An Operational State Aggregation Technique for Transmission Expansion Planning Based on Line Benefits

Quentin Ploussard, Luis Olmos, *Member, IEEE*, and Andrés Ramos

***Abstract*—This paper provides a novel technique to represent in a reduced, or compact, way temporal variability in transmission expansion planning (TEP). This reduction is handled by means of "snapshot selection." Instead of taking into account all the possible operational states and their associated optimal power flow, a reduced group of them is selected that is representative of all the states that should have an influence on investment decisions. Considering this reduced group of operational states should lead to the same investment decisions as if all the snapshots in the target year were considered. Original operational states are compared in the space of the benefits produced by potential reinforcements considered, which are relevant drivers for investment decisions. The benefits produced by these potential reinforcements are computed based on the incremental change in operation costs resulting from their installation. A clustering algorithm is used to group together those operational states where similar line benefits are realized. Our algorithm has been tested on a modified version of the standard IEEE 24 bus system. The method produces promising results and proves to be more efficient than other snapshot selection methods used until now in computing an accurate enough selection of snapshots representing system operation variability in TEP.**

***Index Terms*—Clustering, dimension reduction, integer linear programming, transmission expansion planning.**

NOMENCLATURE

Indices & Sets

i, j	Node
c, cc, ce	Existing or candidate circuit
s	Snapshot
g	Generator
k	Target number of clusters
CC	Candidate circuits
EC	Existing circuits
CONV	Conventional generators
RES	Renewable Energy Source (RES) generators
N_B	Total number of nodes
N_S	Total number of snapshots
N_G	Total number of generators ($[1;N_G] = CONV \cup RES$)
N_{CC}	Total number of candidate circuits ($N_{\mathrm{CC}} = card(CC)$)

Parameters

(i,j)	Existing or candidate circuits between nodes *i* and *j*
$\Omega_i^{\mathrm{start}}$	Existing or candidate circuits for which *i* is a start node
Ω_i^{end}	Existing or candidate circuits for which *i* is an end node
Ω_i^G	Conventional or RES generators connected at node *i*
Y_c	Admittance of circuit *c* (p.u.)
cg_g	Generation cost of generator *g* (€/MWh)
$d_{i,s}$	Demand level at node *i* during snapshot *s* (MW)
$f_c^{\max}$	Maximum capacity of a circuit *c* (MW)
$p_{g,s}^{\max}$	Generation capacity of RES generator *g* during snapshot *s* (MW)
$p_g^{\min}$	Minimum generation capacity of conventional generator *g* (MW)
$p_g^{\max}$	Maximum generation capacity of conventional generator *g* (MW)
$av_{g,s}$	Availability state of a conventional generator *g* during snapshot *s* {0;1}
ρ_s	Weight of snapshot *s* (h)
C^{ENS}	Cost of energy non-served (€/MWh)
C_c	Investment cost of candidate circuit *c* (€)
S_B	Base power (MW)
γ	Fix annual charge rate (p.u.)
M_c	Big-M parameter of candidate circuit *c* (p.u.)

Variables

x_c	Decision to invest or not in candidate circuit *c* {0;1}
$p_{g,s}$	Power production of generator *g* during snapshot *s* (MW)
$pns_{i,s}$	Power non-served at node *i* during snapshot *s* (MW)
$\theta_{i,s}$	Voltage angle at node *i* during snapshot *s* (p.u.)
$f_{c,s}$	Power flow through circuit *c* during snapshot *s* (MW)

Manuscript received February 12, 2016; revised June 8, 2016 and September 21, 2016; accepted September 24, 2016. Date of publication October 4, 2016; date of current version June 16, 2017. The work of Q. Ploussard was supported through an Erasmus Mundus Ph.D. Fellowship. Paper no. TPWRS-00232-2016.

The authors are with the Technological Research Institute, Comillas Pontifical University, Madrid 28015, Spain (e-mail: quentin.ploussard@iit.comillas.edu; luis.olmos@iit.comillas.edu; andres.ramos@comillas.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2016.2614368

I. INTRODUCTION

THE aim of Transmission Expansion Planning (TEP) studies is to determine which, where, and when new lines should be constructed at the minimum total cost. The high temporal variability of demand and intermittent generation (wind, solar) increases the number of operational states to consider and, therefore, the complexity of an already intricate problem. This raises the need for methods able to reduce the size of the problem while providing a similar solution to the one that would be obtained considering the full size problem.

Techniques to express the diversity of operational states in a compact way are used to reduce the size of this problem and make TEP resolution tractable. This article provides a novel technique to express the diversity of operation situations in a compact way when conducting static TEP. Static TEP deals with investment decisions that have to be implemented by the target year (e.g. the year 2030). “Snapshot selection” is here used to represent in a compact way the temporal variability of system operation. This involves clustering the 8760 operation hours that comprise the one-year period of analysis for TEP into a reduced number of groups of snapshots that are considered independent. When clustering the operational states, they are divided into two groups: those states (or operation situations) that are most relevant for TEP purposes and the rest of the states. In order to cluster the most relevant states into groups, they are compared in the space of the benefits produced by potential network reinforcements. Less relevant operational states are clustered in the space of operation cost savings.

In the context of this article, the benefit produced by a potential network reinforcement, also called benefit of this new line (line benefit), is defined as the contribution made by this individual potential network reinforcement to the global operation cost savings brought about by the whole potential network investment plan. Network reinforcements, whose benefits are considered to cluster operational states, are termed “potential” because they are computed as continuous investments and, therefore, may not fully correspond to the optimal discrete ones. Network reinforcements are computed as continuous investments because the 8760 hourly operation situations of the target year of analysis need to be considered when computing them. This is so because, being used as clustering variables, the benefits produced by these reinforcements need to be computed before clustering operation situations into groups of them. Computing the discrete network expansion plan considering the 8760 hourly operation situations of the target year is not possible, since the resulting TEP problem is not tractable (it is because this problem is not tractable that we are clustering operation situations into groups of them).

