Réduction d'un modèle de système électrique pour des études technico-économiques

Nuno Pinto Marinho

université PARIS-SACLAY

NNT: 2018SACLC051

CentraleSupélec

Reduction of an electrical power system model for techno-economic studies

Thèse de doctorat de l'Université Paris-Saclay
préparée à CentraleSupélec

École doctorale n° 575 EOBE
Spécialité de doctorat: Génie électrique

Thèse présentée et soutenue à Gif-sur-Yvette, le 19 juin 2018, par

Nuno Filipe Pinto Marinho

Composition du jury:

Samson Lasaulce Directeur de Recherche — L2S	Président
Keith Bell Professeur — University of Strathclyde	Rapporteur
Goran Strbac Professeur — Imperial College of London	Rapporteur
Tanguy Janssen Ingénieur de Rechercher — RTE	Examinateur
Martin Hennebel Maître de Conférence — CentraleSupélec	Examinateur
Jean-Claude Vannier Professeur — CentraleSupélec	Directeur de thèse
Jean-Yves Bourmaud Ingénieur de Recherche — RTE	Invité
Adrien Atayi Ingénieur de Recherche — EDF	Invité
Yannick Phulpin Ingénieur de Recherche — EDF	Invité

Thèse de doctorat

"If I have seen further it is by standing on ye shoulders of Giants."

Isaac Newton, *Letter to Robert Hooke*, 15 February 1676.

Abstract

The simulation of complex processes in large scale power systems needs the reduction of the problem. How to reduce the spatial complexity of a large scale power network while minimizing information loss? To answer this question we have divided this work in three main steps: 1) network buses aggregation; 2) modelling of the clusters' links; 3) defining the equivalent branches maximum exchange capacity.

The bus aggregations in a cluster implies that it will be treated as a coppper-plate by the market model. Therefore, the most frequent network congestions must be identified ideally placed at the clusters frontiers. After the reduction, the same power flow repartition must be found in both reduced and complete model. To do that, a methodology to define a PTDF matrix was developed. For economic purpose studies, the branches maximum capacity is a key parameter, to define this value, a methodology is proposed that estimates the equivalent transmission capacities using historical system operating set points.

These approaches were applied to the European transmission network and allowed to define a reduced model that minimises the information loss.

Key words: Electricity markets, Network reduction, Locational Marginal Prices, Power Transfer Distribution Factors.

Résumé

La simulation des processus complexes dans des réseaux de transport d'électricité de grande taille nécessite la réduction de la dimension du problème. Comment réduire la complexité spatiale d'un réseau de grande dimension en gardant un bon niveau de précision ? Pour répondre à cette question nous avons divisé ce travail en trois grandes étapes : 1) la réduction par agrégation du nombre de noeuds; 2) la modélisation des liaisons entre ces clusters de noeuds et 3) le calcul des capacités des lignes équivalentes.

L'agrégation des noeuds dans un cluster implique que celui-ci sera traité comme une plaque de cuivre par le modèle de marché. En conséquence, pour l'agrégation des noeuds, les congestions récurrentes dans le réseau sont identifiées et placées idéalement aux frontières des clusters. Après la réduction, la même répartition des flux dans le réseau complet et dans le modèle réduit du réseau doit être trouvée. Pour ce fait une méthodologie d'estimation d'une matrice PTDF a été développée. Pour les études économiques la limite thermique des lignes est un paramètre clé. Pour résoudre ce problème, nous proposons une méthodologie qui estime les capacités équivalentes à partir des points de fonctionnement historiques du système complet.

Les approches présentées dans ce travail ont été appliquées sur un modèle du réseau continental européen et ont permis d'obtenir un modèle simplifié qui minimise la perte d'information.

Mots-clés : Marché d'électricité, Réduction de réseau, Locational Marginal Prices, Power Transfer Distribution Factors.

Acknowledgements

I would like to thank Professor Keith Bell and Professor Goran Strbac for carefully reviewing this work, as well as the other members of the jury for their valuable remarks that improved the quality of this manuscript.

I would also like to thank my academic and industrial supervisors whose guidance was fundamental to the completion of this work. A special word must go to Yannick, your drive and enthusiasm were always an inspiration throughout this work and your contributions were unmeasurable both professional and personally.

Throughout these 3 years my time was spent working at EDF and I must thank all my colleagues for making me feel at home since day one. A special word must go to all the *DRIIME*r's that have crossed my way and that were fundamental for my professional and personal development.

Working at (Centrale)Supélec was also a real pleasure, and I must thank all my PhD colleagues for being so welcoming and making the *Département Énergie* such a pleasing environment to work and our weekend getaways were also important to recharge the energies needed to finish this PhD.

Throughout my stay in Paris at *Maison de Portugal* I had the opportunity to meet amazing people that were always supportive and understanding of my work. It was great to feel at home so many kilometres away from it. Also, even if the geographical distance did not allow meeting frequently, I can not forget my friends from FEUP that were also present from the beginning until the end of this work.

Last but not least, my family has always been the foundation that allowed me to progress every day and without them none of this would be possible.

Obrigado Pai, Mãe, Catarina, Hugo e Gabriela!

Nuno

Contents

List of Figures

List of Tables

Abbreviations

Acronym	Full form
AC	Alternating Current
ACER	Agency for the Cooperation of Energy Regulators
ATC	Available Transfer Capability
CHP	Combined Heat and Power
CWE	Central Western Europe
DC	Direct Current
ENS	Energy Not Supplied
ENTSO-E	European Network of Transmission System Operators for Electricity
EU	European Union
GSK	Generation Shift Key
HP RoR	Hydro Power Run-of-River
HVDC	High Voltage Direct Current
LCOE	Levelised Cost Of Electricity
LMP	Locational Marginal Price
NRMSE	Normalized Root Mean Square Error
OPF	Optimal Power Flow
PST	Phase Shifting Transformers
PTDF	Power Transfer Distribution Factor
PSDF	Phase Shifter Distribution Factor
PV	Photovoltaic
RES	Renewable Energy Sources
RMSE	Root Mean Square Error
TSO	Transmission System Operator
TYNDP	Ten-Year Network Development Plan

UC	Unit Commitment
UCTE	Union for the Coordination of Transmission of Electricity
VaR	Value at Risk
VOLL	Value Of Loss Load
WP On	Wind Power Onshore
WP Off	Wind Power Offshore

Symbols

Sets

L	set of branches in the full system
$\bar{L}$	set of branches in the reduced system
N	set of buses in the full system
N_g	set of generators connected to the system
$\bar{N}$	set of clusters in the reduced system
S_T	set of training scenarios
S_E	set of evaluation scenarios
u_s	set of control variables of the optimization problem for scenario s
u^*	set of the optimal control variables of the global optimization problem
$\hat{u}$	set of the optimal control variables for each subproblem of the optimization methodology
x_s	set of state variables of the global optimization problem for scenario s
x^*	set of the optimal state variables of the optimization problem

Indices

g	index of the generators ranging from 1 to N_g
l	index of branch from the L branches of the full system
$\bar{l}$	index of branch from the $\bar{L}$ branches of the reduced system
n	index of bus from the N buses of the full system
$\bar{n}$	index of bus from the $\bar{N}$ buses of the reduced system

s	index of scenarios ranging from 1 to S_x

Variables

δ	PST tap position in degrees
F^{max}	branches' thermal limit
$F_{l,s}$	observed flow on branch l at scenario s
$\bar{F}_{\bar{l},s}$	estimated flow on the equivalent branch $\bar{l}$ at scenario s
f^0	loop flows between clusters
Γ	matrix of size $N \times \bar{N}$ that establishes a correspondence between the original buses and the ones they were aggregated into
ω	power reduction on a given branch due to a change in a PST tap position
P_s^{inj}	array containing the injected power for all buses N of the complete system at scenario s
$P_s^{inj^r}$	array containing the aggregated injected power for all clusters $\bar{N}$ of the reduced system at scenario s
$P_{n,s}^{inj}$	injected power in bus n at scenario s
$P_{\bar{n},s}^{inj^r}$	aggregated injected power in cluster $\bar{n}$ at scenario s
ΔP	power variation in a given asset
P_{gen_g}	power generated by generator g
Ψ	PTDF matrix of the full system of dimensions $L \times N$
Ψ_r	PTDF matrix of the reduced system of dimensions $\bar{L} \times \bar{N}$

Functions

$f(u_s, x_s)$	cost function of the whole system, representing the overall surplus
$g(u_s, x_s)$	function representing the equality constraints of the system
$h(u_s, x_s)$	function representing the inequality constraints of the system
$p_{\bar{n}}(u_s, x_s)$	function determining the net position of every cluster $\bar{n}$

General introduction

General context and motivation

Technological and regulatory evolution have severely increased the complexity of power system's operation. In a now competitive world, each system stake-holder must face this complexity, frequently, with fragmented information. Nevertheless, to perform both dynamic and static prospective studies of power systems, stakeholders rely generally on reduced system models, whose settings have a significant impact on the final results. The definition of a reduced model can be difficult, as a trade-off between results' accuracy and computation tractability needs to be performed. Therefore, an important question arises:

How to reduce the complexity of a system, while keeping an acceptable level of accuracy?

Throughout this thesis, a methodology to define a reduced system model for economic purposes is proposed. It focuses essentially on a topological simplification and can be divided into three main steps.

First, the network is clustered into a reduced number of buses, by aggregating buses connected along uncongested branches. Second, system branches are reduced, and a single link to connect clusters is defined using a Power Transfer Distribution Factor (PTDF) matrix. Third and final step, the reduced links' maximum transfer capacities are determined so they match the ones obtained with the original full model.

Thesis structure

The work developed throughout this thesis is organized in 5 chapters.

Chapter I defines the problem addressed in this thesis, and presents the literature review of the three main reduction trends, namely, temporal and spatial reduction and

optimisation linearisation. The motivation to perform spatial reduction throughout this work is also addressed, and a formal presentation of a large scale power system that will be used as test case throughout this work is done.

Chapter II details the first step to define a reduced model, namely the network clustering. An overview of the literature review is presented, the three most promising techniques are applied to a test case and a framework to rank the different approaches is proposed.

Chapter III presents a methodology to represent the links between the clusters obtained in Chapter II. A literature review is performed, stressing out their main limitations, and a new approach is proposed to overcome them.

Chapter IV describes the problem of defining equivalent transmission capacities in reduced models and the proposed solutions in the literature. A new methodology to estimate the reduced network's operational limits based on the full model's one is proposed, allowing the full characterization of the clusters' links by determining the equivalent transmission capacity between them.

Chapter V stresses the importance of power flow control devices in power systems, and proposes a methodology to explicitly represent them in reduced network models.

This document ends with the General Conclusions, where the key findings and further research are detailed.

Appendix A and B complement the main document. Appendix A presents a methodology to estimate generation variable costs, using only publicly available data, that are a key parameter for the economic analysis performed throughout this work, and Appendix B presents the main characteristics of the Optimal Power Flow (OPF) formulation in Power Factory.

Chapter I

How and why to reduce complexity?

* * *

This chapter is organized as follows. Section I.1 presents an overview of the complexity of power systems' optimization, Section I.2 discusses the main approaches proposed to reduce temporal and spatial complexity, as well as the motivations of this work to focus on spatial reduction. Finally, Section I.3 describes an example of a large scale power system, based on the European transmission network, that will be used as a benchmark throughout this work.

Contents of chapter I

I.1 The complexity of power systems' optimization

Until the liberalization of the energy sector, the optimization of large-scale power systems had been managed by a unique operator, responsible for all the assets within its own control area. After Article 9 of the Directive 2009/72/EC, the assets repartition through different system players has increased the uncertainty regarding system's operation. Transmission System Operator (TSO) must assure an economic development of the transmission system [1, 2] while ensuring the security of supply. Utilities should compete to maximize their profit while respecting the units' technical restrictions [3].

In this framework, two tools stand out in the literature as the most applied to power system's optimization: the Optimal Power Flow (OPF) and Unit Commitment (UC). Both optimize the generation unit's injections to respect given demands levels but vary in their purpose and mathematical formulation.

UC problems optimize the generation units' schedule to respect a given demand level over consecutive time steps. The goal of this optimization is to minimize operation costs or maximize producers' profit while respecting competition rules and the units' operational limitations. Those include inter-temporal constraints, corresponding to the unit's physic limitations such as ramp-rate, minimum up-times and warm-up period time. For example, starting a coal generation unit requires to keep it running for eight hours, and this can, therefore, impact the economic optimization of the producers' portfolio. More details on the mathematical formulation of this problem can be found in [4], for example.

This complex mathematical formulation poses challenges to the computational tractability when the system size and modelling detail increases, therefore, frequently, they do not include network models and rely instead on a copper plate approach, with the possible implications this may have in the final results.

OPF calculation determines the best output level for a given set of generation units to meet demand for a single time step. The optimization minimizes generation costs and ensures that steady-state network constraints are respected. Opposite from the UC problem, it is often assumed that no inter-temporal constraints are modelled, and each available unit can produce at full power. On the one hand, this assumption significantly eases the mathematical problem of production fleet optimization and consequently reduces its computation tractability, a new set of constraints are added that are not present

in the UC problem - steady-state network constraints. Those are key in the limitation of the economic optimum, as generators' dispatch must be adjusted to respect network elements' limits. When applied to large-scale power systems, these constraints significantly impact the optimization problem formulation and computation tractability.

The computation tractability problems become even more severe when applied to transmission/generation expansion problems. When performing long-term studies, the evolution of demand scenarios can degrade the reliability of the system or its capacity to satisfy the demand of all the consumers. This obliges to add network branches and generation units as optimization variables, increasing the mathematical complexity observed in the UC and OPF problems. Even using reduced network models, these problems can take up to 40h of simulation [5] and have therefore been a focus of intense research regarding its simplification [6].

The observation of these two trends allows to understand that both inter-temporal and steady-state network constraints have an important impact on the simulation of power systems. Therefore, an effort to reduce power systems' complexity has been made to allow the integration of simpler network models in more detailed economic simulations, such as the UC.

I.2 Well-known methods to reduce complexity

Three main types of simplifications are proposed when simulating large power systems, namely optimisation linearisation, temporal and spatial reduction. Comparing the three different approaches is difficult, as their results depend on the purpose of the studies, and no clear conclusions can be obtained [7].

Linearisation approaches focus on relaxation or omission of constraints specific to the optimization problems [8, 9, 10], being the most common the Direct Current (DC) linearisation (as detailed in [11]) which will be used throughout this work[1]. In this work, with the goal of defining a reduced network model that can be used independently of the optimization problem (i.e. the same approach allows to perform both UC and OPF calculations), a focus is made on both temporal and spatial reductions.

[1]Throughout this work OPF and UC will always refer to the DC versions of these problems

I.2.1 Temporal reduction

In adequacy assessment studies, temporal reduction has been performed empirically, as the most critical situation of the system was considered to be the winter's peak hour. The large penetration of Renewable Energy Sources (RES) in the system has increased the uncertainty in the generation and loads' forecasts.

The increase in uncertainty of these exogenous quantities has reinforced the need to use stochastic optimization problems (e.g. [12]), raising challenges regarding the computational tractability and forcing the reduction of temporal complexity in power systems.

On the generation side, the clustering of renewable energy sources has received particular attention in recent years due to its pronounced variability, and different methodologies were proposed to determine a reduced number of representative scenarios [13, 14, 15]. The introduction of load management programs has increased its uncertainty, forcing the need to consider multiple scenarios of load and therefore, some approaches to select the most representative scenarios have been proposed to ease the computational burden [16, 17]. An extensive review of load forecasting methods can be found in [18]. In addition, the reduction of temporal complexity has been applied to reduce the number of operational scenarios including load and generation, such as [19], or to reduce the number of outages scenarios to be considered, such as [20].

I.2.2 Spatial reduction

Spatial reduction methods are differentiated between those intended to perform system dynamic or steady-state analysis. For economic studies, the focus is on steady-state simulations, and therefore only those methods were considered in this section.

Historically, two methods were used to perform network reduction, the one proposed by J.B. Ward in [21], commonly known as Ward reduction, and the one proposed by P. Dimo in [22], commonly known as Radial,Equivalent,Independent (REI) reduction.

Ward reduction relies on the definition of an internal network, whose buses remain intact, and of frontier buses(which are the buses between the internal and the external network, whose buses and branches are reduced. The reduction is performed by Gaussian elimination of the admittance matrix, and loads/generators of the external network

would be distributed by the buses at the frontier. This uneven distribution is an important limitation of this methodology, as for example, in contingency analysis it would be hard to estimate the contribution of a specific generator of the external network.

In the REI reduction, the same definition of internal, external and frontier buses applies. The concept is to create a virtual frontier bus, containing all the generation and load of the external buses. The virtual buses are connected through a virtual REI network to the internal buses, the admittance matrix of this network is calculated so the flows in the internal network match those of the original system. The main limitation of these methodologies are the errors caused by a change in the operating set point.

Different methods were proposed to overcome those limitations, including a holomorphic embedding technique [23] to improve the accuracy of the models over a broader range of operating conditions. Nevertheless, this approach adds a layer of computation complexity in the reduction process that is not suitable for the reduction of large-scale power systems.

Variations were proposed to deal with economic studies that would mimic the Ward reduction, where generators contained in the reduced buses would be replicated in the internal system by calculating an average cost function representing the existing and reduced units. The internal branches' thermal limit were incrementally adjusted to the generation plan would mimic the one of the full model. This would allow for economic analysis of the reduced system, but the results were not sufficiently accurate as a change in the operating set point would require a new adjustment of the internal branches' thermal limit [24]. Similarly, another approach was proposed, denominated "inverse power flow" [25], that compensated the movement of the generators into the internal system. Loads were also reallocated so that the reduced system's power flows would match the ones of the complete system. Despite the promising results obtained, its applicability to large-scale power systems can be limited, in particular because:

- There is a need to have *a priori* information about network constraints, that limit the system throughout many hours, as an input;
- After the first Ward reduction, branches with an impedance bigger than 5 p.u. are removed, which is not particularly justified and can remove structural branches from the model. Some workarounds to this problem were presented [26], but

results tend to show that eliminating those branches from the system tends to severely increase the error associated with the model;

- A second ward reduction is performed to reallocate the generators using Dijkstra's algorithm [27] to move it to the closest (electrically) bus. This can pose problems regarding computational tractability, as Dijkstra's algorithm has a time complexity of $O(n^2)$, where n is the number of buses;
- Connecting generators based on electrical distance may distort the geographic data of the system (e.g. attach a generator to a different country), and therefore have an impact in economic studies;
- The load reallocation to mimic the power flows of the complete system may also end with a distribution that is not geographically coherent.

In recent years, some approaches with the same principle of the REI reduction have emerged. The idea is to gather all the generation and load of the reduced network on equivalent buses and then find a way to properly represent the flows in a way that matches the complete network. The most common way of representing the flows in the reduced model is by means of the Power Transfer Distribution Factor (PTDF) matrix [28, 29], but other works suggest also an aggregation based on physical parameters, such as electrical distance, and keeping the original system branches, as is the case in [30].

I.2.3 Motivations for spatial reduction

Both spatial and temporal reductions are viable solutions to reduce the complexity of large-scale power systems simulations and avoid the computational tractability problem. In prospective economic studies of power systems, it is important to have a temporal coherency, as for example, in unit commitment studies inter-temporal constraints are key in the obtained results. Also, in a case where different input scenarios need to be considered (e.g. different load or generation mix hypothesis), the temporal reduction requires a computation for each entry, as with spatial reduction, the goal is to have a model robust to different input parameters while ensuring computation tractability.

In this work, a focus on steady-state analysis of power system is made with the goal of performing prospective economic studies. Therefore, throughout this thesis, the focus will be on the reduction of spatial complexity, and the term "reduction" will refer specifically to the spatial steady state reduction of power systems.

As highlighted in Section I.2.2, existing approaches to reduce spatial complexity present nevertheless some limitations. Namely, methodologies to calculate equivalent branches' transmission capacity are scarce and limited, and therefore, reduction techniques tend to be unreliable to perform economic studies and this work seeks to fill this important gap in the existing literature.

Also, most of the reduction techniques proposed aim at a single operating point optimization and results present significant deviations when the system operating point changes and for the methodologies that present promising results, important challenges remain in their application to large-scale power systems as the number of variables to reduce increases significantly and so does computation time.

I.3 Examples of complex systems

Throughout the years, several factors have increased power systems' complexity: for example, the number of power flow control devices included in the network, the increase in interconnections between different control zones, the range of regulation and technologies applying to network users, etc.

The European transmission network is an intriguing example of a complex system whose optimisation poses significant computation challenges. A quick overview of the network topology is depicted in Figure I.1.

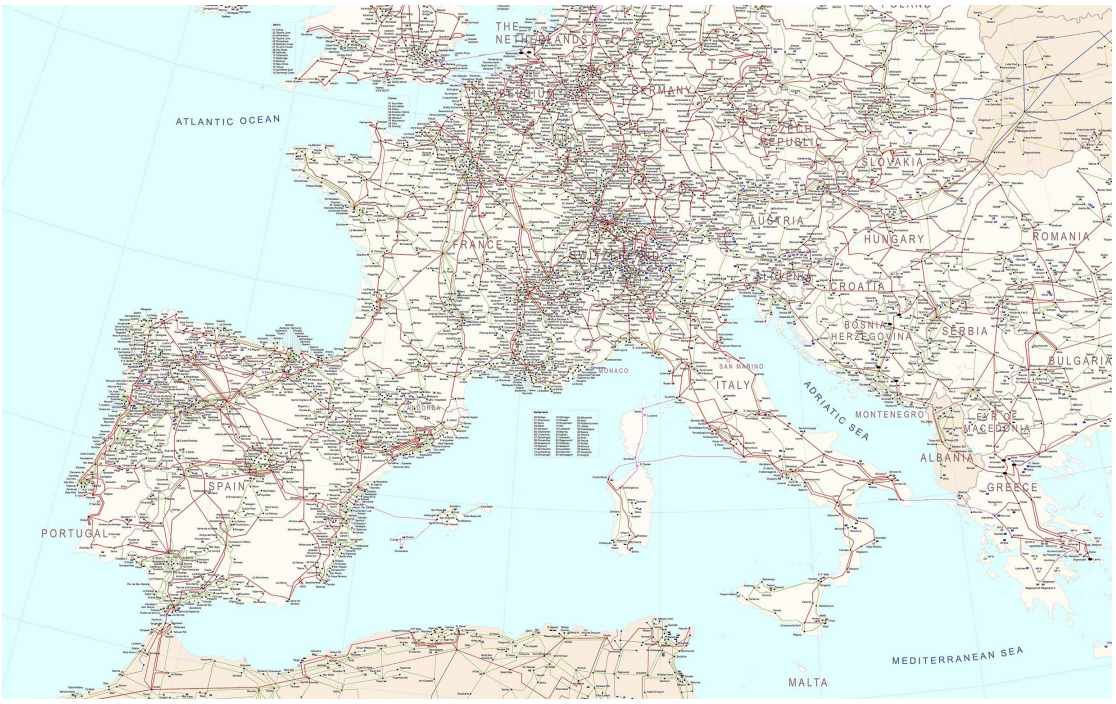

FIGURE I.1: Overview of the European transmission system. [31]

I.3.1 Complexity of large-scale power systems

The operation of large scale power systems undertakes to comply with several assets' technical restrictions. Some of them are translated into inter-temporal constraints such as generators' minimum up and down time which represents the minimum time period the generator must produce after being started and the time period it must stay offline when stopped, or the generators' ramp up/down constraints which define the output variation rate.

Different assets' settings can also change at each time step such as lines' thermal limit which defines the maximum apparent power that it can transmit at a given period, buses voltage limits which typically should be around $\pm 10\%$ of the nominal base voltage, Phase Shifting Transformers (PST) tap settings and topological configuration.

The optimization of all these operational problems in large-scale power systems is computationally challenging and, therefore, model simplifications are compulsory. In the next section, the network model that will be used as a benchmark to validate the proposed approaches throughout this thesis, and assess them in an operational problem with realistic features, will be presented.

I.3.2 Benchmark system

This section presents the model based on the Union for the Coordination of Transmission of Electricity (UCTE) 2008 - network model [32] that represents a vast majority of the Central and Eastern European transmission network (countries with reduced interconnection capacities have a simplified representation, as detailed below). Its main characteristics are detailed below as well as a data set that allows the simulation of the system operating conditions throughout a year, namely its load and generation characteristics. A quick overview of the network topology is depicted in Figure I.2, using the geographical data provided in [33] the assets were drawn over a Google Maps image of continental Europe.

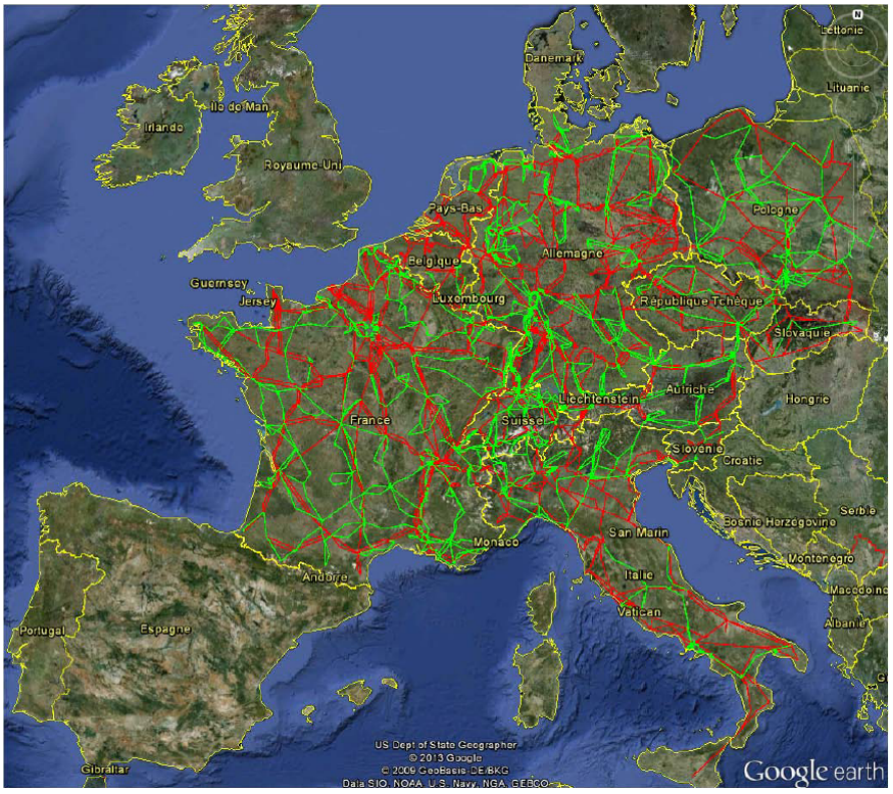

FIGURE I.2: Modified UCTE 2008 network model - Red lines represent the 400kV assets and green lines the 225kV network.

