e-HIGHWAY 2050

Modular Development Plan of the Pan-European Transmission System 2050

Contract number	308908	Instrument	Collaborative Project
Start date	1st of September 2012	Duration	40 months
WP 8	Enhanced Pan-European Transmission Planning Methodology		
D 8.4.a	Enhanced methodology to define the optimal modular plan		

Revision: [1.3 FINAL]

Due date of delivery: [December 2014]

		Date & Visa
Written by	Sergeï AGAPOFF, RTE Leif WARLAND, SINTEF	06/01/2015
Checked by	Patrick PANCIATICI, WP8 Leader Camille PACHE, RTE Jean MAEGHT, RTE Alexander WEBER, TU Berlin	19/01/2015
Validated by	Brahim BETRAOUI, RTE Gérald SANCHIS, Coordinator	12/02/2015

Project co-funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Document information

General purpose

Deliverable 8.4.a (D8.4.a) presents the methods to achieve the goal of task 8.4: optimize the Transmission Expansion Planning on a large geographical scale for several time-horizons and several scenarios.

Three processes are proposed for this purpose: Snapshot Selection to find representative operational situations of the grid, Candidate Selection to reduce the pool of expansion possibilities to a meaningful set and Transmission Expansion Planning optimization to find the optimal expansion for each of those candidates on the considered snapshots.

Those three processes are presented in this document along with a small test case to assess their performances in different situations. The results of the test case are provided as well as recommendations and conclusions of the work done in task 8.4.

The implementation of the different processes is presented in deliverable D8.4.b "Prototype for optimal modular plan to reach 2050 grid architectures".

Change status

Revision	Date	Changes description	Authors
V0.0	02/09/2014	Initial version	S. AGAPOFF L. WARLAND
V1.0	11/12/2014	First draft	S. AGAPOFF
V1.1	19/12/2014	Minor typography modifications Simplification and re-organization of the test objectives and test cases	S. AGAPOFF
V1.2	02/01/2015	Minor typography modifications Expansion Indicators modifications Update of the results and conclusions	S. AGAPOFF
V1.3	06/01/2015	Conclusions modified	S. AGAPOFF
V1.4	02/02/2015	Modifications after review from coordinators	C. PACHE
V1.5	11/02/2015	Addition of new test case results	C. PACHE

EXECUTIVE SUMMARY

Transmission Expansion Planning (TEP) is traditionally processed over a short period (usually 10 years) and results in a single-step evolution of the transmission grid. This short-term process is efficient to define precisely the needs of the power system for the next years and to decide reinforcements, but it cannot prepare the grid for long-term energy developments. Such developments (technology breakthroughs, radically new energy policies or major societal changes) induce a long adaptation of the grid structures and they should be taken into account when planning a multi-step evolution of the grid: several milestones (time-horizons) should be considered to obtain a modular development of the grid. However, planning for such major changes can only be handled through scenarios which describe the possible evolutions of energy requirements, technologies and policies. With multiple scenarios a mitigation method is necessary to ensure a common development in the first time-horizons, able to provide propositions to decision-makers.

Task 8.4 aims at defining the methodology needed to obtain such a modular development plan for a large system (European power system) and on a long-term perspective (from 2020 to 2050). The basic idea of a Transmission Expansion Planning is to optimize the investments in given candidates so that the global operation of the system is more efficient (reduce the energy not served, the congestions etc.) than without the investments. This statement induces two questions: how are defined the candidates and how is assessed the impact of a given investment on system operation? The expansion candidates are usually picked by experts. Uncertainties in the system can be taken into account through Monte Carlo simulations, while the operational cost of the reinforced transmission grid mirrors the impact of investments on system operation. We need specific methods to ensure that the choice of the candidates is driven by the current and future needs of the grid.

Transmission Expansion Planning optimization is a well-known problem, and we based the optimization problem on existing models [13]-[14]. The ideal approach would be optimizing grid investments while taking into account operational costs over all the snapshots. In a long-term perspective where several scenarios are considered, this would lead to millions of snapshots: the size of the TEP optimization problem would be too large. Thus, we want to drastically reduce the number of snapshots without degrading the grid expansion results. The main focus of our work is to find a Snapshot Selection method meeting these requirements.

Three processes are proposed to achieve the goal of task 8.4:

- Snapshot Selection to find representative operational situations of the system
- Candidate Selection to reduce the pool of expansion possibilities to a meaningful set
- Transmission Expansion Planning optimization to find the optimal expansion for each of those candidates on the considered snapshots.

The Snapshot Selection is based on grouping similar snapshots together and choosing one in each group to be the representative snapshot. Different similarity features were proposed based on price-differences between zones and non-controllable generation and demand in each zone. Those features were studied as local values (one value for each zone or pair of zones) and as statistical values (minimum, maximum, average and standard deviation over the grid).

The Candidate Selection uses a simplified TEP optimization to successively identify relevant candidates which seem to be profitable. We based our approach on the work of Lumbreras, Ramos and Sanchez [11] from Comillas.

The TEP optimization answers the questions of when and how much a candidate should be invested while taking into account the impacts of such investments on system operation. To obtain a complete modular development plan, this optimization takes into account all scenarios and all time-horizons.

The main contributions of our work are:

- Development of a recursive algorithm around a clustering method to find representative snapshots
- Choice of the criteria for the Snapshot Selection
- Common definitions and file formats to ease the integration of the different processes in a single methodology

At the time of writing, only the Snapshot Selection has been tested on a small system (further studies are suggested in the deliverable conclusion). Based on the obtained results, we suggest that the statistical price-differences can be used as a clustering feature for the Snapshot Selection. It uses 4 values (instead of the number of pairs of zones) to describe the snapshots and hence should reduce drastically the complexity of the clustering.

On a 6-zones system the computation time results were not very conclusive: tests on a larger system should show the benefit of using statistical values instead of local values.

TABLE OF CONTENT

List of figures

List of tables

Acronyms and definitions

AC/DC	Alternative/Direct Current
ARI	Adjusted Rand Index: *Similarity index between two partitions of the same dataset*
CSI	Candidate Selection Index: *Similarity index between two solutions of the Candidate Selection*
ESI	Expansion Solution Index: *Similarity index between two TEP solutions (investment decision for each candidate)*
ERD	Expansion Relative Difference: *Error index between two TEP solutions (optimal expansion of each candidate)*
MILP	Mixed Integer Linear Programming
NC D/G	Non-controllable Demand (typical load) or Generation (production of intermittent power sources etc.)
OM	Operation & Maintenance cost (independent from operations of the grid)
PST	Phase-Shifter Transformer
p.u.	Per Unit
TAD	Time-horizon Absolute Difference: *Error index between two TEP solutions (time-horizon of investment for each candidate)*
TEP	Transmission Expansion Planning

<u>Zone</u>

A zone is an *electrical area* of the system, in which several types of production and consumption are represented and aggregated. A zone can be assimilated to a cluster from WP2 (see D2.2 “European cluster model of the pan-European transmission grid”). The intra-zonal constraints are not taken directly into account: they are accounted for in the calculation of the transfer capacities and reactance values: this work is performed in task 8.3.

<u>Time-horizon</u>

This methodology considers a defined study period that goes from 2020 to 2050. The system is considered to be in its initial state in 2020 and the methodology will propose developments up to 2050. After 2020, a time-horizon is a given year of the study period and it is used as a *milestone* for the optimization process. In our document we consider 6 time-horizons for which grid developments have to be optimized: 2025, 2030, 2035, 2040, 2045 and 2050. Each time-horizon represents a given number of years (depending on the number of years between two time-horizons). We consider that each of those years has the same behaviour as the "representative" year: the maintenance and the operational costs are the same and the investments are done at the start of the time-horizon and are available for all years within the time-horizon.

The behaviour of the time-horizon is described through time-series of production and consumption obtained through several Monte-Carlo hourly simulations. For each time-horizon, several simulated years of 8760 time-steps are available to represent the operational situations of this given time-horizon.

Scenario

A scenario describes the evolution of the energy system in each time-horizon and each zone. Several independent scenarios are considered in this methodology. Their probability of occurrence is measured by a weight.

Snapshot

Each time-step of a given time-horizon and scenario is called a snapshot. In this methodology we propose a method to select representative snapshots for the TEP optimization: those selected snapshots represent all the snapshots related to the considered time-horizon and scenario. The weight of each representative snapshot for the considered time-horizon and scenario is defined depending on the size of its cluster (pool of similar snapshots).

Injection

An injection is the power (MW) of a given production/consumption entity in a given zone. We consider different injection types defined by their sign factor (+1 for production, -1 for consumption). The main difference between injection types is their controllability (wind generation is much less controllable than coal generation for example) and this is mirrored in the min/max values that those injections can take and their up/down Marginal Costs of variation. Injections in the system are the main characteristics to define the system's behaviour in a given snapshot.

Corridor

Corridor technology: AC or DC.

Corridor type: Corridor technology, Maximal capacity, Reactance per unit of length, Costs (Investment and Maintenance) and Maximal length. The Maximal capacity, the Reactance per unit of length and the Maximal length are only used for candidate definition (the corridors in the existing grid have their own capacity and reactance values).

A corridor is defined as a combination of pair of zones and a corridor type: there can be several corridors of the same type between the two same zones.

Only the existing corridors are defined: end zones, type, actual capacity and reactance.

Phase-shifter transformers

A phase-shifter transformer (PST) is a flexibility asset used on AC corridors to optimize the angle phase difference on the considered corridor. We consider that no new PST can be installed during the TEP process: only the existing PSTs are taken into account. All the PSTs are given in a list with their characteristics and each corridor using a PST is linked to a member of that list.

Candidate

A candidate is defined as a technical upgrade (capacity increase and, if applicable, reactance increase) of a given "c*orridor type*" between two zones. If the difficulty of building a candidate has to be taken into account (for example: sub-sea cables) a "land-use" factor, specific to each candidate, is proposed to multiply the investment and/or maintenance costs.

Our optimization model allows switching an existing corridor to another corridor type. For this purpose we use "Combined Candidates".

Unit

A unit is defined as an increment of candidate. Therefore "building n times the candidate k in the time-horizon h" is equivalent to "adding n units of candidate k in the time-horizon h". The decision variable $n_k(s,h)$ is derived from this definition and represents the optimal number of units for the candidate k to be installed in the scenario s and the time-horizon h. The maximal number of units available for a given candidate in each time-horizon is a parameter given when defining the candidate k.

Nomenclature

This section presents every symbol used in this document. For symbols taking a specific value, we indicate the unit in which it should be provided: $(\mathrm{p.u.})$ indicates "per unit" values and $(\emptyset)$ indicates unit-less values (e.g. a ratio). We indicate the few symbols used as optimization variables by adding [var] in their description.

α_c^r	[var] Phase Shifter angle on corridor c
$ArchFactor_k$	Architecture-related factor for a candidate $(\emptyset)$. 3 architectures are proposed
$\beta^{s,h}(k,b)$	[var] Binary variable related to the decomposition of the expansion variable $n_k(s,h)$ $(b \in [\![1, n_k^{Max}]\!])$
$\mathcal{C}$	Candidate Pool ($\mathcal{C}_{AC/DC}$ are AC and DC subsets)
$Cap_{AC/DC}(z_A, \mathrm{z_B})$	Initial AC or DC capacity between zones z_A and $\mathrm{z_B}$
Cap_k^{inc}	Capacity for one unit of candidate k (MW)
$\mathrm{Card}(A)$	Size of a given set A (i.e. number of elements in the set)
CDH	Common Development Horizons
$CombCand$	Combined Candidates pool
$Corridors$	Set of initial corridors
$Corr_{PST}$	Subset of initial AC corridors with phase-shifter transformer
$CorrTypes$	Set of Corridor Types
CTr	Computation time ratio between two runs of a given process $(\emptyset)$
$DF(T)$	Discount Factor for a period of T years $(\emptyset)$
$\Delta I_{+/-}$	[var] Variations of the injections
F_i^m	Clustering feature set
$f_{AC/DC}(z_A, z_B)$	[var] AC or DC Power flow between two zones
$FlowVar_{threshold}$	Flow Variation threshold for the Candidate Analysis $(\emptyset)$
$\mathcal{H}$	Set of time-horizons $(h \in \mathcal{H})$, H is the final time-horizon
h_{step}	Number of years between each time-horizon
I_0^r	Initial Injections for each zone and each injection type, based on the snapshot r
I	[var] Injection array describing Power Injections for each type and each zone (MW)
I_{min}/I_{max}	Injection boundaries for each zone and each injection type (MW)
$InjTypes$	Injection types
Inv_{type}	Total investment cost for a given $type$ of corridor $(€/\mathrm{km})$
L_{ref}	Reference length for the architecture constraints (km)
Len_k	Length of a candidate k (km)

LT_{type}	Life-Time of a given $type$ of corridor (years)
$MaxLength_{type}$	Maximal Length of a given $type$ of corridor (km)
$MC_{+/-}$	Marginal Costs of variation (Up/Down) for each zone and each injection type (€/MW)
$n_k(s,h)$	[var] Number of expansion units for candidate k in scenario s and time-horizon h
n_k^{Max}	Maximal number of units for candidate k
OM_{type}	Annual Operation & Maintenance Cost for a given $type$ of corridor (€/(km · year))
$P_k^{s,h}$	Profitability of a given candidate k for the scenario s and the time-horizon h
$\mathcal{R}(s,h)$	Set of representative snapshots for the scenario s and the time-horizon h
$\mathcal{S}$	Set of scenarios ($s \in \mathcal{S}$), n_s is the number of scenarios
$\sigma_{+/-}$	Penalization for the architecture constraints (∅)
$sg(i)$	Sign factor of an injection type $i \in InjTypes$ (production: $+1$; consumption: -1) (∅)
S_n	Nominal Power used for the per unit system (MW)
$StableAssign_{threshold}$	Cluster Stability limit for the Snapshot Selection
$SwBus$	Set of Swing Buses needed for the DCOPF
τ	Discount rate (∅)
θ	[var] Phase angle in each zone
$\tilde{\theta}_z^{r,h}(k,b)$	[var] Phase angle variable replacing $\theta_z^r \cdot \beta^{s,h}(k,b)$
U_n	Nominal Voltage used for the per unit system (kV)
w_r	Probabilistic weight of the snapshots (∅)
w_s	Probabilistic weight of the scenarios (∅)
y_0	Initial year
$Y(z_A, z_B)$	Initial admittance between zones z_A and $\mathrm{z_B}$ (p.u.)
Y_c	Initial admittance for a specific corridor c (p.u.)
Y_k^{inc}	Admittance for one unit of candidate k (p.u.)
Z	Set of all the zones
Z_n	Nominal Reactance used for the per unit system (Ω)
$ZoneA/ZoneB$	End zones of a specific corridor or candidate
$(x)^+$	Positive value of x (0 if x negative, x elsewise)

Conventions and Assumptions

(a) We consider that each time-horizon h represents a period of h_{step} years and that all expenses paid in time-horizon h have to be discounted from this "representative" year to the first year of the study period (y_0). We also consider that investments made during this time-horizon are directly available, which means that the capacity installed in time-horizon h is accounted for in the maintenance cost and in flow constraints of time-horizon h and future ones. Thus, investments are made in year h.

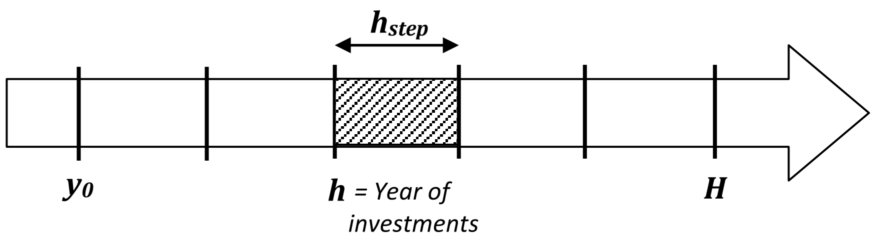

Figure 1 – Decomposition into time-horizons

(b) Every Injection is a positive value: a Consumption or Production injection is indicated by the sign factor of its type ($sg(i) \in \{-1, +1\}$).

(c) For a given corridor, the flow capacity is the same for each direction of the flow.

(d) The per-unit system (p.u.) is defined as follows for the flows (F in MW), voltages (U in kV) and impedances (Z in Ω):

$$F = S_n \cdot f$$

$$U = U_n \cdot u$$

$$Z = Z_n \cdot z$$

The subscripted symbols (S_n, U_n, Z_n) represent nominal values and the small case symbols (f, u, z) are the p.u. values.

Usually the nominal power rating S_n is defined for the entire grid and we use the value of 100MVA. U_n and Z_n on the other hand depend on the considered corridor: for each type of corridor the nominal voltage U_n is considered given (technical data) and the nominal impedance is obtained as follows:

$$Z_n = \frac{{U_n}^2}{S_n}$$

(e) Resistances are neglected.

(f) Losses are not modelled in the TEP optimization. Taking them into account would add quadratic terms to the optimization problem, which makes it extremely complex to solve. They will be considered in the nodal expansion, when the different technologies are compared and chosen. At the TEP level, it is only decided whether or not a capacity should be built between two zones, AC or DC, how many MW and when.

(g) *DC approximation of the flows*

For 2 zones A and B linked by a line with complex impedance $Ze^{i\delta}$, the flow is defined as follows (by neglecting reactive power flows):

$$F_{A\to B} = \frac{U_A \cdot U_B}{Z} \cdot cos(\theta_A - \theta_B - \delta)$$

With the following assumptions:

Voltage is uniform in the system	$U_A = U_B = U_n \cdot u = U_n$
Resistances are neglected ($R \ll X$)	$\delta \sim \pi/2$
Angle difference between two consecutive nodes is small	$\theta_A - \theta_B \ll 1$

Using the per-unit system defined above, we approximate the flow as follows:

$$F_{A\to B} = \frac{S_n}{z} \cdot (\theta_A - \theta_B)$$

(h) No storage management is taken into account.

We consider that the storage has been modelled in the copperplate simulation: our goal is to minimize the deviation of the achievable injections from the "copperplate" injections: the flexibility of the storage facilities is limited to only allow a small deviation.

