DOCTORAL THESIS

MADRID, SPAIN 2019

Efficient reduction techniques for a large-scale Transmission Expansion Planning problem

Quentin Ploussard

Efficient reduction techniques for a large-scale Transmission Expansion Planning problem

Quentin Ploussard

Thesis supervisors:

Luis Olmos, Andrés Ramos

Members of the Examination Committee:

YYYYY

YYYYY

YYYYY

YYYYY

TRITA xxxxxxxx (will be provided by KTH)

ISBN xxxxxxxxxx (will be provided by your home university)

Printed by: [Publisher]

[OPTIONAL TEXT is allowed here]

Efficient reduction techniques for a large-scale Transmission Expansion Planning problem

Sustainable concrete infrastructure design

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus prof.dr.ir. T.H.J.J. van der Hagen
Chair of the Board for Doctorates
to be defended publicly on
weekday day month year at xx:xx o'clock

by

Quentin PLOUSSARD

Electrical Engineer,
SUPÉLEC, France
born in Aix en Provence, France

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	Chairman
Prof. dr. ir. M.P.C. Weijnen	Delft University of Technology, the Netherlands, promotor
Prof.dr. V. Leclerc	Comillas Pontifical University, Spain, promotor
Dr. ir. R.A. Hakvoort	Technische Universiteit Delft, the Netherlands, copromotor

Independent members:

Prof. dr. I.L. Hancher	Universiteit van Tilburg, the Netherlands
Prof. P. Carrion-Gordon	Comillas Pontifical University, Spain
Prof. G.E.G. Beroggi	KTH Royal Institute of Technology, Sweden
Dr. T. Kapetanovic	Energie-Control GmbH, Austria
Prof. dr. J. Groene	Delft University of Technology, the Netherlands, reserve member

The doctoral research has been carried out in the context of an agreement on joint doctoral supervision between Comillas Pontifical University, Madrid, Spain, KTH Royal Institute of Technology, Stockholm, Sweden and Delft University of Technology, the Netherlands.

Keywords: energy, electricity, regulation, European Union, market integration, interconnector investment, congestion management, market transparency, European governance, European modes of regulation, regulatory change

ISBN (will be provided by your home university)

Printed in the Netherlands

SETS Joint Doctorate

The Erasmus Mundus Joint Doctorate in ***Sustainable Energy Technologies and Strategies***, SETS Joint Doctorate, is an international programme run by six institutions in cooperation:

- Comillas Pontifical University, Madrid, Spain
- Delft University of Technology, Delft, the Netherlands
- KTH Royal Institute of Technology, Stockholm, Sweden
- Florence School of Regulation, Florence, Italy
- Johns Hopkins University, Baltimore, USA
- University Paris-Sud 11, Paris, France

The Doctoral Degrees issued upon completion of the programme are issued by Comillas Pontifical University, Delft University of Technology, and KTH Royal Institute of Technology.

The Degree Certificates are giving reference to the joint programme. The doctoral candidates are jointly supervised and must pass a joint examination procedure set up by the three institutions issuing the degrees.

This thesis is a part of the examination for the doctoral degree. The invested degrees are official in Spain, the Netherlands and Sweden respectively.

SETS Joint Doctorate was awarded the Erasmus Mundus **excellence label** by the European Commission in year 2010, and the European Commission's **Education, Audiovisual and Culture Executive Agency**, EACEA, has supported the funding of this programme.

The EACEA is not to be held responsible for contents of the thesis.

(Acknowledgement)

"No problem can be solved until it is reduced to some simple form. The changing of a vague difficulty into a specific, concrete form is a very essential element in thinking."
J. P. Morgan, Banker

"All models are wrong, but some are useful."
George E. P. Box, Mathematician

Table of Contents

NOMENCLATURE xvi

1. Introduction 1
   - 1.1. Background 1
   - 1.2. Research motivation 2
   - 1.3. Thesis objectives and research methodology 3
   - 1.4. Thesis outline and organization 4
2. Literature review 6
   - 2.1. Snapshot selection techniques 6
   - 2.2. Network reduction techniques 7
     - 2.2.1. Network partition methods and similarity measures 8
     - 2.2.2. Bus aggregation methods 9
     - 2.2.3. Equivalencing techniques 9
     - 2.2.4. Network reduction in a TEP context 10
   - 2.3. Search space reduction techniques 12
3. TEP model assumptions and general reduction methodology 15
   - 3.1. Network representation in TEP models 15
   - 3.2. Programming aspects of the TEP problem 17
   - 3.3. TEP model considered in this thesis 18
   - 3.4. General reduction methodology 21
     - 3.4.1. Problem size reduction and information theory 21
     - 3.4.2. Linear relaxation of the TEP problem 22

3.4.3. General input data reduction methodology based on the relaxed version of the TEP problem ... 24

4. Snapshot selection for a TEP problem ... 25

4.1. Preliminary discussion ... 25

4.2. Methodology ... 26

4.2.1. Relaxed TEP problem and related optimal investments ... 26

4.2.2. Computation of line benefits for each operation situation ... 28

4.2.3. Dimension reduction of the line benefit space ... 30

4.2.4. Clustering algorithm ... 31

4.3. Case study ... 32

4.3.1. Systems description ... 32

4.3.2. Application of the methodology ... 34

4.3.3. Validation of results ... 36

5. Network reduction for a TEP problem ... 42

5.1. Preliminary discussion ... 42

5.2. Methodology ... 42

5.2.1. Identification of the critical pairs of buses ... 43

5.2.2. Network partition ... 44

5.2.3. Bus elimination ... 48

5.2.4. Equivalencing ... 49

5.3. Case study ... 50

5.3.1. Systems description ... 50

5.3.2. Application of the methodology ... 51

5.3.3. Validation of results ... 53

5.4. Comparison with alternative methods ... 55

5.4.1. Network partition based on alternative metrics ... 55

5.4.2. Bus aggregation method based on PTDF matrix 57

6. Search space reduction for a TEP problem 58

6.1. Preliminary discussion 58

6.1.1. Preliminary candidate lines filter 58

6.1.2. Efficient search space reduction technique in a TEP context 59

6.2. Method 60

6.2.1. Relaxed TEP problem with an unbounded number of candidate lines per corridor 60

6.2.2. Relaxed TEP problem with bounded number of candidate lines per corridor 63

6.2.3. Binary representation of candidate lines 65

6.3. Case study 66

6.3.1. System description 66

6.3.2. Application of the methodology 67

6.3.3. Validation of results 69

7. Conclusion 73

7.1. Contributions and conclusion 73

7.2. Combination of the presented reduction methods 75

7.3. further works 76

Bibliography 77

CURRICULUM VITAE 85

LIST OF RELEVANT PUBLICATIONS 86

NOMENCLATURE

List of abbreviations

AC	Alternative Current
ACLF	AC Load Flow
ATC	Available Transfer Capacity
BB	Branch and Bound
BC	Branch and Cut
BNS	Binary Numerical System
CHA	Constructive Heuristic Algorithm
CO	Combinatorial Optimization
CP	"Copper Plate"
DC	Direct Current
DCLF	DC Load Flow
ENS	Energy Non-Served
ENTSO-E	European Network of Transmission System Operators for Electricity
GEP	Generation Expansion Planning
HEPS	Hydroelectric pump-storage
HVDC	High Voltage Direct Current
KCL	Kirchhoff's Current Law
KVL	Kirchhoff's Voltage Law
LP	Linear Programming
MILP	Mixed Integer Linear Programming
MINLP	Mixed Integer Non-Linear Programming
NLP	Non-Linear Programming
NP	Nondeterministic Polynomial time
OPF	Optimal Power Flows
PCA	Principal Component Analysis
PINT	"Put IN one at a Time"
PNS	Power Non-Served
PT	Power Transfer
PTDF	Power Transfer Distribution Factor
REI	Radial, Equivalent, Independent
RES	Renewable Energy Source
RTS	Reliability Test System
SP	Shortest Path
TEP	Transmission Expansion Planning
TEPUNCL	TEP problem with unbounded number of candidate lines per corridor
TEPBNCL	TEP problem with bounded number of candidate lines per corridor
TLF	Transportation Load Flow
TOOT	"Take Out One at a Time"
UC	Unit Commitment

Indices & Sets

i, j, k, l	Node
(i, j)	Corridor (pair of nodes)
(i, j, c)	Triplet identifying a unique circuit (AC or DC, existing or candidate) between two nodes i and j
s	Snapshot
g	Generator

k	Target number of clusters
AC	Set of AC circuits
CC	Set of candidate circuits
EC	Set of existing circuits
$CONV$	Conventional generators
RES	Renewable Energy Source (RES) generators
N_B	Total number of nodes
N_S	Total number of snapshots
N_G	Total number of generators ($[1;N_G] = CONV \cup RES$)
N_{CC}	Total number of candidate circuits ($N_{CC} = card(CC)$)
(i,j)	Existing or candidate circuits between nodes *i* and *j*
Ω_i^G	Conventional or RES generators connected at node *i*

Parameters

Y_{ijc}	Admittance of AC circuit (i,j,c) (p.u.)
$Y_{i,j}$	Equivalent admittance of corridor (i,j) in the initial network (p.u.)
$Y_{i,j,1}^{cand}$	Admittance of a single candidate AC circuit in corridor (i,j) (p.u.)
cg_g	Generation cost of generator *g* (€/MWh)
$d_{i,s}$	Demand level at node *i* during snapshot *s* (MW)
$\overline{f_{ijc}}$	Maximum capacity of circuit (i,j,c) (MW)
$\overline{f_{i,j}}$	Equivalent maximum capacity of corridor (i,j) in the initial network (MW)
$\overline{f_{i,j,1}^{cand}}$	Maximum capacity of a single candidate AC circuit in corridor (i,j) (MW)
$\overline{p_{g,s}}$	Generation capacity of RES generator *g* during snapshot *s* (MW)
$\overline{p_g}$	Maximum generation capacity of conventional generator *g* (MW)
$av_{g,s}$	Availability state of a conventional generator *g* during snapshot *s* {0;1}
ρ_s	Weight of snapshot *s* (h)
C^{ENS}	Cost of energy non-served (€/MWh)
C_{ijc}	Investment cost of candidate circuit (i,j,c) (€)
$C_{i,j,1}$	Investment cost of a single candidate AC circuit in corridor (i,j) (€)
S_B	Base power (MW)
γ	Fix annual charge rate (p.u.)
M_{ijc}	Big-M parameter of candidate circuit (i,j,c) (MW)

Variables

x_{ijc}	Decision to invest or not in candidate circuit (i,j,c) (p.u.)
x_{ij}	Number of candidate circuits installed in corridor (i,j) (p.u.)
$p_{g,s}$	Power production of generator *g* during snapshot *s* (MW)
$pns_{i,s}$	Power non-served at node *i* during snapshot *s* (MW)
$\theta_{i,s}$	Voltage angle at node *i* during snapshot *s* (p.u.)
$f_{ijc,s}$	Power flow through circuit (i,j,c) during snapshot *s* (MW)
$f_{i,j,s}^{cand}$	Total power flow through of a set of candidate circuits installed in corridor (i,j) during snapshot *s* (MW)
$Y_{i,j}^{cand}$	Equivalent admittance of a set of candidate circuits installed in corridor (i,j) (p.u.)

$\overline{f_{i,j}^{cand}}$ Equivalent maximum capacity of a set of candidate circuits installed in corridor (i,j) (MW)

1. Introduction

1.1. Background

The aim of Transmission Expansion Planning (TEP) studies is to decide which, where, and when new grid elements should be built in order to minimize the total system cost. This problem has been widely explored in the last decades. However, the recent policy objectives and regulation applied in the electricity sector and the new technologies introduced in modern power systems have resulted in an increase in the complexity and the size of TEP problems.

Identifying the most efficient investment decisions, i.e. the ones leading to minimum total costs, requires adequately representing within the TEP problem several aspects of the power system, its functioning, and its evolution, such as:

- the structure of the existing electrical network,
- the spatial and temporal distribution of the power system load and energy sources,
- the set of candidate grid elements that may be installed.

When attempting to solve the TEP problem with a reasonable level of accuracy, each of these aspects of the power system functioning and development should be appropriately represented.

In traditional power systems, TEP studies have been carried out at a national level, limiting the size of the network to consider to hundreds of buses (Gallego, Alves, Monticelli, & Romero, 1997). The power consumption has been the main source of temporal variability. The load hourly profile is usually well represented by a couple of operating hours (for instance peak and off-peak hours in summer and winter) (Boîteux, 1949). Candidate grid elements have usually been limited to a couple of AC lines to potentially install in existing or new corridors suggested in advance by the planner.

On the other hand, in modern power systems, the new policy objectives drive the installation of large amounts of RES-based generation, which largely increases the sources of temporal variability in the power system and, thus, the number of operational states to consider in TEP studies (Desta Z. Fitiwi, de Cuadra, Olmos, & Rivier, 2015). For instance, solving a static TEP problem with a fully accurate, hourly detailed, temporal variability, requires to model the entire target year with 8760 operational states. The recent trend of integrating the operation and planning of the development of grids in neighboring systems, through some form of coordination processes involving System and Market Operators, or planning authorities in general, has resulted in a significant increase in the size of the network to consider in TEP studies (Lee, Ng, Zhong, & Wu, 2006). Partly due to this, the number of candidate lines to consider in these studies has also increased substantially. The latter is also related to the fact that the number of technological choices for transmission upgrades has increased with the possible deployment of HVDC and other power flow controlling devices (Sara Lumbreras & Ramos, 2016).

For all these reasons, the size of the TEP problem has drastically increased in several dimensions: the representation made of the grid (spatial dimension), the representation made of the relevant operation situations (temporal representation), and the number of candidate grid elements to consider. (Fig. 1.1)

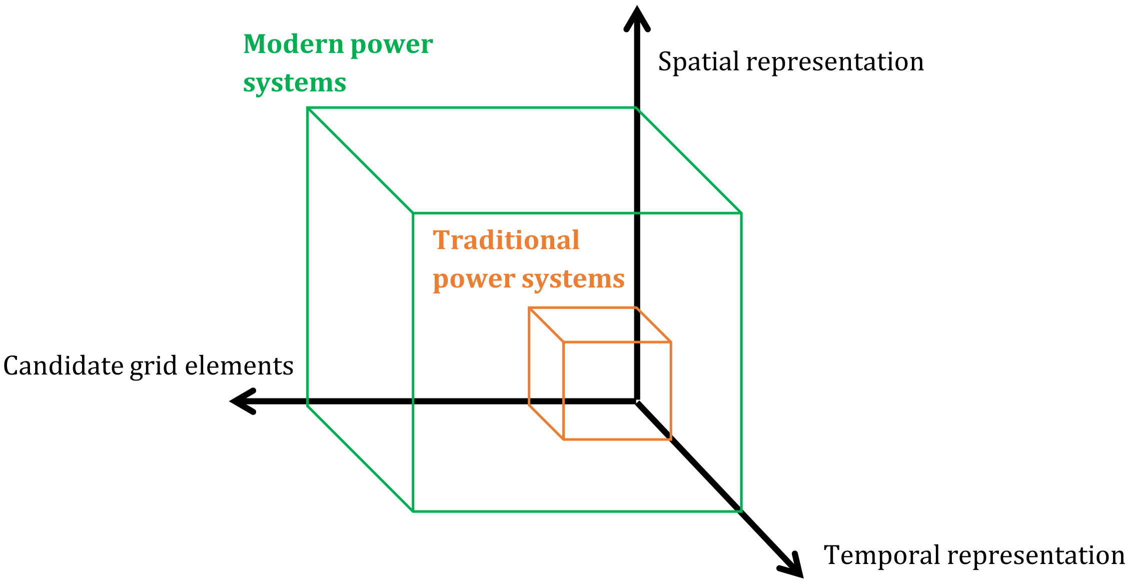

Fig. 1.1 Traditional and modern power system representation when solving the TEP problem

1.2. RESEARCH MOTIVATION

The computation time required to solve the TEP problem increases with the size of this problem. This raises the need for methods able to reduce the size of the TEP problem in all the aforementioned dimensions while preserving crucial information necessary to make the same optimal investment decisions.

These reduction techniques, already explored previously in power system studies, can be categorized into three dimensions (Fig. 1.2):

1. "Snapshot selection" techniques are used to reduce the set of operational states which represent the temporal variability in operation conditions in a power system,
2. "Network reduction" techniques are methods to represent the electrical network in a compact way,
3. "Search space reduction" techniques help identifying the most relevant expansion projects to take into account in TEP studies.

Even though these techniques have been widely explored in the literature, the vast majority of those already developed are based on a rule-of-thumb, while few of them are actually adapted to the TEP problem. For instance, on the one hand, snapshot selection

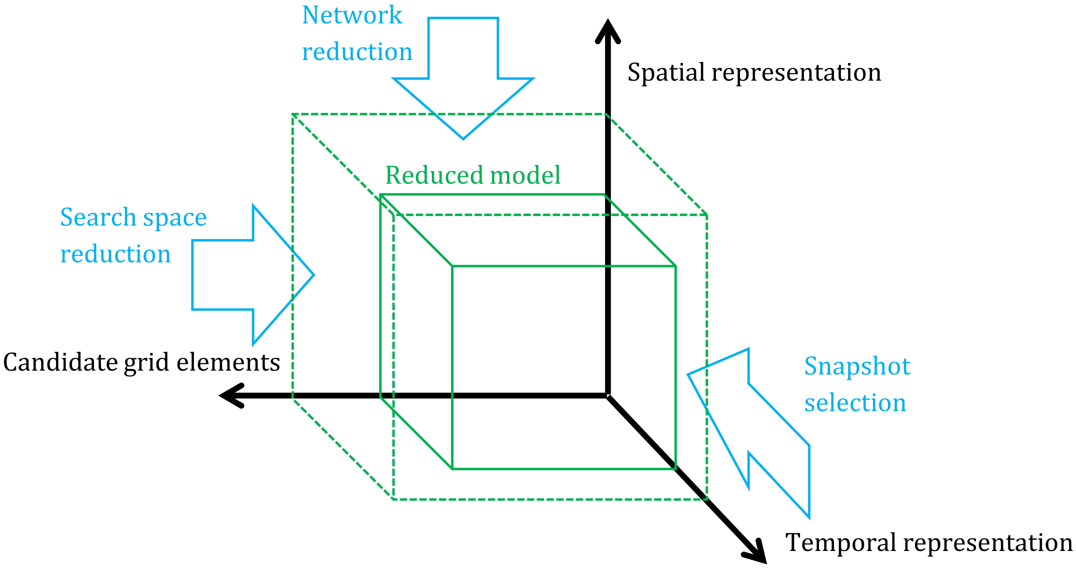

Fig. 1.2 Reduction techniques applied to reduce the size of the TEP problem

methods have traditionally been based on the selection of peak load and off-peak hours (Boîteux, 1949). However, other operation snapshots may become relevant when large amounts of intermittent renewable energy based generation are installed in the system, since the number and identity of the most stressful situations for the grid may be conditioned by the availability of the primary energy resource used by this new generation. Regarding the representation made of the grid, most of the existing network reduction methods are just based on the application of electrical distance criterion (Fezeu, Bell, Ding, Panciatici, & Debry, 2014), i.e. they do not take into account the power system operation.

Moreover, even when some reduction techniques have been specifically designed to be applied in a TEP context, their ability to produce an efficient reduced version of the TEP problem has almost never been tested.

1.3. Thesis objectives and research methodology

The objective of this thesis is to find methods to reduce the computational time of solving the TEP problem by reducing the size of its input data.

It should be noted that there are several other strategies to reduce the computational time of solving the TEP problem, such as the use of meta-heuristic methods (Rathore, Roy, Sharma, & Patel, 2013), (Torres & Castro, 2012), (Gallego et al., 1997), or the use of mathematical programming techniques (Binato, Pereira, & Granville, 2001). However, these strategies are out of the scope of this thesis, and the present research work focuses on the strategy of producing a TEP problem with a reduced size to be solved with classical optimization methods.

More specifically, the objective of this thesis is to build efficient, more suitable, reduction techniques to be applied in a TEP context. A novel reduction technique will be developed for each dimension of the TEP problem, that is:

- the set of candidate grid elements to consider,
- the representation made of the network,
- the temporal dimension, or representation made of the relevant operation situations.

With the term efficient, we refer to how the set of optimal investment solutions of the reduced TEP problem defined using the techniques here proposed compares to the set of optimal investment solutions of the complete, detailed, TEP problem. Moreover, a comparison should also be made of the TEP solution computed by reducing the TEP problem according to the techniques here proposed with the TEP solutions obtained considering a reduced version of the TEP problem defined considering alternative reduction techniques in the literature. The procedure followed to assess the efficiency of the reduction techniques proposed in this thesis work is depicted in Fig. 1.3. This procedure is followed to assess each of the reduction techniques presented in this thesis.

1.4. Thesis outline and organization

This thesis document is organized as follows:

- Chapter 2 reviews the existing literature regarding reduction techniques applied to power system studies in general, and TEP problems in particular,
- Chapter 3 describes the assumptions made in our formulation of the TEP problem and the general relaxation method used when applying all the reduction techniques proposed in this thesis,
- Chapter 4 introduces a tailor-made snapshot selection method to be used in a TEP context,
- Chapter 5 presents a network reduction method designed for TEP problems,
- Chapter 6 describes and discusses the application of a new method to reduce the space of candidate grid elements in a TEP context,
- Chapter 7 gathers the main findings of this thesis in the form of conclusions, summarizes the main contributions of the thesis by revisiting the thesis objectives, provides recommendation about the way to combine the proposed reduction techniques, and finally draws some directions for extending this thesis work.

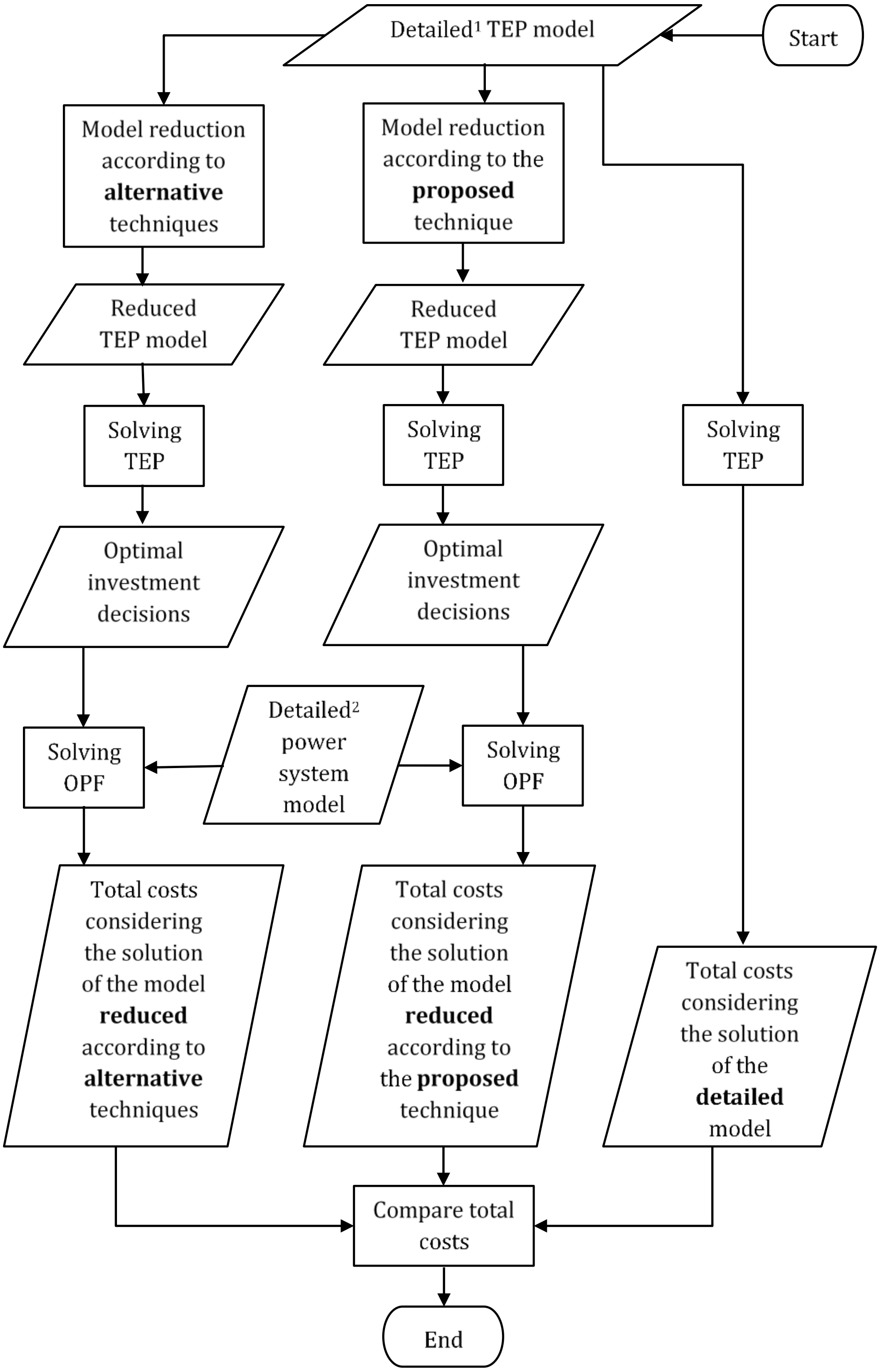

Fig. 1.3 General procedure followed to assess the efficiency of each reduction technique.

[1] With the term "detailed" model, we refer to the original, non-reduced model, regarding the dimension to be reduced in each case.

[2] The "detailed" model here is the original detailed TEP model.

2. Literature review

The TEP problem has been widely explored topic in the last decades. However, this thesis focuses only on the literature related to reduction techniques that can be applied to the TEP problem. Reduction techniques are well known in power system studies and can be categorized into three types, related to three dimensions of the TEP problem:

- "Snapshot selection" techniques are used to reduce the set of operational states which represent the temporal variability in a power system,
- "Network reduction" techniques are methods to represent the electrical network in a compact way,
- "Search space reduction" techniques help identifying the most relevant expansion projects to consider as candidates in the TEP problem.

Next, the most relevant works related to the reduction techniques of each type are discussed.

2.1. Snapshot selection techniques

Because in modern power systems, numerous sources of temporal variability exist, the number of operation situations to consider in order to solve the TEP problem can be huge, which makes this problem hardly tractable. However, several operation situations may have similar impacts on investment decisions. This raises the need for the use of snapshot selection methods to detect such similarities. A snapshot selection method aims to identify a reduced set of snapshots that is able to accurately represent all the operation situations that are relevant from the TEP perspective.

The oldest approaches to snapshot selection known are solely based on demand level. The simplest approach involves the consideration of a few "peak" hours during summer or winter (Grijalva, Dahman, Patten, & Visnesky, 2007). Sometimes, a few "load blocks" based on the demand duration curve are taken into account (Boîteux, 1949). Most of the power system studies which have been conducted in the following years and are making use of snapshot selection are based on variations of this idea (Gu, Ni, & Bo, 2012). With the high penetration of renewable energy sources in the power systems, some authors extend this idea by grouping together operating hours with similar load and renewable generation (Özdemir, Munoz, Ho, & Hobbs, 2016), (F. D. Munoz, Hobbs, & Watson, 2016).

According to a recent study within the e-HIGHWAY 2050 project, the snapshots to be considered should be defined by taking into account not only the load level and intermittent generation outputs but also other economic variables related to the power system operation, namely nodal prices (Agapoff & Warland, 2014). Authors in (Desta Z. Fitiwi et al., 2015) go a step further by grouping together operation situations where there is a similar pattern of network congestion, which is considered a more relevant driver of TEP investment decisions.

The previous references use the k-means algorithm to group together similar snapshots. On the contrary, authors in (Liu, Sioshansi, & Conejo, 2018) use a hierarchical clustering algorithm to identify representative hours.

Lastly, authors in (Wogrin, Duenas, Delgadillo, & Reneses, 2014) recommend clustering snapshots while preserving the chronological link among operation situations (reproducing the features of the Unit Commitment problem). In (Wogrin et al., 2014), part of the chronological information related to system operation remains available through the use of a transition matrix while reducing the size of the problem. Even though the chronological interdependency between snapshots may be relevant in a TEP problem to model the unit commitment constraints of some power plants in the power system (Zhi Wu, 2016), preserving this information is out of the scope of this thesis.

A summary of snapshot selection methods used in the literature is represented in Table 2.1.

TABLE 2.1
SUMMARY OF SNAPSHOT SELECTION METHODS USED IN THE LITERATURE

Clustering variables	Clustering algorithm	References
Load only	Peak and off-peak hours	(Grijalva et al., 2007)
	Load duration curve	(Boîteux, 1949), (Gu et al., 2012)
Load and intermittent generation	k-means	(Özdemir et al., 2016), (Munoz et al., 2016)
	Hierarchical clustering	(Liu et al., 2018)
Nodal prices	k-means	(Agapoff et al., 2014)
Network congestion patterns	k-means	(Desta Z. Fitiwi et al., 2015)

Snapshot selection methods categorized according to the clustering variables and the clustering algorithm used.

2.2. NETWORK REDUCTION TECHNIQUES

Techniques to model the transmission network in a compact way, also called "network reduction" methods, are used to reduce the size of this problem and make TEP resolution tractable.

In power systems, network reduction processes can be broken down into three steps: a network partitioning step to divide the buses into groups, or areas; a bus aggregation step to reduce the number of buses inside each area, and an equivalencing step to compute the equivalent features of the links and buses in the reduced network obtained.