I.3.2.a Network data

The benchmark network model contains a total of 2842 buses, 1820 generators and 3739 branches. This represents mostly the 225kV and 400kV networks of twelve areas, namely,

France (FR), Belgium (BE), Netherlands (NL), Germany (DE), Austria (AT), Switzerland (CH), Italy (IT), Slovenia (SL), Poland (PL), Czech Republic (CZ), Slovakia (SK) and West Denmark (DK). Some lower voltages are also represented as is the case with some 150kV branches in DE and NL and 110kV in NL and AT.

Besides these areas, five other are represented in a more simplified way, namely the United Kingdom, Greece, Sweden, Norway and (East) Denmark. Given that those areas have a limited interconnection capacity or are connected only by High Voltage Direct Current (HVDC) cables, their network is modelled through a single bus.

To build the benchmark system, the original UCTE 2008 network model was updated to account for the reinforcements and new branches reported on the 2014 Ten-Year Network Development Plan (TYNDP) [34], and each bus was georeferenced using the data available in [33].

I.3.2.b Load data

The original UCTE 2008 model only provides load data for a single operation point, to perform a multi-period simulation, load demand data were downloaded from European Network of Transmission System Operators for Electricity (ENTSO-E) transparency platform [35]. In particular, it was considered the 2013 hourly load profiles for each area, and load factors were applied to each bus, keeping the original load repartition. These values, as detailed in ENTSO-E transparency platform [35], corresponds to the sum of power generated by plants on both TSO/DSO networks, therefore corresponding to the total electrical load. The total energy values per country can be found in Table I.1.

TABLE I.1: Total load highest and peak for each country in 2013.

Country	Total load (TWh)	Highest Peak (GW)
DE	590,8	91,8
AT	64,4	10
BE	86,6	13,4
FR	445,9	82,8
NL	103,7	16,4
PL	143,1	23,2
CZ	66,6	10,1
CH	61,9	9,8
IT	290,5	52,5
SL	12,3	1,9
SK	24,6	3,8
DK	30,5	5,6

I.3.2.c Generators data

The generator database was updated to a 2013 scenario using the commercial database *PLATTS* [36] containing some technical information required, namely its technology and maximum power output. The minimum power output values were not available and are therefore assumed to be zero. Table I.2 presents the installed capacity of the dispatchable generation per country.

TABLE I.2: Capacity of dispatchable generation in *GW*.

Country	Nuclear	Coal	Fuel	Gas	Lignite	Reservoir
DE	12	18.8	2.5	18.1	20.5	7.8
AT	0.8	0	0.4	3.3	4.3	8
BE	3.9	0.9	1.2	5.7	0	1.3
FR	63	7	6.9	4	0	18.5
NL	0.5	3.9	13.2	0	0	0
PL	0	15.3	0.4	0.3	9.7	1.7
CZ	3.8	0.8	0.0	0.1	4.9	1.7
CH	3.2	0	0.1	0.1	0	11.7
IT	0	7.2	11.6	47	0.2	15.2
SL	0.7	0.5	0	0	0.8	0.8
SK	1.9	0.4	0	0.5	0.5	1
DK	1.7	0	0.1	1	0	0

Table I.3 presents the installed capacity of the non-dispatchable generation per country, namely, Photovoltaic (PV), Wind Power Onshore (WP On), Wind Power Offshore (WP Off), Combined Heat and Power (CHP) and Hydro Power Run-of-River

(HP RoR). Since these technologies are not optimisation variables as their production is imposed, annual profiles for each country and non-dispatchable technology were collected from [35]. Nevertheless, the modelling proposed in this work considers that all these technologies can be re-dispatched downwards as a last resort to solve network constraints.

TABLE I.3: Capacity of non-dispatchable generation in *GW*.

Country	PV	WP On	WP Off	CHP	HP RoR
DE	32	31	0.1	17.5	2.3
AT	0.4	1.3	0	1.7	4.3
BE	2.7	1	0.4	1.3	0.1
FR	3.6	8	0.1	3.2	6.1
NL	0.4	2.3	0.2	4.2	0
PL	0	2.3	0	5.9	0.2
CZ	2	0.3	0	5.2	0.2
CH	0.2	0.1	0	0.2	1.6
IT	15.2	6.7	0	4.1	2.3
SL	0.2	0	0	0.2	0.3
SK	0.5	0	0	1	1.4
DK	0.3	2.6	0.4	0.7	0

I.3.2.d Definition of variable costs

For the complete characterisation of the generation portfolio, and to capture as accurately as possible cross-border flows, it is key to have realistic figures for the variable production costs.

As described in [37], a generation unit is characterized by its fixed and variable costs. The fixed costs correspond roughly to the investment and maintenance expenditures of the generation unit and do not vary with the power output. Variable costs correspond to the expenditures needed to run it, that vary directly with the power output (i.e. primary fuel, CO_2 emissions and operating costs) [38].

Variable costs of generation units are key parameters in prospective economic studies of power systems, likely to motivate evolutions in economic policy, for the massive integration of RES [39] for example, or for updating the delineation of the bidding zones in Europe [40]. However, these costs are difficult to characterize, as they depend on multiple time-varying factors such as fuel cost (including transportation), CO_2 emissions' cost, units' efficiency, dynamic constraints, etc. Furthermore, generation units' variable

costs are private, commercially sensitive information that is unlikely to be disclosed in extent.

To avoid high complexity, variable costs are generally defined as assumptions in economic studies. Two main approaches are used to set their values, namely i) based on experts' advice [41], or ii) chosen from public databases, such as [42], and it usually leads to a cost per technology and per country. For prospective economic studies, this kind of information can be sufficient, but for studies that consider the power flows between different regions [43], it is important to differentiate the variable costs with a finer granularity.

Some reports detail an estimation of the costs of generating electricity [41],[44]–[45] per technology. But all of these works calculate the Levelised Cost Of Electricity (LCOE), which considers both variable and fixed costs over the lifetime of the generation units [44]. This indicator is useful as it allows to compare the costs of different technologies on the long run, but for OPF computations it is generally considered that producers present short term costs to the market. In an optimal power flow study, the generation units' costs are defined to mimic the producers bid in a market environment.

In this work, for simplicity purposes, and with the goal of miming the producers bid in a market environment, each generator is therefore characterized by its marginal production cost. A statistical analysis of the main European electricity markets was performed to estimate realistic marginal costs per technology and per country. The analysis consisted in comparing the electricity hourly spot market prices with the actual generation per production type (e.g. nuclear, gas, coal).

The increase in transparency forces market parties to release publicly available information about market prices, generation units' production and availability at a local and unit scale that can be useful to estimate the generation variable costs.

The approach to estimate the generation units' variable costs by technology and at a national, regional and unit scale is described in Appendix A.

The corresponding values which served as an input to the benchmark system used throughout this work, are presented in Table I.4.

TABLE I.4: Parameters for marginal costs of generation units in €/MWh.

Country	Gas	Coal	Nuclear	Source
DE	45	20	10	EEX
AT	49	30	-	EEX
BE	50	35	10	EEX
FR	40	35	10	RTE
NL	40	30	-	EEX
PL	45	40	10	ENTSOE
CZ	40	25	10	EEX
CH	45	40	10	-
IT	45	40	-	-
SL	45	40	10	-
SK	45	40	10	-
DK	45	40	10	-

I.3.3 Training and evaluation samples

Given this system complexity and also the complexity of the optimization tools as described in Section I.1, some temporal simplifications were made to be able to represent the networks' operating conditions. Using the benchmark system, each OPF is an optimization problem with 2689 variables and 9660 constraints that if computed using DigSilent's Power Factory 15.2 [46] on an Intel Xeon at 2.40 GHz using an one hour time step, the computation time for each OPF simulation is in average 3 minutes.

Ideally, one would perform the UC computation taking all this data into account. Nevertheless, with this level of complexity, a simplification needed to be performed despite some inaccuracies it might introduce in the result. Geographic characterisation was favoured in detriment of the inter-temporal constraints, and therefore, the unit commitment problems were modelled as a series of independent OPF computations.

Namely, the proposed approaches will use a set of 300 training operating situations (S_T) and a set of 300 evaluation situations (S_E), that will be used to calibrate the proposed methodologies and to assess the quality of the reduction, respectively. Each scenario corresponds to a different operating historical set-point corresponding to different levels of load, conventional generation and renewable power generation (maintaining the cross relation between all those variables).

To ensure the representativeness of the system's operating conditions through various situations, a monotone of the net load (load - RES) was calculated and the evaluation

and training data set were picked uniformly through that curve.

Chapter II

How to define clusters?

* * *

This chapter is organized as follows. Section II.1 presents the approaches proposed in the literature and details three of the most used techniques, Section II.2 presents the proposed metric based on the redispatch effort to compare the methodologies, Section II.3 applies them to the test case and the proposed metric is used to rank them and finally, Section II.4 draws the principal findings.

Contents of chapter II

II.1 Reducing spatial complexity

In this chapter, a focus on the aggregation of network models for prospective economic studies will be performed, specifically on aggregating buses without network congestions in between them, and leaving congestions at the clusters' frontiers.

Several academic works have addressed the problem of defining appropriate bus clusters according to network congestions. Some did not use any economic information, and buses were grouped regarding their impact on the flows on congested branches [47]. While others applied clustering algorithms to 8760 point time series of Locational Marginal Price (LMP) to define new clusters (e.g. [48] and [49]), or even used Monte-Carlo's approach to generate the LMPs time-series before clustering [50].

The European project eHighway 2050 [43] proposed, for instance, a methodology to determine network clusters that does not implicitly consider any network technical information. An improved methodology was latter presented [51] where the network technical characteristics were implicitly considered, namely with the calculation of electrical distances similar to the work presented in [30]. Other than this, different bus selection methods were also considered based on congestion profiles [52] or power transfer distribution factors [53].

In this work, it was chosen to focus on the three most used clustering techniques proposed in the literature, detailed hereafter, given the goal of identifying congestions to define the clusters' frontiers, using mostly LMP as a distance metrics in detriment of Power Transfer Distribution Factor (PTDF) or electrical distances. For simplification purposes, the approaches proposed by each author will be named after the clustering methodology used.

The approaches differ themselves not only by the clustering methodology used but also by the metric used to measure the distance between network buses, it is therefore hard to compare them and assess which one suits best the purpose of this work. Therefore, a framework to assess the quality of the aggregation is proposed that can compare the approaches regardless of the methodology and metric used.

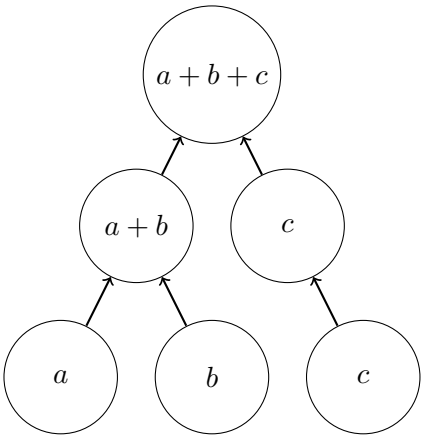

FIGURE II.1: Graphical representation of hierarchical clustering.

II.1.1 Hierarchical approach

Considering the approach developed by Burstedde [48], using a connectivity-based clustering algorithm that performs a bottom-up approach: each observation starts as an isolated cluster, and at each step, two clusters are aggregated until all the observations N belong to the same cluster. The aggregation A is defined considering the local best scenario (minimum distance) at each stage, and only electrically connected buses can be aggregated. A graphical representation of this algorithm is depicted in Figure II.1.

$$A = min(dist_{a,b}) \;\; \forall a \in N, \; \forall b \in N \tag{II.1}$$

The sum over all the scenarios of the squared Euclidean distances is used to determine the distance between pairs of observations, containing the LMPs of each bus for S_T scenarios s as:

$$dist_{a,b} = \sum_{s}^{S_T} ||LMP_{a,s} - LMP_{b,s}||^2 \tag{II.2}$$

II.1.2 K-means clustering

Considering the approach developed by Imran et Bialek [54], who applied the k-means algorithm using LMPs as distance metrics to define the clusters, the k-means algorithm is used to divide P points ($P = p_1, p_2, p_3, ..., p_{S_T}$, where P can be an array of observations which can be weighted by a factor γ_n) into $\bar{N}$ clusters, by minimizing the sum of the squared Euclidean distance between each point $p_n \in P$ and a cluster center (centroid)

$c\bar{N}$. The ex-ante definition of the total number of clusters is an important feature of k-means and is further discussed in Section II.3.

$$min(\sum_{\bar{n}}^{\bar{N}} \sum_{p \in cK_i} ||p - cK_{\bar{n}}||^2) \ , \tag{II.3}$$

where $p = \gamma_1 \times p_1 + \gamma_2 \times p_2 + \cdots + \gamma_{S_T} \times p_{S_T}$.

To determine the clusters and its centroids, Lloyd's algorithm [55] was used with the following steps:

1. Randomly initialize the $\bar{N}$ centroids $c\bar{N}$ from within the dataset P;
2. Calculate the sum of the squared Euclidean distance from P to each centroid $c\bar{N}$;
3. Attribute each point to the closest centroid;
4. Update the centroids as the average of all the points within the centroid (the newly defined centroid does not need to be contained in the dataset);
5. Repeat points 2) to 4) until convergence (no changes in the centroids from one iteration to the next one) is reached.

A P set of observations was used as an input, containing the LMPs of each bus for S_T different scenarios and their latitude and longitude coordinates. Given its top down approach, it is important the algorithm has some form of bus connectivity information, the geographical coordinates needed to be added as an extra dimension so that buses with similar LMPs, but geographically distant in the network, were not aggregated in the same cluster. Given the different magnitude between LMPs and geographic coordinates, all the entry data must be normalized, this was performed using Frobenius norm.

Given its random initialisation, it is possible that some results represent only a local optimum. To avoid that, the clustering algorithm was run several times for each $\bar{N}$, and only the best result was saved.

II.1.3 K-medoids clustering

The k-medoids algorithm is presented in [51] as a clustering approach using a combination of 50% electrical distance and 50% geographical distance to measure the distance between network buses. Given the difficulty in calculating the electrical distance, a first step in the network aggregation was performed were all the 225kV buses were assigned to the geographically closest 400kV bus.

This methodology only differs from the k-means approach in the definition of the centroid. In k-medoids, the centroid is the observation with the lowest average distance to the algebraic center, as opposed to the k-means where the centroid is the algebraic center. In their modifications to the method, the authors propose to define the medoids at each step, based on their connectivity (number of electrical connections), being the most connected bus selected as the new medoid.

II.2 Metrics for quality assessment and limits of the proposed modelling

A successful aggregation of the network buses will allow to have similar results, regarding the generators dispatch, when performing an Optimal Power Flow (OPF) using both the reduced and complete model. Each cluster is treated as a copper plate, with no network constraints applied to the internal exchanges and therefore should not contain any structural network constraint at its interior.

To assess the quality of the aggregation a metric is proposed, aiming to provide an order of magnitude of the redispatching actions, i.e. changing the schedule of operational units such as power plants, when the dispatch resulting from the optimization with a given bus aggregation definition is likely to create congestions.

It is thus defined as a quantitative assessment of a difference between two system states:

- Optimal dispatch resulting from a centralized optimisation of the whole region considering, in particular, all network constraints, described in Section II.2.1;

- Dispatch supposed to result from the optimization with the bus aggregation definition under consideration and neglecting the network constraints, as described in Section II.2.2.

Even though the limitations in the approach, as described in Section I.3.3 (such as the DC approximation), the assessment framework described in the following sections can be applied to different systems regardless of the approximations or tools used to perform the simulations. Therefore, the general formulations are presented in the next subsections.

II.2.1 Optimal dispatch simulation

First, the optimal dispatch that would be obtained with the full model is computed.

In practice, the optimal dispatch can be obtained from a transmission-constrained unit commitment, as described in [56], written as the following optimization problem.

$$\min f(u_s, x_s) \; \forall s \in S_T \tag{II.4}$$

such that:

$$g(u_s, x_s) = 0 \tag{II.5}$$

$$h(u_s, x_s) \leq 0 \tag{II.6}$$

where u_s represents the control variables (e.g. status and output of each power plant for each scenario s), x_s the state variables (e.g. power flows at each branch for each scenario s), $g(u_s, x_s)$ and $h(u_s, x_s)$ equality and inequality constraints (e.g. ramp limits and minimum up and down times), respectively, and $f(u_s, x_s)$ the cost function of the whole system, reflecting generally the overall surplus.

Let us denote u^* and x^* the control and state variables, respectively, corresponding to the optimal solution of Problem (II.4)-(II.6).

II.2.2 Clustered dispatch simulation

Second, the dispatch that would be obtained with the reduced model (using the clustering method under consideration) is computed.

For a given bus aggregation definition with $\bar{N}$ clusters, x and u can be formally organized cluster by cluster as follows.

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_{\bar{N}} \end{bmatrix}, \; u = \begin{bmatrix} u_1 \\ \vdots \\ u_{\bar{N}} \end{bmatrix} \tag{II.7}$$

Similarly, $f(u,x)$, $g(u,x)$ and $h(u,x)$ can be decomposed in $f_1(u,x)$,...,$f_{\bar{N}}(u,x)$, $g_1(u,x)$,... ,$g_{\bar{N}}(u,x)$, $h_1(u,x)$,...,$h_{\bar{N}}(u,x)$ according to the location of the assets considered in the functions.

Furthermore, one can assess a set of functions $p_{\bar{n}}(u,x)$ that determine the net position of every cluster $\bar{n} \in [1,\ldots,\bar{N}]$. It corresponds in practice to the sum of power flows on branches connecting each cluster with its neighbours.

To determine the dispatch resulting from the bus aggregation definition under consideration, an internal dispatch within every cluster $\bar{n} \in [1,\ldots,\bar{N}]$ is simulated, subject to a new equality constraint on the net position of this cluster. This constraint makes $f_{\bar{n}}$ and $g_{\bar{n}}$ sensitive to $u_{\bar{n}}$ and $x_{\bar{n}}$ only, such that the optimisation problem of every cluster $\bar{n}$ can be written as follows.

$$\min f_{\bar{n}}(u_{\bar{n}}, x_{\bar{n}}) \tag{II.8}$$

such that

$$g_{\bar{n}}(u_{\bar{n}}, x_{\bar{n}}) = 0 \tag{II.9}$$

$$h_{\bar{n}}(u_{\bar{n}}, x_{\bar{n}}) \leq 0 \tag{II.10}$$

$$p_{\bar{n}}(u_{\bar{n}}, x_{\bar{n}}) = p_{\bar{n}}(u^*_{\bar{n}}, x^*_{\bar{n}}) \tag{II.11}$$

Network-related constraints in $h(u_s, x_s)$ are not accounted for at this stage, as the clustered dispatch does not consider transmission network constraints.

The set of optimal settings for control variables of each sub-problem is denoted $\hat{u}$, such that:

$$\hat{u} = \begin{bmatrix} \hat{u}_1 \\ \vdots \\ \hat{u}_{\bar{N}} \end{bmatrix} \tag{II.12}$$

The goal of performing this clustered dispatch simulation is to assess the dispatch within a cluster without the network constraints, and compare it with dispatch obtained in Section II.2.1 (considering all the network constraints). In the clustered model, one knows the total load and available generation inside each cluster, but exchanges between clusters are yet to be determined and should be an output of the optimisation problem. At this stage, no exchange capacity between clusters is defined (links between clusters are not defined yet), making it impossible for the optimisation problem to determine the exchanges for each cluster.

In order to overcome this limitation and make computation more straightforward, it is defined that each cluster should export/import the same amount as in the full optimal dispatch simulation described in Section II.2.1. This assumption might be considered optimistic, as in the clustered model the net position can significantly vary due to the lack of internal network constraints, and it was therefore tested to assess its impact in the final results. Tt was observed that it may have an impact on the magnitude of the index but no changes were observed in the rank of the tested solutions. This is further illustrated in Section II.3.3.

II.2.3 Index assessment

The redispatch effort index RE is based on the difference between u^* and $\hat{u}$, corresponding in practice to changes in the dispatch level of each unit, for every scenario s. For a given scenario $s \in S_E$ (corresponding to different historical set-points as defined in

Section I.3.3), it is assessed as the ratio between the volume of redispatch to satisfy the new constraints and the total amount of energy dispatched.

$$RE = \frac{\sum_{g=1}^{N_g} ||\hat{P_{gen_g}} - {P_{gen_g}}^*||}{\sum_{g=1}^{N_g} {P_{gen_g}}^*} \tag{II.13}$$

where N_g corresponds to the total number of generators, and P_{gen_g} is the power generated by generator g.

As the impact on the units' redispatch depends on the system configuration at scenario s, three different metrics are considered to summarize the variation of the RE over the S_E scenarios, namely:

- Value at Risk (VaR) 95%, $I_1(S_E)$: $VaR_{95\%}(RE) = x \in \mathbb{R} : F_{RE}(x) \geq 95\%$ where $F_{RE}(x)$ is the cumulative density function of the Redispatch Effort
- the average quadratic distance $\forall s$, $I_2(S_E)$: $\frac{\sum_{s=1}^{S_T} RE_s^2}{S_E}$
- the maximum distance $\forall s$, $I_3(S_E)$: $maximum(RE_s \ \forall \ s \in S_E)$

II.3 Application to the benchmark system

The clustering approaches using the different distance metrics, as presented in Section II.1, were used to aggregate the buses of the benchmark system presented in Section I.3. This aggregation was performed regardless of the actual borders of each market zone or country. The three approaches were assessed using the index presented in Section II.2 as a function of the number of clusters.

First, in Section II.3.1 the assessment was performed for a single period, and then, in Section II.3.2 using 300 different evaluation scenarios (S_E), as defined in Section I.3.3.

II.3.1 Single period analysis - corresponding to a winter peak: 16 January 2013 at 18:00

The three approaches were applied and assessed for a single period corresponding to a winter peak hour. This corresponds to the best case scenario for the clustering methodologies, as the algorithm has to deal with only one load and generation pattern. Nevertheless, defining a reduced model based on a single point is far from being optimal, as real power systems' operating conditions are constantly varying.

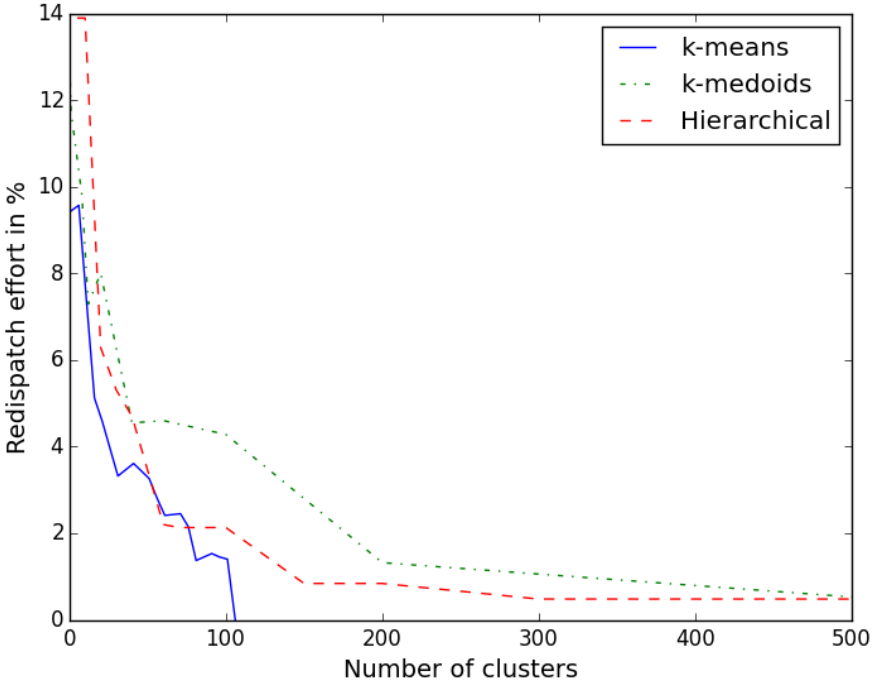

FIGURE II.2: Redispatch effort as a function of the number of clusters for three bus clustering methodologies. Results for a single period at the winter peak: 16 January 2013 at 18:00.

Figure II.2 depicts the redispatch effort obtained as a function of the number of clusters with each clustering technique. All the methods converge to an almost zero redispatch effort, as soon as the number of clusters exceeds a few hundreds and they all demonstrate the same decreasing trend. K-means reaches the zero redispatch effort around 100 clusters (18 generators per cluster in average), hierarchical reaches a redispatch effort below 2% around 100 clusters while k-medoids reaches that threshold around 180 clusters.

II.3.2 Multi-period analysis

The three approaches were used to define network clusters using 300 training scenarios S_T, as defined in Section I.3.3, that correspond to different levels of load and renewable power generation. This directly impacts the ones using LMPs as metrics, in this case k-means and hierarchical, as those are likely to change at each scenario.

To assess the quality of the aggregation, namely its robustness regarding the system operating set-points, the three approaches are assessed using 300 evaluation scenarios S_E. For the approach where the metrics takes only into account the electrical distance, in this case k-medoids, even though the input does not change at each scenario, when varying the load and generation levels different congestions can be formed within the defined clusters and therefore impact the RE.

