Electric Distribution Network Expansion Under Load-Evolution Uncertainty Using an Immune System Inspired Algorithm

Eduardo G. Carrano, Frederico G. Guimarães, Ricardo H. C. Takahashi, Oriane M. Neto, and Felipe Campelo, *Student Member, IEEE*

***Abstract*—This paper addresses the problem of electric distribution network expansion under condition of uncertainty in the evolution of node loads in a time horizon. An immune-based evolutionary optimization algorithm is developed here, in order to find not only the optimal network, but also a set of suboptimal ones, for a given most probable scenario. A Monte-Carlo simulation of the future load conditions is performed, evaluating each such solution within a set of other possible scenarios. A dominance analysis is then performed in order to compare the candidate solutions, considering the objectives of: smaller infeasibility rate, smaller nominal cost, smaller mean cost and smaller fault cost. The design outcome is a network that has a satisfactory behavior under the considered scenarios. Simulation results show that the proposed approach leads to resulting networks that can be rather different from the networks that would be found via a conventional design procedure: reaching more robust performances under load evolution uncertainties.**

***Index Terms*—Artificial immune systems, load evolution uncertainty, multiobjective sensitivity analysis, network optimization, power distribution planning.**

I. Introduction

Electric distribution networks, from time to time, are expanded and re-designed, in order to follow the changes that occur in the load [1]–[3]. The design of such networks must take into account not only the present load, but also the load that is expected to exist within some time horizon. This means that the design problem is intrinsically endowed with some uncertainty. Performing a design procedure that under-estimates the load growth means that the system will no longer be able to supply the demand and a new expansion will be necessary early. It leads to a global cost that is much greater than if the expanded network were able to supply the future load with the configuration of the first design. On the other hand, a design procedure performed with over-estimated loads would lead to a costly network that does not make a reasonable usage of the resources employed for its installation. These considerations lead the electric utility companies to perform the planning of electrical distribution network expansion according to some forecast of the demand for a given time horizon, considering some "mean" or "most likely" scenario for the load expansion [2], [3].

The design of a distribution network is a complex problem: it is of combinatorial nature, with nonlinear functionals [4], [5]. This problem consists of finding an optimal configuration for the network, including the topology (connection configuration) and conductor setup (capacity of each conductor in a specific topology), with constraints related to technical specifications such as load demands and capacity of the distribution lines [5]–[8]. Finding the exact optimal configuration of such system becomes a hard computational task even for a moderate number of nodes [7], [8]. This possibly explains why most of the design methodologies just consider a single future scenario (the "most likely" one), instead of considering a set of possible load scenarios [6], [7], [9]. Furthermore, the approaches that consider multiple scenarios just take into account the simplified situations in which the load in all nodes grow with identical rates [1].

In fact, the variation of the load in each system node is a stochastic process, leading to a set of load probability distributions in the system nodes. The variances of these distributions grow with time [10], [11]. Consequently, a representative discrete set of possible load scenarios in each system node leads to a combinatorial set of load scenarios for the whole system. In addition, some factors may cause the net energy tax in each system node to become different from the other nodes. For instance, the final consumer load location can evolve with different geographical patterns, which leads to different costs and different losses in the secondary distribution systems.

Since the cardinality of the set of possible scenarios grows exponentially with the number of nodes, it becomes clear that any attempt for directly performing the (combinatorial) optimization of the network configuration considering the whole set of scenarios is infeasible, even for small instances of problems.

In this paper, the probability distribution of load evolution is supposed to be known for each node of a system to be expanded. The computation of the optimal network is performed considering the mean load for each node, and a unique expected mean energy tax. Several suboptimal solutions whose costs are not far from the optimal one are simultaneously found, leading to a

Manuscript received July 26, 2006; revised December 25, 2006. This work was supported by CNPq, CAPES, and FAPEMIG – Brazil. Paper no. TPWRS-00495-2006.

E. G. Carrano, F. G. Guimarães, and O. M. Neto are with the Department of Electrical Engineering, Universidade Federal de Minas Gerais, Belo Horizonte, MG 31270-010, Brazil (e-mail: carrano@cpdee.ufmg.br, fgg@ufmg.br; oriane@cpdee.ufmg.br).

R. H. C. Takahashi is with the Department of Mathematics, Universidade Federal de Minas Gerais, Belo Horizonte, MG 30123-970, Brazil (e-mail: taka@mat.ufmg.br).

F. Campelo is with the Laboratory of Hybrid Systems, Graduate School of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan (e-mail: pinto@emsi.eng.hokudai.ac.jp).