2.2.1. NETWORK PARTITION METHODS AND SIMILARITY MEASURES

Network partitioning deals with the problem of dividing the network into groups of similar nodes. Network partitioning is a well-known problem in graph theory (Buluc, Meyerhenke, Safro, Sanders, & Schulz, 2013). In some power system studies, this step is discarded and authors rely on existing, well-defined, partitions, such as administrative regions (Svendsen, 2015), to identify a partition.

The most common way to divide the nodes into groups is to assign a weight to each line, or edge, connecting two nodes. The weight $W_{i,j}$, of an edge connecting nodes i and j, is a quantity that measures how similar nodes i and j are, i.e. how likely it is that these two nodes end up belonging to the same cluster, or area. This problem is then defined as an optimization problem where the objective is to maximize the total weight of intra-cluster lines or, alternatively, minimize the total weight of inter-cluster lines, while defining a given number of clusters k (2.1).

$$\min_{\substack{\text{all possible} \\ \text{k}-\text{clusters}}} \left(\sum_{\substack{\text{i and j belong to} \\ \text{different clusters}}} W_{i,j} \right) \tag{2.1}$$

In power system studies, some authors define a similarity between two buses based on the first quadrant of the power flow Jacobian (Cotilla-Sanchez, Hines, Barrows, Blumsack, & Patel, 2013), which represents the sensitivity of power injections with respect to the voltage angles in these buses. Yet, this electrical sensitivity may not be able to properly reflect the congestion in the grid, which must be considered when partitioning the network in a TEP context. Therefore, an adequate similarity measure should be defined according to the problem to be solved and not only the electrical sensitivity of the network. Thus, a market-oriented study uses a similarity measure based on nodal price differences to find an adequate partition (Singh & Srivastava, 2005), whereas the TEP-oriented study in (Shayesteh, Hobbs, Söder, & Amelin, 2015) partitions the electrical network considering the available transfer capacity (ATC)[1] between each two buses. A similarity measure based on the ATC between nodes is able to better reflect congestion in the grid than a measure based on the sensitivity of power injections with respect to voltage angles but may still miss relevant new lines likely to be installed. Relevant lines in a TEP context are congested lines and candidate lines that are likely to be built. From now on, and according to the terms used by authors in (S. Lumbreras et al., 2015), these relevant lines will be referred to as “critical lines” and their pairs of end nodes will be referred to as “critical pairs of buses”.

Besides, the optimization problem in (2.1), which has been widely used in the literature to partition electrical networks, has many drawbacks. First, a target number of clusters should be predefined, even though the most appropriate number of clusters to define might not be known in advance. Thus, numerous iterations, with various target numbers of clusters, may be required to find the most appropriate one. Second, since the

[1] The Available Transfer Capability is the measure of the available room in the physical transmission network, for transfers of power for further commercial activity, over and above already committed uses (Ejebe et al., 1998).

problem in (2.1) includes binary variables and is NP-hard, it is often relaxed. Then, a heuristic algorithm like k-means converts the solution of the relaxed problem into an integer solution that does not guarantee achieving a low integrality gap. Moreover, there is no guarantee that the partition obtained in this way leads to connected clusters (Hamon, Shayesteh, Amelin, & Söder, 2015); there might be two buses, belonging to the same cluster that are not connected by any path including only buses from this cluster. Since areas should correspond to connected components of the network, the areas end up being divided into extra areas corresponding to their connected components, resulting in more areas than initially wanted. Finally, this formulation merely implies that very dissimilar buses are more likely to belong to different areas whereas, in a TEP context, it might be relevant to force very dissimilar buses, such as critical pairs of buses, to belong to different areas.

2.2.2. Bus aggregation methods

Once a partition has been computed, there are various ways to reduce the number of buses within each of the areas defined. The most intuitive one is to aggregate all buses inside each area into a single equivalent bus. The equivalent reduced network obtained this way is called, in graph theory, "quotient graph" (Buluc et al., 2013). This is the choice made by authors in (Oh, 2012) to reduce the network. However, although this method is very efficient in terms of the reduction achieved in the size of the network, drastically decreasing the number of buses, the resulting equivalent network, whose features are obtained by using the equivalencing techniques described in 2.2.3, often fails to represent the original one with a high level of accuracy.

On the other hand, preserving some buses in the clusters, while eliminating others, can lead to a more accurate equivalent network than the aggregation method described above. The preserved buses are generally the "border buses", i.e. the buses connected to an inter-cluster line, whereas the bus to be eliminated are the "non-border buses", i.e. the buses not connected to any inter-cluster line. Preserving the border buses results in the inter-cluster lines also being preserved during the reduction process. Gaussian elimination (Dorfler & Bullo, 2013) is used to eliminate non-border buses to produce an electrically equivalent subnetwork inside each area. The REI method and the WARD method derive from it. In the REI method, two new buses are added after eliminating the non-border buses: one including their accumulated load and the other their overall generation, see (Oatts, Erwin, & Hart, 1990). In the WARD method, no virtual bus is added, and the load and generation from the eliminated buses are allocated to the remaining ones (Ward, 1949).

2.2.3. Equivalencing techniques

Equivalencing techniques aim to compute the equivalent features of the reduced network, i.e. the load and generation of buses and the admittance and capacity of lines, and depend on the bus aggregation method used. When using the "quotient graph" method, the bus features (load and generation) are easy to compute, whereas the line features (admittance and capacity) are more difficult to compute. The load of the representative bus in each area is computed as the sum of the nodal loads inside that area. Generators from the same technology in a given area are usually aggregated into a single one. The equivalent admittances of the inter-area corridors are set so that the reduced Power Transfer Distribution Factor (PTDF) matrix results in inter-area flows that are as close as possible to the ones existing in the original network under the same operating conditions (S. Lumbreras et al., 2015), (Oh, 2012). Recent advances have been made in the representation of the equivalent capacity of the inter-area corridors. Authors in (Oh, 2012) choose to keep apart inter-area lines that are often congested while grouping together non-congested lines in the same corridor. Authors in (S. Lumbreras et al., 2015) deduce the equivalent inter-area capacities by computing the maximum amount of power that can flow between each pair of areas in several sets of operating conditions representing situations of maximum stress of the network. The main drawback of these reduction methods is that the reduced PTDF matrix is always computed in an error minimization process. Then, this PTDF matrix does not generally result, in multiple situations characterized by different sets of operating conditions, in exactly the same inter-area power flows as those in the original network.

However, the Kron reduction, and the methods WARD and REI deriving from it, are based on the Gaussian elimination of non-border buses. Using these, the features of the new network should represent accurately the original one, and the inter-area flows in it should be exactly the same as those of the original network (Dorfler & Bullo, 2013). It can be mathematically proven that, for a given operating condition, the Gaussian elimination of a non-border bus from a given area preserves the power flows in any line exterior to the area it belongs to or crossing the border of that area. As a matter of fact, the inter-area lines, their admittance, and their capacity, are not affected by the elimination of the non-border buses. As for the new intra-area lines generated by this elimination, their admittance can be deduced by the Kron reduction formulas, as well as for the new generation and load in the remaining buses (Dorfler & Bullo, 2013). However, the capacity of the new intra-area lines resulting from this elimination remains to be defined. Authors in (Jang, Mohapatra, Overbye, & Zhu, 2013) have focused on this problem. They propose a method to compute the equivalent capacity of the new lines based on the PTDF matrix computed for the original and reduced network and the ATC between border buses computed for the original network. Experimentally, this method has proven to result in a single value of the capacity of most of the new lines, while, for the remaining lines, an upper bound to their capacity is computed.

2.2.4. Network reduction in a TEP context

Even though network reduction is a largely explored topic in the power system literature, few articles focus on the network reduction in a TEP context. Most of the authors applying network reduction for TEP purposes only describe their proposed reduction method without assessing it within a TEP case study (Shayesteh et al., 2015), (Fezeu et al., 2014). As for authors in (S. Lumbreras et al., 2015), they explicitly compute a TEP with the reduced network produced. However, since they are using equivalent candidate lines, it is not easy to deduce, from the optimal equivalent lines installed within the reduced network, the lines that should be installed in the original network. Authors in (J. Jia, 2014) explicitly compare the TEP solution obtained with the reduced network to the TEP solution obtained with the original network from the IEEE 24-bus power system. Although the solutions computed for both networks turn out to be similar, they are never exactly the same. This might be due to the fact that the network partition they use is not specifically adapted to a TEP problem.

The network partition method used by authors in (Fezeu et al., 2014) is not well-suited for a TEP problem, since it relies only on the structure of the network considered (electrical distance) and does not take into account the operating conditions. However, the operating conditions to be considered should have a large impact on the TEP solution computed. Authors in (Shayesteh et al., 2015) compute a partition of the network based on the ATC between each pair of buses, which is a similarity indicator that is more relevant for a TEP problem. However, the method they employ for this is based on the formulation in (2.1) and, therefore, might fail to preserve some critical lines, as inter-area ones, which are the lines that can be explicitly represented in the reduced network. The concept of critical lines is introduced in (S. Lumbreras et al., 2015). These are lines that, due to their features, should be preserved in the reduced network in order to acutely reflect the operation of the original network for network expansion planning purposes. However, critical lines in (S. Lumbreras et al., 2015) are only defined based on the congestions occurring in the network before its expansion. Then, the method proposed by authors in (S. Lumbreras et al., 2015) may miss some lines, or corridors, that should be defined as critical. Moreover, the heuristic clustering method proposed in (S. Lumbreras et al., 2015), which is based on k-medoids, might produce more areas than initially defined.

A summary of network reduction techniques used in the literature is represented in Table 2.2

TABLE 2.2
SUMMARY OF NETWORK REDUCTION TECHNIQUES USED IN THE LITERATURE

Network partition method	Similarity or distance measure	Bus aggregation method	Equivalencing techniques	References
Use existing, well-defined, regions	*NA*	Quotient graph	PTDF matrix	(Svendsen, 2015), (Oh, 2012)
	NA	REI	Kron reduction	(Oatts et al., 1990), (Jia, 2014)
Specify the number of clusters in advance and minimize the total inter-cluster weight	Jacobian	*NA*	*NA*	(Cotilla-Sanchez et al., 2013)
	Electrical distance	*NA*	*NA*	(Fezeu et al., 2014)
	Nodal price difference	Quotient graph	PTDF matrix	(Singh et al., 2005)
	ATC	REI	Kron reduction	(Shayesteh et al., 2015)
	Electrical distance and geographical distance	Quotient graph	PTDF matrix	(Lumbreras et al., 2015)

Network reduction techniques categorized according to the clustering variables and the clustering algorithm used.

NA: Not applicable

2.3. SEARCH SPACE REDUCTION TECHNIQUES

In TEP studies, the number of corridors that can be expanded increases with the square of the number of buses in the network. Moreover, the number of possible expansion alternatives increases exponentially with the number of candidate grid elements. This raises the need for the use of candidate line selection methods to identify the most promising expansion projects. Techniques to select the most appropriate candidate grid assets are called "search space reduction" methods.

In the literature, the list of candidate lines suggested in TEP studies consists most of the times in lines that are parallel to congested lines in the existing network (Grijalva et al., 2007), (Z. Wu, Du, Gu, Zhang, & Li, 2018), (Gomes & Saraiva, 2016). Although this method is very intuitive and simple, it has also major drawbacks. On the one hand, this method can only identify necessary reinforcements in existing corridors, and not promising new corridors. On the other hand, this method is not able to provide the potential benefit of reinforcing non-congested corridors. However, nodal prices provide advanced information in this regard. A high nodal price difference between two buses reveals a high potential benefit of installing a new line between these buses, at least for the first MW installed. This information makes it possible to compute a benefit-to-cost ratio[2] between any pair of buses, not only existing corridors, and to select promising candidate lines as those having the highest benefit-to-cost ratio (Z. Wu et al., 2018), (S. Lumbreras, Ramos, & Sánchez, 2014), (Zhang & Conejo, 2018). The main drawback of this sensitivity method, however, is that it estimates candidate lines' benefits based on marginal information, that is, the benefits brought by marginal reinforcement. Because of this, it fails to provide a good estimate of the required additional capacity. On the contrary, authors in (Majidi-Qadikolai & Baldick, 2015) and (Villasana, Garver, & Salon, 1985) rely on incremental methods that capture more accurately the required capacity expansion in corridors. In (Majidi-Qadikolai & Baldick, 2015), authors relax the line capacity constraints when solving an optimal power flow in the existing network and identify promising candidate lines as those in which overflows are occurring and whose reinforcement would not have negative impacts on other congested lines. Yet, similarly to methods based on congested lines (Grijalva et al., 2007), (Z. Wu et al., 2018), (Gomes & Saraiva, 2016), this method only provides information about existing corridors. Moreover, the investment cost of incremental capacity is not taken into account, and this method does not take into account the change in the corridors' equivalent admittance when reinforcing them. Authors in (Villasana et al., 1985), on the other hand, solve a relaxed "hybrid" TEP problem in which all possible corridors, together with their expansion costs, are taken into account. The solution obtained indicates which corridors should be reinforced and to which extent. The main drawback of their method, however, is that candidate lines in their relaxed TEP problem obey the transportation load flow model, which is not realistic for AC lines.

[2] The benefit-to-cost ratio of a candidate line is defined as the ratio between the economic benefit brought by the installation of the line and its investment cost. The line's benefit can be estimated thanks to its ends' nodal price. An upper-estimate of the benefit of installing a candidate line is the product of the nodal price difference by the capacity of the candidate line (S. Lumbreras, Ramos, & Sánchez, 2014).

Authors in (S. Lumbreras et al., 2014) and (Zhang & Conejo, 2018) iteratively solve the mixed-integer linear TEP problem and suggest additional candidate lines, based on congestion patterns or nodal prices, in each iteration. Authors in (Vinasco, Rider, & Romero, 2011), on the other hand, solve the mixed-integer linear TEP problem independently for each stage of the multistage TEP problem, that is, each year of the planning horizon, to find a promising set of candidate lines. The main drawback of these three articles is to include the resolution of smaller, yet complex, mixed-integer linear problem in the search space reduction process, making the reduction technique hard to apply.

Authors in (Sanchez, Romero, Mantovani, & Rider, 2005), (Rider, Gallego, Romero, & Garcia, 2007), and (Mendonça, Silva Junior, Dias, & Marcato, 2016) use constructive heuristic algorithms (CHA) to find good quality TEP solutions. CHA are iterative algorithms that solve a relaxed TEP problem while forcing the installation of new lines at each step of the process. The advantage of this algorithm is that it converges quickly towards a final step in which no new candidate lines improve the quality of the solution. The drawback of this method, however, is that the solution is often a local minimum instead of an optimal global one (Sanchez et al., 2005).

Finally, metaheuristic algorithms use a combination of random choices and knowledge of previous results to find a good enough solution of the TEP problem (Rathore et al., 2013), (Torres & Castro, 2012), (Gallego et al., 1997). Strictly speaking, these methods do not aim to reduce the search space of the TEP problem, but rather to travel through it until a good solution is found. Contrary to all the previous methods above, metaheuristic algorithms do not rely on optimization techniques.   A summary of search space reduction techniques used in the literature is represented in Table 2.3.

TABLE 2.3

SUMMARY OF SEARCH SPACE REDUCTION TECHNIQUES USED IN THE LITERATURE

References	Criterion used to identify expansion needs	Can identify new corridors	Solve a MILP TEP problem to identify candidate lines	Use CHA	Use metaheuristics methods
(Grijalva et al., 2007), (Z. Wu et al., 2018), (Gomes et al., 2016)	Congested existing lines				
(Z. Wu et al., 2018), (Lumbreras et al., 2014), (Zhang et al., 2018)	Benefit-to-cost ratio based on nodal prices	✓			
(Majidi-Qadikolai et al., 2015)	Overflows in existing lines				
(Villasana et al., 1985)	Candidate corridors partially expanded	✓			
(Lumbreras et al., 2014), (Zhang et al., 2018), (Vinasco et al., 2011)	NA	NA	✓		
(Sanchez et al., 2005), (Rider et al., 2007), (Mendonça et al., 2016)	NA	NA		✓	
(Rathore et al., 2013), (Torres et al., 2012), (Gallego et al., 1997)	NA	NA			✓

NA: Not applicable

3. TEP model assumptions and general reduction methodology

3.1. Network representation in TEP models

As discussed in the introduction, one aspect of the complexity of the TEP problem lies in the size of the problem, i.e. the amount of data, or input parameter values, to take into account when solving the problem. The present thesis focuses on this aspect.

Another aspect of its complexity lies in the network model representation, i.e. the mathematical constraints representing the physical laws that govern the way the components of the power system work. As a matter of fact, the tractability of the TEP problem strongly depends on the mathematical formulation used to represent it. In the TEP context, the main component whose representation is critical is the network and the laws that govern the power flows through the lines.

Authors in (Desta Zahlay Fitiwi, 2016) carry out a very complete review of the alternative network representations in a TEP context. This subsection provides a summary of the most common network representations.

Regarding the network representation, power flows in AC lines should formally satisfy AC load flow (ACLF) constraints. In this model, power flows are rigorously represented by the following equations:

$$P_{ij} = {V_i}^2 G_{ij} - V_i V_j \left(G_{ij} \cos\left(\theta_i - \theta_j\right) + B_{ij} \sin\left(\theta_i - \theta_j\right)\right) \tag{3.1}$$

$$Q_{ij} = -{V_i}^2 B_{ij} - V_i V_j \left(G_{ij} \sin\left(\theta_i - \theta_j\right) - B_{ij} \cos\left(\theta_i - \theta_j\right)\right) \tag{3.2}$$

$${P_{ij}}^2 + {Q_{ij}}^2 \leq S_{ij}^{\max 2} \tag{3.3}$$

Where P_{ij} and Q_{ij} are, respectively, the net active and reactive power flows in line *(i,j)* as seen from bus *i*, G_{ij} and B_{ij} are, respectively, the real and imaginary part of this line admittance, V_i is the voltage level in bus *i*, and θ_i is the voltage angle in bus *i*.

Thanks to this formulation line losses P_{ij}^{L} can be naturally represented through equation (3.4):

$$P_{ij}^{L} = P_{ij} + P_{ji} = G_{ij}\left({V_i}^2 + {V_j}^2 - 2V_iV_j\cos\left(\theta_i - \theta_j\right)\right) \tag{3.4}$$

Because equations (3.1)-(3.4) are non-linear, their introduction in a TEP problem produces a Mixed Integer Non-Linear Programming (MINLP) problem that is extremely hard to solve. Because of this, authors using ACLF model to solve a TEP problem generally avoid using classical optimization methods and use, instead, metaheuristic methods to find high quality solutions (Rider, Garcia, & Romero, 2007), (Hooshmand, Hemmati, & Parastegari, 2012).

A common simplification of the ACLF model is the DC load flow (DCLF) model. In this model, we make the following assumptions:

- lines resistance is negligible: $G_{ij} \approx 0$,
- voltage angle difference between the two ends of each AC line is very low: $\sin\left(\theta_i - \theta_j\right) \approx \left(\theta_i - \theta_j\right)$, $\cos\left(\theta_i - \theta_j\right) \approx 1$,
- voltage level difference between the two ends of each AC line is very low: $V_i \approx V_j$.

These assumptions allow equations (3.1)-(3.4) to be reformulated as:

$$P_{ij} = -P_{ji} = -{V_i}^2 B_{ij}\left(\theta_i - \theta_j\right) \tag{3.5}$$

$$Q_{ij} = 0 \tag{3.6}$$

$$-S_{ij}^{\max} \leq P_{ij} \leq S_{ij}^{\max} \tag{3.7}$$

$$P_{ij}^{L} = 0 \tag{3.8}$$

Contrary to equations (3.1)-(3.4), equations (3.5)-(3.8) are linear. These linear equations are the ones used in most TEP studies because of the significantly lower complexity of this network model (C. Munoz, Sauma, Contreras, Aguado, & Torre, 2012), (Park, Baldick, & Morton, 2015). Moreover, the assumptions from which these equations derive are realistic for most electrical transmission networks. Equation (3.5) is also known as the Kirchhoff's voltage law (KVL) under DCLF assumptions.

A more drastic model simplification is the Transportation Load Flow (TLF) model. In this model, the KVL is discarded. Therefore, the net active power flows in transmission lines should only satisfy power flow capacity constraints (3.7) and the power balance constraints in each node. These assumptions generally produce power flows that are unrealistic for real-life AC transmission lines (Alonso et al., 1991). However, this may provide a good approximation of the power flows occurring in DC transmission networks depending on the controllability of flows in branches there.

Finally, the simplest network representation is the Copper Plate (CP) model. In this model, all network effects, including capacity constraints (3.7), are discarded. Even though this representation turns out to be well suited for some energy market or

Generation Expansion Planning (GEP) studies (Palchak et al., 2017), it is fundamentally unable to represent the flows in the network and thus to identify the network expansion needs. Therefore, this representation is not appropriate for TEP studies.

A summary of the assumptions made for each network model is represented in table 3.1.

TABLE 3.1
SUMMARY OF NETWORK REPRESENTATION ALTERNATIVES

Assumptions	ACLF model	DCLF model	TLF model	CP model
$V_i \approx V_j$		X	X	X
$\theta_i - \theta_j \ll 1$		X	X	X
$G_{ij} \approx 0$		X	X	X
No KVL			X	X
$S_{ij}^{\max} \approx \infty$				X

Most common transmission network representation alternatives. A cell with an "X" means that the model of the associated column satisfies the assumption of the associated row.

3.2. PROGRAMMING ASPECTS OF THE TEP PROBLEM

The aim of solving the TEP problem is to identify optimal transmission assets investment decisions. The decision to invest and install a transmission asset in the network is a binary decision, not a continuous one. Therefore, regardless of the network or generator representation, formulating a TEP problem necessarily requires the use of binary variables. Thus, TEP problems should be formulated as mixed integer optimization problems, either MINLP or MILP. Moreover, since the number of combinations of new transmission assets to install is finite or countably infinite, these problems belong to the class of Combinatorial Optimization (CO) problems (Papadimitriou & Steiglitz, 1998).

CO problems are NP-hard and extremely difficult to solve beyond a certain problem size. There are two main methods to solve this type of problems. The first method relies on the use of metaheuristic algorithms (Blum & Roli, 2003). Metaheuristic algorithms are procedures used to find good quality solutions based on previous tested solutions. Even though these algorithms are efficient to find good solutions with low computational effort, the optimality of the solutions found this way is not guaranteed, and it is generally not possible to know how close the solutions found are from the globally optimal ones.

On the other hand, classical optimization methods converge towards a globally optimal solution. These methods can also measure how close feasible solutions found during the search procedure are from globally optimal solutions. In these methods, both MINLP and MILP problems are solved through the use of a search tree (Dakin, 1965). In this search tree, each node corresponds to a modified, non-integer (continuous), version of the original problem in which all the integer (binary) variables are linearized, and some of them are bounded by integer (binary) values. For a given node representing a specific optimization problem, a child of this node represents a modified version of this problem in which new bounds are introduced on one of the linearized integer variables, when using the Branch and Bound (BB) algorithm (Little, Murty, Sweeney, & Karel, 1963). Apart from these bounds, additional constraints, or "cuts", may be introduced to tighten the search space of the problem represented by the child node, when using the Branch and Cut (BC) algorithm (Padberg & Rinaldi, 1991). An example of such search tree is depicted in Fig. 3.1.

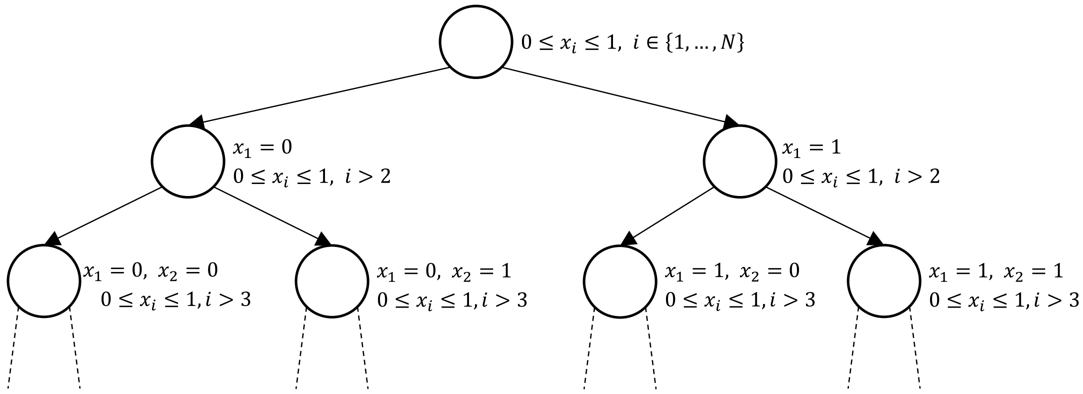

Fig. 3.1. Illustration of the search tree when using the BB algorithm to solve a MILP problem. Here, the MILP problem is composed of binary variables $x_i, i \in \{1, \dots, N\}$. In the root node, on the top, all binary variables are relaxed. The deeper we explore the tree, the more bounds we introduce.

In each node of the search tree, the corresponding problem is a Linear Programming (LP) problem if the problem to be solved is a MILP problem, or a Non-Linear Programming (NLP) problem if the problem to be solved is a MINLP problem. Solving a NLP problem is difficult since it requires computationally demanding convex optimization methods, when the NLP problem is convex. When the NLP problem is nonconvex, it is usually divided into convex sub problems, independently solved through a BB procedure. On the other hand, LP problems can be efficiently solved with well-known methods, such as the simplex algorithm (Nash, 2000).

Since solving LP problems requires significantly lower computational efforts than solving NLP problems, solving MILP problems requires, consequently, less computational efforts than solving MINLP problems.

3.3. TEP MODEL CONSIDERED IN THIS THESIS

In this thesis, we aim to solve a TEP problem through the use of classical optimization methods, instead of metaheuristic methods. According to the previous section, a MILP representation of the TEP problem is preferred over a MINLP representation, for computational reasons and given that we are focused on large-scale case studies. Therefore, the network model in section 3.1 should be of an LP type. The DCLF model is the most accurate LP representation of the power flows in AC lines and transformers. As for the DC lines and converters, their power flows do not obey KVL and the TLF model can be adopted instead, assuming the full controllability of flows in them. The assumptions made here in this regard are summarized in table 3.2. Moreover, to limit the number of binary variables and make the MILP TEP problem tractable, only transmission asset investment decisions should be represented as binary variables.

TABLE 3.2
SUMMARY OF THE TEP MODEL ASSUMPTIONS

	Mathematical representation	Class of optimization problem
AC lines and trasnsformers	DCLF	LP
DC lines and converters	TLF	LP
Transmission investment decisions	Binary variables	IP
Complete TEP model		MILP

Apart from the transmission investment decisions, represented as binary variables, all the other element of the TEP problem are represented via a LP formulation. The resulting TEP model is a MILP problem with binary variables only related to transmission investment decisions.

For the sake of simplicity, the TEP problem considered in the present thesis is a static TEP problem. Static TEP deals with investment decisions that have to be implemented by a single target year (e.g. the year 2030), whereas dynamic TEP deals with investment decisions that can be implemented in several time horizons.

According to the assumptions made above, the static optimization TEP problem can be formulated as follows:

$$\min\left\{\sum_{s=1}^{N_S}\rho_s\left(\sum_{g=1}^{N_G}p_{g,s}cg_g\right)+\sum_{s=1}^{N_S}\rho_s\left(\sum_{i=1}^{N_B}pns_{i,s}C^{ENS}\right)+\ \gamma\left(\sum_{(i,j,c)\in CC}x_{ijc}C_{ijc}\right)\right\} \tag{3.9}$$

Subject to:

$$\sum\nolimits_{g\in\Omega_i^G}p_{g,s}-d_{i,s}+pns_{i,s}+\sum\nolimits_j f_{jic,s} \\ -\sum\nolimits_j f_{ijc,s}=0;\ \forall i\in[\![1;N_B]\!],\forall s\in[\![1;N_S]\!] \tag{3.10}$$

$$f_{ijc,s}=S_BY_{ijc}\big(\theta_{i,s}-\theta_{j,s}\big);\ \forall(i,j,c)\in EC\cap AC,\forall s\in[\![1;N_S]\!] \tag{3.11}$$

$$f_{ijc,s}=x_{ijc}S_BY_{ijc}\big(\theta_{i,s}-\theta_{j,s}\big)\ ;\ \forall(i,j,c)\in CC\cap AC,\forall s\in[\![1;N_S]\!] \tag{3.12}$$

$$-\overline{f_{ijc}}\le f_{ijc,s}\le\overline{f_{ijc}};\ \forall(i,j,c)\in EC,\forall s\in[\![1;N_S]\!] \tag{3.13}$$

$$-x_{ijc}\overline{f_{ijc}}\le f_{ijc,s}\le x_{ijc}\overline{f_{ijc}};\ \forall(i,j,c)\in CC,\forall s\in[\![1;N_S]\!] \tag{3.14}$$

$$0\le pns_{i,s}\le d_{i,s};\ \forall i\in[\![1;N_B]\!],\forall s\in[\![1;N_S]\!] \tag{3.15}$$

$$0\le p_{g,s}\le\overline{p_g};\ \forall g\in\mathrm{CONV},\forall s\in[\![1;N_S]\!] \tag{3.16}$$

$$0\le p_{g,s}\le\overline{p_{g,s}};\ \forall g\in\mathrm{RES},\forall s\in[\![1;N_S]\!] \tag{3.17}$$

Equation (3.9) provides the total cost of system operation and transmission investments, which aims to be minimized. The first two terms of this equation are related to operation costs and, more specifically, to generation variable production costs and the cost of non-served energy, respectively. A weight ρ_s is associated with each snapshot (operation situation) s representing its duration in hours. These weights are all equal to 1 hour in the original problem, i.e. the TEP problem before selecting a reduced set of representative snapshots. Afterwards, weights ρ_s are used to represent the number of hours of occurrence of each representative snapshot selected through the snapshot clustering process proposed in 4.