(i) The slack buses are the same for all the snapshots.

INTRODUCTION

Transmission Expansion Planning (TEP) is traditionally processed over a short period (usually 10 years) and results in a single-step evolution of the transmission grid. This short-term process is efficient to define precisely the needs of the power system for the next years, but it cannot prepare the grid for long-term energy developments. Such developments (technology breakthroughs, radically new energy policies or major societal changes) induce a long adaptation of the grid structures and they should be taken into account when planning a multi-step evolution of the grid. However, planning for such major changes can only be handled through scenarios which describe the possible evolutions of energy requirements, technologies and policies. With multiple scenarios a mitigation method is necessary to ensure a common development in the first time-horizons, able to provide propositions to decision-makers.

The expansion candidates are usually picked by experts and are mainly linked to the current difficulties encountered in the grid. Uncertainties in the system can be taken into account through Monte Carlo simulations and hence impact of investments on system operation can be assessed through the expected operation cost of the actual grid enhanced by the decided investments. We need specific methods to ensure that the choice of the candidates is driven by the future needs of the grid and that the expected operation cost in the "enhanced grid" is representative of the actual operation cost that we could obtain by simulating the operation in this enhanced grid on an hourly time-step.

This task 8.4 "Enhanced Modular Development Plan" is a subtask of the Work Package (WP) 8 "Enhanced Pan-European Transmission Planning Methodology" of the e-Highway2050 project. WP8.4 is therefore linked to other subtasks dedicated to achieve WP8's goal. These subtasks are organized as presented in the following figure where task 8.4 is described as "STEP 4 – Optimal grid expansions at zonal level form today to 2050".

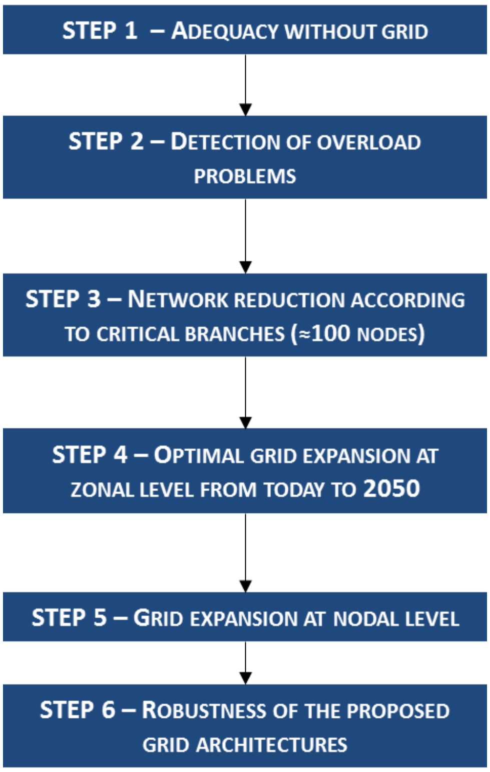

Figure 2 - WP8 Global Methodology

Transmission Expansion Planning optimization is a well-known problem, and we based the optimization problem on existing models [13]-[14]. The ideal approach would be optimizing grid investments while taking into account operational costs over all the snapshots. In a long-term perspective where several scenarios are considered, this would lead to millions of snapshots: the size of the TEP optimization problem would be too large. Thus, we want to drastically reduce the number of snapshots without degrading the grid expansion results. The main focus of our work is to find a Snapshot Selection method meeting these requirements.

This document describes the methodology we propose to take into account long-term evolution of the global energy picture and to suggest different grid evolutions according to different energy scenarios for several time-horizons within this time-frame. This "Modular Development Plan" methodology should be able to handle large networks such as the European power system on a long-term period: selection techniques are required to reduce the computational burden of this optimization.

The document first presents the Snapshot Selection method, which aims at choosing representative operation situations among all the simulated hours for each scenario and time-horizon. Then the Candidate Selection is proposed to only use potentially profitable candidates in the TEP optimization. The model we use for this TEP optimization is then described and adapted to our needs, with a specific method to obtain the Common Development in the first horizons. Finally a small test case and its associated results are presented to illustrate and validate the behaviour of the proposed models.

1. General Method

The goal of the Enhanced Modular Development Plan methodology is to propose grid developments based on a detailed analysis of a grid for several time-horizons and several scenarios. In this document we consider a large network represented at zonal level, i.e. zones are an aggregation of the underlying nodes. For convenience we will indicate the starting year of the study by 2020 and the target year by 2050, but the proposed methodology can easily be adapted to other temporal scopes. The purpose of Transmission Expansion Planning is to provide grid expansion candidates that will allow the grid to take advantage of most of the generating units while reducing the load curtailments at a minimal investment cost. It is a trade-off between economic benefit (minimizing investments) and social benefit (minimizing operational troubles). It is less expensive to building new network components than add new generation units to the system. Thus, we first optimize the production dispatch in a copperplate grid where flows are not constrained, and then we decide grid reinforcements by minimizing the operational cost deviations from the copperplate solution.

Before optimizing the Transmission Expansion Planning two components must be defined: the candidates in which the model is allowed to invest and the snapshots on which the impacts on system operation are assessed. Nowadays most of the TEP methods consider that candidates are provided by experts based on which corridors need reinforcements and where should the most profitable candidates be built. The snapshots are also defined manually and usually reflect the system critical states based on the variations of load and intermittent generation like wind or solar on a few situations: those states usually only describe worst case scenarios or high and low load situations. Both processes can introduce flaws in the problem's definition and then result in suboptimal solutions:

- Hand-picking the candidates prevents us from testing candidates that would become profitable once others have already been implemented. This is even more important when the TEP problem is solved over a long period with several time-horizons: it becomes difficult to estimate the profitability of a candidate without a proper methodology.
- If the chosen snapshots are only based on critical situations, they can lead to pessimistic estimations of operations and over-investment.

For both components we believe that a rigorous methodology should be followed: potential candidates should be chosen for their feasibility and profitability while representative snapshots should be chosen to reflect as much as possible the actual operations of the grid.

The Transmission Expansion Planning optimization in itself is often considered as a two stages problem where the investments are first optimized and then the operations are calculated through an Optimal Power Flow constrained by the previously optimized transfer capacities. This decomposition can be suited when the goal is to plan a "one-step" development of the grid, where the decision variables are optimized only for the target year and the following years are not considered. However when several time-horizons are considered within the span of the study period, it cannot be solved one horizon at a time: all horizons must be taken into account to dynamically optimize the development of the grid. Since we consider a long study period (30 years in our test cases) we need to take into account several scenarios of evolution for the energy mix throughout the years. Those scenarios account for the long-term uncertainties in the development of the system, but short-term uncertainties (e.g. hourly wind energy generation) should also be considered: the stochastic complexity is handled through Monte Carlo simulations of the generation facilities and loads behaviour over one year. All those simulated years are an input to our methodology.

Because grid evolution is a long administrative and technical process, Transmission System Operators usually plan their future expansions 10 years ahead: the grid development cannot be optimized independently for each scenario and we consider that Transmission Expansion Planning in the first two time-horizons (2025 and 2030) should be common to all scenarios. For that purpose we present a formulation of the TEP problem as a single optimization over all the scenarios and all the time-horizons.

We propose to perform the different processes in the following order:

1) Snapshot Selection: find representative snapshots among hourly simulations of the grid
2) Candidate Selection: choose profitable candidates based on the operational savings they could bring (calculated on the selected snapshots)
3) TEP optimization: optimize when and how much the candidates should be invested in (operational consequences of investments are calculated on the selected snapshots)

We consider the following inputs:

- Initial zonal grid (2020) description:
    - Localization of zones and their connections
    - Grid equipment description (AC links capacity and reactance, DC links capacity, phase-shifters flexibility...)
- Zonal injections information
    - Hourly copperplate injections for each time-horizon, scenario and Monte Carlo year (nodal injections from adequacy without grid, aggregated at the zonal level)
    - Injection Boundaries (minimum and maximum)
    - Hourly injections and Local Marginal Prices for each time-horizon, scenario and Monte Carlo year (Optimal Power Flows on the reduced zonal 2020 grid)
- Costs
    - Variation cost (marginal cost of modifying an injection up or down) for each time-horizon, scenario and time-step
    - Investment and Maintenance costs of corridor types
    - Penalization cost (related to the long/short distances architecture)
- Candidates Selection rules (candidates' length threshold...)

As presented in Figure 3, the General Method can be divided into two steps. First the representative snapshots and the potential candidates are selected. Then the Transmission Expansion Planning is optimized under a given architecture focus. Three architecture focuses are considered to propose different options to the decisions-makers:

- Optimal solution: minimization of investment, maintenance and operational costs
- Super-grid: penalization of short distance candidates
- Local development: penalization of long distance candidates

For each architecture focus the method will perform an optimization of the modular Transmission Expansion Planning. This process chooses which successive grid enhancements result in the least cost development for all scenarios, according to the investment cost but also the operational expenses (maintenance and remaining congestions management).

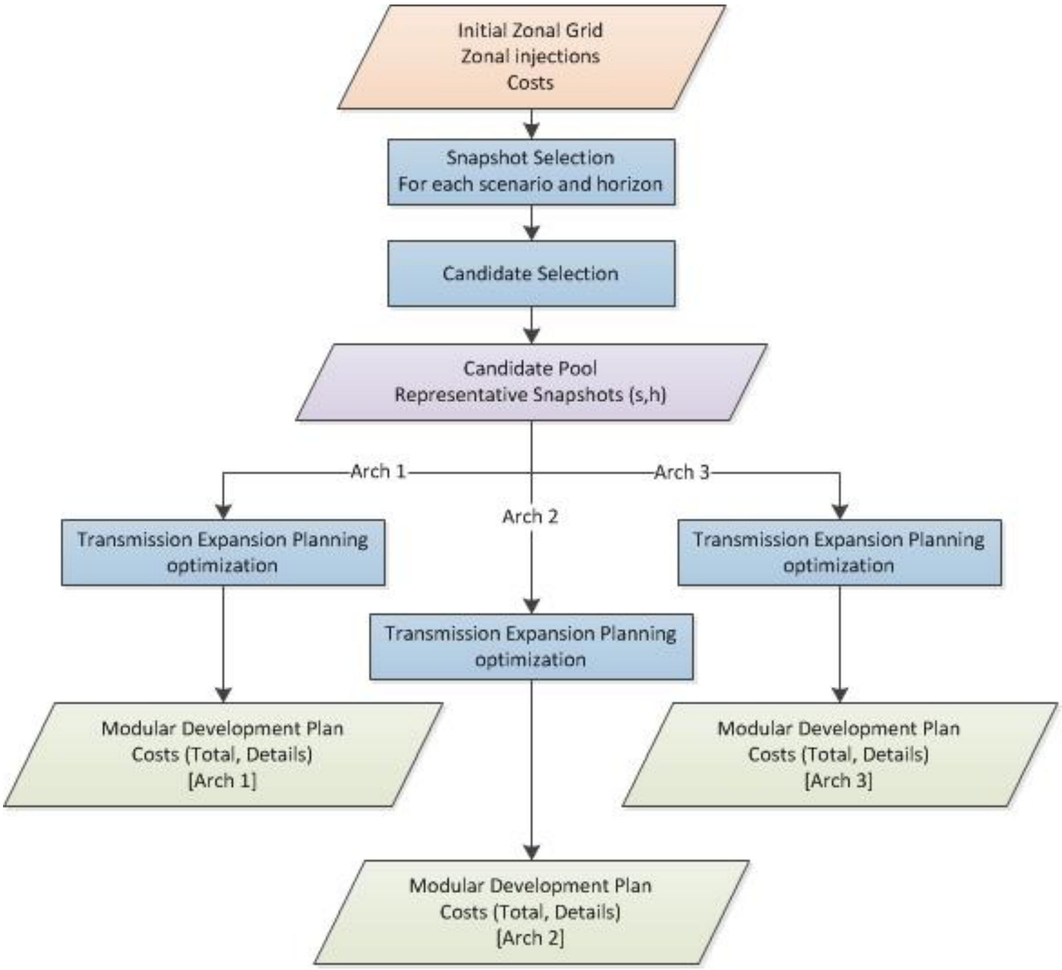

Figure 3 - General Method workflow

2. Snapshot Selection

2.1. Objective and criteria

The aim of the Snapshot Selection is to find a relatively small number of snapshots to represent a whole set of them in all their aspects (not only critical situations). Those representative snapshots will be used to reduce the computation time in the Grid Reduction (task 8.3.1.4) and in the Transmission Expansion Planning optimization (task 8.4). Those two processes have different objectives and thus the ideal representatives might be different for each of them. This section presents a selection method based on clustering theory. Different options are proposed for this method and we only focus on the Snapshot Selection for TEP optimization. Snapshot Selection through clustering has been proposed for studies in security of supply [1], demand analysis [2] or energy prices [3], and it is sometimes applied in TEP works.

The global method – based on clustering theory – is to find groups of similar snapshots within the initial pool obtained with Monte-Carlo simulations for a given scenario and a given time-horizon. Since the number of final clusters K depends a lot on the actual data and is related to the magnitude of computation time reduction we want to achieve, we consider for now that this number is not known and the methods will be tested with different values of K. When the clusters have been formed we need to find a representative object in each of them: the pool of clusters' representatives is the pool of "selected snapshots". To select the representative in each cluster, we follow the medoid definition [4]:

> *The medoid of a cluster is the member which minimizes the quantization error of the cluster: the sum of distances between cluster members and the medoid is minimal.*

The weight w_r of a representative snapshot r is calculated as the ratio between the number of hours that it represents and the total number of hours considered. If the total operational cost in snapshot r is described by $OPEX_r(G)$ for a given grid G, the total operational cost for a given time-horizon and scenario represented by the snapshots in $\mathcal{R}(s,h)$ is obtained as follows:

$$OPEX_{(s,h)}(G) = 8760 \cdot \sum_{r\ in\ \mathcal{R}(s,h)} w_r \cdot OPEX_r(G) \tag{1}$$

The objective of the Snapshot Selection is to find snapshots that can represent the whole pool of grid simulations so that operational expenses can be estimated. To be consistent, the selected snapshots should lead to similar results through the considered process than when it is run with the whole pool of snapshots.

Since our methods are based on clustering they have two main components:

- clustering feature: used to assess the similarity between the snapshots
- clustering algorithm: used to group the snapshots together based on their similarity

In this section we will consider several features and only one algorithm.

The metric used to measure distances between objects is crucial for the outcome of any clustering method. In this work we only consider the Euclidean distance as defined in Equation 2, where p_i is an object of the dataset and $p_i[f]$ is the value of that object for the feature f.

$$d(p_i, p_j) = \sqrt{\sum_{f \in features} \left(p_i[f] - p_j[f]\right)^2} \quad (2)$$

The target number of clusters is also a parameter that we can fix. Some tools can be used to identify the best value for K depending on the data to be clustered [5], [6]. Those techniques usually require running the clustering method for different values of K and analysing the obtained partitions to choose the best K. In this work we consider that this number of clusters is given and we will use a set of 12 arbitrary values of K for our test case.

2.2. Features

In a power system the features can be of various natures: balance-related (load, wind or solar generation), congestion-related (congestion duration, load shedding) etc. We decided to concentrate our study on two different criteria, one balance-related and one congestion-related. However, other criteria could have been studied if more time was available. Net-load is often a good indicator of the balance in the system and we propose to use its components as a base for our features: expected non-controllable demand (NC-D) and expected non-controllable generation (NC-G). On the other hand price-differences between zones can give a feeling of the congestion on the corresponding line and they are triggers for investment in new lines: we propose price-differences as another base for our features. This is based on the assumption that similar price-differences patterns imply similar behaviours in the investment optimization. To capture variations of price-differences and of non-controllable load and generation, we propose a hybrid feature combining them:

- "Hybrid": Local NC-D and NC-G are multiplied by the local energy price to obtain new local values.

Load, generation and price-differences are local values: each clustering feature is associated to a zone or a pair of zones. Thus, the number of features is exponential when the size of the system grows: we propose to replace local features by geographical statistical indicators, such as the mean of price-differences values over the system. Both local and statistical features are implemented and the results of statistical features will be compared to results of local features.

In our methodology, we consider that such features are provided by an external process: to obtain the zonal prices, demand and updated generation, a DC optimal power flow has to be run on the zonal network.

For convenience we use the following abbreviations:

- *Local*: local values
- *Avg*: average value in the system
- *Min*: minimum value in the system
- *Max*: maximum value in the system
- *Std*: standard deviation of all the values in the system
- *Total*: sum of the values in the system

The following are the feature sets we have implemented. For each feature set we specify the number of values it involves for each snapshot (in a system of N zones).

Table I - Features for the Snapshot Selection

Code	*Type*	*Values*	*Number*
$\boldsymbol{F_1^a}$	Price-differences	*Local*	$N \cdot (N-1)/2$
$\boldsymbol{F_1^b}$	Price-differences	*Avg, Min, Max, Std*	4
$\boldsymbol{F_2^a}$	NC-D and NC-G	*Local*	$2 \cdot N$
$\boldsymbol{F_2^b}$	NC-D and NC-G	*Avg, Min, Max, Std*	8
$\boldsymbol{F_2^c}$	NC-D and NC-G	*Total*	2
$\boldsymbol{F_4^a}$	Hybrid	*Local*	$2 \cdot N$
$\boldsymbol{F_4^b}$	Hybrid	*Avg, Min, Max, Std*	8

Notes:

- *Demand represents the initial demand in a given zone from copperplate simulations*
- *If there are several types of non-controllable generation types (wind and solar for example) they are summed up in each zone to obtain the zonal NC-Generation (we do the same for demand).*

2.3. Algorithm

We propose to consider one of the most popular algorithms for clustering data: K-means. This algorithm will partition the data into groups and its iterative resolution technique is readily implemented in the machine learning package "Scikit-learn" available for Python [7]. Other popular algorithms could also be used, such as K-medoids, which usually provides better results than K-means, and Agglomerative Clustering, which builds up the clusters by successively merging the closest clusters until the target number of clusters is reached. Those two alternatives are more described in Appendix A.

