatch $(y_{k,t}^{p})$ and it is the amount of flexibility that the customer will provide to the network, i.e. specify how much power demand is required to the customer to increase or decrease respect to the non-interruptible customer demand, and finally the last term is related with the uncertain variable of to the EV consumption $\mathbf{z}$, which is an additional power consumption of the customer affected by uncertainty due to the fact the EV charging sessions can materialise

at request of the customer in any instant of time.

5.3.3 Flexibility Equations

Inequality constraint (5.13) is related with customer flexibility. This equation involves both the dispatch decision variable $\mathbf{y}$ and the flexibility commitment decision variable (binary decision) $\mathbf{x}$ with the upper and lower bounds for each customer in terms of flexibility y_k^{max}, y_k^{min}. This equation works in such a way that if a customer connected to a specific node and phase is committed to provide flexibility in a certain time-step, the commitment variable will take unitary value and the dispatch decision is allowed to take values between a maximum y_k^{max} bound and a minimum y_k^{min} bound for the flexibility dispatch. Note that these bounds are constant in time and are fixed based on the capabilities of the customer to provide flexibility. If a customer does not have a high value of power contracted or does not have enough appliances to increase their power consumption, that needs to be reflected in the definition of the upper bound. On the contrary, if the customer have critical appliances, the upper bound will be a lower quantity. Note that the dispatch variable has been separated in positive dispatch variable and negative dispatch variable and so the equation (5.13) is formulated for both of them is such a way: $\mathbf{y}^+ \leq x_{k,t} \cdot y_k^{max}$ and $\mathbf{y}^- \geq x_{k,t} \cdot y_k^{min}$. The objective of the flexibility scheduling problem is to reduce power losses of the network and so it is explicitly defined in inequality constraint (5.14). In that equation is defined that the total power losses in each time period resulting from the flexibility scheduling implemented are constrained to be upper bounded by a maximum allowed L_t^M, which is a parameter fixed by the decision-maker.

Note that the average power losses levels in a unbalanced LV smart grids usually takes values between 2-4% of the power supply, therefore the maximum power loses L_t^M has to be fixed below those threshold to activate the scheduling of flexibility. Power losses of the network are calculated using the phase current flowing in each feeder section in real and imaginary components. Power losses correspond with the product of the square of the phase magnitude current that flows by the feeder section that connects nodes (r) and (j) in phase (p) (calculated with the real component $i_{k,j}^{p,re}$ defined in (5.15) and the imaginary component $i_{k,j}^{p,im}$ defined in (5.16) for the phase current), the linear resistance of the same

phase feeder section $r_{k,j}^p$ (Ω/m) and the length of the same phase feeder section, $e_{k,j}^p$ (m).

Real and imaginary components of the current that flows through the phase conductor (p) from node (k) to node (j) can be described in terms of the phase voltages and admittance line components (conductance and susceptance). These components are written in terms of the state variable phase voltages as follow: $g_{k,j}^p$ is the conductance of the feeder section between nodes (k) and (j) of the phase (p), $b_{k,j}^p$ is the susceptance of the feeder section between nodes (k) and (j) of the phase (p), $u_j^{p,re}$, $u_j^{p,im}$ are the real and imaginary components (respectively) of the phase voltage magnitude of node (j), phase (p) and $u_k^{p,re}$, $u_k^{p,im}$ are the real and imaginary components of the phase voltage magnitudes of node (k), phase (p).

Regarding the solution of the unbalance power flow, one of the network buses is chosen as the slack (and is denoted as k^*) and their specified (sp) power injection is fixed to be equal to the net demand plus the power losses of the network. The slack node is assigned to the LV side of the SS bus where voltage is fixed to the nominal value. In addition, the voltage angle of the slack bus is fixed to zero.

5.3.4   Uncertainty Modelling

To deal with the uncertainty sources namely intermittent PV injections, and sudden or unexpected EV load demand charging, in this section a Robust Optimization (RO) approach is adopted. The objective with this is to find a solution (flexibility schedule, both commitment and dispatch) that minimises the objective function for among all the feasible solutions and for all the realisations of the uncertain sources (PV and EV). That robust solution has a priori ensured feasibility when the uncertain sources vary within an uncertainty set, which may be large. The uncertainty set is a non-empty compact set, which means that the set is closed (containing all the limit points) and bounded (having all its points lie within some fixed distance among each other). The compactness assumption of the uncertainty set implies that uncertainty is bounded, which reflects the reality. PV injections and EV demand are considered to be affected by uncertainty, meanwhile flexibility is assigned to the customer demand, and results in decision variable to achieve the optimal operation, being the output of the model the optimal day-ahead scheduling

for that variables. Unlike Stochastic Optimization (SO) or Probabilistic Optimization (PO), RO does not rely on statistical information nor depends on expensive discretisation of uncertainty in scenarios. Instead, RO models uncertainty by decision variables which are allowed to take values within a pre-specified uncertainty set. In such a way, RO provides an optimal solution which is immune to all possible realisations of the uncertainty variables (sources) within the defined uncertainty set. In the case of this work, it will provide the flexibility schedule which is optimal (from the economic point of view of the flexibility) and feasible (from the technical point of view) for any realisation of the uncertainty sources (variable) under consideration. But more than that, the solution found is optimal and feasible for the worst-case realisation of the uncertainty variables (i.e. the combination of PV injections and EV demand that produce the maximum losses and technical contingencies).

The uncertainty set represents all possible materialisations of the uncertain sources. The worst-case values, are one of the extreme of the corresponding uncertainty set intervals. Uncertainty sets are key building blocks of the proposed robust model. The first step to build a robust model is to define the uncertainty sets for each uncertainty variable under consideration, and this corresponds with the expressions (5.17)-(5.18). In the flexibility scheduling problem the uncertain parameters are the PV injections and the EV power consumption during charging sessions. Therefore each of these uncertainty parameters has to be described in terms of a uncertainty set. Considering that, in this thesis the uncertainty set is modelled as a budget constrained polyhedral defined by equation (5.17) in the case of the PV uncertain variable and by the equation (5.18) in the case of the EV uncertain variable. The uncertainty set chosen it is based on the deviation of the uncertainty parameters from their expected value. In those expressions N_z, N_u are the set of nodes affected by the uncertain variables, $\mu_{\mathbf{z},t}$, $\mu_{\mathbf{u},t}$ are the vectors of mean expected values for the uncertain variables in each time period; and $\sigma_{\mathbf{u},t}$, $\sigma_{\mathbf{z},t}$ are the vectors of standard deviations for the uncertain variables in each time period. Note that each uncertain variable $z_{k,t}^p$ and $u_{k,t}^p$ is allowed to take values in the range $\{\mu_{(\bullet),k,t} - \sigma_{(\bullet),k,t} \, , \, \mu_{(\bullet),k,t} + \sigma_{(\bullet),k,t}\}$ and the total variability in each time period is constrained by a uncertainty budget $\varphi_{d,t}$ which is a parameter to control the level of robustness allowed to the model. It takes

values between 0 and $N_{(\bullet)}$ and its purpose is to control the total deviation of all uncertainty variables from their nominal values, weighted by the standard deviation. If the uncertainty budget is null ($\varphi_{(\bullet)} = 0$), then the vector of uncertainty variables takes only the expected values and the situation corresponds with the determinist case. Choosing a low uncertainty budget leads to more deterministic solutions since the uncertainty variables takes values closer to their mean values, meanwhile large uncertainty budgets lead to more conservative solutions and so more robust solutions against any feasible realisation of the uncertainty variables. As the uncertainty budget get larger, the uncertainty set gets wider. For this reason, robust sets are generally modelled as boxes or polyhedral constraints.

5.3.5   Statutory Limits

Finally equations (5.19) and (5.20) define the technical limits for the phase current magnitude in each feeder line section and the phase voltage magnitude (respectively) in each time instant. Note that both equations are expressed in terms of the real and imaginary components of the phase current and the phase voltage, and additionally the phase current can be expressed in terms of voltage magnitude by replacing with equations (5.11) and (5.12).

5.3.6   Solution Procedure

The robust optimization problem presented in this chapter comprise equations (5.1)-(5.20), and represents a max-min bilevel programming problem whose solution procedure rely in decomposition techniques as well as iterative procedures [Conejo and Wu, 2021]. To formulate and solve the robust optimization model the PYTHON based library RO-MODEL [Wiebe and Misener, 2021] has been selected in which a cutting planes solver is used to implement an iterative strategy for solving the robust model. This library replaces each uncertain constraint and objective by a cut which contains the nominal constraint. The solver iteratively solves the master problem (minimisation) and produces cut off solutions which are not robustly feasible. A solution $\mathcal{X}^*$ is considered to be ro-

bustly feasible when for each uncertain constraint $g_i(\mathcal{X}^*) \leq 0$ the value of the objective function is smaller than some tolerance OF^{min}. The solver used in the master problem is COUENNE [Belotti et al., 2009] which is a branch&bound algorithm able to find global optima of non-convex MINLPs. This solver implements linearisation, bound reduction, and branching methods within a branch-and-bound framework.

$$\mathcal{F}(\mathcal{X}) \leq OF^{min} \tag{5.31}$$

5.4 Case Study

5.4.1 Data

The test network chosen for implementing the robust flexibility scheduling is the modified European LV Test Feeder (ELVTF) [Report, 2015] illustrated in Fig. 5.4. This feeder was selected as an appropriate test for validating the proposed formulation since it represents an appropriate benchmark for researchers who want to study unbalanced LV smart grids, which are common in Europe Countries. The test feeder has a purely radial topology with a single connection point to the SS. The single feeder have a length of 1.4 km with 53 buses (connections points). The nominal line-to-line feeder voltage level is 416 V(line-to-line) with a base frequency of 50 Hz which is typical in the European LV distribution networks. The feeder is connected to the MV system through a distribution transformer located at the SS, which steps the voltage down from 11 kV to 416 V and has a power rating of 630 kVA.