The final term in (3.9) is related to the network investment costs. In order to obtain an annualized cost that can be directly compared to the operational one, a fixed charge annualization rate γ is applied to these investment costs. Equation (3.10) represents the power balance at each node. This includes the power produced by a virtual generation plant representing non-served power. Equations (3.11) and (3.12) correspond to the DCLF model used to represent the flow of power through existing and candidate AC circuits, respectively. Equations (3.13) and (3.14) refer to the maximum power flow that can go through existing and candidate circuits, respectively, both AC and DC. Equation (3.15) limits non-served power in each node to be positive and always lower than the actual demand level in this node. Equations (3.16) and (3.17) represent the power generation limits of thermal and RES generators, respectively. The main difference between the two is the time-dependency of RES generators capacity.

Constraints (3.12) involve the multiplication of a binary variable x_{ijc} with a linear variable $\left(\theta_{i,s} - \theta_{j,s}\right)$ and, therefore, do not satisfy the MILP requirements. To cope with this, (3.12) can be reformulated as disjunctive constraints (Sharifnia & Aashtiani, 1985).

$$-M_{ijc}\left(1 - x_{ijc}\right) \leq f_{ijc,s} - S_B Y_{ijc}\left(\theta_{i,s} - \theta_{j,s}\right) \leq M_{ijc}\left(1 - x_{ijc}\right); \quad \forall (i,j,c) \in CC \cap AC, \forall s \in [\![1; N_S]\!] \tag{3.18}$$

Constraints (3.18) impose the condition that the relationship between voltage angles at the two extreme nodes of a candidate line and the flow in the line is the same as for existing lines when this candidate line is installed ($x_{ijc} = 1$). On the contrary, assuming M_{ijc} big enough, it does not constrain in any sense voltage angles at the extreme nodes when the line is not built.

Authors in (Binato et al., 2001) provide a method to set the lowest possible value for this parameter. For a candidate line (i,j,c), they compute M_{ijc} as:

$$M_{ijc} = S_B Y_{ijc} D_{i,j} \tag{3.19}$$

Where $D_{i,j}$ is the maximum possible angle difference between buses i and j. To find this value, they compute a shortest path problem between buses i and j. The graph in which they compute this shortest path is the weighted graph whose structure is the one of the initial (not expanded) transmission network, and whose line weights $W_{k,l}$ are the maximum angle difference between the two ends of the corresponding lines, that is,

$W_{k,l} = \frac{\overline{f_{k,l}}}{S_B Y_{k,l}}$. Because the shortest path problem is defined for graph without parallel lines, the parameters $D_{i,j}$, $W_{k,l}$, $\overline{f_{k,l}}$, and $Y_{k,l}$, are defined for each pair of buses, not for each circuit. $\overline{f_{k,l}}$ and $Y_{k,l}$ are respectively the equivalent maximum power flow and equivalent admittance of the corridor (k, l), and they can be calculated as $Y_{k,l} = \sum_c Y_{klc}$ and $\overline{f_{k,l}} = Y_{k,l} \min_c \left(\frac{\overline{f_{klc}}}{Y_{klc}}\right)$.

Thanks to this value, the resulting TEP problem described by equations (3.9)-(3.11) and (3.13)-(3.18) is MILP and tight.

3.4. GENERAL REDUCTION METHODOLOGY

The MILP TEP problem formulation provided in the previous section strikes a balanced trade-off between accuracy and tractability. However, as explained in the introduction, for modern power systems, its size makes this problem difficult to solve. This is the reason why reduction techniques should be applied to this problem.

As discussed in section 2, there are many reduction techniques applied in the literature targeting each dimension of the problem: the temporal variability, the network size, and the candidate network asset search space. However, the vast majority of these techniques are either based on a rule-of-thumb or not adapted to the problem at hand (TEP). To be efficient, a reduction technique should be tailor-made to the problem we aim to solve with the reduced model.

3.4.1. PROBLEM SIZE REDUCTION AND INFORMATION THEORY

The reduction of the size of a mathematical problem involves the reduction of the amount of data to take into account when solving the problem or the complexity of the formulation of this problem. Problem size reduction is therefore, many times, equivalent to data compression.

Authors in (Sayood, 2005) define data compression as the procedure of *"discarding irrelevant information"* from a data set, or signal. Going further, authors in (Tishby, Pereira, & Bialek, 2000) define the "relevant information" of a signal as *"being the information that this signal provides about another signal"*. More specifically, the goal of data compression is to *"extract the information from one variable that is relevant for the prediction of another one."* (Tishby et al., 2000)

In the specific case of a TEP problem, the data we aim to compress is the set of input parameters that we have previously categorized into three dimensions, namely the temporal variability, the network structure, and the network investment search space. On the other hand, the variables we aim to predict are the investment decision variables. Therefore, an efficient reduction technique applied to a TEP problem involves compressing the problem input data enough while preserving the relevant information required to compute the investment decisions.

The key here is to determine the relevance of the input parameters for the computation of the decision variables. This requires capturing the impact of the input parameters on the decision variables. The obvious way to do this is solving the TEP problem. However, solving the original MILP TEP problem to capture this information would be meaningless, since we want to use this information to reduce the size of the problem before solving it. Then, one reasonable alternative is solving a modified, more tractable, version of the TEP problem, and thus obtaining useful information on the potential impact of input parameters on decision variables.

3.4.2. Linear relaxation of the TEP problem

The linear relaxation of a MILP problem is a modified, LP, version of the original problem in which the integer (binary) variables have been converted into continuous ones. When applying linear relaxation to the TEP problem described in 3.3, binary constraints $x_{ijc} \in \{0,1\}$ are replaced by linear constraints $0 \leq x_{ijc} \leq 1$. The main advantage of this alternative LP problem is that it can be solved without exploring the humongous search tree described in 3.2. However, despite being linear, the relaxed TEP problem is still very large, and not tractable for LP solvers based on the simplex method. Thankfully, the alternative interior point method (Gondzio & Makowski, 1995) turns out to be very efficient when solving large-scale LP problems.

In the specific case of the TEP problem, the linear relaxation relies on the assumption that candidate circuits can be "partially built", and that the capacity, admittance, and investment cost of these partially built circuits are proportional to the quantity of line built. These assumptions, along with the equations they are related to, are summarized in table 3.3.

TABLE 3.3
COMPARISON OF INVESTMENT DECISIONS COMPUTED IN THE NON-RELAXED AND THE RELAXED TEP MODEL

Characteristics of the circuit c	Non-relaxed TEP model	Relaxed TEP model	Equations related to this relaxation
Quantity built	0 or 1	$0 \leq x_{ijc} \leq 1$	
Investment cost	C_{ijc}	$x_{ijc}C_{ijc}$	(3.9)
Capacity	$\overline{f_{ijc}}$	$x_{ijc}\overline{f_{ijc}}$	(3.14)
Admittance	Y_{ijc}	$x_{ijc}Y_{ijc}$	(3.18), (3.27)

The admittance of the "partially built" candidate circuits would be exactly proportional to the quantity built if, in the relaxed problem, we could include equation (3.12). This equation represents the KVL for partially built candidate circuits (with an admittance $x_{ijc}Y_{ijc}$). However, due to the fact that it includes a product of linear variables x_{ijc} and $\left(\theta_{i,s} - \theta_{j,s}\right)$, this equation is non-linear. Thus, including this equation in the relaxed TEP model would prevent us from using LP programming to solve it. For that reason, instead of enforcing equation (3.12), the product $x_{ijc}\left(\theta_{i,s} - \theta_{j,s}\right)$ is represented by an envelope of it for the ranges of variables x_{ijc} and $\left(\theta_{i,s} - \theta_{j,s}\right)$ to be considered in the TEP problem, while the value of the term $\frac{f_{ijc,s}}{S_B Y_{ijc}}$ is enforced to lie within this envelope.

The McCormick envelope (McCormick, 1976) of a product of variables xy for which the variable x is bounded between $\underline{x}$ and $\overline{x}$, and the variable y is bounded between $\underline{y}$ and $\overline{y}$, is represented by equations (3.20) to (3.23). This envelope is depicted in Fig. 3.2.

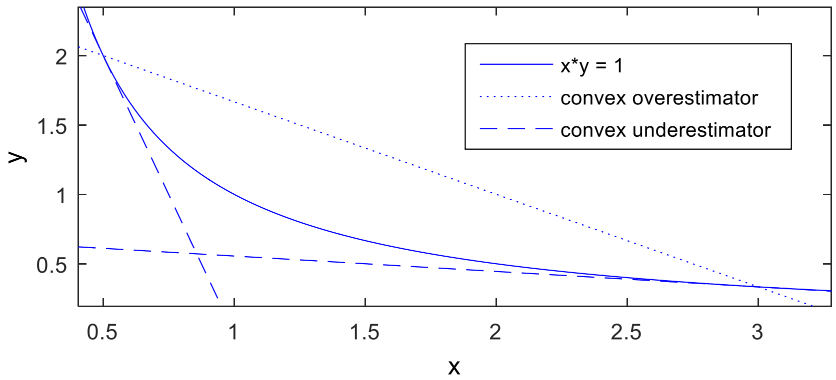

Fig. 3.2. McCormick envelope of bilinear variable *xy*, when variable *x* is assumed to be bounded between 1/2 and 3 and variable *y* is assumed to be bounded between 1/3 and 2.

$$xy \leq \overline{y}x + \underline{x}y - \underline{x}\overline{y} \tag{3.20}$$

$$xy \leq \overline{x}y + \underline{y}x - \underline{y}\overline{x} \tag{3.21}$$

$$xy \geq \underline{y}x + \underline{x}y - \underline{x}\underline{y} \tag{3.22}$$

$$xy \geq \overline{y}x + \overline{x}y - \overline{x}\overline{y} \tag{3.23}$$

In section 3.3, according to (Binato et al., 2001), $(\theta_{i,s} - \theta_{j,s})$ can be bounded between $\pm D_{i,j} = \pm \frac{M_{ijc}}{S_B Y_{ijc}}$. Besides, x_{ijc} is bounded between 0 and 1. Consequently, according to equations (3.20) to (3.23), the McCormick envelope of the product $x_{ijc}(\theta_{i,s} - \theta_{j,s})$ is represented by equations (3.24) and (3.25).

$$-x_{ijc}\frac{M_{ijc}}{S_B Y_{ijc}} \leq x_{ijc}(\theta_{i,s} - \theta_{j,s}) \leq x_{ijc}\frac{M_{ijc}}{S_B Y_{ijc}} \tag{3.24}$$

$$-\frac{M_{ijc}}{S_B Y_{ijc}}(1 - x_{ijc}) \leq x_{ijc}(\theta_{i,s} - \theta_{j,s}) - (\theta_{i,s} - \theta_{j,s}) \leq \frac{M_{ijc}}{S_B Y_{ijc}}(1 - x_{ijc}) \tag{3.25}$$

Since $x_{ijc}(\theta_{i,s} - \theta_{j,s})$ can be replaced by $\frac{f_{ijc,s}}{S_B Y_{ijc}}$, equation (3.25) is exactly the disjunctive constraints (3.18) and is redundant in the relaxed TEP model. However, equation (3.24) can be reformulated as (3.26) and these additional constraints tighten the relaxed TEP problem, generating higher quality optimal decision variables.

$$-x_{ijc}M_{ijc} \leq f_{ijc,s} \leq x_{ijc}M_{ijc} \tag{3.26}$$

Constraints (3.14) and (3.26) can be merged into the single set of constraints (3.27).

$$-x_{ijc}\min(M_{ijc}, \overline{f_{ijc}}) \leq f_{ijc,s} \leq x_{ijc}\min(M_{ijc}, \overline{f_{ijc}}) \tag{3.27}$$

Finally, the relaxed version of the TEP problem to be solved to obtain the optimal "partially built" candidate lines, using the McCormick envelope, is the one described by equations (3.9)-(3.11), (3.13), (3.15)-(3.18), (3.27). The network flows computed considering this problem in the case study of section 4.3 have shown to be closely related to those resulting from enforcing the KVL once network investments have been decided. Then, using the McCormick envelope has proven to affect only to a small extent the solution of the problem and can, therefore, be regarded as a reasonable relaxation of the optimal continuous network reinforcement problem (the problem expressed by equations (3.9)-(3.17) considering continuous investment decision variables).

As aforementioned, these newly built circuits do not correspond, in general, to real discrete reinforcements, but to virtual ones whose investment cost, capacity, and admittance are largely proportional to the value of the corresponding investment decision. However, the identity and size of new circuits built as continuous reinforcements is relevant information on the impact that input parameters have on discrete investment variables.

3.4.3. General input data reduction methodology based on the relaxed version of the TEP problem

The linear relaxation of the TEP problem, described in the previous section, is the problem we solve to obtain relevant information about the impact of input parameters on the decision variables. This information can be used to determine which input data should be preserved when determining the optimal network investments and which can be discarded.

For every reduction technique presented in this thesis, the first step consists in solving the relaxed TEP problem. Then, the following information, related to the network investment decision variables, is used to reduce the size of the TEP problem for each dimension:

- The potential benefits brought by candidate lines is employed to reduce the temporal variability considered in the problem,
- The identity of the congested lines and partially installed lines is employed to reduce the size of the network,
- The identity of the expanded corridors and the amount of new capacity built in them is employed to reduce the search space (set of candidates to consider).

4. Snapshot selection for a TEP problem

4.1. Preliminary discussion

This chapter corresponds to our article “An Operational State Aggregation Technique for Transmission Expansion Planning Based on Line Benefits”.

To our knowledge, no previous work on the clustering of operational states in a TEP context considers new line benefits as clustering variables. The temporal variability and scenario reduction techniques already available in the literature for the selection of snapshots either consider clustering variables that are related to the causes of network reinforcements (like the increase in the nodal load level), or they consider an incomplete, or less relevant, subset of the effects in the system of network reinforcements.

We believe that the benefits to be produced by network investments represent the most relevant drivers of investment decisions in the context of the TEP problem, since the decision to undertake a new transmission expansion project directly depends on the operation cost reduction (i.e. the kind of benefits considered in the analysis presented here) the corresponding new assets would bring about if installed in the system. It is an estimate of these economic benefits what drives, together with reinforcement costs, the selection for their construction of some specific reinforcements instead of others. By finding the most representative snapshots according to the operation costs reduction achieved in them through the expansion of the network, the proposed approach for the selection of snapshots is based on the effects, or results, caused by network investments, or the outputs of the TEP problem, instead of the primary causes of these investments, or inputs to the TEP problem, i.e. the changes in the pattern of demand and generation in the system creating new network reinforcement needs. Here, a novel snapshot selection method for TEP is proposed based on the consideration of the similarities and differences that exist among operation situations in terms of the benefits produced by candidate reinforcements in them.

As seen in 2.1, some snapshot reduction techniques (Wogrin et al., 2014) allow to partially preserve information about chronological inter-dependency between them, which is useful when modeling chronological constraints such as the unit commitment constraints in the TEP problem. However, this is out of the scope of this thesis and, for the sake of simplicity, such chronological constraints are not considered in our TEP model.

4.2. Methodology

In this section, the methodology applied to choose the representative snapshots and their respective weights to consider in the TEP problem is discussed.

This methodology is divided into the steps that follow:

1. Running a relaxed version of the TEP problem to compute a proxy to the optimal set of network reinforcements;
2. Computation of the hourly benefits produced by each of these reinforcements;
3. Reduction of the dimensionality of the space of hourly benefits produced by reinforcements;
4. Clustering of operation situations according to the compact representation of the benefits that reinforcements produce in them.

A flowchart of the methodology applied is shown in Fig. 4.1.

4.2.1. Relaxed TEP problem and related optimal investments

Computing the benefits produced in each operation situation by all the candidate lines provided a priori by the network planner may probably not be representative enough of the actual benefits to be obtained from network expansion in each of these situations. Some candidate lines may produce too small benefits compared to their cost, while others may even have a negative impact on the overall benefit of the expansion plan (S. Lumbreras et al., 2014). Thus, computing the benefits produced by all candidate lines may result in a poor estimate of the benefits produced by the real expansion plan. Besides, the size of the reinforcement benefit space considering the whole set of candidate lines to be built may be huge, thus making the clustering of snapshots in this space more difficult and inaccurate.

Ideally, we should compute the benefits produced by the optimal set of reinforcements (candidate lines selected to be built), i.e. the one minimizing the total annualized network investments plus operation costs of the system throughout the target year. However, this would require solving the original TEP problem considering all hourly snapshots in this year, which is exactly what we wish to avoid. For that reason, we instead consider the benefits produced in each hourly snapshot by each reinforcement within a set of most promising ones, which we compute by solving a relaxed version of the TEP problem.

Then, we characterize each snapshot by the benefits produced in it by each of the optimal reinforcements computed in the relaxed version of the TEP problem. These benefits represent a proxy to the benefits produced in this same snapshot by the discrete reinforcements identified as optimal in the original TEP problem.

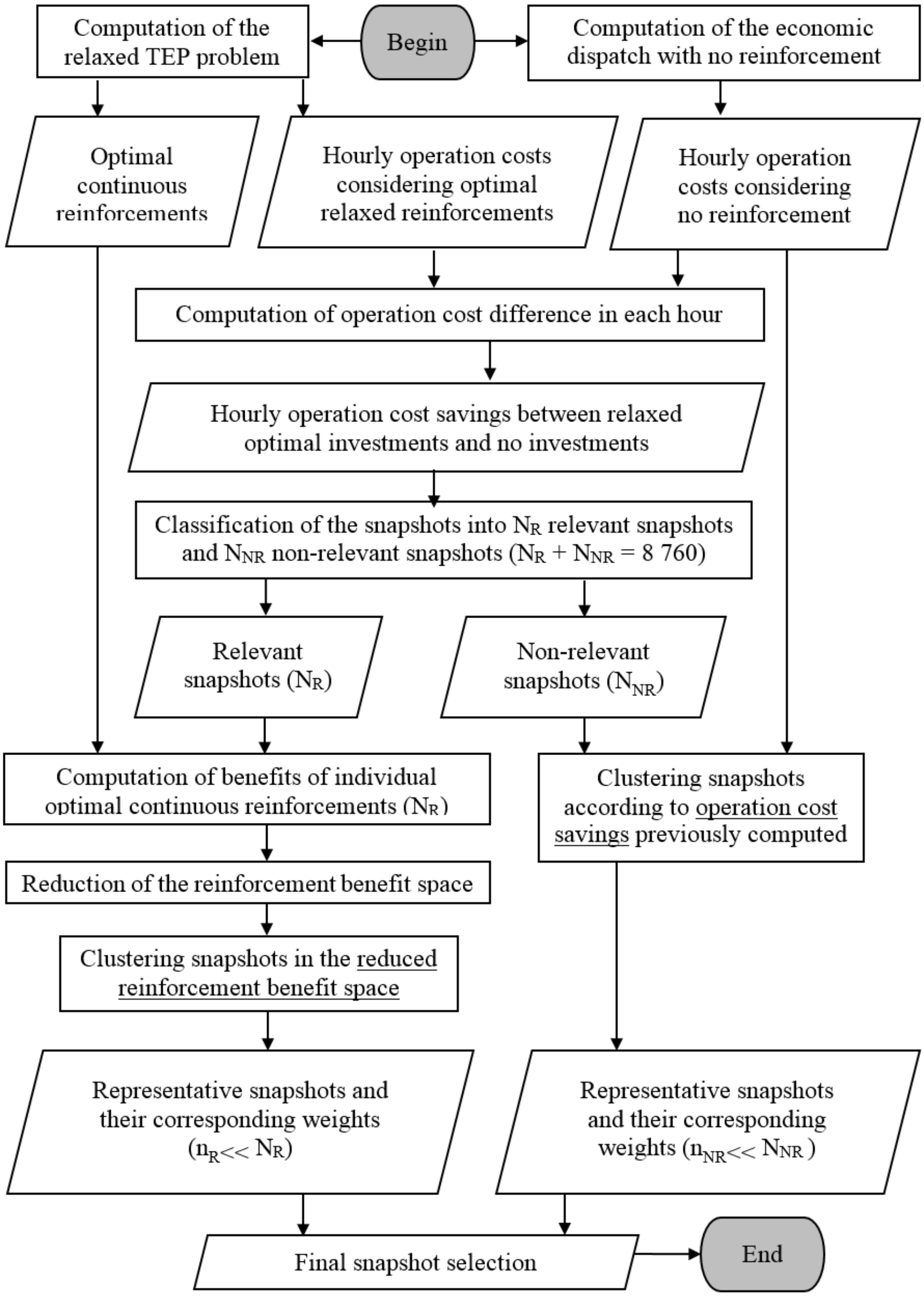

Fig. 4.1. Flowchart of the methodology. First, operation situations are classified into relevant and non-relevant ones according to the potential operation cost savings that can be realized in them by reinforcing the network with optimal continuous reinforcements. Then, relevant snapshots are clustered according to line benefits, whereas non-relevant ones are clustered according to operation cost savings. Figures N_x in brackets represent the numbers of snapshots considered in each step.

In the relaxed version of the TEP model, binary investment decision variables x_{ijc} are transformed into linear ones bounded between 0 and 1. From now on, this model is referred to as "relaxed TEP model", corresponding to a LP problem whose computational burden is much smaller than that of the TEP problem with binary investment decision variables, which is a MILP problem and is referred to as "non-relaxed TEP model".

Each hourly operation situation, or snapshot, is to be characterized by the benefits (reductions in operation costs) produced in it by those reinforcements computed in the relaxed TEP model. As aforementioned, these newly built circuits do not correspond, in general, to real discrete reinforcements, but to virtual ones whose investment cost, capacity, and admittance are largely proportional to the value of their associated investment decision.

Candidate lines that are not built (those for which the investment decision variable is equal to 0 in the solution of the relaxed TEP model) are not taken into account for the computation of line benefits when characterizing snapshots.

4.2.2. COMPUTATION OF LINE BENEFITS FOR EACH OPERATION SITUATION

Once reinforcements to consider are computed by solving the relaxed TEP model, the operation cost savings brought about by the overall set of reinforcements in each snapshot, as well as the benefit brought about by each of them individually, are computed.

The operation cost savings in each snapshot resulting from the deployment of this expansion plan are obtained by computing the difference, in terms of hourly operation costs, between the case in which we force all investment decisions to be equal to zero (initial Optimal Power Flow OPF) and the case in which we force all investment decisions to be equal to the ones which are solution of the relaxed TEP model (OPF when implementing optimal continuous reinforcements).

The benefit produced by a new line in an operation situation is, in principle, to be computed as the operation cost savings it brings about in this operation situation. These savings should, in principle, be computed as the difference between the system operation costs in the case where this line is not built and the operation costs in the case where the line is built, all other things being equal. However, defining the benefit produced by a specific line, when this line is part of a group of lines assumed to be installed simultaneously, i.e. an expansion plan, is not an easy task. Indeed, the operation cost savings produced by this line depend on the other lines being installed together with it and their order of deployment.

Several approaches have been proposed in the literature (Báñez Chicharro, Olmos Camacho, Ramos Galán, Canteli, & María, 2016) to compute the benefits of a specific line within an expansion plan. One first approach involves computing the benefit of installing a line as the reduction in operation costs caused by the installation of this line when no other reinforcement in the plan has yet been undertaken (PINT approach, for "Put IN one at a Time"), see equation (4.1). (ENTSOE, 2015) Another, more recent, approach involves computing the benefit of installing a line as the benefits produced by its installation when all the other reinforcements in the plan have already been undertaken (TOOT approach, for "Take Out One at a Time"), see equation (4.2). (ENTSOE, 2015)

$$\begin{aligned} B_{s,ijc}^{\mathrm{PINT}} &= \mathrm{Op.\,costs\ in\ s\ when\ no\ line\ is\ installed} \\ &- \mathrm{Op.\,costs\ in\ s\ when\ only\ line\ (i,j,c)\ is\ installed} \end{aligned} \tag{4.1}$$

$$B_{s,ijc}^{\text{TOOT}} = \text{Op. costs in s when only line (i, j, c) is not installed}$$
$$- \text{Op. costs in s when all lines are installed} \quad (4.2)$$

However, in reality, network reinforcements within the expansion plan may be deployed in any possible order. Besides, one must take into account the fact that benefits produced by each reinforcement when it is installed in the first place can be deemed to be largely complementary of those produced by this reinforcement when it is installed in the last place within the expansion plan. This is so because benefits produced by a reinforcement when it is installed alone (as in PINT) correspond to those benefits that only require the installation of this line to be realized, while they may also be produced by other reinforcements in the expansion plan. On the other hand, benefits produced by a reinforcement when it is installed in the last place within the plan include those benefits that are contingent on the undertaking of other reinforcements in the plan together with the one concerned.

Because of this, it seems appropriate to combine the benefits assigned to each individual reinforcement by the TOOT and PINT approaches in order to compute a more accurate estimate of the benefits produced by this reinforcement. This implicitly involves assuming that benefits to be allocated to each reinforcement include both part of the benefits it brings when it is the first one to be installed and part of the incremental benefits resulting from its installation in the last place within the expansion plan. Results obtained using this approach have proven to be reasonably accurate. Thus, the benefits here allocated to the individual network reinforcements comprising the expansion plan computed by solving the relaxed TEP model are a weighted average of the benefits allocated to them in the PINT and the TOOT approaches, see equation (4.3). In the case examples considered, the coefficient α in (4.3) is given a value of 0.5. This is so because, generally speaking, undertaking a specific reinforcement within the plan in the first place is neither more probable nor less probable than undertaking it is the last place. Actually, in most cases, this reinforcement should be undertaken in between some other reinforcements within the plan. However, considering all the possible orders of deployment of reinforcements within a large expansion plan is not computationally feasible.

$$B_{s,ijc} = \alpha B_{s,ijc}^{TOOT} + (1-\alpha) B_{s,ijc}^{PINT} \quad (4.3)$$

More sophisticated approaches are proposed by authors in (Báñez Chicharro et al., 2016), but they are more difficult to implement and computationally burdensome. Because the proposed scheme for the computation of the benefits produced by each new line should produce reasonable results, and in order not to increase substantially the computational burden of the approach proposed here for the selection of snapshots, the application of sophisticated methods for the computation of the benefits of individual reinforcements has been discarded.

Operation costs considered when computing the benefits of reinforcements correspond to the sum of variable production costs and the cost of non-served energy from equation (3.21). We can thus represent operation situations within the space of reinforcement benefits. In this space, benefits in each axis correspond to those obtained in the corresponding operation situation from the undertaking of a specific reinforcement within the expansion plan. Each operation situation is represented by a point in this space.

Moreover, results computed in the case example presented in section 4.3, as well as those computed in other analyses, show that reinforcements are only bringing about significant benefits in a reduced set of snapshots. Then, in order to characterize snapshots and cluster them into groups more accurately, snapshots where benefits from reinforcements are significant, which are here referred to "relevant snapshots", have been clustered separately from the remaining ones, which are referred to as "non-relevant snapshots". Besides, in order to reduce the computational burden of computing the benefits of individual reinforcements in individual snapshots, these benefits are only considered as clustering variables in the snapshot selection process conducted for "relevant snapshots". Non-relevant snapshots are clustered according to the overall operation cost savings produced in them by the deployment of the whole expansion plan.

In the case studies below, relevant snapshots have been chosen to be the smallest set of snapshots possible where, at least, 90% of the overall annual operation cost savings resulting from the optimal expansion of the network are achieved. According to the results computed in the case studies, this threshold value for the fraction of the overall benefits achieved in relevant snapshots has proven to represent a good compromise between capturing a large enough portion of overall operation cost savings in these snapshots and having a small enough subset of relevant snapshots.

4.2.3. Dimension reduction of the line benefit space

The sparsity of the set of network reinforcement benefits across lines and snapshots, and the relevant correlation factors probably existing among the benefits produced by several reinforcements, advices identifying the main underlying sources of the variability existing in the benefits of reinforcements across snapshots, and representing the variability in these benefits across snapshots in a more compact way, as the authors in reference (Aggarwal, Wolf, Yu, Procopiuc, & Park, 1999) explain. In high dimensional spaces, distances among snapshots become relatively uniform. Then, distances defined become meaningless, which makes clustering more difficult. This raises the need for efficiently reducing the dimension of the benefit space, while retaining the major part of information available about the variability of reinforcement benefits across operation situations.

In order to reduce the dimension of the benefit space, Principal Component Analysis (PCA) is applied to original reinforcement benefits. The PCA algorithm transforms a dataset in a new one that provides the same information as the original one but in a more compact way. This is achieved through the use of a new frame of reference where each dimension, x, is termed principal component x. The transformation of the original dataset is defined in such a way that the first principal component of the transformed dataset has the largest possible variance (that is, accounts for as much of the variability in the dataset as possible), and each succeeding component has, in turn, the highest variance possible under the constraint that it is orthogonal to the preceding components. The major part of the variability across operation situations existing in the initial reinforcement benefit space is captured using the first Principal Components.

Thanks to this process, the dimension of the reinforcement benefit space can be drastically reduced without incurring in important loss of relevant information about the operation situations to be clustered.