After applying the clustering procedure, one representative of each group of snapshots is selected. The set of representative snapshots obtained represents in a compact way the initial whole set of operation situations in the context of the TEP problem.

The rest of the paper is arranged as follows: Section II presents a short review of snapshot reduction methods used in the past and the main contributions of our work. Section III presents the mathematical formulation of the TEP problem. Section IV introduces the proposed methodology of snapshot selection. Two case studies are presented in Section V to analyze and validate the proposed methodology using the standard IEEE 24 bus Reliability Test System. Section VI concludes.

II. Literature Review and Main Contribution From Our Work

TEP has been widely explored in the last decades. However, this paper focuses only on the literature related to snapshot reduction techniques applied to TEP.

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

The oldest approach to snapshot selection known [1] involves the consideration of a few “load blocks,” that is, clustering the operation situations according to the global load level of the system in them. Most of the studies on snapshot selection which have been conducted in the following years are based on variations of this idea.

According to a very recent study [2] within the e-HIGHWAY 2050 project, the snapshots to be considered should be defined by taking into account not only the global load level but also other operational and economic variables related to the power system operation, namely nodal prices and intermittent generation outputs. Then, operation situations are classified into clusters according to these variables by using a K-means algorithm [3]. Authors in [4] go a step further by grouping together operation situations where there is a similar pattern of network congestion, which is considered a more relevant driver of TEP investment decisions. Lastly, authors in [5] recommend clustering snapshots while preserving the chronological link among operation situations (reproducing the features of the Unit Commitment problem). In [5], part of the chronological information related to system operation remains available through the use of a transition matrix while reducing the size of the problem.

No previous work on the clustering of operational states in a TEP context considers new line benefits as clustering variables. The temporal variability and scenario reduction techniques already available in the literature for the selection of snapshots either consider clustering variables that are related to the causes of network reinforcements (like the increase in the nodal load level), or they consider an incomplete, or less relevant, subset of the effects in the system of network reinforcements. We believe that the benefits to be produced by network investments represent the most relevant drivers of investment decisions in the context of the TEP problem, since the decision to undertake a new transmission expansion project directly depends on the operation cost reduction (i.e. the kind of benefits considered in the analysis presented here) the corresponding new assets would bring about if installed in the system. It is an estimate of these economic benefits what drives, together with reinforcement costs, the selection for their construction of some specific reinforcements instead of others. By finding the most representative snapshots according to the operation costs reduction achieved in them through the expansion of the network, the proposed approach for the selection of snapshots is based on the effects, or results, caused by network investments, or outputs of the TEP problem, instead of the primary causes of these investments, or inputs to the TEP problem, i.e. the changes in the pattern of demand and generation in the system creating new network reinforcement needs. Here, a novel snapshot selection method for TEP is proposed based on the consideration of the similarities and differences that exist among operation situations in terms of the benefits produced by candidate reinforcements in them.

Main contributions of our work are listed next:

1. Definition of a novel method for the selection of the snapshots to be considered in the transmission expansion planning (TEP) problem. Novel features of this method that make it suitable to the problem ta hand are discussed next.
2. Application of this method to a relevant case study (based on the IEEE-24 RTS) to show its implementability.
3. Comparison of the method proposed to others that can potentially be applied to this same problem showing the weak and strong points of each.

III. Formulation of the TEP Problem

The static optimization TEP problem can be expressed as follows:

$$\min \left(\sum_{s=1}^{N_S} \rho_s \left(\sum_{g=1}^{N_G} p_{g,s} cg_g \right) + \sum_{s=1}^{N_S} \rho_s \left(\sum_{i=1}^{N_B} pns_{i,s} C^{\mathrm{ENS}} \right) + \gamma \left(\sum_{c \in CC} x_c C_c \right) \right) \quad (1)$$

Subject to:

$$\sum_{g \in \Omega_i^G} p_{g,s} - d_{i,s} + pns_{i,s} + \sum_{c \in \Omega_i^{\mathrm{end}}} f_{c,s} - \sum_{c \in \Omega_i^{\mathrm{start}}} f_{c,s} = 0;\ \forall i \in [\![1; N_B]\!],\ \forall s \in [\![1; N_S]\!] \quad (2)$$

$$f_{c,s} = S_B Y_c \left(\theta_{i,s} - \theta_{j,s} \right);\ \forall c \in (i,j) \cap EC,\ \forall s \in [\![1; N_S]\!] \quad (3)$$

$$\left| f_{c,s} - S_B Y_c \left(\theta_{i,s} - \theta_{j,s} \right) \right| \le M_c \left(1 - x_c \right);\ \forall c \in (i,j) \cap CC,\ \forall s \in [\![1; N_S]\!] \quad (4)$$

$$|f_{c,s}| \le f_c^{\max};\ \forall c \in (i,j) \cap EC,\ \forall s \in [\![1; N_S]\!] \quad (5)$$

$$|f_{c,s}| \le x_c f_c^{\max};\ \forall c \in (i,j) \cap CC,\ \forall s \in [\![1; N_S]\!] \quad (6)$$

$$0 \le pns_{i,s} \le d_{i,s};\ \ \forall i \in [\![1; N_B]\!],\ \forall s \in [\![1; N_S]\!] \quad (7)$$

$$0 \le p_{g,s} \le av_{g,s} p_g^{\max};\ \forall g \in CONV,\ \forall s \in [\![1; N_S]\!] \quad (8)$$

$$0 \le p_{g,s} \le p_{g,s}^{\max};\ \forall g \in RES,\ \forall s \in [\![1; N_S]\!] \quad (9)$$

The temporal variability of system operation conditions is assumed to be captured by the diversity of operation situations taking place in the time period of analysis (the target year). Power flows through the lines are computed according to the DC load flow model. This model does not consider losses in lines. However, the proposed method can directly account for them if included in the optimization problem.