Electrical magnitudes are indicated in per unit values using the transformer power rating as power base and the nominal line to line voltage as voltage reference. The original ELVTF network is modified to apply the robust flexibility scheduling model proposed. It is considered that residential and commercial customers are connected to each bus. Each customer have a contractual power that range from 3.45 kW to 9.20 kW (which are the normalised power contracted in Spain). The load consumption patterns of residential and commercial customers are extracted as baseline from the OSIRIS project [UC3M et al., 2017] which includes a large-scale area of LV smart grids. The load con-

Figure 5.4: European low-voltage test feeder topology representation

sumption data includes customer connection phase as indicated in Table A.2 (in Appendix) where it is indicated the node number in which the customer is connected, the phase to which is connected (A, B, or C), the average load power consumption (active power) and the daily expected value for PV generation and the EV consumption pattern. PV facilities have a peak power that range between 1 kW to 4 kW (which results in a realistic rooftop array between 4 and 16 PV panels respectively, each one of 250 W of peak power). Meanwhile EV charging points are modelled to have nominal values between 2.3 and 3.3 kW. This last series is used to build the uncertainty boxes for the model. In Fig. 5.5 it is shown the hourly mean expected value for the load demand of the network per phase, based on the historical data in [Velasco et al., 2020] which includes four years of smart meters data. It can be seen that the aggregated load consumption in each phase does not follow the same pattern since the load is not balanced between phases. This case is specially more clear in the phase A (blue pattern).

In Table A.1 (in Appendix) it is indicated the main metrics for analysing the unbalance situation in the network under study. In Table A.1 it can be seen that phases B and C are more loaded than phase A. This results in unbalanced operation with leads to over-voltage's and overloading's. The unbalance degree CIGRE-2 is calculated by the

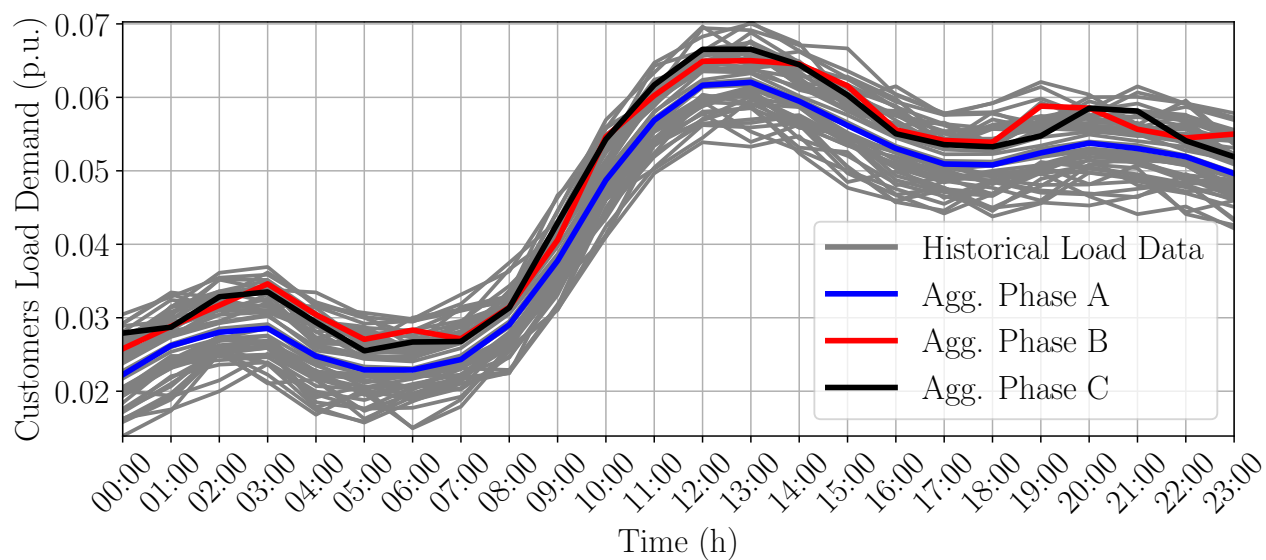Figure 5.5: Aggregated customers load demand expected values

Table 5.1: Principal unbalanced network metrics without flexibility

Metric	Value
Load Unbalance ratio phase A	29 %
Load Unbalance ratio phase B	35 %
Load Unbalance ratio phase C	36 %
Unbalance Degree CIGRE-2 DQV	3.45 %
Max. phase voltage mag. (p.u.)	1.14
Max. phase current mag. (p.u.)	0.16

dispersion of the average phase voltage values respect to the nominal values, as indicate
(5.32)-(5.33), where ($\hat{u}_p$) is the expected phase voltage value in p.u. This unbalance de-
gree results in 3.45 % which clearly indicates a deviation for the desired balance conditions
(the maximum deviation allowed is 3%).

$$DQV = \sqrt{(2/3) \cdot \sum_{p \in (a,b,c)} \delta_p^2} \tag{5.32}$$

$$\delta_p = |\hat{u}_p - 1| \tag{5.33}$$

The historical data of the smart meters is used to forecast the load demand customers
of the day-ahead using a time rolling forecasting technique as is done in [Hyndman et al., 2018].
For simplicity, but without lack of accuracy, unitary power factor is considered in such
a way that the reactive power demand for each customer is zero, and PV injections are
considered purely active power. Regarding PV generation, in Fig. 5.6 is shown the un-
certainty model for the PV injection of the node 27, phase C. It can be seen that the
expected value for the PV injections is plotted in solid black line and the uncertainty box
is delimited by the upper and lower limits in dashed and dotted lines respectively. The
grey area delimits any possible day-ahead injection power.

Figure 5.6: PV Generation expected value and uncertainty area covered by the uncertainty
box

PEV units are considered to have a nominal power charging that range between 2.3

and 3.7 kW (which provides charging times between 7-3 hours respectively Fig. 5.7). This corresponds with the average charging period for the current EV models in the European market [Fotouhi et al., 2019].

Figure 5.7: PEV charging pattern duration and battery capacity scatterplot

Regarding the modelling of the PEV devices, in Fig. 5.8 it is shown the expected demand value (in solid black line) for the PEV unit connected to bus 30 phase C. The uncertain box of the given PEV unit is delimited by the upper and lower limits in dashed and dotted lines (respectively). Any of the points within the grey represents the uncertainty box associated to the worst-case robust flexibility scheduling. As it can be seen, users usually connect their vehicles when they arrive at home, consequently, the battery is charged at night and they unplug the vehicle when they leave home in the morning.

The selection of the uncertainty budget is aligned with the spread of the data set of each uncertainty unit. Related to the flexibility cost, Table 5.2 shows the prices for the flexibility provided by customers, in terms of economic value per unit of energy. It can be seen that the policy is that flexibility provided in central hours of the day is more expensive that in early or lately hours of the day, due to the fact that in the middle of the day, a higher amount of PV injections will take place and so the probability of technical contingencies or high power losses increases.

Figure 5.8: PEV charging pattern expected value and uncertainty box

Table 5.2: Hourly prices for the flexibility scheduling

Time interval	λ_t (c\$/kWh)	Time interval	λ_t (c\$/kWh)
00:00 - 01:00	65	13:00 - 14:00	79
01:00 - 02:00	66	14:00 - 15:00	78
02:00 - 03:00	66	15:00 - 16:00	78
03:00 - 04:00	67	16:00 - 17:00	77
04:00 - 05:00	67	17:00 - 18:00	77
05:00 - 06:00	68	18:00 - 19:00	76
06:00 - 07:00	68	19:00 - 20:00	76
07:00 - 08:00	70	20:00 - 21:00	74
08:00 - 09:00	74	21:00 - 22:00	73
09:00 - 10:00	76	22:00 - 23:00	67
10:00 - 11:00	77	23:00 - 00:00	66
11:00 - 12:00	79		
12:00 - 13:00	79		

5.4.2 Initial Situation

To have an idea of the initial situation of the network before the application of the robust flexibility scheduling, an unbalanced power flow is carried out. The unbalanced power flow entails to solve equations (5.9) and (5.10), considering the uncertain variables **u** and **z** as input parameters. In particular, it is used the expected values for PV and PEV shown in Fig. 5.6 and Fig. 5.8. By this way the total power losses, calculated by left side of inequality (5.14), as well as phase voltage magnitude and phase current magnitudes are calculated under the expected value conditions for demand (Fig. 5.5), PV injections (Fig. 5.6) and PEV demand (Fig. 5.8). As a result, the unbalanced power flow gives the following power losses pattern illustrated in Fig. 5.9. It can be noted that total power losses are superior than 3% in the absence of any flexibility mechanism. Additionally, it can be seen that peak power losses occur at midday according to the periods where PV units injects more power to the grid.

Figure 5.9: Power losses of the network per phase without flexibility

In Fig. 5.10 it is shown the maximum phase current magnitude of the network (node 34) as a result of the unbalance power flow calculation (in absence of any flexibility mechanism). It can be seen that the network operates in unbalanced conditions and surpass the thermal limit of the feeder section at midday (when the PV injects the power).

It can be seen that overloading's and over-voltage are correlated with the power losses

problem, and so alleviating technical contingencies is related to the problem of power losses minimisation. Moreover, note that the resolution of the problem affects to the precision of the solution achieved. In this sense, if a high resolution is used, the power losses as well as the phase voltage and current unbalanced calculated will be more accurate and realistic. One important fact to understand is that, after the robust flexibility scheduling is applied the unbalanced situation will be minimised by mean of the flexibility calculated.

Figure 5.10: Max. phase current magnitude without flexibility schedule

5.4.3   Methods Comparison

For comparison purposes a stochastic version of the flexibility scheduling model is applied and a deterministic version is also presented in which there is no uncertainty. The stochastic version of the proposed model is composed by the minimisation of the expected value using the same objective function (5.34) subject to the same constraints (5.2)-(5.6), (5.8)-(5.16), (5.19)-(5.20) so excluding equations (5.17) and (5.18) where the uncertainty boxes are defined, and (5.6)-(5.7) where the uncertain variables $\mathbf{u}$ and $\mathbf{z}$ are defined.

$$\min_{\mathbf{x},\mathbf{y}}\quad \sum_{\omega \in \Omega_s} \pi_\omega \left\{ \sum_{t \in T} \sum_{k \in V} x^p_{k,t} \cdot \xi_t + y^{p,+}_{k,t} \cdot \lambda_t + y^{p,-}_{k,t} \cdot \lambda_t \right\} \qquad (5.34)$$

Where Ω_s is the set of scenarios considered and π_ω is the probability of scenario $\omega \in \Omega_s$.

Figure 5.11: Max. phase voltage magnitude without flexibility schedule

Instead of limiting the uncertainty variables **u**, **z** to take values within uncertainty boxes, the variables are considered as parameters, in such a way that they take values given a certain scenario. The objective function (5.34) minimises the expected value of the costs over the scenarios considered. Fifteen scenarios are considered for the day-ahead, one of them corresponds to the expected value of a SARIMA forecast and the remaining scenarios correspond to variations of the forecasted values considering different forecasting errors. The deterministic model corresponds with the expected values for load demand of the customers, PV injections and EV charging demand. For the operation of the proposed flexibility scheduling model, the load demand can be forecasted using a SARIMA model [Hyndman et al., 2018] and the smart meter measurements are collected by the smart metering infrastructure.