4.2.4. Clustering algorithm

4.2.4.1. Deterministic computation of the K centroids

The K-means algorithm is applied to cluster operation situations within the reduced reinforcement benefit space (comprising the principal components of the original reinforcement benefit space). The Euclidean distance is used to compare operation situations. There are different possible criteria to measure the goodness of a clustering of samples (operation situations) for a given number of clusters K. The most classical one, which is the one used in this article, aims to minimize the sum of distances of samples to their associated centroid (center of their cluster), i.e., it aims to minimize the intra-cluster distance.

The K-means clustering problem is NP-difficult and computationally hard to solve. Therefore, heuristic algorithms are preferred over classical optimization methods to solve this problem. Most heuristic clustering algorithms start from an initial set of K samples, defined as the initial set of K centroids, and iteratively try to increase the goodness of the clustering of operation situations by modifying the set of centroids. The choice of the initial set of centroids is of major concern. Improving the initial choice of centroids leads to computing a good partition of snapshots in fewer iterations and avoiding trivial partitions (those where some clusters contain only one snapshot). The K-means++ algorithm (Arthur & Vassilvitskii, 2007) addresses this concern by increasing the spread of the initial set of centroids. A first centroid is chosen randomly, considering a uniform distribution for the probability of choosing any snapshot as the first centroid. Then, each subsequent centroid is chosen among the remaining snapshots when each of these snapshots is assigned a probability of being chosen proportional to the square distance in the reduced benefit space from this snapshot to the closest snapshot already chosen as a centroid. This process ends when the initial set of K centroids are chosen. Authors in (Arthur & Vassilvitskii, 2007) prove the efficiency of such an initial choice of centroids.

Here we have applied an adapted, deterministic, version of the K-means++ algorithm. We choose the first centroid to be the very first sample (the first snapshot) of the dataset. This is an arbitrary choice. Any other would be possible. Making explicit here the choice made of the first centroid should allow the reader to reproduce results computed, if needed. Then, each next centroid chosen is the snapshot that is farthest from the previously chosen centroids within the reduced line benefit space. Taking the maximum number of iterations as an input parameter in the clustering algorithm, and an initial set of K centroids, a deterministic clustering of snapshots can be obtained.

4.2.4.2. SELECTION OF THE REPRESENTATIVE SNAPSHOTS AND THEIR RESPECTIVE WEIGHT IN THE TEP PROBLEM

At the end of the application of the K-means algorithm, the centroid of a cluster is computed as the average principal component values over all the operation situations making the cluster. Thus, centroids do not correspond to realistic system states. They cannot be taken as the representatives of clusters defined. Only one of the real operation situations making a cluster can be taken as the representative of this cluster. But the cluster representative should have features that are close to the average ones for this cluster. Therefore, for each cluster, we choose its representative as the operation situation within this cluster that is closest to its centroid in the reduced space of line benefits.

Given that each of the operation situations considered represents one specific hour of the year, i.e. all of them have the same probability of occurrence, the representative snapshot of each cluster is assigned a probability of occurrence, or weight, in the TEP problem equal to the number of snapshots in this cluster. Thus, each snapshot within the original set of them has a weight $\rho_s = 1$. Then, once the representatives of clusters have been chosen according to the clustering results, each original snapshot s is given a weight in the TEP problem equal to $\rho_s = n$ if the snapshot s is the representative of a cluster of n snapshots, and equal to $\rho_s = 0$ if the snapshot s is not a cluster representative.

From now on, the TEP model formulated considering the representatives of the clusters of snapshots computed is referred to as the “reduced TEP model”, while the TEP model formulated considering all snapshots in the year is referred to as the “non-reduced TEP model”.

4.3. CASE STUDY

The proposed method for the selection of snapshots to be considered in TEP has been applied to select the most relevant ones for a case study based on the standard IEEE 24-bus Reliability Test System (RTS) (Grigg et al., 1999). The power system is illustrated in Fig. 4.2. The original data of this power system can be found at (“Power Systems Test Case Archive,” n.d.). Based on the RTS system, analyses have been conducted considering hourly time series for main system variables. Two case studies have been built in this way, which differ in the level of demand. Case A is deemed the base one, while case B considers 5% of extra demand in each node.

4.3.1. SYSTEMS DESCRIPTION

Most of the system features are the ones of the standard IEEE-24 RTS. However, transmission losses are neglected. Besides, only variable production costs are considered for generation units, and the minimum output for all units is set to zero. For each existing line in this system, an additional candidate reinforcement with the same features is considered. Hence, the number of candidate lines considered is 34. The investment costs of the candidate lines are the ones considered in (Alguacil, Motto, & Conejo, 2003). The annualized factor, or rate, applied to line investment costs to compute annualized ones is γ equal to 7.94% for all candidate lines. According to data in (Desta Z. Fitiwi et al., 2015),

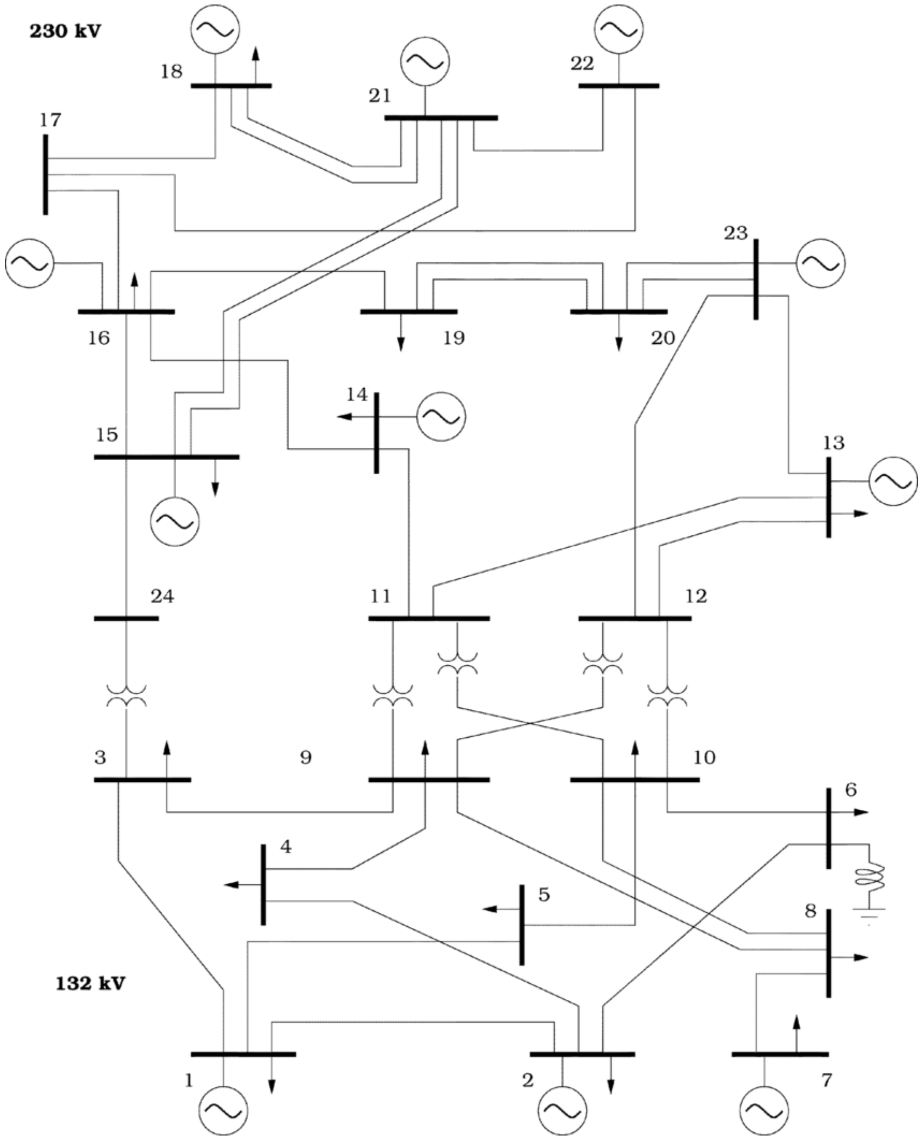

Fig. 4.2. Illustration of the standard IEEE 24-bus Reliability Test System (Grigg et al., 1999)

both in case A and case B, a solar farm with a maximum output of 500 MW and two wind farms with maximum outputs of 1 500 MW and 1 000 MW have been included in buses 4, 13 and 22, respectively. The time series for the availability state of generators has been produced as a random binary one (1 representing the corresponding unit is available and 0 representing it is not) where the average value (across all snapshots) has been constrained to lie between 0.85 and 0.95 depending on the generator technology. A one-year scope, hourly detailed, load profile has been produced for each bus. The load profiles and the RES production output profiles have been generated based on real, hourly-detailed, load and RES output profiles from the European power system, which have been scaled to be well adapted to the load level of the IEEE 24 bus system. Its average value in case A is equal to the demand level defined for this bus in the original RTS power system. The average load level in case B is 5% higher. The level of correlation between the load level in two nodes ranges from 0.55 to 0.96. This results in a large enough variety of operation situations.

4.3.2. APPLICATION OF THE METHODOLOGY

The relaxed TEP model is solved for the two case studies. The value of 11 network investment decision variables is strictly positive both in case A and case B. Reinforcements computed are provided in table 4.1.

TABLE 4.1
OPTIMAL CONTINUOUS REINFORCEMENTS FROM RELAXED TEP MODEL

Case A: 24-bus system		Case B: 24-bus system with 5% more demand	
Line	Quantity built (%)	Line	Quantity built (%)
4-9	4.9	4-9	14.9
6-10	1.0	6-10	4.4
8-9	17.8	8-9	22.7
9-11	4.7	9-11	8.1
10-11	4.6	10-11	6.6
11-13	17.9	11-13	35.3
14-16	22.6	14-16	23.0
15-21	48.2	15-21	56.4
16-17	73.4	16-17	73.1
20-23	14.6	20-23	17.6
21-22	50.5	21-22	58.6

Optimal values of investment decisions obtained from the relaxed TEP model.

First, the operation cost savings produced in each snapshot by reinforcements computed in the relaxed TEP model are determined. Relevant snapshots have been determined as those for which total operation cost savings from reinforcements add up to 90% of the total annual savings achieved. (Fig. 4.3.) According to operation cost savings computed, there are 1350 relevant snapshots in case A, and 1050 relevant snapshots in case B, out of the 8760 hourly snapshots that exist in a whole year. The most relevant reason why increasing the demand level may result in having fewer relevant snapshots has to do with the way we have defined relevant and non-relevant snapshots. Defining an absolute threshold value of operation cost savings to identify relevant snapshots would have probably led to obtaining a higher number of relevant snapshots the larger demand is. However, considering relevant snapshots as those where a certain fraction of savings are achieved has led, in this specific case, to the number of these snapshots being lower when demand is larger. This is because, due to the increase in demand, savings from reinforcing the grid have grown much more significantly in some few snapshots than in the rest of them. Therefore, most relevant savings have concentrated in a small number of snapshots.

For each relevant snapshot, individual reinforcement benefits are computed according to equations (4.1), (4.2) and (4.3). Relevant snapshots can be represented in the space of line benefits, whose dimension is 11 both in cases A and B.

PCA is applied to express the variability of the vector of reinforcement benefits in a more compact way, i.e. using a lower number of dimensions. Thus, the Principal Components of the reinforcement benefit dataset are determined. The smallest set of principal components of reinforcement benefits capturing at least 99% of the variability

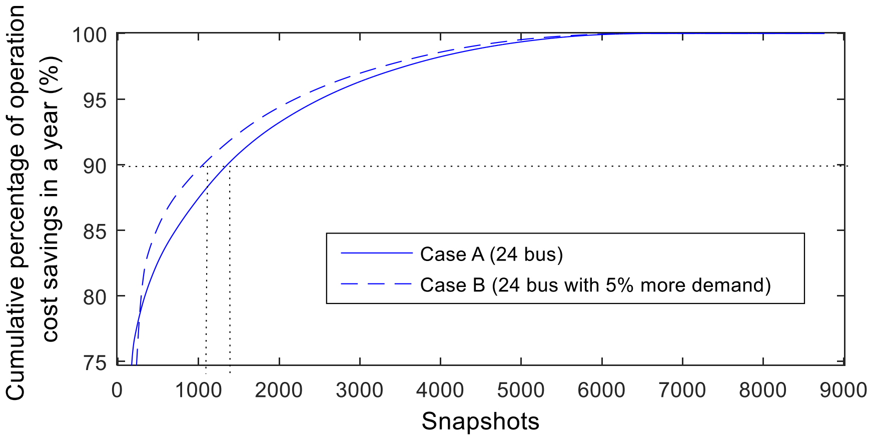

Fig. 4.3. Cumulative sum of the operation cost savings achieved in all snapshots to the left with respect to total savings throughout the whole year when sorting snapshots in decreasing order of operation cost savings achieved in them.

of these reinforcement benefits is retained. This corresponds to the set of the first four Principal Components both in case A and case B.

Relevant snapshots and non-relevant snapshots are clustered independently, as described in the methodology section. The results of clustering relevant snapshots in case A considering 10 clusters are depicted in Fig. 4.4.

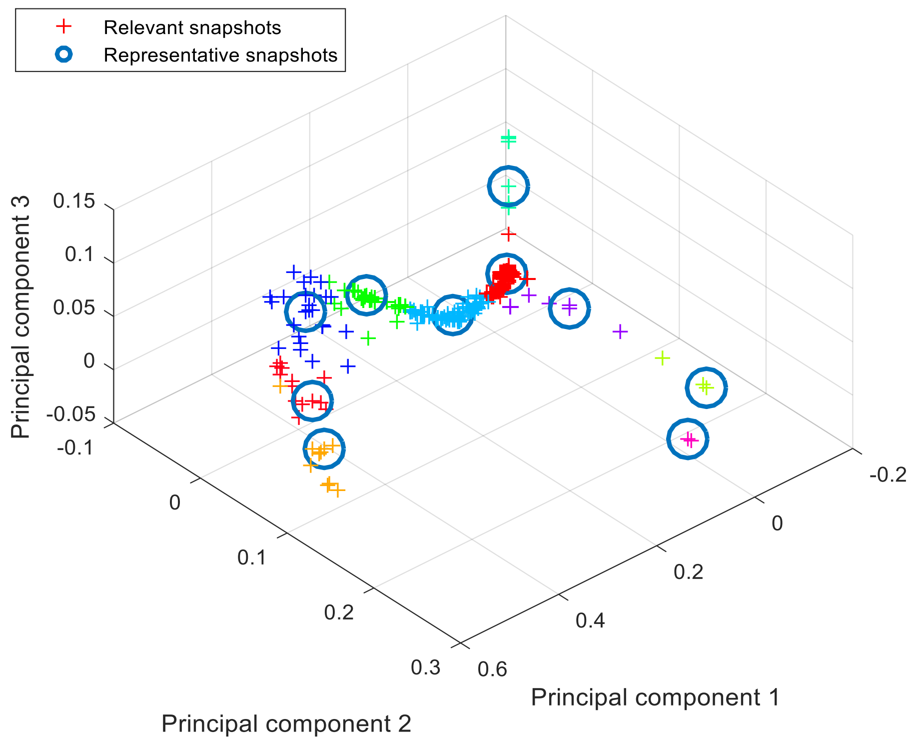

Fig. 4.4. Case A: representation of the relevant snapshots, the cluster they belong to, and their representatives in the space of the first three principal components of the reinforcement benefit space. Each color is associated with a specific cluster.

One may notice that most of the initial snapshots represented in the space of principal components are spread close to each other and aligned along the main axes. This is especially the case for the snapshots depicted in blue, red, and green, in the center of Fig. 4.4. These "compact straight line" patterns can be explained by the following reasons:

- The compactness of these groups of snapshots can be explained by the fact that, from an hour to another, load and intermittent generation among nodes hardly vary and, consequently, the benefit of installing a given line barely changes as well.
- The straight-line shape taken by some groups of snapshots can be explained by the correlation of benefit among multiple candidate lines. The fact that these alignments appears along the principal component axes is a natural result of the PCA, that is essentially rotating the original space such that the first axes are aligned with the main directions of variance. A positive correlation, that is, the benefit of two or more candidate lines simultaneously increasing from a snapshot to another, could mean that these candidate lines play a similar function in the system, at least for the group of snapshots for which this correlation occurs. A negative correlation, that is, the benefit of one candidate line decreasing when the benefit of another candidate line increases, on the contrary, could mean that these candidates play a largely different function.

The directions of the principal components could be used to identify transmission lines with similar function. This research work, however, is outside of the scope of the present thesis.

4.3.3. VALIDATION OF RESULTS

In order to assess how the number of clusters considered affects the accuracy of the solution computed, as well as to compare the merits of the clustering approach proposed with those of other approaches, the whole process of selection of snapshots here described is repeated several times, each one considering a different number K of clusters. Besides, snapshots are also classified into clusters, for several numbers of clusters, according to three other approaches.

In the first alternative approach considered, snapshots are clustered according to the net demand in nodes (demand net of RES generation).

The second alternative clustering approach applied is based on the scenario reduction technique proposed by authors in (Morales, Pineda, Conejo, & Carrion, 2009), whereby representative snapshots are selected according to the value for each snapshot of the total system operation and network investment costs. Network investment costs are annualized ones computed only considering the average snapshot throughout the year. Therefore, they are common to all snapshots. System operation costs for each snapshot are computed considering that this snapshot is the only one occurring in the whole year (the 8760 operation hours in the year are equal to this snapshot).

Lastly, the third alternative clustering approach applied here corresponds to the scenario reduction technique that groups together snapshots according to the demand, intermittent generation output, and conventional generator availability considered separately, which are the input parameters of the TEP problem, following the method proposed by authors in (Dupačová, Gröwe-Kuska, & Römisch, 2003).

These four methods are compared according to the efficiency of the network expansion plan computed, both in cases A and B, considering the representative snapshots and weights for them determined with the three methods. Then, for each method, the reduced, non-relaxed, TEP problem is solved as many times as different numbers of clusters considered. Each time, a new set of clusters, and, therefore, a new set of representatives and weights for them ρ_s, is computed with the method concerned. Considering these representatives and their weights, discrete transmission expansion plans are determined. Reinforcements computed in the reduced TEP problem are included in the grid to compute the resulting operation of the system considering the initial 8760 snapshots. Thus, the performance of each method for each case and number of clusters is assessed based on the total system costs resulting from implementing the corresponding TEP solution. Total system costs include the annualized investment costs of these reinforcements and the annual system operation costs resulting from the deployment of the former.

The evolution of total system costs with the number of clusters of snapshots computed, corresponding to the snapshot selection approach presented in this article and the three other approaches, is shown in Fig. 4.5 for case A. The same evolution, in case B, is shown in Fig. 4.6.

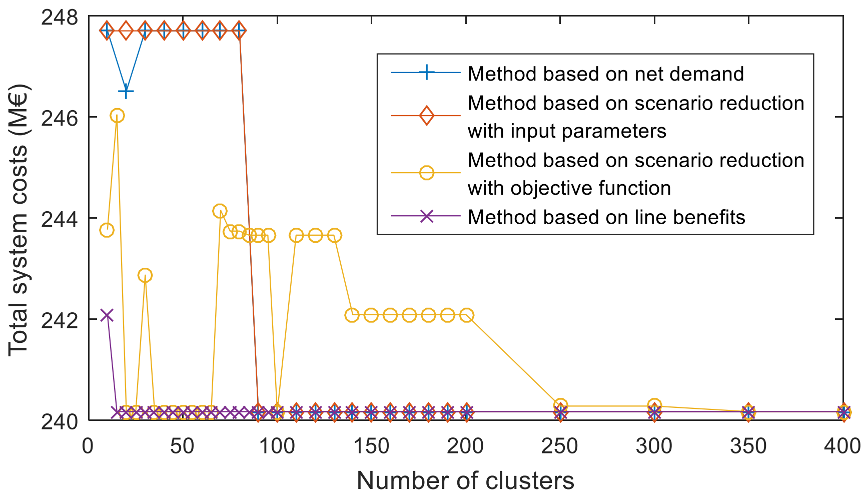

Fig. 4.5. Case A: Evolution of total system costs with the number of snapshots considered when these are selected 1) with the method proposed here, 2) according to the net demand in the system nodes (1st alternative approach), 3) with the scenario reduction technique based on total system costs (2nd alternative approach), and 4) with the scenario reduction technique based on input parameters (3rd alternative approach).

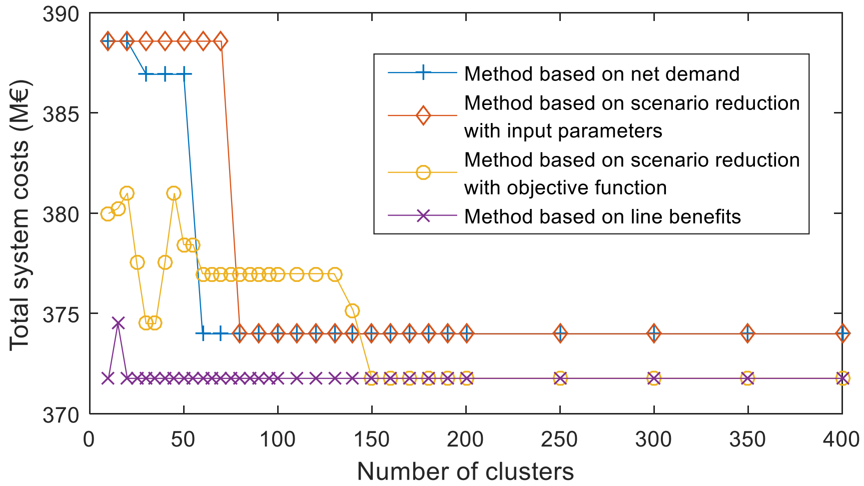

Fig. 4.6. Case B: Evolution of total system costs with the number of snapshots considered when these are selected 1) with the method proposed here, 2) according to the net demand in the system nodes (1st alternative approach), 3) with the scenario reduction technique based on total system costs (2nd alternative approach), and 4) with the scenario reduction technique based on input parameters (3rd alternative approach).

In case A, the optimal total costs associated with the network expansion plan computed solving the non-relaxed, non-reduced, TEP problem are equal to 240.17 M€. These are the minimum total system costs possible. The optimal TEP solution comprises reinforcements to lines 14-16, 15-21, 16-17, and 21-22. When applying the approach here proposed, non-relevant snapshots are always represented by 5 selected snapshots. As it can be seen in Fig. 5, both when snapshots are clustered according to their nodal net demand and when the scenario reduction technique taking as classification variables the input parameters of the TEP problem (the demand, RES generation output, and conventional generation availability) is applied, considering 90 clusters of snapshots is needed to compute the optimal expansion plan.

Selecting snapshots to be considered in the TEP problem according to the scenario reduction technique that we have taken as the 2nd alternative approach (making use of total system costs for each snapshot as classification variable) allows one to compute the optimal network expansion plan for some specific, relatively low, numbers of clusters. Thus, taking 20, 25, or between 35 and 65 clusters would result in the computation of the optimal expansion plan in the TEP analysis. However, for larger numbers of snapshots selected according to this approach, the expansion plan computed is no longer optimal.

This is in contrast with the fact that 15 snapshots selected according to the approach here proposed are enough to compute the optimal expansion plan. These snapshots include 10 relevant snapshots and 5 non-relevant ones.

The fact that the method classifying snapshots according to the nodal net demand and the scenario reduction technique proposed in (Dupačová et al., 2003) (the 3rd alternative approach assessed here) provide similar results is not surprising. After all, they both select representative snapshots according to similar parameters: the demand and the intermittent generation output, or the net of them. On the contrary, characterizing snapshots according to the total system costs for them, as in (Morales et al., 2009), and doing so according to the benefits network investments produce in each snapshot, allows one to group together those snapshots that have a similar impact on the objective function value and the decision variables in the TEP problem, respectively. Both the TEP objective function value and the decision variables (network investments) are outcomes of the TEP problem (though a simplified version of this problem is considered in both cases, of course). The instability in the goodness of the selection of snapshots made with the 2nd alternative approach may be explained by the fact that this approach was devised to select a reduced set of scenarios, while the problem at hand here is to select a representative set of snapshots. Scenarios comprise a multiplicity of snapshots occurring over a certain period of time. Computing the expansion plan to consider, in the characterization of snapshots, according to the average snapshot only may not be appropriate due to the large differences existing among snapshots occurring throughout a year. Besides, computing total annual system operation costs according to a single snapshot (the one being characterized in each case) is probably not accurate enough. Therefore, grouping operation situations according to the total system costs in them may not be the most appropriate approach to select representative snapshots, while it is probably much better adapted to the selection of representative scenarios.

In case B, the optimal expansion plan includes the same reinforcements as in case A plus the reinforcement of line 11-13. The total system costs of this optimal expansion plan are equal to 371.76 M€. First, clusters and their representatives to be considered in TEP studies are computed according to the net demand in the nodes in each snapshot, and according to the scenario reduction technique considering input parameters (the demand, RES generation output, and conventional generation availability) of the TEP problem as classification variables in each snapshot. In both cases, total system costs resulting from the deployment of the corresponding network expansion plan are still higher than those corresponding to the optimal expansion plan for very large numbers of snapshots selected (as large as 400, since total costs for the expansion plan computed considering 400 snapshots are still the same as those for the expansion plan computed considering 80).

Considering in TEP studies the snapshots selected according to the 2nd alternative approach in case B does not lead to the computation of the optimal expansion plan until the number of snapshots considered is 150. On the other hand, when applying the clustering approach here proposed, only 20 snapshots are needed to compute the optimal expansion plan (including 5 non-relevant ones).

The analysis of results is the same as that carried out of those obtained for case A. However, contrary to what occurs in case A, in case B the snapshot reduction methods focusing on the inputs to the TEP problem, instead of the outcomes of it, do not produce a selection of snapshots leading to the computation of the optimal expansion plan, even for very large numbers of snapshots considered.

The efficiency of a snapshot selection method can also be assessed according to the computational burden of the minimum size TEP problem considering the snapshots selected with this method that results in the optimal network expansion plan. Both the brute force method and the approach here proposed have been applied to compute the optimal expansion plan in case A using a machine with an Intel® Xeon® X5570 processor running at 2.93 GHz and 36 GB of RAM. Solving the non-reduced, non-relaxed, TEP problem, has taken more than 46 hours, whereas computing the same optimal expansion plan following the approach here proposed has taken about 80 minutes. More specifically, computing the relaxed TEP model has taken 30 minutes; computing the initial OPF and the OPF with optimal continuous reinforcement has taken 10 minutes each; computing the benefits of reinforcements within the relaxed expansion plan in each snapshot has taken 9 minutes; computing the principal components of the benefits of reinforcements across all relevant snapshots has only taken a few seconds; computing the representatives of the clusters of relevant snapshots and their weights for the seven clustering analyses conducted corresponding to seven different numbers of clusters has also taken a few seconds; and solving the reduced TEP problem considering the sets of snapshot representatives has taken less than 1 minute for each number of representatives. Lastly, computing the system economic dispatch in the 8760 snapshots of the whole year, when deploying the expansion plans computed, has taken 10 minutes.

Regarding the comparison with the rest of approaches assessed here, it has taken around 30 minutes in total in case A to compute, according to the 2nd alternative approach, the total system costs (classification variable) for all the 8760 snapshots in the target year considered separately, whereby a separate estimate of costs is computed for each snapshot. As for the method proposed here, computing the clustering variables (line benefits and operation cost savings) for all the snapshot in the year has taken around 60 minutes in total. On the other hand, grouping snapshots together and selecting the representative ones according to alternative approaches 2 and 3 has taken, on average, 3 minutes and a half for each number of clusters, while it has only taken a few seconds to select representative snapshots with the method here proposed starting from the clustering variables previously computed. The time complexity and accuracy of the several snapshot reduction methods applied to this case study have been summarized in table 4.2.

It can be noted that, although the time needed to compute (or select) the clustering variables is longest in our method, it becomes competitive in terms of computational burden when compared to the two methods based on scenario reduction techniques if the clustering algorithm is run 20 or more times to select the appropriate number of clusters. This is because the k-means clustering algorithm employed in our approach is much less computationally demanding than the scenario reduction algorithm used to aggregate snapshots in the two methods based on scenario reduction.

The two snapshot selection methods that take input system parameters, like the net demand, as clustering variables do not need to select, or compute, these clustering variables. However, these methods are highly inaccurate in terms of the representativeness of the snapshots selected when compared to the method proposed in this article. Thus, the number of snapshots to be selected to compute an efficient transmission expansion plan is much higher when making use of these two snapshot selection methods than when the method proposed is applied for this.

Therefore, although more computationally demanding in computing the clustering variables, the method proposed can be considered highly competitive against others in terms of the time it requires to select a good enough set of representative snapshots.

TABLE 4.2
TIME COMPLEXITY AND ACCURACY OF THE SNAPSHOT SELECTION METHODS

	Method based on net demand	Method based on scenario reduction with input parameters	Method based on scenario reduction with objective function	Method based on line benefits
Time needed to compute the clustering variables	0 min	0 min	30 min	60 min
Time needed to run the clustering algorithm (for each target number of clusters defined)	< 1 min	3,5 min	3,5 min	< 1 min
Total time needed (when running the clustering algorithm 20 times to select the appropriate number of clusters to define)	< 1 min	70 min	100 min	60 min
Accuracy: number of clusters needed to obtain the same solution as the non-reduced TEP problem	90 clusters	90 clusters	Obtained from 20 clusters but unstable before 250 clusters	15 clusters

Comparison of the time complexity and the accuracy of the different snapshot selection methods.