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

The final term in (1) is related to the network investment costs. In order to obtain an annualized cost that can be directly compared to operational ones, a fixed charge (annualization) rate γ is applied to these investment costs.

Equation (2) represents the power balance at each node. This includes the power produced by a virtual generation plant representing non-served power.

Equations (3) and (4) correspond to the DC-load flow model used to represent the flow of power through existing and candidate circuits, respectively. Equation (4) is a disjunctive constraint formulated in such a way that it satisfies MIP programming conditions. This constraint imposes the condition that the relationship between voltage angles at the two extreme nodes of a candidate line and the flow in the line is the same as for existing lines when this candidate line is installed. On the contrary, assuming M_c big enough, it does not constrain in any sense voltage angles at the extreme nodes when the line is not built. Authors in [6] provide a method to set an appropriate value for this parameter.

Equations (5) and (6) refer to the maximum power flow that can go through existing and candidate circuits, respectively. Equation (7) limits non-served power in each node to be positive and always lower than the actual demand level in this node. Equations (8) and (9) represent the power generation limits of conventional and RES generators, respectively. The main difference between the two is the time-dependency of RES generators capacity.

Temporal variability sources have been integrated into the model representing them by hourly parameters that vary over the year (across snapshots). These include the availability state of conventional generators $av_{g,s}$ in equation (8) and the maximum output of RES generators $p_{g,s}^{\max}$ in equation (9). Load level hourly variations throughout the planning horizon have also been accounted for via the parameter $d_{i,s}$ in equation (2).

IV. Methodology

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

This methodology is divided into the steps that follow: A) Running a relaxed version of the TEP problem to compute a proxy to the optimal set of network reinforcements; B) Computation of the hourly benefits produced by each of these reinforcements; C) Reduction of the dimensionality of the space of hourly benefits produced by reinforcements; D) Clustering of operation situations according to the compact representation of the benefits that reinforcements produce in them. A flowchart of the methodology applied is shown in Fig. 1.

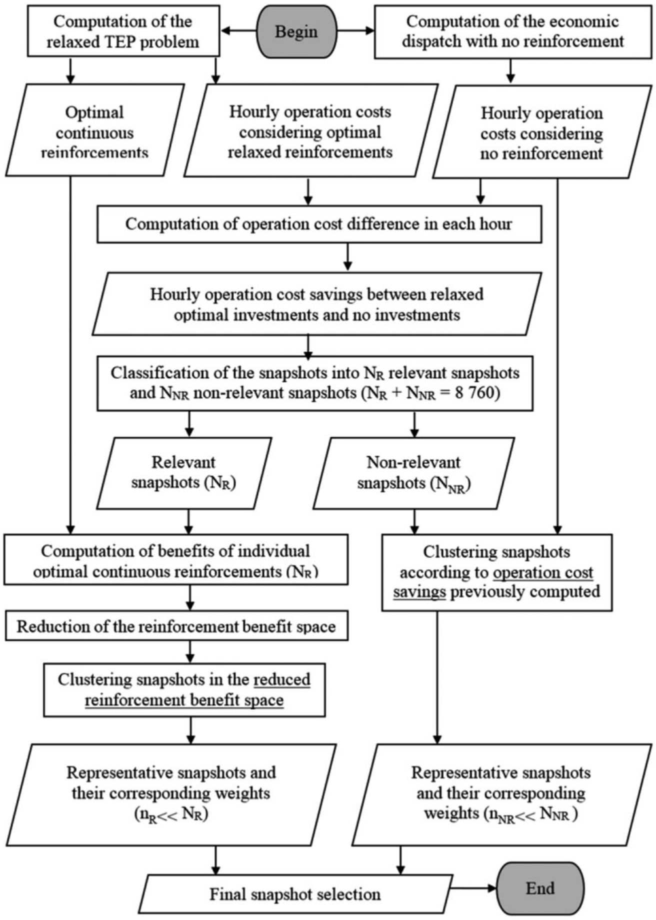

Fig. 1. Flowchart of the methodology. First, operation situations are classified into relevant and non-relevant ones according to the potential operation cost savings that can be realized in them by reinforcing the network with optimal continuous reinforcements. Then, relevant snapshots are clustered according to line benefits, whereas non-relevant ones are clustered according to operation cost savings. Figures N_x in brackets represent the numbers of snapshots considered in each step.

A. Relaxed TEP problem and related optimal investments

Computing the benefits produced in each operation situation by all the candidate lines provided a priori by the network planner may probably not be representative enough of the actual benefits to be obtained from network expansion in each of these situations. Some candidate lines may produce too small benefits compared to their cost, while others may even have a negative impact on the overall benefit of the expansion plan [7]. Thus, computing the benefits produced by all candidate lines may result in a poor estimate of the benefits produced by the real expansion plan. Besides, the size of the reinforcement benefit space considering the whole set of candidate lines to be built may be huge, thus making the clustering of snapshots in this space more difficult and inaccurate.