The ideal K-means clustering is the minimization of the distortion Δ obtained by the following formula where the mean of each cluster C is indicated by μ_C and its coordinates are averaged over the coordinates of the cluster members:

$$\Delta = \sum_{C} \sum_{i \in C} d(i, \mu_C)^2 \tag{3}$$

Optimizing the cluster assignment to reach the lowest Δ is a NP-hard problem and an algorithm has been proposed to undertake the optimization [8]. This algorithm takes K initial means as an input and iteratively assigns objects to their closest mean and re-computes the means.

Due to this technique the obtained partition is only a local minimum dependent on the initial partition. The initial partition is often obtained by randomly choosing K objects to be the initial means of the K clusters. The algorithm should be performed several times to find the optimal partition or to deduce the expected behaviour of the method.

Since this algorithm only provides the mean of each cluster, we need to find the clusters' medoids to obtain the representatives. A representative can be far from its related mean and therefore be far from representing the other cluster members: we propose to perform once again the K-means algorithm on the same dataset taking the previously obtained representatives as means for the initial partition. This is repeated until the cluster assignment is stable. The stability is defined using the Adjusted Rand Index [9] value between the last partition and the current one: if ARI is above a given threshold ($StableAssign_{threshold}$), the assignment is considered stable. The following figure sums this process up.

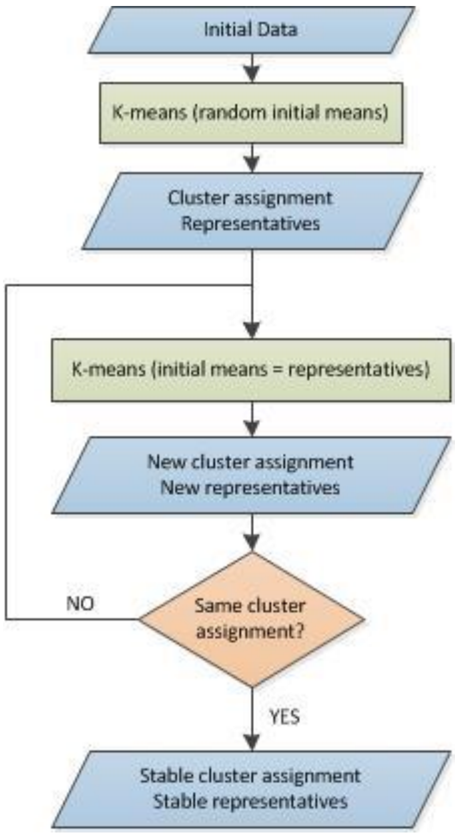

Figure 4 - Algorithm based on K-means

The initial means of the K-means algorithm are obtained through an initialization method: in this work we only consider the "random" initialization, where the K initial means are randomly chosen among the objects to be clustered. Other methods are available such as K-means++ which chooses the initial means so that they are well distributed among the data [10].

2.4. Example

The following figures show an example of Snapshot Selection for the "Garver-like" test case (see Section 6). The clustering feature is the statistical feature for price-differences (F_1^b) and we fixed the number of representative snapshots to 10. Both figures present the 8760 snapshots (marked by a cross) and the 10 representative snapshots (marked by a circle) but under different perspectives: Figure 5 represents the data with respect to the minimum and maximum values in the system, while Figure 6 represents the data with respect to the average and standard deviation values in the system. The values have been scaled so that for each set (average value of price-differences for example) the mean over the snapshots is 0 and the variance is 1. Each cluster has a different colour.

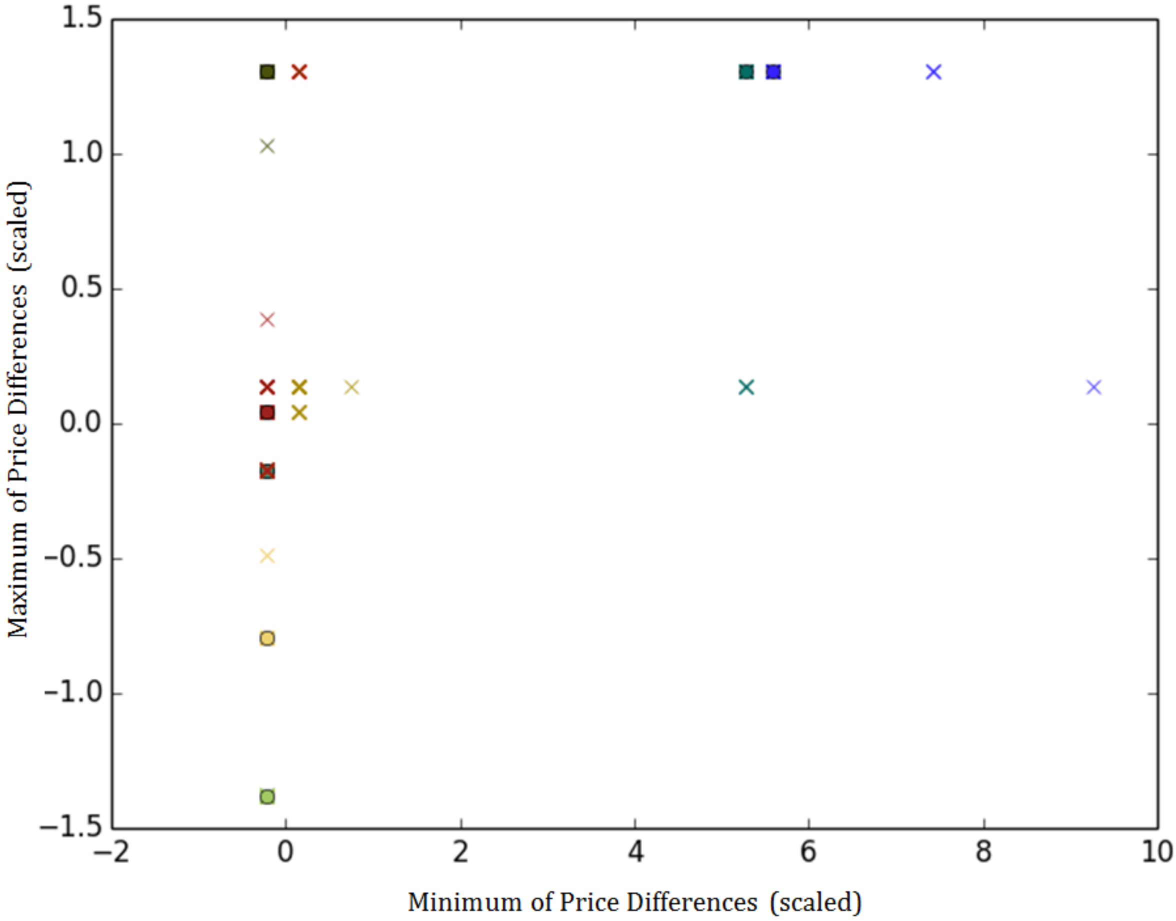

Figure 5 - Example of Snapshots and Representative Snapshots (max VS min)

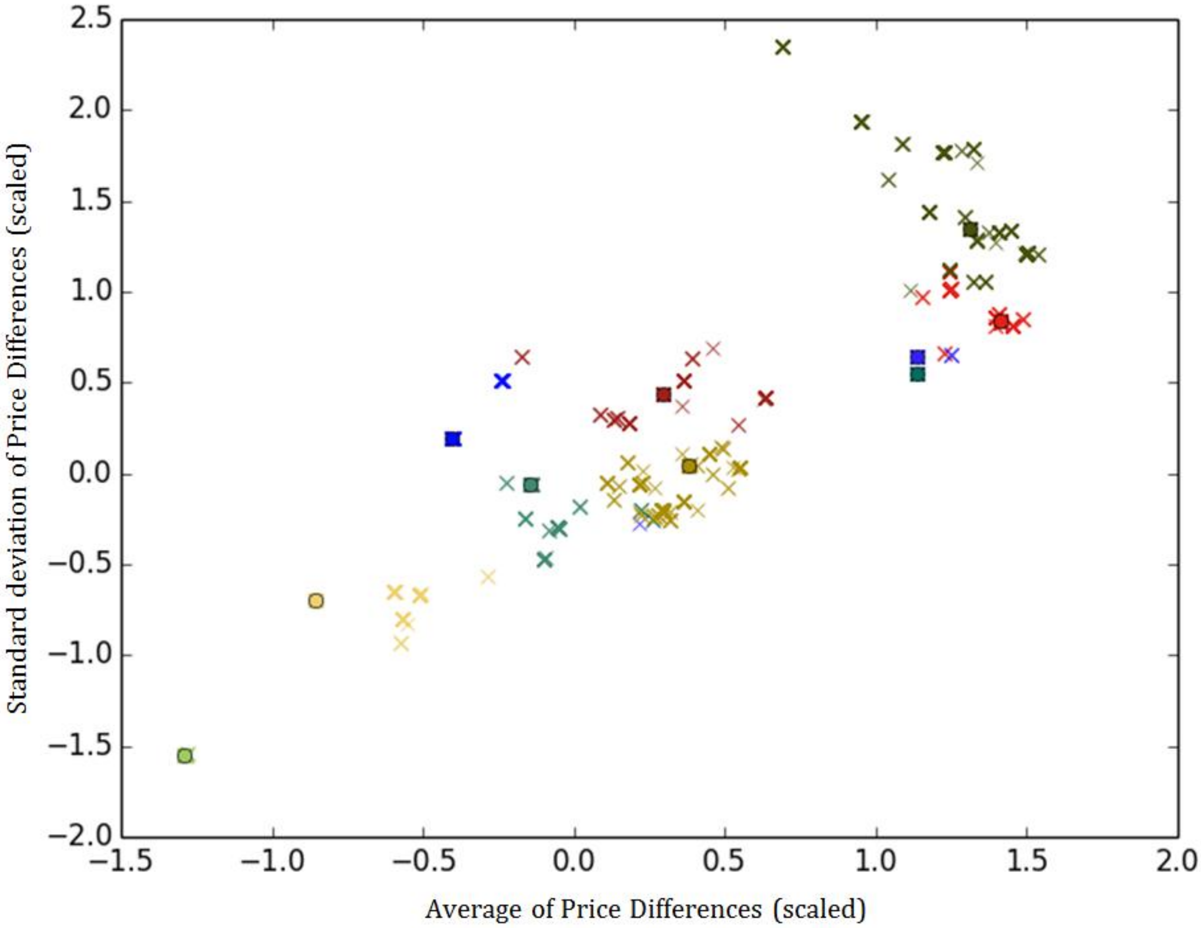

Figure 6 - Example of Snapshots and Representative Snapshots (std VS avg)

Since in our test case local energy prices can only take their values in a small set (formed by marginal costs of variation), price-differences values are also limited to a small set of values: it is very likely that different snapshots end up having the exact same values of price-differences (and hence of price-differences statistics). This explains why we do not actually see 8760 crosses on those figures and why some representative snapshots are not visible in the "maximum-minimum" perspective.

3. Candidate Selection

3.1. Objective

In order to reduce the computation time of the Transmission Expansion Planning (TEP) optimization we want to reduce the number of candidates proposed to the optimization. Without this Candidate Selection the optimization would have to deal with thousands of candidates (and all the combination possibilities) for the grid expansion. As an example, in a 100-zone system with only 2 corridor types, there are $\frac{99\times100}{2}$ pairs of zones, which lead to 9900 possible candidates. Many of those initial candidates are not technically feasible (too long connections, land-use constraints...) and many of the feasible ones are not profitable because they cost more than the potential benefit they can bring. The proposed method operates as a filter by sorting the feasible and profitable candidates out of the initial pool.

Since the TEP optimization considers several scenarios and time-horizons, we propose to find the final pool of candidates by merging the final pools related to all the scenarios and time-horizons: this makes the Common Development in the first time-horizons easier and allows the optimization to pick candidates in a broader pool and find unexpected investments.

3.2. Method

Based on the work of Lumbreras, Ramos and Sanchez [11], this method can be divided into three main processes: identification of the feasible candidates, candidate management and candidate analysis. The first process analyses the initial pool of candidates and removes any candidate which violates a technical constraint: in our work we only consider the length of the candidate. The second process iteratively checks the profitability of the feasible candidates, identifies the candidates to install (simplified TEP on the most profitable candidates), and installs them in the grid. When no new candidates are installed the candidate analysis is performed to find complementary or substitute candidates among the profitable candidates which have not been invested in.

The following figure explains how the three processes are performed. The method is designed so that candidates are selected for each time-horizon and each scenario and then the final candidate pool is obtained by merging the individual pools: this allows the subsequent TEP optimization to use candidates in a time-horizon even though they were not specifically discovered for it and it will help the parallelization when implementing the methodology.

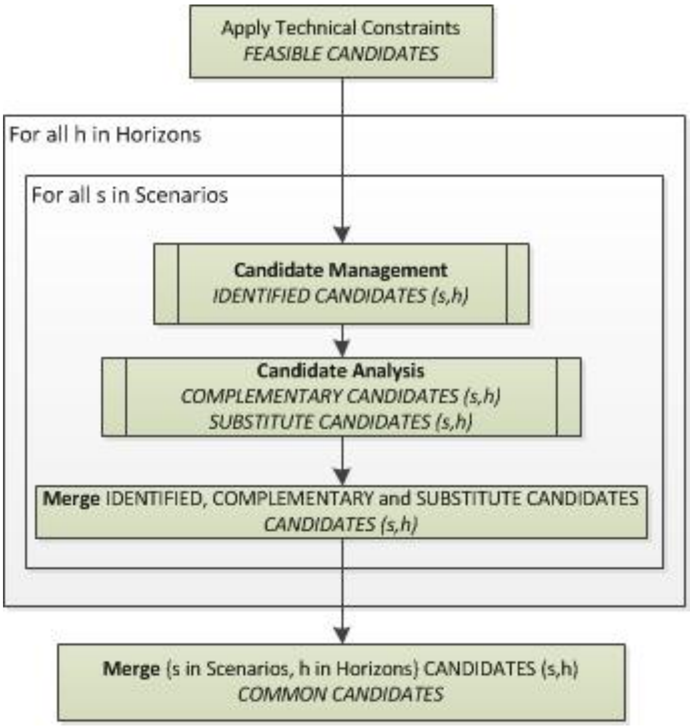

Figure 7 - Candidate Selection general method

The following figures describe the processes of candidate management and candidate analysis.

The candidate management method (Figure 8) is based on analysing the potential benefit of the technically feasible candidates. This Potential Benefit of a candidate is calculated from the price-difference $\Delta Price_r$ between the two connected zones during the snapshot r and the maximal potential flow $\bar{f}$ on the candidate. The profitability P of a candidate $cand$ is then the ratio between the Potential Benefit and the annualized Investment Cost (Equation 4).

$$P_{cand}^{s,h} = \frac{8760 \cdot \sum_{r \in \mathcal{R}(s,h)} w_r \cdot \bar{f}(cand) \cdot |\Delta Price_r(cand)|}{Inv(cand)/LifeTime(cand)} \tag{4}$$

Equation 5 gives the definition of $\bar{f}$ for DC and AC candidates. For DC candidates, the flow is limited by the transfer capacity of the candidate and for AC candidates the flow is constrained by the phase difference between the zones, which is in theory limited by π.

$$\bar{f}(cand) = \begin{cases} Cap_{cand}^{inc} & \text{if } cand \text{ is } DC \\ S_n \cdot Y_{cand}^{inc} \cdot \pi & \text{if } cand \text{ is } AC \end{cases} \tag{5}$$

A candidate with a profitability higher than 1 brings more benefit than it costs in investment, we consider only those candidates to identify the final candidates. Based on the profitable candidates the "Relaxed TEP" optimizes the grid development under relaxed constraints (the proposed MILP resolution method for the TEP optimization presented in Section 4.2 is relaxed and a constraint is added to ensure a proper relaxation, see constraint Relaxation Addition in this final formulation). The goal of the Candidate Selection is to find potential candidates for the TEP optimization: we do not need to obtain optimal expansion solution at this point of the methodology. The optimal expansions obtained under relaxed constraints are therefore rounded to their closest integer value and implemented in the grid so that the next iteration calculates the zonal prices (and hence profitability) taking into account the candidates already identified.

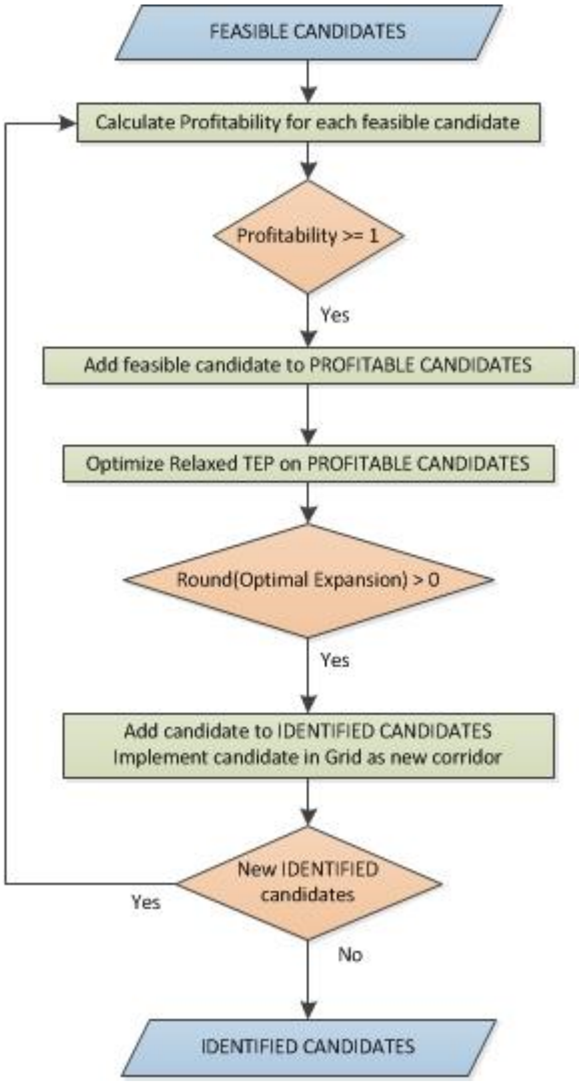

Figure 8 - Candidate Management method

The candidate analysis method (Figure 9) is based on the variation of the utilization of a given installed candidate (IDENTIFIED CANDIDATES pool) when one of the profitable candidates is installed. If the installation of that profitable candidate makes the absolute flow in the considered identified candidate go up, those candidates are complementary; if it goes down they are substitute. Complementary candidates are interesting because they help the system using the full potential of the identified candidates. Substitutes on the other hand are considered as alternative options for the grid development.