5. Network reduction for a TEP problem

5.1. Preliminary discussion

This chapter corresponds to our article "An Efficient Network Reduction Method for Transmission Expansion Planning using Multicut Problem and Kron Reduction".

In the context of the TEP problem, the structure of the network is certainly the most delicate dimension to reduce. Actually, the network is the structure in which candidate lines may or may not be installed, and congestion occurring within the network are the reasons why transmission expansion is required. Therefore, a network reduction procedure, necessarily achieved through the fusion or elimination of some buses, as seen in 2.2.2, produces a network representation in which the ends of relevant candidate lines may either no longer be represented, in the case of bus elimination, or be merged together, in the case of bus merging.

In (S. Lumbreras et al., 2015), authors cope with this issue by considering equivalent candidate lines in the reduced network. Once the TEP problem has been solved considering the reduced network model, the optimal set of equivalent candidate lines is later converted in line reinforcements in the fully detailed network.

In the present thesis, we aim to avoid this step and aim to keep the most relevant original candidate lines preserved within the reduced model. To do so, we first identify critical pairs of buses that should be preserved thanks to the linear relaxation described in section 3.4.2. Then, a network partition is performed, in which these critical pairs of buses are border buses, i.e. connected to inter-area lines. Finally, a bus elimination is achieved in which most of non-border buses are removed from the original network.

5.2. Methodology

In this section, the methodology applied to reduce the network to be used in TEP applications is described and discussed. The description of this methodology is divided into the steps that follow:

1. Identification of the critical pairs of buses;

2. Computation of the network partition;

3. Elimination of the non-border buses; and
4. Computation of the equivalent features of the links in the reduced network obtained.

A flowchart of the methodology applied is shown in Fig. 5.1.

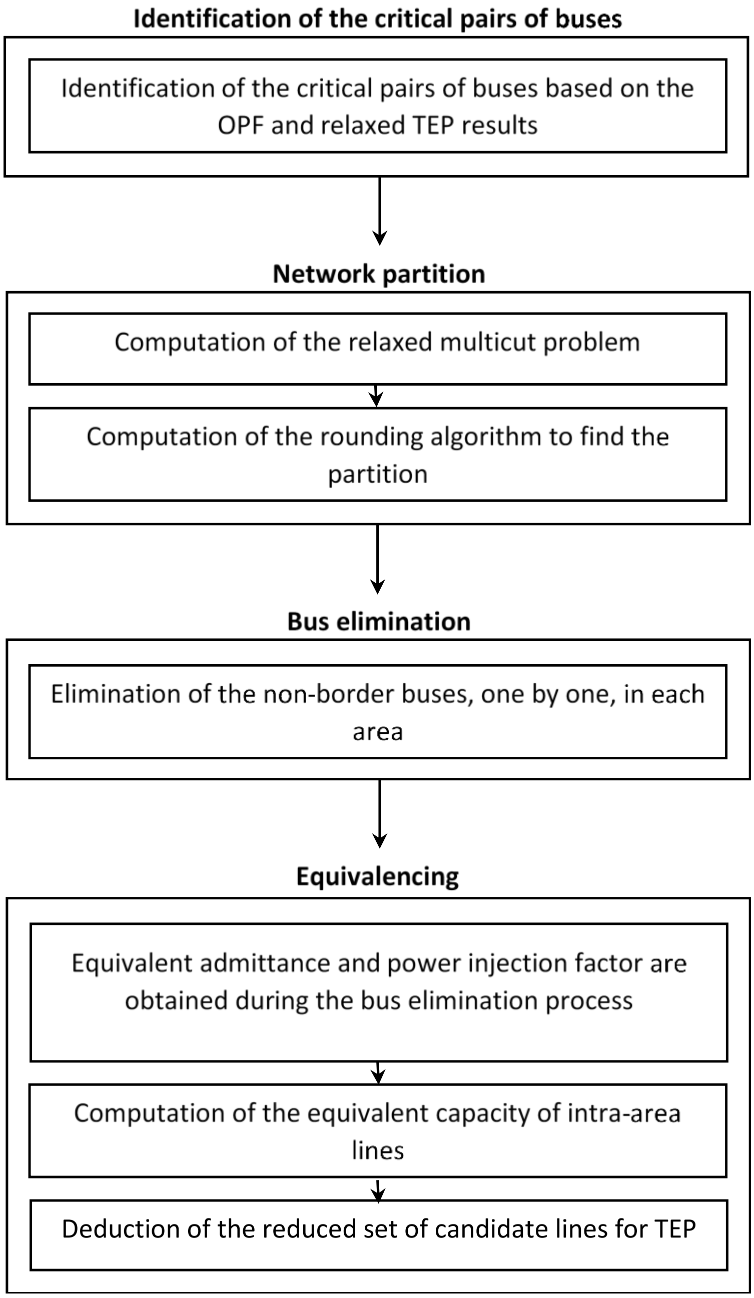

Fig. 5.1. Flowchart of the methodology applied for network reduction.

5.2.1. IDENTIFICATION OF THE CRITICAL PAIRS OF BUSES

When reducing the size of an electrical network in order to decrease the computation time of solving a TEP problem, it is critical to keep intact those lines that are the most impacted by the network expansion and be able to compute flows in these lines that are as close to the ones computed for the original network as possible. Lines that are congested in the network before its expansion are relevant, but the congestion of a few lines in it may disappear by installing a single line in a new corridor. This is why it is necessary to capture the potential effects of the network expansion on the operation of the network, and not only its causes (Desta Z. Fitiwi et al., 2015). In order to capture the effects of the network expansion on system operation, while reducing as much as possible the computational burden of the original TEP problem, a relaxed version of this problem, in which investment decisions are represented by continuous variables instead of binary ones, is considered to identify critical lines. The solution of this relaxed problem not only provides useful information about the possibly congested lines in the expanded network, but also reveals which new lines, in existing or new corridors, are likely to be built.

The optimal power flow (OPF), i.e. the power flows minimizing the operation costs, in the network before its expansion is computed. The relaxed TEP problem is also computed. We define the critical pairs of buses as those complying with any of the following conditions:

- Those buses at the two ends of those lines whose flow is reaching 100% of their capacity in at least one operating situation, according to the OPF solution in the network before its expansion,
- Those buses at the two ends of those lines whose flow is above 80% of their capacity in at least one operating situation, according to the solution of the relaxed TEP problem,
- Those buses at the two ends of the candidate lines which are partially built according to the solution of the relaxed TEP problem, i.e., those candidate lines for which the optimal investment decision variable in this problem is strictly greater than 0,
- Those buses at the two ends of already installed or candidate power flow controlling devices, such as direct current lines.

Regarding the choice of the congestion threshold, there is a tradeoff to be made between the reduction in the network model considered and the accuracy in representing the network congestion and the impact on it of the network investments in TEP analyses. On the one hand, this threshold should be as low as possible in order to keep as much information as possible on the congestions that can occur in the network and, therefore, to compute a TEP that is as efficient as possible when considering the reduced network. On the other hand, the number of critical buses to keep and, thus, the size of the reduced network and the TEP problem, increases when the congestion threshold decreases. The choice of the congestion threshold made above proved to be a good compromise for the two case studies described in section 5.3.

5.2.2. Network partition

5.2.2.1. Minimum multicut problem and appropriate weight

Our objective is to reduce as much as possible the size of the network while keeping intact the critical pairs of buses identified in 5.2.1 and representing accurately the flow in them. To keep these pairs of buses intact, we have concluded in section 2.2.2 that a Gaussian elimination (Dorfler & Bullo, 2013) (WARD (Ward, 1949) or REI (Oatts et al., 1990)) should be carried out instead of a quotient graph reduction. Moreover, the two ends of each critical pair should belong to different clusters, or areas, so that we can eliminate non-border buses in each area without affecting the potential power flows going through these critical lines.

Besides complying with the former requirements, in order to minimize the size of the reduced network, one may initially think of finding a partition of the network for which the number of inter-area links is as small as possible. However, minimizing the size of the network actually involves minimizing the number of border nodes, not that of inter-area links, since only border buses within each area are supposed to be kept in the network reduction process. Besides, the only nodes that we know in advance that need to be represented with certainty in the reduced network are the ones that are part of at least one critical pair of buses, which we shall call "critical buses" from now on. Then, the number of additional border buses, besides the critical ones, corresponding to each additional inter-cluster link defined is the number of non-critical buses in it. This implies that minimizing the size of the reduced network should involve assigning the links among clusters similarity weights that are proportional to the number of non-critical buses they are connected to and minimizing the overall similarity among clusters, or areas. Next, we provide the formulation of the problem to be solved to compute the envisaged network partition.

Let $\left(W_{i,j}\right)_{i,j\,\in\{1,\dots,N\}}$ be a matrix describing a weighted network of N buses, and let $(s_k, t_k)_{k\,\in\{1,\dots,K\}}$ be a list of K "sink-source" pairs of buses within this network. The weighted multicut problem (L. R. Ford, Jr. & D. R. Fulkerson, 1956) aims to find a partition of this network in which each bus s_k belongs to a cluster different from the one t_k belongs to, while minimizing the sum of the weight of the links connecting two different clusters. This problem can be formulated as in (5.1).

$$\min_{\substack{\text{all possible}\\ \text{clustering}}} \left(\sum_{\substack{\text{i and j belong to}\\ \text{different clusters}}} W_{i,j} \right) \tag{5.1}$$

s.t. s_k and t_k belong to different clusters

The multicut problem is particularly appropriate for the network partition problem we aim to solve when considering as "sink-source" pairs the critical pairs of buses defined above, and taking the initial electrical network as the one to consider. However, it is not the number of inter-cluster lines what we want to minimize, but rather the number of border buses, including critical ones, which need to be preserved. This can be easily achieved by defining, for each link (i, j) in the initial network, a weight $W_{i,j}$ corresponding to the number of non-critical buses in it. An example of a network for which these weights have been computed is depicted in Fig. 5.2.

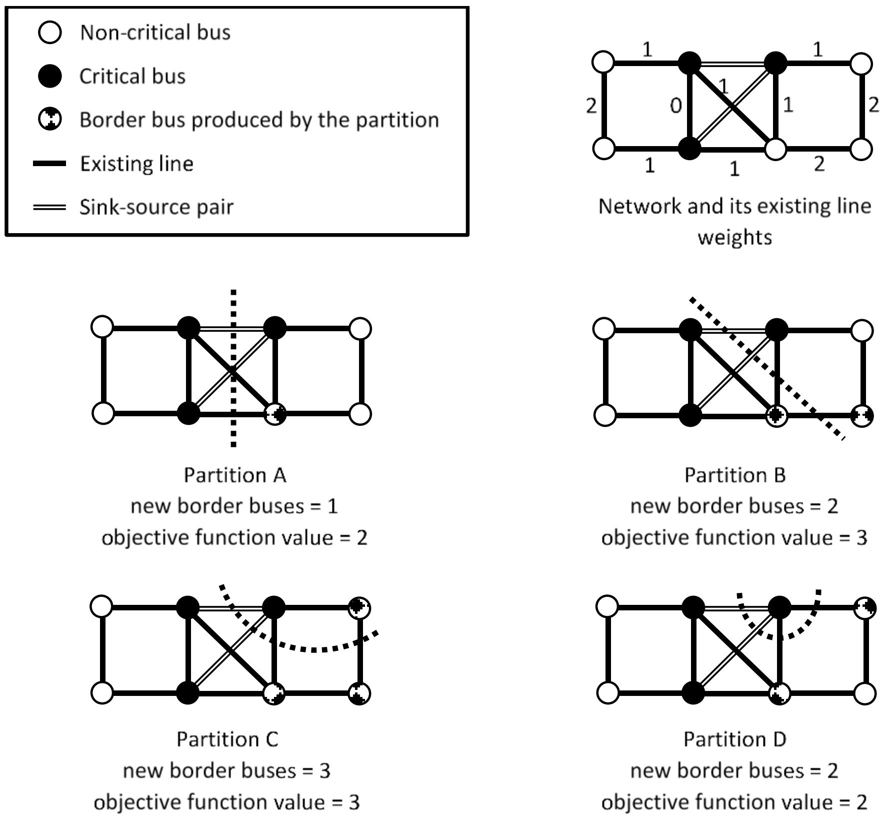

Fig. 5.2. Various partitions of a network computed solving the multicut problem. The weight of existing lines, computed according to the number of critical buses they are connected to, is represented in the first network at the top. The four partitions satisfy the sink-source pair constraints. The objective function value from equation (3) is always greater or equal to the number of border buses produced and, in partitions C and D, the two values are equal. Even though the partition producing the least amount of border buses is clearly partition A, the optimal solutions according to equation (3) will be both partitions A and D.

If there is no line directly connecting buses *i* and *j* in the initial network, separating these two buses in different clusters does not generate more border buses. Therefore, we can assign a weight $W_{i,j} = 0$ to these pairs of buses. However, if there is a line directly connecting buses *i* and *j* and the number of critical buses among them is $p \in \{0,1,2\}$, then separating buses *i* and *j* in different clusters will create at most $2 - p$ new border buses.

The above problem formulation in (5.1) may be considered appropriate to minimize the number of border buses within a partition. In reality, the number of additional border buses to be defined due to the inclusion of a link within the set of inter-area ones may not be the same as the number of non-critical buses in this link as it happens in partition A in Fig. 5.2. But, at least, the objective function in (5.1) is an upper estimate of the number of border buses corresponding to a certain partition which is reasonably close to the real number of border buses.

5.2.2.2. Relaxation of the multicut problem and a rounding algorithm

A practical formulation of the weighted multicut problem is that in (5.2)-(5.5), where distance variables $d(i,j)$ are introduced (Chekuri & Madan, 2017). The variable $d(i,j)$ is equal to 0 when buses i and j belong to the same area, whereas it is equal to 1 when they belong to different areas.

$$\min\left(\sum_{(i,j)\,\in\,E} d(i,j)\,W_{i,j}\right) \tag{5.2}$$

$$\text{s.t.} \quad d(s_k,t_k) \geq 1,\ \ k \in \{1,\dots,K\} \tag{5.3}$$

$$d(u,v) + d(v,w) \geq d(u,w),\ \ u,v,w \in \{1,\dots,N\} \tag{5.4}$$

$$d(u,v) \geq 0,\ \ u,v \in \{1,\dots,N\} \tag{5.5}$$

Equation (5.2) is an alternative way to express the objective function in (5.1) using distance variables. Here, E is the set of lines in the initial network, i.e. the set of existing corridors. Equation (5.3) guarantees that buses s_k and t_k belong to different clusters. Equations (5.4) and (5.5) ensure that $d(i,j)$ is behaving like a distance by being positive and by satisfying the triangle inequality. Indeed, a distance between two objects is always positive. This property is ensured by equation (5.5). Moreover, if a bus u belongs to the same area as a bus v, and bus v belongs to the same area as a bus w, then the buses u and w are deemed to belong to the same area. This property is enforced by equation (5.4).

The formulation (5.2)-(5.5), where $d(i,j)$ are binary variables, is just a way to express the problem (5.1) that is “easy to understand” by the solver, and the problem formulated in (5.2)-(5.5) is strictly the same as the one formulated in (5.1).

Formally, $d(i,j)$ is a binary variable, and the problem described by (5.2)-(5.5) is a MILP problem. This problem is NP-hard (Garg, Vazirani, & Yannakakis, 1996) and, therefore, computationally intractable for large networks. In the literature, the problem is usually relaxed by linearizing the variable $d(i,j)$, which allows linear programming (LP) solvers to quickly find an approximate solution. Then, a rounding algorithm (Garg et al., 1996) is applied which converts the optimal solution of the relaxed problem into an integer solution.

This rounding algorithm guarantees finding an integer solution within an integrality gap of $O(\ln(K))$ (Garg et al., 1996). Finally, the resulting network partition is represented by the connected components of the initial network, after removing all the links (i,j) for which $d(i,j) = 1$. Each connected component of this modified network represents an area of the partition. The connected components of an undirected graph can be computed relatively fast, in $O(N + |E|)$ (Tarjan, 1971).

The main advantages of this partition method over the common network partition methods based on (2.1) are the following:

- It guarantees that a set of pairs of buses is preserved in the subsequent reduction process, by forcing their ends to belong to different clusters; this feature is particularly interesting for TEP problems,

- There is no need to specify the number of clusters in advance; the best number of clusters is found by the algorithm,
- The last step of the rounding algorithm ensures that all the nodes within each cluster are connected by nodes of this same cluster; there is no need of post-processing the clusters to ensure this is achieved,
- The integrality gap between the integer solution of the multicut problem obtained with the rounding algorithm and the optimal solution of the relaxed multicut problem is guaranteed to be relatively low.

5.2.3. BUS ELIMINATION

Once a partition has been computed and the border buses have been identified, non-border buses are eliminated through Gaussian elimination (Dorfler & Bullo, 2013). The Gaussian elimination of the non-border buses in a given area can be performed via Kron reduction, either in a single step, or in multiple steps, one bus at a time (Dorfler & Bullo, 2013). The latter iterative method results in an update of the admittance matrix of the network, and its operating condition, in each step according to equation (5.6) and (5.7), respectively.

$$Y_{i,j}' = Y_{i,j} - \frac{Y_{i,k}Y_{k,j}}{Y_{k,k}},\ i,j\ \in \{1, \dots, k-1, k+1, \dots, N\} \tag{5.6}$$

$$P_m' = P_m - \frac{Y_{m,k}P_k}{Y_{k,k}},\ m \in \{1, \dots, k-1, k+1, \dots, N\} \tag{5.7}$$

Where $\left(Y_{i,j}\right)_{i,j\ \in\{1,\dots,N\}}$ and $(P_m)_{m\ \in\{1,\dots,N\}}$ are the admittance matrix and the power injection vector of the N-bus electrical network before removing bus k, and $\left(Y_{i,j}'\right)_{i,j\ \neq k}$ and $(P_m')_{m\ \neq k}$ are the admittance matrix and the power injection vector of the $(N$-$1)$-bus electrical network after removing bus k. As seen in 3.3, even if the electrical network is composed of parallel existing lines, the two-dimensional matrix $\left(Y_{i,j}\right)_{i,j\ \in\{1,\dots,N\}}$ can be defined by calculating the equivalent admittance $Y_{i,j}$ in each corridor (i,j) as $Y_{i,j} = \sum_c Y_{ijc}$.

Computing Kron reduction in multiple steps presents two advantages. First, it avoids computing an inverse matrix, which is advisable for complexity reasons. Second, eliminating all the non-border buses from a given area may lead to a dense subnetwork, which comprises many lines (Rommes & Schilders, 2010). However, reducing the size of a network is interesting only if the density of the reduced network stays relatively low. Therefore, it is advisable to avoid eliminating the last non-border buses if their elimination dramatically increases the number of lines. The application of these conditions is illustrated in Fig. 5.3.

When eliminating one bus at a time, the order in which buses are eliminated is important. One of the methods suggested by authors in (Tinney & Walker, 1967), and commonly used in the literature, is to eliminate, at each step, the non-border bus with the lowest degree, i.e. the non-border bus with the lowest number of lines connected to it.

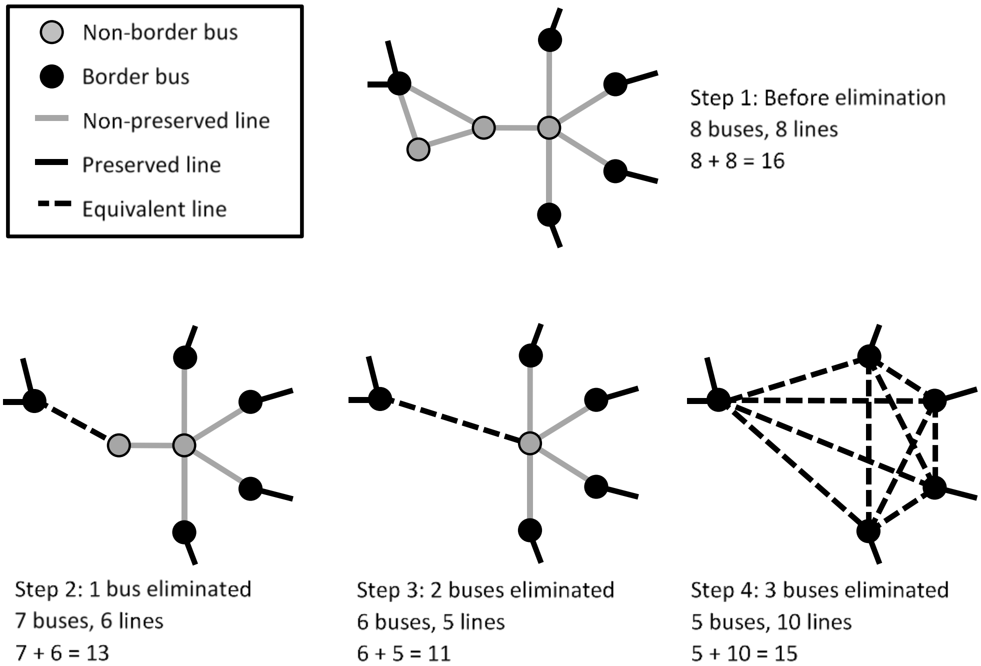

Fig. 5.3. Iterative Kron reduction process in an area. The elimination of all the non-border buses, at step 4, results in a dense subnetwork, with every pair of buses connected by a line. To avoid that, it is advisable to stop the reduction at step 3, in which the total number of buses and lines is minimum.

This method has been experimentally proven to result in subnetworks with low numbers of buses and lines. This is the elimination order considered in our work.

For a given area, we iteratively eliminate non-border buses until none of them remains. At each step, we record the number of lines and buses remaining in that area. Finally, the network used to represent that area is the one for which the total number of lines and buses in that area is minimum and the density of the grid remains low. This corresponds to step 3 in Fig. 5.3.

5.2.4. Equivalencing

5.2.4.1. Buses and existing lines in the reduced network

The admittance of the equivalent lines is iteratively computed according to equation (5.6). The fraction of the demand and generation in an eliminated bus allocated to each remaining bus is iteratively computed according to equation (5.7). The admittance and capacity of the lines crossing two areas are not affected by the Gaussian elimination of non-border buses. However, an equivalent capacity should be computed for the intra-area lines. Their capacity is computed according to the method described in (Jang et al., 2013). As explained in (Jang et al., 2013), the capacity computed in this way is most of the times the exact one that enables the same transfers of power as in the original network. Otherwise, an upper estimate is provided by the method in (Jang et al., 2013).

5.2.4.2. Candidate lines in the reduced network

The advantage of applying the Kron reduction to the non-border buses inside each area is that it preserves the operating conditions of the lines crossing two areas. Therefore, if a new line is built between two areas in the reduced network, the power flow going through it should be exactly the same as the one that would be computed for the same new line installed in the non-reduced network. The candidate lines that can be built in the reduced network are all those whose ends are border buses located in two different areas. The other candidate lines are discarded because their installation within an area would change the outcome of the Kron reduction inside this area. Discarding these candidate lines, however, should not have a dramatic impact on the TEP solution knowing that, according to the definition of critical pairs of buses described in section 5.2.1, they are neither partially built according to the solution of the relaxed TEP problem, nor parallel to any congested line.

5.3. CASE STUDY

The proposed method for the reduction of the network to be considered in TEP has been applied to one case study based on the standard IEEE 118-bus system, and one 2088-bus case study based on parts of the European power grid.

5.3.1. SYSTEMS DESCRIPTION

5.3.1.1. CASE BASED ON IEEE 118-BUS SYSTEM

Most of the system features are the ones of the standard IEEE 118-bus system. (“IEEE 118-bus, 54-unit, 24-hour system,” n.d.) The power system is depicted in Fig. 5.4.

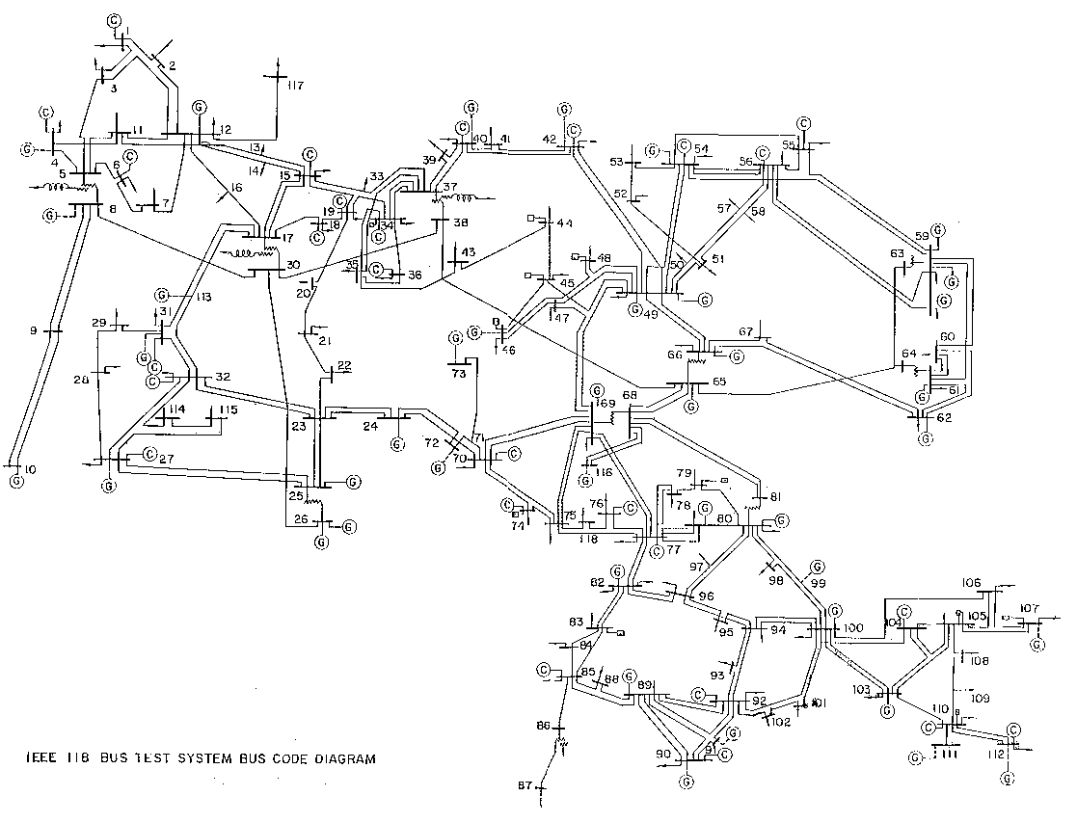

Fig. 5.4. Illustration of the standard IEEE 118-bus Test System (“IEEE 118-bus, 54-unit, 24-hour system,” n.d.)

This power system comprises 118 buses and 179 lines. Generation parameters including the bus location of the generators, their variable production cost, and their production capacity, have been obtained from (L. Wang et al., 2015). Besides, the 9 candidate lines defining new corridors have also been obtained from (L. Wang et al., 2015). Furthermore, for each of the 179 existing line in this system, an additional candidate reinforcement with the same features is considered. We have assumed in this paper that the investment costs of these candidate reinforcements are proportional to the product of their capacity by their reactance. This is because the investment cost of a line increases with its thermal capacity. Moreover, the investment cost of a line is generally proportional to its distance, which is roughly proportional to its reactance. Hence, the total number of candidate lines considered is 188. Transmission losses are neglected. Besides, only variable production costs are considered for generation units. In order to have sufficient operation variability, 300 operating situations, accounting for a total duration of 8760 hours, are taken into account. The load profiles and the RES production output profiles have been generated based on real load and RES output profiles for the European power system, which have been scaled to be well adapted to the load level of the IEEE 118-bus system. Its average value in each bus is set to be equal to the demand level defined for this bus in the original power system. More details about the candidate line and generation data can be found in (L. Wang et al., 2015).

5.3.1.2. CASE BASED ON PARTS OF THE EUROPEAN POWER GRID

ENTSO-E publishes an interactive map of the European electrical network ("ENTSO-E Transmission System Map," 2018). Authors in ("GridKit extract of ENTSO-E interactive map," 2016) have extracted the network data from this map. This data has been considered in this. The high voltage grid topology, together with the bus coordinates, the number of circuits per corridor, and the voltage level of these circuits, were directly available from the network data extraction. Line features (capacity and admittance) have been deduced based on their voltage level, their number of circuits and their length. The location and features of most generators, including their technology, and their capacity were, as well, available from this grid data extraction. Data for the rest of generators, including the solar and wind generation, has been deduced based on information retrieved from several sources provided by Transmission System Operators in Europe, as well as from the ENTSO-E website. Twenty operating situations, accounting for a total duration of 8760 hours, are taken into account. The profile of the load in each bus and each snapshot has been deduced based on real load and RES output profiles for the European power system. In order for the case study to be tractable when addressed formulating a MILP TEP problem, only five European countries have been considered within it, namely Portugal, Spain, France, the United Kingdom, and the Republic of Ireland. In total, the system in the case study comprises 2088 buses and 2953 lines, including AC lines and DC lines. The total number of candidate lines considered is 212.

5.3.2. APPLICATION OF THE METHODOLOGY

The OPF in the network before its expansion, and the relaxed TEP problem, are solved for the two case studies in order to identify the critical pairs of buses in each. In the 2088-bus case, the congestion threshold for the lines' flows in the solution of the relaxed TEP problem is set to 90% instead of 80%, in order to increase the network reduction ratio.

