Optimization of Cascade-Resilient Electrical Infrastructures and its Validation by Power Flow Modelling

Yiping Fang, Nicola Pedroni, Enrico Zio

Optimization of Cascade-resilient Electrical Infrastructures and its Validation by Power Flow Model

Yiping Fang[1], Nicola Pedroni[1], Enrico Zio[1,2]

[1]*Chair on Systems Science and the Energetic challenge*
Ecole Centrale Paris and Supelec, France

[2]*Energy Department, Politecnico di Milano, Italy*

ABSTRACT

Large scale outages on real-world critical infrastructures (CIs), although infrequent, are increasingly disastrous to our society. In this paper, we are primarily concerned with power transmission networks and we consider the problem of allocation of generation to distributors by rewiring links under the objectives of maximizing network resilience to cascading failure and minimizing investment costs. The combinatorial multi-objective optimization is carried out by a non-dominated sorting binary differential evolution (NSBDE) algorithm. For each generators-distributors connection pattern considered in the NSBDE search, a computationally-cheap, topological model of failure cascading in a complex network (named, the Motter-Lai (ML) model) is used to simulate and quantify network resilience to cascading failures initiated by targeted attacks. The results on the 400kV French power transmission network case study show that the proposed method allows to identify optimal patterns of generators-distributors connection which improve cascading resilience at an acceptable cost.

To verify the realistic character of the results obtained by the NSBDE with embedded ML topological model, a more realistic but also more computationally-expensive model of cascading failures is adopted, based on optimal power flow (namely, the ORNL-Pserc-Alaska (OPA) model). The consistent results between the two models provide impetus for the use of topological, complex network theory models for analysis and optimization of large infrastructures against cascading failure with the advantages of simplicity, scalability and low computational cost.

KEY WORDS: critical infrastructure, power transmission network, cascading failures, complex network theory model, power flow model, optimization

1 INTRODUCTION

Our modern society has come to depend on large-scale critical infrastructures (CIs) to deliver resources and services to consumers and businesses in an efficient manner. These CIs are complex networks of interconnected functional and structural elements. Large scale outages on these real-world complex networks, although infrequent, are increasingly disastrous to society, with estimates of direct

costs up to billions of dollars and inestimable indirect costs. Typical examples include blackouts in power transmission networks [1-3], financial bankruptcy [4], telecommunication outages [5], and catastrophic failures in socio-economic systems [6-7].

Cascading failures are initiated typically when a small part of the system fails for some reasons, and the load on that part (i.e. the flow passing through it) must be redistributed to other parts in the system. This redistribution may cause other components to exceed their capacity causing them also to fail. Hence, the number of failed or stressed components increases, propagating throughout the network. In particularly serious cases, the entire network is affected. Research regarding modeling, prediction and mitigation of cascading failures in CIs, whereby small initial disturbances may propagate through the whole infrastructure system, has addressed the problem in different ways [4-6, 8-13].

Albert et al. [14] demonstrated that the vulnerability of modern infrastructure networks (e.g., power transmission networks) is inherent to their structure. Thadakamalla [15] revealed that the topology of a supply infrastructure has great impact on its resilience. Then, much attention has been paid in recent years in the direction of network topology optimization, with the purpose of achieving desired targets of reliability and/or resilience [16-19, 26]. Shao et al. [17] proposed a shrinking and searching algorithm to maximize the reliability of a distributed access network with constrained total cost; however, the intense computational cost for evaluating network reliability prohibits the application of the model to large size networks. Gutfraind [18] introduced a multi-objective optimization method for constructing cascade resilient networks based on the structure of terrorist networks. Besides, Newth et al. [19] used a modified Metropolis evolutionary algorithm to evolve failure resilient networks with the objective of maximizing the average network efficiency. Cadini et al. [20] investigated the problem of optimizing the transmission reliability efficiency of an existing power transmission network with least cost by adding new connection links.

In practical cases, the cost of knocking down an existing network and reconstructing it from scratch is prohibitive, especially for CIs like the power transmission network. A more practicable alternative is to reconfigure parts of the network topology, e.g. by reallocation of the links which connect production facilities to consumers.

The primary objective of this paper is to propose a methodology for optimal allocation of the links connecting generators and distributors in a power transmission network for obtaining high resilience to cascading failures while keeping the investment costs low. Formulated as a large-scale, nonlinear and combinatorial multi-objective optimization problem, the facility allocation problem is solved by an evolutionary method, i.e., the non-dominated sorting binary differential evolution (NSBDE) algorithm [21, 22].

The search by the NSBDE requires also: (i) the construction of a model to describe the cascading failure process in the network of interest, and (ii) the repeated evaluation of the model for every possible generators-distributors configuration proposed by the algorithm during the search. With respect to the model, two approaches are typically considered in the analysis of power transmission systems: *complex network theory* models, such as the Motter-Lai (ML) model [8, 9] and *artificial power flow models*, such as the ORNL-Pserc-Alaska (OPA) model [10-12]. These approaches provide different tradeoffs between the (relatively low) computational cost associated to the model evaluation (allowing real-time applications to large scale power grids) and the (high) level of detail in the system description (including physical characteristics and power flows constraints), respectively.

The OPA model seeks to faithfully describe the dispatching dynamics of the power flows during the evolution of the failure propagation following the initial disturbances, by explicitly incorporating the standard DC power flow equations and minimizing generation cost and load shedding [11]. Embracing this more physical description and solving the constrained linear optimization functions associated to the model, result in a significant increase in the computational burden, rendering its application extremely difficult for realistic networks with large number of elements [13].

For these reasons, topological models based on complex network theory (e.g. the ML model) have emerged in recent years [8, 9, 23-25]. In particular, the ML model is a relatively simple and abstract model relying on the resemblance of complex networks to electrical infrastructure systems (in terms of graph theory). It has the advantage of modelling cascading dynamics with few parameters, so that its application to realistic, large-scale networks is feasible and certainly more readily than OPA [23]. However, ML abstracts the power flow laws and constraints of the electrical system. Inevitably, then, it cannot provide direct physical measures of blackout size, but rather abstract measures such as efficiency loss. This has posed questions on whether or not it is adequate in practice, due to its abstract nature, although it has been recognized to offer a new and interesting perspective on the study of cascading failures on power grids [24].

It is worth mentioning that studies tackling the problem of validation of network-centric approaches are few in literature. Some studies [13, 24] have provided qualitative comparisons between complex network theory models and power flow models – identifying similarities and differences, and evaluating advantages and disadvantages. Most recently, Correa and Yusta [26] conclude on the appropriateness of graph theory techniques for the assessment of electric network vulnerability by means of comparisons between physical power flow models and scale-free graph statistic indexes. Cupac et al. [27] have presented a method to quantitatively compare a network-centric model (CLM) and the power flow model OPA, finding that the CLM model exhibits overall properties which are consistent with the more realistic OPA fast-scale model. On the other hand, Fitzmaurice et al. [40] find that the topological nearest neighbor cascading failure model (namely, the TC model) shows different

characteristics from other two Kirchhoff models (namely, LD and QSS). Hines et al. [41] conclude that evaluating vulnerability in power networks using purely topological metrics may be misleading under some circumstances. Furthermore, Cotilla-Sanchez et al. [42] propose a new method for representing electrical structure using electrical distances rather than geographic connections.

In the present paper, we embrace the topological ML cascading failure model and embed it in the NSBDE for optimally solving the problem of generators-distributors link allocation. For exemplification, we apply the method to the 400 kV French power transmission network, under the objectives of maximizing network resilience to cascading failures and minimizing investment costs [28]. We, then, tackle the problem of realistic significance of the results that can be obtained with the proposed methodology. For this reason, the OPA is performed on the optimal network topologies found. To the authors' knowledge, this is the first study addressing the validation of optimization based on topological cascade model (namely, the ML model), by applying a more realistic power flow model (namely, the OPA model).