Since the expansion decision is integer (and not binary), for a specific profitable candidate the variation of the flow in a given installed candidate is averaged over all the possible increments of the profitable candidate. This averaged variation is then analysed to assess whether the profitable candidate has a complementary or substitute behaviour.

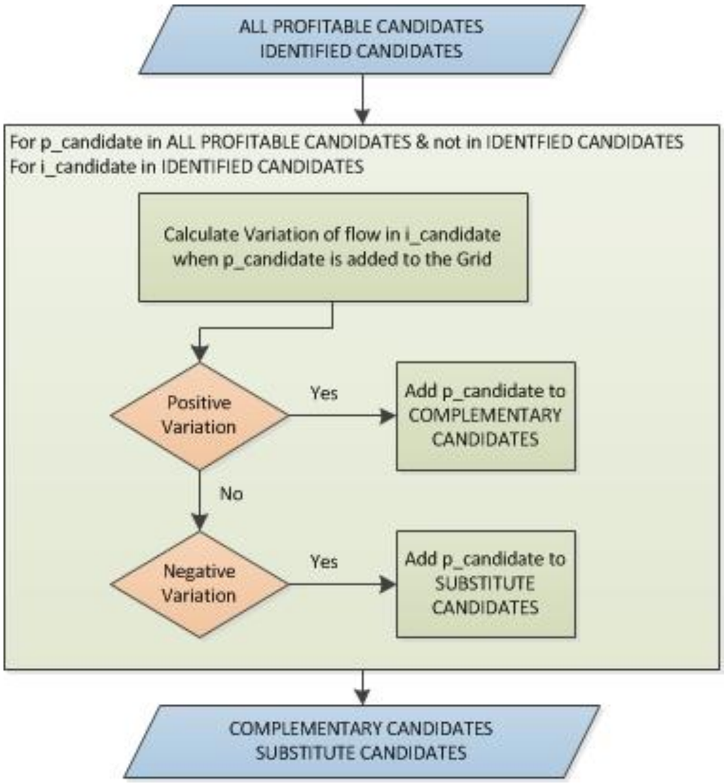

Figure 9 - Candidate Analysis method

To avoid candidates that only slightly modify the flows in identified candidates, we consider a variation threshold $FlowVar_{threshold}$: the candidate is considered worthy of interest (complementary or substitute) only if the variation it causes is greater than the threshold.

3.3. Combined candidates

In this project we add the possibility of switching an existing corridor to another type of corridor (e.g. switch an AC corridor to DC). Two combined candidates are used to achieve such a switch:

- 1 "negative candidate" of the same technology than the existing corridor (i.e. AC or DC) such that it nullifies the previous capacity (and reactance) of the corridor
- 1 "positive candidate" of the new corridor type with the new capacity (and reactance) needed

Since the entire corridor has to be shut down, we only consider the full capacity to be shut down and replaced at once (i.e. the maximal number of units available for those candidates is 1).

The negative capacity and reactance of the "negative candidate" should be calculated from the corridor to be removed. Given the study period (30 years) and the usual life-time of grid assets, there is no reason for building new candidates between two zones and then switching them to another corridor type in the same study period. We consider that combined candidates should only remove existing corridors (i.e. from the grid of the starting year).The capacity of the "positive candidate" is calculated as a multiple of the incremental capacity of the related corridor type so that the new capacity is larger than the decommissioned capacity.

Since the capacity of transfer between two zones is calculated from the initial capacity and the cumulative capacities added by candidates, the capacity related to the initial corridor is reduced to zero thanks to the "negative candidate" from the time-horizon where the switching is decided until the end of the study: that means that assessments of the impacts on system operation take into account the switching without other variables or parameters.

Regarding the cost, we define specific corridor types which are "negative types" for existing regular types: they have the same technology but the investment cost is the cost of removing 1 km of the regular type and OM is the opposite value of the regular type's OM cost. We also consider that a negative candidate has a life-time of 1 year, so that no rest-value can be reclaimed for this investment at the end of the study, only the rest-value of the positive candidate should be considered.

Based on work done in WP3 [12], we consider that the investment cost of a regular candidate is the sum of the costs of equipment, installation, civil works, project management, authorizations and rights of way. Those costs can be expressed as fractions of the total investment cost. For a "negative candidate" we consider that project management, authorizations and rights of way are accounted for in the positive candidate (simultaneous investment) and that equipment cost is zero (we are removing the equipment here). For the sake of simplicity we consider that the cost of removing the equipment is equivalent to the cost of installing it. To model the fact that the infrastructure (towers...) can be reused we consider that the civil work costs of the positive candidate can be reduced by the civil works costs of the regular type related to the negative candidate.

Finally the investment cost of the negative type can be expressed as a fraction ($Removing_{share}$) of the regular type's total investment cost as shown in Equation 6

$$Inv(Neg\ Cand) = Removing_{share} \cdot Inv(Initial\ Corr) \tag{6}$$

As the civil work ($CivilWork_{share}$) has already been done for the corridor that is going to be decommissioned (mainly towers), this cost can be saved from the investment cost of the candidate which will replace the decommissioned corridor (see Equation 7) and the main part of removing an equipment is then the "installation cost" ($Install_{share}$). Since total investment costs of the positive candidate and the regular type to be decommissioned are different, the civil works costs of the combined candidate will not be zero: the difference can be interpreted as an adaptation cost.

$$Removing_{share} = Install_{share} - CivilWork_{share} \tag{7}$$

To include the combined candidates in the Candidate Selection process, we consider that for each regular type of corridors the "negative" type and the "replacement" type are known: when an AC corridor from the existing grid is longer than an arbitrary length, a switch is automatically proposed. Negative types should be well defined so that the existing corridor can actually be "shut down" by a negative candidate. The capacity of the positive candidate is calculated as an integer number of upgrades of the related corridor type so that this capacity covers the decommissioned capacity. When the regular candidates are generated, the Candidate Selection also generates the combined candidates and they are considered "feasible" if the positive candidate is feasible (length and land-use constraints).

The negative and positive candidates of a "Combined Candidate" are treated as one single candidate for the candidate management:

- To compute the profitability, the maximal potential flow $\bar{f}$ introduced in Equation 4 is calculated as follows in each snapshot r for a Combined Candidate cc:

$$\bar{f}^r(cc) = \bar{f}(PosCand[cc]) - f^r(InitialCorr[cc]) \tag{8}$$

  $f^r(InitialCorr[cc])$ is the flow on the corridor to be decommissioned in the particular snapshot r. This value is defined in a similar way than we defined $\bar{f}$ in Equation 5:

$$f^r(InitialCorr[cc]) = \begin{cases} Cap_{DC}(InitialCorr) & \text{if } InitialCorr \text{ is } DC \\ S_n \cdot Y(InitialCorr) \cdot \Delta\theta^r[cc] & \text{if } InitialCorr \text{ is } AC \end{cases} \tag{9}$$

  $\Delta\theta^r[cc]$ is the phase difference between the two considered zones. The actual value might be different when the initial corridor is not constrained, but if there is no congestion between the two zones the local prices are equal and the Potential Benefit becomes zero.
  Finally the obtained Potential Benefit is divided by the annualized sum of both investment costs (negative candidate and positive candidate).

- Their optimal expansion variables are bound to be equal in the relaxed TEP optimization

In the Candidate Analysis, we analyse the impact of installing a switch if it has been found profitable (i.e. its negative and positive candidates are in POTENTIAL CANDIDATES) but has not been installed in the Candidate Management. We also consider the impact of all potential candidates (regular or negative/positive) on "positive candidates" related to installed switches (i.e. related negative and positive candidates are in IDENTIFIED CANDIDATES).

3.4. Example

Figure 10 shows an example of Candidate Selection for the "Garver-like" test case (see Section 6). We only present candidates identified during the Candidate Management phase (red dashed lines are AC candidates and red continuous lines are DC candidates). Given the size of this system and the operational constraints of the case, almost all other candidates are identified during the Candidate Analysis phase. The representative snapshots used for this process are the ones obtained in the example presented in Section 2.4 and no corridor switch is allowed.

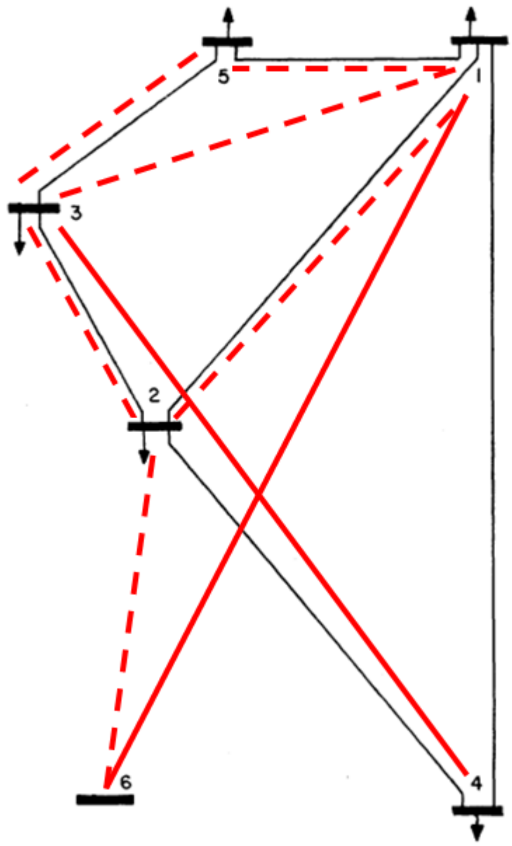

Figure 10 - Example of Candidate Selection

Since the studied system is already really constrained, most of the proposed candidates concern the zones 1 to 5. Only two candidates are proposed to connect zone 6 (new wind production facility) to the rest of the grid.

4. Transmission Expansion Planning optimization

This section presents the mathematical formulation related to the model for the optimization of the Transmission Expansion Planning (TEP) over several time-horizons in different scenarios. The main driver of Transmission Expansion Planning is the cost of the new investments and the benefit we can expect from them when operating the system. These consequences on system operations can also be optimized as a function of the different power components of the system (power flows, injections and angles) subject to the grid constraints in the reinforced network.

We developed a model for those two optimizations and proposed a Mixed Integer Linear Problem formulation as a resolution method in line with formulations of Bahiense et al. [13] and Romero et al. [14].

4.1. High-level description of the problem

The purpose of our methodology is to decide in each scenario and time-horizon which candidates are the most beneficial under operational constraints. We assess the cost implied by the investment in each candidate but also the impacts on system operation that we can expect. We consider that the ideal situation is when production and consumption in each zone are at their "copperplate" level: the operational consequences of reinforcements are the cost of deviation from these ideal injections. Once an investment has been decided, we model the impacts on system operation as the difference between the costs of operating with and without the grid (obtained from task 8.2 "Adequacy without grid"). Therefore an investment is profitable if its cost and its operational consequences are lower than the operational consequences of not investing. The expansion stops when the operational cost has reached the copperplate solution, however we consider that for each time-horizon and each scenario, investments cannot be higher than a fixed limit: this maximum investment could be arbitrarily defined or based on the difference between the expected operational cost with the initial grid and the copperplate cost for the considered time-horizon and scenario.

The operational consequence of investments cannot be assessed for only one operating state of the system: we need to take into account as many situations as possible (from regular states to crisis states). We consider representative snapshots (r) obtained from the Snapshot Selection and we model the impact of investments on system operation over a year as the weighted sum of the operational consequences on those individual representative snapshots.

For this high-level description we use simplified symbols:

- $CAPEX\left(NewG_{s,h}\right)$: Investment cost of optimal grid enhancements (maintenance is included)
- $OPEX_r\left(TotalG_{s,h}\right)$: Impact on system operation of investments represented by $TotalG_{s,h}$ (total capacity $TotalCap_{s,h}$ and admittance $TotalY_{s,h}$ in a given scenario and time-horizon) in the snapshot r
- $NewG_{s,h}$: Available candidates which are installed in a given scenario and time-horizon (composed of the capacity $NewCap_{s,h}$ and the admittance $NewY_{s,h}$of the installed candidate)
- $VarCost(I^r, I_0^r)$: Cost of modifying the dispatch from the reference I_0^r to the optimal I^r found under the grid constraints in snapshot r.

Since the desired methodology should handle several scenarios we formulate the problem with scenario weights so that every scenario can be considered in the same optimization. Every time-horizon is also taken into account so that the developments of a given time-horizon depend on the optimal development of the previous ones: this is the modular development plan.

Also, because grid evolution is a long administrative and technical process, Transmission System Operators usually plan their future expansions 10 years ahead: we consider that Transmission Expansion Planning in the first two time-horizons (the first 10 years of the study period) should be common to all scenarios (Common Development Horizons CDH).

We present a high-level formulation of the TEP problem as a bi-level optimization over all the scenarios and all the time-horizons in the following problem.

$$\min_{NewG} \sum_{s,h} w_s \cdot \left(CAPEX\left(NewG_{s,h}\right) + E_{\mathrm{r} \in \mathcal{R}(s,h)}\left[OPEX_r\left(TotalG_{s,h}\right)\right]\right)$$

subject to:

$$CAPEX\left(NewG_{s,h}\right) \leq InvMax_{s,h}$$

$$TotalCap_{s,h} = TotalCap_{s,h-h_{step}} + NewCap_{s,h}$$

$$TotalY_{s,h} = TotalY_{s,h-h_{step}} + NewY_{s,h}$$

$$OPEX_r\left(TotalG_{s,h}\right) = \min_{I^r} VarCost(I^r, I_0^r)$$

$$\begin{aligned} &\text{subject to:} \\ &|Flows(I^r)| \leq TotalCap_{s,h} \\ &LocalBalance(I^r) = 0 \\ &ACFlows(I^r) = Angles(I^r) \cdot TotalY_{s,h} \end{aligned}$$

Problem 1 – High-Level Description

In this simple formulation, $E_{\mathrm{r} \in \mathcal{R}(s,h)}$ is the expectation over the snapshots related to the considered scenario and time-horizon. I^r is the optimal dispatch of generating units allowed by the grid in snapshot r. The objective of the second optimization is to match the optimal dispatch with the reference dispatch I_0^r ("copperplate" injection dispatch in the snapshot r) by minimizing the variation cost caused by this dispatch modification.

From this high-level description and since investments only depend on the equipment ($NewG_{s,h}$), we can split the problem into two levels:

- Investment Level (1st optimal function), where only the grid is optimized,
- Operational Level (2nd optimal function), where the injections are optimized and constrained by the grid.

When those two levels are combined, products between decision variables and angle variables appear: a basic formulation of the full optimization becomes non-linear. Since the considered products relate to integer variables on the one hand and continuous variables on the other hand, we can easily modify the formulation through a "disjunctive model" [13] so that the optimization can be written as a Mixed Integer Linear Programming problem. The two levels and this modification are fully formulated in Appendix B.

4.2. Final Formulation

Architecture Focus

To obtain different expansion plans, we propose an architecture factor which multiplies the investment cost of a given candidate with respect to its length. This penalization is computed as follows for each candidate k:

$$ArchFactor_k = \left(1 + \sigma_+ \cdot \frac{\left(Len_k - L_{ref}\right)^+}{L_{ref}} + \sigma_- \cdot \frac{\left(Len_k - L_{ref}\right)^-}{L_{ref}}\right) \qquad (10)$$

In this formulation, Len_k is the length of the considered candidate while the other symbols represent parameters: L_{ref} is the "reference" length (i.e. the threshold above or under which a candidate is considered too long or too short) and $\sigma_{+/-}$ are penalization "costs" (unit-less).

Three architecture focuses are defined:

- Optimal solution: minimization of investment, maintenance and operational costs
- Super-grid: penalization of short distance candidates to only keep long reinforcements
- Local development: penalization of long distance candidates

Table II presents the range of the parameters needed to obtain the desired architecture.

Table II - Range of values for the architecture factor parameters

Architecture Focus	$\boldsymbol{L_{ref}}$	$\boldsymbol{\sigma_+}$	$\boldsymbol{\sigma_-}$
Optimal solution	/	0	0
Super-grid	$]0;+\infty[$	0	$]0;+\infty[$
Local development	$]0;+\infty[$	$]0;+\infty[$	0

For example with $\{L_{ref} = 100km\ ;\ \sigma_+ = 1\ ;\ \sigma_- = 0\}$ a candidate of 200km would be twice as expensive in the "Local development" focus while a candidate of 50km would not be affected.

Mixed Integer Linear Programming formulation

When we combine the two levels of Problem 1, non-linear terms involving angle and decision variables appear and a disjunctive model is used to obtain a linear formulation (see Appendix B for details). We obtain the Mixed Integer Linear Programming formulation presented in Problem 2.