The results in terms of VaR (I_1), expectation (I_2) and maximum redispatch effort (I_3) are illustrated in Figures II.3, II.4 and II.5 respectively.

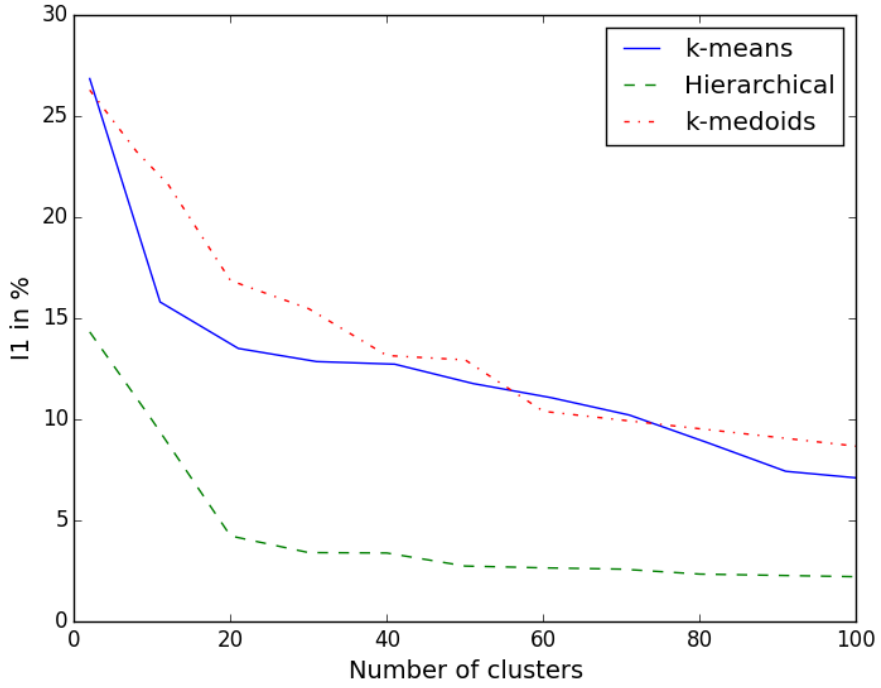

FIGURE II.3: Redispatch effort's VaR with 95% confidence, as a function of the number of clusters, for 300 scenarios corresponding to different levels of net load.

Results tend to show that for a multi-scenario approach, all the studied strategies present the same trend. However, the hierarchical approach tends to perform better overall. It can also be observed that the performance of the three methods tends to stabilize with

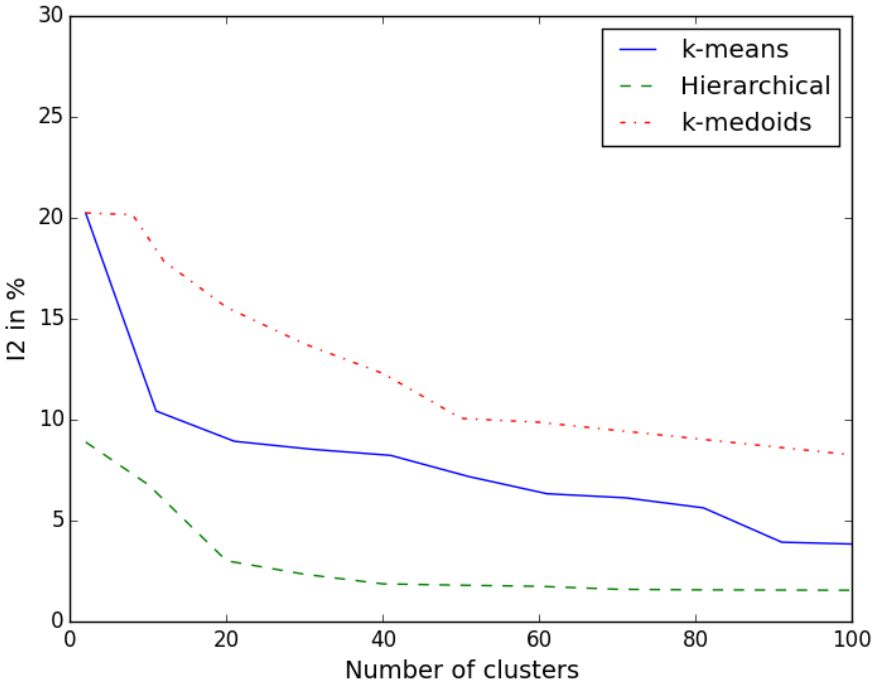

FIGURE II.4: Average redispatch effort, as a function of the number of clusters, for 300 scenarios corresponding to different levels of net load.

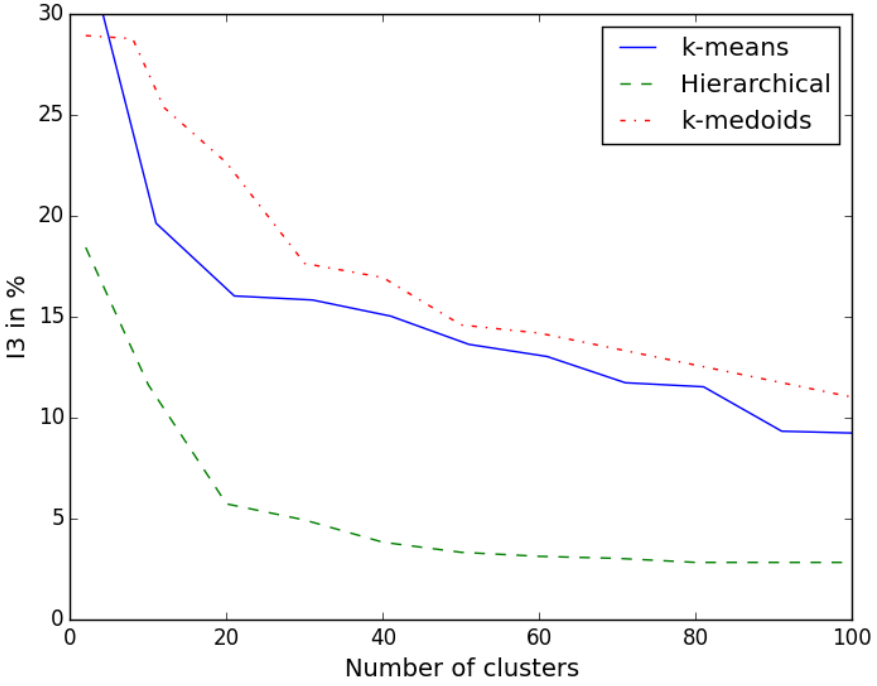

FIGURE II.5: Maximum redispatch effort, as a function of the number of clusters, for 300 scenarios corresponding to different levels of net load.

more than 20 clusters, making the trade-off between performance and number of clusters more challenging.

Even though the performance of the k-medoids and k-means approaches using their particular distance metrics are similar, within its implementation, the first one requires a deeper knowledge of the studied network (definition of critical branches) and a more intensive computation effort (calculation of the electrical distance using the elements of the inverse of the admittance matrix) when comparing with k-means approach and its chosen metrics (only needs to calculate the LMP). Also, for the k-medoids approach, to reduce the complexity of the problem (number of buses to select), an *ex-ante* reduction of the 225kV buses to the most geographically close 400kV bus was performed, as suggested in [51]. This step is likely to cause slight deviations in the final results, but as demonstrated also in [30], computing the electrical distance for all the voltage levels poses problems for the computation tractability and *ex-ante* operations to reduce the complexity are always required.

The hierarchical approach, that presents lower redispatch effort, also demands a strong computational effort, as its bottom-up approach forces to evaluate, at each step, all the possible connections between the different clusters. The time complexity of most hierarchical clustering algorithms is quadratic $O(n^2)$ which opposes to the linearity of k-means and k-medoids $O(n)$ [57].

II.3.3 Assessment of the net position constraint

In Section II.2.2, it was considered that the net position of each cluster would be the same in the clustered dispatch as it was with the optimal dispatch model. This was made to overcome the limitation imposed by the lack of characterization of the connections between clusters at this stage. In a clustered dispatch where all the connections are well-defined, the lack of internal constraints could boost the exchanges with the neighbouring countries and this could affect the results of the proposed assessment framework.

To assess its impact, a simulation was performed to demonstrate the sensitivity of the results to a relaxation of this constraint and its possible impacts on the methodologies rank.

A zonal dispatch can allow higher exchanges than the nodal dispatch given the relaxation of the network constraints internal to the cluster. In the developed approach the net position of each cluster was scaled up, by a factor of ϕ, according to their margins

between optimal exchange and maximum import (all dispatchable generation offline) for importing clusters, and maximum export (dispatchable generation at full power) for exporting clusters.

Figure II.6 depicts the impact of increased exchanges between the clusters, namely for $\phi = 0\%$ (optimal net positions), $\phi = 10\%$, $\phi = 30\%$ and $\phi = 50\%$. It shows that moving away from optimal exchanges significantly increases the volume of the redispatch to satisfy the network constraints. The more clusters under consideration, the higher the difference in magnitude of the index.

This result emphasizes that the assumption on the overall export position of each cluster may impact the absolute value of the indicator but does not affect the classification of clustering techniques for every number of clusters. This tends to confirm that the proposed index provides an insightful measure of the difference in redispatching effort deriving from the different clustering methodologies.

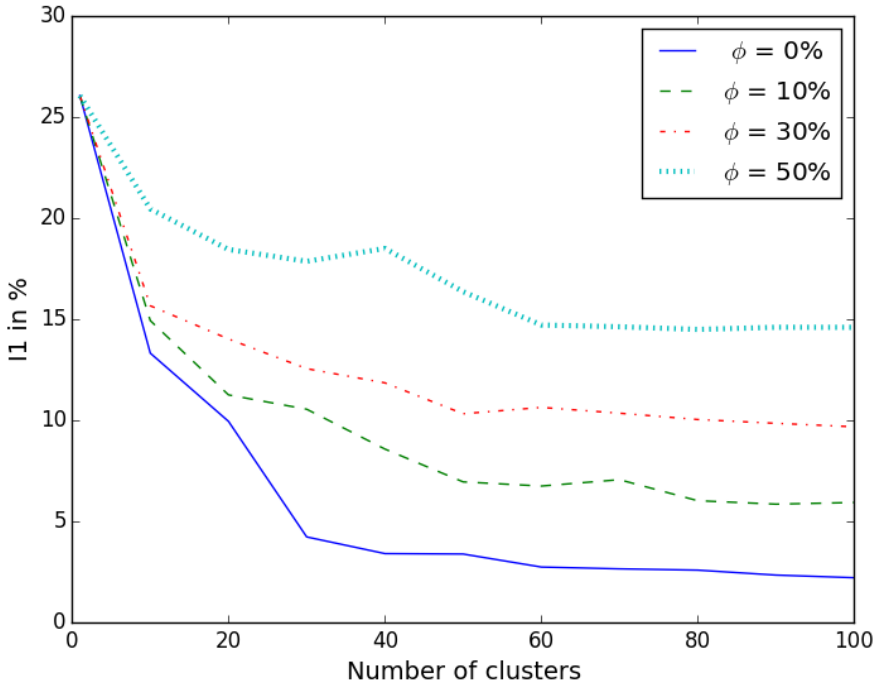

FIGURE II.6: Redispatch effort's VaR with 95% confidence using Hierarchical clustering, as a function of the exchanges' augmentation between clusters, for 300 scenarios corresponding to different levels of net load.

II.4 Key findings

This chapter discusses the techniques to perform network buses aggregation. A literature review was performed, and three main clustering techniques were identified to reduce power networks based on economic parameters, namely k-means, hierarchical and k-medoids.

To determine the most suitable technique to apply, an evaluation approach based on a new index is proposed. The index assesses the information loss when performing the aggregation by quantifying the difference between:

- the dispatch resulting from the complete model (equivalent to a nodal pricing dispatch);
- the dispatch resulting from the reduced model (equivalent to zonal dispatch, where each cluster represents a zone from the full model), considering a constant net position for each cluster with respect to the full model.

Quantitative assessment based on the proposed index makes it possible to rank the clustering techniques, allowing to observe that:

- for a single period (winter peak), the k-means outperforms the hierarchical and k-medoids approaches;
- for multiple periods (300 different points covering a broad range of situations faced in 2013), the hierarchical approach proves to be more effective. At first sight, most of the benefits in terms of redispatch effort are captured with less than 20 clusters.

Given the aforementioned results, hierarchical clustering using LMPs and geographical coordinates as distance metrics was the approach used in the course of this work to perform the aggregation of power system's buses and allow the complexity reduction in its simulations.

Furthermore, the proposed index can also be of interest in the analysis of market bidding zones adequacy, as proposed in [58].

Chapter III

How to represent power flows?

* * *

This chapter is organized as follows. Section III.1 details the proposed approaches from the literature, and emphasises their main limitations, namely the robustness of the system's operating conditions and high computational effort. Section III.2 introduces the proposed methodology to overcome the literature's limitations and Section III.3 proposes some indicators to assess the quality of the reduction. Section III.4 presents the results of the proposed methodology when applied to the benchmark system and, finally, Section III.5 presents the key findings.

Contents of chapter III

III.1 Establishing a relationship between injected power and power flows

To define a reduced network model, once bus clusters are defined, it is key to define branches between them that reflect the flow repartition of the original network subject to Kirchhoff's laws.

Most techniques for branch definition rely on the physical characteristics of the original system, using the electrical distance as an input [30], whereas others rely on more empirical considerations. For example, Schwippe et al. propose in [59] a simplified model of the pan-European network, where in a first step the connections between the clusters are considered to have all the same impedance. This has a strong impact on the reduced network's simulated power flows and on the exchange capacities between the different bus clusters. To overcome this limitation the authors then present a network model based on a Power Transfer Distribution Factor (PTDF) matrix.

PTDF matrix of the full network is one of the most common input data for branch definition [53]. PTDFs describe a linear relationship between the flow repartition through the entire system and a change in power injection at a given bus. As an example, [29] and [53] develop equivalent network models based on the PTDF matrix of the system. These methods present however some limitations, namely, a significant computational effort to calculate the complete system PTDF matrix, and also a simplified model that is optimized for a single network operation point. These limitations can lead to significant errors when applied to different operating conditions.

The definition of a reduced network based on the PTDF matrix, where each cluster aggregates different generation units, is similar to the setting of a zonal PTDF matrix. A change in the generation pattern inside a cluster can influence the flow repartition to the neighbouring clusters, and therefore, impact the PTDF coefficients. This is similar to the definition of the Generation Shift Key (GSK) coefficients in zonal PTDF models [28]. GSKs, which changes over time, are used to translate any change in the net position of one cluster into a change of injections in the buses of that cluster [60].

Therefore, in order to improve its robustness to the variations of the operation set points, different scenarios must be considered when determining the PTDF matrix of the reduced model [28]. To address this problem, a methodology to build a reduced

network model robust to changes in the system's operating conditions is proposed, based on an optimization problem. The approach is computationally easy as it does not require calculating the full system PTDF matrix, and takes into account different system operating points, which improves the robustness regarding system's operation points and allows for the reduction of dynamic components of the network such as High Voltage Direct Current (HVDC) cables and Phase Shifting Transformers (PST).

III.2 Optimizing a PTDF matrix representing multiple scenarios

The proposed methodology consists in determining, through an optimization problem, a PTDF matrix (Ψ) of the reduced system and a set of loop flows (f^0), while minimizing the difference between observed (F) and estimated flows ($\bar{F}$) for each scenario s.

Loop flows are the flows between two clusters induced by the load/generation pattern within another cluster [61]. When aggregating generation inside the cluster, the information regarding its repartition through the aggregated buses is lost, which can impact the flow repartition between neighbour clusters. Figure III.1 illustrates how a cluster can have the same net position with a different distribution of the generation pattern at its interior. This will necessarily add some deviations to the results of the reduced model, therefore, to take all this into account, the variable f^0 was added to the optimization problem.

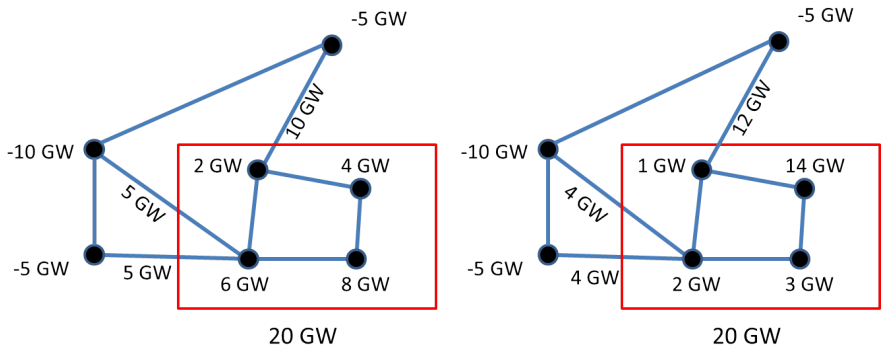

FIGURE III.1: Illustrative example of generation variation within a cluster.

This approach is inspired by the work presented in [43], but presents clear differences regarding the connection modelling where a reduced network PTDF matrix is computed instead of equivalent impedances.

For a given bus aggregation definition that converts a system with N buses and L branches into $\bar{N}$ clusters and $\bar{L}$ branches and considering a set of scenarios S_T, the optimization problem is formulated as follows:

$$min\{\Psi_r, f^0\} \ (F_{l,s} - \bar{F}_{\bar{l},s})^2 \tag{III.1}$$

s.t.

$$\forall s \in S_T, \forall \bar{l} \in \bar{L} \quad \bar{F}_{\bar{l},s} = \sum_{\bar{n}=1}^{\bar{N}} \Psi_{\bar{l},\bar{n}} \times P_{\bar{n},s}^{inj^r} + {f^0}_{\bar{l}} \tag{III.2}$$

$$\forall \bar{n} \in \bar{N}, \forall \bar{l} \in \bar{L} \quad |\Psi_{r\bar{l},\bar{n}}| \leq 1 \tag{III.3}$$

Where:

- Variables
  - Ψ_r is the PTDF matrix of dimension $\bar{L} \times \bar{N}$;
  - $\bar{F}_{\bar{l},s}$ is the estimated power flow in branch $\bar{l}$ for scenario s;
  - ${f^0}_{\bar{l}}$ is the power flow's estimated error in branch $\bar{l}$ due to the aggregation of generation, denominated loop flows.
- Parameters
  - $F_{l,s}$ is the observed flow in branch l for scenario s;
  - $P_{\bar{n},s}^{inj^r}$ is the power injected in bus $\bar{n}$ for scenario s;

For this optimization problem, the variable f^0 that represents the loop flows that might result from the changes in internal GSK is also a good indicator of the average power flow error for each branch.

III.3 Metrics for quality assessment and limits of the proposed modelling

As a new methodology is proposed for a relatively classical problem, there is a need to make sure that the approach is of interest for practical application. Being the goal of this approach to accurately represent the flows' repartition of the complete model, the following performance assessment process is proposed.

For a given reduction technique, which translates a system as depicted in Figure III.2 (left) with L branches and N buses into a reduced system as depicted in Figure III.2 (right) with $\bar{L}$ branches and $\bar{N}$ clusters, a comparison between the estimated flows $\bar{F}_{\bar{l},s}$ for each branch of the reduced system $\bar{l} \in \bar{L}$, and the observed flows $F_{l,s}$ in the corresponding branches $l \in L$ of the full system is proposed where s is a given scenario from the set of all considered evaluation scenarios S_E.

According to the example provided in Figure III.2, for a given scenario s, the observed flow between cluster a and b, F_{a,b_s}, is calculated as follows:

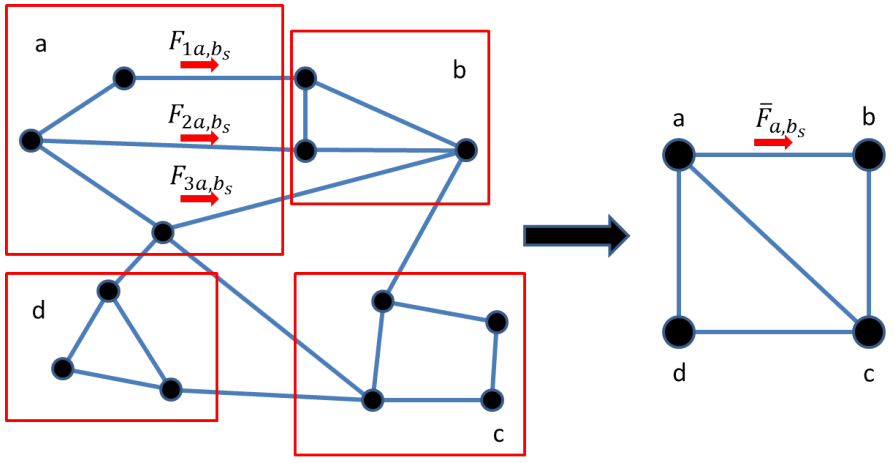

FIGURE III.2: Illustrative example of the performance assessment for a reduction approach.

$$F_{a,b_s} = F_{1_{a,b_s}} + F_{2_{a,b_s}} + F_{3_{a,b_s}} \tag{III.4}$$

The estimated power flows $\bar{F}_{\bar{l},s}$ of all the branches in the reduced system are calculated with the PTDF matrix issued from the optimization problem, and can now be compared with $F_{l,s}$ using the Root Mean Square Error (RMSE), the Normalized Root Mean Square

Error (NRMSE) and the absolute value of the difference between $\bar{F}_{\bar{l},s}$ and $F_{l,s}$, e^{abs} as defined hereafter:

$$RMSE_{\bar{l}} = \sqrt{\frac{\sum_{s=1}^{S_E}(F_{l,s} - \bar{F}_{\bar{l},s})^2}{S_E}} \tag{III.5}$$

$$NRMSE_{\bar{l}} = \frac{RMSE_{\bar{l}}}{avg\ \bar{l}} \times 100 \tag{III.6}$$

where $avg\ \bar{l}$ represents the average of the absolute values of the flows in branch $\bar{l}$.

$$e^{abs}_{\bar{l},s} = |F_{l,s} - \bar{F}_{\bar{l},s}| \tag{III.7}$$

III.4 Application to the benchmark system

Considering the cluster definition obtained with the methodology define in Section II.3, corresponding to 50 clusters and 91 branches and using the 300 training scenarios' power flows and injected powers, as defined in Section I.3.3, the optimisation problem formulated in Section III.2 was defined to calculate the reduced system's Ψ_r and f^0.

Once the model was defined on a training set, a evaluation set of 300 scenarios (S_E) was used to assess the quality of the model, noted as the superscript $(.)^{test}$. The new estimated flows $\bar{F}_{\bar{l},s}{}^{test}$ for each of the 300 scenarios s were calculated using Ψ_r:

$$\bar{F}^{test} = \Psi_r \times P_s^{inj^r} \tag{III.8}$$

Where $P_s^{inj^r}$ is an array containing the test sets' injected power at each cluster $\bar{n}$.

III.4.1 Assessing the performance of network reduction

Figure III.3 presents the NRMSE for the 91 branches of the system, over the considered 300 test scenarios. It can be observed that most of the branches present a NRMSE lower than 15%, but an important number of branches present a significant NRMSE, in some cases larger than 30%.

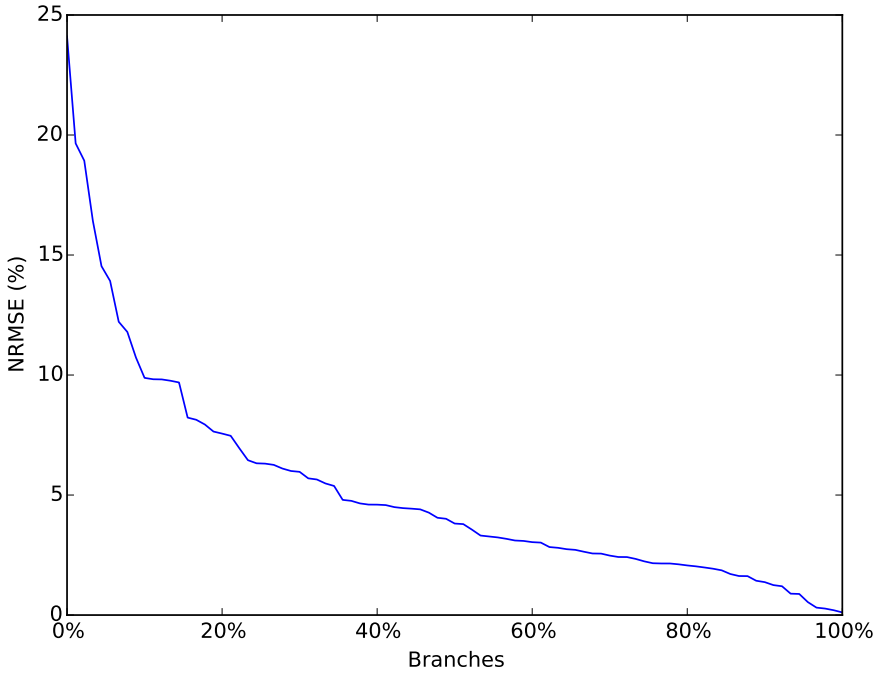

FIGURE III.3: NRMSE in % for all branches of the reduced system $\bar{L}$.

For illustration purposes, a focus on only the branches for which RMSE $\geq$ 100MW and NRMSE $\geq$ 5% is presented corresponding to a total of 15 branches out of the 91 which constitute the reduced system.