The optimization problem considered is addressing the network topology and in the specific case study we have considered for exemplification purpose the topology abstracted from the 400kV French power grid. In the abstraction, any station (generator, transmission/distribution substation) is regarded as one individual topological node in the network model, whereas the internal structure and functional logic of the specific station are ignored. Then, how the transmission lines interconnect with lower voltage networks has not been considered in this study, similar to what has been done in prior studies on these analyses [18-20]. The purpose of performing these analyses in this way is to leverage the simplicity and low computational cost of the topological (cascading failure) model used within the (evolutionary) network optimization, which otherwise would be very costly.

The remainder of this paper is organized as follows. In Section 2, we introduce the ML and OPA cascading failure models in detail. We, then, formulate the multi-objective optimization problem taking investment costs and failure resilience into account in Section 3. Section 4 unveils the detailed procedure of the proposed NSBDE algorithm. Section 5 illustrates the French 400kV power transmission network case study and the analysis and evaluation of the results. Discussion and conclusion are drawn in Section 6.

2 CASCADING FAILURE MODELS CONSIDERED IN THIS WORK

Modelling the dynamic evolution of system-wide cascading failure processes poses a number of challenges due to the diversity of mechanisms which can initiate the initial failure and influence the subsequent propagation of breakdowns in the power system [13]. Various cascading failure models have been proposed; these can be divided into two main categories: those based on complex network theory

analysis and those using power flow analysis, often including optimal economic power dispatch after each failure in the propagation, e.g., by linear optimal power flow (OPF) [27].

Complex network theory models, including the ML model adopted in this work and described in Section 2.1 below, abstract the representation of a power grid as a graph and then study the connectivity characteristics, the propagation mechanisms through the graph connections and their relationships. They typically consider flows of discrete packets that are injected and removed from all nodes and follow least distance paths, and the importance of links or nodes is measured by "betweenness", which is proportional to the number of least distance paths through the link or node [13]. Among these, the ML model is the widely used and relatively simple one. These types of models have proved to provide a good understanding of the specific grid dynamics of cascading failures [30]. However, in these models the assumptions only abstract the real loading of the components and the flow distribution through the connections. For this reason, it is necessary to ascertain the meaningfulness of the results for real electrical infrastructures.

Power flow models, on the contrary, are based on realistic power flow equations to describe the flow dispatching dynamics and failure evolution after the initial disturbances in the power grid. The OPA model, which is the most commonly used of this type of models, is introduced in Section 2.2 below and is based on the linearized or DC power flow approximation, which has been proved to be able to give a good approximation of active power flows in the network [29]. Another power flow model is the CASCADE model [43], though it is considered ''too simple'' in that it ''disregards the system structure, neglects the times between adjacent failures and generation adaptation during failure'' [44].

2.1. The ML model

The ML model has been proposed by Motter and Lai [8], with extensions to differentiate generators and loads [23]. The power transmission network is represented as an undirected graph Q with a set of N nodes representing N_G generators and N_D loads representing distribution substations, interconnected by a set of edges representing transmission lines. The structure of the network is identified by an $N \times N$ interaction matrix W, whose element w_{ij} is 0 if node i and j are not connected directly; otherwise it is assigned 1 for an unweighted network or a numerical value between i and j for a weighted network.

The ML model assumes that at each time step, one unit of the relevant quantity (electrical flow for power grids) is exchanged between every pair of generator and distributor nodes, and transmitted along the shortest path connecting them. The flow at one node is, then, the number of shortest paths passing through it. More precisely, the flow L_k passing through node k is quantified by the node betweenness calculated as the fraction of the generator-distributor shortest paths passing through that node:

$$L_k = \frac{1}{N_G N_D} \sum_{i \in V_G, j \in V_D, i \neq j \neq k} \frac{n_{ij}(k)}{n_{ij}} \quad (1)$$

where n_{ij} is the number of shortest paths between generator nodes and distributor nodes, and $n_{ij}(k)$ is the number of generator-distributor shortest paths passing though node *k*.

The capacity of node *k* is assumed to be proportional to its initial node L_k with a network tolerance parameter α,

$$C_k = (1 + \alpha) L_k \quad (2)$$

The concept of the tolerance parameter α ($\alpha > 0$) can be understood as an operating margin allowing safe operation of the component under potential load increment. The occurrence of a cascading failure is initiated by removal of a node, which in general changes the distribution of shortest paths. Then the load at a particular node can change and if it increases and exceeds its capacity, the corresponding node fails. Any failure leads to a new redistribution of loads and, as a result, subsequent failures can occur. It should be noted that the single failure mechanism applied here does not attempt to simulate a realistic trigger event of cascading failure; instead, it is only a manner of starting the cascading failure simulation for the ML model (and the OPA model introduced below).

Using this cascading failure model, the vulnerability of network *Q* can be characterized by the fraction of network efficiency lost in the cascading failure:

$$Vul(Q) = \frac{E(Q) - E(\bar{Q})}{E(Q)} \quad (3)$$

where $Vul(Q) \in (0,1)$ and $\bar{Q}$ represents the residual network structure after the initial failure. $E(Q)$ measures the network efficiency based on the node pair shortest path distance between generators and distributors. For its computation all pairs of nodes $i \in V_G$, and $j \in V_D$ are weighted by the inverse of their distance:

$$E(Q) = \frac{1}{N_G N_D} \sum_{i \in V_G} \sum_{j \in V_D} \frac{1}{d(i,j)} \quad (4)$$

where $d(i,j)$ is the number of edges for an unweighted network or the sum of edge weights for a weighted network in the shortest path from *i* to *j*.

The geodesic vulnerability $Vul(Q)$ measures the functionality of a network when subjected to a contingency due to cascading link disruption with regard to its steady state (base case). As $Vul(Q)$ increases, the impact on the network due to cascading failure also increases, as some components become disrupted. $Vul(Q)$ has been proved to be a well-defined index being capable of providing results consistent with those of physical model indices [(26)].

The detailed simulation of the ML cascading failure model proceeds as follows:

Step 1. Apply equation (1) to compute the initial load of each node for a proposed network by Floyd's shortest paths algorithm [31] and calculate the capacity of each node based on equation (2).

Step 2. Trigger the initial failure. In the optimization, one of the top five most loaded nodes is chosen as failed and, thus, is removed from the network.

Step 3. Recur to equation (1) and Floyd's shortest paths algorithm to recalculate the load of each working node in the network.

Step 4. Test each node for failure: for each node k ($k \in N$) of the network, if $L_k > C_k$ then node k is regarded as failed and, thus, is removed from the network.

Step 5. If any working node fails, return back to step 3. Otherwise, terminate the cascading simulation and evaluate the vulnerability of the network using equation (3).

Complex network theory models, such as the ML that we use within our optimization framework in Section 3, have no direct physical relation to the mechanisms of realistic power grids, but they have the key advantage that by utilizing techniques from graph theory they can be applied to analyze large-scale networks. For this reason, this modelling approach is seeing increasing applications for modelling cascading failure processes in power grids.

2.2. The OPA model

The OPA model has been proposed by researchers at Oak Ridge National Laboratory (ORNL), Power System Engineering Research Center of Wisconsin University (PSerc), and Alaska University (Alaska) [10-12]. The OPA model is built upon the Self-Organized Criticality (SOC) theory, contains two interdependent time scale dynamics, i.e., fast power flow dispatching dynamics and slow power grid growth dynamics, and describes the complexity and criticality of power systems. The slow time scale dynamics describes how the system evolves as demand changes over longer timeframes (e.g., over days), and subsequent system upgrades in response to demand vary and blackouts. On the other hand, the fast time scale dynamics depicts cascading failures of transmission lines over very short time (e.g., over seconds) during the slow dynamics. It is a novel and powerful tool for analyzing power systems. Our analyses focus on the fast power flow dynamics, in order to ensure comparability with the ML model shortest path assumption.