In this formulation the objective function is the sum of the following items (in this order)

- Investment Cost for the candidates
- Maintenance Cost for the candidates

- Impact on system operation averaged on all the snapshots
- Rest value of the investments

The last constraint (Relaxation Addition) is proposed to ease the relaxation of the TEP optimization in the Candidate Management.

$$\min_{n,\theta,I,f} \left\{ \sum_{s \in \mathcal{S}} w_s \sum_{h \in \mathcal{H}} \left[DF(h - y_0) \left[\sum_{k \in \mathcal{C}} n_k(s,h) \cdot Inv_{Type(k)}(s,h) \cdot Len_k \cdot ArchFactor_k \right.\right.\right.$$
$$+ h_{step} \left(\sum_{\substack{j \in \mathcal{H} \\ j \leq h}} \sum_{k \in \mathcal{C}} n_k(s,h) \cdot OM_{Type(k)}(s,h) \cdot Len_k \right.$$
$$\left.\left. + 8760 \sum_{r \in \mathcal{R}(s,h)} w_r \sum_{z \in Z} \sum_{i \in InjTypes} MC_+^r(z,i) \cdot \Delta I_+^r(z,i) + MC_-^r(z,i) \cdot \Delta I_-^r(z,i) \right) \right]$$
$$- DF\left(H + h_{step} - y_0\right) \sum_{k \in \mathcal{C}} n_k(s,h) \cdot \frac{Inv_{Type(k)}(s,h)}{LT_{Type(k)}} \cdot Len_k \cdot ArchFactor_k$$
$$\left.\left. \cdot \left(h + LT_{Type(k)} - \left(H + h_{step}\right)\right)^+ \right] \right\}$$

subject to:

$n_k(s,h)\ integer$

$\beta^{s,h}(k,b)\ binary$

$Dev_k(s,h)\ binary$

$\forall s \in [\![2..N_S]\!], \forall h \in CDH, \forall k \in \mathcal{C},$

$$n_k(s,h) = n_k(s-1,h)$$

(Common Development)

$\forall s \in [\![1..N_S]\!], \forall h \in \mathcal{H}, \forall (k_1,k_2) \in CombCand, \qquad n_{k_1}(s,h) = n_{k_2}(s,h)$

(Combined Candidates)

$\forall s \in \mathcal{S}, \forall h \in \mathcal{H}, \forall k \in \mathcal{C},$

$$0 \leq n_k(s,h) \leq n_k^{Max} \cdot Dev_k(s,h)$$

(Investment increment)

$\forall s \in \mathcal{S}, \forall k \in \mathcal{C},$

$$\sum_{h \in \mathcal{H}} Dev_k(s,h) \leq 1$$

(Limited development)

$$\forall s \in \mathcal{S}, \forall h \in \mathcal{H}, \sum_{k \in \mathcal{C}(h)} n_k(s,h) \cdot Inv_k(s,h) \cdot Len_k \leq INV_{max}(s,h)$$

(Investment maximum)

$\forall s \in \mathcal{S}, \forall h \in \mathcal{H}, \forall r \in \mathcal{R}(s,h), \forall z \in Z,$

$$\sum_{z_A \in Z} f_{DC}^r(z_A, z) - \sum_{z_B \in Z} f_{DC}^r(z, z_B) + \sum_{z_A \in Z} f_{AC}^r(z_A, z) - \sum_{z_B \in Z} f_{AC}^r(z, z_B) + \sum_{i \in InjTypes} sg(i) \cdot I^r(z,i) = 0$$

(Power Balance)

$\forall s \in \mathcal{S}, \forall h \in \mathcal{H}, \forall r \in \mathcal{R}(s,h), \forall (z_A, z_B) \in Z \times Z,$

$$|f_{AC}^r(z_A, z_B)| \leq Cap_{AC}(z_A, \mathrm{z_B}) + \sum_{\substack{j \in \mathcal{H} \\ j \leq h}} \sum_{\substack{k \in \mathcal{C}_{AC} \\ ZoneA(k) = z_A \\ ZoneB(k) = z_B}} n_k(s,j) \cdot Cap_k^{inc}$$

(AC Flow Capacity)

$\forall s \in \mathcal{S}, \forall h \in \mathcal{H}, \forall r \in \mathcal{R}(s,h), \forall (z_A, z_B) \in Z \times Z,$

$$|f_{DC}^r(z_A, z_B)| \leq Cap_{DC}(z_A, \mathrm{z_B}) + \sum_{\substack{j \in \mathcal{H} \\ j \leq h}} \sum_{\substack{k \in \mathcal{C}_{DC} \\ ZoneA(k) = z_A \\ ZoneB(k) = z_B}} n_k(s,j) \cdot Cap_k^{inc}$$

(DC Flow Capacity)

$\forall s \in \mathcal{S}, \forall h \in \mathcal{H}, \forall r \in \mathcal{R}(s,h), \forall (z_A, z_B) \in Z \times Z,$

$$f_{AC}^{r}(z_A, z_B) = \sum_{\substack{j \in \mathcal{H} \\ j \leq h}} \sum_{\substack{k \in \mathcal{C}_{AC} \\ ZoneA(k)=z_A \\ ZoneB(k)=z_B}} \sum_{b=1}^{n_k^{Max}} S_N \cdot Y_k^{inc} \cdot \left(\tilde{\theta}_{z_A}^{r,j}(k,b) - \tilde{\theta}_{z_B}^{r,j}(k,b)\right) + S_N \cdot \left[Y(z_A, z_B) \cdot \left[\theta_{z_A}^{r} - \theta_{z_B}^{r}\right] + \sum_{\substack{c \text{ in } Corr_{PST} \\ ZoneA(c)=z_A \\ ZoneB(c)=z_B}} Y_c \cdot \alpha^r(c) \right]$$

(AC Flow Definition)

$\forall s \in \mathcal{S}, \forall h \in \mathcal{H}, \forall r \in \mathcal{R}(s,h), \forall c \in Corr_{PST}$,

$$\alpha_{min}^{r}(c) \leq \alpha^r(c) \leq \alpha_{max}^{r}(c)$$

(Phase Shifter Boundaries)

$\forall s \in \mathcal{S}, \forall h \in \mathcal{H}, \forall r \in \mathcal{R}(s,h), \forall z \in Z, \forall i \in InjTypes$,

$$I^r(z,i) - I_0^r(z,i) = \Delta I_+^r(z,i) - \Delta I_-^r(z,i)$$

(Injection Variations)

$\forall s \in \mathcal{S}, \forall h \in \mathcal{H}, \forall r \in \mathcal{R}(s,h), \forall z \in Z, \forall i \in InjTypes$,

$$I_{\min}^{r}(z,i) \leq I^r(z,i) \leq I_{\max}^{r}(z,i)$$

(Injection Boundaries)

$\forall s \in \mathcal{S}, \forall h \in \mathcal{H}, \forall r \in \mathcal{R}(s,h), \forall z \in SwBus$,

$$\theta_z^r = 0$$

(Swing Buses Angle)

$\forall s \in \mathcal{S}, \forall h \in \mathcal{H}, \forall r \in \mathcal{R}(s,h), \forall j \leq h, \forall k \in \mathcal{C}_{AC}, \forall b \in \llbracket 1, n_k^{Max} \rrbracket, \forall z \in \{ZoneA(k), ZoneB(k)\}$,

$$-\beta^{s,j}(k,b) \cdot M \leq \tilde{\theta}_z^{r,j}(k,b) \leq \beta^{s,j}(k,b) \cdot M$$

$$-\left(1 - \beta^{s,j}(k,b)\right) \cdot M \leq \theta_z^r - \tilde{\theta}_z^{r,j}(k,b) \leq \left(1 - \beta^{s,j}(k,b)\right) \cdot M$$

$$\sum_{b=1}^{n_k^{Max}} \beta^{s,h}(k,b) = n_k(s,h)$$

(Extended Disjunctive Model)

$\forall s \in \mathcal{S}, \forall h \in \mathcal{H}, \forall r \in \mathcal{R}(s,h), \forall k \in \mathcal{C}_{AC}, \forall b \in [\![1, n_k^{Max}]\!]$,

$$-\beta^{s,j}(k,b) \cdot Cap_k^{inc} \leq S_N \cdot Y_k^{inc} \cdot \left(\tilde{\theta}_{ZoneA(k)}^{r,j}(k,b) - \tilde{\theta}_{ZoneB(k)}^{r,j}(k,b)\right) \leq \beta^{s,j}(k,b) \cdot Cap_k^{inc}$$

$$\text{(Relaxation Addition)}$$

Problem 2 – MILP Formulation of the TEP optimization

<u>Variables to Optimize</u>

$n_k(s,h)$: Number of installed units for candidate k in scenario s and time-horizon h

$Dev_k(s,h)$: Binary variable which controls that a candidate is used only once in each scenario

$\beta^{s,h}(k,b)$: Binary decomposition of the related number of units

I : Injection array describing Power Injections for each type and each zone

$\Delta I_{+/-}$: Variations of the injections

θ : Phase Angle in each zone

α : Phase Shifter angle on each initial corridor with PST

$\tilde{\theta}_z^{r,j}$: Non-linear replacement of the product "number-of-units x angles"

$f_{AC/DC}$: Power flow between two zones

4.3. Example

With the snapshots selected in Section 2.4 and candidates (including complementary and substitute candidates) selected in Section 3.4 we are able to run the TEP optimization for the "Garver-like" test case (see Section 6). The following figure shows the optimal expansion of the candidates for the "optimal solution" architecture focus mentioned above. AC candidates are represented with red dashed lines and DC candidates with red continuous lines. The numbers show the optimal amount of installed units for each candidate.

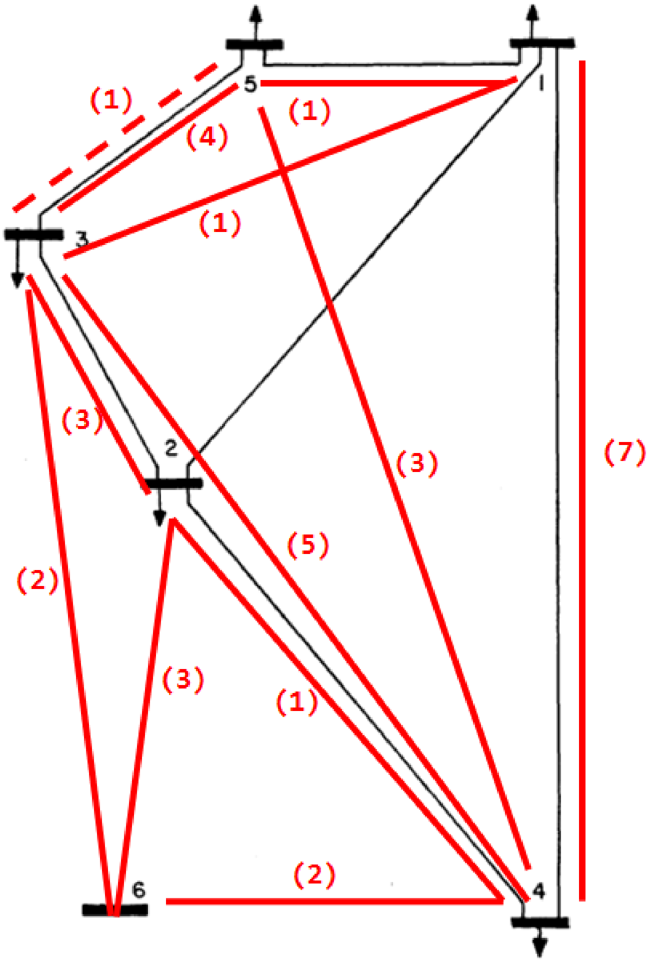

Figure 11 - Example of TEP optimization

When comparing with the candidates proposed during the Candidate Management on the same test case and with the same representative snapshots (see Section 3.4), we see that most of the identified pairs of zones are actually connected (except (1,2) and (1,6) which are not installed). All the other installed candidates come from the Candidate Analysis: this shows the necessity of this phase of the Candidate Selection to find complementary and substitute candidates. However, almost all installed candidates are DC whereas most of the identified candidates during Candidate Management are AC. This behaviour could be an indication either that the profitability of AC candidates is over-estimated during Candidate Management or that the model of DC flows is too simple in the TEP optimization.

5. Tests objectives

The proposed methodology is split into three modules: Snapshot Selection, Candidate Selection and TEP optimization. The main parameters to be defined are related to the Snapshot Selection: the clustering feature (F_i^v presented in Section 2.2) and the number of clusters (K). The features are of different natures (price differences or demand and generation) and use different values (local or statistical values). In this section we first present the indicators used to monitor the performance of each module and then the methodology used to compare the results obtained with different choices of features and numbers of clusters.

5.1. Performance indicators

For each run of the whole methodology, we record the following information after completion of the different modules:

- Snapshot Selection
    - Final Clustering solution: cluster assignment for each snapshot
    - Computation time
- Candidate Selection
    - Set of selected candidates when considering only the selected representative snapshots
    - Computation time
- TEP optimization
    - Expansion solution when considering only the selected representative snapshots and the selected candidates
    - Computation time

We propose to use this information to compare the outcome of each module obtained with different choices of similarity features and numbers of clusters.

As a reminder $Card(set)$ is the number of elements in the considered set.

5.1.1. <u>Adjusted Rand Index</u>

The Adjusted Rand Index (ARI) [9] is designed to compare two clustering solutions of a given set of objects. ARI measures the degree of correspondence between two clustering solutions and is based on the comparison of assignment of every pair of objects in the dataset between the two solutions. Given two solutions U and V there are 4 types of pair of objects (see Figure 12 below where three clusters are represented – grey, red and green – and 4 pairs of objects have been identified – one of each type):

1. Objects are in the same cluster in U and in the same cluster in V
2. Objects are not in the same cluster in U and not in the same cluster in V
3. Objects are not in the same cluster in U and in the same cluster in V
4. Objects are in the same cluster in U and not in the same cluster in V

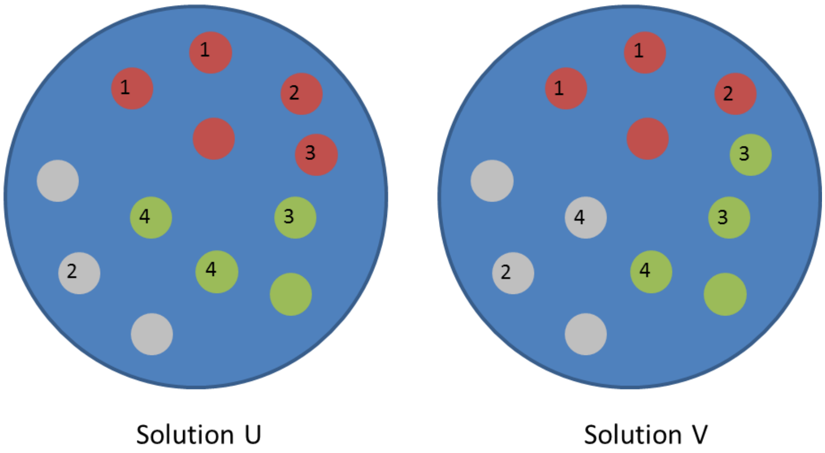

Figure 12 - Example of two clustering solutions

Pairs of types 1 and 2 are agreements between solutions U and V, while types 3 and 4 are disagreements.

The Rand Index (RI) is the probability of agreement as described in the following formula, where A is the number of agreements (number of pairs of types 1 and 2) and n is the number of objects in the data set.

$$RI = A/\binom{n}{2} \tag{11}$$

The Rand Index has been corrected so that it takes a constant value when the partitions are chosen under an appropriate null model [9] and the resulting ARI is obtained as follows:

$$ARI = \frac{RI - E(RI)}{1 - E(RI)} \tag{12}$$

$E(RI)$ is the expected value of the Rand Index under hypergeometric distribution assumption, i.e. the value taken by RI when U and V are drawn randomly with a fixed number of clusters and a fixed number of elements in each cluster [15].

Note: clustering solutions do not need to have the same number of clusters.

5.1.2. <u>Candidate Selection Index</u>

We propose to measure the difference between two Candidate Selection solutions by comparing the set of selected candidates for each solution. Since the candidates are selected from an initial pool composed of all the feasible connections, the selected candidates are a subset of this initial pool. We define the following similarity index between two sets of selected candidates SC_1 and SC_2:

$$CSI = \frac{Card(SC_1 \cap SC_2)}{Card(SC_1 \cup SC_2)} \tag{13}$$

With this measure we simply calculate how many candidates the two sets have in common.

5.1.3. TEP solution indicators

We propose to measure the difference between two expansion solutions by comparing the optimal integer decisions for each expansion candidate. The TEP is optimized over several time-horizons and a candidate can only be expanded once. Since those decisions are integer and can be taken in different time-horizons, we consider three levels of comparison:

- Agreement on whether or not the considered candidate should be built (i.e. expansions are both zero or both strictly positive)
- If the solutions agree on the candidate expansion: compare the values of the optimal expansion for the considered candidate in both solutions
- If the solutions agree on the candidate expansion (and both expansions are strictly positive): compare when the considered candidate is built in both solutions.

The TEP is optimized over different scenarios: indicators are calculated for each scenario and averaged using scenario weights.

We consider two solutions S_1 and S_2 obtained on two candidate sets SC_1 and SC_2:

- $h_{cand}(S)$ is the time-horizon where the optimal expansion for $cand$ is non-zero (if any) in solution S.
- $inv_{cand}(S)$ is the optimal expansion decision for candidate $cand$ in solution S (strictly speaking it is the sum of optimal decisions for that particular candidate over all time-horizons for the considered solution).

For simplicity we use the following symbols:

$$\rho_{12} = inv_{cand}(S_1) \cdot inv_{cand}(S_2)$$
$$\varepsilon_{12} = |inv_{cand}(S_1) - inv_{cand}(S_2)|$$
$$\eta_{12} = |h_{cand}(S_1) - h_{cand}(S_2)|$$

SC_1 and SC_2 can be different: if a candidate is in SC_1 but not in SC_2 we consider that its optimal expansion for solution S_2 is 0 (and vice-versa).

Expansion Solution Index

The first comparison is handled through the Expansion Solution Index (ESI), calculated as follows for the considered solutions:

$$ESI = \sum_{cand} \frac{\gamma_{cand}(S_1, S_2)}{Card(SC_1 \cup SC_2)} \tag{14}$$

γ_{cand} can take two different values:

$$\gamma_{cand}(S_1,S_2) = \begin{cases} 0 & \text{if } \rho_{12} = 0 \text{ and } \varepsilon_{12} \neq 0 \\ 1 & \text{else} \end{cases} \tag{15}$$

The values of γ_{cand} depend on the agreement between the two investment solutions for the considered candidate:

- If expansion decisions are equal ($\varepsilon_{12} = 0$) then γ_{cand} is 1
- If expansion decisions are different but none of them is 0: solutions agree on investing in the candidate then γ_{cand} is also 1
- If expansion decisions are different and one of them is 0, i.e. one solution invests in the candidate while the other does not ($\varepsilon_{12} \neq 0$ and $\rho_{12} = 0$) then γ_{cand} is 0

A value of ESI close to 1 indicates similar expansion solutions while a value close to 0 indicates really different investment solutions.