SARIMA is a widely used forecasting technique that combines Seasonal Auto-Regressive models (AR) with Moving-Average (MA) applied for non-stationary time series by using integrated moving average. Since PEV load demand and PV generation (solar irradiation and temperature) are not-stationary processes, this model is chosen to perform the forecasting [Montgomery et al., 2008]. SARIMA is an extension to ARIMA that supports the direct modelling of the seasonal component of the time-series. SARIMA$(p, d, q)(P, D, Q, s)$ process at time $t + \tau$ (τ period in the future) can be expressed as is indicated in (5.35)

$$\hat{y}_{t+\tau} = \delta + \sum_{i=1}^{p+d} \phi_i y_{t+\tau-i} + \varepsilon_{t+\tau} - \sum_{j=1}^{q} \theta_j \varepsilon_{t+\tau-j} \tag{5.35}$$

Where:

- p and seasonal P: indicate the number of autoregressive terms (i.e lags of the stationarized series).

- d and seasonal D: indicate differencing degrees required (i.e stationarize series).

- q and seasonal Q: indicate the number of moving average terms (i.e lags of the forecast errors).

- s: indicates seasonal length in the data.

In the above expression, δ its a constant term that depends on the mean value of the time series $\delta = \mu_y (1 - \phi_0)$, ε_t is a white noise process and $y_{t+\tau}$ is the time series that represents the variables to be forecasted. The value of the time series forecasted in time τ is $\hat{y}_{t+\tau}$ (i.e. the time series variables to be forecasted such as ambient temperature $\widetilde{T}_{amb}$ or solar irradiance $\widetilde{G}$ for the stochastic programming approach). The degree of the model (p, d, q) is defined by p, which represents the order of the auto-regressive part, d which denotes the degree of the first differencing involved and q which denotes the order of the moving average. The degree of the model as well as the coefficients have to be fitted using the python-based library STATS MODELS [Seabold and Perktold, 2010]. The later, is an open-.source tool widely used in statistical studies. To select the appropriate combination of the model parameters, a grid search evaluation of SARIMA models have been carried out using the *AIC* statistic (*Akaike Information Criteria*) as a performance metric indicated in the expression (5.36).

$$AIC = 2K - 2ln(L) \tag{5.36}$$

Where K is the the number of parameters of the model and L is the value of the maximum likelihood function of the SARIMA model. The *AIC* statistic quantifies the goodness of fit as well as the simplicity of the model. Between two models, the one with the lowest AIC value is the better to be fitted [Burnham and Anderson, 2002].

Weather data has been collected from several weather stations in the Madrid area provided by AEMET (Agecia Estatal de Meteorologia) [Agencia Estatal de Meteorologia, 2015].

By this way the DSO will have a day-ahead scheduling flexibility tool in which the only model parameter will be the uncertainty budget for the uncertainty variables (PV and PEV), being more expensive the flexibility as much as uncertainty budget is configured.

5.4.4 Results

The proposed robust flexibility scheduling formulation has been applied in the case study and results are presented in this section. The formulation has been implemented using Python as the programming language and the optimisation model is formulated as a MINLP using the python libraries PYOMO [Hart et al., 2011] as the optimization framework to define the min-max problem as well as the ROMODEL [Wiebe and Misener, 2021] and COUENNE [Belotti et al., 2009] that have been used as solvers. The optimization problem is executed in a dual-core 2.5 GHz Intel Core i5 with 16 GB RAM memory. In Fig. 5.12 it is shown the day-ahead aggregated flexibility scheduling (in p.u.) using different uncertainty budgets for the uncertain variables. The first observation that has to be made is that the dashed red line indicates the flexibility scheduling considering null uncertainty budget, which corresponds with a deterministic case (i.e. using the expected value of the PV injections and the PEV charging showed in Fig. 5.6 and Fig. 5.8 , respectively). As it can be seen, this case corresponds with the lower total flexibility cost (denoted with Z). Then, if the uncertainty budget increases, the required amount of flexibility will increase, since there is more uncertainty to tackle in the system, i.e. the flexibility scheduler is more conservative and request more load flexibility just in case the realisation of the uncertain sources lead to a worse situation. This same behaviour continues until the maximum uncertainty budget is achieved, and so the one with the maximum flexibility cost. The solid blue line indicates the chosen flexibility scheduling since it results in a compromise between flexibility cost and uncertainty coverage.

Note that the flexibility scheduling follows a pattern similar to the aggregated PV injections. This is motivated by the fact that a high amount of energy is being injected in the network, and it has to be absorbed. In Table 5.3 is shown the main comparative results
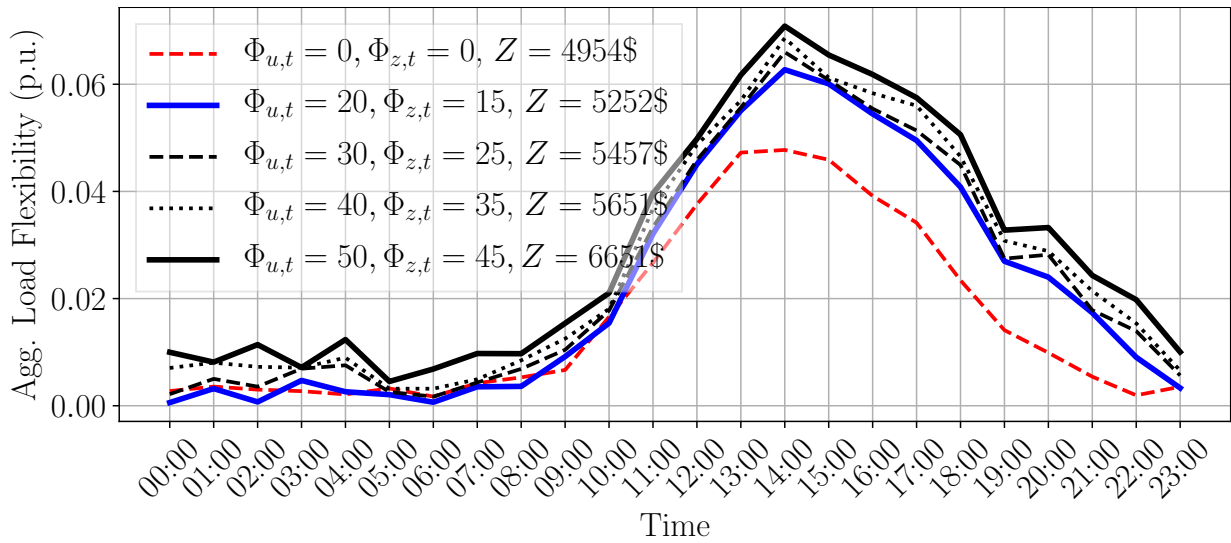
Figure 5.12: Aggregated flexibility scheduling for the day ahead with different uncertainty budgets

among the implemented robust formulation, the deterministic version (without robust treatment of uncertainty sources, i.e. null uncertainty budget) and the stochastic optimal power flow with 5 scenarios . It can be noted that total power losses by the proposed day-ahead robust formulation results lower values compared to both deterministic and stochastic models.
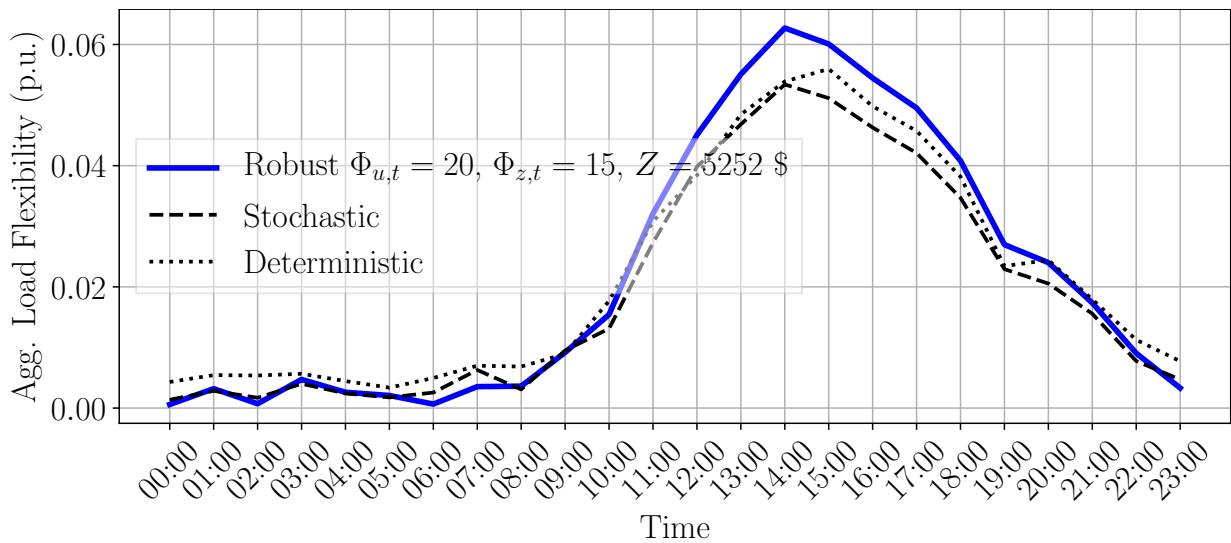
Figure 5.13: Flexibility scheduling comparison