The cascading failure model is based on the standard DC power flow equation,

$$F = AP \tag{5}$$

where F is a vector whose N_E components are the power flows through the lines, F_{ij} (N_E is the total number of links in the network), P is a vector whose N-1 components are the power injection of each node, P_i (N is the total number of nodes in the network), with the exception of the reference generator, P_0, and A is a constant matrix that depends on the network structure and impedances (see Ref. (11) for

details about the computation of *A*). The reference generator power is not included in the vector *P* to avoid singularity of *A* as a consequence of the overall power balance.

The generator power dispatch is solved using standard linear programming methods. Using the input power demand, the power flow equation (5) is solved with the condition of minimizing the following cost function:

$$Cost = \sum_{i \in V_G} P_i(t) + K \sum_{j \in V_D} P_j(t) \quad (6)$$

where V_G and V_D are the sets of generators and distributors, respectively. This definition gives preference to generation shift whilst assigning a high cost (set K = 100) to load shedding, and it is assumed that all generators operate at the same cost and that all loads are served with equal priority. The minimization is done with the following constraints:

(1) Generator power injections are generally positive and limited by installed capacity limits: $0 \leq P_i \leq P_i^{max}, i \in V_G$.

(2) Loads always have negative power injections: $P_j^{dem} \leq P_j \leq 0, j \in V_D$.

(3) The absolute flow through links is limited by link capacities: $|F_{ij}| \leq F_{max}$.

(4) Total power generation and consumption remain balanced: $\sum_{i \in V_G \cup V_D} P_i = 0$.

After solving the linear optimization by using the simplex method as implemented in Ref. (32), we examine which lines are overloaded. A line is considered to be overloaded if the power flow through it is within 1% of the limit capacity F_{max}. Each overloaded line may outage with probability p_1 (p_1 is set as 1 in the case study to ensure its comparability with ML). If an overloaded line experiences an outage, its power flow limit F_{max} is divided by a very large number k_1 to ensure that practically no power may flow through the line. This action can avoid the infeasibility of the power flow optimization stem from topological islands of the system by removing the component directly. Besides, to avoid a matrix singularity from the line outage, the impedances of failed lines are multiplied by a large number k_2, resulting in changes of the network matrix *A*.

Load shedding is utilized to quantify the damage of the cascading failure. For an individual node, load shedding is defined as the difference between its power injection and demand:

$$S_i = P_i^{dem} - P_i \quad (7)$$

Subsequently, total load shedding for the system is:

$$S = \sum_{i \in V_D} S_i \quad (8)$$

Finally, system load shedding is normalized by its total demand *D* and used as a measure of cascading vulnerability:

$$S/_D = \frac{\sum_{i \in V_D} S_i}{\sum_{i \in V_D} P_i^{dem}} \quad (9)$$

The fact that simulation results from OPA model are consistent with historical blackout data for real power systems has justified its effectiveness [12]. However, the applications of OPA have generally been limited to networks with a relatively small number of nodes compared to real power grids [24], due to the computational efforts involved.

3 OPTIMIZATION MODEL

For a given network, cascading failure resilience could be enhanced in many ways. In this paper, we focus on choosing the connecting patterns between generators and distributors of a realistic power transmission network, so as to optimize resilience to cascading failures. In this study, system vulnerability to cascading failure (i.e. system functionality loss in cascading failure) is regarded as a reverse measure of system resilience: the less the functionality loss, the higher the system resilience. Given the goal of analyzing a realistic-size network, the ML cascading failure model is used to evaluate the resilience of a pattern of connections. By associating a cost to each link posed in the network, the optimization also seeks to minimize the total cost.

The network is modeled as a weighted graph, in which the edge weights are given by their physical distances which we assume directly related to the transmitting cost of the link. We define the variables to be optimized as the links of generation nodes to the different distribution nodes:

$$X_{ij} = \begin{cases} 1, \text{if } i \text{ is connected with } j \text{ directly} \\ 0, \text{otherwise} \end{cases} \quad (10)$$

for all $i \in V_G$ and $j \in V_D$. Two constraints have to be met when rewiring generators and distributors: (1) each distributor node is required to connect with at least one generator node or other distributor node, to make it accessible to the power supplying generators; (2) each generator node has to connect at least with one distributor node.

We assume that the cost associated with each connection cutting and rewiring is linearly proportional to the physical length of the linkage, with coefficient φ. The total investment cost of a reconstructed pattern X in the power transmission network can be defined as

$$C = \sum_{i \in V_G} \sum_{j \in V_D} \varphi X_{ij}\, d(i,j) \quad (11)$$

where $d(i,j)$ is the physical distance between i and j.

The cascading failure resilience of each reconstructed pattern X can be quantified by the vulnerability of the new network, given by equation (3). It should be noted that the effect of the type of initial event could significantly influence the cascading failure result: the efficiency loss of a cascade triggered by

the failure of a critical component could be much more severe than that originated by the failure of a normal component. Therefore, we consider a worst-case scenario in this study by choosing the failure of one of the top five most loaded nodes as initial failure in each cascade process simulation and, then, the results are averaged on the number of simulations.

Through the quantification of the connection pattern cost and cascading failure vulnerability, the facility allocation problem is formulated as a multi-objective optimization problem:

$$\begin{cases} \min C(X_{ij}) & (12a) \\ \min Vul\left(Q_{X_{ij}}\right) & (12b) \end{cases}$$

$$s.t. \begin{cases} \sum_{i \in V_G \cup V_D} X_{ij} > 0 \ \forall j \in V_D & (12c) \\ \sum_{j \in V_D} X_{ij} > 0 \ \forall i \in V_G & (12d) \end{cases}$$

The objective function (12a) is the sum of the fixed rewiring costs; (12b) expresses the resilience objective. The two constraints mentioned above are enforced by formulas (12c) and (12d), respectively. Observe that the least costly generator allocation is simply that with no links among facilities and consumers.

In our work, the multi-objective optimization problem (12a) – (12d) is tackled by the Non-dominated Sorting Binary Differential Evolution (NSBDE) algorithm presented in the next Section 4.

4 NON-DOMINATED SORTING BINARY DIFFERENTIAL EVOLUTION ALGORITHM FOR TOPOLOGY OPTIMIZATION

In this section, the operative procedures of the Non-dominated Sorting Binary Differential Evolution (NSBDE) algorithm are proposed for solving the multi-objective optimization problem introduced in Section 3 above. The starting point is the standard Differential Evolution (DE) algorithm, initially proposed as a population-based global optimization method for real-valued optimization problems, which has been found to outperform other optimization algorithms in various applications [21, 33, 34]. In order to solve the combinatorial multi-objective problem of interest, the fast non-dominated sorting, ranking and elitism techniques used in non-dominated sorting genetic algorithm-II (NSGA-II) [35] are introduced into a modified binary differential evolution (MBDE), which is a binary version of DE developed to tackle single-objective binary-coded optimization problems [36]. The NSBDE proceeds as follows: [21]

Step 1. Initialization of parameters

Set the values of the population size *NP*, the crossover rate *CR*, the scaling factor *F*, and the maximum generations N_{max}.

Step 2. Generation of initial population and evaluation

Initialize each individual in the population which is represented as a bit-string and denoted as $px_i^t = \{px_{ij}^t, |px_{ij}^t \in \{0,1\}; i = 1,2,\dots,NP, j = 1,2,\dots,M\}$, where *NP* is the population size and *M* is the dimensionality of the solutions. Each individual is also called a chromosome and forms a candidate solution to the problem. Each bit of each initial chromosome takes a value from the set {0, 1} with probability equals to 0.5: the bit takes '1' if the corresponding generator node and distributor node are connected, '0' otherwise.

Each of the *NP* chromosomes is evaluated by computing the two objective functions, i.e. formula (12a) and (12b).