In each case, the line's weight to be used in the multicut problem is defined according to the connection of these lines with critical buses, as explained in the methodology section. Then, the relaxed version of the weighted multicut problem, described by equations (5.2)-(5.5), is solved. The value of the objective function represents a lower bound of the optimal integer solution. Applying the rounding algorithm in (Garg et al., 1996), a good enough integer solution of this problem is obtained. The value of the objective function for the integer solution is an upper estimate of the number of new border buses really produced by the partition computed. The largest areas produced by the associated partition are illustrated in Fig. 5.5.

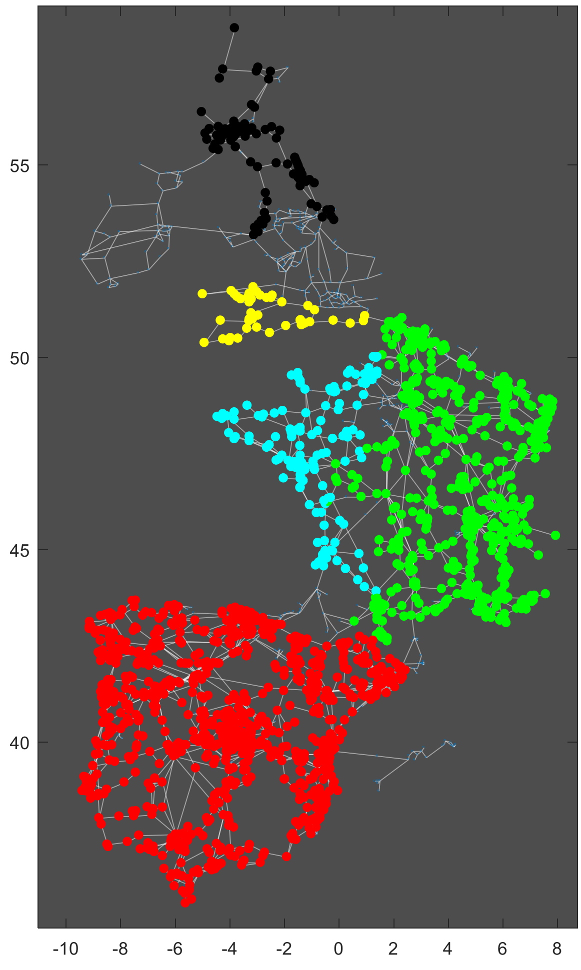

Fig. 5.5. Illustration of the network partition generated by the proposed method on the 2088-bus case. For the sake of clarity, only the 5 largest areas have been represented by coloring their buses in red, green, blue, black, and yellow. Numerous areas are composed of a single node and some of them are located in Spain, in the middle of the large red area.

Afterwards, in each area, non-border buses are eliminated one by one, according to the elimination order described in (Tinney & Walker, 1967), as long as the total number of buses and lines is decreasing. In practice, only 2 non-border buses are left in the reduced network in the 118-bus case, whereas 55 non-border buses are left in the 2088-bus case. This results in reduced networks comprising 48, and 518 buses, respectively, as shown in table 5.1. The resulting number of existing lines is 91, and 1067, respectively. The admittance matrix of the reduced network, together with the power injection factor, is updated, at each step of the bus elimination process according to equations (5.6) and (5.7). The capacities of the lines interconnecting the several areas remain the same as in the original network, whereas the capacity, or an upper estimate of the capacity, of the equivalent intra-area lines defined is computed following the method in (Jang et al., 2013). Only those candidate lines from the original network whose ends are border buses and belong to two different areas are proposed as candidate lines to be considered in the TEP problem to be computed for the reduced network, which correspond (table 5.1) to 48, and 114 candidate lines, in the 118-bus case, and the 2088-bus case, respectively. The other candidate lines in the original network are discarded. More details about the reduction process are provided in table 5.1.

TABLE 5.1
SUMMARY OF THE REDUCTION PROCESS

	118-bus case	2088-bus case
Initial number of buses	118	2088
Initial number of existing lines	179	2953
Initial number of candidate lines	188	212
Number of critical pairs of buses	22	346
Number of critical buses	30	392
Solution of the relaxed multicut	11	66
Solution after rounding	19	76
Number of areas in the partition	11	175
New border buses generated	16	71
Total number of border buses	46	463
Buses in the reduced network	48	518
Existing lines in the reduced network	91	1067
Candidate lines in the reduced network	48	114

Summary of the reduction process

In the 118-bus case, the number of buses and lines in the reduced network is between 2 and 3 times smaller than in the detailled network. In the 2088-bus case, the number of buses and lines in the reduced network is between 3 and 4 times smaller than in the detailled network. The resulting reduced network of the 2088-bus case is depicted in Fig. 5.6.

5.3.3. VALIDATION OF RESULTS

In order to assess the efficiency of the network reduction process here described, the MILP TEP problem is solved for the two case studies, both considering the original network and the reduced network. The allowed integrality gap is fixed to 0.05% for the 118-bus case, and to 5.5% for the 2088-bus case. The larger integrality gap in the 2088-

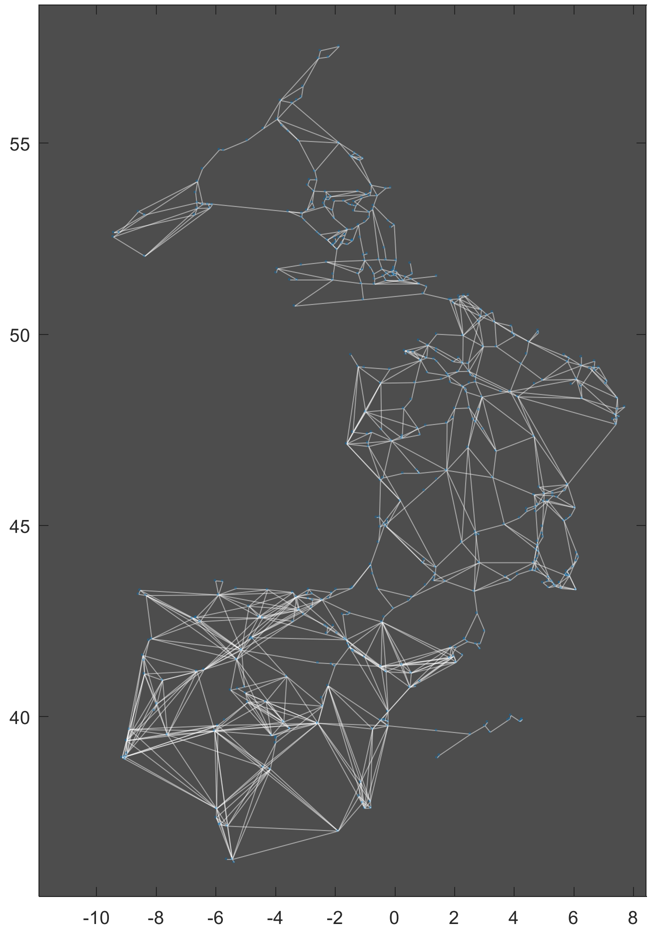

Fig. 5.6. Illustration of reduced network generated by the proposed method on the 2088-bus case.

bus case was necessary in order to solve the TEP problem in a reasonable amount of time. The solutions are compared in table 5.2. The time required to apply the reduction process, and to solve the TEP problem considering the reduced network, on the one hand, and to solve the TEP problem considering the complete network, on the other hand, are also shown in this table.

In the 118-bus case, the network investment decisions computed as optimal when considering the reduced network, within a 0.05% error gap, are exactly the same as the ones computed as optimal when considering the complete network. In the 2088-bus case, the optimal network investments computed when considering the reduced network are different from the optimal network investments computed when considering the complete network. However, the system total costs computed for the complete network when implementing the set of network investments computed considering the reduced network amount to 44575 M€. These are slightly lower than the system total costs resulting from implementing the set of network investments computed considering the complete network. This is because the integrality gap of 5.5% causes the network investment

TABLE 5.2
TIME COMPLEXITY AND ACCURACY OF THE NETWORK REDUCTION METHOD

<table>
<tr><th></th><th colspan="2"></th><th>118-bus case</th><th>2088-bus case</th></tr>
<tr><td rowspan="3">Results of the TEP problem</td><td colspan="2">Comparison of investment decisions between the complete network and the reduced network</td><td>same investment decisions</td><td>different investment decisions</td></tr>
<tr><td rowspan="2">Total costs (M€)</td><td>with complete network</td><td>446.37</td><td>44717</td></tr>
<tr><td>with reduced network</td><td>446.31 (446.37)*</td><td>44654 (44575)*</td></tr>
<tr><td rowspan="7">Time needed to compute (min)</td><td rowspan="4">Reduction process</td><td>Identification step (OPF and relaxed TEP)</td><td>6</td><td>2</td></tr>
<tr><td>Relaxed multicut problem</td><td>1.5</td><td>65</td></tr>
<tr><td>Rounding algorithm</td><td>< 1</td><td>< 1</td></tr>
<tr><td>Kron reduction and equivalencing</td><td>< 1</td><td>< 1</td></tr>
<tr><td colspan="2">Total reduction process</td><td>8</td><td>67</td></tr>
<tr><td colspan="2">TEP problem with reduced network</td><td>30</td><td>463</td></tr>
<tr><td colspan="2">TEP problem with complete network</td><td>180</td><td>2482</td></tr>
</table>

Time complexity of the network reduction method, and comparison of the TEP results in terms of investment decisions and total costs, with and without reducing the network. In the 2088-bus case, the TEP solution considering the complete network differs from the one considering the reduced network, but both solutions satisfy the integrality gap of 5.5%.

* Regarding the total costs of the solution computed with the reduced network, the first value is the total cost of this solution when considering the reduced network, whereas the second value in parenthesis is the total cost when considering the investment decisions of this solution and the complete network.

solution computed when solving the TEP problem considering the complete network to be suboptimal. The investment decisions computed in the 2088-bus case when solving the TEP problem considering the reduced network are, therefore, slightly better than the ones computed when solving the TEP problem considering the complete network. However, the solutions computed considering the complete and the reduced network models can be deemed to be equally "optimal", subject to an integrality gap of 5.5%.

The errors made when computing the total costs in the 118-bus case, and the 2088-bus case, are, respectively, equal to 0.01%, and 0.2%. The biggest error made, corresponding to the larger case, may be due to the higher congestion threshold we used to identify critical pairs of lines.

The total computation time needed to compute the reduced network model and solve the TEP problem considering the reduced network is, in both cases, 5 times shorter than the computation time needed to solve the TEP problem when considering the complete network.

5.4. COMPARISON WITH ALTERNATIVE METHODS

We compute a reduced network model of the 2088-bus case according to four alternative partition methods from (Shayesteh et al., 2015) and (Cuffe & Keane, 2017), and one alternative bus aggregation method from (Oh, 2012).

5.4.1. NETWORK PARTITION BASED ON ALTERNATIVE METRICS

We compute the network partition of the 2088-bus case based on the Thevenin impedance distance (Z^{thev}), the power transfer distance (PT), and the shortest path distance (SP), as described in (Cuffe & Keane, 2017). We compute, as well, the network partition based on the ATC between each pair of nodes and applying spectral clustering, following the procedure in (Shayesteh et al., 2015). A summary of the results computed is presented in table 5.3.

TABLE 5.3
NETWORK PARTITIONS BASED ON ALTERNATIVE METRICS

Alternative metric	Z^{thev}	PT	SP	ATC
Number of connected clusters	179	187	180	214
Number of border buses	696	803	917	741
Buses in the reduced network	724	816	964	748
Existing lines in the reduced network	1291	1376	1557	1271
Candidate lines in the reduced network	26	57	48	58
Total costs of the TEP solution with reduced network (M€)	52729	53622	46691	44652
Total costs of the TEP solution with complete network (M€)	71775	71818	52889	52286

Summary of the alternative network partition results

Because we use the k-means algorithm to cluster the network buses, the number of clusters has to be defined beforehand. We have considered the same number of clusters as when applying our methodology to compute the network partition, i.e. 175. For any of the aforementioned partition methods, the number of internally connected areas eventually defined is slightly higher than that of clusters. Once the areas have been defined, we apply Kron reduction to reduce the number of buses considered. Once a reduced network model has been computed, the only candidate lines that can be taken into account are those whose end nodes are explicitly represented in this reduced network and belong to different network areas. Finally, the TEP is computed considering this reduced network model. We take the total system operation plus network investment costs corresponding to the TEP solution computed for each network partition as being representative of the accuracy of this partition. The aforementioned costs are determined for the fully detailed network when setting the network investment decisions to those in the TEP computed considering the reduced network for each distance metric.

System operation plus network investment costs are lowest when partitioning the network based the ATC amongst nodes. Total costs in this case amount to 52286 M€, which are 17% higher than those corresponding to the TEP computed considering the reduced network model resulting from applying our method, which amount to 44575 M€. This gap is due to the fact that, when considering the ATC-based distance among buses, only a small fraction of all the critical pairs of buses are preserved (explicitly represented) in the reduced network, since 50% of the congested lines have their two ends within a single area. The fraction of preserved congested lines is even lower when partitioning the network using the three other metrics. The congestion affecting those critical lines that are not preserved cannot be accurately represented in the search for an optimal expansion plan. On the contrary, the likely-to-get-congested lines are preserved in the network reduction process when applying our method by enforcing that their two ends must be within two different areas.

5.4.2. Bus aggregation method based on PTDF matrix

Starting from the partition of the 2088-bus network obtained in our methodology, we follow the procedure described in (Oh, 2012) to build the quotient graph and its equivalent line and bus features.

The equivalent admittance of the lines in the reduced network, which represent the inter-area connections in the complete network, are computed by solving an error-minimization problem. We compute the capacity of the equivalent line between two areas as the maximum net amount of power flowing between these two areas in any of the operation situations considered, in any of the two directions, when considering the original, fully detailed, network. Regarding the definition of the set of candidate lines to be considered for the reduced network, there is no clear indication in the literature on how their features should be computed. For the sake of simplicity, we keep, in the reduced network, the original features of the candidate line in the complete network. Candidate lines whose two ends are in the same area cannot be represented in the reduced network and, therefore, they are discarded when computing the TEP considering the reduced network.

When computing the TEP considering the reduced network, the optimal total costs are 10 times higher than the optimal total costs when computing the TEP with the complete network.

The very inaccurate results obtained with this network reduction method can be explained by various factors: first of all, the equivalent line features in the reduced network are computed thanks to an error-minimization problem. Given that the error made is non-zero, we are necessarily incurring in some inaccuracies when trying to represent the flows in the detailed network by those in the reduced one. Moreover, there is no reason for the candidate lines to consider in the reduced network to have the same features as the candidate lines considered in the fully detailed network. However, there are not, so far, appropriate methods to compute the features of the equivalent candidate lines to be considered in the reduced network when computing its expansion.

6. Search space reduction for a TEP problem

6.1. Preliminary discussion

6.1.1. Preliminary candidate lines filter

In this case, we are considering a TEP problem in which candidate transmission assets are AC lines, AC/DC converters, and DC lines. Compared to AC lines, DC lines have two major advantages, according to our assumptions: the power flows in them are fully controllable, i.e. they do not obey KVL, and their installation cost over a certain line distance is lower than that of AC lines. However, DC lines can only be connected to the AC grid through an AC/DC converter. What is more, fully controlling the flow on each DC line separately requires installing this type of devices. Because of the installation cost of such converters, it is, normally, not cost-effective to install DC lines below a certain length, as authors explain in (Kenzelmann, Rufer, Dujic, Canales, & Novaes, 2015). This critical distance, also known as the "break-even" distance, as represented in Fig. 6.1, is lying between 500 and 800 km for overhead lines and between 150 and 200 km for underground cables (Kenzelmann et al., 2015). For submarine lines, the break-even distance was much shorter, around 50 km, which is why the vast majority of transmission projects crossing the sea are submarine DC lines ("ENTSO-E Transmission System Map," 2018).

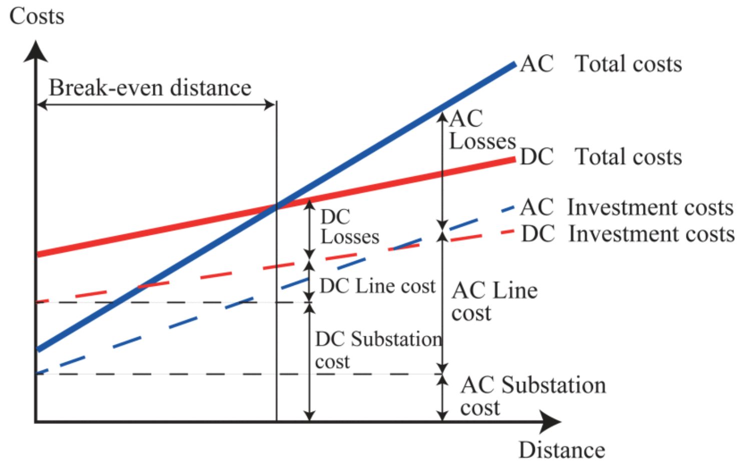

Fig. 6.1 Proportional break-even distance of DC lines versus AC lines (Kenzelmann et al., 2015)

Above this break-even distance, candidate AC lines have no advantages over candidate DC lines, neither from an economic, nor from a flexibility point of view. This is why, in TEP studies, onshore AC lines whose length is above the break-even distance, and offshore AC lines, are generally discarded from the search space.

Apart from these considerations, other environmental or political considerations may result in a further reduction of the set of potential corridors to reinforce within the transmission grid (Battaglini, Komendantova, Brtnik, & Patt, 2012). This allows us to exclude numerous candidate lines before applying a search space reduction method.

6.1.2. Efficient search space reduction technique in a TEP context

The goal of the present section is to describe an efficient search space reduction technique that relies on optimization techniques, instead of metaheuristic algorithms. As discussed in section 2.3, an interesting option to guide the search space reduction is considering the solution of a "hybrid" relaxed TEP model where the candidate circuit capacity to be built in each corridor is continuous and unbounded. This TEP model is called "hybrid" because, within it, the flows in the existing AC lines satisfy the DCLF constraints, whereas the flows in candidate AC lines satisfy the TLF ones. However, as discussed in 3.1, the TLF model is not accurate enough for AC lines. Therefore, there is a need to find a way to enforce, to the extent possible, the KVL in candidate AC lines. Moreover, the reduction process should not involve computing a MILP version of the TEP problem, to keep the search space reduction problem tractable. Finally, avoiding forcing the installation of certain candidate lines during the candidate line selection process is necessary in order not to compute local optima, such as in CHA methods.

In order to limit the lack of accuracy resulting from employing the TLF model for candidate AC lines, the method proposed in this thesis computes a linear relaxation of the TEP problem in which the KVL in candidate AC lines is enforced to the extent possible. We achieve this by solving, in the first iteration, a "hybrid" relaxed TEP problem in which only flows in existing AC lines satisfy DCLF constraints, while flows in candidate DC lines satisfy TLF ones. Then, in each of the subsequent iterations, making use of the optimal network investment solution computed in the previous iteration, we refine the problem formulation by including disjunctive constraints, similar to constraints (3.18), in order to enforce, as much as possible, DCLF constraints in relevant candidate AC lines.

The advantages of the proposed method are listed below:

- we identify the corridors where reinforcements should be most relevant based on the optimal value computed for the relaxed investment decision variables. This information not only identifies which corridors should be expanded, but also to which extent (which line capacity) they should be expanded, striking a tradeoff between the benefits these investments produce and their investment costs,
- we, progressively, include disjunctive constraints that lead the flows in candidate AC lines to increasingly satisfy KVL, making the optimal investment decisions more accurate,

- the problems solved during the search space reduction process are all LP, allowing the reduction process to be conducted quickly,
- no investment decision variable is fixed during the search process, thus avoiding, as much as possible, converging towards a local optimum.

6.2. METHOD

In this section, the method applied to reduce the search space of the TEP problem is described and discussed. The description of this method is divided into the steps that follow:

1. Computation of the relaxed TEP problem solution considering an unbounded number of candidate lines per corridor, and a TLF model in candidate AC lines;
2. Iterative computation, until convergence is achieved, of the relaxed TEP problem solution considering a bounded number of candidate lines per corridor, and a relaxed KVL model in candidate AC lines; and
3. Computation of the equivalent candidate lines to consider when using binary network investment variables.

A flowchart of the methodology applied is shown in Fig. 6.2.

6.2.1. RELAXED TEP PROBLEM WITH AN UNBOUNDED NUMBER OF CANDIDATE LINES PER CORRIDOR

For the sake of simplicity, we assume for the rest of the chapter that candidate lines are all AC lines, and that candidate AC lines in the same corridors, i.e. between the same pairs of buses, have the same technical features, that is, the same power capacity and the same reactance. Thanks to this, we can refer to each individual candidate line by the pair of buses *(i,j)* between which it can be installed. The general case that considers both AC and DC candidate lines, and parallel AC candidate lines with different features, can be addressed analogously to this specific case.

The equivalent, overall, admittance $Y_{i,j}^{cand}$, respectively equivalent capacity $\overline{f_{i,j}^{cand}}$, of a set of parallel candidate lines installed in a given corridor is calculated as the admittance $Y_{i,j,1}^{cand}$, respectively capacity $\overline{f_{i,j,1}^{cand}}$, of a single candidate line in this corridor multiplied by the number of candidate lines installed $x_{i,j}$ (6.1), (6.2).

$$Y_{i,j}^{cand} = x_{i,j} Y_{i,j,1}^{cand} \tag{6.1}$$

$$\overline{f_{i,j}^{cand}} = x_{i,j} \overline{f_{i,j,1}^{cand}} \tag{6.2}$$

Because of this, equations (3.12) and (3.14) that rule the flow of power through individual candidate lines can be generalized for the case of multiple candidate lines per corridor, assuming that:

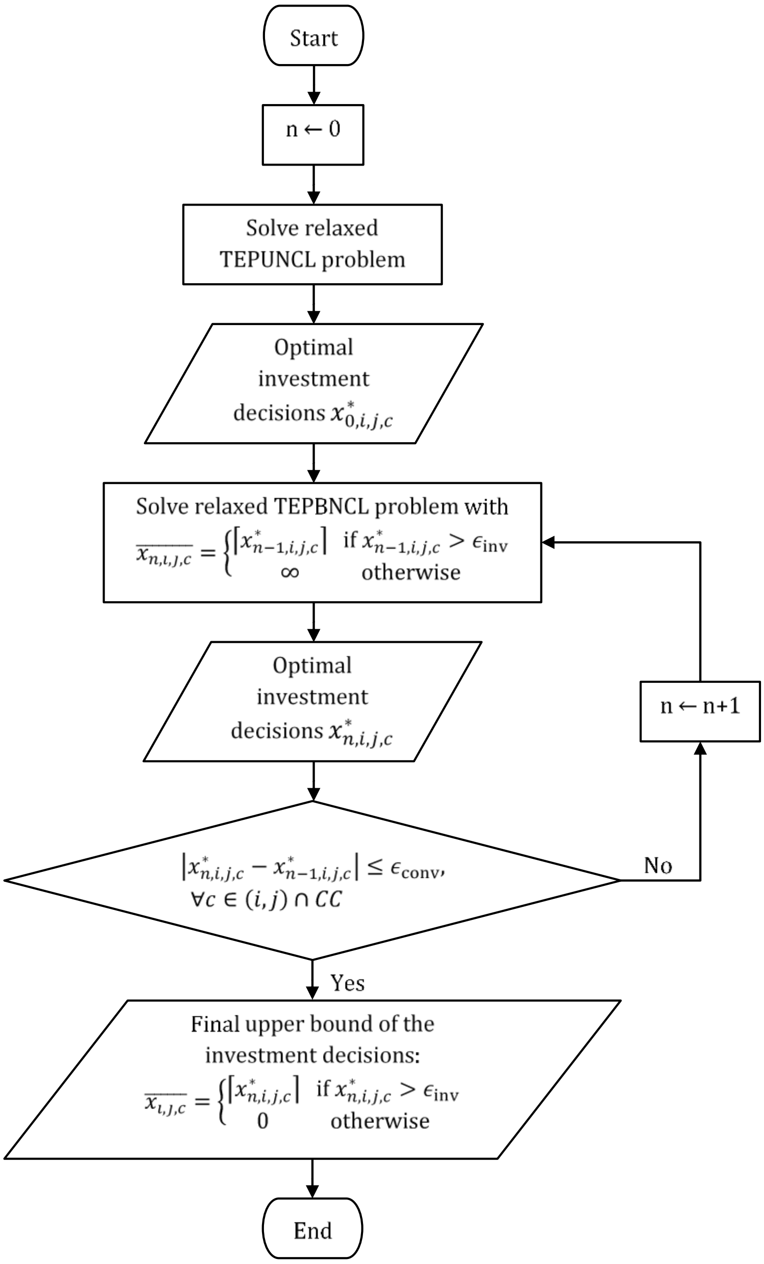

Fig. 6.2 Flowchart of the methodology applied for search space reduction.

- $f_{ijc,s}$ is replaced by $f^{cand}_{i,j,s}$, which represents the total flow through all the candidate lines installed in corridor (i,j),
- x_{ijc} is replaced by $x_{i,j} \in \{0, 1, 2, \dots\}$, which represents the number of candidate lines installed in corridor (i,j),
- Y_{ijc} and $\overline{f_{ijc}}$ are replaced, respectively, by $Y^{cand}_{i,j,1}$ and $\overline{f^{cand}_{i,j,1}}$, the admittance and capacity of a single candidate line in corridor (i,j).

Therefore, the TEP problem with an unbounded number of candidate lines per corridor (TEPUNCL) can be formulated as (3.9)-(3.17), and (6.3).

$$x_{i,j} \in \{0, 1, 2, \dots\} \tag{6.3}$$

As in 3.4.2, we aim to define a linear relaxation of the TEPUNCL problem whose solution provides us with relevant information about the relevant search space of the original TEP problem considering integer investment variables. For this, we need to linearize the integer variables $x_{i,j}$ and find a convex feasible space that includes the space defined by (3.10)-(3.17).

In the linear relaxation of the TEPUNCL problem, the investment decisions variables satisfy equations (6.4):

$$x_{i,j} \in [0; +\infty[\tag{6.4}$$

Constraints (3.9)-(3.11) and (3.13)-(3.17) are linear. Therefore, the space delimited by them is convex and encompasses the feasible space of the TEPUNCL problem. On the other hand, constraint (3.12), enforcing the KVL for candidate AC lines, is not linear because it is defined in terms of the product of an integer variable and a linear variable $x_{i,j}\left(\theta_{i,s} - \theta_{j,s}\right)$. Disjunctive constraints (3.18) are not a valid alternative to equation (3.12) when considering unbounded continuous investment variables. In 3.4.2, we were able to find the smallest convex space delimiting the product of variables in (3.12) by making use of the McCormick envelop of the product of variables. This was previously possible because the variable $x_{i,j}$ was bounded between 0 and 1. Here, the variable $x_{i,j}$ has no upper bound and, therefore, the narrowest convex region for the product of these variables is:

$$-x_{i,j} \frac{M_{i,j}}{S_B Y_{i,j,1}^{cand}} \leq x_{i,j}\left(\theta_{i,s} - \theta_{j,s}\right) \leq x_{i,j} \frac{M_{i,j}}{S_B Y_{i,j,1}^{cand}} \tag{6.5}$$

Where $M_{i,j} = S_B Y_{i,j,1}^{cand} D_{i,j}$, and $D_{i,j}$ is the shortest path distance between buses i and j, as seen in 3.3.

Equation (6.5) can be reformulated as:

$$-x_{i,j} M_{i,j} \leq f_{i,j,s}^{cand} \leq x_{i,j} M_{i,j} \tag{6.6}$$

Equation (6.6) is not limiting the flow on the candidate line in corridor (i,j) more than the capacity constraint. Therefore, it is useless in this case. An illustration of the convex region delimiting the values of the variables $x_{i,j}$ and $f_{i,j,s}^{cand}$ is depicted in Fig. 6.3.

Then, the relaxed TEPUNCL problem must be formulated as (3.9)-(3.11), (3.13), (3.15)-(3.17), (3.27), and (6.4). It can be noticed that the relaxed TEPUNCL is a “hybrid” problem, in which existing AC lines satisfy the DCLF model, and candidate AC lines satisfy the TLF model. However, solving the relaxed TEPUNCL problem provides relevant information about the upper bounds of the decision variables.