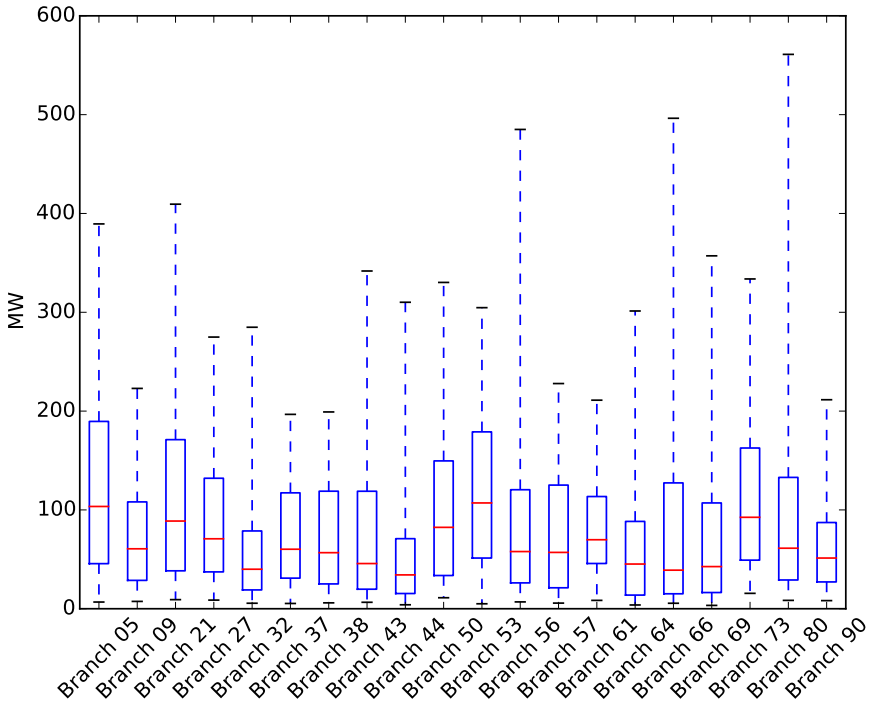

FIGURE III.4: Boxplot of the e^{abs} with whiskers from the 5^{th} to the 95^{th} percentile, optimizing the loop flows for all the scenarios. The red line represents the median.

Figure III.4 depicts the distribution of e^{abs} of each branch for all the scenarios. One can observe that, even though these branches correspond to the highest values of NRMSE, the median value of e^{abs} is in a large majority inferior to 100 MW. This low average flow explains the high NRMSE observed in some branches.

Also, it can be observed that some branches present a difference larger than 400 MW between the median e^{abs} and its 95^{th} percentile, which shows that for extreme scenarios, where the GSK within the cluster is severely changed, the estimated power flows might still present a significant error.

III.4.2 Modelling multi-period loop flows

As it was described in [43], loop flows can have an important impact in the estimation of the power flows between clusters. To try to mitigate this problem, the formulation of the problem in Equation III.2 was modified, so that f^0 could be optimized for each scenario s:

$$\forall s \in S_T, \forall \bar{l} \in \bar{L} \quad \bar{F}_{\bar{l},s} = \sum_{\bar{n}=1}^{\bar{N}} \Psi_{r\bar{l},\bar{n}} \times P_{\bar{n},s}^{inj^r} + {f^0}_{\bar{l},s} \tag{III.9}$$

Figure III.5 depicts the obtained results with this new formulation. Those highlight that for this particular case-study, the upgrade of the optimization function makes almost no difference when compared to the one in Equation III.2. Indeed, when setting the threshold for RMSE $\geq$ 100MW and NRMSE $\geq$ 5%, almost all the same branches appear in Figure III.4 and Figure III.5, except for Branch 61. And for almost all the branches, the values of e^{abs} and its 95^{th} percentile remain the same, with exception of some negligible variations for a reduced number of branches.

Given the results presented, for this specific test-case, there is no relevant interest in optimizing a f^0 for each scenario.

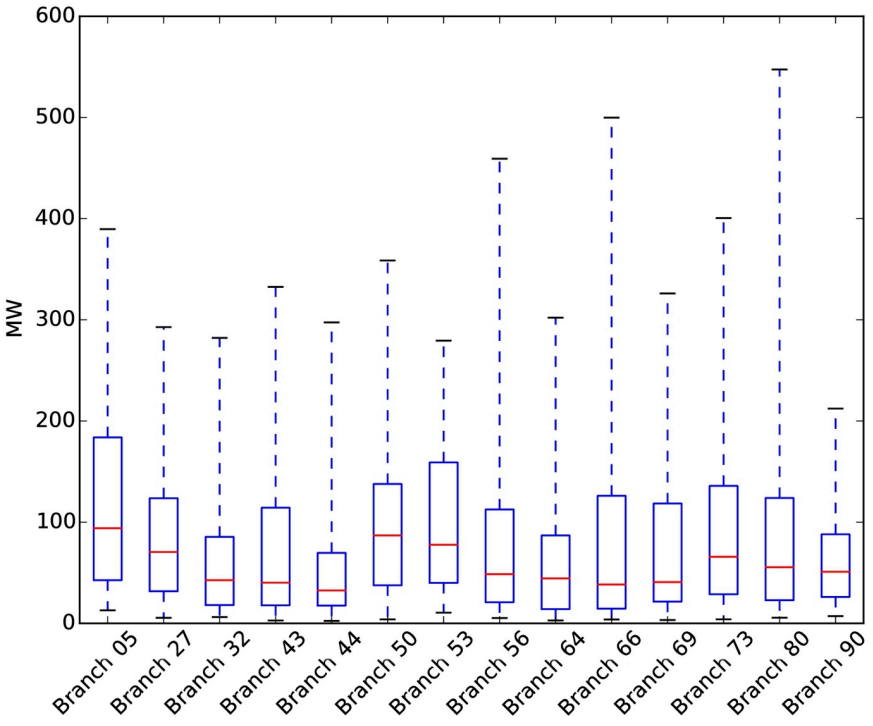

FIGURE III.5: Boxplot of the e^{abs} with whiskers from the 5^th to the 95^th percentile, optimizing the loop flows for each scenario. The red line represents the median.

III.4.3 The impact of varying the number of clusters

The results presented in the previous section were performed using a 50 cluster reduced model of the benchmark system. Given that, as detailed in Section III.2, cluster definition can have an important impact on the results of the proposed methodology, in this section five different reduced models, namely with 10, 50, 100, 150 and 200 clusters, were used to assess the results' sensitivity to the variation of the number of clusters.

Figure III.6 depicts the NRMSE for all the branches of the reduced systems. It can be observed that the most significant errors are obtained with the 10 clusters model and that a significant reduction occurs when using the 50 clusters model (already presented in Figure III.3). Aggregating all the generation units in 10 clusters can have a strong impact on loop flows and therefore affect the results' accuracy. A less significant reduction occurs when comparing the 50 and 100 cluster model, and the error variations between the 100, 150 and 200 clusters model are almost unnoticeable. It can also be observed that independently of the number of clusters formed, there are always a significant number of branches with NRMSE superior to 5%. Similarly to the results presented in

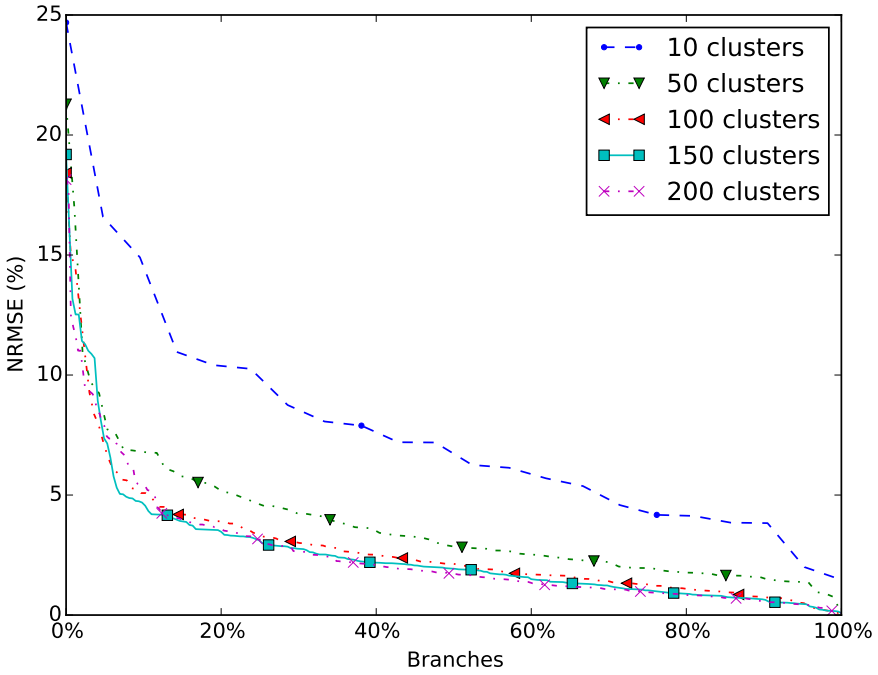

FIGURE III.6: NRMSE in % for all branches of different clusters definition.

Section III.4.1, this is due to the low average flows of those specific branches that tend to increase the value of NRMSE.

These results show the same trend observed in Section II.3, as reduced models with more clusters present lower errors and after 50 clusters there is no direct relationship between an increase in the model detail and accuracy improvement.

III.5 Key findings

The proposed methodology allows to characterize the connections between clusters and differentiates itself by its optimisation for different operation set-points, that improves its robustness regarding the system operation conditions. By applying it to a real test case it allows to identify challenges that are not easily addressed with simplified test systems.

To evaluate the accuracy of the proposed methodology, a performance assessment index is also defined that assesses the deviations on branches' power flows. This index can also be applied to assess and compare other reduction methodologies.

The accuracy of the proposed methodology was assessed using the benchmark system and results for different reduced models definitions were presented. The results show that a good representation of the full models' power flows can be obtained, but the number of clusters of the reduced model has a strong impact on its accuracy, specially for models with less than 50 clusters.

The presented results are encouraging, nevertheless, the proposed methodology would benefit from a comparison with different reduction methodologies, and with an exposure to a larger set of operating conditions. And it is also important to stress that all the results were obtained using the DC approximation, and therefore should be carefully interpreted, as with a full Alternating Current (AC) analysis results might differ.

Ideally, in a cluster definition that does not aggregate congested branches, the changes in dispatch within each cluster (for a given export) should only induce flows on its internal branches, which are not represented in the reduced model. The criteria used in this work to determine the clusters, as detailed in Section II.1.1, focuses on network congestions and not necessarily on loop flows, the standard deviation in the errors of some branches can be explained by this phenomenon. Therefore, the definition of clusters based on this criteria could be a topic of interest for further work.

Chapter IV

How to define branches' operational limits?

* * *

This chapter is organized as follows. Section IV.1 describes the problem of defining operational limits for reduced models' equivalent branches and presents the proposed approaches in the literature emphasizing their main limitations. Section IV.2 details the newly proposed methodology to estimate equivalent branches' transmission capacity. Section IV.3 presents the metrics to assess the quality of the estimation obtained in Section IV.4, while using the benchmark system. Finally, Section IV.6 presents the key findings and gives some insights for further works.

Contents of chapter IV

IV.1 From a load flow to an optimal power flow

Load flow is a classical problem in the power systems literature and its basic formulation is well known [62, 63]. Its calculation determines the static operating conditions of a system (i.e. power flows, voltages, angles, etc.), for which, the injected power at each bus is known, and therefore, no optimization is performed.

Optimal Power Flow (OPF) calculations typically aim at minimizing the system's cost by setting the production plan taking into account steady-state constraints. In a system where there are no energy transmission limits, the results would be trivial, as the OPF would dispatch each unit from the cheapest to the most expensive one, successively, while respecting its minimum and maximum production limits. Transmission branches' thermal rating limitations along with the branches' power flow repartition impose the constraints that limit the transmission capacity and therefore the economic optimization of the system, making it a key parameter to perform OPF calculations.

With the methodologies proposed in Section II and Section III, a reduced model of the transmission network can be obtained that actually represents the flow repartition and therefore allows to perform load flow studies. In this framework, a methodology to set maximum flow limits in the branches of the reduced network is developed.

Most of the reduced networks models proposed in the literature are intended to perform load flow studies, and therefore, the equivalent transmission capacity of the newly modelled branches between the new clusters is often neglected, for example [28, 25, 64, 65]. Other approaches are proposed, for example, by forcing branches connecting two clusters to remain separated avoiding the definition of equivalent thermal capacity [66]. This solution avoids the equivalent capacity calculation but, as a drawback, leads to a limited reduction in the number of branches, e.g. if two clusters have five branches connecting them, all of them must be kept. Moreover, it also neglects the effect that the clusters' internal branches can have on its exchange capacity.

Some works have proposed different methodologies to estimate the equivalent transmission capacities. For example, some simplistic approaches are proposed, such as the sum of the thermal capacities of the existing branches connecting the clusters [67], or the average of the historical flows of those branches [53]. These approaches give a gross estimation of steady-state constraints that can be misleading as the impact of Kirchhoff's

law and the limitations of the clusters' internal network can significantly impact the total transmission capacity available. The same problem may arise with approaches that estimate the Available Transfer Capability (ATC) between buses [68, 69], as the ATC is calculated as the difference between the existing flow and the branches' thermal capacity, the impact of adjacent branches is neglected. Figure IV.1 illustrates this problem, as the calculated ATC from bus a to b would be of 50MW, corresponding to the difference between branches' thermal limit (100MW) and the actual power flow (100MW), but this exchange is limited by the branch connecting bus a to c that is already at its limit.

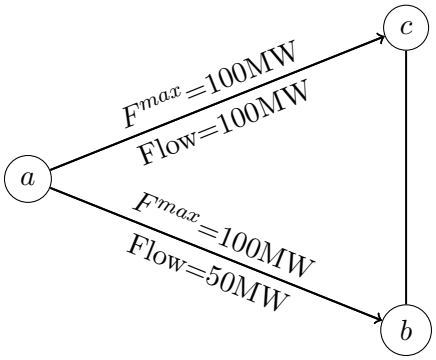

FIGURE IV.1: Graphical representation of the ATC-method.

More complex methodologies have been proposed [51, 70], where authors propose to perform specific simulations in the full network, by increasing the load and generation in the buses corresponding to the importing/exporting clusters, and in this way determine their maximum import/export capacity. Even though the latter are clear improvements to the more simplistic methods proposed in the literature, they still present some limitations, namely regarding the assumptions that need to be made when assessing the branches' thermal capacity by simulation. Even though the maximum increase in load-/generation can give a good estimation of the transmission capacity between clusters, it requires performing iterative simulations until the final values are reached. This can add an extra layer of complexity for large power systems with an important number of clusters.

Considering the reduced number of alternatives and the limitations of the existing ones, this chapter introduces a new methodology to estimate equivalent branches' transmission capacity that can accurately represent the steady-state constraints of the full network model. This methodology uses only the full system's historical operating set-points (no need to perform new simulations), while maintaining a robustness to changes in the

operating conditions and not needing assumptions regarding generation or load shift factors.

IV.2 Computing branches' equivalent capacities

Given a configuration of aggregated buses, as defined in Section III.4, connected by branches defined according to the methodology defined in Section III.2, the proposed methodology to calculate the branch's maximum capacity is detailed hereafter.

Considering a classical formulation of a transmission-constrained unit commitment, as described in Section II.2.1. At this stage, given the approach followed in Chapter III, the network is represented by a Power Transfer Distribution Factor (PTDF) matrix and therefore a slight adjustment regarding the formulation is needed.

In Section II.2.1 one of the constraints represented by Equation II.6, corresponds to the branches' thermal limit (${F^{max}}_l$). In this section, for a branch l, the branch's thermal limit can be defined as follows:

$$\|\Psi_{l,n_1} \times P^{inj}_{n_1} + \ldots + \Psi_{l,n_N} \times P^{inj}_{n_N}\| \leq {F^{max}}_l \tag{IV.1}$$

Where Ψ_{l,N_1} is the PTDF of bus N_1 on branch l and $P^{inj}_{N_1}$ is the injected power at bus N_1.

Considering a complete network model with L branches, there will be a total of $L \times 2$ constraints as defined in Equation IV.1. These constraints will define a domain of feasible solutions for the optimisation problem, regarding all steady-state constraints considered in the OPF. This domain is illustrated in Figure IV.2 for a two dimensional system. In a system with N buses, there will be N dimensions that will geographically define a polytope[1].

Considering that all the information from the full system needed to perform an OPF is known, it is possible to determine the domain using its PTDF matrix Ψ, that dictates the slope of the restriction, and the corresponding F^{max}, that gives the y-intercept point.

[1] A polytope is a generalisation in any number of dimensions of the three-dimensional polyhedron.

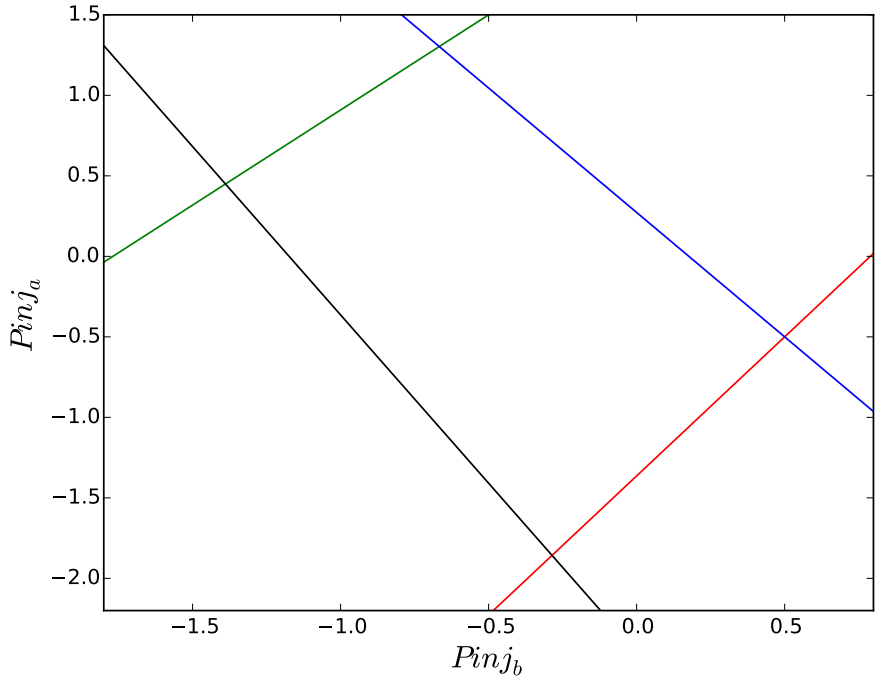

FIGURE IV.2: Illustration of a 2D optimisation domain between a zone a and b. P_{inj} represents the net position of each zone.

On the other hand, for the reduced system, the same cannot be done, as the F^{max} for each new aggregated branch is not known.

To have a reduced system that accurately represents the full model, this thesis proposes to define transmission capacities of the reduced model so that its domain of feasibility aligns with the range of solutions of the OPF on the full model, i.e. having a similar area/volume and being centred around the same coordinates. Based on this alignment principle it is proposed to set the steady-state constraints of the reduced system reflecting those considered in the OPF of the full system. Knowing the PTDF matrix of the reduced system (Ψ_r), the F^{max} will be determined to approximate as much as possible the reduced model's domain to the full one.

To this end, the interception with the axis of each plane must be adjusted to allow the same flow transfers as in the full model. Using the historical operating set points of the full system, represented as red dots in Figure IV.3, the restrictions that define the admissible domain are displaced to accommodate the same flows on the equivalent branches.

Ideally, the historical operating points of the European system would take into account the different assumptions taken by each Transmission System Operator (TSO) of each

area in order to optimize their own system. Given the lack of information regarding TSOs procedures, in this work, the historical operating points are approximated by OPF computations that minimizes generation costs, while respecting both generators and network limits and maximizes the utilisation of network capacity.

For a given scenario s, the injected power at each bus of the full network model (P_s^{inj}) was already obtained through an OPF calculation and is now converted to the reduced model ($P_s^{inj^r}$) as detailed in Equation IV.2, where Γ represents a matrix of size $N \times \bar{N}$ that establishes a correspondence between the original buses and the ones they were aggregated into.

$$P_s^{inj^r} = P_s^{inj} \times \Gamma \tag{IV.2}$$

Using a PTDF defined through an optimization problem that relates injected power and branches power flows from historical scenarios, as detailed in Section III.2, the expected flows in the reduced model can be calculated for each scenario. The estimated flows for the reduced system's branches are calculated as follows:

$$\Psi_r \times P_s^{inj^r} = \bar{F}_{\bar{l},s} \; \forall s \in S_T \tag{IV.3}$$

where Ψ_r is the PTDF matrix of the reduced system and $\bar{F}_{\bar{l},s}$ is the estimated flow on the equivalent branch $\bar{l}$ that represents the connection between two clusters.

This calculation allows to establish the maximum admissible flows in the equivalent branches, limited by the maximum admissible exchanges obtained with the original network. The maximum flow for the equivalent branch is equal to the maximum value found in the array $\bar{F}$ that contains all the results from Equation IV.3.

Algorithm 1 describes the iterative procedure to define each branches' maximum capacity based on the system's historical operating set-points, where $\Psi_{r,\bar{l}}$ represents column $\bar{l}$ of

the PTDF matrix Ψ_r.

for *every scenario* s **do**
  for *every branch* $\bar{l}$ **do**
    if $P_s^{inj^r} \times \Psi_{r,\bar{l}} \leq {F^{max}}_{\bar{l}}$ **then**
      next branch;
    else
      ${F^{max}}_{\bar{l}} = P_s^{inj^r} \times \Psi_{r,\bar{l}}$;
    end
  end
end

Algorithm 1: Defining F^{max} for each branch $\bar{l}$ of the reduced system.

Figure IV.3 illustrates the same process in a simplified two dimensional view. Considering that the dotted lines correspond to the initial domain of the reduced model, the red points are the historical operating points and the solid lines correspond to the domain after at the end of the algorithm. It can be observed that the constraints are adjusted so all operating points are inside the feasibility domain, and by doing that, a coherent set of constraints are being delineated for other operating points. In the case where the exchange between clusters was not limited in the full system, it will be constrained by the feasibility domain defined by the other branches.

The goal of a reduced system is to significantly diminish the complexity of the full system, therefore the approximation of the original domain is done with a reduced number of planes, it is expected that not every constraint has a perfect correspondence in the reduced domain. Therefore, in order to allow some flexibility to the approximation methodology, the value for the reduced model was chosen so that a given percentile η of exchanges observed with the real system is inside the feasibility domain. In practice, for $\eta = 95\%$ this means that the reduced model may not manage 5% of all the historical exchanges between clusters.

In Algorithm 1, that is equal to replace Equation IV.4:

$${F^{max}}_{\bar{l}} = P_s^{inj^r} \times \Psi_{r,\bar{l}} \tag{IV.4}$$

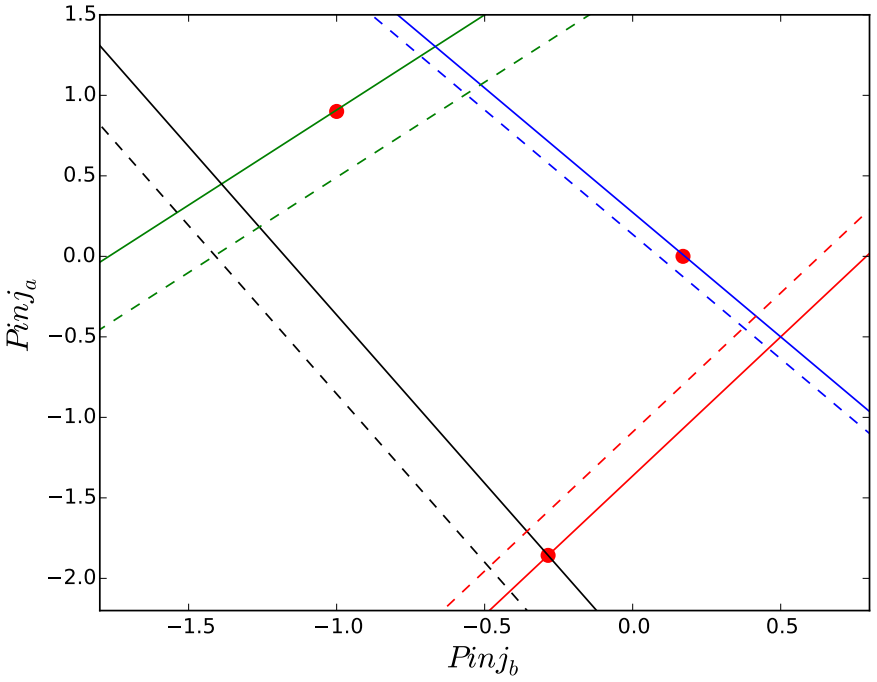

FIGURE IV.3: Representation of the methodology on a 2D optimisation domain between a zone a and b. The red dots represent historical operation set-points, the dotted lines the initial domain and the solid lines the domain after being adjusted to the historical operation set-points.

With the following expression that finds the percentile η of all the scenarios:

$$F^{max}{}_{\bar{l}} = percentile_{\eta}(P_s^{inj^r} \times \Psi_r \forall s \in S) \tag{IV.5}$$

And graphically it represents adjusting the constraints to fit with only η of the observations.

IV.3 Metrics for quality assessment and limits of the proposed modelling

A new methodology is proposed to determine the reduced model's branches' steady-state constraints, therefore an assessment of the newly calculated values is needed. Given that there is no reference to compare with obtained values, the assessment is done by comparing simulations' results obtained with the full and the reduced models.

Given that the results of the full model's OPF are already known for S_E scenarios, those were compared with the OPF's outputs when using the reduced model for the

same scenarios. The reduced model was subject to the same system conditions (i.e. the same load and generation level), and an OPF was performed with the newly calculated equivalent steady-state constraints. It is assumed that the difference between the reduced and full model OPF calculation is due to the variation of the steady-state constraints.

The results obtained with both models were compared regarding the system's total cost and the net export level of each cluster. Also, when estimating the equivalent transmission capacity for the reduced grid model, there is a risk that an insufficient capacity may result in Energy Not Supplied (ENS). To take this into account, a value of 3000€/MWh was set to reflect this variable when analysing only the system's cost and moreover, an assessment of the total ENS for all the calculated scenarios was performed. This is modelled as a fake generator in the OPF computation with the cost of 3000€/MWh.