Step 3. Generation of trial population

Apply the binary tournament selection operator [35] to the population PX^t to generate a trial population PV^t, which undergoes the evolution operations of mutation and crossover.

Step 3.1 Mutation

The following probability estimation operator *P*(*px*) is utilized to generate the mutated individuals according to the information of the parent population:

$$P\left(px_{ij}^t\right) = \frac{1}{1+e^{-\frac{2b[px_{r1,j}^t + F\left(px_{r2,j}^t - px_{r3,j}^t\right) - 0.5]}{1+2F}}} \tag{13}$$

where *b* is a positive real constant, usually set as 6; *F* is the scaling factor; $px_{r1,j}^t$, $px_{r2,j}^t$ and $px_{r3,j}^t$ are the *j*-th bits of three randomly chosen individuals at generation *t*. According to the probability estimation vector $P\left(px_i^t\right) = [px_{i,1}^t, px_{i,2}^t, \dots\ px_{i,N}^t]$ created by equation (13), the corresponding offspring pu_i^t of the current target individual px_i^t is generated as equation (14).

$$pu_{ij}^t = \begin{cases} 1, & \text{if } rand \le P(px_{ij}^t) \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

where *rand* is a uniformly distributed random number within the interval [0,1].

Step 3.2 Crossover

The crossover operator is used to mix the target individual and its mutated individual. The trial individual $pv_{ij}^t = (pv_{i,1}^t, pv_{i,2}^t, \dots\ pv_{i,N}^t)$ can be obtained by the crossover operator as follows,

$$pv_{ij}^t = \begin{cases} pu_{ij}^t, & \text{if } randj \le CR \text{ or } j = randi \\ px_{ij}^t, & \text{otherwise} \end{cases} \tag{15}$$

where *randj* $\in (0,1]$ is a uniform random value, *CR* is the crossover rate, and *randi* is a uniform discrete random number in the set {1, 2, ..., *NP*}.

Step 4. Evaluation

Evaluate each of the *NP* chromosomes in the population PV^t by computing its rewiring cost (12a) and resilience to cascading failures (12b) by performing the ML cascade process simulation procedure presented in Section 2.2.

Step 5. Union and Sorting

Combine the parent and trial populations to obtain a union population $PU^t = PX^t \cup PV^t$. Rank the individuals in the union population by the fast non-dominated sorting algorithm [33] with respect to the objective values, and identify the ranked non-dominated fronts F_1, F_2, …, F_k where F_1 is the best front, F_2 is the second best front and F_k the least good front.

Step 6. Selection

Select the first *NP* individuals from PU^t to create a new parent population PX^{t+1}. The crowding distance is used in this step to choose the individuals with the same front, where crowing refers to the density of solution present in a neighborhood of an individual of specified radius [35]: we prefer the individual which is located in a region with least number of individuals. The algorithm stops when it reaches the predefined maximum generations N_{max}.

5 CASE STUDY AND RESULTS ANALYSIS

5.1. Case study and parameters setting

In this paper, the 400kV French power transmission network (FPTN400) (Figure 1) is taken for exemplification of the proposed approach. The network is built from the data on the 400 kV transmission lines of the RTE website [37]. It has 171 nodes (substations) and 220 edges (transmission lines). We distinguish the generators, which are the source of power, from the other distribution substations, that receive power and transmit it to other substations or distribute it in local distribution grids. By obtaining the power plants list from EDF website [38] and relating them with the ID of the buses in the transmission network, we have 26 generators and 145 distributors. Only the nuclear power plants, hydroelectric plants and thermal power plants whose installed capacities are larger than 1000 MW, are considered.

For reallocation of the power generating nodes to the other nodes, the NSBDE algorithm introduced in the previous section is applied. The parameters values used to run the NSBDE algorithm are reported in Table I. The tuning parameters are chosen based on trial-and-improvement for fast convergence of the algorithm [28]. The network tolerance parameter α is set to 0.3 to simulate the normal operating condition; linkage cost parameter φ is set to 1.

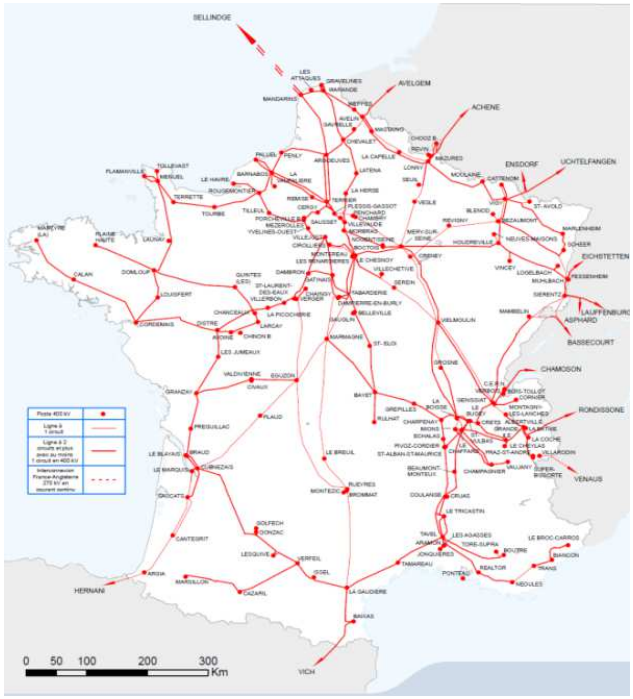

Fig. 1. The 400kV French power transmission network (FPTN400) [37]

At the beginning of the simulation, all 55 links among generators and distributors in the FPTN400 are cut off. The population is initialized by randomly assigning 0 or 1 to each bit of each chromosome in the population, forming a group of potential rewiring solutions. For evaluating the cascading vulnerability of a given generators-distributors allocation pattern, the ML cascading failure model is run starting from failing one of the top five most loaded (largest betweenness) nodes in repeated cascading simulations at the end of which the vulnerability values are averaged.

Table I . Parameters of the NSBDE algorithm

Parameters	Values
Population size *NP*	25
Dimensionality of solution *M*	3770
Crossover rate *CR*	0.9
Scaling factor *F*	0.2
Maximum generation N_{max}	300

5.2. Topological optimization results

Figure 2 reports the convergence plots of one run of the NSBDE algorithm. The top and bottom panels show the two optimal solutions with regard to the two objectives (12a) and (12b), respectively. It is observed that the algorithm is able to converge after around 150 generations.

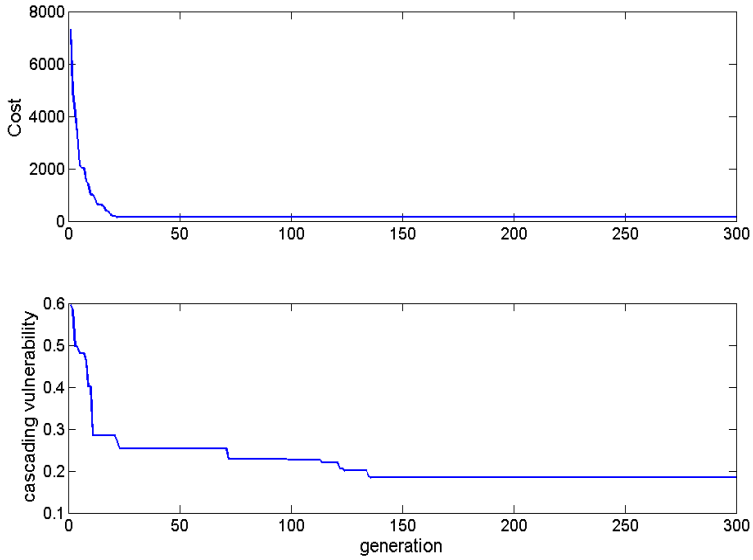

Fig. 2. Convergence plots of objective functions (12a) (top) and (12b) (bottom) during the evolution of NSBDE