Digital Object Identifier 10.1109/TPWRS.2007.894847

candidate-solution set. The candidate-solution set is employed within a sensitivity analysis, under a set of load-evolution conditions (including load level and net energy tax) which is representative of the combinations of possible load evolutions in each node. The final solution is chosen as the one that minimizes the cost function weighted by the probability distributions, while keeping the probability of feasibility[1] and the fault cost[2] above a specified level. The case studies presented here reveal an interesting pattern: the mean-scenario optimal solution is sometimes more sensitive than other solutions for such variations, and often is not the final solution to be chosen.

Due to the nature of network design problems, of combinatorial optimization with nonlinear functional, evolutionary optimization techniques have become the main class of techniques applied in such problems [12]. Genetic algorithms (GAs) [5], [6], simulated annealing methods [9], ant colony system algorithms [7], and other evolutionary techniques have been used successfully for searching optimal configurations for electric distribution networks. These algorithms, however, usually deliver a single final solution, that is expected to be near the global optimum. *Artificial Immune Systems* (AIS) are computational techniques inspired by theoretical immunology that, differently from other evolutionary algorithms, deliver not only a single solution (the optimal solution) at the end of the optimization procedure, but also an entire set of suboptimal solutions (local optima) that are explicitly kept and evolved along the optimization process [13], [14]. This motivates applying AIS algorithms for developing the approach used in this paper[3]. The suboptimal solutions delivered by AIS are nearly optimal related to the criterion that was employed for the optimization. The point is that they can be useful as design alternatives under perturbed load conditions not considered explicitly in the optimization stage.

This paper is structured as follows.

- Section II presents the problem to be dealt with, defining its representation, objective functions, and constraints.
- Section III presents the sensitivity analysis approach which has been employed to deal with the uncertainty scenario. The uncertainty model employed is discussed in this section too.
- Section IV presents the characteristics and structure of the AIS algorithms developed here. The metric which is used in the developed algorithms is also shown in this section.
- Section V shows the results of the proposed algorithms in two systems.

II. Problem Statement

A. The Distribution System Design

The distribution system can be represented by a planar tree graph in which each possible connection between nodes represents a variable. It is a multibranch graph, once many conductor types can be used to connect the nodes. Two main factors must be considered in optimizing a distribution system:

- Minimization of energy losses.
- Minimization of investment in new facilities and distribution lines.

Besides, some constraints must be considered:

- Line capacity.
- Voltage level in the load buses.
- Graph connectivity (all nodes in the tree must be connected to the root of the graph, directly or indirectly).
- Radiality of the network (the network is a *tree*, or a fully connected graph without loops).

The radiality constraint can be expressed as $c = n - 1$, where c is the number of connections and n is the number of nodes.

An important feature of the objective functions considered in the optimization of distribution networks is the strong interaction they present with respect to changes in network topology and changes in the type of conductors to be used in each connection. This precludes the application of very simple algorithms, such as algorithms that find firstly the "optimal topology", and then the "optimal conductor set" for that topology, in a two-step approach.

The main objective functions mentioned previously can be joined in a single one through the computation of the "present value" of the energy losses along the time horizon [5], [6]. This objective function is presented in (1)

$$f_{mc}(N) = \sum_{j=1}^{m} Y_j^N \left[FC(N_j)\right] + \sum_{t=1}^{at} \left(\sum_{j=1}^{m} Y_j^N \left[MC(N_j) + LC(N_j)\right] \right) (1 - int_{rt})^{t-1} \qquad (1)$$

where

$f_{mc}(N)$	present monetary cost of network N;
m	number of possible connections;
Y_j^N	1 if branch j is present in network N or 0 elsewhere;
$FC(N_j)$	fixed cost of branch j in network N;
at	analysis time;
$MC(N_j)$	maintenance cost of branch j in network N;
$LC(N_j)$	loss cost of branch j in network N;
int_{rt}	interest rate.

Due to text length limitations, it is not considered here an objective function that accounts for a reliability cost, in which case an encoding of alternate paths for energy restoration could be included. Such formulation could be easily adapted to the proposed methodology via a weighted objective function approach, and the same objective function and variable encoding that have been used in [5].

B. Representation of Variables

Each variable represents each possible connection between two nodes of the graph. A graph with n nodes in which any node

[1]A network is feasible if it is able to supply a demand within the technical specifications.

[2]The fault cost is used to estimate the reliability of the network.