Consequently, customers load demand (including flexibility and PEV realisation) is

higher in the robust model, since the flexibility provided is positive in all the time horizon (and all the customers) and so it means an increase in the power consumption. This increase in the power consumption can be realised in the use of some household appliances that works as an energy storage device such as hot water devices, or by load shifting to the hours in which the load flexibility required. In terms of computation time, the robust proposed model requires a higher amount of time in comparison with the (simple) stochastic model and the deterministic (with no uncertainty treatment). In Fig. 5.14 and Fig. 5.15 are shown day-ahead maximum phase voltage and current magnitudes when the selected robust flexibility schedule is applied.
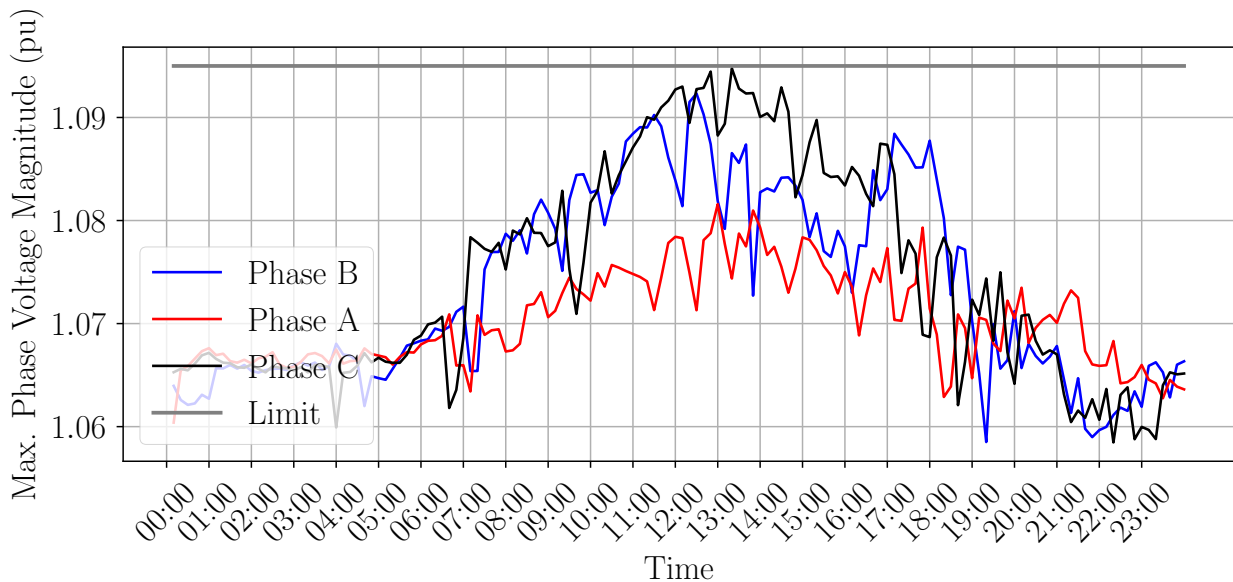
Figure 5.14: Max. phase voltage magnitude with flexibility schedule

Regarding the statutory limits, In Table 5.3 it is indicated the principal unbalance network metrics after the flexibility schedule is applied. Both robust and stochastic models provide flexibility scheduling that keep voltage below the maximum limit. As well the maximum phase current magnitudes obtained are below the maximum limited allowed. As it can be seen, the unbalance situation has improved respect to the situation in which no flexibility is applied. However, the reduction in overloading's and over-voltage's achieved by the robust model is higher than the stochastic and the deterministic versions.
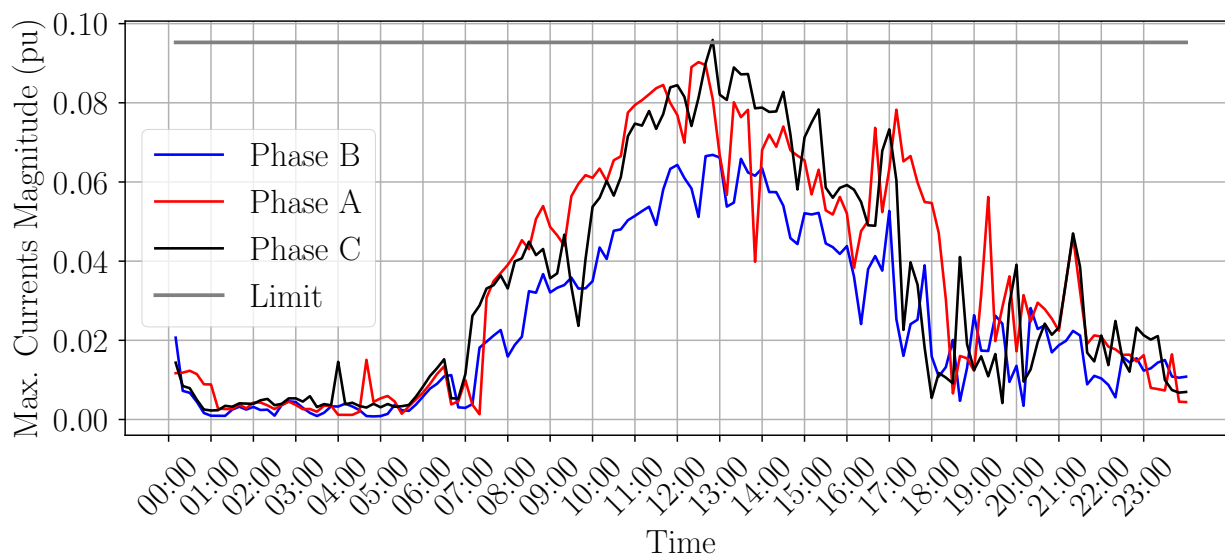
Figure 5.15: Maximum phase current with flexibility schedule

Table 5.3: Results comparison between robust flexibility scheduling and deterministic power flow

	Robust FS	Stochastic FS	Deterministic FS
Total Power Losses (kWh)	197.52	229.40	238.67
Total Energy Imp. (kWh)	264.87	332.32	316.45
Total Load Demand (kWh)	1443.78	1314.21	1352.45
Total PV Production (kWh)	1617.20	1439.75	1464.21
Total Flexibility (kWh)	240.77	228.45	189.45
Computation time (min)	674	267	6
Max. phase voltage mag. (p.u.)	1.06	1.09	1.14
Max. phase current mag. (p.u.)	0.07	0.09	0.16

FS: Flexibility Scheduling

5.5 Discussion & Conclusions

In this chapter, it is proposed a robust flexibility scheduling for LV unbalanced smart grids considering uncertainty in PV injections as well in PEV consumption and considering demand response as flexibility mechanism. The objective is to minimise the economic cost associated to the required flexibility to reduce total power losses below a certain maximum limit and at the same time to keep the system within safe operation conditions in terms of voltage and current limits. The model considers fully unbalance operation of the network and model uncertainty of PV and PEV devices through a polyhedral uncertainty sets such as the cardinality-constrained uncertainty set which provides the worst case uncertainty realisation in one of the polyhedron vertex that represent the set. Although a simple demand response mechanism has been presented in this chapter, it should be noted that the main solution procedure can incorporate other aspects related to the generation, consumption and storage of electricity in terms of flexibility such as considering the presence of BES systems which can provide more flexibility capacity and efficiency. Future research will address the incorporation of those devices in the network using ARO frameworks or Stochastic-Robust optimization frameworks.

These can help DSOs for decision-making operations about the optimal number of BES units or their optimal locations. Moreover, the proposed methodology can be used for dispatch decisions about the required flexibility to be provided in the network to keep system within normal operation conditions in terms of voltage, current and power losses and improving, at the same time, the PV hosting capacity under uncertain situations.

Additionally, further refinements of the model could be the modelling of correlation between robust sets across time and space. For instance power demand in a specific location and time can vary in a specific interval, and this variation is correlated with the variation of the energy prices in the same place and time.

Relevant conclusions from this work are the following:

1. Modelling uncertainty in PV generation has been proved to result in a key component in the success of flexibility scheduling since it strongly depends on the intermittent behaviour of the weather conditions such as solar radiation and ambient

temperature, as well as other aspects such as cloudy conditions, orientation or main-tenance. The uncertainty related to PEV consumption is also fundamental but due to the valley charging period, it has no relevant impact in network performance.

2. A data-driven uncertainty sets proposed in this work enable a customised control of robustness in combination with the choice of the uncertainty budget. Since the uncertainty sets depends on uncertainty budget, which defines the limits, a rigorous exploratory data analysis of the historic time series can improve the quality of such uncertainty sets since it will be more realistic and hence the decision-making process will be data-driven.

3. In terms of feasibility, it is worthy to mention that the proposed framework provides feasible and implementable solutions in unbalanced smart grids which by definition are ill-conditioned problems.

4. Demand response results in a well convenient mechanism to leverage flexibility in order to increase the hosting capacity of smart grids and keep the system as efficient as possible.

Chapter 6

Conclusions

6.1   Main Conclusions

This thesis contributes to the field of power loss estimation and optimization in low-voltage (LV) smart grids in large-scale distribution areas. To contextualize the importance of the research, it has been necessary to explain the unbalanced nature of low voltage distribution areas where there is a huge deployment of smart meter rollout, and there is also uncertainty related to renewable energy generation. Main results of the thesis have been applied in two smart grid research projects: the national project OSIRIS (*Optimización de la Supervisión Inteligente de la Red de Distribución*) and the European project IDE4L (*Ideal Grid For All*).

The topic of the power losses optimization through the deployment of flexibility scheduling was examined. Particular attention was paid to the uncertainty that involves the intermittent behavior of DERs sources. An uncertainty characterization was performed to model both the demand and generation sources of uncertainty.

A comprehensive and reliable power loss estimation method for unbalanced large-scale LV smart grids under uncertainty was proposed. The method comprises a preliminary data preprocessing step based on PCA, a K-means++ based clustering process for the LV feeders and a deep neural network regressor to infer the values of power loss from a set of representative feeders. The proposed method represents a timely tool for DSOs to infer the power loss levels in large-scale distribution areas using big data and deep learning

technologies.

A robust formulation for the flexibility scheduling in unbalanced LV smart grids was also proposed. The aim is to reduce power losses allowing, at the same time, the maximum integration of renewable-based distributed resources and keeping the system operation within normal conditions. This formulation model explicitly examines the uncertainty of demand and DERs sources.

6.2   Contributions

The main contributions of this thesis are the following:

- Objective 1: Power losses estimation in unbalanced LV smart grids under uncertainty conditions.

  ✓ An optimization-based procedure to estimate load consumption of non-telemetered customers.

  ✓ A Markov chain-based process to estimate intra-hour load demand for data having a low resolution and for non-telemetered customers or customers which smart meters provide incorrect measurements.

- Objective 2: Power losses estimation in unbalanced LV smart grids in large-scale areas with a presence of DERs.

  ✓ A data mining approach to reduce a high-dimensionality dataset in smart grids to yield a reduced set of relevant features.

  ✓ A clustering process to obtain representative feeders within a large-scale distribution area of smart grids.

  ✓ A deep learning-based power losses estimator for large-scale LV smart grids. The method is formulated as a deep neural network that uses as input features the power load demand and power generation of a set of representative feeders. The model gives, as output, the power losses of the whole area.

- Objective 3: Flexibility scheduling for power losses minimization in unbalanced smart grids under uncertainty.

  - ✓ A robust optimization model for the flexibility scheduling optimization model for unbalanced smart grids with distributed resources, such as PV panels and PEV devices.