Ideally, we should compute the benefits produced by the optimal set of reinforcements (candidate lines selected to be built), i.e. the one minimizing the total annualized network investments plus operation costs of the system throughout the target year. However, this would require solving the original TEP problem considering all hourly snapshots in this year, which is exactly

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

Characteristics of the circuit c	Non-relaxed TEP model	Relaxed TEP model	Equations related to this relaxation
Quantity built	0 or 1	$0 \leq x_c \leq 1$	
Investment cost	C_c	$x_c\ C_c$	(1)
Capacity	$f_c^{\max}$	$x_c\ f_c^{\max}$	(6)
Admittance	Y_c	$x_c Y_c$	(17), (18)

what we wish to avoid. For that reason, we instead consider the benefits produced in each hourly snapshot by each reinforcement within a set of most promising ones, which we compute by solving a relaxed version of the TEP problem.

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

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

This relaxation relies on the assumption that candidate circuits can be "partially built", and that the capacity, admittance, and investment cost of these partially built circuits are proportional to the quantity of line built. These assumptions, along with the equations they are related to, are summarized in Table I.

1) Relaxation of the 2nd Kirchhoff law for candidate circuits: The admittance of the "partially built" candidate circuit would be exactly proportional to the quantity built if, in the relaxed problem, we could include equation (10).

$$f_{c,s} = x_c S_B Y_c \left(\theta_{i,s} - \theta_{j,s}\right); \ \forall c \in (i,j) \cap CC, \ \forall s \in [\![1; N_S]\!] \tag{10}$$

Equation (10) represents the 2nd Kirchhoff law for partially built candidate circuits (with an admittance $x_c\, Y_c$). However, due to the fact that it includes a product of linear variables x_c and $(\theta_{i,s} - \theta_{j,s})$, this equation is non-linear. Thus, including this equation in the relaxed TEP model would prevent us from using LP programming to solve it. For that reason, instead of enforcing equation (10), the product $x_c(\theta_{i,s} - \theta_{j,s})$ is represented by an envelope of it for the ranges of variables x_c and $(\theta_{i,s} - \theta_{j,s})$ to be considered in the TEP problem, while the value of the term $f_{c,s}/(S_B Y_c)$ is enforced to lie within this envelope. The McCormick envelope [8] of a product of variables xy for which the variable x is bounded between $\underline{x}$ and $\bar{x}$, and the variable y is bounded between $\underline{y}$ and $\bar{y}$, is represented by equations (11) to (14). This envelope is depicted in Fig. 2 for the case where variable x is assumed to be bounded between $1/2$ and 3 and

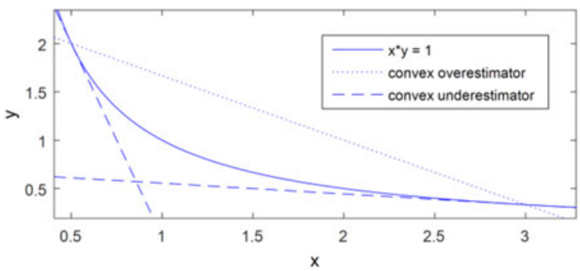

Fig. 2. McCormick envelope of bilinear variable *xy*, when variable *x* is assumed to be bounded between $1/2$ and 3 and variable *y* is assumed to be bounded between $1/3$ and 2.

variable y is assumed to be bounded between $1/3$ and 2.

$$xy \leq \bar{y}x + \underline{x}y - \underline{x}\bar{y} \tag{11}$$

$$xy \leq \bar{x}y + \underline{y}x - \underline{y}\bar{x} \tag{12}$$

$$xy \geq \underline{y}x + \underline{x}y - \underline{x}\underline{y} \tag{13}$$

$$xy \geq \bar{y}x + \bar{x}y - \bar{x}\bar{y} \tag{14}$$

We shall assume that the nodes *i* and *j* to be connected by each candidate circuit *cc* (with a capacity $f_{\mathrm{cc}}^{\max}$ and an admittance Y_{cc}) are already connected by an existing circuit *ce* (with a capacity $f_{\mathrm{ce}}^{\max}$ and an admittance Y_{ce}). Then, according to equations (3) and (5), $(\theta_{i,s} - \theta_{j,s})$ is bounded between $\pm f_{\mathrm{ce}}^{\max}/(Y_{\mathrm{ce}} S_B)$. Besides, x_{cc} is bounded between 0 and 1. Consequently, according to equations (11) to (14), the McCormick envelope of the product $x_{\mathrm{cc}}(\theta_{i,s} - \theta_{j,s})$ is represented by equations (15) and (16).

$$|x_{\mathrm{cc}}\left(\theta_{i,s} - \theta_{j,s}\right)| \leq x_{\mathrm{cc}} \frac{f_{\mathrm{ce}}^{\max}}{Y_{\mathrm{ce}} S_B} \tag{15}$$

$$|x_{\mathrm{cc}}\left(\theta_{i,s} - \theta_{j,s}\right) - \left(\theta_{i,s} - \theta_{j,s}\right)| \leq \frac{f_{\mathrm{ce}}^{\max}}{Y_{\mathrm{ce}} S_B}\left(1 - x_{\mathrm{cc}}\right) \tag{16}$$

Finally, considering a relaxed version of equation (10), for which the term $x_{\mathrm{cc}}(\theta_{i,s} - \theta_{j,s})$ is represented by its envelope as in equations (15) and (16), results in constraints (17) and (18).

$$|f_{\mathrm{cc,s}}| \leq x_{\mathrm{cc}} \frac{f_{\mathrm{ce}}^{\max} Y_{\mathrm{cc}}}{Y_{\mathrm{ce}}} \tag{17}$$

$$|f_{\mathrm{cc,s}} - S_B Y_{\mathrm{cc}}\left(\theta_{i,s} - \theta_{j,s}\right)| \leq \frac{f_{\mathrm{ce}}^{\max} Y_{\mathrm{cc}}}{Y_{\mathrm{ce}}}\left(1 - x_{\mathrm{cc}}\right) \tag{18}$$

Thus, in the relaxed version of the TEP problem provided in Section III, equations (17) and (18) should replace the disjunctive constraint (4).