Expansion Absolute Difference

The second comparison is handled through the Expansion Absolute Difference (EAD), calculated as follows for the considered solutions over the previously found Agreed Candidates:

$$EAD = \sum\nolimits_{cand} \frac{\varepsilon_{12}}{Card(AgreedCandidates_{12})} \tag{16}$$

To avoid comparing a positive expansion with a zero expansion, only the agreed candidates (i.e. for which $\rho_{12} \neq 0$ or $\varepsilon_{12} = 0$) are taken into account.

A value of EAD close to 0 indicates similar expansion solutions for Agreed Candidates.

Time-horizon Absolute Difference

The third comparison is handled through the Time-horizon Absolute Difference (TAD), calculated as follows for the considered solutions over the previously found Agreed Candidates:

$$TAD = \sum\nolimits_{cand} \frac{\lambda_{cand}(S_1,S_2)}{Card(AgreedCandidates_{12})} \tag{17}$$

The values of λ_{cand} depend on the difference between the two expansion solutions for the considered candidate. Only the agreed candidates (i.e. for which $\rho_{12} \neq 0$ or $\varepsilon_{12} = 0$) are taken into account. The difference in time-horizons λ_{cand} for a given candidate is assessed as follows:

$$\lambda_{cand}(S_1,S_2) = \begin{cases} 0 & \text{if } \varepsilon_{12} = 0 \\ \eta_{12} & \text{if } \rho_{12} \neq 0 \end{cases} \tag{18}$$

A value of TAD close to 0 indicates similar decisions on *when* investing in the Agreed Candidates.

5.1.4. Averaged indicators

Since a random initialization method is used for the first K-means clustering in the Snapshot Selection, we will perform several runs of the methodology for a given feature: the whole process described in Figure 3 (with only one architecture focus) is considered as one run. Raw information (final cluster assignment, selected candidates, expansion solution, computation times) and the computed indicators ARI, CSI, ESI, EAD and TAD are stored for each run. To understand the impact of the different features, we average some of that information over the runs.

Since ARI, CSI, ESI, EAD and TAD are symmetric and involve 2 solutions (clustering, candidate set or expansion), we need to calculate the average of those indicators on each pair of solutions as follows:

$$\tilde{I}\big(S(f_1);S(f_2)\big) = \frac{1}{Card(runs_1)\cdot Card(runs_2)}\cdot \sum_{\substack{r\in runs_1\\ q\in runs_2}} I\left(S_r(f_1);S_q(f_2)\right) \tag{19}$$

$\tilde{I}\big(S(f_1);S(f_2)\big)$ represents the average value of the considered indicator I between the solutions obtained with feature f_1 and feature f_2. $Card(runs_1)$ is the number of runs performed for feature f_1 and $S_r(f_1)$ is the actual solution obtained in a given run r for feature f_1.

The computation time values are averaged as follows:

$$\widetilde{CT}(f) = \frac{1}{Card\big(runs_f\big)}\cdot \sum_{r\in runs_f} CT(f,r) \tag{20}$$

$CT(f,r)$ is the computation time of the considered module for feature f_1 for the specific run r.

5.2. Comparison methodology

(a) Validity of the statistical features

The goal of this comparison is to assess the error done when using statistical features rather than local features for the Snapshot Selection

For each value of the number of clusters K in a given ordered set $[K_1, K_2 \dots K_m]$ we consider the following "reference" results:

- **Reference partition:** partition obtained with local features
- **Reference candidates:** result of Candidate Selection on the reference representative snapshots (obtained from the reference partition)
- **Reference expansion solution:** optimal expansion for the reference candidates on the reference representative snapshots (obtained from the reference partition)

We propose to use ARI as a measure of the accuracy of the "statistical partition" (i.e. obtained with statistical features), CSI as a measure of the accuracy of the "statistical candidates set" (i.e. obtained on the "statistical partition") and ESI, EAD, TAD as measures of the accuracy of the investment solution obtained on the statistical partition with the statistical candidates set. The Computation Time ratio (CTr) between statistical and local runs is checked globally and for each module to ensure that using statistical features actually reduces the computation time.

The following table summarizes the features used to obtain reference results and those used to obtain their estimated counterparts.

Table III - Comparisons for the validity of the statistical features

Feature of the Reference Solution	*Feature of the Estimated Solution*
$\mathbf{F}_1^{a}$	F_1^{b}
$\mathbf{F}_2^{\mathbf{a}}$	F_2^{b}
$\mathbf{F}_2^{\mathbf{a}}$	$\mathrm{F}_2^{\mathrm{c}}$
$\mathbf{F}_3^{\mathbf{a}}$	F_3^{b}
$\mathbf{F}_3^{\mathbf{a}}$	F_3^{c}
$\mathbf{F}_4^{\mathbf{a}}$	$\mathrm{F}_4^{\mathrm{b}}$

(b) Sensitivity of the Expansion Solution to the value of K

The goal of this comparison is to assess the sensitivity of the expansion solution (obtained through TEP optimization) to the number of representative snapshots in order to find which feature induces the lowest sensitivity. Each expansion solution is compared to the one obtained with all the snasphots (reference), giving an idea of the error brought by Snapshot Selection.

Using the TEP optimization model presented in Section 4, for each feature set ($\mathbf{F}_i^{v}$) we consider that the **expansion solution obtained with all the snapshots (no Snapshot Selection) is the reference expansion solution**. We propose to evaluate the sensitivity of the expansion solution to the value of K by calculating the ESI values between the solutions obtained for different K values and the reference solution (with all snapshots, i.e. K = 8760).

Table IV - ESI Comparisons for the K-sensitivity of TEP

Reference Solution (all snapshots, i.e. K = 8760)	*Number of snapshots for the Estimated Solution*
8760	K_1
8760	K_2
8760	…
8760	K_m

We propose to observe the variability of ESI when the value of K decreases. A good feature would make the TEP optimization the least sensitive to the value of K.

By nature statistical features should show higher values of ESI than local features. However, the value of the ESI is not our focus here, we are mainly interested in its variability: we analyse the sensitivity brought by Local features ($\mathbf{F}_1^{a}$, $\mathbf{F}_2^{a}$, $\mathbf{F}_3^{a}$, $\mathbf{F}_4^{a}$) together and by Statistical features ($\mathbf{F}_1^{b}$, $\mathbf{F}_2^{b}$, $\mathbf{F}_3^{b}$, $\mathbf{F}_4^{b}$) together.

6. Small test case based on Garver System: "Garver-like" test case

In this section we propose a small test case adapted from the Garver test case [16]. We identified the Garver zones with European areas so that historical data, such as load, can be directly used to describe operational situations.

To fairly compare results of the test case, we need to set the general parameters (Table V).

Table V - General parameters

Parameter	*Value*	*Comment*
$\boldsymbol{\tau}$	0.1	Discount rate is 10%
$\boldsymbol{K}$	[5,10,15,20,25,30,35,40,45,50,75,100]	12 values of K are tested
$\boldsymbol{StableAssign_{threshold}}$	0.9	Cluster assignment is considered stable when the similarity between two successive clustering solutions is higher than 90%
$\boldsymbol{FlowVar_{threshold}}$	0.5	In the Candidate Analysis, complementary and substitute candidates are only considered if they change the flow of more than 50% in at least one of the identified candidates
$\boldsymbol{\sigma_+}$; $\boldsymbol{\sigma_-}$; $\boldsymbol{L_{ref}}$	0 ; 0 ; /	Architecture focus is "optimal solution"

For the sake of simplicity, we relax the maximum investment constraint.

6.1. Network data

In his paper, Garver proposed a power system of 5 existing zones and one new zone as we can see on Figure 13. In order to fit this structure, we grouped some European countries together.

On this representation, the percentages reflect the share of total demand of each zone (initial load for Garver). Each zone can roughly be related to one set of countries regarding this share of total demand and European countries consumption (ENTSO-E 2012 Consumption [18]).

Zone 6 (with no demand) is kept to model a new production site (offshore wind or large PV farms for example).

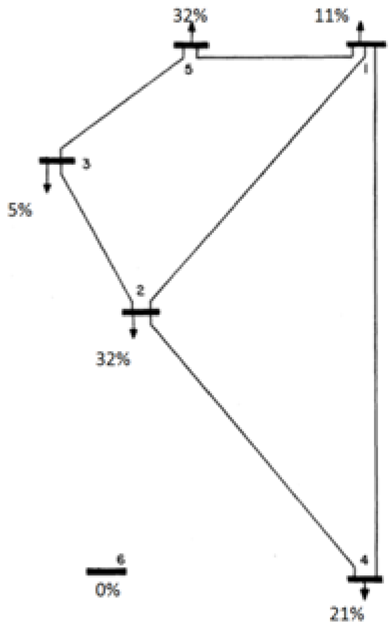

Figure 13 - Garver system with demand shares

Based on the distances between zones provided in the Garver's paper and the disposition of the Garver system, we estimated the zones' position as proposed in Table VI.

Table VI - Zones' Coordinates

Zone	*X-coordinate [km]*	*Y-coordinate [km]*
1	0	0
2	45	45
3	60	16
4	0	96
5	32	0
6	48	96

For our methodology we need to choose a swing (or slack) bus – the reference zone for phase angles. As this choice does not have any influence on the results, it is possible to pick out any of the zones. However the usage is to choose the most meshed zone. In our test case, zone 1 and zone 2 have the same number of connections to other zones, but from its location zone 2 is more likely to develop more connections during the expansion of the power system. The swing bus is then chosen to be zone 2 for the whole optimization.

We propose two types of corridor as presented in Table VIITable VIII: one AC and one DC. The technical details are sourced from the work of e-Highway2050 WP3 [12]. Type 1 is based on AC OHL 400kV (2050) and type 2 is based on DC OHL 500kV (2050).

Table VII - Corridor Types

Index	Technology	Annual Maintenance [k€/km/year]	Total Investment [k€/km]	Reactance [p.u./km]	Capacity [MW]	Life Time [years]	Maximal Length* [km]
1	AC	27	1350	0.0004371	4300	100	100
2	DC	11	1770	0	4000	100	1000

* Those corridor types do not have technical constraints on their length: we chose arbitrary values for test purposes.

For the possible switch of a corridor from type 1 to type 2, we consider the negative corridor type presented in Table VIII. It has been built to fit the method described in Section 3.3.

Table VIII - Negative Corridor Type

Index	Technology	Annual Maintenance [k€/km/year]	Total Investment [k€/km]	Reactance [p.u./km]	Capacity [MW]	Life Time [years]	Maximal Length [km]
3	AC	-27	337.5	0	0	1	100

In the system, we consider the nominal power (Sn) of 100 MVA. It is used to calculate per unit values of reactance.

The initial corridors (Table IX) are based on the way Garver zones are connected and capacities are aggregated Net Transfer Capacities (NTC) [19]. No PST is considered in this test system.

Table IX - Initial Corridors Configuration

Index	Zone A	Zone B	Type	Capacity [MW]	Reactance [p.u.]
1	1	2	1	3405	$7.2 \cdot 10^{-3}$
2	1	4	2	650	/
3	1	5	1	700	$3.6 \cdot 10^{-3}$
4	2	3	1	100	$3.7 \cdot 10^{-3}$
5	2	4	1	7935	$7.6 \cdot 10^{-3}$
6	3	5	1	850	$3.6 \cdot 10^{-3}$

The reactance values are based on the distances between zones, which do not reflect actual distances between European areas.

6.2. Injections data

For this test case we consider that each zone can operate three types of injection: non-controllable generation (Pnc), controllable generation (Ptc) and non-controllable demand (Cnc) as listed in Table X. Only one year (8760 snapshots) is considered.

Table X - Injection Types

Name	*Sign*	*Controllable*
Pnc	+1	0
Ptc	+1	1
Cnc	-1	0

We simulate the system in a copperplate state at the zonal level in each snapshot (hourly simulation): the reference injections for Cnc, Pnc and Ptc are obtained for each zone and each snapshot. For simplicity, we consider that for all injections the lower bound (I_{min}) is 0 and the upper bound (I_{max}) is the initial value (value before simulation for Cnc and Pnc, generation capacity for Ptc).

This test case has not been checked for adequacy: it is possible that even the initial system is not able to serve the load on a “copperplate network”.

7. Results

At the time of writing, the "Garver-like" test case has only been run with features F_1^a, F_1^b, F_4^a and F_4^b: the Snapshot Selection with features F_2^a and F_2^b took too much time compared to the others and we preferred stopping the computation and analyze the available results. The complete method (snapshot selection, candidate selection and TEP optimization) was tested on the "Garver-like" test case for the first analysis comparing local and statistical features. However, we arbitrarily used a set of 9 candidates based on the expansion solution proposed in Garver's paper [16] for the second analysis, as the reference simulation considering all the snapshots was taking too much time. In the "Garver-like" test case only one time-horizon is considered, the Time-horizons Absolute Difference is also meaningless and is not presented. Only the computation time of the Snapshot Selection is studied.

For each feature and each value of K, 50 runs of the methodology have been considered to obtain the averaged values (see Section 5.1.4 for details). One run of the TEP was also made considering all the snapshots (reference simulation), and 50 other runs were performed with a random selection of 5 snapshots (worst case).

7.1. Validity of the statistical features

The goal of this comparison is to assess the error done when using statistical features rather than local features for the Snapshot Selection.

Feature F_1 (price differences)

The local feature represents the system with 15 values and the statistical feature with 4 values (see Table I to calculate the number of values for each feature). The comparisons of results for ARI, CSI, ESI, EAD and Snapshot Selection computation times between F_1^a and F_1^b are presented in Table XI. Those results can be interpreted as follows:

- The obtained partitions show relatively high ARI values (especially for $K \geq 15$), which means that statistical and local partitions are similar. ARI values seem to stabilize around 0.88 after $K = 20$: the difference in assignment between local and statistical partitions is about 12%.
- The CSI values show that the sets of candidates obtained with statistical and local partitions are close, especially for $K \geq 20$ from which the index is above 0.8.
- The ESI values show that the agreement error between Expansion Solution obtained through statistical partition and through local partition is lower than 15% after $K = 25$. The index seems to stabilize around 0.9 for $K \geq 45$.
- When the "statistical" and "local" solutions agree on whether a candidate should be built or not, the expansion error is quite small (less than 1 unit)
- Snapshot Selection with the statistical feature takes almost the same time than with the local feature (the average running times are identical).

Table XI - Results for F_1^a vs F_1^b

K	ARI	CSI	ESI	EAD	CTr
5	0.55	0.41	0.63	0.29	0.93
10	0.69	0.69	0.50	0.56	1.18
15	0.86	0.74	0.74	0.51	0.98
20	0.87	0.80	0.77	0.44	1.02
25	0.87	0.87	0.85	0.53	1.00
30	0.89	0.82	0.84	0.68	0.98
35	0.87	0.84	0.84	0.68	1.00
40	0.87	0.83	0.86	0.76	0.95
45	0.87	0.86	0.90	0.58	1.00
50	0.88	0.89	0.92	0.62	0.98
75	0.86	0.88	0.85	0.71	1.00
100	0.89	0.96	0.91	0.62	1.00

Feature F_4 (hybrid: non-controllable generation and demand multiplied by local marginal price)

The local feature represents the system with 12 values and the statistical feature with 8 values (see Table I to calculate the number of values for each feature). The comparisons of results for ARI, CSI, ESI, EAD and Snapshot Selection computation times between F_4^a and F_4^b are presented in Table XII. Those results can be interpreted as follows:

- The obtained partitions show relatively low ARI values, which means that statistical and local partitions are not similar: the difference in assignment between local and statistical partition is about 50-70%.
- The obtained sets of candidates show relatively high CSI values for $K \geq 25$, which means that the proposed candidates are similar. The error is below 15% after $K = 40$.
- The ESI values show that the agreement error between Expansion Solution obtained through statistical partition and through local partition is lower than 15% after $K = 35$ and the index is around 0.9 for $K \geq 45$.
- When the "statistical" and "local" solutions agree on whether a candidate should be built or not, the expansion error is quite small (less than 1 unit)
- Snapshot Selection with the statistical feature seems to be slower than with the local feature, although it can be faster sometimes: it can take up to 46% less time ($K = 35$) and up to 248% more time ($K = 75$).

Table XII - Results for F_4^a vs F_4^b

K	ARI	CSI	ESI	EAD	CTr
5	0.42	0.42	0.58	0.29	2.03
10	0.50	0.62	0.58	0.20	2.38
15	0.46	0.72	0.81	0.65	1.95
20	0.39	0.63	0.77	0.68	2.60
25	0.32	0.69	0.73	0.80	1.89
30	0.33	0.68	0.79	0.67	1.46
35	0.38	0.72	0.83	0.69	0.54
40	0.35	0.88	0.89	0.61	2.05
45	0.36	0.94	0.92	0.85	0.72
50	0.37	0.89	0.89	0.61	2.03
75	0.33	0.94	0.95	0.54	3.48
100	0.32	0.97	0.91	0.48	1.71

7.2. *Sensitivity of the Expansion Solution to the value of K*

The goal of this comparison is to assess the sensitivity of the expansion solution (obtained through TEP optimization) to the number of representative snapshots in order to find which feature induces the lowest sensitivity.

Unlike the previous section, we compare each expansion solution (ESI) to the reference one, without Snapshot Selection. However, simulating the complete method with all the 8760 snapshots would take too much time. Thus, we predefined a set of 9 candidates based on the expansion solution proposed in Garver's paper [16] in order to skip the candidate selection process.

Local features

The results for ESI between the reference expansion solution (all 8760 snapshots) and the estimated expansion solutions (between 5 and 100 representative snapshots) for the local features are presented in Figure 14. The number of representative snapshots is displayed on the x-axis, the ESI on the y-axis.