6.3 Future Work

Regarding recommendations for future research in the area of unbalanced LV smart grids under uncertainty and with the presence of DERs, the following topics seem a natural extension of the current study:

- Non-technical losses modeling that is based on deep-learning anomaly detection techniques.

- Modeling the massive integration of BES systems as flexibility providers to take full advantage of the flexibility offered by LV customers.

- Modeling the flexibility in an MV network and the distributed resources connected to that network.

- Modeling that incorporates air conditioners, heaters and hot water devices into the flexibility equations. Here, the constraints could include comfort temperatures and the dwelling's consumption.

Appendix A

Network Data

Table A.1: Customers power contracted and DERs ratings for the modified IEEE European LV test feeder [Report, 2015]

Bus	Phase	$\overline{p}^p_{d,k}$	$\widetilde{\mu}_{\mathbf{u},k}$	$\widetilde{\mu}_{\mathbf{z},k}$	Bus	Phase	$\overline{p}^p_{d,k}$	$\widetilde{\mu}_{\mathbf{u},k}$	$\widetilde{\mu}_{\mathbf{z},k}$
1	A	5.75	1.5	3.3	29	B	8.05	1.5	3.3
2	B	4.60	2.0	2.3	30	C	6.90	2.0	2.3
3	C	5.75	2.5	3.3	31	A	5.75	1.0	3.3
4	A	5.75	1.5	3.3	32	B	5.75	2.0	2.3
5	B	8.09	1.0	3.3	33	C	4.60	2.5	3.3
6	C	5.75	2.5	3.3	34	A	4.60	3.0	3.3
7	A	5.75	3.5	2.3	35	B	8.05	2.5	2.3
8	B	5.75	2.5	3.3	36	C	9.20	1.5	2.3
9	C	5.75	1.0	3.3	37	A	5.75	1.5	3.3
10	A	5.75	1.0	2.3	38	B	8.05	2.5	3.3
11	B	4.60	1.0	6.3	39	C	9.20	2.0	2.3
12	C	5.75	2.5	3.3	40	A	4.60	3.5	3.3
13	A	4.60	2.0	3.3	41	B	4.60	3.5	3.3
14	B	5.75	1.5	3.3	42	C	4.60	4.0	3.3
15	C	6.90	2.5	3.3	43	A	8.05	1.5	3.3
16	A	8.05	1.5	3.3	44	B	5.75	2.0	3.3
17	B	9.20	2.5	3.3	45	C	5.75	1.0	3.3
18	C	5.75	1.0	3.3	46	A	4.60	2.5	3.3
19	A	5.75	1.5	3.3	47	B	9.20	2.5	2.3
20	B	6.90	3.5	3.3	48	C	4.60	1.0	3.3
21	C	8.05	3.5	3.3	49	A	5.75	2.5	3.3
22	A	5.75	4.0	3.3	50	B	4.60	1.0	3.3
23	B	5.75	2.5	3.3	51	C	5.75	3.5	3.3
24	C	6.90	2.0	3.3	52	A	4.60	3.5	3.3
25	A	8.05	3.5	3.3	53	B	8.09	4.0	3.3
26	B	9.20	2.0	3.3					
27	C	4.60	2.0	3.3					
28	A	4.60	2.0	3.3					

Table A.2: OSIRIS Case Study Network Data [UC3M et al., 2017]

Feeder/Phase	Sending Bus (i)	Receiving Bus (j)	Resistance (Ω)	Reactance (Ω)
1	1	2	0,0048	0,0028
1	2	3	0,0033	0,0004
1	2	4	0,0005	0,0003
1	4	5	0,0021	0,0003
1	4	6	0,0015	0,0009
1	6	7	0,0021	0,0003
1	6	8	0,0008	0,0004
1	8	9	0,0021	0,0003
1	8	10	0,0015	0,0009
1	10	11	0,0021	0,0003
1	10	12	0,0007	0,0004
1	12	13	0,0021	0,0003
1	12	14	0,0017	0,0010
1	14	15	0,0021	0,0003
1	14	16	0,0007	0,0004
1	16	17	0,0021	0,0003
1	16	18	0,0016	0,0009
1	18	19	0,0021	0,0003
1	19	20	0,0004	0,0002
1	18	21	0,0006	0,0003
1	21	22	0,0021	0,0003
1	22	23	0,0025	0,0003
1	21	24	0,0017	0,0010
1	24	25	0,0021	0,0003
1	24	26	0,0007	0,0004
1	26	27	0,0021	0,0003
1	26	28	0,0016	0,0009
1	28	29	0,0021	0,0003
1	29	30	0,0028	0,0004
1	29	31	0,0028	0,0004
1	28	32	0,0006	0,0004
1	32	33	0,0021	0,0003
1	33	34	0,0028	0,0004
1	32	35	0,0016	0,0009
1	35	36	0,0021	0,0003
1	35	37	0,0004	0,0003
1	37	38	0,0006	0,0003
1	38	39	0,0041	0,0005
1	38	40	0,0006	0,0004
1	40	41	0,0014	0,0008
1	41	42	0,0004	0,0002
1	42	43	0,0059	0,0008
1	42	44	0,0010	0,0006
1	44	45	0,0076	0,0043
1	45	46	0,00212	0,0003
1	46	47	0,0035	0,0020
2	1	48	0,0086	0,0049

2	48	53	0,0037	0,0021
2	48	49	0,0021	0,0003
2	49	50	0,0052	0,0007
2	49	51	0,0056	0,0007
2	49	52	0,0054	0,0007
2	53	54	0,0021	0,0003
2	53	55	0,0037	0,0021
2	55	56	0,0065	0,0037
2	56	57	0,0021	0,0003
2	56	58	0,0073	0,0042
3	1	59	0,0035	0,0020
3	59	60	0,0028	0,0004
3	59	61	0,0057	0,0032
3	61	62	0,0012	0,0007
3	62	63	0,0022	0,0003
3	62	64	0,0074	0,0042
3	64	65	0,0050	0,0007
4	1	66	0,0187	0,0107
4	66	67	0,0035	0,0005
4	66	68	0,0018	0,0010
4	68	69	0,0007	0,0004
4	68	70	0,0005	0,0003
5	1	71	0,0133	0,0076
5	71	72	0,0010	0,0005
6	1	73	0,0068	0,0039
6	73	74	0,0048	0,0006
6	73	75	0,0052	0,0007
6	73	76	0,0065	0,0009
7	1	77	0,0014	0,0008
7	77	78	0,0034	0,0005

References

[Abadi et al., 2015] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2015). TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Software available from tensorflow.org.

[Abdi et al., 2017] Abdi, H., Beigvand, S. D., and Scala, M. L. (2017). A review of optimal power flow studies applied to smart grids and microgrids. *Renewable and Sustainable Energy Reviews*, 71:742–766.

[Agencia Estatal de Meteorologia, 2015] Agencia Estatal de Meteorologia, A. (2015). Agencia Estatal de Meteorologia (AEMET). accessed on 14 August 2016.

[Altman, 1992] Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185.

[Andrew Watters and Boslaugh, 2018] Andrew Watters, P. and Boslaugh, S. (2018). *Statistics in a Nutshell*. O'Reilly Media.

[Anna Veronika Dorogush, 2018] Anna Veronika Dorogush, Vasily Ershov, A. G. (2018). CatBoost: gradient boosting with categorical features support.

[Baldick and Wu, 1991] Baldick, R. and Wu, F. F. (1991). Approximation Formulas for the Distribution System: The Loss Function and Voltage Dependence. *IEEE Transactions on Power Delivery*, 6(1):252–259.

[Baran and Fernandes, 2016] Baran, A. R. and Fernandes, T. S. (2016). A three-phase optimal power flow applied to the planning of unbalanced distribution networks. *Electrical Power and Energy Systems*, 74:301–309.

[Belotti et al., 2009] Belotti, P., Lee, J., Liberti, L., Margot, F., and Wachter, A. (2009). Branching and bounds tightening techniques for non-convex MINLP. *Optimization Methods and Software*, 24(4-5):597–634.

[Ben-Tal et al., 2009] Ben-Tal, A., Ghaoui, L. E., and Nemirovski, A. (2009). *Robust Optimization*. Princeton: Princeton Univ. Press.

[Bengio et al., 2013] Bengio, Y., Boulanger-Lewandowski, N., and Pascanu, R. (2013). Advances in optimizing recurrent networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 8624–8628.

[Bergstra and Bengio, 2012] Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13:285–305.

[Bertsimas and Thiele, 2006] Bertsimas, D. and Thiele, A. (2006). Robust and data-driven optimization: Modern decision-making under uncertainty. *INFORMS Tutorials in Operations Research: Models Methods and Applications for Innovative Decision Making*.

[Bishop, 2006] Bishop, C. M. (2006). *Pattern Recognition and machine learning*. Springer.

[Bletterie et al., 2018] Bletterie, B., Kadam, S., and Renner, H. (2018). On the Classification of low voltage feeders for network planning and hosting capacity studies. *Energies*, 11(651).

[Borghetti, 2013] Borghetti, A. (2013). Using mixed integer programming for the volt/var optimization in distribution feeders. *Elect. Power Syst. Res.*, 27(89).

[Bracale et al., 2012] Bracale, A., Caldon, R., Celli, G., Coppo, M., Dal Canto, D., Langella, R., Petretto, G., Pilo, F., Pisano, G., Proto, D., Scalari, S., and Turri, R. (2012). Analysis of the Italian distribution system evolution through reference networks. In *IEEE PES Innovative Smart Grid Technologies Conference Europe*.

[Breiman, 2001] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

[Brownlee, 2016] Brownlee, J. (2016). Supervised and unsupervised machine learning algorithms. *Mathematical Programming*, 16(3).

[Buller and Woodrow, 1928] Buller, F. H. and Woodrow, C. A. (1928). Load factor equivalent hours values compared. *Electr. World*.

[Burnham and Anderson, 2002] Burnham, K. and Anderson, D. (2002). *Model Selection and Multimodel Inference A Practical Information-Theoretic Approach*. Springer-Verlag.

[Carreiro et al., 2017] Carreiro, A. M., Jorge, H. M., and Antunes, C. H. (2017). Energy management systems aggregators: A literature survey. *Renewable and Sustainable Energy Reviews*, 73:1160–1172.

[Carvalho et al., 2018] Carvalho, P. M. S., Ferreira, L. A. F. M., Santana, J. J. E., Dias, A. M. F., and Machado, J. A. C. (2018). Combined effects of load variability and phase imbalance onto simulated lv losses. *IEEE Transactions on Power Systems*, 33(6):7031–7041.