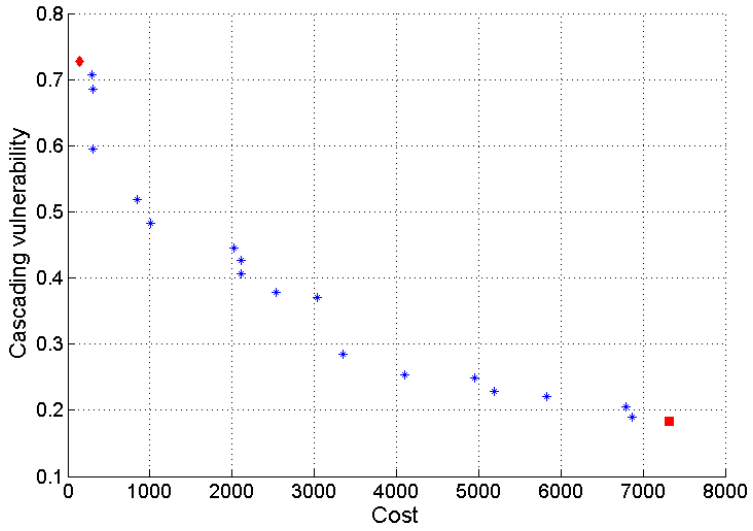

Fig. 3. Pareto front reached by a population of 25 chromosomes evolving for 300 generations

The Pareto front obtained by the NSBDE algorithm at convergence is illustrated in Figure 3. The diamond point in Figure 3 represents the current network with the present pattern of connecting links, which is also the least costly network; the square point is the most resilient network, whose cascading vulnerability is 0.184. It is not unexpected that the original network is the least costly one, since the electrical transmission lines and substations are placed with geographical constraints and connections between two distant substations are avoided. Actually, cost-effectiveness is a major consideration in constructing real power transmission networks.

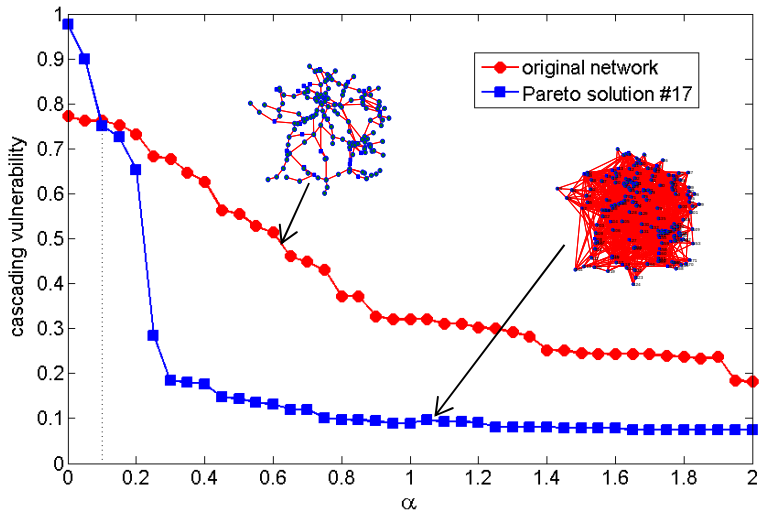

Fig. 4. Comparison of the cascading vulnerability between the original and the most resilient networks under different network tolerance values

It is also noted from Figure 3 that the cascading failure resilience of the FPTN400 can be improved significantly by properly rewiring the generator-distributor connections, though at a cost; the network vulnerability is decreased from 0.728 to 0.184 (when α=1.3) with an increased cost of 7.3×10^3 (i.e., 53.16 times increase). Figure 4 reports the cascading vulnerability comparison between the original network and the most resilient one (Pareto solution #17) with different tolerance parameters. It shows that when the network tolerance is very low, i.e. $0< \alpha < 0.1$, the optimized network loses most of its efficiency, i.e., it is quite vulnerable to intentional attacks possibly due to its intensive loading condition. However, when $\alpha \geqslant 0.3$ (which is generally the normal operating condition [13]), the optimized network loses less than 20% of its efficiency during a cascading failure initiated by intentional attack.

Albeit a substantial improvement of the cascading failure resilience of the FPTN400 is possible by adding redundant links, a tradeoff between the cost and resilience improvement is necessary for rational decision-making. Along the Pareto frontier of the potential solutions, there are some points at which a small sacrifice of cost gives a large gain of cascading resilience. More generally, by taking a network solution and its neighbor on the frontier (the less costly one), one can define a rate of change of cascading resilience with respect to cost: $|\Delta Vul/\Delta cost|$. This rate can be utilized as a reference to choose the optimized network: the larger the ratio, the more preferred the network is.

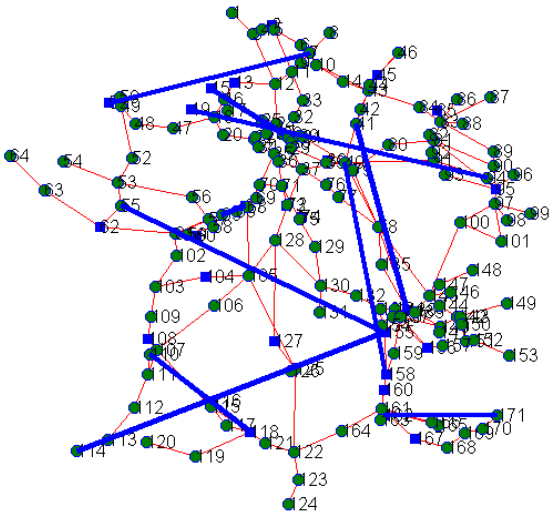

Fig. 5. The topology of the Pareto solution #3 and its difference with the original network

Figure 5 reports the topology of the network corresponding to the Pareto solution #3 (310.6, 0.59) whose $|\Delta Vul/\Delta cost|$ value is comparatively large. The bold links represent the 10 added connections with respect to the original real network: notice that only 10 links are required to be rewired for the original network to gain a 19.2% cascading resilience improvement (the cascading vulnerability is decreased from 0.73 to 0.59). Besides, it is noted from Figure 5 that the newly added links tend to connect distant generator and distributor pairs, indicating that the installation of power lines between remote power substations can improve the resilience of the system, although at larger costs.

5.3. Validation by the OPA model

All the optimization results presented in the previous section are based on the ML model which abstracts basic power flow constraints and electrical characteristics of the power transmission network. In this section, the more realistic OPA model introduced in Section 2.2 is utilized to verify the optimal results found.

The verification is not straightforward due to the differences of the two models in the way of representing system capacity, in the iterative algorithms they rely on, and in the way of measuring the damage produced by the cascading failure. Accordingly, some assumptions and adjustments to the OPA model (as described in Section 5.3.1) are necessary to ensure its applicability to assess the optimization solutions obtained based on the ML model [27].

5.3.1 *OPA Adjustments*

Five representative solutions (i.e., the least cost network FPTN400, the Pareto solution #17 (7300, 0.184) which is the most resilient, together with the solutions #3 (310.6, 0.59), #5 (3344.3, 0.28) and #13 (1003.8, 0.48) whose $|\Delta Vul/\Delta cost|$ values are comparatively large) along the Pareto front are chosen as the basic network topologies to be verified by the OPA model. To facilitate comparability with the ML model, all the generators are assumed to have equal capacity, and all the loads are

assumed to have equal constant demand (we use 26, i.e. the number of generators in the simulation). Furthermore, all edge impedances are calculated using the typical reactance value 0.28 ohm/km at 50 Hz [39]. This heterogeneous impedance setting aligns with the weighted edge initialization in the ML model.