Even if there is not an existing circuit directly connecting the two nodes that a candidate circuit is to lie in between, the angular difference between these nodes can be deemed not to be larger than the sum of the angular differences along an alternative, already existing, electrical path in the network between these nodes [6]. Then, the McCormick envelope can also be used to enforce a relaxed version of 2$^{\text{nd}}$ Kirchhoff law for candidate circuits that may be partially built.

Network flows computed using the McCormick envelope in the case study discussed below have shown to be closely related to those resulting from enforcing the 2$^{\text{nd}}$ Kirchhoff law once network investments have been decided. Then, using the McCormick envelope has proven to introduce a very low distortion in the solution of the problem and can, therefore, be regarded as a reasonable relaxation of the optimal continuous reinforcement problem (the problem expressed by equations (1) to (9) considering continuous investment decision variables).

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

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

B. *Computation of line benefits for each operation situation*

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

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

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

Several approaches have been proposed in the literature [9] to compute the benefits of a specific line within an expansion plan. One first approach involves computing the benefit of installing a line as the reduction in operation costs caused by the installation of this line when no other reinforcement in the plan has yet been undertaken (PINT approach, for "Put IN one at a Time"), see equation (19). Another, more recent, approach involves computing the benefit of installing a line as the benefits produced by its installation when all the other reinforcements in the plan have already been undertaken (TOOT approach, for "Take Out One at a Time"), see equation (20).

$$B_{s,c}^{\text{PINT}} = Op.\ costs\ in\ s\ when\ no\ line\ is\ installed$$
$$-\ Op.\ costs\ in\ s\ when\ only\ line\ c\ is\ installed \qquad (19)$$
$$B_{s,c}^{\text{TOOT}} = Op.\ costs\ in\ s\ when\ only\ line\ c\ is\ not\ installed$$
$$-\ Op.\ costs\ in\ s\ when\ all\ lines\ are\ installed \qquad (20)$$

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

Because of this, it seems appropriate to combine the benefits assigned to each individual reinforcement by the TOOT and PINT approaches in order to compute a more accurate estimate of the benefits produced by this reinforcement. This implicitly involves assuming that benefits to be allocated to each reinforcement include both part of the benefits it brings when it is the first one to be installed and part of the incremental benefits resulting from its installation in the last place within the expansion plan. Results obtained using this approach have proven to be reasonably accurate. Thus, the benefits here allocated to the individual network reinforcements comprising the expansion plan computed by solving the relaxed TEP model are a weighted average of the benefits allocated to them in the PINT and the TOOT approaches, see equation (21). In the case examples considered, the coefficient α in (21) is given a value of 0.5. This is so because, generally speaking, undertaking a specific reinforcement within the plan in the first place is neither more probable nor less probable than undertaking it is the last place. Actually, in most cases, this reinforcement should be undertaken in between some other reinforcements within the plan. But considering all the possible orders of deployment of reinforcements within a large expansion plan is not computationally feasible.

$$B_{s,c} = \alpha B_{s,c}^{\text{TOOT}} + (1-\alpha)\, B_{s,c}^{\text{PINT}} \qquad (21)$$

More sophisticated approaches are proposed by authors in [9], but they are more difficult to implement and computationally burdensome. Due to the fact that the proposed scheme for the computation of the benefits produced by each new line should produce reasonable results, and in order not to increase substantially the computational burden of the approach proposed here for the selection of snapshots, the application of sophisticated methods for the computation of the benefits of individual reinforcements has been discarded.

Operation costs considered when computing the benefits of reinforcements correspond to the sum of variable production costs and the cost of non-served energy from equation (1). We can thus represent operation situations within the space of reinforcement benefits. In this space, benefits in each axis correspond to those obtained in the corresponding operation situation from the undertaking of a specific reinforcement within the expansion plan. Each operation situation is represented by a point in this space.

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

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

C. *Dimension reduction of the line benefit space*

The sparsity of the set of network reinforcement benefits across lines and snapshots, and the relevant correlation factors probably existing among the benefits produced by several reinforcements, advices identifying the main underlying sources of the variability existing in the benefits of reinforcements across snapshots, and representing the variability in these benefits across snapshots in a more compact way, as the authors in reference [10] explain. In high dimensional spaces, distances among snapshots become relatively uniform. Then, distances defined become meaningless, which makes clustering more difficult. This raises the need for efficiently reducing the dimension of the benefit space, while retaining the major part of information available about the variability of reinforcement benefits across operation situations.

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

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

D. Clustering algorithm

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

The K-means clustering problem is NP-difficult and computationally hard to solve. Therefore, heuristic algorithms are preferred over classical optimization methods to solve this problem. Most heuristic clustering algorithms start from an initial set of K samples, defined as the initial set of K centroids, and iteratively try to increase the goodness of the clustering of operation situations by modifying the set of centroids. The choice of the initial set of centroids is of major concern [11]. Improving the initial choice of centroids leads to computing a good partition of snapshots in fewer iterations and avoiding trivial partitions (those where some clusters contain only one snapshot). The K-means++ algorithm addresses this concern by increasing the spread of the initial set of centroids. A first centroid is chosen randomly, considering a uniform distribution for the probability of choosing any snapshot as the first centroid. Then, each subsequent centroid is chosen among the remaining snapshots when each of these snapshots is assigned a probability of being chosen proportional to the square distance in the reduced benefit space from this snapshot to the closest snapshot already chosen as a centroid. This process ends when the initial set of K centroids are chosen. Authors in [11] prove the efficiency of such an initial choice of centroids.