[3]A case study presented in the last section of this paper shows that a genetic algorithm is not suitable for developing the procedure proposed here.

can be linked to any other one has m possible connections, with m given by

$$m = \frac{n \cdot (n-1)}{2}. \quad (2)$$

From (2), it can be noticed that the increment of one node in the graph increases the problem dimensionality by n new variables, where n is the number of graph nodes before the node insertion. Problems with a large number of nodes become hard to be solved, due to this combinatorial characteristic, since the search inside this growing variable space is also combinatorial: the number of trees joining the nodes (n_{trees}) increases exponentially with the number of nodes, as shown in (3)

$$n_{\text{trees}} = n^{n-2}. \quad (3)$$

Nevertheless, some characteristics of the problem can be used for reducing the number of variables. Here, the strategy proposed in [5], named Controlled Greedy Encoding, has been adopted. This heuristic comes from the observation that distant nodes are not usually linked in optimal systems, which means that a node can be reasonably linked only to some neighbor nodes. In order to account for rout limitations, some links can be excluded from the encoding.

Therefore, an encoding scheme can be defined by considering just a pre-determined number of nearest connections for each node, without loss of optimal solutions. Only the "possible connections" become encoded as variables. In this encoding scheme, each variable x_i can assume integer values from 0 to *brtp* (where $x_i = 0$ means that the nodes are not connected and values from $x_i = 1$ to $x_i = brtp$ mean that the connection is performed via a conductor of type corresponding to x_i). Thus, each possible tree can be represented by a vector of integer values as shown in (4)

$$\begin{array}{lcccccccc} \text{from}: & & 1 & 1 & \dots & 2 & \dots & n-1 & \\ \text{to}: & & 2 & 3 & \dots & 3 & \dots & n & \\ N & = [& x_1 & x_2 & \dots & x_k & \dots & x_m &] \end{array} . \quad (4)$$

Every node must be connected to at least one other node, due to the connectivity constraint, and this vector must contain exactly $n-1$ nonzero components, due to the radiality constraint. It is straightforward to check these conditions in this vector: every initial candidate solution that is employed for starting the proposed algorithm is subject to these tests, and such candidate solution is accepted only in the case these conditions hold. As the algorithm evolves, the conditions are implicitly kept by the operations that are applied; in this sense, the encoding guarantees that only radial connected trees are considered.

III. Multiobjective Sensitivity Analysis of Solutions

The sensitivity analysis proposed here is based on a feature of the immune-based algorithms used in the optimization process: they can find and maintain some suboptimal solutions together with the current best solution. This sensitivity analysis is performed for changes in the operating conditions of a distribution network, considering each such suboptimal solution as a candidate solution. The reasoning behind the proposed procedure is the following.

- The mean scenario, although being probably wrong when compared with the *a posteriori* real scenario, is the most probable one. It is reasonable to perform the network optimization, for a time horizon, considering this mean scenario.
- The aim is to keep, in addition to the optimal network for the mean scenario, some other ones that are also still near-optimal under this mean scenario. These networks are expected to keep a nearly optimal behavior under the most probable situation and also under several variants of such situation, and are the candidate solution networks.
- It is expected that there will be some deviation from the mean scenario. It is possible that some networks that are good in a single analysis under mean scenario reveal to be very sensitive, with strongly degraded performance, for small variations of this mean scenario. Other networks with similar performances under the mean scenario can still maintain a reasonable performance under perturbed scenarios.

The proposed sensitivity analysis procedure is as follows.

1) Generate a set of scenarios (a Monte Carlo simulation), considering perturbed node loads and energy taxes along the time horizon. These variables should be generated with joint probability distributions that are assumed to be known.

2) Evaluate all candidate solutions in all scenarios, according to four pre-established merit functions: cost of network for nominal conditions; infeasibility rate of solutions; mean financial cost of solutions in feasible scenarios; and mean fault cost of solutions in feasible scenarios.

3) These merit functions are employed to extract a "subset of efficient solutions" from the large set of candidate solutions, based on a *dominance analysis*.

4) Finally, two operations are performed in this "efficient set".

a) check which networks have become infeasible in several scenarios, and discard them;

b) select one solution among the remaining networks, based on the merit functions and on an "expert opinion".

This procedure is explained in the remainder of this paper.

A. Uncertainty Modeling

In the present paper, two sets of parameters are addressed as uncertain variables:

— future load in the nodes.

— energy tax in the nodes.