IV.4 Application to the benchmark system

The proposed methodology was applied to the benchmark that had already been aggregated and reduced using the methodology presented in Chapter II and Chapter III, respectively. Once the equivalent steady-state constraints were calculated, the accuracy of the methodology was assessed using the criteria detailed in Section IV.3.

With this data, one can perform an OPF calculation using the reduced network model. Figure IV.4 depicts the absolute difference between the dispatch for each cluster for $\eta = 80\%$. Dispatch difference is defined as $\frac{|Dispatch_{reduced} - Dispatch_{full}|}{Dispatch_{full}} \times 100$. For illustration purposes, a focus on only the clusters whose average generation over all training scenarios is greater than 200MW is presented. One can observe that even though most of the clusters present low differences in average, there are some outliers with higher medians and the 95^{th} percentile over 50%. The most evident is cluster number 16, which presents a median above 20% and the 95^{th} percentile reaching almost 80%.

Also important as a result from an OPF is the system's total cost. It depends mainly on the cost of the producing generation units, and of the ENS penalty, as referred in Section IV.3. Figure IV.5 depicts the boxplot of the absolute difference of the system's total cost between the full and reduced model, for different values of η. Cost variation is defined as $\frac{|Cost_{reduced} - Cost_{full}|}{Cost_{full}} \times 100$.

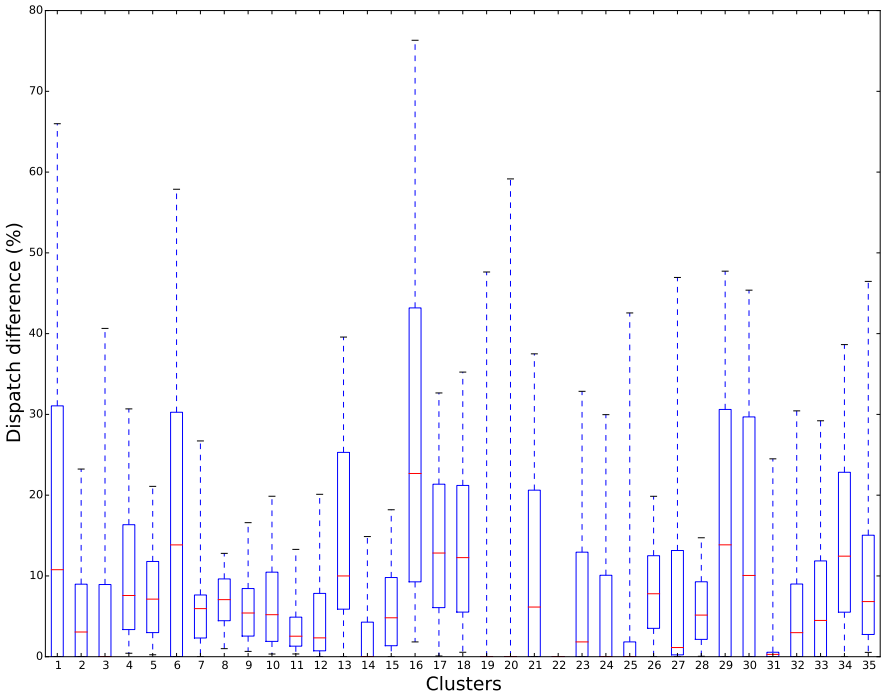

FIGURE IV.4: Boxplot of the absolute difference between the full and reduced model dispatch within each cluster with $\eta = 80\%$. The red line represents the median.

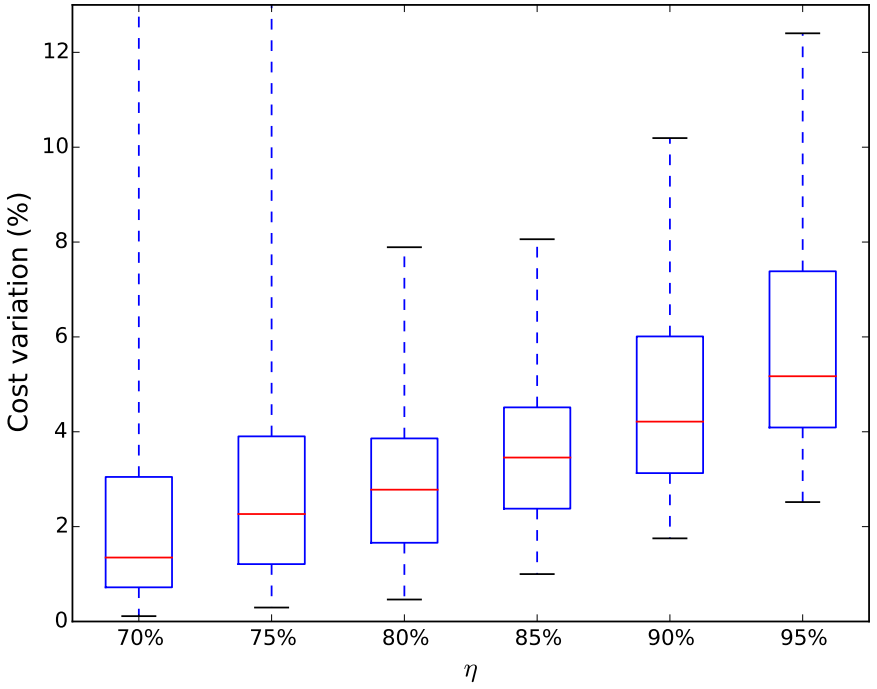

FIGURE IV.5: Boxplot of the absolute difference between the full and reduced new system cost for different values of η. The red line represents the median.

For visualization purposes the 95^{th} percentile of $\eta = 70\%$ and $\eta = 75\%$ do no appear in the figure but they correspond to 46% and 40% respectively. It can be observed that the highest median of cost variation is for η=95% and that it then decreases along with

the value of η selected. Even though the median continues to decrease for η=70% and η=75%, it presents a remarkable increase in the 95^{th} percentile, compared with the rest of the set. This shows that for these values, even though the median of the cost tends to reduce, there are a significant number of scenarios with a high value of ENS, equivalent to a non-convergence, that makes the system cost increase. On the other hand, relaxing too much the problem (higher values of η) leads to a higher media of the systems' cost difference.

Figure IV.6, depicts the monotone of the system's cost difference for both the complete and reduced model throughout the 300 considered scenarios. The cost difference is defined as $\frac{Cost_{reduced}-Cost_{full}}{Cost_{full}} \times 100$, where the subscript "reduced" refers to the OPF ran with the reduced network model and "full" the OPF ran with full network model. A positive difference means that the reduced model's cost is higher than the full model, and therefore the branches' steady-state constraints are too constrained, and the opposite (the difference is negative) means that the reduced model's cost is lower and therefore, the system is not enough constrained.

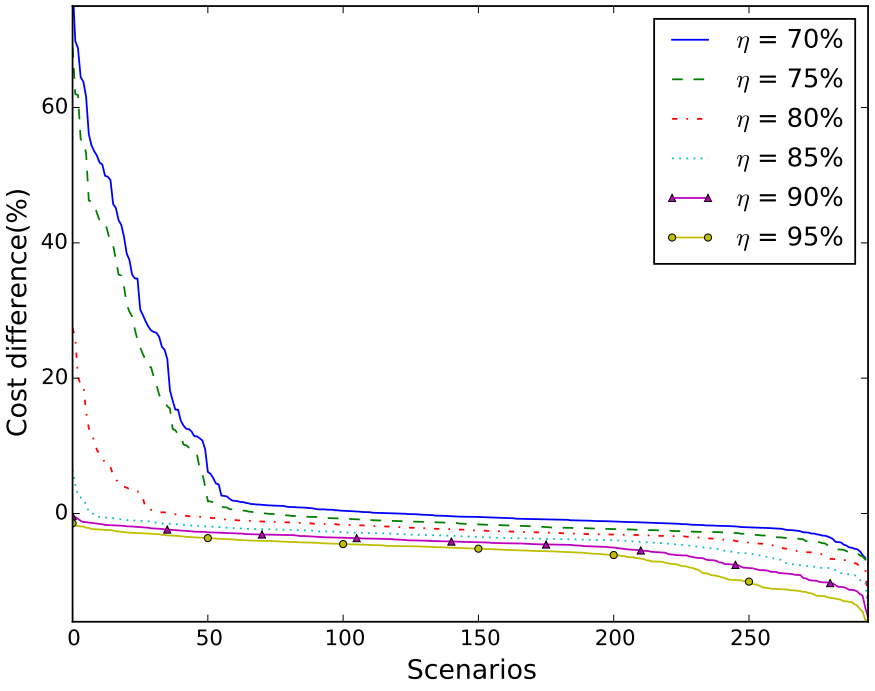

FIGURE IV.6: Monotone of the difference between the full and reduced system cost for different values of η.

For the first 50 scenarios, a remarkable increase can be observed in the reduced model's cost for a value of η inferior to 85%, that is likely to correspond to situations where the

system is too constrained and generation cannot meet demand, incurring in a penalty for the ENS, as detailed in Figure IV.7. It can also be noted that for higher values of η there are no noticeable ENS costs for the first 50 scenarios, but on the other hand, there is a significant deviation in the last 50 scenarios. In this case, as the difference is negative, it means that the reduced model should be more constrained, this puts into evidence the difficulty of defining the right value of η that is capable of minimizing the cost's deviations throughout all the scenarios.

Figure IV.7 confirms that, for values of η lower than 85% there is a significant increase in the ENS, and almost no ENS for the higher values of η. This value is extremely important to take into account when defining the right threshold to apply, as the average system cost can hide an important characteristic of the system.

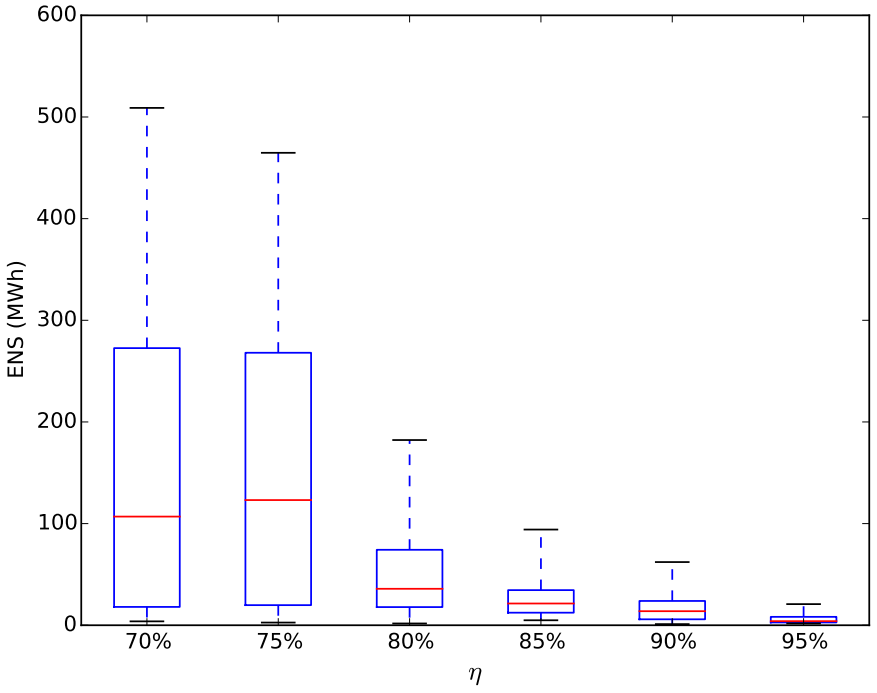

FIGURE IV.7: Energy not supplied for different values of η. The red line represents the median.

When analysing the different results, it can be concluded that even though the differences in the dispatch for each cluster are significant, they tend to fade when looking at the variation in the system's total cost. Therefore, it can be concluded that the variations in the clusters' dispatch are geographic (two close clusters have similar generation units) and not economic (no inversion in the merit-order) as system costs remain the same.

This is more related to the cluster or generation costs definition than with the equivalent branches' transmission capacity's estimation.

IV.5 The impact of varying the number of clusters

The results presented in the last section were obtained using the benchmark system reduced to a 50 clusters equivalent. As discussed in Chapter III.2, the number of clusters can have an impact on the results' accuracy. To assess its impact, in this section, the same methodology was applied to different reduced models ranging from 10 to 200 clusters.

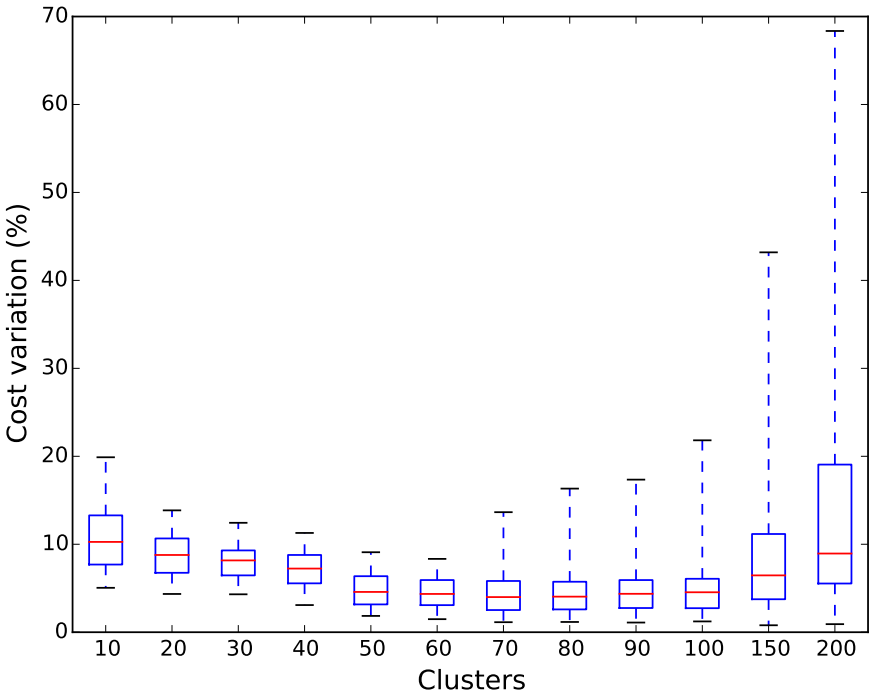

FIGURE IV.8: Boxplot of the absolute difference between the full and reduced system cost for different cluster numbers with $\eta = 85\%$. The red line represents the median.

Figure IV.8 pictures the boxplot of the absolute difference between the full and reduced system when varying the cluster number for a value of $\eta = 85\%$. In accordance with the results obtained before, for a smaller number of clusters, the cost differences tend to increase, and the median hits its minimum for 70 clusters. Nevertheless, when the number of clusters increases, the 95^{th} percentile increases significantly and for 200 clusters the median of the cost difference is higher than with 100 clusters, which is contrary to the expectations as is further discussed in the next paragraphs, as more clusters should

mean a reduced network model more close to the full one, and therefore less difference in the system's costs.

Figure IV.9 depicts the ENS for the same clusters definition, and helps to understand the cost difference variation shown in Figure IV.8. When the number of clusters increases, specially after 60, the quantity of ENS starts to increase, reaching its highest values for 150 and 200 clusters, having a strong impact on system's cost. Therefore, the definition $\eta = 85\%$ appears not to be suitable for a larger number of cluster definition.

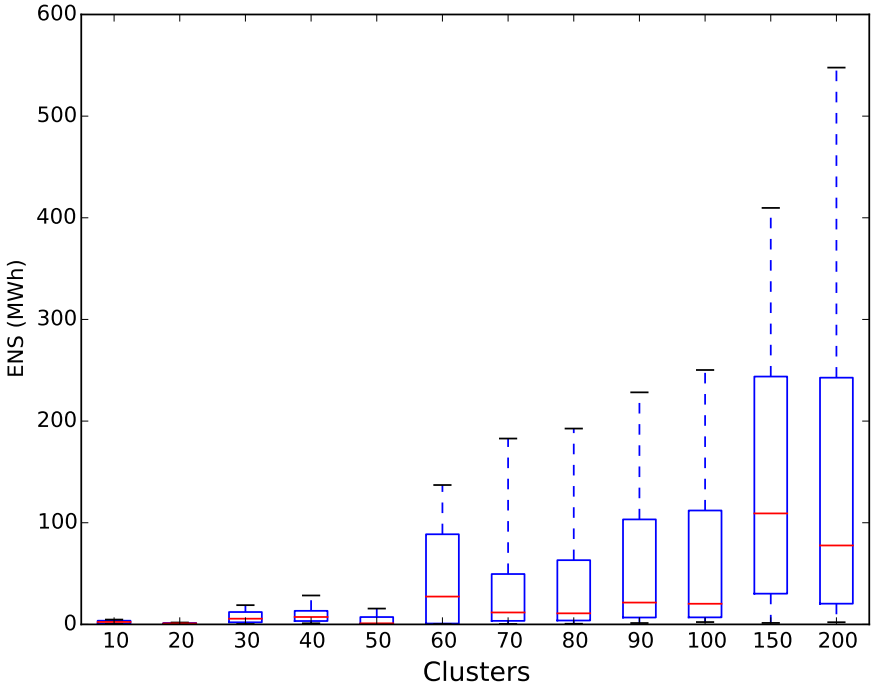

FIGURE IV.9: Boxplot of the energy not supplied different cluster numbers. The red line represents the median.

Figure IV.10 shows the boxplot of the absolute difference between the full and reduced system for a number of clusters between 70 and 200. Different values of η were tested to reduce the system's cost difference observed in Figure IV.8. For each cluster definition, the same simulations were performed changing only η from 85%(which was the used value in the previous simulations) to 95%.

It can be observed that for $\eta = 95\%$ the 95^{th} percentile reduces but the median of the cost tends to increase. This means that increasing the value of η will allow to relax the system's constraints in order to reduce the ENS, but if the value of η is too high, the cost variation will also increase, as the reduced system will be closer to the copper plate model. The proper choice of η relies in the value where the ENS is minimum, as this

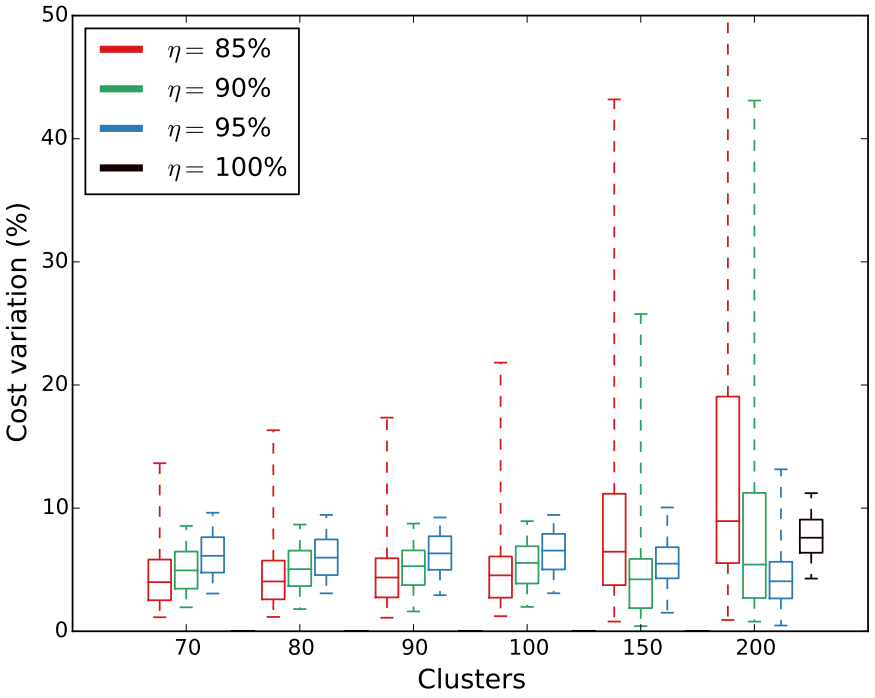

FIGURE IV.10: Boxplot of the absolute difference between the full and reduced system cost for different cluster numbers and η values.

means the system not too constrained (and therefore ENS costs do not impact it) and neither too unconstrained (as for some scenarios load shedding needs to be done).

Also, an important observation is that, following the trends observed in the previous chapters, doubling the clusters numbers from 100 to 200 does not significantly impact the accuracy of the model and the trade-off between the number of variables and accuracy is less interesting.

IV.6 Key findings

In this chapter, a methodology is presented that can accurately represent the steady-state constraints of the full model in the reduced one, by approaching their domain of admissible solutions. This allows to determine the reduced model's equivalent branches' transmission capacity.

This approach distinguishes itself from the previous works for needing neither specific simulations with the full model, nor assumptions regarding generation or load shift factors. Using only the full system's historical operating set-points, the followed approach

aims at proposing values that are independent of the system's operating conditions. To make it possible to fine-tune the model, an η is introduced so the more extreme scenarios do not bias the results.

The methodology is then applied to the benchmark system, and the results of the OPF calculations are compared with the ones obtained with the full model. Different indicators are proposed to evaluate the methodology, and results tend to show that even though some significant differences may occur in the production plan of each cluster, the changes in the overall cost of the system remain acceptable for prospective economic studies of power systems.

A sensitivity analysis to the number of clusters was performed, and results showed that for more detailed reduced models (i.e. larger number of clusters) the η coefficient must be adapted, otherwise, significant values of ENS impact the system's final cost. Also, it can be observed that for reduced models with more than 60 clusters, increasing the number of variables does not translate into a significant increase in the model's accuracy.

The application of this methodology should be extended to different case studies and enlarge the variety of situations to better assess its accuracy. This framework can be integrated into other reduction methodologies and can be of relevancy for optimisation problems applied to large-scale power systems, such as asset valuation for investments in transmission, generation, demand response, and storage infrastructures that require reduced but still accurate representations of the network.

Chapter V

Robustness assessment to power flow control devices

* * *

This chapter is organized as follows. Section V.1 describes the importance of power flow control devices in the operation of power systems, Section V.2 describes the proposed methodology to model them in reduced network models, Section V.3 presents the metrics to assess its quality when applied to the benchmark system in Section V.4, and finally, Section V.5 presents the key findings.

Contents of chapter V

V.1 The impact of phase-shifters on reduced networks

To deal with the increasing uncertainty in system operation, Transmission System Operator (TSO) rely more and more frequently on those devices [71]. These can be used as non-costly remedial actions to solve congestions by redirecting power flows, while avoiding redispatching or countertrading, or on a longer-term the costly reinforcement of the network.

It is therefore important to consider those assets appropriately when proceeding with network reduction, e.g. to perform simulation of complex processes in large-scale power systems. Indeed, reduced networks generally aim to reflect the main steady-state features of the full system. For example, different network reduction techniques have been proposed for systems' stability assessment [72] and transmission network expansion planning [29]. Most of the works perform static reductions based on a single operation point and therefore do not consider potential setting variations for the power flow control devices [53, 34]. In the approach developed in Chapter III of this work, multiple operation points in network reduction were considered, but the final result is a reduced model based on static Power Transfer Distribution Factor (PTDF), independent of the setting of power flow control devices. Therefore, the methodologies proposed in the previous chapters did not allow the fine-tuning of power flow control devices such as Phase Shifting Transformers (PST) or High Voltage Direct Current (HVDC).

Given the deployment of more and more controllable devices in transmission networks, the question arises whether reduced models should be upgraded, by explicitly representing PSTs and to be able to adjust their parameters according to the operation conditions.

Different approaches are proposed in the literature to model a PST in load flow computations [73], either by modelling relating the changes in nodal voltages angles to changes in active power transmitted [74], or by modelling its impact as an injected power, using a Phase Shifter Distribution Factor (PSDF) matrix, that establishes a relationship between the injected power and flow distribution through other network elements [75].

A limitation for these kinds of approaches is the need to know the electrical parameters of the system, namely branches' impedance or buses' angles. This can be a challenge when

dealing with reduced network models, which are often described using only a PTDF matrix.

To address this problem, a methodology to emulate and assess the impact of PSTs in reduced network models is proposed. PSTs are represented as an extra variable that can be adjusted subject to the systems operating point, whereas the other network components are represented by a static PTDF matrix defined with the methodology presented in Chapter III. To this end, multiple scenarios considering different operating conditions were simulated in the full model and compared to the representation in the reduced model. Errors are assessed in terms of power flows between clusters when applied to the benchmark system.

V.2 Defining a PSDF matrix for phase shifters

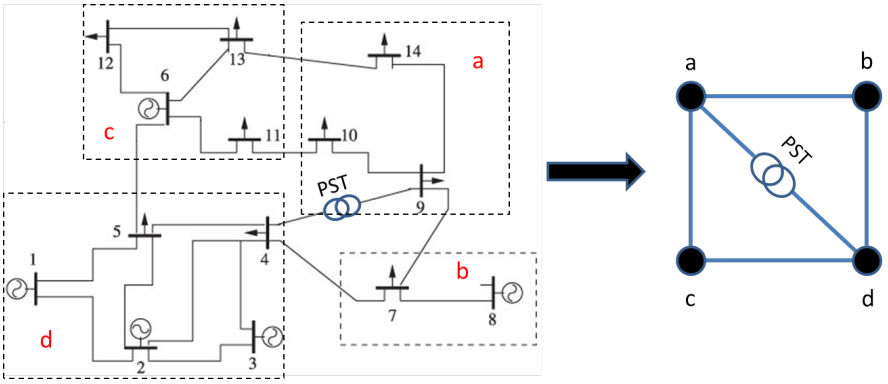

FIGURE V.1: Illustrative example of the performance assessment for a reduction approach.

A PST introduces a difference in voltage angle between two buses, that can be modelled as a power injection through the branch where it is installed. This increase/ decrease in the branch's flow affects the entire system with a redistribution over the other assets.

In other words, considering that the tap position of a PST, corresponding to an angle δ, would reduce/increase the flow in a given branch l of a given system, to comply with Kirchoff's laws, this power variation ΔP_l should be distributed through the remaining lines of the system.

This variation could be calculated using the coefficient from the system's PTDF matrix (Ψ) considering a as the injection bus and l the impacted branch. As an example $\Psi^a_{c,d}$

will be the coefficient of the matrix representing the influence of the injection in bus a over the branch connecting buses c and d.