The ML model uses the parameter α to represent network tolerance, while regarding the OPA model, prior studies set the initial limits (demand, generator capacity, line flow limits) by evolving the network using combined fast-slow dynamics until the network reaches a steady state [11]. Considering that we limit the scope of the OPA evaluation to fast dynamics, we use a simpler initialization strategy (proposed by Cupac et al. [27]) which does not require the slow power grid growth dynamics, and apply the parallel capacity setting (the α model) to facilitate the comparison. In particular, the values of the initial flows $\overline{F_{ij}}(0)$ and of the link capacities F_{ij}^{max} are determined as follows: demand for all distributor nodes is fixed to a constant amount, as mention above, and total generation capacity is set to be equal to total demand, and equally divided among the generators. Then, the power flows along the lines are estimated by assuming that every distributor node would obtain an equal amount of power from every generator. The initial flows are calculated by selecting a generator (one at a time), setting all other generator capacities to 0 and then computing power flows to each distributor node. The sum of the power flows over all the generators results in the estimated initial flow along each link, $\overline{F_{ij}}(0)$. Analogous to the initialization process in the ML model, the maximum capacity for a link connecting nodes *i* and *j* is given by

$$F_{ij}^{max} = (1 + \alpha)\left|\overline{F_{ij}}(0)\right| \tag{16}$$

It is noted that the values of the initial flows $\overline{F_{ij}}(0)$ are only used to set the link flow capacities F_{ij}^{max} in such a way that they are comparable to the capacities C_k used by the ML model. The network tolerance parameter is set to $0 \leq \alpha \leq 2$ in our approach, parallel to the ML model, representing excess transmission capacity. Then, the node transmission capacity is modelled as the sum of link flow capacities of adjacent links $\sum_{j \in V_j} F_{ij}^{max}$ where V_j is the set of nodes directly connected to node *i*.

In the OPA implementation, the probability of an overloaded link is set to $p_1 = 1$ (identical with that in Cupac et al. [27]), to ensure comparability with ML, where an overloaded node fails and is removed from the network with certainty. This setting will not change the OPA validation results where only the relative ranking of cascade vulnerability for each network is considered, although it has probably changed all the absolute values of cascade vulnerability. Besides, we initiate the cascade in the same manner that we do in the ML model, as stated in Section 3.

5.3.2 *Validation Results*

Figure 6 reports the landscapes of the node transmission capacities C_k and F_{ij}^{max} under both ML model and OPA model, respectively, for the five chosen networks (with $\alpha = 0$). It shows that node capacities in ML are highly correlated with node capacities in OPA model for the FPTN400, Pareto solution #3, #5 and #13 (actually, the correlation coefficients are 0.904, 0.890, 0.862 and 0.914 respectively); for Pareto solution #17, the linear correlation of node transmission capacities still exists (with correlation coefficient 0.619). This indicates that the initialization strategy is consistent for ML and OPA models: nodes with high capacity in ML tend to have high capacity in OPA, and nodes with low capacity in ML also tend to have low capacity in OPA [27].

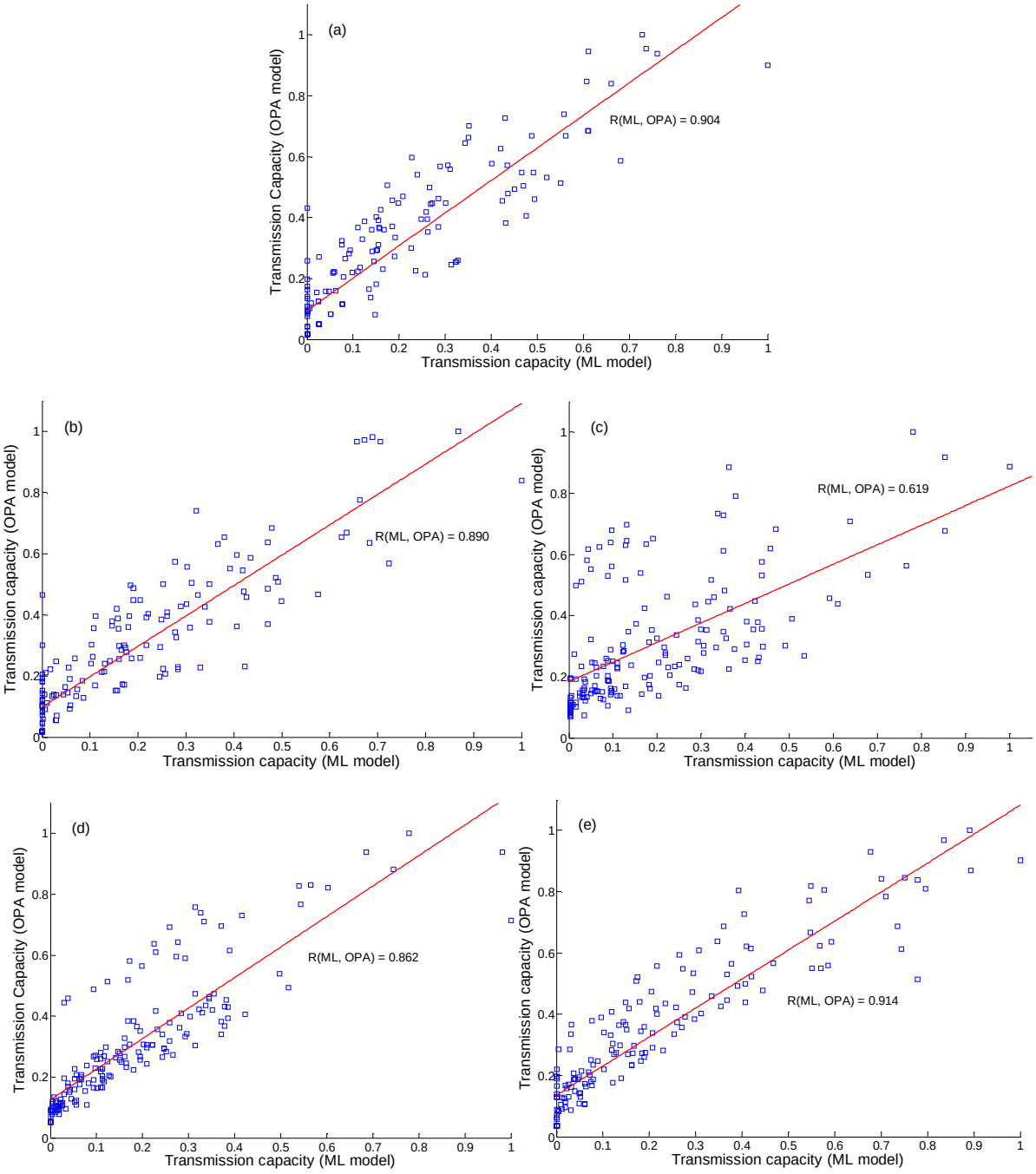

Fig. 6. Scatterplot of normalized node transmission capacity in ML versus OPA model using, (a) the original FPTN400; (b) Pareto solution #3 network; (c) Pareto solution #17 network; (d) Pareto solution #5; (e) Pareto solution #13. Node transmission capacity in OPA is highly correlated with transmission capacity in ML, the correlation coefficient are 0.904, 0.890, 0.619, 0.862 and 0.914 for the five networks, respectively. The solid lines represent the best fits.

In Figure 7, we plot the curves of normalized load shedding S/D versus network tolerance α obtained by applying the OPA model to the five representative networks selected from the Pareto front. The OPA simulation is triggered by removing one of the top five most loaded nodes (i.e., targeted initial failure). Analogous to the ML model (Figure 4), the network damages decreases when network tolerance increases for all the networks. When network tolerance value is high enough ($\alpha > 1.2$), any small intentional disturbance on the network would tend to cause quite low damage to the functioning of the network ($< 1\%$). Most importantly, it is observed that in the OPA simulation, the network corresponding to Pareto solution #3 (310.6, 0.59) (green triangle curve) is more resilient, i.e., it presents less load shedding than the original network (red circle curve) over a wide range of network tolerance α (i.e., $0 < \alpha < 1.2$); in addition, solution #13 (1003.8, 0.48) (magenta diamond curve) generally outperforms the solution #3, while solution #5 (3344.3, 0.28) (grey star curve) outperforms #13 in terms of cascade resilience. Finally, Pareto solution #17 (7300, 0.184) (which is the most resilient network according to the ML model) presents the lowest load shedding among the five networks over the entire range of α values considered. This ranking of cascading failure resilience is consistent with the simulation results based on ML model.