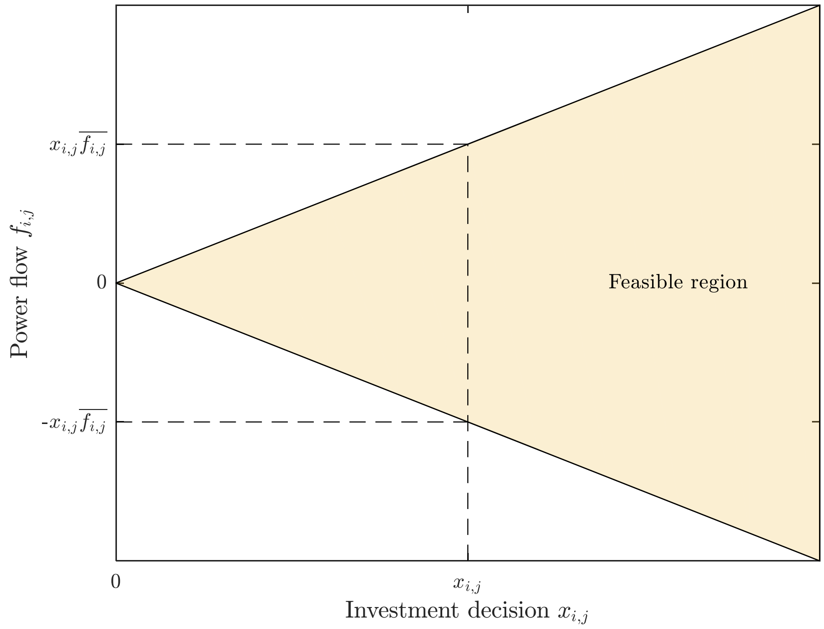

Fig. 6.3 Convex region delimiting the values of $x_{i,j}$ and $f_{i,j,s}^{cand}$ in the relaxed TEPUNCL problem

6.2.2. Relaxed TEP problem with bounded number of candidate lines per corridor

When solving the relaxed TEPUNCL problem, the optimal investment decisions computed correspond to the best possible tradeoff, for the relaxed problem, between the benefits produced by these investments and their cost. The value of these optimal investment decisions can be used to define an upper bound of the number of candidate lines that should be built in each corridor. We shall assume here, as a first approximation, that the number of candidate lines that should be considered for a given corridor, which is an integer, should not be greater than the smallest integer that is greater than the optimal value of the investment variable for this given corridor computed solving the relaxed TEPUNCL problem.

With this in mind, we can define a new relaxed TEP problem with a bounded number of candidate lines per corridor (TEPBNCL) in which investment decision variables are bounded by the following upper bound:

$$\overline{x_{n+1,i,j}} = \begin{cases} \lceil x_{n,i,j}^* \rceil & \text{if } x_{n,i,j}^* > \epsilon_{\text{inv}} \\ \infty & \text{otherwise} \end{cases} \quad (6.7)$$

With $\overline{x_{n+1,i,j}}$ being the upper bound of the investment decision variable $x_{n+1,i,j}$ in the relaxed TEPBNCL problem solved at iteration $n+1$, $x_{n,i,j}^*$ being the optimal value of the investment decision variable for corridor (i,j) computed solving the relaxed TEPUNCL, or TEPBNCL problem, at iteration n, and ϵ_{inv} being the investment decision threshold above which a candidate corridor is considered to be reinforced.

For those corridors whose investment decision variables $x_{n+1,i,j}$ have a finite upper-bound, the following constraints can be introduced:

$$-M_{i,j}\left(\overline{x_{n+1,i,j}}-x_{n+1,i,j}\right)\leq f_{i,j,s}^{cand}-S_B\overline{x_{n+1,i,j}}Y_{i,j,1}^{cand}\left(\theta_{i,s}-\theta_{j,s}\right)$$
$$\leq M_{i,j}\left(\overline{x_{n+1,i,j}}-x_{n+1,i,j}\right) \qquad (6.8)$$

Equation (6.8) is a generalization of the disjunctive constraint (3.18) for any maximum number of candidate lines per corridor $\overline{x_{n+1,i,j}}$. Equation (6.8) together with equation (3.26) represent the convex envelop of the product $x_{i,j}\left(\theta_{i,s}-\theta_{j,s}\right)$ when $x_{i,j}$ is bounded by:

$$0\leq x_{i,j}\leq\overline{x_{n+1,i,j}} \qquad (6.9)$$

Finally, the relaxed TEPBNCL problem can be formulated as (3.9)-(3.11), (3.13), (3.15)-(3.17), (3.27), (6.8), and (6.9). An illustration of the convex region delimiting the values of the variables $x_{i,j}$, $f_{i,j,s}^{cand}$ and $\left(\theta_{i,s}-\theta_{j,s}\right)$ is depicted in Fig. 6.4.

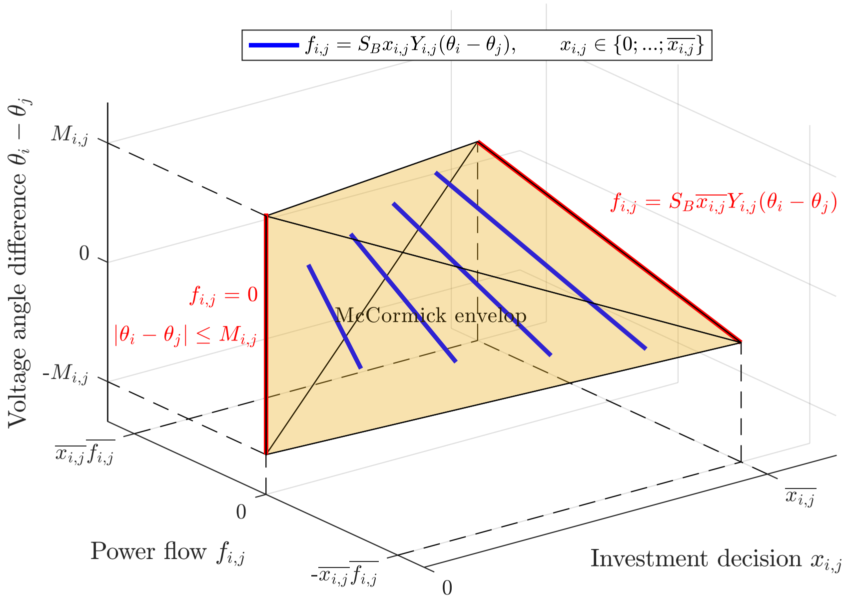

Fig. 6.4 McCormick envelop delimiting the values of $x_{i,j}$, $f_{i,j,s}^{cand}$ and $\left(\theta_{i,s}-\theta_{j,s}\right)$ in the relaxed TEPBNCL problem. The envelop, depicted by the orange tetrahedron, is the smallest convex region covering both the feasible region for which no candidate line is installed ($x_{i,j}=0$) and the feasible region for which the maximum number of candidate lines is installed ($x_{i,j}=\overline{x_{i,j}}$), in red. The feasible regions enforcing the KVL for an integer number of candidate lines greater than 0 and lower than $\overline{x_{i,j}}$, are depicted using blue lines. These blue lines all lie in the McCormick envelop.

It can be noticed that the projection of the convex region depicted in Fig. 6.4 on the two-dimensional subspace $\left(x_{i,j},f_{i,j,s}^{cand}\right)$ corresponds to the sub-region of the one depicted in Fig. 6.3 that is limited by the constraint (6.9). Therefore, the feasible space illustrated in Fig. 6.4 is tighter than the feasible space illustrated in Fig. 6.3.

The relaxed TEPBNCL problem is solved for each of a sequence of iterations. In each iteration, new optimal network investment decisions are computed, and equations (6.7) are employed to define the new upper bounds of the investment decision variables to be considered in the relaxed TEPBNCL problem to be solved in the next iteration. In the last iteration, n_{final}, the following convergence criterion must be met:

$$\left|x^*_{n,i,j} - x^*_{n-1,i,j}\right| \leq \epsilon_{\text{conv}}, \forall\ (i,j) \tag{6.10}$$

This involves that the difference between the optimal number of new lines to be built in each corridor computed in this and the previous iteration must be lower than a certain threshold, i.e. the development of the network computed in these two subsequent iterations must be similar enough. Once convergence has been achieved, the maximum number of candidate lines to be considered in each corridor is computed as in equation (6.11):

$$\overline{x_{i,j}} = \begin{cases} \lceil x^*_{n,i,j} \rceil & \text{if } x^*_{n,i,j} > \epsilon_{\text{inv}} \\ 0 & \text{otherwise} \end{cases} \tag{6.11}$$

In other words, at this stage of the process, the maximum number of candidate lines to be considered for those corridors where optimal continuous investments computed in the last iteration are lower than the investment threshold, ϵ_{inv}, are set to zero.

6.2.3. Binary representation of candidate lines

Making use of the algorithm presented above, the search space of the TEP problem has been reduced from

$$x_{i,j} \in \{0, 1, 2, 3, \dots\} \tag{6.12}$$

to

$$x_{i,j} \in \left\{0, 1, 2, \dots, \overline{x_{i,j}}\right\} \tag{6.13}$$

It should be noted that, contrary to what occurs in the TEP model described in 3.3, the reduced search space computed here assumes that the network investment decision variables are integer variables, not binary ones. However, the integer representation of investment variables considered in (6.13) can be easily converted into a binary representation by making use of the binary numerical system (BNS), described by the authors in (Rahmani, Romero, & Rider, 2013). Using the BNS, integer variables $x_{i,j}$ can be decomposed into a set of binary variables according to:

$$x_{i,j} = \sum_{p=0}^{N_{i,j}-1} 2^p z_{p,i,j} \tag{6.14}$$

Where

$z_{p,i,j} \in \{0; 1\}$ (6.15)

$$N_{i,j} = \begin{cases} \lfloor \log_2(\overline{x_{i,j}}) \rfloor + 1, & \text{if } \overline{x_{i,j}} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (6.16)$$

According to the BNS representation of candidate reinforcements, $N_{i,j}$ candidate lines are considered for each corridor (i, j) with the features described in table 6.1.

TABLE 6.1
CANDIDATE LINE CHARACTERISTICS CONSIDERING THE BNS

	1st candidate circuit	2nd candidate circuit	...	${N_{i,j}}^{th}$ *candidate circuit*
Capacity	$2^0 \overline{f_{i,j,1}^{cand}}$	$2^1 \overline{f_{i,j,1}^{cand}}$	...	$2^{N_{i,j}-1} \overline{f_{i,j,1}^{cand}}$
Admittance	$2^0 Y_{i,j,1}^{cand}$	$2^1 Y_{i,j,1}^{cand}$	...	$2^{N_{i,j}-1} Y_{i,j,1}^{cand}$
Investment cost	$2^0 C_{i,j,1}$	$2^1 C_{i,j,1}$	...	$2^{N_{i,j}-1} C_{i,j,1}$

Technical and economic features of the set of candidate circuits suggested in each corridor (i, j) making use of the BNS.

Considering the BNS, the total number of binary variables to consider in the TEP problem is equal to:

$$\sum_{i,j} N_{i,j} \quad (6.17)$$

6.3. CASE STUDY

6.3.1. SYSTEM DESCRIPTION

The proposed method for the reduction of the search space to be considered in TEP has been applied to the 2088-bus case study based on parts of the European power grid described in ("ENTSO-E Transmission System Map," 2018), ("GridKit extract of ENTSO-E interactive map," 2016). This power system has been previously described in 5.3.1.2.

Following the discussion in 6.1.1, the initial search space is the following one:

- AC circuits can be installed in any inland corridor with a length lower than 200 km,
- Onshore DC circuits can be installed in any inland corridor with a length greater than 200 km,
- Offshore DC circuits can be installed in any submarine corridor whose two ends are on a coast,
- Converters can be installed at any 400 kV bus.

DC circuits can only be connected to the AC network through converters. However, two DC circuits can be directly connected to each other, and full controllability of the flow in each DC corridor is assumed. This is depicted in Fig. 6.5 below.

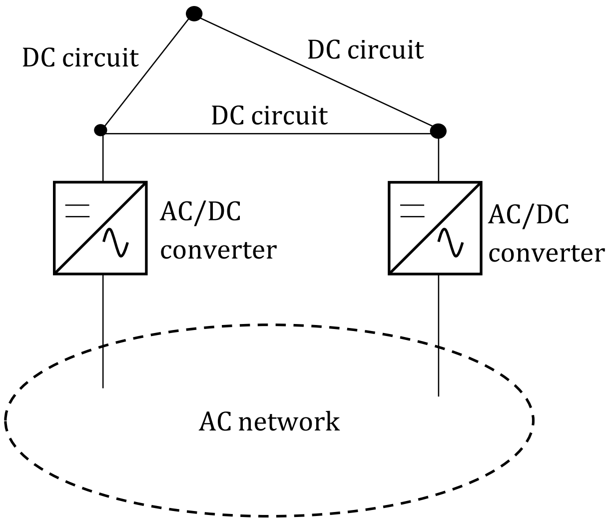

Fig. 6.5 Illustration of a possible AC and DC network structure.

In total, there are 54742 candidate corridors, comprising:

- 11936 corridors for candidate AC circuits,
- 42198 corridors for candidate DC circuits,
- 84 corridors for candidate Offshore DC circuits, and
- 524 buses with candidate converters.

The investment cost per MW and km of AC and DC circuit, as well as the investment cost per MW of converter, is provided in table 6.2 below together with the size of candidate circuits of each type and converters.

TABLE 6.2
CANDIDATE LINE INVESTMENT COSTS

Type of candidate	Investment cost (M€/MW/km)	Capacity per circuit (MW)
Candidate AC lines	1.21	1020
Candidate onshore DC lines	0.44	1000
Candidate submarine DC lines	1.36	1000
Candidate AC/DC converters (M€/MW)	75.75	1000

For the sake of simplicity, from now on, we will refer to the candidate corridors for AC circuits as “candidate AC corridors”, and we will refer to the candidate corridors for onshore and offshore DC circuits and the candidate buses for converters as “candidate DC corridors”.

6.3.2. APPLICATION OF THE METHODOLOGY

In the initial step, the relaxed TEPUNCL problem is solved. The investment decision threshold is set at $\epsilon_{\text{inv}} = 10^{-3}$. According to the solution computed, there are only 175 candidate corridors for which the optimal investment decision variable is greater than this threshold, including 96 candidate AC corridors.

In the following step, when solving the relaxed TEPBNCL problem, the investment decision variables related to these 96 candidate AC corridors are bounded by the nearest integer greater than the optimal value of the decision variable found for those corridors when solving the relaxed TEPUNCL. Thanks to this, we can force the flows in these corridors to meet constraint (6.8), which involves partially enforcing the KVL.

Flows in the 79 candidate DC corridors, on the other hand, do not have to satisfy the KVL. Therefore, their investment decision variable can remain unbounded in the TEPBNCL problem.

The optimal investment decision variable is lower than ϵ_{inv} for the vast majority of the candidate corridors in the relaxed TEPUNCL problem. The investment decision variables for these candidate corridors are kept unbounded when solving the relaxed TEPBNCL problem. In this way, the search space is not prematurely reduced. Candidate corridors that were not worth expanding in the relaxed TEPUNCL problem may be worth expanding when adding constraint (6.8) affecting partially built candidate AC corridors.

The relaxed TEPBNCL problem is solved for as many iterations as needed, updating in each iteration the upper bounds of investment variables, until the maximum change in an optimal investment decision variable over all the candidate corridors is lower than the convergence threshold ϵ_{conv}. The convergence threshold is set at $\epsilon_{\text{conv}} = 10^{-2}$. After 3 iterations involving solving the TEPBNCL problem the convergence threshold is reached. Solving the TEPUNCL problem and the 3 iterations of the TEPBNCL problem has taken around 1 hour overall. A summary of the results obtained is provided in table 6.3.

TABLE 6.3
SUMMARY OF THE REDUCTION PROCESS

	Iteration 1	Iteration 2	Iteration 3	Iteration 4
Problem being solved	Relaxed TEPUNCL	Relaxed TEPBNCL		
Number of candidate AC corridors with $x^*_{n,i,j} > \epsilon_{\text{inv}}$	96	106	109	112
Number of candidate DC corridors with $x^*_{n,i,j} > \epsilon_{\text{inv}}$	79	75	81	84
Total number of candidate corridors with $x^*_{n,i,j} > \epsilon_{\text{inv}}$	175	181	190	196
Total costs (M€)	33454	33462	33468	33470
Computation time (min)	14	13	19	16

Once convergence is reached, the reduced search space to be considered when solving the interger TEP problem is set to comprise all those candidate corridors for which $x^*_{n_{\text{final}},i,j} > \epsilon_{\text{inv}}$ in the final iteration. The maximum number of candidate circuits that can be installed in each of these corridors is set to be equal to $\overline{x_{n_{\text{final}},i,j}}$. In total, there are 247 candidate circuits to be considered for 196 corridors. These results are depicted in Fig. 6.6.

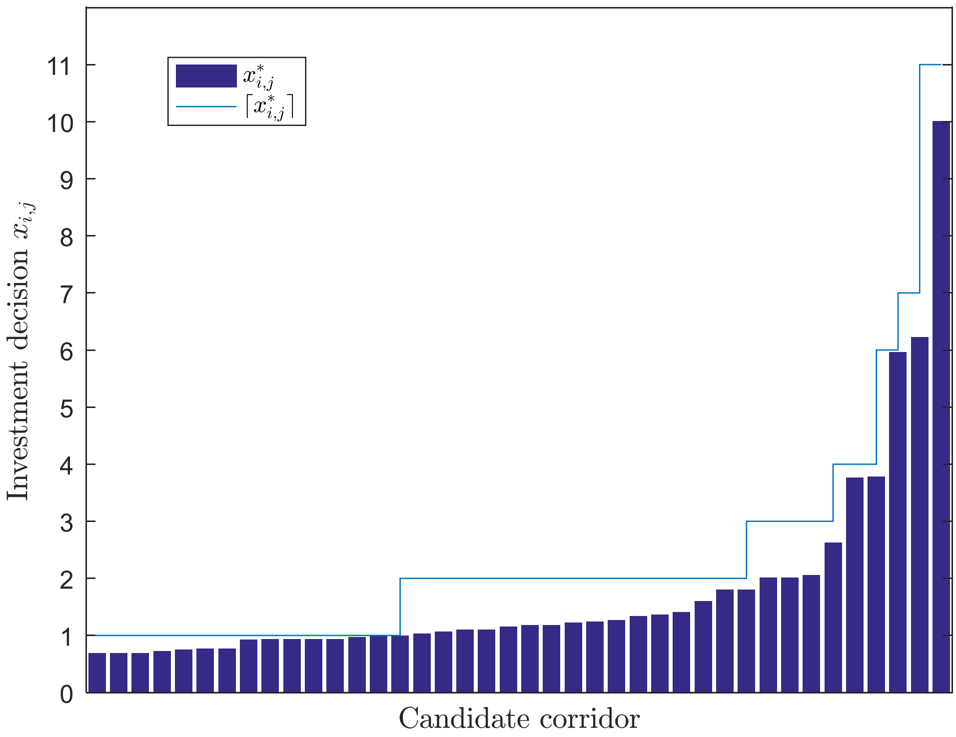

Fig. 6.6 Bar graph: Optimal value of the network investment decision variables $x^*_{i,j}$ in the last iteration of the TEPBNCL problem for those 40 variables whose value is largest (bar graph). Investments variables are displayed in increasing value order. Staircase graph: number of candidate circuits to be considered in the integer TEP problem for the corresponding corridor $\lceil x^*_{i,j} \rceil$.

Finally, the integer investment decision variables to be considered are converted into binary decision variables using the BNS. The number of equivalent candidate lines associated with binary investment decision variables amounts to 227.

6.3.3. Validation of results

In order to assess the efficiency of the search space reduction process here described, the MILP TEP problem is solved taking into account the list of 227 equivalent candidate lines found above. The integrality gap is set at 1%.

It has taken 8 hours to solve the MILP TEP problem taking into account this list of equivalent candidate lines. The total costs corresponding to the optimal solution found amount to 33769 M€. Overall, 99 corridors have been reinforced building 111 new circuits. Moreover, 16 out of the 99 reinforced corridors are associated with investment decision variables in the relaxed TEPUNCL problem whose optimal value was lower than the investment threshold. This proves the relevance of refining, step by step, the set of candidate circuits to consider through the process proposed, which involves updating the convex envelop of the KVL to be considered for the flows in the relevant corridors according to the solution computed in each iteration of the TEPBNCL problem.

Contrary to what happens for the case studies considered when testing the snapshot selection method and the network reduction method, here it is not possible to compare the solution of the MILP TEP problem computed when considering the reduced search space with the solution of this problem computed when considering the non-reduced search space, since solving this problem in the latter case is not possible (the problem becomes intractable). However, we can compare the quality of the solution obtained when applying the proposed search space reduction method with the quality of those solutions of the MILP TEP problem resulting from applying other search space reduction methods in the literature. Thus, the proposed method is compared with the method used by the authors in (Zhang & Conejo, 2018), based on nodal prices, and the method used by the authors in (Villasana et al., 1985), based on the relaxed hybrid TEP model.

6.3.3.1. METHOD BASED ON NODAL PRICES (ZHANG & CONEJO, 2018)

The authors in (Zhang & Conejo, 2018) iteratively enlarge the search space of the TEP problem making use of the information provided by the nodal prices computed in each iteration. In each step (iteration) k of the method they propose, they solve the MILP TEP problem considering a list of candidate lines $\Omega_C^{(k)}$. Then, they compute the benefit-to-cost ratio of every possible new candidate line based on the new nodal prices computed. The candidate lines with the largest benefit-to-cost ratio are added to the list $\Omega_C^{(k)}$ to build the list $\Omega_C^{(k+1)}$. This is carried out time after time until there are no remaining candidate lines with a benefit-to-cost ratio larger than 1. The list of candidate lines considered in the very first step $\Omega_C^{(0)}$ is empty. Then, in this step the authors simply solve an OPF problem. As the authors in (Zhang & Conejo, 2018) point out, the number of candidate lines added in each step should strike a trade-off between number of iterations required and the computational burden of each.

We have applied their method to the 2088-bus case study. We have chosen a number of candidate lines equal to 50, which corresponds to a suitable fraction (not very large, not very small) of the number of suitable candidate lines computed when applying our search space reduction method.

In theory, the benefit-to-cost ratio computed for candidate DC circuits should be based on the nodal price at their ends, as well as for candidate AC circuits. However, since most of the buses in the existing network are not connected yet to an AC/DC converter, the ends of most candidate DC circuits are not connected to the rest of the network. As a result, the nodal prices at the end nodes of most candidate DC circuits are not to be used to calculate their benefit-to-cost ratio. To cope with this issue, we have calculated, instead, the benefit-to-cost ratio of a “DC project” composed of a 1 GW DC circuit together with two 1 GW AC/DC converters at both ends of this circuit. The potential benefit of this DC project is calculated based on the nodal prices at the two buses in the AC network to which the 2 converters would be connected. Then, the benefit-to-cost ratio is computed as the ratio between the aforementioned potential benefit and the sum of the investment cost of the DC circuit and the investment cost of the two converters.

Then, if, at a given step, we identify a DC project as one of the 50 most promising candidate lines, we add the DC circuit and the two converters it comprises to the list of candidates.

We have applied the method in (Zhang & Conejo, 2018) as described. We have computed 9 iterations of the algorithm. The main results computed are summarized in table 6.4. In the first iterations, the total operation-plus-investment cost quickly decreases as the list of candidate lines grows. On the other hand, the computation time required to solve the TEP problem increases with the iteration number, and dramatically increases in iteration 4. From iteration 7 on, it takes more than 24 hours to solve this problem. Even though there are still lines with a benefit-to-cost ratio larger than 1 in iteration 9, we stop the algorithm at this point because of the fact that the decrease in total costs achieved by then in each iteration is insubstantial compared to the computational effort of this iteration.

TABLE 6.4

TEP SOLUTION FOR THE METHOD BASED ON NODAL PRICES

Iteration number	Number of candidate lines considered	Total costs of the MILP TEP solution (M€)	Computation time to solve the TEP problem (min)
1	0	92,681	5
2	50	60,857	6
3	100	45,094	314
4	150	40,738	975
5	200	39,262	238
6	250 + 6cv	37,490	568
7	300 + 6cv	37,034	1421
8	350 + 12cv	35,839	1236
9	400 + 12cv	35,620	2060

Evolution of the number of candidates, the quality of the TEP solution, and the computational effort, with the iteration number when applying the search space reduction method based on nodal prices to the 2088-bus case study. The term “cv” refers to the candidate converters added to the list of candidate lines.

In total, it has taken 79 hours to build a search space composed of 412 candidate lines (time to compute the TEP in the first 8 iterations). The search space computed using the method proposed here is almost 2 times smaller, and it has taken us only 1 hour to compute it. Besides, the best TEP solution computed considering the search space defined based on the method in (Zhang & Conejo, 2018) corresponds to a total cost of 35,620 M€, which is 5.5% larger than the total cost corresponding to the best TEP solution computed considering the search space defined using the method proposed here.

6.3.3.2. METHOD BASED ON THE RELAXED HYBRID TEP MODEL (VILLASANA ET AL., 1985)

In (Villasana et al., 1985), the authors solve a relaxed hybrid TEP problem. The investment decisions in it are linearized and the power flows in the candidate AC lines are computed according to the TLF model. As a matter of fact, the relaxed hybrid TEP problem solved by the authors in (Villasana et al., 1985) is exactly the same as the relaxed

TEPUNCL problem described in 6.2.1, which is solved in the very first iteration of the search space reduction algorithm proposed in this thesis.

Authors in (Villasana et al., 1985) assume, and state, that the optimal values of the investment decision variables computed when solving the relaxed hybrid TEP problem correspond to the required additional capacity in each corridor (Villasana et al., 1985). Contrary to what is carried out in the method proposed in this chapter, they do not use the values of investment decision variables computed when solving the relaxed hybrid problem to reduce the search space considered in the MILP TEP problem. Instead, they directly build an integer TEP solution out of the solution of the relaxed hybrid problem. Given that the optimal values of the investment decision variables in the relaxed hybrid TEP problem are not integer, they are converted into integer values to generate a feasible MILP TEP solution. This can be done in one of the following several ways:

- The optimal values of decision variables $x^*_{i,j}$ are rounded to the next larger integer $\lceil x^*_{i,j} \rceil$,
- The optimal values of decision variables $x^*_{i,j}$ are rounded to the next smaller integer $\lfloor x^*_{i,j} \rfloor$,
- The optimal value of decision variables $x^*_{i,j}$ are rounded to the nearest integer $\lfloor x^*_{i,j} \rceil$.

To assess the efficiency of the method in (Villasana et al., 1985), the optimal values of the investment decision variables in the relaxed TEPUNCL problem considered at the end of the first iteration in our search space reduction method are rounded according to the 3 rounding options above. Then, for each option, the number of candidate circuits installed in each corridor is set to the corresponding rounded number and an OPF is run for the resulting grid. The number of candidate circuits installed, and the total investment plus operation costs obtained for each rounding method are provided in table 6.5.

TABLE 6.5

TEP SOLUTION FOR THE METHOD BASED ON RELAXED HYBRID TEP MODEL

Rounding method	$x_{i,j} = \lceil x^*_{i,j} \rceil$	$x_{i,j} = \lfloor x^*_{i,j} \rfloor$	$x_{i,j} = \lfloor x^*_{i,j} \rceil$
Number of candidate circuits installed	235	60	103
Total costs (M€)	34684	52855	36659

The best MILP TEP solution computed according to any of the rounding options, obtained by rounding the optimal value of the investment decision variables to the next larger integer, results in total costs that are 2.7% higher than the total costs computed for the TEP solution found with the method proposed in this chapter. This MILP TEP solution is a high quality one. Moreover, it has taken us only 14 minutes to find this solution (time required to solve the relaxed TEPUNCL problem), as shown in table 6.3. However, the TEP solution computed using the method proposed in this chapter to reduce the search space still is noticeably better than the one being discussed.

7. Conclusion

7.1. Contributions and conclusion

The recent technological innovations and policies in modern power systems have resulted in a very large increase in the complexity of the TEP problems to be solved. Because of this, there is a critical need to reduce the size of the TEP problem in its three main dimensions:

1. the representation made of the operation situations that are relevant (temporal dimension),
2. the network representation, and
3. the list of candidate transmission projects considered.

Such reduction methods have already been investigated in the literature. However, either they are generic, i.e. they are not adapted to the specific problem to be solved (TEP), or they are not sufficiently tested, i.e. when applied, the quality of the TEP solution computed using them is not compared to the quality of the TEP solution of the non-reduced problem.

We rely on information theory to develop tailor-made efficient methods to reduce the size of the TEP problem regarding each of the aforementioned three dimensions while preserving the most relevant information about the impact of input system parameters on the most important decision variables, the transmission investment decision ones. To achieve this, we make use of a linear relaxation of the TEP problem. Considering the relaxed TEP problem allows one to compute a solution of the TEP problem in a considerably lower amount of time, thanks to the availability of extremely powerful linear solvers. At the same time, the solution of the relaxed TEP problem includes relevant information about the impact of input system parameters on the investment decision variables. We have developed consistent methods that make use of this information to guide the process of reducing the size of the TEP problem.

Besides, the performance of the network reduction method and the snapshot selection method proposed in this thesis project is assessed by comparing the system costs for the TEP solution computed considering the reduced TEP problem obtained by applying these reduction methods with the system costs for the TEP solution computed considering the complete, fully-detailed, TEP problem. Moreover, we systematically compare the performance of each of the three the reduction methods proposed in this thesis with that of the reduction methods proposed in the literature, proving that the method we propose is the most efficient one when applied in a TEP context. A summary of the features of the reduction techniques proposed in this thesis project, together with the reduction factor of the TEP problem achieved when applying them, is presented in table 7.1.

TABLE 7.1
REDUCTION METHODS PROPOSED IN THIS THESIS PROJECT

Dimension reduced	Temporal variability	Network structure	Set of candidate lines
Information captured	Candidate lines' potential benefits	Congested lines and candidate lines likely to be built	Candidate corridors likely to be reinforced, together with their refinforced capacity
Method used	k-means clustering of the potential lines' benefits	Multicut network partition preserving critical lines and kron reduction of non-border buses	TEP problem iteratively solved while refining the convex envelop of KVL constraints for candidate AC corridors
Data reduction achieved	Number of representative snapshots divided by 400*	Number of buses divided by 4**	Number of candidate corridors divided by 300**
Computation time reduction achieved	Computation time divided by 35*	Computation time divided by 5**	Difficult to estimate but potentially very large due to its combinatorial explosion with the number of candidate lines

* For the modified version of the 24-bus system. ** For the 2088-bus system.