[Cespedes et al., 1983] Cespedes, R., Duran, H., Hernandez, H., and Rodriguez, A. (1983). Assessment of Electrical Energy Losses in the Colombian Power System. *IEEE Transactions on Power Apparatus and Systems*, PAS-102(11):3509–3515.

[Chang, 1968] Chang, N. E. (1968). Determination of Primary-Feeder Losses. *IEEE Transactions on Power Apparatus and Systems*, 87(12):1991–1994.

[Chang et al., 2010] Chang, Y.-W., Hsieh, C.-J., Chang, K.-W., Ringgaard, M., and Lin, C.-J. (2010). Training and testing low-degree polynomial data mappings via linear svm. *Journal of Machine Learning Research*, 11:1471–1490.

[Chen et al., 1994] Chen, C. S., Hwang, J. C., Cho, M. Y., and Chen, Y. W. (1994). Development of simplified loss models for distribution system analysis. *IEEE Transactions on Power Delivery*, 9(3):1545–1551.

[Chen et al., 2005] Chen, C. S., Lin, C. H., Huang, M. Y., Chem, H. D., Kang, M. S., and Huang, C. F. (2005). Development of distribution feeder loss models by artificial neural networks. *IEEE Power App. Syst*, PAS-1:164–170.

[Chen and Guestrin, 2016] Chen, T. and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA. ACM.

[Chiang et al., 1997] Chiang, H. D., Wang, J. C., and Miu, K. N. (1997). Explicit loss formula, voltage formula and current flow formula for large-scale unbalanced distribution systems. *IEEE Transactions on Power Systems*, 12(3):1061–1067.

[CNMC, 2019] CNMC (2019). Incentivo a la reduccion de perdidas. Technical Report RDC/DE/003/21, Comision Nacional de los Mercados y la Competencia.

[Conejo and Wu, 2021] Conejo, A. and Wu, X. (2021). Robust optimization in power systems: a tutorial overview. *Optimization and Engineering*.

[Cortes and Vapnik, 1995] Cortes, C. and Vapnik, V. N. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.

[C.T.Gaunt et al., 2017] C.T.Gaunt, E.Namanya, and R.Herman (2017). Voltage modelling of LV feeders with dispersed generation: Limits of penetration of randomly connected photovoltaic generation. *Electric Power Systems Research*, 143:1–6.

[D. C. Montgomery, 1999] D. C. Montgomery, G. C. R. (1999). *Applied Statistics and Probability for Engineers*. New York. Wiley, Boston, 4 edition.

[DallAnese et al., 2013] DallAnese, E., Zhu, H., and Giannakis, G. B. (2013). Distributed optimal power flow for smart microgrids. *IEEE Trans. Smart Grid*, 4(3).

[Dashtaki and Haghifam, 2013] Dashtaki, A. K. and Haghifam, M. R. (2013). A new loss estimation method in limited data electric distribution networks. *IEEE Transactions on Power Delivery*, -(4):2194–2200.

[de Souto et al., 2008] de Souto, M. C. P., de Araujo, D. S. A., Costa, I. G., Soares, R. G. F., Ludermir, T. B., and Schliep, A. (2008). Comparative study on normalization procedures for cluster analysis of gene expression datasets. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 2792–2798.

[Deisenroth et al., 2020] Deisenroth, M. P., Faisal, A. A., and Soon, C. (2020). *Mathematics for Machine Learning*. Cambridge University Press.

[Ding et al., 2017] Ding, T., Li, C., Yang, Y., Jiang, J., Bie, Z., and Blaabjerg, F. (2017). A two-stage robust optimization for centralized-optimal dispatch of photovoltaic inverters in active distribution networks. *IEEE Transactions on Sustainable Energy*, 8(2):744–754.

[Dortolina and Nadira, 2005] Dortolina, C. A. and Nadira, R. (2005). The loss that is unknown is no loss at all: A top-down/bottom-up approach for estimating distribution losses. *IEEE Transactions on Power Systems*, 20(2):1119–1125.

[European Comission, 2019] European Comission (2019). Directive (EU) 2019/944 of the European Parliament and of the Council of 5 June 2019 on common rules for the internal market for electricity and amending Directive 2012/27/EU .

[Fang et al., 2019] Fang, L., Ma, K., Li, R., Wang, Z., and Shi, H. (2019). A statistical approach to estimate imbalance-induced energy losses for data-scarce low voltage networks. *IEEE Transactions on Power Systems*, 34(4):2825–2835.

[Flammini et al., 2019] Flammini, M. G., Prettico, G., Julea, A., Fulli, G., Mazza, A., and Chicco, G. (2019). Statistical characterisation of the real transaction data gathered from electric vehicle charging stations. *Electric Power Systems Research*, 166:136–150.

[Flaten, 1988] Flaten, D. L. (1988). Distribution system losses calculated by percent loading. *IEEE Transactions on Power Systems*, 3(3):1263–1269.

[Fotouhi et al., 2019] Fotouhi, Z., Hashemi, M. R., Narimani, H., and Bayram, I. S. (2019). A general model for ev drivers charging behavior. *IEEE Transactions on Vehicular Technology*, 68(8):7368–7382.

[Franco et al., 2018] Franco, J. F., Ochoa, L. F., and Romero, R. (2018). Ac opf for smart distribution networks: An efficient and robust quadratic approach. *IEEE Transactions on Smart Grid*, 9(5).

[Fu et al., 2016] Fu, X., Chen, H., Cai, R., and Xuan, P. (2016). Improved LSF method for loss estimation and its application in DG allocation. *IET Generation, Transmission and Distribution*, 10(10):2512–2519.

[Garcia et al., 2000] Garcia, P. A. N., Luiz, J., Pereira, R., Carneiro, S., Da Costa, M., and Martins, N. (2000). Three-Phase Power Flow Calculations Using the Current Injection Method. *IEEE Transactions on Power Systems*, 15(2).

[Geron, 2017] Geron, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow*. O'Reilly Media, Inc.

[Goldberger, 1964] Goldberger, A. S. (1964). Econometric theory. new york: John wiley and sons. *Journal of Machine Learning Research*.

[Goldstein and Wooff, 2007] Goldstein, M. and Wooff, D. (2007). *Bayes Linear Statistics, Theory and Methods*. Wiley.

[Goodfellow et al., 2016a] Goodfellow, I., Bengio, Y., and Courville, A. (2016a). *Deep Learning*. MIT Press.

[Goodfellow et al., 2016b] Goodfellow, I., Bengio, Y., and Courville, A. (2016b). *Deep Learning*. MIT Press.

[Gouveia et al., 2021] Gouveia, A. M. V., Carvalho, P. M. S., and Dias, A. M. F. (2021). Effects of high resolution load modelling onto simulated lv distribution losses. *IEEE Transactions on Power Systems*, 36(2):1537–1545.

[Greene, 2003] Greene, W. H. (2003). *Econometric Analysis (Fifth ed.)*. Prentice-Hall.

[Grigoras and Scarlatache, 2013] Grigoras, G. and Scarlatache, F. (2013). Energy losses estimation in electrical distribution networks with a decision trees-based algorithm. In *2013 8th International Symposium on advanced topics in electrial engineering (ATEE)*, pages 1–4. IEEE.

[Grus, 2015] Grus, J. (2015). *Data Science from Scratch. First principles with python.* O Reilly Media.

[Guha et al., 1999] Guha, S., Rastogi, R., and Shim, K. (1999). Rock: a robust clustering algorithm for categorical attributes. In *Proceedings 15th International Conference on Data Engineering (Cat. No.99CB36337)*, pages 512–521.

[Guo et al., 2019] Guo, Z., Zhou, Z., and Zhou, Y. (2019). Impacts of Integrating Topology Re-configuration and Vehicle-to-Grid Technologies on Distribution System Operation. *IEEE Transactions on Power Systems*, 29(1).

[Gustafson, 1983] Gustafson, M. W. (1983). Demand, Energy, and Marginal Electric System Losses. *IEEE Power Engineering Review*, PAS-102(9):3189–3195.

[Gustafson and Baylor, 1988] Gustafson, M. W. and Baylor, J. S. (1988). The equivalent hours loss factor revisited. *IEEE Transactions on Power Systems*, 3(4):1502–1508.

[Gustafson and Baylor, 1989] Gustafson, M. W. and Baylor, J. S. (1989). Approximating the system losses equation. *IEEE Power Engineering Review*, 4(3):850–855.

[Gutierrez-Lagos et al., 2019] Gutierrez-Lagos, L., Ochoa, L. F., and . (2019). OPF-Based CVR operation in PV-Rich MV-LV distribution networks. *IEEE Transactions on Power Systems*, 34(4).

[Haim, 2006] Haim, Y. B. (2006). *Info-Gap Decision Theory*. California: Academic Press.

[Harrington, 2012] Harrington, P. (2012). *Machine Learning in Action.* Manning.

[Hart et al., 2011] Hart, W. E., Watson, J.-P., and Woodruff, D. L. (2011). Pyomo: modeling and solving mathematical programs in Python. *Mathematical Programming Computation*, 3(3):219–260.

[Hartigan and Wong, 1979] Hartigan, J. A. and Wong, M. A. (1979). Algorithm as 136: A k-medias clustering algorith. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 28:100–108.

[Hastie, 1990] Hastie, T. J.; Tibshirani, R. J. (1990). *Generalized Additive Models.* Chapman and Hall/CRC.

[Hastie and Friedman, 1991] Hastie, T. and Friedman, R. T. J. (1991). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer Series in Statistics.

[Hastie et al., 2017] Hastie, T., Tibshirani, R., Witten, D., and James, G. (2017). *Introduction to Statistical Learning: with Applications in R.* Springer.

[Hayter, 2012] Hayter, A. (2012). *Probability and statistics for engineers and scientists.* Brooks/Cole.

[Heckmann et al., 2013] Heckmann, W., Barth, H., Reimann, T., Hamann, L., Dasenbrock, J., Scheidler, A., Braun, M., and Ma, C. (2013). Detailed analysis of network losses in a million customer distribution grid with high penetration of distributed generation. In *22nd International Conference and Exhibition on Electric- ity Distribution (CIRED 2013)*.

[Henrion et al., 2009] Henrion, R., Kuchler, C., and Romisch, W. (2009). Scenario reduction in stochastic programming with respect to discrepancy distances. *Comput. Optim. Appl.*, 1(43):67–93.