Considering a PST installed on the branch connecting buses a and d, as illustrated in Figure V.1, the percentage of the new flow ($F^1_{(a,d)}$) that will be transferred to the branch connecting buses (c,d), $\omega_{(a,d)\rightarrow(c,d)}$ is determined as follows.

$$\omega_{(a,d)\rightarrow(c,d)} = \frac{\Delta F_{(c,d)}}{F^1_{(a,d)}} \tag{V.1}$$

The PST will induce an extra flow $\Delta P(\delta)$, that depends on the PST's angle installed on branch (a,d). Therefore $F^1_{(a,d)}$ can be calculated as:

$$F^1_{(a,d)} = F^0_{(a,d)} + \Psi^a_{(a,d)} \times \Delta P(\delta) \tag{V.2}$$

which is the original flow on the branch ($F^0_{(a,d)}$) plus the power injected by the PST ($\Delta P(\delta)$) times the coefficient of the PTDF matrix for bus a in branch (a,d).

Considering that in a extreme case, the injection by the PST equals the original flow on the branch, $F^0_{(a,d)} = \Delta P(\delta)$, Equation V.2 becomes:

$$\Delta P(\delta) = \frac{F^1_{(a,d)}}{1 - \Psi^a_{(a,d)}} \tag{V.3}$$

The flow's increment on branch (c,d) due to the PST tap change is:

$$\Delta F_{(c,d)} = \Psi^a_{(c,d)} \times \Delta P(\delta) \tag{V.4}$$

which is the power injected by the PST ($\Delta P(\delta)$) times the coefficient of the PTDF matrix for bus a in branch (c,d).

Therefore, the impact of the PST can be calculated as:

$$\omega_{(a,d)\rightarrow(c,d)} = \frac{\Psi^a_{(c,d)}}{1 - \Psi^a_{(a,d)}} \tag{V.5}$$

For all the branches L of the system, the new flow F^1 can be calculated as:

$$F_l^1 = F_l^0 - \omega_l^{PST\ bus} \times \Delta P(\delta) \ \ \forall l \in L \tag{V.6}$$

With exception of the actual branch where the PST is installed, where its impact will be -1 and will therefore be:

$$F_l^1 = F_l^0 + \Delta P(\delta) \tag{V.7}$$

V.2.1 Illustrative example

As illustrated in Figure V.1, the Matpower's 14 bus test case [76] was modified in order to accommodate a PST between the buses 4 and 9. The newly reduced system is composed of 4 buses connected by 5 branches, where the PST is installed in the branch connecting buses a and d. To assess the results the flows of the five branches of the aggregated model will be compared with the expected exchanges observed in the original system.

Considering that all branches have the same impedance, the PTDF matrix, $\Psi, \in \mathbb{R}^{\mathrm{L}\times\mathrm{N}}$ of the system is:

$$\Psi = \begin{bmatrix} 0 & -0.82 & -0.25 & -0.45 \\ 0 & -0.08 & -0.59 & -0.28 \\ 0 & -0.09 & -0.14 & -0.26 \\ 0 & 0.18 & -0.25 & -0.45 \\ 0 & -0.09 & 0.39 & -0.4 \end{bmatrix}$$

For a given set-point the injected power $[P_{inj}]$ and flows on the system branches F are:

$$P_{inj} = \begin{bmatrix} -53.4 & 8.5 & -34.3 & 79.2 \end{bmatrix}$$

$$F = \begin{bmatrix} -38.22 & -5.22 & -18.97 & -29.72 & -40.69 \end{bmatrix}$$

Applying the Equation V.5, a new matrix is defined that establishes the impact of the PST on every branch of the system:

$$PSDF = \begin{bmatrix} -0.36 & -0.22 & 1 & -0.36 & -0.24 \end{bmatrix}$$

Since in the reduced model the representation does not deal with the degrees, a choice has been made to represent the MW value caused by the tap change. In order to do this, in the original system a change of ± 1 deg in the PST installed in branch 3 was made to find the associated power injection. Table V.1 one can observer that $+1$ deg corresponds to an injection of 2.53MW.

TABLE V.1: Power flows for a PST with $\omega = \pm$ 1 degree.

	-1 deg	0 deg	+1 deg
Branch 3	-17.4 MW	-14.87 MW	-12.34 MW

Table V.2 shows the results for the same PST tap position in both the original and the reduced system. It can be observed that in both models, the branch where the PST is installed has almost the same value, but the same is not true for the other branches of the system where slight deviations appear. Even so, a satisfactory representation of the PST is obtained.

TABLE V.2: Power flows for a PST with $\omega = \pm$ 2 degree in both the original and reduced model.

	Original		Reduced	
	-2 deg	+2 deg	-5.06 MW	+5.06 MW
Branch 1	-28.2 MW	-34.9 MW	-34.7 MW	-31.3 MW
Branch 2	-2.8 MW	-1.9 MW	-3.0 MW	-0.9 MW
Branch 3	-19.9 MW	-9.8 MW	-19.4 MW	-9.8 MW
Branch 4	-19.7 MW	-26.4 MW	-26.2 MW	-22.8 MW
Branch 5	-40.7 MW	-43.5 MW	-38.3 MW	-36.1 MW

V.3 Impact assessment metric

To assess the quality of the reduced model, a comparison between the estimated flows between the clusters $\bar{F}_{\bar{l},s}$ and the observed flows of the full model $F_{l,s}$ will be made. To perform these simulations two different sets of injected power are used:

1. P_{zero}^{inj}: which corresponds to the injected power issue of the full model simulation without any PST optimisation;

2. P^{inj}_{optim}: which corresponds to the injected power issue of the full model simulation with the optimisation of the PST. During this calculation, all PSTs of network were considered as an optimisation variable that could vary between ±30 degrees.

Also, three different reduced model representations will be considered:

1. Ψ_{PST}: which is calculated using the methodology presented in Section III, and having as input P^{inj}_{optim}. This model does not explicitly includes any variable to represent the PSTs. In other words, this is a static PTDF matrix that was built using the injected power issue of the full model simulation with the optimisation of the PST.

2. Ψ_{static}: which is calculated using the methodology presented in Section III, and having as input P^{inj}_{zero}.

3. $\Psi_{static} + PSDF$: Besides the Ψ_{static} matrix, a $PSDF$ matrix is also calculated using the methodology described in Section V.2.

With these cases, it is intended to highlight the effects of the PSTs on different reduced network models.

A first assessment is performed, using a PTDF matrix following the methodology described in Chapter III (Ψ_{static}). As this does not explicitly models the PST, one can assess the error of the proposed methodology when PSTs are optimized. In order to do that, the different scenarios where simulated, one where PSTs were optimized (P^{inj}_{optim}) and other they were not considered (P^{inj}_{zero}).

Once the accuracy of the model when PSTs are neglected is known, the accuracy of explicitly representing PSTs in the reduced model is assessed, as suggested in Section V.2. To do that a PTDF matrix Ψ_{static}, plus a PSDF matrix, that describes the impact of PSTs, are used to describe the system.

The differences between the estimated $\bar{F}_{\bar{l},s}$ and observed $F_{l,s}$ flows will be compared using the Root Mean Square Error (RMSE):

$$RMSE = \frac{1}{S_E} \sum_{s=0}^{S_E} \sqrt{\frac{\sum_{l,\bar{l}=0}^{L} (F_{l,s} - \bar{F}_{\bar{l},s})^2}{L}} \tag{V.8}$$

Where S_E is the set of evaluation scenarios and L the total number of interconnectors.

Also, to have an overview of the methodology performance over the extreme scenarios of the proposed cases, a Value at Risk (VaR) index is calculated to assess the risk of extreme under performance for a reduced set of scenarios

V.4 Application to the benchmark system

V.4.1 Obtaining the reduced model

The hierarchical clustering, as defined in Section II.1.1, was applied to aggregate the network, using the S_T scenarios. Given the purpose of this approach, an exceptional restriction was introduced in order to avoid the aggregation of areas connected by PSTs, meaning that buses connected by a PST must always be in different clusters. Therefore, the algorithm would stop when all the observations at the interior of a zone, were grouped except for those connected by a PST. This resulted in an equivalent model of the benchmark system with 54 buses and 82 branches.

V.4.2 Assessing the impact of PST modelling

Given the high complexity of analysing the entire power system, for simplification purposes a focus was made on a single PST located in the border between Germany and the Netherlands.

As detailed in Section V.3, three different reduced network models and two set of injected power sets are used to assess the proposed methodology.

First, the static PTDF matrix Ψ_{static} is assessed regarding scenarios where no PSTs were optimized (P^{inj}_{zero}) and scenarios where PST was optimized (P^{inj}_{optim}). The goal is to assess the capacity of this network representation to represent both scenarios.

Second, the static PTDF matrix Ψ_{PST}, that was built using scenarios where the PSTs were optimized, is assessed for the two injected power scenarios. The goal being to assess the capacity of a PTDF matrix that was built based on "PST optimized scenarios" to represent both the optimized and non PSTs scenarios.

Lastly, the third representation of the reduced network model, consisting of the static PTDF matrix Ψ_{static} plus a PSDF matrix is also assessed using both the injected power scenarios, namely P^{inj}_{zero} and P^{inj}_{optim}.

The assessment consists in comparing the flows obtained with the reduced model simulations with the ones obtained using the complete model.

Table V.3 presents the error of the flows obtained with the reduced network model for Ψ_{static} and Ψ_{PST} PTDF matrix modelling. The error was calculated using Equation V.8 with the evaluation scenarios S_E.

TABLE V.3: Error values for the power flows comparison between the full and reduced model, without an explicit PST modelling.

	P^{inj}_{zero}	P^{inj}_{optim}
Ψ_{static} (MW)	143.3	200.6
Ψ_{PST} (MW)	204.7	138.5

Table V.3 demonstrates the impact of the optimization of the PST in the reduced model. The Ψ_{static} matrix performs well when the input scenarios do not consider the optimisation of the PST, but the error tends to increase when the P^{inj}_{optim} is used. On the other hand the Ψ_{PST} matrix can reduce the error for the case where the PST is optimized P^{inj}_{optim}, with a RMSE of only 138.5 MW per branch per scenario, but the error rapidly increases when P^{inj}_{zero} is applied.

The results in Table V.3 show that the static PTDF matrix representation performs well under the scenarios from which it was built. When the "non PST optimized" injected powers (P^{inj}_{zero}) are applied to the "PST optimized" PTDF matrix (Ψ_{PST}) the results are less accurate than when applied to the "non PST optimized" PTDF matrix (Ψ_{static}) representation and vice-versa.

The same trend can be observed in Table V.4, which presents the VaR of 5% for the flows calculated using the reduced model. It can be remarked that the values of VaR are similar for both P^{inj}_{zero} and P^{inj}_{optim} when using the Ψ_{static}, but more significant values arise when applying P^{inj}_{zero} to the Ψ_{PST} matrix.

Finally, the reduced model including an explicit modelling of the PST ($\Psi_{static} + PSDF$) is tested. For the injected power set where PSTs were optimized (P^{inj}_{optim}), the PST of the reduced model mimics its behaviour, in other words, an injected power is multiplied

TABLE V.4: VaR of 5% for the flows comparison between the full and reduced model, without an explicit PST modelling.

	P^{inj}_{zero}	P^{inj}_{optim}
Ψ_{static} (MW)	273.1	279.3
Ψ_{PST} (MW)	307.2	248.9

by the $PSDF$ matrix. Similar to has been done in Section V.2.1, a relationship between degrees of the PST and injected power was established and the same power is injected by the PST in the reduced model.

TABLE V.5: Error values for the flows comparison between the full and reduced model, with an explicit PST modelling.

$\Psi_{static} + PSDF$	P^{inj}_{zero}	P^{inj}_{optim}
RMSE (MW)	143.3	194.2
VaR (MW)	273.1	275.3

Table V.5 shows the error results for the case where an explicit modelling of the PST was done ($\Psi_{static} + PSDF$). As it can be observed, for the case where the PST was not optimized (P^{inj}_{zero}), the error is the same as presented in Table V.3, as the injected power of the PST of the reduced model will be set to zero.

When considering the case where the PST was optimized (P^{inj}_{optim}) the error shows a slight reduction issue to the explicit PST modelling. Also, when looking into the VaR, it can be observed that with the explicit modelling the VaR tend to be similar independently of the considered case.

Comparing both the results obtained with and without the explicit PST modelling, it can be observed that when relying only on the optimization of the PTDF matrix using a set of data where PST were optimized, it tends to under perform for the case where the PST were not optimized and vice versa. When adding the explicit modelling of the PSTs, the results for the case using P^{inj}_{optim} are not so accurate as the ones obtained with the Ψ_{PST} matrix, but are more accurate for the case where P^{inj}_{zero} is applied.

Overall, the proposed model with an explicit modelling of the PSTs loses some accuracy for a specific case P^{inj}_{optim}, but compensates by not under perform for the opposite case P^{inj}_{zero}, as it is proven by the VaR values presented in Table V.5.

V.5 Key findings

A methodology to model PSTs in reduced network models and its effects in transmission networks are presented.

A performance assessment index is proposed that allows to assess the pertinence of representation in a reduced model, different cases are studied, including the one where no PST representation exists.

Preliminary results tend to show that there is an interest in modelling PSTs in reduced network models. Simulations show that there is a clear impact of this kind of devices in the network, and that explicitly modelling them along with the PTDF matrix can increase the level of accuracy of the reduced model when comparing to the approach based only on the definition of the PTDF matrix.

The proposed methodology can be of interest for the development of reduced static model of a large scale power network, such as the ones performed in [34]. The increase of these devices over the network, poses more and more difficulties to a static representation of the power system over its different operation conditions.

It is important to stress that given the specificities of power systems, the results are conditioned by the choice of the simulated PST. A PST in a more central position or next to critical bottlenecks can have a different impact on the system flows and production plan, in the same way as a more decentralized PST can cause the inverse. Also, it is important to remember that all the results were obtained using the Direct Current (DC) approximation, and therefore should be carefully interpreted, as the with a full Alternating Current (AC) analysis results might differ.

Therefore, a key direction for further work is to expose the proposed methodology to a larger set of operation conditions, and equipment specificities. It would also be of interest to assess the suitability of such methodology to model HVDC cables and assess its impact on reduced network models.

General conclusion and perspectives

Overview

The evolution of the regulatory framework has lead to different challenges in power systems' operation and optimization. Different stakeholders are responsible to optimize the system independently and, to cope with the interactions between market and network models, optimizations problems become more and more complex posing problems to computational tractability. To overcome those problems system's stakeholders must rely more and more on accurate reduced systems to perform their prospective studies. Given the importance of keeping a temporal coherency when performing prospective economic studies and also to obtain a simplification robust to different input parameters, in this work, a choice was made to perform spatial complexity reduction of the network model. Therefore, the work developed in this thesis proposes to determine a reduced network model for economic studies with as few clusters as possible while maintaining an acceptable level of accuracy.

From this perspective, three of the most promising network clustering techniques with different distance metrics are studied and applied to a benchmark system representing the European power system. A framework is proposed to rank them and identify the most suitable one to apply. Results show that hierarchical clustering (using Locational Marginal Price (LMP) as metrics) outperforms k-means (using LMPs and geographic coordinates) and k-medoids (using electrical distance and geographic coordinates) in a multi-scenario analysis and also that after 50 clusters, the trade-off between accuracy and model simplification is less interesting.

Once the network is clustered, a methodology is proposed to determine the branches between them, that accurately represent the power flow repartition of the full model. A common way to represent those branches is by using a Power Transfer Distribution Factor (PTDF) matrix, but existing approaches tend to focus on a single operation point, conditioning the robustness of this representation in different system operating conditions. A new methodology is proposed that defines a PTDF based on multiple scenarios. The obtained results validate the approach, as branches' flows are accurately represented through different system operating conditions.

With the network clustering performed and the branches between them characterized through a PTDF matrix, a first reduced model is obtained. Still, to perform power system economic analysis a key parameter is missing: branches' maximum transmission capacity. Most of the approaches developed in the literature focus on load flow analysis of the network, and maximum transmission capacity is often neglected or approximated using empiric approaches. A new approach to approximate the reduced system's operational limits to the full system's one is proposed. With this approach, equivalent branches' maximum transmission capacity can be estimated and a full characterization of the reduced network for economic analysis is obtained. Results show that this characterization can accurately represent the full system, as results from an economic analysis remain accurate through different operating conditions.

Main contributions

Four main contributions can be taken from this thesis:

- The development of a framework to rank clustering methodologies based on the redispatch of generating units. This methodology compares the generation dispatch using the full model with the one obtained using the reduced model. The aggregation of network buses will eliminate branches' constraints whenever two buses are aggregated and a copper plate approach is considered.

  Besides the use proposed in this thesis, this approach can also be used in a market environment to assess the adequacy of the bidding zone definition in power markets, as presented in [58].

- The definition of a methodology to estimate a multi-scenario fitted PTDF matrix that represents the links between clusters. This methodology presents a clear

improvement regarding the existing approaches in the literature, as considering multiple scenarios in the definition of the PTDF matrix improves the robustness of this representation for different operating conditions.

- The definition of a methodology allowing to represent the full system's steady-state constraints on a reduced model. This approach that can be applied to determine the different variables that limit the feasibility domain of the system's optimization, is used in this work to fully characterize the connection between clusters and estimate their maximum transmission capacity.

- The definition of a methodology to assess the robustness of static reduced network models to the optimization of power flow control devices such as Phase Shifting Transformers (PST). This can be useful to assess the error in power system's operation and optimization studies that use reduced models without explicitly model PSTs.

Perspectives

The main contributions and approaches developed in this work can be applied to different studies namely:

- The developed methodology throughout this work allows to define reduced models with different clusters numbers and also to assess the error associated with those models. This allows to observe that the benchmark system based on the European network can be accurately represented using a 60 cluster reduced model. Outside the scope of this thesis, this error quantification can also be of help to better interpret some economic studies using extremely reduced network models, and whose conclusions are strongly linked with the network model simplification.

- Different economic analyses of assets evaluation are performed using reduced models to overcome the computational tractability problem of large-scale power systems. Those reduced models are often determined by empiric approaches without a clear assessment of the error introduced by the defined reduced network model used, and therefore, the approaches developed in this thesis could be of interest to determine a reduced network model that guarantees the computational tractability of the optimization while assessing the impact of the network constraints in the assets economical evaluation.

- Bidding zone review studies should be periodically performed to assess the adequacy of the European network bidding zone configuration. A clear methodology to assess it is missing and the existing ones are frequently limited by the definition of strong assumptions that clearly influence the final results. In this way, the framework proposed in this thesis could be of interest to assess the adequacy of bidding zone delineation regarding network structural congestions.
- The proposed methodology allows to determine a reduced network model from a series of independently simulated Optimal Power Flow (OPF), that can then be used to perform more complex simulations. The obtained results throughout this thesis tend to confirm that the error associated with the reduction can be sufficiently reduced to provide coherent results in more complex simulations, such as the Unit Commitment (UC).

Further works

Throughout the steps of the work developed in this thesis, different aspects were identified for future works:

- The bus clustering was performed using mainly the LMP as input data, and once the aggregation was performed, its quality was assessed using the redispatch index as criteria. The clustering results could be improved if the redispatch index was explicitly taken into account when performing the bus clustering, as this would take into account more information than just the LMPs.
- The aggregation of network buses using congested branches as the criteria to determine the clusters, revealed to be efficient but when determining the PTDF matrix that characterizes the links between them, it is clear that besides the congested branches, other criteria can influence the quality of this representation: loop flows. Taking this variable into consideration in the clusters' definition could positively impact the results obtained with the PTDF representation.
- The definition of the PTDF matrix based on multiple scenarios proved to be efficient in improving the robustness to the system's operating conditions, nevertheless, the proposed formulation requires the knowledge of multiple historical operation points in order to define the reduced PTDF matrix. By slightly changing the proposed formulation to take into account the differences in injected power

between scenarios and not the injected power itself, as PTDFs can be seen as translating the differences in power injections into branches' flows, an improvement in the results for the same number of scenarios could be obtained.

- The proposed methodology determines and assesses the reduced model based on simulations of historical set-points. Even though, this allows to assess the lost information in the reduction, it does not allow a comparison with real historical operating points taking into account the Transmission System Operator (TSO) decisions. Using real historical data to train and evaluate the model would allow that comparison, and therefore, assess how well the reduced model can reflect reality. Also, the impact of the approximation of the UC computation by a series of independent OPF should be assessed and different approximation strategies could be studied, such as the proposed in [77].

- All the proposed methodologies were defined and compared with a system without the *N-1* security criteria. Assessing the robustness of the reduced model, when considering this security criteria, would be of interest. Moreover, this would also be an interesting challenge to the proposed method to estimate branches' equivalent capacities.

List of publications

International journals

N. Marinho, Y. Phulpin, D. Folliot, and M. Hennebel, "Redispatch index for assessing bidding zone delineation," in IET Generation Transmission and Distribution, 2017.

N. Marinho, Y. Phulpin, A. Atayi, and M. Hennebel, "Equivalent transmission capacity in reduced network models," to be submitted

International conferences

N. Marinho, Y. Phulpin, D. Folliot, and M. Hennebel, "Approaching generation variable costs from publicly available data," in 2016 13th International Conference on the European Energy Market (EEM), June 2016.

N. Marinho, Y. Phulpin, D. Folliot, and M. Hennebel, "Network reduction based on multiple scenarios," in 2017 IEEE Manchester PowerTech, June 2017, pp. 1-6.

N. Marinho, Y. Phulpin, A. Atayi, and M. Hennebel, "Phase shifters in reduced network models," to be submitted

Appendix A

Defining generation variable costs

This appendix is a support for the thesis reader, it provides the methodology used to determine generation variable costs.

A.1 The importance of generation variable costs

A simplified model to estimate the generation units' variable costs by technology and at a national, regional and unit scale is proposed. Given a function f_i that determines the real variable cost y_i of the generation unit i:

$$f_i(G_i, X) = y_i \tag{A.1}$$

where G_i is an array with the specific characteristics of unit i (such as unit's efficiency, salaries, etc.) and X is an array with all the external variables (such as fuel cost, CO_2 emissions cost, etc.).

A function $f'_{S,T}$ will be determined for each geographic scope S and technology T that estimates the approximative variable cost y':

$$f'_{S,T}(p, g_{S,T}, d_{S,T}, h) = y' \tag{A.2}$$

where p, g, d and h represent time series of publicly available data namely price, actual energy output, available capacity and hour, respectively.

This approach is based on the assumption that the electricity spot markets are fair and efficient, with no market power abuse and that the market clearing price represents the marginal cost of the most expensive unit required to meet demand [78].

The goal is to approximate y_i with $y^{'}$ at a unit, regional and national scale S for each technology T. Different strategies to process the entry data will be tested. To validate the methodology, the results are compared with the realizations of generation for different market prices in the past.

A.1.1 Methodology

a) Generation data pre-processing

In a raw analysis of the electricity spot prices, different phenomena such as the seasonal variability of the generation portfolio and operation strategies of generation units can deteriorate the correlation between price and actual power output of each generation unit, making it difficult to extract useful information.

Using the information of the amount of available capacity for each hour, the spot market price is related with the ratio between the power output and the available power output. The advantage of this strategy can be clearly seen for the nuclear power, where the availability of the nuclear fleet can vary significantly, depending on the season. In some countries the available generation information is not known, and in those cases the analysis is performed considering only the actual generation.

When establishing a relation between spot market price and power output, all the points below and above a certain threshold of the power output, p_{low} and p_{high} respectively, will be excluded, depending on the technology under consideration and the geographic scope. For the upper threshold, the value is chosen to exclude all the situations where the technology is not marginal, and therefore, is no longer defining the market price. For the lower threshold, the value is defined to exclude all the situations where the generation unit is not responding to the markets signals (e.g. auxiliary services, operational testing, start-up and shut-down periods, etc.)

b) Price data pre-processing

With all the selected data, it may still be difficult to make a robust estimate based only on the hourly spot market prices and generation unit power output, especially with base technologies, as their operators often define generation schedules over larger periods. This can be mostly influenced by its high start-up costs and difficulty to quickly vary its production over a small period of time, that can force the units to produce during periods where the market price is below its variable costs. For example, a hard coal fuelled generation unit can produce during the night where the market price is low to be able to participate in the morning peak and receive the high market prices of that period.

To study this phenomena, it is proposed to consider the market prices over larger periods of time, to simulate the case where the operator decides to produce, even though it is not profitable yet, so that he can receive a higher price in the following hours. The enlarged interval, should always consider the price for hour h as the commercial losses during that period should always be considered.

In that way, the production in one hour is correlated to the maximum price of a sliding window average of duration b around the selected hour h, and the price p_h for each hour is defined as follows and illustrated in Figure A.1:

$$p_h = max(m_{n1}, \ldots, m_{nb}) \tag{A.3}$$

where m_{ni} corresponds to:

$$m_{ni} = \frac{1}{b} \sum_{k=0}^{b-1} x_{n-b+k+i} \quad , \quad i \in [1, b] \tag{A.4}$$

where x_n is the original market price for hour n.

c) Variable costs estimation

At a national and regional scale:

When the data under consideration contains information of several generation units (i.e. national or regional scale), the estimation of the variable costs is given by a regression

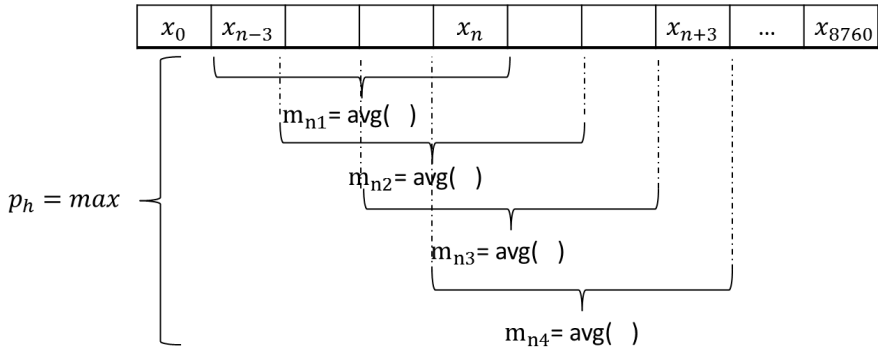

FIGURE A.1: Scheme of the sliding window average methodology with $b = 4$.

on the r percentile between p_{low} and p_{high}. These three parameters are adjusted based on the technology and geographic scale under consideration, and the output of the generation units is normalized by a step of 5 MWh.