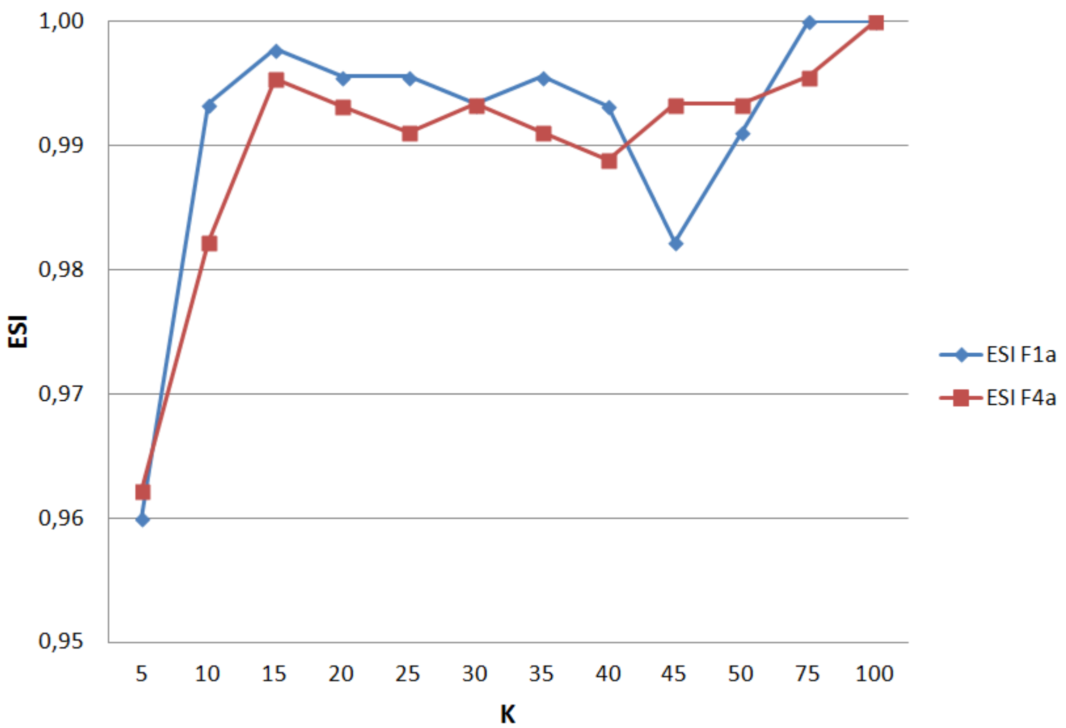

Figure 14 - Results for local features

The variability of the results decreases when $K \geq 15$ and most of the ESI values are higher than 0.99 above this limit. The expansion solutions of both local features F_1^a and F_4^a present a similar trend. However, F_1^a seems to show better results.

To check the validity of the results, we compared these ESI values to the one obtained for the worst case (50 runs using 5 randomly selected snapshots). The ESI value between the reference expansion solution and the random expansion solution is 0.59, which is a lot lower than the values we obtained with the local features (ESI = 0.96).

The standard deviation of the series of values is given in Table XIII. Feature F_4^a seems to be a bit more sensitive than feature F_1^a.

Table XIII - Standard deviations for ESI (local features)

Feature	Standard deviation
F_1^a	0.024
F_4^a	0.028

Statistical features

The results for ESI between the reference expansion solution (all 8760 snapshots) and the estimated expansion solutions (between 5 and 100 representative snapshots) for the statistical features are presented in Figure 15. The number of representative snapshots is displayed on the x-axis, the ESI on the y-axis.

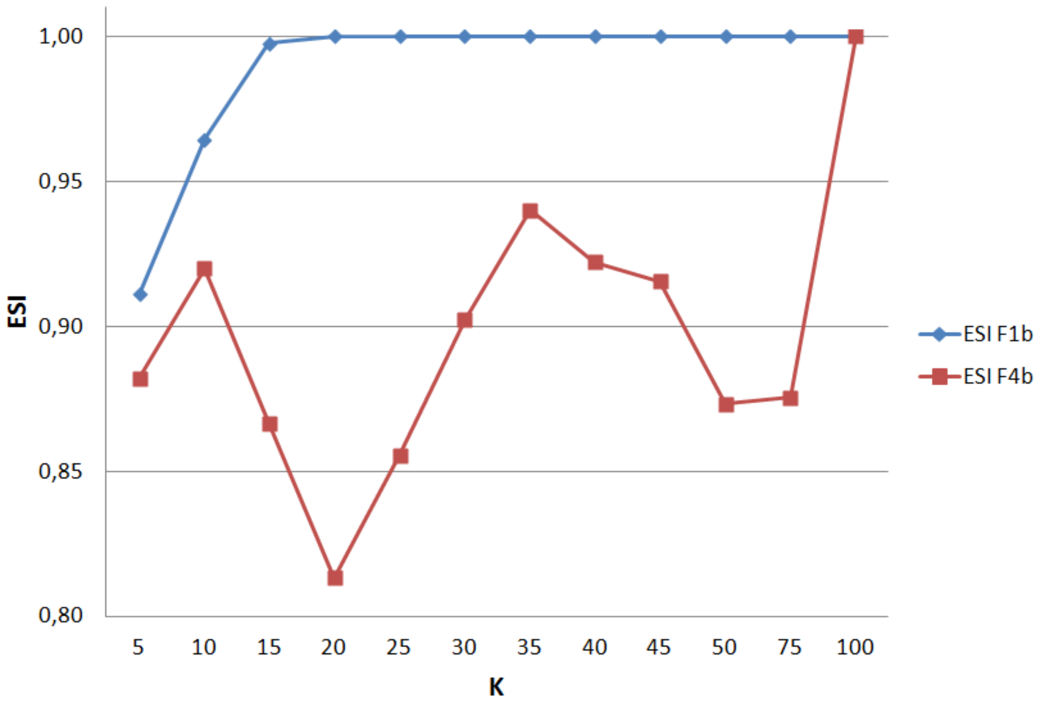

Figure 15 - Preliminary results for statistical features

Unlike local features, the expansion solutions of statistical features are very different. ESI values for F_1^b present a stable trend, and expansion solutions are identical to the reference one for $K \geq 20$. However, the results for F_4^b are chaotic and do not seem to stabilize, although most of them are higher than 0.85, which is a lot better than the ESI value obtained for randomly chosen snapshots (ESI = 0.59).

The standard deviation of the series of values is given in Table XIV. Feature F_4^b seems to be a lot more sensitive than feature F_1^b.

Table XIV - Standard deviations for ESI (statistical features)

Feature	Standard deviation
$\boldsymbol{F_1^a}$	0.010
$\boldsymbol{F_4^a}$	0.081

Those last results are in favour of feature F_1^b, which presents very good ESI values, even better than the local features.

7.3. Tests conclusions

On this small test case (6 zones), using statistical values instead of local values as clustering features seems to be valid (especially for values of K higher than 25 for F_1, and higher than 35 for F_4) from the investment agreement point of view (ESI): the local and statistical expansion solutions agree on about 85% of the candidates. Besides, expansion differences (EAD) are very small once both statistical and local solutions agree on whether a candidate should be built or not: the absolute difference is lower than one unit.

Regarding Snapshot Selection, the obtained partitions for statistical and local $F_1$are more similar than for F_4. Indeed, the difference between partitions stabilizes around 12% for F_1while it is about 50-70% for F_4.

From the Candidate Selection point of view, the sets of candidates for statistical and local features are close and they have more than 80% of common candidates for values of K higher than 20 for F_1 (resp. 40 for F_4).

The second analysis comparing expansion solutions for different values of K to the reference solution considering all the snapshots shows very conclusive results regarding the statistical price-differences feature F_1^b. For this small test case, expansion solutions with Snapshot Selection (F_1^b) are almost identical to the ones without Snapshot Selection for values of K higher than 15. In contrast, the statistical hybrid feature F_4^b present very chaotic results which do not seem to stabilize.

The computation time of Snapshot Selection seems to depend on each individual run and it is not possible for now to conclude on the benefit of using statistical features over the local ones. However it is known that computation time evaluation in a multiprocessors and multiusers environment is not easy.

The features only based on non-controllable demand and generation (F_2^a and F_2^b) seem to take much more time to compute than the other features.

All these different tests and analyses tend towards the same conclusion: the statistical price-differences feature presents good results and expansion solutions using this feature are very close to the one without Snapshot Selection. However, more tests should be performed on larger systems as the small size of this test case may bias the simulations.

8. Conclusions and future research

To achieve the goal of task 8.4 we proposed three processes:

- Snapshot Selection to find representative snapshots among all the simulated operational situations
- Candidate Selection to identify potential expansion candidates without any prior knowledge but their technical constraints
- Transmission Expansion Planning optimization to know when and how much those candidates should be invested in while taking into account the impacts of such investments on system operation over the selected snapshots

Since we consider several scenarios and several time-horizons, the outcome of such a methodology is a modular development plan of the zonal grid from 2020 to 2050 with a common development in the first time-horizons. This common development will be analyzed in task 8.3.2 to propose different expansions at the nodal level for the first time-horizons.

The different analyses we carried on the small test case suggest that the *statistical price-differences* can be used as a clustering feature for the Snapshot Selection with values of K higher than 15. This feature uses 4 values (instead of the number of pairs of zones) to describe the snapshots and hence should reduce drastically the complexity of the clustering. However on a 6-zone system the computation time results were not conclusive: tests on a larger system should show the benefit of using statistical values instead of local values.

To complete our work, the feature based on non-controllable generation and demand should be studied and the complete method should be tested on a larger network.

To provide a better methodology, some areas could be more developed:

- The different clustering features were arbitrarily picked among popular indicators for system operation. Correlation of the proposed features (and other potential features) with the value of OPEX or with the output of TEP optimization should be checked to see how the different parameters (price-differences, demand, generation etc.) influence the outcome of the processes. This will help find "a-priori" good features.
- The method for Snapshot Selection could be linked to the subject of "reference case selection for Transmission Expansion Planning". Nahmmacher et al. [20] recently presented an approach similar to the Snapshot Selection and used it to select representative days for long-term power system models.
- The number of clusters used in the Snapshot Selection was considered given in our methodology: the user has to decide the number of representative snapshots he wants to find. However, methods (based on silhouette values [21] for example) can be developed to guess a good number of clusters by analyzing the data to be clustered.
- The Candidate Selection method can be enhanced by taking into account the land-use constraints of the corridor types (e.g. use specific underwater types where it is needed or modify the investment cost when the considered candidate has to go through mountains or around protected areas).

9. References

[1] H. Kile, “Evaluation and Grouping of Power Market Scenarios in Security of Electricity Supply Analysis,” 2014.

[2] R. Green, I. Staffell, and N. Vasilakos, “Divide and Conquer? K-Means Clustering of Demand Data Allows Rapid and Accurate Simulations of the British Electricity System,” *IEEE Trans. Eng. Manag.*, vol. 61, no. 2, pp. 251–260, May 2014.

[3] F. Martínez-Álvarez, A. Troncoso, J. C. Riquelme, and J. M. Riquelme, “Partitioning-Clustering Techniques Applied to the Electricity Price Time Series,” in *Intelligent Data Engineering and Automated Learning - IDEAL 2007*, H. Yin, P. Tino, E. Corchado, W. Byrne, and X. Yao, Eds. Springer Berlin Heidelberg, 2007, pp. 990–999.

[4] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons, 2009.

[5] A. M. Mehar, K. Matawie, and A. Maeder, “Determining an optimal value of K in K-means clustering,” in *2013 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2013, pp. 51–55.

[6] R. Tibshirani, G. Walther, and T. Hastie, “Estimating the number of clusters in a data set via the gap statistic,” *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 63, no. 2, pp. 411–423, Jan. 2001.

[7] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, “Scikit-learn: Machine Learning in Python,” *J. Mach. Learn. Res.*, vol. 12, p. 2825−2830, Oct. 2011.

[8] A. K. Jain, *Algorithms for clustering data*. Englewood Cliffs, N.J: Prentice Hall, 1988.

[9] L. Hubert and P. Arabie, “Comparing partitions,” *J. Classif.*, vol. 2, no. 1, pp. 193–218, Dec. 1985.

[10] D. Arthur and S. Vassilvitskii, “K-means++: The Advantages of Careful Seeding,” in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, Philadelphia, PA, USA, 2007, pp. 1027–1035.

[11] S. Lumbreras, A. Ramos, and P. Sánchez, “Automatic selection of candidate investments for Transmission Expansion Planning,” *Int. J. Electr. Power Energy Syst.*, vol. 59, pp. 130–140, Jul. 2014.

[12] A. Vafeas, T. Pagano, and E. Peirano, “D3.1 Assessment from 2030 to 2050,” e-Highway2050, European Commission, report, Aug. 2014.

[13] L. Bahiense, G. C. Oliveira, M. Pereira, and S. Granville, “A mixed integer disjunctive model for transmission network expansion,” *IEEE Trans. Power Syst.*, vol. 16, no. 3, pp. 560–565, Aug. 2001.

[14] R. Romero, A. Monticelli, A. Garcia, and S. Haffner, “Test systems and mathematical models for transmission network expansion planning,” *IEE Proc. - Gener. Transm. Distrib.*, vol. 149, no. 1, p. 27, 2002.

[15] S. Wagner and D. Wagner, *Comparing Clusterings- An Overview*. 2007.

[16] L. L. Garver, “Transmission Network Estimation Using Linear Programming,” *IEEE Trans. Power Appar. Syst.*, vol. PAS-89, no. 7, pp. 1688–1697, Sep. 1970.

[17] K. Brunix, D. Orlic, D. Couckuyt, N. Grisey, B. Betraoui, Y. Surmann, T. Anderski, N. T. Franck, G. Keane, B. Hickman, D. Huertas-Hernando, R. Jankowski, and M. Wilk, “D2.1 G/D/E Scenarios 2050,” e-Highway2050, European Commission, report, Jul. 2014.

[18] ENTSO-E, “Monthly Consumption of all countries in 2012.” 23-Jan-2014.

[19] ENTSO-E, “Indicative values for Net Transfer Capacities (NTC) in Europe - Summer 2010,” 07-Jun-2010. [Online]. Available: https://www.entsoe.eu/publications/market-reports/ntc-values/ntc-matrix/Pages/default.aspx.

[20] P. Nahmmacher, E. Schmid, L. Hirth, and B. Knopf, “Carpe Diem: A Novel Approach to Select Representative Days for Long-Term Power System Models with High Shares of Renewable Energy Sources,” Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2537072, Dec. 2014.

[21] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987.

[22] B. F. Hobbs, G. Drayton, E. B. Fisher, and W. Lise, “Improved Transmission Representations in Oligopolistic Market Models: Quadratic Losses, Phase Shifters, and DC Lines,” *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1018–1029, Aug. 2008.

[23] C. Audet, P. Hansen, B. Jaumard, and G. Savard, “Links Between Linear Bilevel and Mixed 0–1 Programming Problems,” *J. Optim. Theory Appl.*, vol. 93, no. 2, pp. 273–300, May 1997.

10. Appendices

A. Alternative algorithms for the Snapshot Selection method

(a) Partitioning Around Medoids ("K-medoids")[1]

This clustering attempts at minimizing the total quantization error:

$$TQE = \sum_{C} \sum_{\substack{i \in C \\ i \neq m_C}} d(i, m_C) \tag{21}$$

Like K-means, this optimization is NP-hard and a sub-optimal iterative technique has to be performed to find acceptable solutions. The most known technique has been proposed by Kaufman and Rousseeuw [4] and proceeds, after an initialization phase, by changing one medoid at each iteration to find a partition which lowers the TQE. Like K-means, K-medoids results are dependent on the initial partition, which is often randomly chosen. We only consider the random initialization.

The representatives are directly obtained by choosing the clusters medoids provided by the algorithm.

(b) Agglomerative Clustering with Complete Linkage (ACCL)[2]

Agglomerative Clustering iteratively builds up the groups from individual clusters to one big cluster with all the data. At each step the algorithm merges the two "closest" clusters already built. The distance between clusters is defined by the chosen linkage, which is usually picked from the three following ones:

- Complete Linkage: merges clusters which have the minimal maximal distance between them
- Single Linkage: merges clusters which have the minimal minimal distance between them
- Average Linkage: merges clusters which have the minimal distance between their centres

We choose to use the Complete Linkage criteria since it is the most used [1]. The distance between two clusters C_A and C_B under Complete Linkage is given in Equation 22.

$$D(C_A, C_B) = \max_{\substack{x \in C_A \\ y \in C_B}} d(x, y) \tag{22}$$

The Agglomerative Clustering algorithm can be represented by a dendrogram, a tree where the leaves are the individual snapshots in their own clusters and the root is only one big cluster with all the data: cutting the dendrogram at a certain height results in a partition of the data (based on the number of clusters we want).

Since no cluster centre is provided during this process we define the representatives as the medoid of each cluster. Unlike K-means or K-medoids, ACCL is deterministic and we only need to perform it once.

---

[1] In Pycluster package for Python

[2] In scikit-learn package for Python

No initialization method is needed for this algorithm.

B. Construction of the MILP formulation for the TEP optimization

Mathematical Formulation of investment and operational level

Investment Level

This level optimizes the number of units needed for each candidate in every situation (s, h) so that the total cost triggered by those decisions is minimal. The total cost has 5 main components:

- **Total Investment Cost** according to the number of increments for each candidate: it is totally paid in the considered time-horizon.
- **Operation and Maintenance Cost** for corridors newly built (current candidates): it is calculated over the time-horizon only.
- **Operation and Maintenance Cost** for the "installed" capacity (candidates built before the considered time-horizon): it is calculated over the time-horizon only. The existing grid only adds a constant maintenance cost and is not considered in the objective function.
- **Expected OPEX** according to the new Capacity and Admittance: it is calculated over the representative snapshots for (s, h).
- **Rest Value of the investments** after the end of the study period (i.e. after H): it is calculated over the remaining Life Time and we consider that this amount of money is "refunded" in year $H + 1$.

<u>Discount Method</u>

Costs from one year to another are not strictly comparable since the value of the money changes with the inflation. In order to carefully compare costs, we need to add discount factors to the cost formulations previously explained to take this inflation into account.

To calculate the "Present Value" of a cost paid T years after the initial year, we apply the following discount factor:

$$DF(T) = \frac{1}{(1 + \tau)^T}$$

Here we consider that all expenses paid in time-horizon h have to be discounted from this "representative" year to the first year of the study period but the Rest Value is discounted from the year it is "refunded" (i.e. h_{step} years after the final time-horizons).