[Heydarian-Forushani et al., 2018a] Heydarian-Forushani, E., Golshan, M. E. H., Shafiekhah, M., and Siano, P. (2018a). Optimal operation of emerging flexible resources

considering sub-hourly flexible ramp product. *IEEE Transactions on Sustainable Energy*, 9(2):916–929.

[Heydarian-Forushani et al., 2018b] Heydarian-Forushani, E., Golshan, M. E. H., and Siano, P. (2018b). Evaluating the operational flexibility of generation mixture with an innovative techno-economic measure. *IEEE Transactions on Power Systems*, 33(2):2205–2218.

[Hoerl, 1970] Hoerl, A. E. R. W. K. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12:55–67.

[Hoke et al., 2011] Hoke, A., Brissette, A., Maksimovic, D., Pratt, A., and Smith, K. (2011). Electric vehicle charge optimization including effects of lithium-ion battery degradation. In *IEEE Vehicle Power and Propulsion Conference, Chicago, IL*, pages 1–8.

[Hong and Chao, 2002] Hong, Y. Y. and Chao, Z. T. (2002). Development of energy loss formula for distribution systems using FCN algorithm and cluster-wise fuzzy regression. *IEEE Transactions on Power Delivery*, 17(3):794–799.

[Hosmer and Lemeshow, 2000] Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression (2nd ed.)*. Wiley.

[Howard and Karlin, 1984] Howard, M. T. and Karlin, S. (1984). *An introduction to Stochastic Modeling,*. London: Academic Press limited.

[Hsu et al., 1995] Hsu, C. T., Tzeng, Y. M., Chen, C. S., and Cho, M. Y. (1995). Distribution feeder loss analysis by using an artificial neural network. *Electric Power Systems Research*, 34:85–90.

[Hyndman et al., 2018] Hyndman, R., Athanasopoulos, G., and (2018). *Forecasting: principles and practice*. OTexts: Melbourne, Australia. OTexts.com/fpp2.

[Hyndman and Fan, 2010] Hyndman, R. and Fan, S. (2010). Density forecasting for long-term peak electricity demand. *IEEE Transactions on Power Systems*, 25(2):1142–1153.

[Ibrahim et al., 2017] Ibrahim, K. A., Au, M. T., Gan, C. K., and Tang, H. J. (2017). System wide MV distribution network technical losses estimation based on reference feeder and energy flow model. *International Journal of Electrical Power and Energy Systems*, 93:440–450.

[Jabr et al., 2012] Jabr, R. A., Singh, R., and Pal, B. C. (2012). Minimum loss network reconfig- uration using mixed-integer convex programming. *IEEE Transactions on Power Systems*, 27(2).

[Jolliffe, 2002] Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer.

[Jungnikel, 2009] Jungnikel, D. (2009). *Graphs, Networks and Algorithms*. Springer-Verlag.

[Kang et al., 2006] Kang, M.-S., Chen, C. S., Lin, C.-H., Huang, C.-W., and Kao, M.-F. (2006). A Systematic Loss Analysis of Taipower Distribution System. *IEEE Transactions on Power Systems*, 21(3):1062–1068.

[Ke-Lin and M. N. S., 2014] Ke-Lin, D. and M. N. S., S. (2014). *Neural Networks and Statistical Learning*. Springer.

[Keane et al., 2013] Keane, A., Ochoa, L. F., Borges, C. L., Ault, G. W., Alarcon-Rodriguez, A. D., Currie, R. A., Pilo, F., Dent, C., and Harrison, G. P. (2013). State-of-the-art techniques and challenges ahead for distributed generation planning and optimization. *IEEE Transactions on Power Systems*, 28(2).

[Kuhn and Johnson, 2013] Kuhn, M. and Johnson, K. (2013). *Applied Predictive Modelling*. Springer.

[Kumar et al., 2017] Kumar, V., Abraham, O. A., and SnÃ¡Å¡el, V. (2017). Metaheuristic design of feedforward neural networks: A review of two decades of research. *Engineering Applications of Artificial Intelligence*, 60:97–116.

[Leal et al., 2009] Leal, A. G., Jardini, J. A., Magrini, L. C., and Ahn, S. U. (2009). Distribution Transformer Losses Evaluation:A New Analytical Methodology and Artificial Neural Network Approach. *IEEE Transactions on Power Systems*, 24(2):705–712.

[Leonard and Peter, 1990] Leonard, K. and Peter, R. J. (1990). *Partitioning Around Medoids (Program PAM)*. Wiley Series in Probability and Statistics John Wiley and Sons, Inc.

[Li et al., 2019] Li, H., Cui, H., and Li, C. (2019). Distribution Network Power Loss Analysis Considering Uncertainties in Distributed Generations. *Sustainability*, 11(1311).

[Li et al., 2016] Li, N., Uckun, C., Constantinescu, E., Birge, J., Hedman, K., and Botterud, A. (2016). Flexible operation of batteries in power system scheduling with renewable energy. In *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pages 1–1.

[Ljungberg, 2006] Ljungberg, L. (2006). *Architecture of distribution networks in the presence of decentralized production - planning under uncertainties and decentralized operating modes*. PhD thesis, Chalmers University of Technology.

[López et al., 2015] López, G., Moreno, J. I., Amarís, H., and Salazar, F. (2015). Paving the road toward Smart Grids through large-scale advanced metering infrastructures. *Electric Power Systems Research*, 115:194–205.

[Lorenzo et al., 1994] Lorenzo, E., Araujo, G., Cuevas, A., Egido, M., Minano, J., and Zilles, R. (1994). *Solar Electricity: Engineering of Photovoltaic Systems*. Progensa Editorial.

[Lu et al., 2013] Lu, H., Venetsanopoulos, A., and Plataniotis, K. N. (2013). *Multilinear Subspace Learning*. CRC Press.

[Mandel, 1990] Mandel, J. (1990). *The Statistical Analysis of Experimental Data*. New York: Interscience.

[Massopust, 2010] Massopust, P. R. (2010). *Interpolation and Approximation with Splines and Fractals*. Oxford University Press.

[Mateo et al., 2018] Mateo, C., Prettico, G., Gómez, T., Cossent, R., Gangale, F., Frías, P., and Fulli, G. (2018). European representative electricity distribution networks. *International Journal of Electrical Power and Energy Systems*, 99:273–280.

[McKenna et al., 2012] McKenna, E., Richardson, I., and Thomson, M. (2012). Smart meter data: Balancing consumer privacy concerns with legitimate applications. *Energy Policy*, 41:807–814.

[McLachlan, 2004] McLachlan, G. J. (2004). *Discriminant Analysis and Statistical Pattern Recognition*. Wiley Interscience.

[McLoughlin et al., 2010] McLoughlin, F., Duffy, A., and Conlon, M. (2010). The Generation of Domestic Electricity Load Profiles through Markov Chain Modelling. *Euro-Asian Journal of Sustainable Energy Development Policy*, 3.

[Miao et al., 2016] Miao, S., Xie, K., Karki, H. Y. R., Tai, H. M., and Chen, T. (2016). A mixture kernel density model for wind speed probability distribution estimation,. *Energy Conversion and Management*, 126(2):1066–1083.

[Mikić, 2007] Mikić, O. M. (2007). Variance-based energy loss computation in low voltage distribution networks. *IEEE Transactions on Power Systems*, 22(1):179–187.

[Monteiro et al., 2018] Monteiro, R. V., Guimarães, G. C., Silva, F. B., da Silva Teixeira, R. F., Carvalho, B. C., Finazzi, A. d. P., and de Vasconcellos, A. B. (2018). A medium-term analysis of the reduction in technical losses on distribution systems with variable demand using artificial neural networks: An Electrical Energy Storage approach. *Energy*, 164:1216e1228.

[Montgomery et al., 2008] Montgomery, D. C., Jennings, C. L., and Kulahci, M. (2008). *Introduction to Time Series Analysis and Forecasting*. Wiley.

[Murphy, 2012] Murphy, K. P. (2012). *Machine Learning A Probabilistic Perspective*. The MIT Press.

[Navarro-Espinosa et al., 2014] Navarro-Espinosa, A., Ochoa, L. F., and Randles, D. (2014). Assessing the Benefits of Meshed Operation of LV Feeders with Low Carbon Technologies. In *Innovative Smart Grid Technologies Conference (ISGT 2015)*, Istambul, Turkey.

[Nelder and Wedderburn, 1972] Nelder, J. and Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 3.

[Ng and Han, 1998] Ng, R. T. and Han, J. (1998). Clarans: A method for clustering objects for spatial data mining. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, 14(5):1003–1016.

[Nijhuis et al., 2017] Nijhuis, M., Gibescu, M., and J.F.G.Cobben (2017). Valuation of measurement data for low voltage network expansion planning. *Electric Power Systems Research*, 151:59–67.

[N.Srivastava et al., 2014] N.Srivastava, G.Hinton, A.Krizhevsky, I.Sutskever, and R.Salakhutdi (2014). Dropout: A simple way to prevent neural network from overfitting. *J. Mach. Learn.*, 15:1929–1958.

[Ochoa and Harrison, 2011] Ochoa, L. F. and Harrison, G. P. (2011). Minimizing energy losses: Optimal accommodation and smart operation of renewable distributed generation. *IEEE Transactions on Power Systems*, 26(1).

[O'Connell et al., 2014] O'Connell, A., Flynn, D., , and Keane, A. (2014). Rolling multiperiod optimization to control electric vehicle charging in distribution networks. *IEEE Transactions on Power Systems*, 29(1).

[Olivier et al., 2016] Olivier, F., Aristidou, P., Ernst, D., and Van Cutsem, T. (2016). Active management of low-voltage networks for mitigating overvoltages due to photovoltaic units. *IEEE Transactions on Smart Grid*, 7(2):926–936.

[Onen, 2016] Onen, A. (2016). Energy Saving of Conservation Voltage Reduction Based on Load-Voltage Dependency. *Sustainability*, 8:803.

[Ostling, 2021] Ostling, L. (2021). Smart Metering in Europe - 16th Edition. Technical report, Research and Markets.

[P. et al., 1977] P., D. A., N., L., and B., R. D. (1977). "maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39:1–38.

[Paszke et al., 2019] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc.

[Pedregosa et al., 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

[Pedro, 2009] Pedro, A. (2009). Reducing Technical and Non Technical Losses in the Power Sector. Technical report, World Bank.

[Pinsky and Karlin, 2011] Pinsky, M. A. and Karlin, S. (2011). 3 - markov chains: Introduction. In Pinsky, M. A. and Karlin, S., editors, *An Introduction to Stochastic Modeling (Fourth Edition)*, pages 79 – 163. Academic Press, Boston, fourth edition edition.