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

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

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

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

V. Case Study

The proposed method for the selection of snapshots to be considered in TEP has been applied to select the most relevant ones for a case study based on the standard IEEE 24-bus Reliability Test System (RTS) [12]. The original data of this power system can be found at [13]. Based on the RTS system, analyses have been conducted considering hourly time series for main system variables. Two case studies have been built in this way, which differ in the level of demand. Case A is deemed the base one, while case B considers 5% of extra demand in each node.

A. Systems description

Most of the system features are the ones of the standard IEEE-24 RTS. However, transmission losses are neglected. Besides, only variable production costs are considered for generation units, and the minimum output for all units is set to zero. For each existing line in this system, an additional candidate reinforcement with the same features is considered. Hence, the number of candidate lines considered is 34. The investment costs of the candidate lines are the ones considered in [14]. The

TABLE II
OPTIMAL CONTINUOUS REINFORCEMENTS FROM RELAXED TEP MODEL

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

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

annualized factor, or rate, applied to line investment costs to compute annualized ones is γ equal to 7.94% for all candidate lines. According to data in [4], both in case A and case B, a solar farm with a maximum output of 500 MW and two wind farms with maximum outputs of 1 500 MW and 1 000 MW have been included in buses 4, 13 and 22, respectively. The time series for the availability state of generators has been produced as a random binary one (1 representing the corresponding unit is available and 0 representing it is not) where the average value (across all snapshots) has been constrained to lie between 0.85 and 0.95 depending on the generator technology. A one-year scope, hourly detailed, load profile has been produced for each bus. The load profiles and the RES production output profiles have been generated based on real, hourly-detailed, load and RES output profiles from the European power system which have been scaled to be well adapted to the load level of the IEEE 24 bus system. Its average value in case A is equal to the demand level defined for this bus in the original RTS power system. The average load level in case B is 5% higher. The level of correlation between the load level in two nodes ranges from 0.55 to 0.96. This results in a large enough variety of operation situations.

B. Application of the methodology

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

First, the operation cost savings produced in each snapshot by reinforcements computed in the relaxed TEP model are determined. Relevant snapshots have been determined as those for which total operation cost savings from reinforcements add up to 90% of the total annual savings achieved. (Fig. 3.) According to operation cost savings computed, there are 1350 relevant snapshots in case A, and 1050 relevant snapshots in case B, out of the 8760 hourly snapshots that exist in a whole year. The most relevant reason why increasing the demand level may result in having fewer relevant snapshots has to do with the way we have defined relevant and non-relevant snapshots. Defining an absolute threshold value of operation cost savings to identify

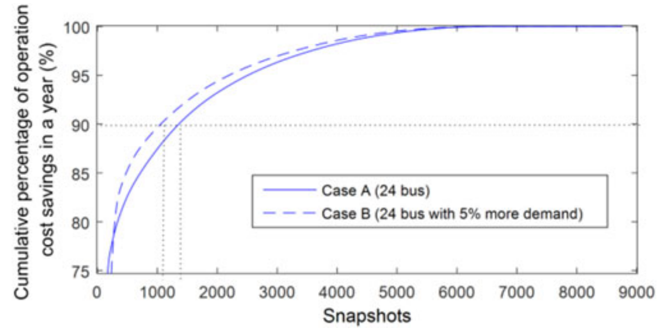

Fig. 3. Cumulative sum of the operation cost savings achieved in all snapshots to the left with respect to total savings throughout the whole year when sorting snapshots in decreasing order of operation cost savings achieved in them.

relevant snapshots would have probably led to obtaining a higher number of relevant snapshots the larger demand is. However, considering relevant snapshots as those where a certain fraction of savings are achieved has led, in this specific case, to the number of these snapshots being lower when demand is larger. This is because, as a consequence of the increase in demand, savings from reinforcing the grid have grown much more significantly in some few snapshots than in the rest of them. Therefore, most relevant savings have concentrated in a small number of snapshots.

For each relevant snapshot, individual reinforcement benefits are computed according to equations (19), (20) and (21). Relevant snapshots can be represented in the space of line benefits, whose dimension is 11 both in cases A and B.

PCA is applied to express the variability of the vector of reinforcement benefits in a more compact way, i.e. using a lower number of dimensions. Thus, the Principal Components of the reinforcement benefit dataset are determined. The smallest set of principal components of reinforcement benefits capturing at least 99% of the variability of these reinforcement benefits is retained. This corresponds to the set of the first four Principal Components both in case A and case B.

Relevant snapshots and non-relevant snapshots are clustered independently, as described in the methodology section. The results of clustering relevant snapshots in case A considering 10 clusters are depicted in Fig. 4.

C. Validation of results

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

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

The second alternative clustering approach applied is based on the scenario reduction technique proposed by authors in

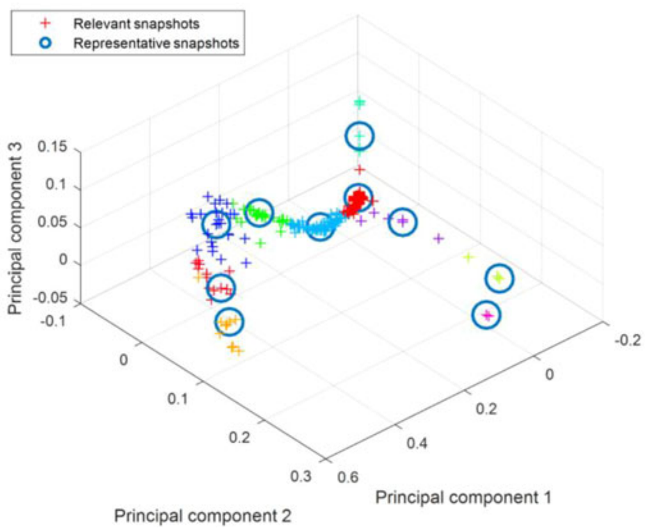

Fig. 4. Case A: representation of the relevant snapshots, the cluster they belong to, and their representatives in the space of the first three principal components of the reinforcement benefit space. Each color is associated with a specific cluster.