Figure 8 shows the results of OPA simulation on the five networks, where the failures are triggered by removing a randomly chosen node (i.e., random initial failure) and the results are averaged over 30 different samples. The ranking of cascade resilience of the five networks here is also parallel with the optimization results based on ML. This demonstrates that a resilience-improved network from the optimization based on the ML model is also more resilient than another one if evaluated by the more realistic OPA cascade simulation, therefore, verifying that the insights gained by the topological optimization approach are valid.

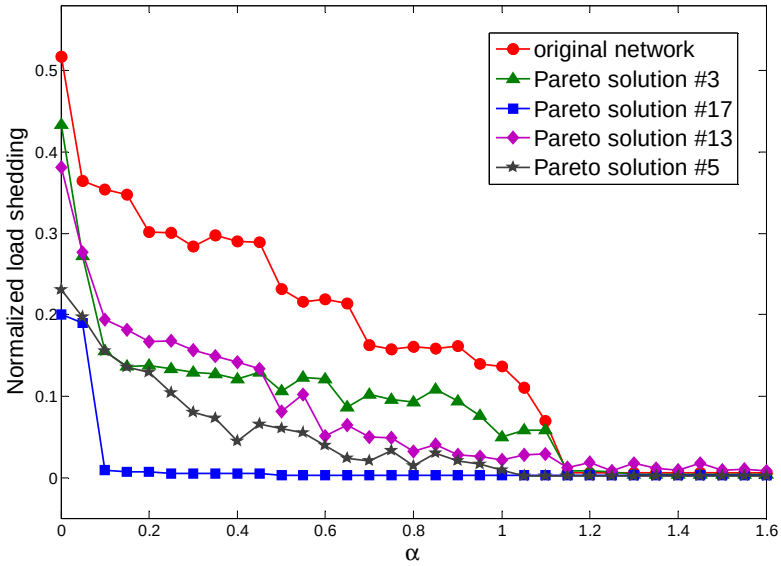

Fig. 7. Cascading vulnerability (normalized load shedding) evaluated by the OPA model for the five chosen networks over a range of network tolerance values α under targeted initial failure.

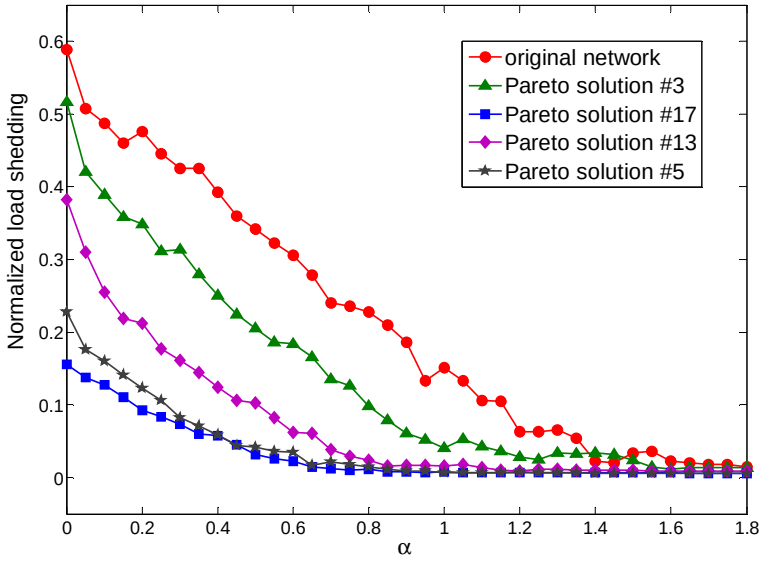

Fig. 8. Cascading vulnerability (normalized load shedding) evaluated by the OPA model for the five chosen networks over a range of network tolerance values α under random initial failure. The results have been averaged over 30 different samples.

Also important is to remember that the results produced by the simple ML topological model are obtained at a much lower computational cost than those of the OPA model: actually, the average time needed to carry out a single cascading failure simulation is 3.9s and 20.8s for the ML and OPA models, respectively, on a double 2.4 GHz Intel CPU and 4 GB RAM computer.

6 DISCUSSION AND CONCLUSIONS

Generally, the structure of power grids emerges through an unplanned growth process to meet service demand and/or results from optimization of costs. However, the increasing threat of large scale

failures, albeit infrequent, makes it vital to think of the design of resilient network systems capable to resist against and recover from cascading failures.

In this paper, we have investigated the allocation of generators to distributor nodes by rewiring links under the objectives of maximizing the network cascading failure resilience and minimizing the investment costs.

In realistic cases of networks of large number of nodes, the problem is a combinatorial multi-objective optimization problem. To effectively tackle the problem, we have proposed a NSBDE multi-objective algorithm, within a Pareto optimality scheme of search for non-dominated solutions. To simulate and quantify the cascading failure resilience of network connection solutions selected during the NSBDE search, a complex network model – namely, the Motter-Lai (ML) model – has been used, to exploit is rapidity of calculation.

Exemplification has been done by considering the 400kV French power transmission network (FPTN400). The results of the case study have shown that generator-distributor allocation can be optimized to improve the cascading resilience of a realistic power transmission network system at an acceptable cost.

To validate the physical significance of the topological optimization results, a detailed and more realistic power flow model – i.e., the ORNL-Pserc-Alaska (OPA) model – has been considered. The OPA model has been applied to five network topologies selected from the Pareto front found by the topological optimization process. The ranking of the five selected networks with respect to their vulnerability to both intentional attacks and random failure is consistent with that of the ML model; in addition, the computational time required by the ML approach is shown to be 5.5 times lower than that of the OPA approach. This verifies (i) the physical meaningfulness of the topological optimization solutions and (ii) the practical usefulness of abstract cascading models in network optimization tasks.

It is noted that this consistency is not insignificant since it demonstrates that one resilience-improved pattern of capacity allocation optimized by the ML model is also of higher resilience if measured by the more realistic OPA model, providing impetus for the use of topological, complex network theory models for ensemble analysis and optimization of large infrastructures against cascading failure with the advantages of simplicity, scalability and low computational cost (e.g., future studies may consider using complex network cascading models to optimize both the topology and electrical/reliability properties of realistic power networks, which may enable unraveling questions such as which type of resource distribution is the most favorable for a network to resist to cascading failure, when the total resource is limited).

The initialization strategy of the OPA model in this paper ensures that we can use the network tolerance parameter α as a common measure of transmission capacity for both models. However, the

actual data could be used in the OPA validation if they are initially applied in the optimization based on the ML model, and if they are available. This could be possible future work. Besides, performing optimizations using directly detailed and computationally intensive power flow models (e.g., embrace Newton Raphson based power flow approaches [45] and/or realistic trigger event such as natural hazard and malevolent targeted disruption [46], into the cascade modelling framework) would enable a more thorough and comprehensive comparison of the two classes of approaches considered in this paper. Furthermore, it may be useful to model variations in generation capacity and to consider situations where generation capacity and demand are not equally distributed, which is aligned with more realistic cases of power grids. Finally, while being relatively small compared to real scenarios with thousand buses due to computational constraints and data availability, the proposed network is sufficient to illustrate the usefulness of the topological optimization methodology in this study. Nevertheless, we believe that application of the topological approach to large-scale networks is interesting and this falls perfectly within the scope of our future research in this direction.