[Poursharif et al., 2018] Poursharif, G., Brint, A., Black, M., and Marshall, M. (2018). Using smart meters to estimate low-voltage losses. *IET Generation, Transmission & Distribution*, 12(5):1206–1212.

[Queiroz et al., 2012] Queiroz, L. M. O., Roselli, M. A., Cavellucci, C., and Lyra, C. (2012). Energy losses estimation in power distribution systems. *IEEE Transactions on Power Systems*, 27(4):1879–1887.

[Quinlan, 1986] Quinlan, J. (1986). Induction of decision trees. *Machine Learning*, 29:81–106.

[Quiros-Tortos et al., 2016] Quiros-Tortos, J., Ochoa, L. F., Alnaser, S. W., and Butler, T. (2016). Control of ev charging points for thermal and voltage management of lv

networks. In *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pages 1–1.

[Rao and Deekshit, 2006] Rao, P. S. N. and Deekshit, R. (2006). Energy Loss Estimation in Distribution Feeders. *IEEE Transactions on Power Delivery*, 21(3):1092–1100.

[Rashid, 2017] Rashid, T. (2017). *Make Your Own Neural Network*. Amazon Media EU.

[Report, 2015] Report, I. D. P. W. G. (2015). The ieee european low voltage test feeder.

[Reynolds, 2009] Reynolds, D. A. (2009). Gaussian mixture models. In *Encyclopedia of Biometrics*.

[Ricciardi et al., 2019] Ricciardi, T. R., Petrou, K., Franco, J. F., and Ochoa, L. F. (2019). Defining customer export limits in pv-rich low voltage networks. *IEEE Transactions on Power Systems*, 34(1).

[Richardson et al., 2013] Richardson, P., Flynn, D., and Keane, A. (2013). Local versus centralized charging strategies for electric vehicles in low voltage distribution systems. In *2013 IEEE Power Energy Society General Meeting*, pages 1–1.

[Rigoni et al., 2016] Rigoni, V., Ochoa, L. F., Chicco, G., Navarro-Espinosa, A., and Gozel, T. (2016). Representative Residential LV Feeders: A Case Study for the North West of England. *IEEE Transactions on Power Systems*, 31(1):348–360.

[Robbins and Monro, 1951] Robbins, H. and Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3):400–407.

[Rousseeuw, 1987] Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.

[Russel and Norving, 1994] Russel, S. and Norving, P. (1994). *Artificial Intelligence: A Modern Approach*. Prentice-Hall.

[Schmidhuber, 2015] Schmidhuber, J. (2015). Deep Learning in neural networks: An overview. *Neural Networks*, 61:65–117.

[Schultz, 1978] Schultz, N. R. (1978). Distribution Primary Feeder I2R Losses. *IEEE Transactions on Power Apparatus and Systems*, PAS-97(2):603–609.

[Seabold and Perktold, 2010] Seabold, S. and Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference.*

[Sereeter et al., 2017] Sereeter, B., Vuik, K., and Witteveen, C. (2017). Newton power flow methods for unbalanced three-phase distribution networks. *Energies*, 10(10).

[Shafie-Khah et al., 2018a] Shafie-Khah, M., Mahmoudi, N., Siano, P., Saha, T. K., and Catalao, J. P. S. (2018a). A comprehensive model to integrate emerging resources from supply and demand sides. *IEEE Transactions on Smart Grid*, 9(4):3883–3896.

[Shafie-Khah et al., 2018b] Shafie-Khah, M., Siano, P., and Catalao, J. S. (2018b). Optimal demand response strategies to mitigate oligopolistic behaviour of generation companies using a multi-objective decision analysis. *IEEE Transactions on Power Systems.*

[Shafie-khah et al., 2018] Shafie-khah, M., Siano, P., Fitiwi, D. Z., Mahmoudi, N., and Catalao, J. (2018). An innovative two-level model for electric vehicle parking lots in distribution systems with renewable energy. In *2018 IEEE Power Energy Society General Meeting (PESGM)*, pages 1–1.

[Shenkman, 1990] Shenkman, A. L. (1990). Energy loss computation by using statistical techniques. *IEEE Transactions on Power Delivery*, 5(1):254–258.

[Short, 2004] Short, T. A. (2004). *Electric Power Distribution Handbook.* CRC Press.

[Silverman, 1986] Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis.* London: Chapman & Hall/CRC.

[Simon S, 1999] Simon S, H. (1999). *Neural Networks: A Comprehensive Foundation.* Prentice Hall.

[Soroudi, 2011] Soroudi, A. (2011). *Loss Variation Calculations in Distribution Grids.* PhD thesis, Universite de Grenoble.

[Soroudi, 2012] Soroudi, A. (2012). Possibilistic-scenario model for dg impact assessment on distribution networks in an uncertain environment. *IEEE Transactions on Power Systems*, 27(3).

[Soroudi and Amraee, 2013] Soroudi, A. and Amraee, T. (2013). Decision making under uncertainty in energy systems: State of the art. *Renewable and Sustainable Energy Reviews*, 28:376–384.

[Soroudi et al., 2017] Soroudi, A., Rabiee, A., and Keane, A. (2017). Distribution network's energy losses versus hosting capacity of wind power in the presence of demand flexibility. *Renewable Energy*, 102:316–325.

[Soroudi et al., 2016] Soroudi, A., Siano, P., and Keane, A. (2016). Optimal dr and ess scheduling for distribution losses payments minimization under electricity price uncertainty. *IEEE Transactions on Smart Grid*, 7(1):261–272.

[Stephen et al., 2014] Stephen, M., Colin, A., Kiprakis, A., and Tsagarakis, G. (2014). Development of Low-Voltage Load Models for the Residential Load Sector. *Power Systems, IEEE Transactions on*, 29.

[Strunz et al., 2009] Strunz, K., Fletcher, R. H., Campbell, R., and Gao, F. (2009). Developing benchmark models for low-voltage distribution feeders. In *2009 IEEE Power and Energy Society General Meeting, PES '09*.

[Sudipto et al., 1998] Sudipto, G., Rajeev, R., and Shim, K. (1998). Cure: An efficient clustering algorithm for large databases. *Information Systems*, 2(26):35–58.

[Sun et al., 1980] Sun, D. I. H., Abe, S., Shoults, R. R., Chen, M. S., Eichenberger, P., and Farris, D. (1980). Calculation of energy losses in a distribution system. *IEEE Transactions on Power Apparatus and Systems*, PAS-99(4):1347–1356.

[Taleski and Rajicic, 1996] Taleski, R. and Rajicic, D. (1996). Energy summation method for energy loss computation in radial distribution networks. *IEEE Transacttions on Power Systems*, 11(2):1104–1111.

[Theodoridis and Koutroumbas, 2008] Theodoridis, S. and Koutroumbas, K. (2008). *Pattern Recognition*. Academic Press.

[Tibshirani, 1996] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (methodological)*, 1.

[UC3M et al., 2017] UC3M, Naturgy, Tecnalia, Orbis, and ZIV (2014-2017). OSIRIS (Optimizacion de la Supervision Inteligente de la Red de DIstirbuticion).

[Union Fenosa Distribucion, 2011a] Union Fenosa Distribucion (2011a). Especificaciones particulares para instalaciones de conexion. instalaciones de enlace en baja tension. `https://formularios.ufd.es/servlet/ficheros/1297159684225/476_279_ES.0100.ES.RE.EICEd.5080911INSTALACIONESDEENLACE.pdf`. accessed on 15/03/2021.

[Union Fenosa Distribucion, 2011b] Union Fenosa Distribucion (2011b). Proyecto tipo lineas electricas aereas de baja tension. `https://www.ufd.es/wp-content/uploads/2018/08/4285C2525CIT.0110.ES_.RE_.PTPEd_.3220711LABT.pdf`. accessed on 23/02/2020.

[Urquhart et al., 2015] Urquhart, A. J., Thomson, M., and . (2015). Impacts of Demand Data Time Resolution on Estimates of Distribution System Energy Losses. *IEEE Transactions on Power Systems*, 30(3).

[Usman et al., 2018] Usman, M., Coppo, M., Bignucolo, F., and Turri, R. (2018). Losses management strategies in active distribution networks: A review. *Electric Power Systems Research*, pages 116–132.

[Velasco et al., 2020] Velasco, J. A., Amaris, H., and Alonso, M. (2020). Deep learning loss model for large-scale low voltage smart grids. *International Journal of Electrical Power and Energy Systems*, 121:106054.

[Velasco et al., 2019] Velasco, J. A., Rigoni, V., Soroudi, A., Keane, A., and Amaris, H. (2019). optimising load flexibility for the day ahead in distribution networks with photovoltaics. In *2019 IEEE Milan PowerTech*, pages 1–6.

[Vempati et al., 1987] Vempati, N., Shoults, R. R., Chen, M. S., and L., S. (1987). Simplified feeder modeling for load flow calculations. *IEEE Transactions on Power Systems*, PWRS-2(1):168–174.

[Wang et al., 2017] Wang, S., Dong, P., and Tian, Y. (2017). A Novel Method of Statistical Line Loss Estimation for Distribution Feeders Based on Feeder Cluster and Modified XGBoost. *Energies*, 10.

[Wiebe and Misener, 2021] Wiebe, J. and Misener, R. (2021). Romodel: Modeling robust optimization problems in pyomo.

[Xu et al., 2018] Xu, J., Wang, J., Liao, S., Sun, Y., Deping, K., Li, X., Liu, J., Jiang, Y., and Wei, C. (2018). Stochastic multi-objective optimization of photovoltaics integrated three-phase distribution network based on dynamic scenarios. *Applied Energy*, 231:985–996.

[Xu et al., 2015] Xu, X., Yan, Z., and Xu, S. (2015). Estimating wind speed probability distribution by diffusion-based kernel density method. *Electric Power Systems Research*, 121(2):28–37.

[Zhang et al., 1966] Zhang, T., Ramakrishnan, R., and Livny, N. (1966). Birch: an efficient data clustering method for very large databases. In *Proceedings of the 1996 ACM SIGMOD international conference on Management of data*, pages 103–114.

[Zou and Hastie, 2005] Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B.*, 2:301–320.

[Zucchini et al., 2009] Zucchini, W., MacDonald, I. L., and Langrock, R. (2009). *Hidden Markov Models for Time Series An Introduction Using R, Second Edition.* Chapman and Hall/CRC.