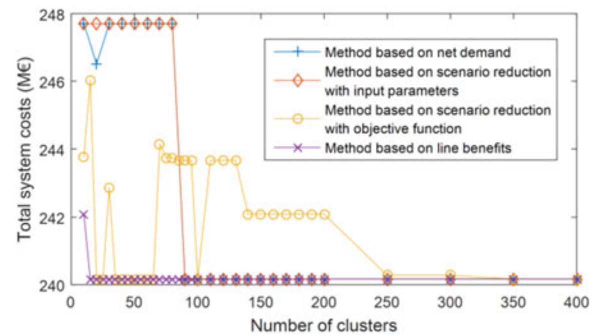

Fig. 5. Case A: Evolution of total system costs with the number of snapshots considered when these are selected 1) with the method proposed here, 2) according to the net demand in the system nodes (1st alternative approach), 3) with the scenario reduction technique based on total system costs (2nd alternative approach), and 4) with the scenario reduction technique based on input parameters (3rd alternative approach).

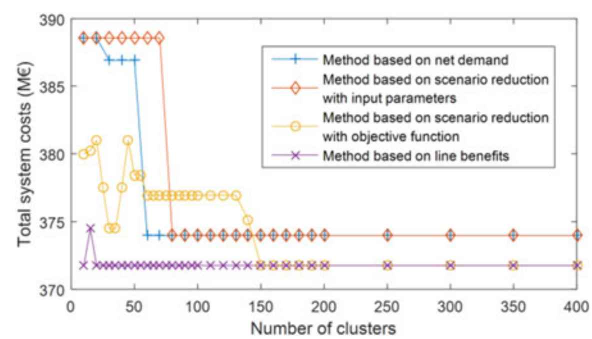

Fig. 6. Case B: Evolution of total system costs with the number of snapshots considered when these are selected 1) with the method proposed here, 2) according to the net demand in the system nodes (1st alternative approach), 3) with the scenario reduction technique based on total system costs (2nd alternative approach), and 4) with the scenario reduction technique based on input parameters (3rd alternative approach).

[15], whereby representative snapshots are selected according to the value for each snapshot of the total system operation and network investment costs. Network investment costs are annualized ones computed only considering the average snapshot throughout the year. Therefore, they are common to all snapshots. System operation costs for each snapshot are computed considering that this snapshot is the only one occurring in the whole year (the 8760 operation hours in the year are equal to this snapshot).

Lastly, the third alternative clustering approach applied here corresponds to the scenario reduction technique that groups together snapshots according to the demand, intermittent generation output, and conventional generator availability considered separately, which are the input parameters of the TEP problem, following the method proposed by authors in [16].

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

The evolution of total system costs with the number of clusters of snapshots computed, corresponding to the snapshot selection approach presented in this article and the three other approaches, is shown in Fig. 5 for case A. The same evolution, in case B, is shown in Fig. 6.

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

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

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

The fact that the method classifying snapshots according to the nodal net demand and the scenario reduction technique proposed in [16] (the 3rd alternative approach assessed here) provide similar results is not surprising. After all, they both select representative snapshots according to similar parameters: the demand and the intermittent generation output, or the net of them. On the contrary, characterizing snapshots according to the total system costs for them, as in [15], and doing so according to the benefits network investments produce in each snapshot, allows one to group together those snapshots that have a similar impact on the objective function value and the decision variables in the TEP problem, respectively. Both the TEP objective function value and the decision variables (network investments) are outcomes of the TEP problem (though a simplified version of this problem is considered in both cases, of course). The instability in the goodness of the selection of snapshots made with the 2nd alternative approach may be explained by the fact that this approach was devised to select a reduced set of scenarios, while the problem at hand here is to select a representative set of snapshots. Scenarios comprise a multiplicity of snapshots occurring over a certain period of time. Computing the expansion plan to consider, in the characterization of snapshots, according to the average snapshot only may not be appropriate due to the large differences existing among snapshots occurring throughout a year. Besides, computing total annual system operation costs according to a single snapshot (the one being characterized in each case) is probably not accurate enough. Therefore, grouping operation situations according to the total system costs in them may not be the most appropriate approach to select representative snapshots, while it is probably much better adapted to the selection of representative scenarios.

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

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

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

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

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

It can be noted that, although the time needed to compute (or select) the clustering variables is longest in our method, it

TABLE III
TIME COMPLEXITY AND ACCURACY OF THE SNAPSHOT SELECTION METHODS

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

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

becomes competitive in terms of computational burden when compared to the two methods based on scenario reduction techniques if the clustering algorithm is run 20 or more times to select the appropriate number of clusters. This is due to the fact that the k-means clustering algorithm employed in our approach is much less computationally demanding than the scenario reduction algorithm used to aggregate snapshots in the two methods based on scenario reduction.

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

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

VI. CONCLUSION

A novel method is proposed to select a reduced set of operation situations to consider in TEP analyses. Snapshots selected are the representatives of the clusters of the operation situations in the target year that are computed taking as classification variables the reduction in operation costs brought about by individual reinforcements to the network determined by solving the relaxed TEP model when considering all the operation snapshots. The reduction in the system operation costs produced by reinforcements are the driver of their deployment. Thus, the proposed approach is sound from a conceptual point of view. The Principal Components of benefits produced by individual network reinforcements are used to represent these benefits in a more compact way.