The following figure illustrates the discount method: we consider a study period from 2020 to 2050 and the time-horizon 2045 (5 years between each time-horizon). The cost function for this time horizon is computed from the **total investment cost** (CAPEX) of candidates installed during this time-horizon, the **annual maintenance cost** (OM) taking in to account the new lines installed from the first time-horizon to the considered time-horizon, the **annual operation cost** (OPEX) of the current grid (existing + all installed candidates) and finally the **rest value** (RV), if any, of the candidates installed during time-horizon 2045

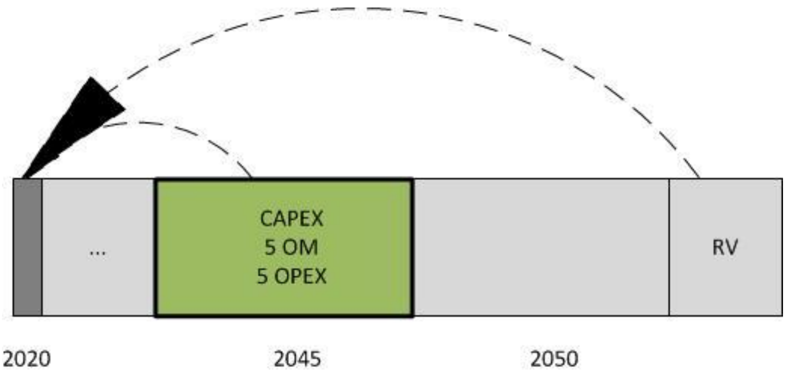

Figure 16 - COST discount method Example

For this time horizon the cost function would be written as follows:

$$DF(25) \cdot [CAPEX + 5 \cdot (OM + OPEX)] - DF(35) \cdot RV$$

We obtain the following formulation for the Investment Level Optimization, where the objective function is organized in the same order as the previous description of the 5 components of the cost.

$$\min_{n} \left\{ \sum_{s\,\in\,\mathcal{S}} w_s \sum_{h\,\in\,\mathcal{H}} \left[DF(h-y_0) \left[\sum_{k\in\mathcal{C}} n_k(s,h) \cdot Inv_{Type(k)}(s,h) \cdot Len_k \right.\right.\right.$$

$$+\, h_{step} \left(\sum_{\substack{j\in\mathcal{H}\\ j\leq h}} \sum_{k\in\mathcal{C}} n_k(s,h) \cdot OM_{Type(k)}(s,h) \cdot Len_k \right.$$

$$\left.\left. +\, 8760 \sum_{r\,\in\,\mathcal{R}(s,h)} w_r \cdot OPEX\big(Cap_{AC}^{s,h}, Cap_{DC}^{s,h}, Y^{s,h}, r\big) \right) \right]$$

$$-\, DF\big(H + h_{step} - y_0\big) \sum_{k\in\mathcal{C}} n_k(s,h) \cdot \frac{Inv_{Type(k)}(s,h)}{LT_{Type(k)}} \cdot Len_k$$

$$\left.\left. \cdot\, \left(h + LT_{Type(k)} - \big(H + h_{step}\big)\right)^{+} \right] \right\}$$

subject to:

$n_k(s,h)\ integer$
$Dev_k(s,h)\ binary$

$$\forall s \in [\![2..N_s]\!], \forall h \in CDH, \qquad n(s,h) = n(s-1,h)$$

(Common Development)

$$\forall s \in [\![1..N_s]\!], \forall h \in \mathcal{H}, \forall (k_1,k_2) \in CombCand, \qquad n_{k_1}(s,h) = n_{k_2}(s,h)$$

(Combined Candidates)

$$\forall s \in \mathcal{S}, \forall h \in \mathcal{H}, \forall k \in \mathcal{C}, \qquad 0 \leq n_k(s,h) \leq n_k^{Max} \cdot Dev_k(s,h)$$

(Investment increment)

$\forall s \in \mathcal{S}, \forall k \in \mathcal{C},$

$$\sum_{h \in \mathcal{H}} Dev_k(s,h) \leq 1$$

(Limited development)

$$\forall s \in \mathcal{S}, \forall h \in \mathcal{H}, \sum_{k \in \mathcal{C}} n_k(s,h) \cdot Inv_{Type(k)}(s,h) \cdot Len_k \leq INV_{max}(s,h)$$

(Investment maximum)

$\forall t \in \{AC, DC\}, \forall s \in \mathcal{S}, \forall h \in \mathcal{H}, \forall (z_A, z_B) \in Z \times Z,$

$$Cap_t^{s,h}(z_A,z_B) = Cap_t^{s,h-h_{step}}(z_A,z_B) + \sum_{\substack{k \in \mathcal{C} \\ Tech(k)=t \\ ZoneA(k)=z_A \\ ZoneB(k)=z_B}} n_k(s,h) \cdot Cap_k^{inc}$$

(Capacity Variation)

$\forall s \in \mathcal{S}, \forall h \in \mathcal{H}, \forall (z_A, z_B) \in Z \times Z,$

$$Y^{s,h}(z_A,z_B) = Y^{s,h-h_{step}}(z_A,z_B) + \sum_{\substack{k \in \mathcal{C} \\ Tech(k)=AC \\ ZoneA(k)=z_A \\ ZoneB(k)=z_B}} n_k(s,h) \cdot Y_k^{inc}$$

(Admittance Variation)

$\forall s \in \mathcal{S}, \forall (z_A, z_B) \in Z \times Z,$

$$Cap_{AC}^{s,y_0}(z_A, z_B) = Cap_{AC}(z_A, z_B)$$
$$Cap_{DC}^{s,y_0}(z_A, z_B) = Cap_{DC}(z_A, z_B)$$
$$Y^{s,y_0}(z_A, z_B) = Y(z_A, z_B)$$

(Initial Grid Configuration)

Problem 3 – Investment Level Optimization

Variable to Optimize

$n_k(s, h)$: number of installed units for candidate k in scenario s and time-horizon h

$Dev_k(s, h)$: binary variable which controls that a candidate is used only once in each scenario

Second Level Call

$OPEX$: Power Adjustment Costs

Remarks

- The snapshots $\mathcal{R}(s, h)$ are only representatives for the situation (s, h). We consider that this is the best estimation that we can have of the production and consumption behaviors in this situation
- In the **Problem 3** formulation there is no constraint on the number of times that a candidate can be invested in (if a candidate is available in several time-horizons, the model could lead to successive upgrades in the same corridor using this candidate)
- The Common Development constraint might lead to an unfeasible problem. An alternative formulation would be to optimize each scenario independently and then analyze the Common Development Horizons to define a-posteriori the common development (see Problem 3bis)

$$\forall s \in \mathcal{S} \ \min_{\boldsymbol{n}} \left\{ \sum_{h \in \mathcal{H}} \left[DF(h - y_0) \left[\sum_{k \in \mathcal{C}} n_k(s,h) \cdot Inv_{Type(k)}(s,h) \cdot Len_k \right.\right.\right.$$

$$+ h_{step} \left(\sum_{\substack{j \in \mathcal{H} \\ j \leq h}} \sum_{k \in \mathcal{C}} n_k(s,h) \cdot OM_{Type(k)}(s,h) \cdot Len_k \right.$$

$$\left.\left. + 8760 \sum_{r \in \mathcal{R}(s,h)} w_r \cdot OPEX\left(Cap_{AC}^{s,h}, Cap_{DC}^{s,h}, Y^{s,h}, r\right) \right) \right]$$

$$- DF\left(H + h_{step} - y_0\right) \sum_{k \in \mathcal{C}} n_k(s,h) \cdot \frac{Inv_{Type(k)}(s,h)}{LT_{Type(k)}} \cdot Len_k$$

$$\left.\left. \cdot \left(h + LT_{Type(k)} - \left(H + h_{step}\right)\right)^{+} \right]\right\}$$

subject to:

Same constraints as **Problem 2** except (Common Development)

Analysis:

$\forall h \in CDH, \boldsymbol{MaximalDvpt(h)} = MAX_{s \in \mathcal{S}}\{\boldsymbol{n(s,h)}\}$

Problem 3bis – Alternative Investment Level Optimization (Example)

MAX : Comparison operator "max" applied component-wise. Example: $MAX \left\{ \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\} = \begin{bmatrix} 3 \\ 5 \\ 2 \end{bmatrix}$

This maximal development is one example of an alternative investment level optimization. Other solutions could be proposed, such as minimal development using the "min" operator or average development using a weighted sum with the scenario weights w_s.

Operational level

This level is a DC Optimal Power Flow (DCOPF) applied on the current state of the grid $\left(Cap_{AC}^{s,h}, Cap_{DC}^{s,h}, Y^{s,h}\right)$ in a given snapshot r related to scenario s and time-horizon h. Starting from an ideal injection distribution (production and consumption in each zone drawn from a copperplate market simulation in the conditions of snapshot r), this model minimizes the injection deviations due to the grid constraints.

As proposed in [22], we consider both AC and DC links in our DCOPF. The flows on DC links are only constrained by the capacity of the links and the flow balance, while the flows on AC links also have to comply with the angle constraints on each end of the links.

$$OPEX\left(Cap_{AC}^{s,h}, Cap_{DC}^{s,h}, Y^{s,h}, r\right) = \min_{\theta,I,f}\left\{\sum_{z\in Z}\sum_{i\in InjTypes} MC_{+}^{r}(z,i)\cdot\Delta I_{+}^{r}(z,i) + MC_{-}^{r}(z,i)\cdot\Delta I_{-}^{r}(z,i)\right\}$$

subject to:

$\forall z \in Z,$

$$\sum_{z_A\in Z} f_{DC}^{r}(z_A,z) - \sum_{z_B\in Z} f_{DC}^{r}(z,z_B) + \sum_{z_A\in Z} f_{AC}^{r}(z_A,z) - \sum_{z_B\in Z} f_{AC}^{r}(z,z_B) + \sum_{i\in InjTypes} sg(i)\cdot I^{r}(z,i) = 0$$

(Power Balance)

$\forall (z_A,z_B) \in Z\times Z,$

$$-Cap_{AC}^{s,h}(z_A,z_B) \le f_{AC}^{r}(z_A,z_B) \le Cap_{AC}^{s,h}(z_A,z_B)$$

(AC Flow Capacity)

$\forall (z_A,z_B) \in Z\times Z,$

$$-Cap_{DC}^{s,h}(z_A,z_B) \le f_{DC}^{r}(z_A,z_B) \le Cap_{DC}^{s,h}(z_A,z_B)$$

(AC Flow Capacity)

$\forall (z_A,z_B) \in Z\times Z,$

$$f_{AC}^{r}(z_A,z_B) = S_n\cdot\left(Y^{s,h}(z_A,z_B)\cdot\left[\theta_{z_A}^{r} - \theta_{z_B}^{r}\right] + \sum_{\substack{c\ in\ Corr_{PST}\\ ZoneA(c)=z_A\\ ZoneB(c)=z_B}} Y_c\cdot\alpha^{r}(c)\right)$$

(AC Flow Definition)

$\forall c \in Corr_{AC},$

$$\alpha_{min}^{r}(c) \le \alpha^{r}(c) \le \alpha_{max}^{r}(c)$$

(Phase Shifter Boundaries)

$\forall z \in Z, \forall i \in InjTypes,$

$$I^r(z,i) - I_0^r(z,i) = \Delta I_+^r(z,i) - \Delta I_-^r(z,i)$$

(Injection Variations)

$\forall z \in Z, \forall i \in InjTypes,$

$$I_{min}^r(z,i) \leq I^r(z,i) \leq I_{max}^r(z,i)$$

(Injection Boundaries)

$\forall z \in SwBus,$

$$\theta_z^r = 0$$

(Swing Buses Angle)

Problem 4 – Operational Level Optimization

Variables to optimize

I : Injection array describing Power Injections for each type and each zone (Prod >0 & Cons <0)

$\Delta I_{+/-}$: Variations of the injections

θ : Phase Angle in each zone

$f_{AC}^r(z_A, z_B)$: Power flow between two given zones

Remarks

- Losses have been neglected for simplicity and consistency but Quadratic Losses could be estimated with a Linear Loss Rate in future developments.

Resolution Method: Mixed Integer Linear Problem

Combining the Investment and Operational Levels

In Section 4.1 we described our problem as a bi-level optimization were the economical solution (investment level) would constraint the technical solution (operational level). Even though it makes sense to separate economical optimization from technical optimization, our problem is simpler than the general bi-level formulation.

The following problem is the typical formulation of a bi-level optimization problem [23].

$$
\begin{aligned}
&\min_{x,y} C_1 \cdot x + D_1 \cdot y \\
&\qquad \mathbf{s.t.}\ A_1 \cdot x \ + \ B_1 \cdot y \leq K_1 \\
&\qquad\qquad x \geq 0 \\
&\qquad\qquad y \in arg\min_{y'} C_2 \cdot x + D_2 \cdot y' \\
&\qquad\qquad\qquad \mathbf{s.t.}\ A_2 \cdot x + B_2 \cdot y' \leq K_2 \\
&\qquad\qquad\qquad\qquad y' \geq 0
\end{aligned}
$$

Problem 5 – General Formulation of a bi-level optimization problem

The resolution of such problems is complex and usually involves the use of the dual problem of the second level into the first level. However, the problem we are trying to solve has the following form where $C \cdot x$ represents the investment (and maintenance) and $D \cdot y$ represents the operational cost:

$$
\begin{aligned}
&\min_{x,y} C \cdot x + D \cdot y \\
&\qquad \mathbf{s.t.}\ A_1 \cdot x \leq K_1 \\
&\qquad\qquad x \geq 0 \\
&\qquad\qquad y \in arg\min_{y'} D \cdot y' \\
&\qquad\qquad\qquad \mathbf{s.t.}\ A_2 \cdot x + B_2 \cdot y' \leq K_2 \\
&\qquad\qquad\qquad\qquad y' \geq 0
\end{aligned}
$$

Problem 6 – General Formulation of the TEP optimization problem

The main differences between Problem 4 and Problem 5 are:

- The "regular" constraints in the first level do not involve variable y
- The objective function in the second level only involves variable y, which has the same coefficient than in the first level (D).

Those differences allow us to solve Problem 5 more easily than Problem 4. It is indeed possible to prove that Problem 5 is actually equivalent to the following simple problem:

$$\min_{x,y} C \cdot x + D \cdot y$$

$$\begin{aligned} \textbf{s.t.}\ & A_1 \cdot x \leq K_1 \\ & A_2 \cdot x + B_2 \cdot y \leq K_2 \\ & x \geq 0 \\ & y \geq 0 \end{aligned}$$

Problem 7 – Simple Formulation of the TEP optimization problem

Then, the two levels of the TEP optimization can be merged together to formulate an equivalent problem which can solve both levels at once.

Relieving the non-linearity

When combining the two levels of the optimization problem and expressing the current admittance with the formulation of the Admittance Variation, the following product of variables n_k and θ, α appears in the AC Flow Definition constraint for a given scenario s, time-horizon h and candidate k:

$$n_k(s,h) \cdot Y_k^{inc} \cdot \left(\theta_{z_A}^r - \theta_{z_B}^r\right)$$

This results in a non-linear problem where the first variable involved (n_k) is integer whereas the other (θ) is continuous.

To relieve this non-linearity we could test each possible value of the integer variable and optimize the continuous one for each of those possibilities. However it is not easy to mirror this idea only with new constraints. A solution – the "standard disjunctive model" – has been provided for the product of a binary variable with a continuous one, as proposed in [13] where binary investments were optimized. We reproduce here this solution with simplified notations:

Let $x \in [a,b]$ be a continuous variable and β be a binary variable. With f, g and h three linear functions and $M \in \mathbb{R}$ a constant such that $[a,b] \subset [-M,M]$, the following two optimization problems are equivalent:

$$P_1 \begin{cases} \min\limits_{x,\beta} f(x,\beta) \\ g(x,\beta) = 0 \\ h(x,\beta,x \cdot \beta) = 0 \end{cases} \qquad P_2 \begin{cases} \min\limits_{\tilde{x},x,\beta} f(x,\beta) \\ g(x,\beta) = 0 \\ h(x,\beta,\tilde{x}) = 0 \\ -(1-\beta) \cdot M \leq x - \tilde{x} \leq (1-\beta) \cdot M \\ -\beta \cdot M \leq \tilde{x} \leq \beta \cdot M \end{cases}$$

Proposition 1 – Standard Disjunctive Model

If we consider an integer variable $\delta \in [\![0, N]\!]$ as the sum of N binary variables β_i we obtain N non-linear products, which can be replaced in the same manner by the previous model. We introduce N new continuous variables $\tilde{x}_i$ to mirror the use of $\tilde{x}$ in the previous proposition. The following extended disjunctive model is obtained:

Let $x \in [a, b]$ be a continuous variable and δ be an integer variable ($\delta \in [\![0, N]\!]$). With f, g and h three linear functions and $M \in \mathbb{R}$ a constant such that $[a, b] \subset [-M, M]$, the following two optimization problems are equivalent:

$$P_1 \begin{cases} \min\limits_{x,\delta} f(x, \delta) \\ g(x, \delta) = 0 \\ h(x, \delta, x \cdot \delta) = 0 \end{cases} \qquad P_2 \begin{cases} \min\limits_{\tilde{x}_1 .. \tilde{x}_N, x, \delta} f(x, \delta) \\ g(x, \delta) = 0 \\ h(x, \delta, \tilde{x}) = 0 \\ \forall i \in [\![1, N]\!], \quad -(1 - \beta_i) \cdot M \leq x - \tilde{x}_i \leq (1 - \beta_i) \cdot M \\ \forall i \in [\![1, N]\!], \quad -\beta_i \cdot M \leq \tilde{x}_i \leq \beta_i \cdot M \\ \delta = \sum\limits_{i=1}^{N} \beta_i \end{cases}$$

Proposition 2 – Extended Disjunctive Model

This extended disjunctive model can be applied to the non-linear product occurring when the two-levels are merged. The obtained linear problem is formulated in Section 4.2.