The main source of uncertainty that affects a network design is the load level in each node. By modeling the energy tax in each system node as an uncertain variable too, it becomes possible to take in account another important uncertainty factor: there is a variable net income per delivered kilowatt hour (kWh), which is different from one node to the other, due to the different costs

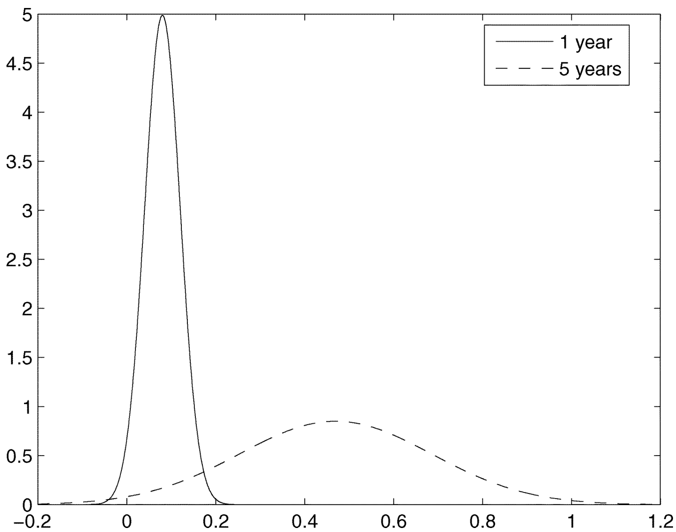

Fig. 1. Example of increase in model uncertainty.

that are incurred in the secondary distribution system that follows each node. This can be used, for instance, for taking in account uncertainties in the geographical distribution of the future loads that will be added to each node. This variability will cause different costs of cable installation and different loses in the secondary system.

A network to be designed has $2n$ uncertain variables (load and the energy tax in each node). Let $V_i \in \mathbb{R}^{2n}$ denote the vector of uncertain variables at year i. The vector V_{i+1} is calculated as

$$V_{i+1} = V_i + \Delta_i. \tag{5}$$

The vector Δ_i is a vector of stochastic variables that represent the increment of each problem variable from one year to the next one. In order to express the dependencies from one variable with the other ones, vector Δ_i is calculated as

$$\Delta_i = D \cdot \delta_i \tag{6}$$

in which δ_i is a vector of "underlying" independent variables and D is a matrix that describes variable dependencies. Vector δ_i is produced, in the Monte Carlo analysis, from within a known probability distribution, leading to Δ_i and to V_{i+1}.

Vector δ_i can have from1 to $2n$ dimensions; its dimensionality can be inferred via a principal component analysis procedure performed on time series data that are representative of the uncertain variables. Matrix D also comes from the principal component analysis procedure.

For the purpose of generating the examples that are presented in this paper, the variables δ_i have been modeled as Gaussian processes with known distributions[4]. The effect of accumulated uncertainty in the variables of V_i, after five years, is illustrated in Fig. 1.

[4]Notice that any other probability distribution could be used without the need of any modification in the proposed procedure.

B. Network Merit Functions

The results of Monte Carlo Simulation are used to evaluate the performance of each solution, considering four relevant criteria.

1) Original cost of network (f_1): this criterion analyzes the cost of the network for the nominal design conditions (most likely scenario).
2) Infeasibility rate of network (f_2): this criterion analyzes the number of times that the network becomes infeasible in the scenarios that have been analyzed, being correlated with a "probability of infeasibility" of the network within the considered time horizon.
3) Mean financial cost of solutions (f_3: this criterion analyzes the mean cost of the networks in the scenarios where they are feasible (notice that this function is, in general, different from f_1, and represents another measure of the expected cost under uncertainty).
4) Mean fault cost of solutions (f_4): this criterion analyzes the costs incurred due to expected network failures, under a given load condition, using the cost function proposed in [15]. This function is evaluated for each scenario in which the network is feasible, and the mean cost for all scenarios is attributed to f_4.

These criteria are expressed as follows:

$$f_1 = CN_{nm}^N \tag{7}$$

$$f_2 = \frac{1}{N_s}\sum_{i=1}^{N_s} FE_i^N \tag{8}$$

$$f_3 = \frac{1}{N_{fs}^N}\sum_{i=1}^{N_{fs}^N} CN_i^N \tag{9}$$

$$f_4 = \frac{1}{N_{fs}^N}\sum_{i=1}^{N_{fs}^N} FN_i^N \tag{10}$$

where

CN_{nm}^N	cost of network N for nominal load conditions (including load value and position);
N_s	number of scenarios analyzed;
FE_i^N	1 if the network N is unfeasible for the load conditions of scenario i or 0 otherwise;
N_{fs}^N	number of scenarios where the network N is feasible;
CN_i^N	cost of network N for the scenario i;
FN_i^N	fault cost of network N for the scenario i [this cost is calculated using (11)]

$$FN^N = \sum_{j=1}^{m} Y_j^N \left[\lambda(N_j)\, \ell(N_j)\, r(N_j)\, P(N_j)\, en_{tx}\right] \tag{11}$$

where

Y_j^N	1 if branch j is present in network N or 0 elsewhere;
$\lambda(N_j)$	failure rate of branch j in network N;

$\ell(N_j)$	length of branch j in network N;
$r(N_j)$	mean fault duration of branch j in network N;
$P(N_j)$	active power of branch j in network N;
en_{tx}	energy tax.