The performance of the snapshot selection method here proposed has been tested by applying it to two case studies based on the standard IEEE 24-bus RTS. The number of snapshots to consider in the reduced TEP problem when these snapshots are chosen according to the method proposed here is much smaller (up to 6 times lower) than the number of snapshots required when they are selected using standard snapshot selection approaches (for example, clustering snapshots according to the net demand in the nodes in each of the former). Besides, for a given number of snapshots, the efficiency of the expansion plan computed considering the snapshots selected with the approach proposed here is higher than the efficiency of the expansion plan computed considering the snapshots selected according to other advanced methods, like that proposed in [15]. Lastly, clustering the operation snapshots according to the proposed approach and solving the non-relaxed TEP model only considering the corresponding snapshot representatives is much less computationally burdensome than the original TEP problem, which considers all snapshots in the year (about 35 times more efficient from a computational point of view in the case study where it has been tested). Hence, the results produced by the proposed method for the selection of snapshots in TEP analyses are promising.

ACKNOWLEDGMENT

The authors would like to express their gratitude towards all partner institutions within the program as well as the European Commission for their support.

REFERENCES

[1] M. Boiteux, "La tarification des demandes en pointe: Application de la théorihtme de la vente au cout marginal," *Revue Generale de l'Electicite*, vol. 58, pp. 321–340, 1949.

[2] S. Agapoff and L. Warland, "Modular development plan of the pan-European transmission system 2050," E-highway 2050 report, 2014. [Online]. Available: http://www.e-highway2050.eu/fileadmin/documents/Workshop5/140407_e-Highway2050_Draft_Agenda_WS_6_May2014.pdf

[3] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. New York, NY, USA: Wiley, 2009.

[4] D. Z. Fitiwi, F. de Cuadra, L. Olmos, and M. Rivier, "A new approach of clustering operational states for power network expansion planning problems dealing with RES (renewable energy source) generation operational variability and uncertainty," *Energy*, vol. 90, pp. 1360–1376, Jul. 2015.

[5] S. Wogrin, P. Dueñas, A. Delgadillo, and J. Reneses, "A new approach to model load levels in electric power systems with high renewable penetration," *IEEE Trans. Power Syst.*, vol. 29, no. 5, pp. 2210–2218, Sep. 2014.

[6] S. Binato, M. Veiga, F. Pereira, and S. Granville, "A new benders decomposition approach to solve power transmission network design problems," *IEEE Trans. Power Syst.*, vol. 16, no. 2, pp. 235–240, May 2001.

[7] S. Lumbreras, A. Ramos, and P. Sánchez, "Automatic selection of candidate investments for transmission expansion planning," *Int. J. Elect. Power Energy Syst.*, vol. 59, pp. 130–140, Jul. 2014.

[8] G. P. McCormick, "Computability of global solutions to factorable nonconvex programs. Part I. Convex underestimating problems," *Math. Program.*, vol. 10, pp. 146–175, Dec. 1976.

[9] F. Bañez-Chicharro, L. Olmos, A. Ramos, and J. M. Latorre, "A benefit-based methodology to rank transmission expansion projects," Working Paper. [Online]. Available: http://www.iit.comillas.edu/docs/IIT-14-079A.pdf, accessed in Apr. 2016.

[10] C. C. Aggarwal, J. L. Wolf, P. S. Yu, C. Procopiuc, and J. S. Park, "Fast algorithms for projected clustering," *ACM SIGMOD Rec.*, vol. 28, no. 2, pp. 61–72, Jun. 1999.

[11] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms*, Philadelphia, PA, USA, 2007, pp. 1027–1035.

[12] C. Grigg *et al.*, "The IEEE reliability test system-1996. A report prepared by the reliability test system task force of the application of probability methods subcommittee," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 1010–1020, Aug. 1999.

[13] (1996). [Online]. Available: http://www.ee.washington.edu/research/pstca/

[14] N. Alguacil, A. L. Motto, and A. J. Conejo, "Transmission expansion planning: A mixed-integer LP approach," *IEEE Trans. Power Syst.*, vol. 18, no. 3, pp. 1070–1077, Aug. 2003.

[15] J. M. Morales, S. Pineda, A. J. Conejo, and M. Carrión, "Scenario reduction for futures market trading in electricity markets," *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 878–888, May 2009.

[16] J. Dupačová, N. Gröwe-Kuska, and W. Römisch, "Scenario reduction in stochastic programming," *Math. Program.*, vol. 95, no. 3, pp. 493–511, Mar. 2003.

Quentin Ploussard was born in Aix-en-Provence, France, in 1989. He received the Electrical Engineer degree from École Supérieure d'Électricité (Supélec), Paris, France, and the M.S. degree in economics from Université Paris Sud XI, Paris, France, both in 2014. He is currently working toward the Ph.D. degree in electrical engineering at Universidad Pontificia Comillas, Madrid, Spain.

Luis Olmos was born in Madrid, Spain, in 1976. He received the degree in industrial engineering and the Ph.D. degree in industrial engineering from the Pontificia Comillas University, Madrid, Spain, in 2000 and 2006, respectively. He belongs to the research staff of the Instituto de Investigación Tecnológica, Pontificia Comillas University. He has extensively researched the regulation of energy markets and the economics of power systems, with a special focus on the regulation of the transmission of electricity and all the problems related to it, i.e., allocation of sunk costs, design of tariffs, management of congestion, and planning of the expansion of the grid.

Andrés Ramos received the electrical engineer degree from Universidad Pontificia Comillas, Madrid, Spain, in 1982 and the Ph.D. degree in electrical engineering from Universidad Politécnica de Madrid, Madrid, in 1990.

He is a Research Fellow at the Technological Research Institute (IIT) and a Full Professor at Comillas' Engineering School, where he is the Head of the Department of Industrial Organization. His areas of interest include the operation, planning, and economy of power systems and the application of operations research to industrial organization.