REFERENCES

(1) Final Report on the August 14, 2003 Blackout in the United States and Canada, US-Canada Power System Outrage Task Force, Tech. Rep., 2004.
(2) Final Report System Disturbance on 4 Nov. 2006, Union for the Coordination of Transmission of Electricity, Tech. Rep., 2007.
(3) Helen Pidd. India blackouts leave 700 million without power. The Guardian, 31 July 2012.
(4) Battiston, Stefano, et al. Credit chains and bankruptcy propagation in production networks. Journal of Economic Dynamics and Control 31.6 (2007): 2061-2084.
(5) Newman, Mark EJ, et al. Email networks and the spread of computer viruses. Physical Review E 66.3 (2002): 035101.
(6) Zhao Kang, et al. Analyzing the resilience of complex supply network topologies against random and targeted disruptions. Systems Journal, IEEE 5.1 (2011): 28-39.
(7) Kempe, David, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 2003.
(8) Motter AE, Y-C Lai. Cascade-based attacks on complex networks. Physical Review E 66.6 (2002): 065102.
(9) Crucitti, Paolo, Vito Latora, and Massimo Marchiori. A model for cascading failures in complex networks. Physical Review E 69.4 (2004): 045104.
(10) Dobson Ian, et al. Complex systems analysis of series of blackouts: cascading failure, criticality, and self-organization. Bulk power system dynamics and control-VI (2004): 22-27.
(11) Dobson I, Carreras BA, Lynch VE, Newman DE. An initial model for complex dynamics in electric power system blackouts. In: Proceedings of the 34th annual Hawaii international conference on system sciences, 2001; 2001. p. 710–8.
(12) Carreras BA, Newman DE, Dobson I, and Poole AB. Evidence for self-organized criticality in a time series of electric power system blackouts. Circuits and Systems I: Regular Papers, IEEE Transactions on 51, no. 9 (2004): 1733-1740.
(13) Baldick R, et al. Initial review of methods for cascading failure analysis in electric power transmission systems IEEE PES CAMS task force on understanding, prediction, mitigation and restoration of cascading failures. Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE, pp. 1-8.
(14) Albert, Réka, István Albert, Gary L. Nakarado. Structural vulnerability of the North American power grid. Physical Review E 69.2 (2004): 025103.
(15) Thadakamaila, HP, et al. Survivability of multiagent-based supply networks: a topological perspective. Intelligent Systems, IEEE 19.5 (2004): 24-31.
(16) Boorstyn, Robert, Howard Frank. Large-scale network topological optimization. Communications, IEEE Transactions on 25.1 (1977): 29-47.

(17) Shao, Fang-Ming, Xuemin Shen, Pin-Han Ho. Reliability optimization of distributed access networks with constrained total cost. Reliability, IEEE Transactions on 54.3 (2005): 421-430.
(18) Gutfraind A. Optimizing topological cascade resilience based on the structure of terrorist networks. PloS one 5.11 (2010): e13448.
(19) Newth, David, Jeff Ash. Evolving cascading failure resilience in complex networks. Proc. of 8th Asia Pacific Symp. on Intelligent and Evolutionary Systems. 2004.
(20) Cadini, F., Zio, E., & Petrescu, C. A. (2010). Optimal expansion of an existing electrical power transmission network by multi-objective genetic algorithms. Reliability Engineering & System Safety, 95(3), 173-181.
(21) Li Y-F, Sansavini G, Zio E. Non-Dominated Sorting Binary Differential Evolution for the Multi Objective Optimization of Cascading Failures Protection in Complex Networks. Reliability Engineering & System Safety (2012).
(22) Zio E., Golea LR, Sansavini G. Optimizing protections against cascades in network systems: A modified binary differential evolution algorithm. Reliability Engineering & System Safety, Volume 103, July 2012, Pages 72-83.
(23) Kinney R, Crucitti P, Albert R, & Latora V. Modeling cascading failures in the North American power grid. The European Physical Journal B-Condensed Matter and Complex Systems, 46.1(2005): 101-107.
(24) Sun K, Han Z-X. Analysis and comparison on several kinds of models of cascading failure in power system. In: Transmission and distribution conference and exhibition: Asia and Pacific, 2005 IEEE/PES; 2005. p. 1–7.
(25) Buldyrev, Sergey V, et al. Catastrophic cascade of failures in interdependent networks. Nature 464.7291 (2010): 1025-1028.
(26) Correa GJ, and Yusta JM. Grid vulnerability analysis based on scale-free graphs versus power flow models. Electric Power Systems Research 101 (2013): 71-79.
(27) Cupac V, Lizier JT, Prokopenko M. Comparing dynamics of cascading failures between network-centric and power flow models. International Journal of Electrical Power & Energy Systems, 49, (2013): 369-379.
(28) Fang, YP, Pedroni, N. and Zio, E. Optimal Production Facility Allocation for Failure Resilient Critical Infrastructures. In ESREL 2013 (2013, September).
(29) Purchala K, Meeus L, Van Dommelen D, Belmans R. Usefulness of DC power flow for active power flow analysis. In Power Engineering Society General Meeting, 2005. IEEE (pp. 454-459). IEEE. (2005, June).
(30) Holmgren AJ. Using graph models to analyze the vulnerability of electric power networks. Risk analysis 26.4 (2006): 955-969.
(31) Floyd RW. Algorithm 97: shortest path. Communications of the ACM 5.6 (1962): 345.
(32) Press WH, Flannery BP, Teukolsky SA, Vetterling WT. Numerical Recipes in C. Cambridge Unversity Press, Cambridge, 1988.
(33) Price KV, Storn RM, Lampinen JA. Differential evolution a practical approach to global optimization. (2005).
(34) Ponsich A, Coello CA. Differential Evolution performances for the solution of mixed-integer constrained process engineering problems. Applied Soft Computing, 11.1(2011): 399-409.
(35) Deb Kalyanmoy, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II. Evolutionary Computation, IEEE Transactions on 6.2 (2002): 182-197.
(36) Wang Ling, et al. A modified binary differential evolution algorithm. Life System Modeling and Intelligent Computing. Springer Berlin Heidelberg, 2010. 49-57.
(37) RTE. Le Réseau de Transport d'Electricité 400 kV, 2011, http://www.rte-france.com
(38) EDF. En direct de nos centrales, http://france.edf.com/france-45634.html, Retrieved Avril, 2013.
(39) Zhou Q, Bialek JW. Approximate model of European interconnected system as a benchmark system to study effects of cross-border trades. Power Systems, IEEE Transactions on, 20.2 (2005): 782-788.
(40) Fitzmaurice R, Cotilla-Sanchez E, Hines P. Evaluating the impact of modeling assumptions for cascading failure simulation. In Power and Energy Society General Meeting, 2012 IEEE (pp. 1-8). IEEE.
(41) Hines P, Cotilla-Sanchez E, Blumsack S. Do topological models provide good information about electricity infrastructure vulnerability? Chaos: An Interdisciplinary Journal of Nonlinear Science, 20.3 (2010), 033122.
(42) Cotilla-Sanchez E, Hines PD, Barrows C., Blumsack S. Comparing the topological and electrical structure of the North American electric power infrastructure. Systems Journal, IEEE, 6.4 (2012), 616-626.

(43) Dobson I, Carreras BA, Newman DE. A loading-dependent model of probabilistic cascading failure. Probability in the Engineering and Informational Sciences, 19.1 (2005), 15-32.
(44) Leelaruji R, Valerijs K. Modeling adequacy for cascading failure analysis. In Power Engineering Conference, 2008. AUPEC'08. Australasian Universities, pp. 1-6. IEEE, 2008.
(45) Wang H, and Thorp JS. Optimal locations for protection system enhancement: a simulation of cascading outages. Power Delivery, IEEE Transactions on, 16(4), pp. 528-533, 2001.
(46) Dueñas-Osorio L, Vemuru SM. Cascading failures in complex infrastructure systems. Structural safety, (2009) 31(2), 157-167.