The aspect to be considered in the TEP problem which is most difficult to represent in a compact, efficient, way is the representation made of the network. This is because the accuracy of the representation made of the network is critical when solving the TEP problem. After all, the network investment decisions made are intimately linked to the candidate lines considered, which must be represented as new components of the reduced network. Besides, accurately computing the expansion needs critically depends on accurately representing the congestion occurring in the network.

On the other hand, according to the numerical results computed, the set of operation snapshots considered and the search space can be drastically reduced without significantly affecting the quality of the TEP solution computed. However, the maximum reduction factor achieved when selecting the relevant operation snapshots may largely depend on the size of the power system. The test case study considered here only includes 24 buses, and, therefore, a significantly lower number of sources of variability in operation conditions, as well as a significantly lower number of corridors to be potentially reinforced, than larger, more realistic, case studies. Therefore, the diversity of the representative operation situations to consider, and thus, the number of snapshots to take into account, may be significantly higher in larger case studies.

The main contributions of the present thesis are summarized below:

1. The reason why the TEP problem became recently more difficult to solve, and why the three dimensions that compose this problem have increased, have been discussed.

2. A unified framework to achieve a data reduction on each of these 3 dimensions, based on information theory, and linear relaxation techniques, has been proposed.
3. A large-scale realistic European power system case study has been developed to solve the TEP problem.
4. A systematic comparison of the proposed reduction techniques with existing ones from the literature has been performed to prove the efficiency of our methods.

7.2. COMBINATION OF THE PRESENTED REDUCTION METHODS

The present thesis introduces separately three reduction techniques that can be independently applied to each of the three dimensions of the TEP problem. However, these three reduction techniques may, and should, be applied together to maximize the reduction achieved of the size of the TEP problem. Combining these different reduction techniques, however, should be done carefully.

First, the snapshot selection method and the network reduction method presented in this thesis both require a finite, and preferably small, set of candidate lines. Therefore, assuming that a set of candidate lines is not already provided by the planner, the search space reduction technique should be the first reduction technique to be applied.

Second, the network reduction technique presented here is more conservative than the snapshot selection technique. As a matter of fact, the procedure used in the network reduction method not only select areas in which congestion is unlikely to occur, but also guarantees that, if such congestion does not occur, the power flows in the inter-area lines (existing and candidate) are the same in the complete and in the reduced network. Therefore, the network reduction technique should largely preserve the relevant information of the TEP problem. On the contrary, the snapshot selection technique aggregates snapshots that are similar from the TEP problem standpoint, but are still quite different from the operation standpoint. Because of this, the information loss in the snapshot selection procedure should be larger than in the network reduction procedure. Thus, the snapshot selection should be applied last to keep this larger information loss in the last step and, thus, avoid misleading the network reduction step.

Finally, based on the discussion above, the ideal order in which the different reduction techniques presented here should be applied to a TEP problem is the following one:

1. Apply the search space reduction technique presented in chapter 6,
2. Apply the network reduction technique presented in chapter 5, using the reduced set of candidate lines from 1.,
3. Apply the snapshot selection presented in chapter 4, using the reduced network from 2. and the reduced set of candidate lines from 1.

7.3. FURTHER WORKS

The TEP problem considered in this thesis work includes a simple representation of the network whereby the flows in AC lines satisfy the DCLF constraints, whereas the flows in DC lines obey the TLF model, and network losses are discarded. In order to increase the accuracy of the representation made of the system operation, and compute higher quality solutions, network losses should be taken into account in further works. This could potentially affect the features of the reduction methods to be applied, since losses should be accurately enough represented in the reduced TEP problem.

The TEP problem considered in our analyses is a static TEP problem. Then, the effect of transmission reinforcements on the system operation has been computed for a single target year. The proposed reduction methods should be adapted to the case where a dynamic TEP problem, considering several time horizons for the analysis of the operation of the system and the installation of new transmission assets, is considered.

What is more, the existing reduction techniques could be modified, or new ones could be developed, to address the case where the development of generation and transmission is jointly planned in a single problem.

Talking about the representation of the relevant operation situations in the TEP problem, preserving the information about the chronological interdependency between snapshots when reducing the temporal variability would allow us to represent with more accuracy chronological constraints such as unit commitment constraints.

Bibliography

Agapoff, S., & Warland, L. (2014). *Modular Development Plan of the Pan-European Transmission System 2050* (No. E-highway 2050 report). Retrieved from http://www.e-highway2050.eu/fileadmin/documents/Workshop5/140407_e-Highway2050_Draft_Agenda_WS_6_May2014.pdf

Aggarwal, C. C., Wolf, J. L., Yu, P. S., Procopiuc, C., & Park, J. S. (1999). Fast Algorithms for Projected Clustering. In *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data* (pp. 61–72). New York, NY, USA: ACM. https://doi.org/10.1145/304182.304188

Alguacil, N., Motto, A. L., & Conejo, A. J. (2003). Transmission expansion planning: a mixed-integer LP approach. *IEEE Transactions on Power Systems*, *18*(3), 1070–1077. https://doi.org/10.1109/TPWRS.2003.814891

Alonso, J. F., Sáiz, A., Martín, L., Latorre, G., Ramos, A., & Pérez Arriaga, I. J. (1991). PERLA: An Optimization Model for Long Term Expansion Planning of Electric Power Transmission Networks. *Red Eléctrica de España S.A.*

Arthur, D., & Vassilvitskii, S. (2007). K-means++: The Advantages of Careful Seeding. In *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (pp. 1027–1035). Philadelphia, PA, USA: Society for Industrial and Applied Mathematics. Retrieved from http://dl.acm.org/citation.cfm?id=1283383.1283494

Báñez Chicharro, F., Olmos Camacho, L., Ramos Galán, A., Canteli, L., & María, J. (2016). A benefit-based methodology to rank transmission expansion projects. Retrieved from https://repositorio.comillas.edu/xmlui/handle/11531/14236

Battaglini, A., Komendantova, N., Brtnik, P., & Patt, A. (2012). Perception of barriers for expansion of electricity grids in the European Union. *Energy Policy*, *47*, 254–259. https://doi.org/10.1016/j.enpol.2012.04.065

Binato, S., Pereira, M. V. F., & Granville, S. (2001). A new Benders decomposition approach to solve power transmission network design problems. *IEEE Transactions on Power Systems*, *16*(2), 235–240. https://doi.org/10.1109/59.918292

Blum, C., & Roli, A. (2003). Metaheuristics in Combinatorial Optimization: Overview and Conceptual Comparison. *ACM Comput. Surv.*, *35*(3), 268–308. https://doi.org/10.1145/937503.937505

Boîteux, M. (1949). La tarification des demandes en pointe: application de la théorie de la vente au cout marginal. *Revue Generale de l’Electricité*, *58*, 321–340.

Buluc, A., Meyerhenke, H., Safro, I., Sanders, P., & Schulz, C. (2013). Recent Advances in Graph Partitioning. *ArXiv:1311.3144 [Cs, Math]*. Retrieved from http://arxiv.org/abs/1311.3144

Chekuri, C., & Madan, V. (2017). Approximating Multicut and the Demand Graph. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms* (Vols. 1–0, pp. 855–874). Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611974782.54

Cotilla-Sanchez, E., Hines, P. D. H., Barrows, C., Blumsack, S., & Patel, M. (2013). Multi-Attribute Partitioning of Power Networks Based on Electrical Distance. *IEEE Transactions on Power Systems*, *28*(4), 4979–4987. https://doi.org/10.1109/TPWRS.2013.2263886

Cuffe, P., & Keane, A. (2017). Visualizing the Electrical Structure of Power Systems. *IEEE Systems Journal*, *11*(3), 1810–1821. https://doi.org/10.1109/JSYST.2015.2427994

Dakin, R. J. (1965). A tree-search algorithm for mixed integer programming problems. *The Computer Journal*, *8*(3), 250–255. https://doi.org/10.1093/comjnl/8.3.250

Dorfler, F., & Bullo, F. (2013). Kron Reduction of Graphs With Applications to Electrical Networks. *IEEE Transactions on Circuits and Systems I: Regular Papers*, *60*(1), 150–163. https://doi.org/10.1109/TCSI.2012.2215780

Dupačová, J., Gröwe-Kuska, N., & Römisch, W. (2003). Scenario reduction in stochastic programming. *Mathematical Programming*, *95*(3), 493–511. https://doi.org/10.1007/s10107-002-0331-0

Ejebe, G. C., Tong, J., Waight, J. G., Frame, J. G., Wang, X., & Tinney, W. F. (1998). Available transfer capability calculations. *IEEE Transactions on Power Systems*, *13*(4), 1521–1527. https://doi.org/10.1109/59.736300

ENTSOE. (2015). *ENTSO-E Guideline for Cost Benefit Analysis of Grid Development Projects* (Guideline). Retrieved from https://docstore.entsoe.eu/Documents/SDC%20documents/TYNDP/ENTSO-E%20cost%20benefit%20analysis%20approved%20by%20the%20European%20Commission%20on%204%20February%202015.pdf

ENTSO-E Transmission System Map. (2018). Retrieved from https://www.entsoe.eu/data/map/

Fezeu, C. A., Bell, K., Ding, J., Panciatici, P., & Debry, M. S. (2014). Simplified representation of a large transmission network for use in long-term expansion planning. In *2014 Power Systems Computation Conference* (pp. 1–7). https://doi.org/10.1109/PSCC.2014.7038298

Fitiwi, Desta Z., de Cuadra, F., Olmos, L., & Rivier, M. (2015). A new approach of clustering operational states for power network expansion planning problems dealing with RES (renewable energy source) generation operational variability and uncertainty. *Energy*, *90, Part 2*, 1360–1376. https://doi.org/10.1016/j.energy.2015.06.078

Fitiwi, Desta Zahlay. (2016). *Strategies, Methods and Tools for Solving Long-term Transmission Expansion Planning in Large-scale Power Systems* (Doctoral thesis). Comillas Pontifical University, Madrid, Spain. Retrieved from https://repositorio.comillas.edu/xmlui/handle/11531/14509

Gallego, R. A., Alves, A. B., Monticelli, A., & Romero, R. (1997). Parallel simulated annealing applied to long term transmission network expansion planning. *IEEE Transactions on Power Systems*, *12*(1), 181–188. https://doi.org/10.1109/59.574938

Garg, N., Vazirani, V. V., & Yannakakis, M. (1996). Approximate Max-Flow Min-(Multi)Cut Theorems and Their Applications. *SIAM J. Comput.*, *25*(2), 235–251. https://doi.org/10.1137/S0097539793243016

Gomes, P. V., & Saraiva, J. T. (2016). Evaluation of the performance of space reduction technique using AC and DC models in Transmission Expansion problems. In *2016 13th International Conference on the European Energy Market (EEM)* (pp. 1–5). https://doi.org/10.1109/EEM.2016.7521351

Gondzio, J., & Makowski, M. (1995). Solving a class of LP problems with a primal-dual logarithmic. *European Journal of Operational Research*, *80,* 184–192.

GridKit extract of ENTSO-E interactive map. (2016, June 17). [Data repository]. Retrieved from https://zenodo.org/record/55853#.WxFJs386-Uk

Grigg, C., Wong, P., Albrecht, P., Allan, R., Bhavaraju, M., Billinton, R., … Singh, C. (1999). The IEEE Reliability Test System-1996. A report prepared by the Reliability Test System Task Force of the Application of Probability Methods Subcommittee. *IEEE Transactions on Power Systems*, *14*(3), 1010–1020. https://doi.org/10.1109/59.780914

Grijalva, S., Dahman, S. R., Patten, K. J., & Visnesky, A. M. (2007). Large-Scale Integration of Wind Generation Including Network Temporal Security Analysis. *IEEE Transactions on Energy Conversion*, *22*(1), 181–188. https://doi.org/10.1109/TEC.2006.889617

Gu, Y., Ni, M., & Bo, R. (2012). Transmission expansion planning considering economic and reliability criteria. In *2012 IEEE Power and Energy Society General Meeting* (pp. 1–8). https://doi.org/10.1109/PESGM.2012.6345475

Hamon, C., Shayesteh, E., Amelin, M., & Söder, L. (2015). Two partitioning methods for multi-area studies in large power systems. *International Transactions on Electrical Energy Systems*, *25*(4), 648–660. https://doi.org/10.1002/etep.1864

Hooshmand, R.-A., Hemmati, R., & Parastegari, M. (2012). Combination of AC Transmission Expansion Planning and Reactive Power Planning in the restructured power system. *Energy Conversion and Management*, *55*, 26–35. https://doi.org/10.1016/j.enconman.2011.10.020

IEEE 118-bus, 54-unit, 24-hour system. (n.d.). Retrieved from http://motor.ece.iit.edu/data/JEAS_IEEE118.doc

J. Jia. (2014). *Transmission Expansion Planning in Large Power Systems Using Power System Equivalencing Techniques* (Master Thesis). KTH Royal Institute of Technology, Sweden. Retrieved from http://kth.diva-portal.org/smash/get/diva2:740848/FULLTEXT01.pdf

Jang, W., Mohapatra, S., Overbye, T. J., & Zhu, H. (2013). Line limit preserving power system equivalent. In *2013 IEEE Power and Energy Conference at Illinois (PECI)* (pp. 206–212). https://doi.org/10.1109/PECI.2013.6506059

Kenzelmann, S., Rufer, A., Dujic, D., Canales, F., & Novaes, Y. R. de. (2015). Isolated DC/DC Structure Based on Modular Multilevel Converter. *IEEE Transactions on Power Electronics*, *30*(1), 89–98. https://doi.org/10.1109/TPEL.2014.2305976

L. R. Ford, Jr., & D. R. Fulkerson. (1956). Maximal flow through a network. *Canadian Journal of Mathematics*, *8*, 399–404.

L. Wang, G. Gross, S. Meliopoulos, B. Chen, W. Fu, E. Polymeneas, & D. Zhao. (2015). *A Framework for Transmission Planning Under Uncertainty* (Final Project Report). PSERC. Retrieved from https://pserc.wisc.edu/documents/publications/reports/2015_reports/M-30_Final-Report_Sept-2015.pdf

Lee, C. w, Ng, S. K. k, Zhong, J., & Wu, F. F. (2006). Transmission Expansion Planning From Past to Future. In *2006 IEEE PES Power Systems Conference and Exposition* (pp. 257–265). https://doi.org/10.1109/PSCE.2006.296317

Little, J. D. C., Murty, K. G., Sweeney, D. W., & Karel, C. (1963). An Algorithm for the Traveling Salesman Problem. *Operations Research*, *11*(6), 972–989.

Liu, Y., Sioshansi, R., & Conejo, A. J. (2018). Hierarchical Clustering to Find Representative Operating Periods for Capacity-Expansion Modeling. *IEEE Transactions on Power Systems*, *33*(3), 3029–3039. https://doi.org/10.1109/TPWRS.2017.2746379

Lumbreras, S., Ramos, A., Olmos, L., Echavarren, F., Banez-Chicharro, F., Rivier, M., … Pache, C. (2015). Network partition based on critical branches for large-scale transmission expansion planning. In *2015 IEEE Eindhoven PowerTech* (pp. 1–6). https://doi.org/10.1109/PTC.2015.7232344

Lumbreras, S., Ramos, A., & Sánchez, P. (2014). Automatic selection of candidate investments for Transmission Expansion Planning. *International Journal of Electrical Power & Energy Systems*, *59*, 130–140. https://doi.org/10.1016/j.ijepes.2014.02.016

Lumbreras, Sara, & Ramos, A. (2016). The new challenges to transmission expansion planning. Survey of recent practice and literature review. *Electric Power Systems Research*, *134*, 19–29. https://doi.org/10.1016/j.epsr.2015.10.013

Majidi-Qadikolai, M., & Baldick, R. (2015). Reducing the number of candidate lines for high level Transmission Capacity Expansion Planning under uncertainties. In *North American Power Symposium (NAPS), 2015* (pp. 1–6). https://doi.org/10.1109/NAPS.2015.7335117

McCormick, G. P. (1976). Computability of global solutions to factorable nonconvex programs: Part I — Convex underestimating problems. *Mathematical Programming*, *10*(1), 147–175. https://doi.org/10.1007/BF01580665

Mendonça, I. M. de, Silva Junior, I. C., Dias, B. H., & Marcato, A. L. M. (2016). Identification of relevant routes for static expansion planning of electric power transmission systems. *Electric Power Systems Research*, *140*, 769–775. https://doi.org/10.1016/j.epsr.2016.05.011

Morales, J. M., Pineda, S., Conejo, A. J., & Carrion, M. (2009). Scenario Reduction for Futures Market Trading in Electricity Markets. *IEEE Transactions on Power Systems*, *24*(2), 878–888. https://doi.org/10.1109/TPWRS.2009.2016072

Munoz, C., Sauma, E., Contreras, J., Aguado, J., & Torre, S. D. L. (2012). Impact of high wind power penetration on transmission network expansion planning. *Transmission Distribution IET Generation*, *6*(12), 1281–1291. https://doi.org/10.1049/iet-gtd.2011.0552

Munoz, F. D., Hobbs, B. F., & Watson, J.-P. (2016). New bounding and decomposition approaches for MILP investment problems: Multi-area transmission and generation planning under policy constraints. *European Journal of Operational Research*, *248*(3), 888–898. https://doi.org/10.1016/j.ejor.2015.07.057

Nash, J. C. (2000). The (Dantzig) simplex method for linear programming. *Computing in Science Engineering*, *2*(1), 29–31. https://doi.org/10.1109/5992.814654

Oatts, M. L., Erwin, S. R., & Hart, J. L. (1990). Application of the REI equivalent for operations planning analysis of interchange schedules. *IEEE Transactions on Power Systems*, *5*(2), 547–555. https://doi.org/10.1109/59.54565

Oh, H. (2012). Aggregation of Buses for a Network Reduction. *IEEE Transactions on Power Systems*, *27*(2), 705–712. https://doi.org/10.1109/TPWRS.2011.2176758

Özdemir, Ö., Munoz, F. D., Ho, J. L., & Hobbs, B. F. (2016). Economic Analysis of Transmission Expansion Planning With Price-Responsive Demand and Quadratic Losses by Successive LP. *IEEE Transactions on Power Systems*, *31*(2), 1096–1107. https://doi.org/10.1109/TPWRS.2015.2427799

Padberg, M., & Rinaldi, G. (1991). A Branch-and-Cut Algorithm for the Resolution of Large-Scale Symmetric Traveling Salesman Problems. *SIAM Review*, *33*(1), 60–100. https://doi.org/10.1137/1033004

Palchak, D., Cochran, J., Deshmukh, R., Ehlen, A., Soonee, S. K., Narasimhan, S. R., … Abhyankar, N. (2017). *Greening the Grid: Pathways to Integrate 175 Gigawatts of Renewable Energy into India's Electric Grid, Vol. I. National Study. Executive Summary* (No. LBNL-2001011). Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States); National Renewable Energy Lab. (NREL), Golden, CO (United States); Power System Operation Corporation (POSOCO), New Delhi (India); US Agency for International Development (USAID), Washington, DC (United States). https://doi.org/10.2172/1393626

Papadimitriou, C. H., & Steiglitz, K. (1998). *Combinatorial optimization : algorithms and complexity*. Mineola, N.Y. : Dover Publications. Retrieved from https://trove.nla.gov.au/version/46593466

Park, H., Baldick, R., & Morton, D. P. (2015). A Stochastic Transmission Planning Model With Dependent Load and Wind Forecasts. *IEEE Transactions on Power Systems*, *30*(6), 3003–3011. https://doi.org/10.1109/TPWRS.2014.2385861

Power Systems Test Case Archive. (n.d.). Retrieved from https://www2.ee.washington.edu/research/pstca/

Rahmani, M., Romero, R., & Rider, M. J. (2013). Strategies to Reduce the Number of Variables and the Combinatorial Search Space of the Multistage Transmission Expansion Planning Problem. *IEEE Transactions on Power Systems*, *28*(3), 2164–2173. https://doi.org/10.1109/TPWRS.2012.2223241

Rathore, C., Roy, R., Sharma, U., & Patel, J. (2013). Artificial Bee Colony Algorithm based static transmission expansion planning. In *2013 International Conference* *on Energy Efficient Technologies for Sustainability* (pp. 1126–1131). https://doi.org/10.1109/ICEETS.2013.6533544

Rider, M. J., Gallego, L. A., Romero, R., & Garcia, A. V. (2007). Heuristic Algorithm to Solve the Short Term Transmission Network Expansion Planning. In *2007 IEEE Power Engineering Society General Meeting* (pp. 1–7). https://doi.org/10.1109/PES.2007.385441

Rider, M. J., Garcia, A. V., & Romero, R. (2007). Power system transmission network expansion planning using AC model. *Transmission Distribution IET Generation*, *1*(5), 731–742. https://doi.org/10.1049/iet-gtd:20060465

Rommes, J., & Schilders, W. H. A. (2010). Efficient Methods for Large Resistor Networks. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, *29*(1), 28–39. https://doi.org/10.1109/TCAD.2009.2034402

Sanchez, I. G., Romero, R., Mantovani, J. R. S., & Rider, M. J. (2005). Transmission-expansion planning using the DC model and nonlinear-programming technique. *Transmission and Distribution IEE Proceedings - Generation*, *152*(6), 763–769. https://doi.org/10.1049/ip-gtd:20050074

Sayood, K. (2005). *Introduction to Data Compression*. Elsevier.

Sharifnia, A., & Aashtiani, H. Z. (1985). Transmission Network Planning: A Method for Synthesis of Minimum-Cost Secure Networks. *IEEE Transactions on Power Apparatus and Systems*, *PAS-104*(8), 2025–2034. https://doi.org/10.1109/TPAS.1985.318777

Shayesteh, E., Hobbs, B. F., Söder, L., & Amelin, M. (2015). ATC-Based System Reduction for Planning Power Systems With Correlated Wind and Loads. *IEEE Transactions on Power Systems*, *30*(1), 429–438. https://doi.org/10.1109/TPWRS.2014.2326615

Singh, H. K., & Srivastava, S. C. (2005). A reduced network representation suitable for fast nodal price calculations in electricity markets. In *IEEE Power Engineering Society General Meeting, 2005* (pp. 2070-2077 Vol. 2). https://doi.org/10.1109/PES.2005.1489092

Svendsen, H. G. (2015). Grid Model Reduction for Large Scale Renewable Energy Integration Analyses. *Energy Procedia*, *80*, 349–356. https://doi.org/10.1016/j.egypro.2015.11.439

Tarjan, R. (1971). Depth-first search and linear graph algorithms. In *12th Annual Symposium on Switching and Automata Theory (swat 1971)* (pp. 114–121). https://doi.org/10.1109/SWAT.1971.10

Tinney, W. F., & Walker, J. W. (1967). Direct solutions of sparse network equations by optimally ordered triangular factorization. *Proceedings of the IEEE*, *55*(11), 1801–1809. https://doi.org/10.1109/PROC.1967.6011

Tishby, N., Pereira, F. C., & Bialek, W. (2000). The information bottleneck method. *ArXiv:Physics/0004057*. Retrieved from http://arxiv.org/abs/physics/0004057

Torres, S. P., & Castro, C. A. (2012). Parallel particle swarm optimization applied to the static Transmission Expansion Planning problem. In *2012 Sixth IEEE/PES Transmission and Distribution: Latin America Conference and Exposition (T D-LA)* (pp. 1–6). https://doi.org/10.1109/TDC-LA.2012.6319053

Villasana, R., Garver, L. L., & Salon, S. J. (1985). Transmission Network Planning Using Linear Programming. *IEEE Transactions on Power Apparatus and Systems*, *PAS-104*(2), 349–356. https://doi.org/10.1109/TPAS.1985.319049

Vinasco, G., Rider, M. J., & Romero, R. (2011). A Strategy to Solve the Multistage Transmission Expansion Planning Problem. *IEEE Transactions on Power Systems*, *26*(4), 2574–2576. https://doi.org/10.1109/TPWRS.2011.2126291

Ward, J. B. (1949). Equivalent Circuits for Power-Flow Studies. *Transactions of the American Institute of Electrical Engineers*, *68*(1), 373–382. https://doi.org/10.1109/T-AIEE.1949.5059947

Wogrin, S., Duenas, P., Delgadillo, A., & Reneses, J. (2014). A New Approach to Model Load Levels in Electric Power Systems With High Renewable Penetration. *IEEE Transactions on Power Systems*, *29*(5), 2210–2218. https://doi.org/10.1109/TPWRS.2014.2300697

Wu, Z., Du, X., Gu, W., Zhang, X. P., & Li, J. (2018). Automatic Selection Method for Candidate Lines in Transmission Expansion Planning. *IEEE Access*, *6*, 11605–11613. https://doi.org/10.1109/ACCESS.2018.2798063

Wu, Zhi. (2016). *Transmission expansion planning and unit commitment with large-scale integration of wind power* (d_ph). University of Birmingham. Retrieved from http://etheses.bham.ac.uk/6738/

Zhang, X., & Conejo, A. J. (2018). Candidate line selection for transmission expansion planning considering long- and short-term uncertainty. *International Journal of Electrical Power & Energy Systems*, *100*, 320–330. https://doi.org/10.1016/j.ijepes.2018.02.024

CURRICULUM VITAE

Quentin Ploussard was born in 1989 in Aix-en-Provence, France. He received the Electrical Engineering degree from École Supérieure d'Électricité (Supélec), Paris, France, and the M.S. degree in Economics from Université Paris Sud XI, Paris, France, both in 2014.

From 2010 to 2014, during his Master studies, he achieved multiple student internships at SNCF (French National Railway Company), ENGIE (GDF Suez), Smart Impulse, and Enedis (French Distribution System Operator).

In 2014, he was selected as a PhD candidate of the Erasmus Mundus Joint Doctorate in Sustainable Energy Technologies and Strategies (SETS) and awarded an Erasmus Mundus Fellowship. This joint PhD program is offered by Comillas Pontifical University, KTH Royal Institute of Technology and Delft University of Technology. He was a visiting PhD student in KTH from September 2016 to May 2017 as part of the mandatory mobility plan of the SETS program.

During his PhD studies, he actively participated in the international SET-Nav project, funded by the European Commission, whose goal is to support strategic decision making in Europe's energy sector.

He is currently working as a Researcher at Argonne National Laboratory, Chicago, United States of America. His research interests include the analysis of energy and power systems, optimization of hydropower plant and reservoir operations, electricity market design and operation, renewable energy integration.

LIST OF RELEVANT PUBLICATIONS

Peer-reviewed Journal Articles Related to the Thesis (Published)

1. Q. Ploussard, L. Olmos, A. Ramos, "An Operational State Aggregation Technique for Transmission Expansion Planning Based on Line Benefits," in IEEE Transactions on Power Systems, vol. 32, no. 4, pp. 2744-2755, July 2017
2. Q. Ploussard, L. Olmos, A. Ramos, "An Efficient Network Reduction Method for Transmission Expansion Planning using Multicut Problem and Kron Reduction," in IEEE Transactions on Power Systems, vol. 33, no. 6, pp. 6120-6130, November 2018

Peer-reviewed Journal Articles Related to the Thesis (Under Review)

3. Q. Ploussard, L. Olmos, A. Ramos, "A Search Space Reduction Method for Transmission Expansion Planning using an Iterative Refinement of the Kirchhoff Voltage Law," in IEEE Transactions on Power Systems, submitted in September 2018

Peer-reviewed Conference Papers Related to the Thesis

4. Q. Ploussard, L. Olmos, A. Ramos, "An uncertainty reduction technique for short-term transmission expansion planning based on line benefits," SIAM Conference on Uncertainty Quantification - UQ16. Lausanne, Switzerland, 5-8 April 2016
5. Q. Ploussard, L. Olmos, A. Ramos, L. Söder, "A novel power grid reduction technique for a transmission expansion planning problem using a weighted signed graph model," 25th Anniversary Conference of the ACO Program - ACO25. Atlanta, United States of America, 9-11 January 2017

Technical Reports Related to the SET-Nav project

6. P. Crespo del Granado, H. Marañon Ledesma, C. Skar, B. Gjorgiev, G. Sansavini, A. Antenucci, L. Olmos, Q. Ploussard, S. Lumbreras, A. Ramos, "Issue Paper on Unlocking Unused Flexibility and Synergy in Electric Power and Gas Supply Systems," Navigating The Roadmap For Clean, Secure And Efficient Energy Innovation, February 2018
7. S. Lumbreras, A. Ramos, L. Olmos, Q. Ploussard, F. Sensfuss, G. Deac, C. Bernath, "Case Study Paper on Centralized vs Decentralized Development of the Electricity Sector," Navigating The Roadmap For Clean, Secure And Efficient Energy Innovation, March 2018
8. F. Holz, T. Scherwath, I. Valentin Kafemann, C. Skar, H. Marañon Ledesma, P. Crespo del Granado, A. Ramos, L. Olmos, Q. Ploussard, S. Lumbreras, A. Herbst, T. Fleiter , "Case Study Report on The Role for Carbon Capture, Transport and Storage in Electricity and Industry in the Future," Navigating The Roadmap For Clean, Secure And Efficient Energy Innovation, April 2018

Back cover

Information required by KTH:

For KTH Royal Institute of Technology:
DOCTORAL THESIS IN ELECTRICAL ENGINEERING
TRITA (will be provided by KTH)
www.kth.se

ISBN (will be provided by your home university)

Spine of the thesis book: Author's name and title