The regression is then extended through all the power range of the generation unit and capped at 0 and P_{max}, as illustrated in Figure A.2 by the dashed green line.

At a generation unit level:

As opposed to the approach followed at a national level, at a generation unit scale there is no interest in analysing a wide range of the power output. Considering that the unit will produce at full power, once the market price is above its variable costs, and has no commercial interest to produce at all once the market price is below, the analysis will focus on the values closer to the maximum power P_{max} ($p_{low} > 80\% \times P_{max}$).

The estimation is given by the r percentile of the price between that interval, allowing to define a unique variable generation cost for the generation unit, instead of an interval as defined for the national and regional scale.

d) Performance indicator:

The parameter b was adjusted to each technology and geographical scale, while observing the accuracy A of the cost approximation and the Root Mean Square Error (RMSE).

The accuracy indicator is defined as the sum of the number of times when the market price is above the cost estimation and the generation unit is producing (above 80% of its maximum output) $count_{up}$, plus the number of times the market price is below the cost

estimation and the generation unit is not producing (output below 20%[1]) $count_{down}$. This is then divided by the total number of points analysed $count_{total}$ to define the accuracy ratio A as:

$$A = \frac{count_{up} + count_{down}}{count_{total}} \tag{A.5}$$

The use of a single value for the cost approximation can lead to an underestimation of the methodology since, in the case of a national analysis given the heterogeneity of the park, it is likely to find different costs for the same technology, as can be seen in Figure A.2.

Therefore, the RMSE was also analysed since it assesses the estimation of the power output having the spot market price as an input and is defined as follows:

$$RMSE = \sqrt{\frac{\sum_{h=1}^{n}(g_{i,h} - g'_{i,h})^2}{n}} \tag{A.6}$$

Where $g_{i,h}$ corresponds to the realized generation output of unit i for the hour h and $g'_{i,h}$ to its approximation.

A.1.2 Results

The proposed methodology was applied focusing in Germany and France, using data from the year 2014, more specifically, an array containing the 8760 spot market prices, actual generation output and available capacity (when disclosed). This information is published in websites such as [79] and [80] that aim to gather the information from multiple countries, and can also be found in some Transmission System Operator (TSO)'s website, such as [81], that publish detailed information about the units connected to its own system. These two countries are of particular interest given their geographic proximity and the key role generation variables costs play in the definition of their exchanges.

a) National level

At the national level the regression was performed considering $r = 50$, and p_{low} and p_{high} corresponding to 20% and 80% of the available capacity P_{max}, respectively.

[1] Any generation below this threshold will be considered limited by dynamic constraints (e.g. cool-down, ramp-rate, etc.), and not correlated with market prices.

In Figure A.2, it can be observed that the regression has a considerable slope referring to the different costs of the generation units analysed within the country, and that the estimated variable cost is contained within an interval of 25€-45€. Since the data under consideration corresponds to a national level, and contains a large variety of generation units that can be geographically disperse (causing an impact in fuel transportation costs, for example) and have different efficiencies that have an impact on their variable cost, it is more appropriate the definition of an interval of variable costs and not a single value. Also, since there was no available information regarding the availability of the generation units in Germany, the analysis was made considering the absolute values of power, which might be a source of errors.

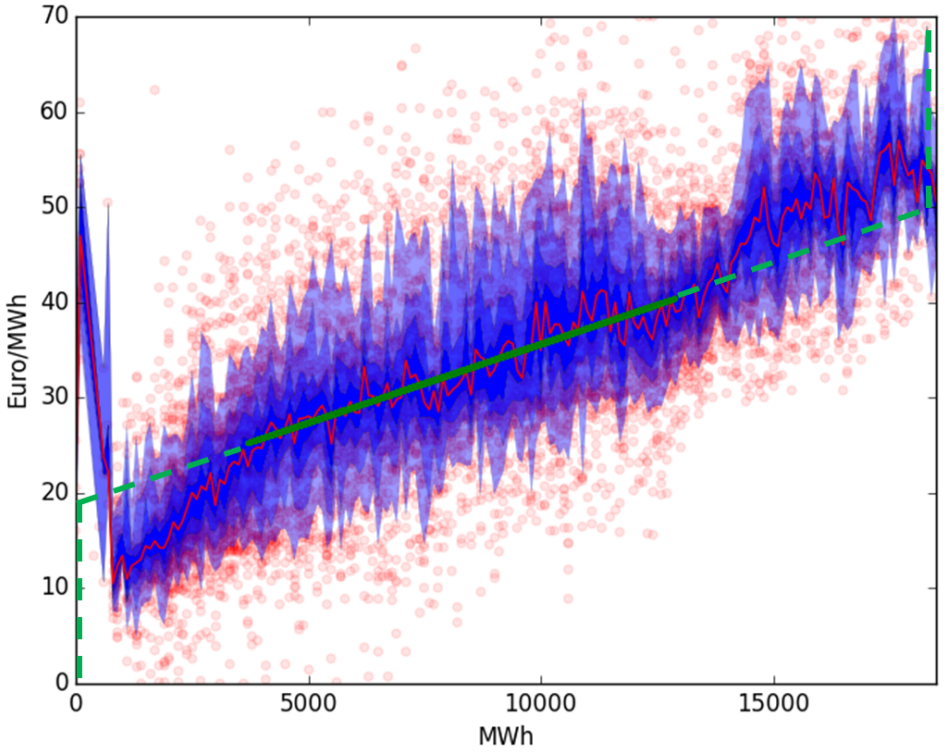

FIGURE A.2: National level variable cost estimation for coal generation in Germany. The red points represent the pair (production, market price). The shades of blue represent the 5% percentiles, the red line represents the 50^{th} percentile and the green line the estimated variable costs with $p_{low} = 20$ and $p_{high} = 80$. The dashed green line represents the extension from 0 to p_{low} and from p_{high} to P_{max}.

In Table A.1 and A.2 it can be observed the accuracy of the estimation for a set of different price representation windows. The consideration of larger time periods does not improve the accuracy for most of the technologies, except for the fuel oil type power plants, that marginally increases it.

TABLE A.1: Accuracy results for generation variable costs at national level in Germany for 2014, with $r = 50$.

b	Gas	Hard Coal	Lignite	Nuclear	Fuel Oil
1	***84.4%***	***69.8%***	***71.8%***	***67.6%***	91.5%
2	84.2%	66.5%	70.0%	67.0%	92.0%
3	84.1%	63.2%	68.3%	66.4%	***92.3%***
4	84.1%	61.0%	66.7%	65.9%	92.3%
5	84.1%	59.0%	65.5%	65.8%	92.2%
6	84.1%	57.7%	64.8%	65.5%	91.9%

Also the RMSE, does not vary significantly with the values of b, and the best estimations are found for its lower values.

TABLE A.2: Root mean squared error for generation at national level in Germany for 2014, with $r = 50$.

b	Gas	Hard Coal	Lignite	Nuclear	Fuel Oil
1	***18.9%***	***22.2%***	29.5%	***17.7%***	24.5%
2	19.4%	22.5%	29.2%	17.7%	22.9%
3	20.2%	23.5%	***28.7%***	17.8%	22.3%
4	19.7%	24.4%	28.8%	17.9%	***22.2%***
5	20.0%	25.4%	28.9%	18.2%	22.9%
6	21.5%	26.3%	28.9%	18.0%	23.5%

In Table A.3 and A.4 the same results can be observed for France. It can be noticed that the accuracy is significantly low for fuel oil, which can be justified by the reduced number of hours this type of generation unit produces throughout the year.

Also, despite the high accuracy shown for the nuclear power units, it should be taken into consideration that these units are near full power most of the time, which makes it hard to define its response to the market prices.

The high RMSE values presented in Table A.4, in particular for fuel oil, demonstrates that the results are just an estimation of the real variable costs. For the nuclear fuelled generation units, the reduced variation in their power output (always running close to full power) underestimates the error committed in the proposed approach justifying the low RMSE values presented.

TABLE A.3: Accuracy results for generation variable costs at national level in France for 2014, with $r = 50$.

b	Gas	Hard Coal	Nuclear	Fuel Oil
1	92.7%	***73.2%***	***89.3%***	***40.0%***
2	92.7%	70.8%	87.4%	39.4%
3	***93.1%***	68.5%	84.6%	39.2%
4	93.0%	67.2%	82.2%	38.7%
5	92.9%	66.1%	79.7%	38.4%
6	92.7%	65.6%	77.9%	38.4%

TABLE A.4: Root mean squared error for generation at national level in France for 2014, with $r = 50$.

b	Gas	Hard Coal	Nuclear	Fuel Oil
1	***27.5%***	23.1%	***4.8%***	73.6%
2	29.2%	***22.8%***	5.1%	72.2%
3	29.2%	23.3%	5.2%	71.8%
4	30.6%	23.5%	5.6%	71.9%
5	31.7%	24.0%	5.8%	***71.7%***
6	30.9%	24.2%	6.6%	78.2%

In Table A.5 the estimated costs for France and Germany are presented, except for fuel oil and nuclear power plants, which as justified above, do not have an accurate value, defined as the minimum and maximum interval of the regression curve, corresponding to p_{low} and p_{high}, respectively. The costs present the expected merit order between the different technologies, and complying with the results of the electricity spot markets for the year of 2014 [82], it can be noticed that the hard coal variable generation costs in France are higher than in Germany.

TABLE A.5: Estimated variable costs per generation type (min - max) at national level for Germany and France for 2014, with $r = 50$.

	Germany	France
Gas	50.1€–55.4€	49.1€–54.7€
Hard Coal	24.7€–45.0€	31.7€–45.7€
Lignite	17.2€–32.2€	–

b) Generation unit level

The analysis was performed for two hard coal fuelled generation units, one in France and the other in Germany (Figure A.3). When producing close to P_{max} the probability that the unit is not marginal and is receiving a higher price increases. Consequently, the percentile in which the regression is performed was reduced to increase the accuracy of the estimation. As it can be see from Table A.6 and A.7, the estimations are more accurate for the lower percentiles. The choice of such a lower percentile is related with the technology analysed, since hard coal is used as a base technology, it is more likely that it is not setting the market price during its production hours.

Table A.6 presents the results for the French hard coal generation unit. The accuracy of the estimation is higher for the lower percentiles, namely the 10^{th} and the 20^{th} and the best results can be found for $b = 3$, but the variations caused by this parameter are slight, and it is essentially the variation of the percentile that influences the accuracy of the estimation.

TABLE A.6: Accuracy results for a single unit variable cost in France with $p_{low} = 0.85$ for 2014.

Percentile	b=1	b=2	b=3	b=4	b=5	b=6
50	73.5%	72.9%	72.1%	71.7%	71.7%	71.9%
40	75.3%	74.0%	73.6%	73.4%	73.3%	73.1%
30	76.5%	75.8%	75.2%	74.8%	74.5%	74.3%
20	***76.9%***	***76.8%***	76.4%	76.2%	76.1%	76.1%
10	75.9%	76.6%	***77.1%***	***77.0%***	***77.0%***	***77.0%***
5	74.1%	75.5%	76.3%	76.6%	76.7%	76.9%
1	67.5%	70.7%	72.7%	74.2%	75.4%	75.7%
0.5	65.6%	67.9%	69.5%	72.3%	72.9%	74.7%

For the German hard coal generation unit analysis presented in Table A.7, the best estimation is always obtained with the percentile $r = 1$ and, with exception from the two lower percentiles, the quality tends to decrease when b increases, following the same trend observed in the national approach.

It is important to stress out, as these results show, that the estimation does not correspond to the actual price of the generation unit, but is an important indicator when comparing to other generation units. In Figure A.3, it is interesting to observe a considerable number of points that deviate from the average, that can be a possible indicator

TABLE A.7: Accuracy results for a single unit variable cost in Germany with $p_{low} = 0.85$ for 2014.

Percentile	b=1	b=2	b=3	b=4	b=5	b=6
50	50.4%	47.9%	45.9%	45.6%	45.6%	45.5%
40	54.9%	52.5%	51.2%	49.6%	49.6%	49.9%
30	59.4%	57.3%	56.0%	55.2%	54.6%	54.4%
20	63.7%	61.8%	60.2%	59.6%	59.2%	59.4%
10	68.7%	66.5%	65.1%	64.2%	63.7%	63.8%
5	73.1%	71.5%	70.2%	68.5%	67.6%	67.3%
1	***74.8%***	***75.2%***	***75.3%***	***75.6%***	***75.6%***	***75.6%***
0.5	73.8%	74.5%	74.9%	74.9%	75.0%	75.0%

of the reduced flexibility of this technology that corresponds to situations where it was not profitable to produce, but it was physically infeasible to reduce the power output.

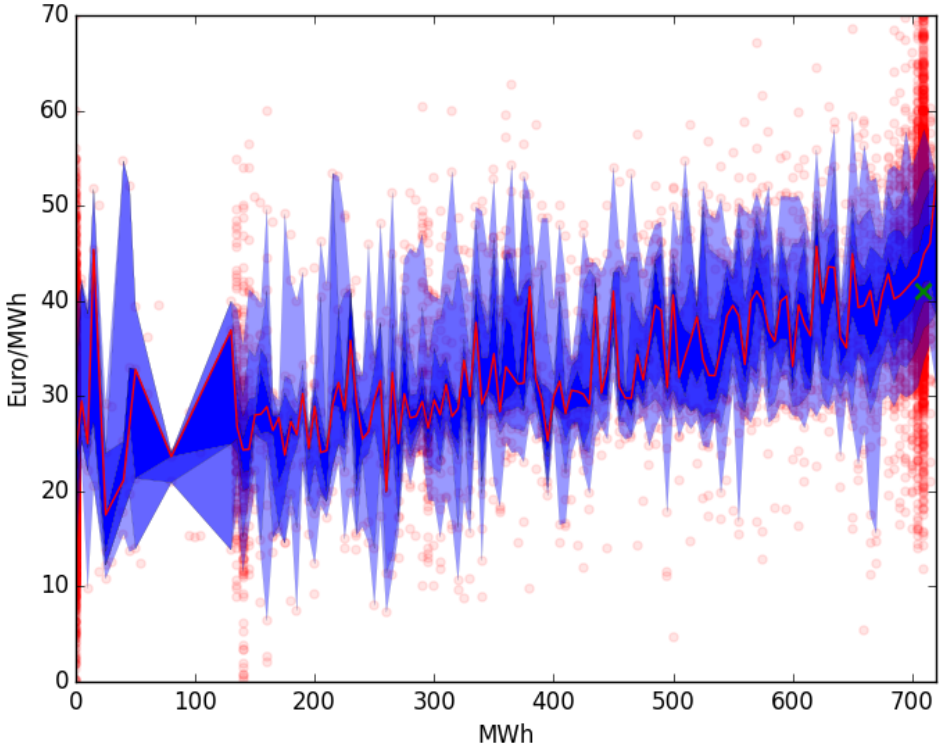

FIGURE A.3: Hard coal powered generation unit in Germany. The red points represent the pair (production, market price). The shades of blue represent the 5% percentiles, the green cross represents the variable cost estimation with $r = 40$.

A.1.3 Conclusions

The presented results tend to show significant costs differences between countries that are geographically close to each other and even electrically well connected, which is the case between France and Germany. This puts into evidence the motivation of this approach in the characterization of production costs per country, area or generation unit.

This approach provides an estimation of variable generation costs that need to be carefully handled. The variable costs of a generation unit depend on a large number of variables that are extremely difficult to simulate/predict and are not within the scope of this work and therefore can bias the estimation.

Apart from network and dynamic constraints that can influence a generation unit's output and that are not considered here, differences in the production schedule defined in the day-ahead markets and all the adjustments that are made throughout the different sessions of the intraday markets and finally also at the balancing market are also part of this works' limitations, and degrade the correlation with spot market price.

Nevertheless, the estimation establishes a consistent merit order between the different technologies, countries and regions that can be an important contribution to economic studies of power systems considering network models. In particular, it is interesting to observe the different prices for the same technology within the same country, that can in some cases inverse the expected merit order (e.g. after starting 80% of the hard coal park it is cheaper to start a gas powered generation unit than to increase the production of hard coal.) and may have an impact in redispatch costs. These variable costs' intervals for the same technology can be allocated based on:

- the age of each generation unit, considering that the older ones are likely the most expensive;
- their location, considering the impact it may have on fuel transportation costs;
- the average of the regression curve as the estimated cost for all generators.

If there is no information that allows the implementation of the previous strategies, the costs can be randomly allocated through the different units.

A key direction for further work is to develop this model to take into account more variables such as the CO_2 price and fuel cost, that would improve the results and allow to predict the variable costs based on forecasts of those external variables.

It would be also challenging to test this methodology for different years and countries with diverse generation mixes.

Appendix B

Power Factory OPF formulation

This appendix is a support for the thesis reader, it provides the details of the Optimal Power Flow (OPF) formulation in DigSilent's Power Factory used throughout this work as stated in its user manual.

B.1 OPF formulation

The OPF calculation is initialised by a load flow, which is calculated using the linear DC load flow method. Power Factory uses a standard LP-solver (based on the simplex method and a branch-and-bound algorithm) which ascertains whether the solution is feasible. The result of the linear optimisation tool includes calculated results for control variables, such that all imposed constraints are fulfilled and the objective function is optimised.

Provided that a feasible solution exists, the optimal solution will be available as a calculation result. That is, the algorithm will provide a DC load flow solution where all generator injections and tap positions are set to optimal values. The DC load flow solution includes the following calculated parameters (parameter names are given in italics):

- For terminals
  - Voltage Angle (phiu [deg])
  - Voltage Magnitude (u [p.u.]; assumed to be 1.0 p.u. in DC calculation)

  - Voltage Magnitude (upc [%]; assumed to be 100% in DC calculation)
  - Line-Ground Voltage Magnitude (U [kV])
  - Line-Line Voltage Magnitude (U1 [kV])

- For branches:
  - Active Power Flow (P [MW])
  - Active Power Losses (Ploss [MW]; assumed to be 0MW in DC calculation)
  - Loading (loading [%]; Loading with respect to continuous rating)

B.2 Objective function

The following parameters are calculated in addition to the results found by the DC load flow:

- For generators:
  - The fixed cost factor [$/ MWh] used in the objective function (i.e. average cost considering the costs at the generator's active power limits);
  - Optimal power dispatch for generator;
  - Production costs in optimal solution.
- For Transformers:
  - Optimal tap position.
- For loads:
  - Optimal load shedding for load.

Even though different objective functions are available when executing a DC Optimisation, in this work a minimisation of costs is used.

The objective is to minimise generation costs. To perform a cost minimisation calculation for each generator, a cost factor needs to be entered corresponding to a cost curve $/MWh per generator element.

The (linear) algorithm uses a fixed cost-factor [$/MWh] per generator. This cost factor is the average cost considering the costs at the generator's active power limits. The selection of this objective function provides the option of calculating the Locational Marginal Prices (LMPs).

B.3 Control variables

The following control variables can be selected when performing a DC OPF:

- Generator Active Power Dispatch

  In generator optimisation, for each selected generator a single control variable is introduced to the system. The total number of generator controls in this case equals the number of selected generators.

- Transformer Tap Positions

  In tap optimisation, for each selected transformer a single control variable is introduced to the system. The total number of tap controls in this case equals the number of selected transformers.

- Allow Load Shedding

  A separate control variable is introduced to the system for each selected load. The total number of load controls in this case equals the number of selected loads. This control variable can be selected in conjunction with any objective function.

B.4 Constraints

When performing a DC OPF three different categories of constraints can be defined:

- Active Power Limits of Generators

  For each synchronous machine, the user may impose up to two inequality constraints, namely, a minimum and maximum value for active power generation specified as MW values.

- Branch Flow Limits (maximum loading)

  Branch flow limits formulate an upper bound on the loading of any branch. Loading limits are supported for lines and 2- and 3-winding transformers.

- Boundary Flow Limits

  Power Factory boundary elements, icon define topological regions in a power system by a user-specified topological cut through the network. Constraints can be defined for the flow of active power in a network (over a defined boundary or between internal and external regions of a boundary), and this constraint can then be enforced in OPF. This can be useful to define maximum exchange capacities between countries.

Résumé

Introduction

La simulation des processus complexes dans des réseaux de grande échelle nécessite la réduction de la dimension du problème. Historiquement, la réduction était souvent temporelle car le point le plus critique du système était traditionnellement l'heure de pointe hiver. L'arrivée des énergies renouvelables pose différents challenges pour la gestion du système, et notamment la réduction de l'échelle temporelle car de nouvelles situations de stress peuvent avoir lieu au delà de l'heure de pointe hiver. Par exemple, ces problèmes sont particulièrement visibles dans le développement de réseau, utilisé pour l'élaboration du TYNDP ou pour des études comme la *bidding zone review*.

Comment réduire la complexité spatiale d'un réseau de grande dimension en gardant un bon niveau de précision ? Pour répondre à cette question nous avons divisé ce travail dans trois grandes étapes : 1) agrégation des noeuds; 2) modélisation des liaisons entre eux et 3) calcul des capacités des lignes équivalentes. Les approches présentées dans ce travail auront comme objectif la définition d'un modèle permettant de réaliser des études économiques prospectives.

Aggrégation des noeuds

L'agrégation des noeuds dans un cluster implique que celui-ci sera traité comme une plaque de cuivre par le modèle de marché. En conséquence, pour l'agrégation des noeuds, nous cherchons à trouver des congestions récurrentes dans le réseau et idéalement les placer toutes aux frontières des clusters.

Plusieurs travaux dans la littérature traitent le problème de définir des zones par rapport à des congestions du réseau. Différentes approches sont considérées en utilisant des indicateurs plutôt physiques, comme les *Power Transfer Distribution Factor* (PTDF),

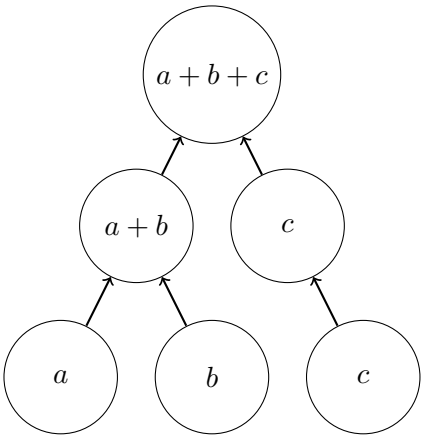

FIGURE B.1: Répresentation graphique de la méthode de clustering hierarchical.

ou économiques, comme les prix nodaux. Dans ces différents travaux, aucune méthode de comparaison est proposée pour évaluer la pertinence de l'agrégation.

Ayant comme référence la réduction pour des études économiques, nous proposons un cadre de comparaison pour évaluer différentes méthodologies de clustering. Parmi ces approches, les 3 plus prometteuses (Hierarchical, k-means et k-medoids) on été testés et meurs performances comparées par rapport au nombre des zones définies. Les résultats obtenus montrent que les différentes méthodes ont des performances similaires, réduisant significativement l'erreur quand le nombre de zones augmente. La méthode Hierarchical présente une réduction de l'erreur plus importante pour un même nombre de zones. Figure B.1 illustre d'une agrégation utilisant la méthode Hierarchical.

Modélisation des liaisons entre clusters

Un des critères d'une bonne réduction est d'avoir la même répartition des flux du réseau complet dans le réseau réduit. Cela veut dire que, pour la même position nette des clusters, on doit retrouver les mêmes échanges entre clusters que ceux observés dans le modèle complet. Dans la littérature, il est proposé de représenter le système réduit en utilisant une matrice PTDF. Ces méthodes présentent des limitations, notamment la matrice est calculée pour un seul point de fonctionnement du système et des erreurs importantes peuvent exister quand le système change. Ce problème a été mis en évidence dans la littérature, où des auteurs montrent que le calcul d'une matrice de PTDF pour un système réduit est équivalent au calcul des PTDF zonaux. En effet, la définition des Generation Shift Keys (GSKs) est un paramètre clé qui varie dans le temps. Figure B.2

illustre ce problème, en démontrant comme la variation de la production à l'intérieur d'un cluster peut impacter les échanges de ce cluster.

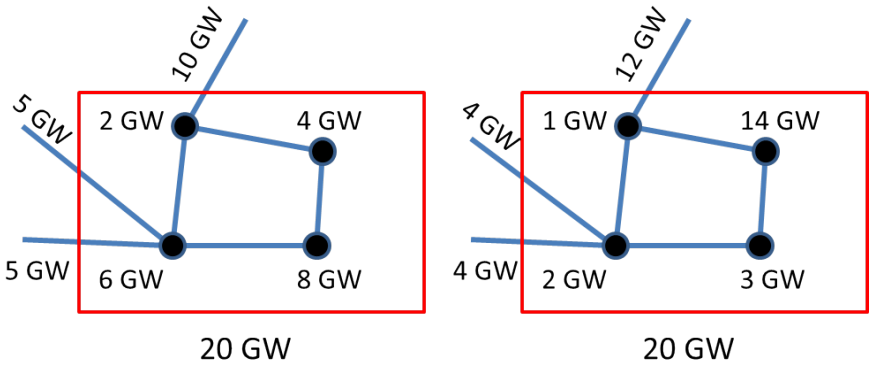

FIGURE B.2: Example de la variation de la production à l'intérieur d'un cluster.

Pour résoudre ce problème, nous proposons une méthodologie pour estimer une matrice PTDF sous forme d'un problème d'optimisation qui minimise l'erreur quadratique moyenne des différences de flux entre le modèle agrégé et le modèle complet. Nous prenons aussi en compte les flux de bouclage qui peuvent avoir lieu suite au changement des groupes de production à l'intérieur de chaque cluster. Pour cela, les flux de bouclage interviennent comme une variable de notre problème d'optimisation. Cette modélisation se révèle efficace: les flux dans les lignes obtenus pour différents points de fonctionnement du système sont proches des flux dans le modèle complet. Figure B.3 illustre la méthode de réduction proposée, où l'ensemble des lignes connectant un clusters sont réduites à une seule.

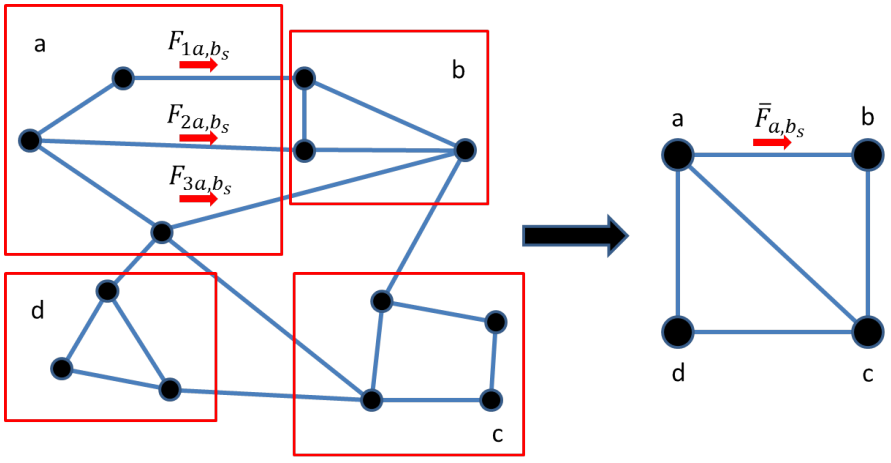

FIGURE B.3: Illustrative example of the performance assessment for a reduction approach.

Estimation des capacités équivalentes

Historiquement, la majorité des méthodologies de réduction étaient utilisées pour des calculs de transit de puissance. Les études économiques utilisant le transit de puissance optimale (OPF), ne prenait pas en compte les modèles de réseau. De ce fait, il y a peu de travaux sur le calcul des capacités équivalentes des lignes. Dans ceux-ci, différents auteurs considèrent que la capacité équivalente des lignes est la somme de la capacité thermique des lignes agrégées, ce qui peut conduire à des erreurs car les lois de Kirchhoff peuvent impacter la capacité totale disponible.

Pour résoudre ce problème, nous proposons une méthodologie qui estime les capacités équivalentes à partir des points de fonctionnement historiques du système complet. En basant notre optimisation sur plusieurs scénarios représentatifs des différents points de fonctionnement du système, nous arrivons à capter les différentes contraintes internes et externes à chaque cluster qui peuvent influencer les limites d'échanges entre les clusters. La Figure B.4 présente une illustration graphique de la méthode proposée, où les contraintes du domaine initial (en pointillée) sont ajustées pour inclure les points de fonctionnement historiques du système (points en rouge).

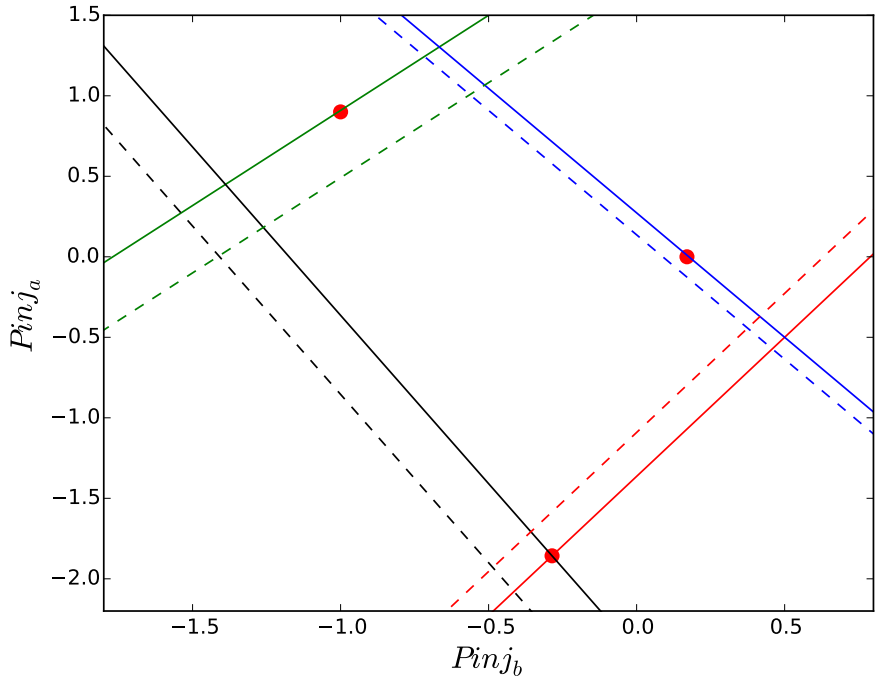

FIGURE B.4: Représentation de la méthodologie dans un domaine d'optimisation à 2D dimension entre la zone a et b. Les points rouges représentent les points de fonctionnement historiques du système.

Cette approche se distingue par la prise en compte de l'ensemble des contraintes du système, y compris les flux de bouclage, et non pas seulement la capacité thermique des

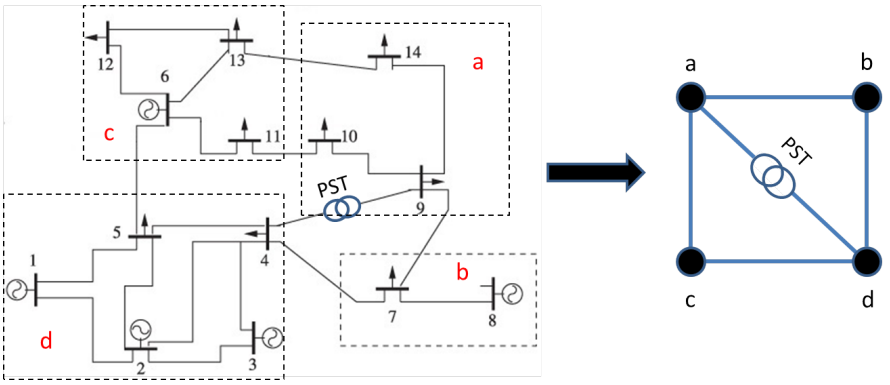

FIGURE B.5: Exemple de l'approche de réduction avec un TD.

lignes.

Impact des TDs sur les réseaux réduits

Pour faire face à l'augmentation de l'incertitude dans les systèmes électriques, les Gestionnaires du Réseau de Transport (GRTs) comptent de plus en plus sur des dispositifs de contrôle du flux de puissance comme des lignes haute tension à courant continu (HVDC) et des transformateurs déphaseurs (TDs). Compte tenu de cela, nous nous demandons si les modèles réduits de réseau doivent être mise à jour pour représenter explicitement des dispositifs. Ici, nous considérons uniquement l'effet des TDs sur le modèle réduit.

Les approches présentes dans la littérature modélisent les TDs comme une impédance variable en série avec un transformateur, ou en utilisant une matrice phase-shifter distribution factor (PSDF) qui établit une relation entre la puissance injectée et sa distribution sur les autres éléments du réseau. Une limitation pour ce genre d'approche est la nécessité de connaitre les paramètres physiques du système, notamment les impédances des lignes et les angles des noeuds. Cela peut être difficile lorsqu'il s'agit des modèles réduits souvent représentés par des matrices PTDF.

Pour résoudre ce problème, nous proposons une méthodologie pour représenter et évaluer l'impact des TDs dans les modèles réduits de réseau. Nous considérons le cas d'un système représenté seulement par une matrice PTDF avec un TD installé sur la ligne connectant les noeuds a et d, comme illustré en Figure B.5, et nous cherchons à caractériser l'impact d'un changement de prise de ce TD sur le flux original $F^0_{a,d}$ et le rapport de flux impactant la ligne (c,d), $TD_{(a,d)\rightarrow(c,d)}$.

$$TD_{(a,d)\to(c,d)} = \frac{\Delta F_{(c,d)}}{F^0_{(a,d)}} \tag{B.1}$$

Le TD induit un flux $\Delta P(\delta)$ dans la ligne (a, d), qui dépend de l'angle. En conséquence, le flux $F^1_{(a,d)}$ dans la ligne s'écrit :

$$F^1_{(a,d)} = F^0_{(a,d)} + PTDF^{(a,d)}_{(a,d)} \times \Delta P(\delta) \tag{B.2}$$

Si nous considérons que dans un cas extrême, le nouveau flux équivaut à l'injection par le PST ($F^1 = \Delta P$). L'Équation B.2 devient :

$$\Delta P = \frac{F^0_{(a,d)}}{1 - PTDF^{(a,d)}_{(a,d)}} \tag{B.3}$$

Sachant que le flux dans la ligne (c, d) avant le changement du TD était :

$$\Delta F_{(c,d)} = PTDF^{(a,d)}_{(c,d)} \times \Delta P \tag{B.4}$$

Par conséquent, l'impact du TD peut être calculé comme :

$$TD_{(a,d)\to(c,d)} = \frac{PTDF^{(a,d)}_{(c,d)}}{1 - PTDF^{(a,d)}_{(a,d)}} \tag{B.5}$$

Pour toutes les lignes L du système, le nouveau flux F^1 peut être calculé comme :

$$F^1_l = F^0_l - TD_{PST\to l} \times \Delta P_{PST} \;\; \forall l \in L \tag{B.6}$$

Ceci est vrai à l'exception de la ligne où le PST est installé. Son impact peut être représenté comme :

$$F^1_l = F^0_l \times -\Delta P_{PST} \;\; \forall l \in L \tag{B.7}$$

Cette méthodologie peut être importante dans l'évaluation des erreurs des études prospectives de l'opération et du système électrique, car l'impact des TDs peut changer la validité des résultats obtenus.

Mots-clés : Marchés électriques, Réseau de transport, Réduction de réseau, Agrégation de réseau.

Bibliography

[1] A. Moreira, G. Strbac, R. Moreno, A. Street, and I. Konstantelos, "A five-level milp model for flexible transmission network planning under uncertainty: A min-max regret approach," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 486–501, Jan 2018.

[2] K. Bell, "Methods and tools for planning the future power system: Issues and priorities," *IET Special Interest Publication for the Council for Science and Technology on "Modelling Requirements of the GB Power System Resilience during the transition to Low Carbon Energy"*, 2015.

[3] P. J. Ramírez, D. Papadaskalopoulos, and G. Strbac, "Co-optimization of generation expansion planning and electric vehicles flexibility," *IEEE Transactions on Smart Grid*, vol. 7, no. 3, pp. 1609–1619, May 2016.

[4] S. Mathieu, M. Petitet, M. Perrot, D. Ernst, and Y. Phulpin, "Sistem, a model for the simulation of short-term electricity markets," *CEEM Working Paper no30*, 2017.

[5] e–HIGHWAY 2050, "D 8.6a - detailed enhanced methodology for long-term grid planning," *Technical Report*, November 2015.

[6] S. Lumbreras and A. Ramos, "The new challenges to transmission expansion planning. survey of recent practice and literature review," *Electric Power Systems Research*, vol. 134, pp. 19 – 29, 2016.

[7] E. Shayesteh, B. F. Hobbs, and M. Amelin, "Scenario reduction, network aggregation, and dc linearisation: which simplifications matter most in operations and planning optimisation?" *IET Generation, Transmission Distribution*, vol. 10, no. 11, pp. 2748–2755, 2016.

[8] R. Bent, G. L. Toole, and A. Berscheid, "Transmission network expansion planning with complex power flow models," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 904–912, May 2012.

[9] R. A. Jabr, "Optimization of ac transmission system planning," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2779–2787, Aug 2013.

[10] C. Coffrin, H. L. Hijazi, and P. V. Hentenryck, "The qc relaxation: A theoretical and computational study on optimal power flow," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 3008–3018, July 2016.

[11] Y. Ji and B. F. Hobbs, "Including a dc network approximation in a multiarea probabilistic production costing model," *IEEE Transactions on Power Systems*, vol. 13, no. 3, pp. 1121–1127, Aug 1998.

[12] Q. P. Zheng, J. Wang, and A. L. Liu, "Stochastic optimization for unit commitment - a review," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1913–1924, July 2015.

[13] J. Sumaili, H. Keko, V. Miranda, A. Botterud, and J. Wang, "Clustering-based wind power scenario reduction technique," *17th Power Systems Computation Conference(PSCC)*, August 2011.

[14] D. Lee and R. Baldick, "Load and wind power scenario generation through the generalized dynamic factor model," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 400–410, Jan 2017.

[15] C. Safta, R. L. Y. Chen, H. N. Najm, A. Pinar, and J. P. Watson, "Efficient uncertainty quantification in stochastic economic dispatch," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2535–2546, July 2017.

[16] Q. Wang, J. Wang, and Y. Guan, "Stochastic unit commitment with uncertain demand response," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 562–563, Feb 2013.

[17] A. Moshari, A. Ebrahimi, and M. Fotuhi-Firuzabad, "Short-term impacts of dr programs on reliability of wind integrated power systems considering demand-side uncertainties," *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2481–2490, May 2016.

[18] S. R. Khuntia, J. L. Rueda, and M. A. M. M. van der Meijden, "Forecasting the load of electrical power systems in mid- and long-term horizons: a review," *IET Generation, Transmission Distribution*, vol. 10, no. 16, pp. 3971–3977, 2016.

[19] Q. Ploussard, L. Olmos, and A. Ramos, "An operational state aggregation technique for transmission expansion planning based on line benefits," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2744–2755, July 2017.

[20] C. Zhao and R. Jiang, "Distributionally robust contingency-constrained unit commitment," *IEEE Transactions on Power Systems*, vol. 33, no. 1, pp. 94–102, Jan 2018.

[21] J. B. Ward, "Equivalent circuits for power-flow studies," *Transactions of the American Institute of Electrical Engineers*, vol. 68, no. 1, pp. 373–382, July 1949.

[22] P. Dimo, *Nodal analysis of power systems*, ser. Abacus Bks. Editura Academiei Republicii Socialisté România, 1975.

[23] Y. Zhu, D. Tylavsky, and S. Rao, "Nonlinear structure-preserving network reduction using holomorphic embedding," *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–1, 2017.

[24] E. H. Allen, J. H. Lang, and M. D. Ilic, "A combined equivalenced-electric, economic, and market representation of the northeastern power coordinating council u.s. electric power system," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 896–907, Aug 2008.

[25] D. Shi and D. J. Tylavsky, "An improved bus aggregation technique for generating network equivalents," in *2012 IEEE Power and Energy Society General Meeting*, July 2012, pp. 1–8.

[26] Y. Zhu and D. Tylavsky, "An optimization-based dc-network reduction method," *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–1, 2017.

[27] E. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[28] X. Cheng and T. J. Overbye, "Ptdf-based power system equivalents," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 1868–1876, Nov 2005.

[29] H. Oh, "A new network reduction methodology for power system planning studies," *Power Systems, IEEE Transactions on*, vol. 25, no. 2, pp. 677–684, May 2010.

[30] C.-A. Fezeu, K. Bell, J. Ding, P. Panciatici, and M.-S. Debry, "Simplified representation of a large transmission network for use in long-term expansion planning," in *Power Systems Computation Conference (PSCC), 2014*, Aug 2014, pp. 1–7.

[31] "ENTSO-E transmission system map," https://www.entsoe.eu/data/map/, accessed: 2018-03-12.

[32] "ENTSO-E Common Grid Model," https://www.entsoe.eu/major-projects/common-information-model-cim/cim-for-grid-models-exchange/standards/Pages/default.aspx, accessed: 2016-05-26.

[33] "PowerGAMA Europe 2014 model," https://zenodo.org/record/54580, accessed: 2016-05-26.

[34] ENTSO-E, "10-year network development plan," 2014.

[35] "ENTSO-E Transparency Platform," https://transparency.entsoe.eu/, accessed: 2016-05-26.

[36] "PLATTS World Electric Power Plants Database," http://www.platts.com/products/world-electric-power-plants-database, accessed: 2016-05-26.

[37] N. Marinho, Y. Phulpin, D. Folliot, and M. Hennebel, "Approaching generation variable costs from publicly available data," in *2016 13th International Conference on the European Energy Market (EEM)*, June 2016.

[38] S. Stoft, *Power System Economics*. Wiley-IEEE Press, 2002.

[39] DENA, "Integration of Renewable Energy Sources in the German Power Supply System," Tech. Rep., 11 2010.

[40] European Commission, "Commission Regulation (EU) 2015/1222 of 24 July 2015 establishing a guideline on capacity allocation and congestion management," July 2015.

[41] IEA, "Projected costs of generating electricity–2015 edition," *Technical Report*, 2015.

[42] "NREL transparent cost database," http://en.openei.org/apps/TCDB/, accessed: 2015-09-30.

[43] e–HIGHWAY 2050, "D 2.2 - european cluster model of the pan-european transmission grid," *Technical Report*, August 2014.

[44] Fraunhofer, "Levelized costs of electricity renewable energy technologies," *Technical Report*, 2013.

[45] G. A. Marrero and F. J. Ramos-Real, "Electricity generation cost in isolated system: the complementarities of natural gas and renewables in the canary islands," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 9, pp. 2808–2818, 2010.

[46] "Dig Silent Power Factory," http://www.digsilent.de/index.php/products-powerfactory.html, accessed: 2016-05-26.

[47] L. Olmos and I. J. Pérez-Arriaga, "Definition of single price areas within a regional electricity system," in *16th Power Systems Computation Conference PSCC*, vol. 8, 2008.

[48] B. Burstedde, "From nodal to zonal pricing: A bottom-up approach to the second-best," in *European Energy Market (EEM), 2012 9th International Conference on the*, May 2012, pp. 1–8.

[49] C. Breuer, N. Seeger, and A. Moser, "Determination of alternative bidding areas based on a full nodal pricing approach," in *Power and Energy Society General Meeting (PES), 2013 IEEE*, July 2013, pp. 1–5.

[50] H. Yang and R. Zhou, "Monte carlo simulation based price zone partitioning considering market uncertainty," in *Probabilistic Methods Applied to Power Systems, 2006. PMAPS 2006. International Conference on*, June 2006, pp. 1–5.

[51] S. Lumbreras, A. Ramos, L. Olmos, F. Echavarren, F. Banez-Chicharro, M. Rivier, P. Panciatici, J. Maeght, and C. Pache, "Network partition based on critical branches for large-scale transmission expansion planning," in *PowerTech, 2015 IEEE Eindhoven*, June 2015, pp. 1–6.

[52] H. Oh, "Aggregation of Buses for a Network Reduction," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 705–712, May 2012.

[53] D. Shi and D. Tylavsky, "A novel bus-aggregation-based structure-preserving power system equivalent," *Power Systems, IEEE Transactions on*, vol. 30, no. 4, pp. 1977–1986, July 2015.

[54] M. Imran and J. Bialek, "Effectiveness of zonal congestion management in the european electricity market," in *Power and Energy Conference, 2008. PECon 2008. IEEE 2nd International*, Dec 2008, pp. 7–12.

[55] S. Lloyd, "Least squares quantization in pcm," *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982.

[56] P. N. Biskas, D. I. Chatzigiannis, G. A. Dourbois, and A. G. Bakirtzis, "European market integration with mixed network representation schemes," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4957–4967, Nov 2013.

[57] S. Firdaus and M. A. Uddin, "A survey on clustering algorithms and complexity analysis," *International Journal of Computer Science Issues (IJCSI)*, vol. 12, no. 2, p. 62, 2015.

[58] N. Marinho, Y. Phulpin, D. Folliot, and M. Hennebel, "Redispatch index for assessing bidding zone delineation," *IET Generation, Transmission & Distribution*, vol. 11, pp. 4248–4255(7), November 2017.

[59] J. Schwippe, A. Seack, and C. Rehtanz, "Pan-european market and network simulation model," in *PowerTech (POWERTECH), 2013 IEEE Grenoble*. IEEE, 2013, pp. 1–6.

[60] ENTSO-E, "Generation and load shift key implementation guide," ENTSO-E, Tech. Rep., September 2016.

[61] THEMA, "Loop flows–final advice," European Comission, Tech. Rep., October 2013.

[62] M. A. Laughton and M. W. H. Davies, "Numerical techniques in solution of power-system load-flow problems," *Electrical Engineers, Proceedings of the Institution of*, vol. 111, no. 9, pp. 1575–1588, September 1964.

[63] B. Stott, "Review of load-flow calculation methods," *Proceedings of the IEEE*, vol. 62, no. 7, pp. 916–929, July 1974.

[64] M. Doquet, "Zonal reduction of large power systems: Assessment of an optimal grid model accounting for loop flows," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 503–512, Jan 2015.

[65] H. G. Svendsen, "Grid model reduction for large scale renewable energy integration analyses," *Energy Procedia*, vol. 80, pp. 349 – 356, 2015, 12th Deep Sea Offshore Wind R&D Conference, EERA DeepWind'2015.

[66] D. Shi, *Power system network reduction for engineering and economic analysis*. Arizona State University, 2012.

[67] L. Pagnier and P. Jacquod, "A predictive pan-european economic and production dispatch model for the energy transition in the electricity sector," in *2017 IEEE Manchester PowerTech*, June 2017, pp. 1–6.

[68] E. Shayesteh, B. F. Hobbs, L. Söder, and M. Amelin, "Atc-based system reduction for planning power systems with correlated wind and loads," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 429–438, Jan 2015.

[69] S. Mohapatra, W. Jang, and T. J. Overbye, "Equivalent line limit calculation for power system equivalent networks," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2338–2346, Sept 2014.

[70] S. Lumbreras, A. Ramos, F. Banez-Chicharro, L. Olmos, P. Panciatici, C. Pache, and J. Maeght, "Large-scale transmission expansion planning: from zonal results to a nodal expansion plan," *IET Generation, Transmission Distribution*, vol. 11, no. 11, pp. 2778–2786, 2017.

[71] M. Belivanis and K. R. W. Bell, "Coordination of phase-shifting transformers to improve transmission network utilisation," in *2010 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)*, Oct 2010, pp. 1–6.

[72] L. Wang, M. Klein, S. Yirga, and P. Kundur, "Dynamic reduction of large power systems for stability studies," *Power Systems, IEEE Transactions on*, vol. 12, no. 2, pp. 889–895, May 1997.

[73] ENTSOE–E, "Phase shift transformers modelling," *Technical Report*, May 2014.

[74] K. R. W. Bell, A. R. Daniels, and R. W. Dunn, "Modelling of operator heuristics in dispatch for security enhancement," *IEEE Transactions on Power Systems*, vol. 14, no. 3, pp. 1107–1113, Aug 1999.

[75] D. V. Hertem, "The use of power flow controlling devices in the liberalized market," *PhD dissertation, KU Leuven, Leuven, Belgium*, January 2009.

[76] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, Feb 2011.

[77] I. Staffell and R. Green, "Is there still merit in the merit order stack? the impact of dynamic constraints on optimal plant mix," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 43–53, Jan 2016.

[78] W. W. Hogan, "Electricity market design and efficient pricing: Applications for new england and beyond," *The Electricity Journal*, vol. 27, no. 7, pp. 23–49, 2014.

[79] "ENTSO-E transparency platform," https://transparency.entsoe.eu/, accessed: 2015-09-30.

[80] "Transparency in energy markets," http://www.eex-transparency.com/homepage, Accessed: 2015-09-30.

[81] "Eco2mix," http://www.rte-france.com/fr/eco2mix/eco2mix, accessed: 2015-09-30.

[82] RTE, "Bilan électrique 2014," *Technical Report*, 2015.

* * *

université PARIS-SACLAY
ÉCOLE DOCTORALE
Physique et ingénierie : électrons, photons, sciences du vivant (EOBE)

Titre: Réduction d'un modèle de système électrique pour des études technico-économiques

Mots clés: Marché d'électricité, Réduction de réseau

Résumé: La simulation des processus complexes dans des réseaux de transport d'électricité de grande taille nécessite la réduction de la dimension du problème. Comment réduire la complexité spatiale d'un réseau de grande dimension en gardant un bon niveau de précision ? Pour répondre à cette question nous avons divisé ce travail en trois grandes étapes : 1) la réduction par agrégation du nombre de noeuds; 2) la modélisation des liaisons entre ces clusters de noeuds et 3) le calcul des capacités des lignes équivalentes.
L'agrégation des noeuds dans un cluster implique que celui-ci sera traité comme une plaque de cuivre par le modèle de marché. En conséquence, pour l'agrégation des noeuds, les congestions récurrentes dans le réseau sont identifiées et placées idéalement aux frontières des clusters. Après la réduction, la même répartition des flux dans le réseau complet et dans le modèle réduit du réseau doit être trouvée. Pour ce fait une méthodologie d'estimation d'une matrice PTDF a été développée. Pour les études économiques la limite thermique des lignes est un paramètre clé. Pour résoudre ce problème, nous proposons une méthodologie qui estime les capacités équivalentes à partir des points de fonctionnement historiques du système complet.
Les approches présentées dans ce travail ont été appliquées sur un modèle du réseau continental européen et ont permis d'obtenir un modèle simplifié qui minimise la perte d'information.

Title: Reduction of an electrical power system model for techno-economic studies

Keywords: Electricity markets, Network reduction

Abstract: The simulation of complex processes in large scale power systems needs the reduction of the problem. How to reduce the spatial complexity of a large scale power network while minimizing information loss? To answer this question we have divided this work in three main steps: 1) network buses aggregation; 2) modelling of the clusters' links; 3) defining the equivalent branches maximum exchange capacity.
The bus aggregations in a cluster implies that it will be treated as a coppper-plate by the market model. Therefore, the most frequent network congestions must be identified ideally placed at the clusters frontiers. After the reduction, the same power flow repartition must be found in both reduced and complete model. To do that, a methodology to define a PTDF matrix was developed. For economic purpose studies, the branches maximum capacity is a key parameter, to define this value, a methodology is proposed that estimates the equivalent transmission capacities using historical system operating set points.
These approaches were applied to the European transmission network and allowed to define a reduced model that minimises the information loss.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France