C. Dominance Analysis

The comparison between two candidate solutions, say, networks a and b, is performed in first place via a *dominance analysis*. Given the merit factor vectors $(f_1(a), \ldots, f_4(a))$ and $(f_1(b), \ldots, f_4(b))$, the network a is said to *dominate* b if $f_i(a) \leq f_i(b)$ for $i = 1, \ldots, 4$, and $f_i(a) < f_i(b)$ for at least one $i = 1, \ldots, 4$. This analysis is useful for eliminating all the *dominated* networks that, under the viewpoints of all merit factors, are worse than some other network. A dominated network should never be chosen, since another network that dominates it is always preferable.

The *dominance analysis* is useful in order to discard several solutions which have been initially found by the algorithm, leaving just the "efficient solutions". The networks which have remained after this "filtering" compose the Pareto-set of the sensitivity analysis.

IV. Proposed Algorithm

Immune-based algorithms rely only on mutation operators, which represent the variation mechanism that creates solution diversity within the search mechanism. In this section, a definition of a distance metric for networks and the distance-based operator supported by this metric are firstly presented. Then, the Artificial Immune System (AIS) algorithm developed here is presented: the Clonal Selection Algorithm for Distribution Networks (CSA-DN) algorithm.

A. Network Distance Metric

Consider a network N_A with n nodes, encoded with m variables (m possible connections). Such a network can be represented as an embedded point in the *vector space* $\mathbb{R}^m$ using the following expression:

$$\overrightarrow{N_A} = \sum_{i=1}^{m} \left[w_i^A \, p_1 + N_A(i) \, w_i^A \, p_2 \right] \overrightarrow{e_i} \tag{12}$$

where

$\overrightarrow{N_A}$	vector which represents network N_A in space $\mathbb{R}^m$;
$N_A(i)$	ith component of the vector of integer variables that encodes the network N_A, as described in (4);
w_i^A	weight of connection i in network N_A – this weight is zero if the connection does not exist, and a number such that $0 < w_i^A \leq 1$: larger if the connection is near the root of the tree, and smaller if the connection is far from the root;
p_1	factor that weights only the position of the connection in network (since it multiplies only w_i^A);
p_2	factor that weights both the position in network (since it multiplies w_i^A) and the type of branch (since it also multiplies $N_A(i)$). This factor is smaller than p_1, making the topological information become more important than the branch type information;
$\overrightarrow{e_i}$	ith vector of canonical base of $\mathbb{R}^m$.

The main idea of expression (12) is to express the fact that connections near to the network root have stronger influence on network characterization than connections that are near to the network end-nodes; this is performed by weight w_i^A. This means that any change in a connection near the network root causes a stronger change in the vector that represents the network than a change in a connection near a network end-point. Notice that this corresponds to what is expected to occur in an electrical network that is connected as a tree: the power flow receives a stronger perturbation due to changes that occur in the branches near the root than the perturbation that would occur due to changes being performed near the network endpoints. In this sense, the proposed distance metric is related to a "power flow similarity" between networks. The factors p_1 and p_2 must be adjusted such that networks with different topologies become "more different" than networks with the same topology, but still guaranteeing that networks with the same topology and different types of branches be represented by different vectors.

Expression (12) embeds a network into the vector space $\mathbb{R}^m$, making the set of networks inherit the properties of such space. For instance, a difference vector (relative position) is defined as

$$\begin{aligned} &\overrightarrow{r_{pos}(N_A, N_B)} \\ &= \sum_{i=1}^{m} \left[p_1 \left(w_i^A - w_i^B \right) + p_2 \left(N_A(i) w_i^A - N_B(i) w_i^B \right) \right] \overrightarrow{e_i}. \end{aligned} \tag{13}$$

Using the inner product that is inherited by this vector representation of networks, the scalar that measures the distance between these two networks can be calculated using an Euclidean norm, as shown in (14)

$$\begin{aligned} &dist(N_A, N_B) \\ &= \left\{ \sum_{i=1}^{m} \left[p_1 \left(w_i^A - w_i^B \right) + p_2 \left(N_A(i) w_i^A - N_B(i) w_i^B \right) \right]^2 \right\}^{1/2}. \end{aligned} \tag{14}$$

It is straightforward to verify that this "distance between networks" fulfills all the properties that any proper definition of a distance must have, such as the triangle inequality.

This representation presents interesting features for network optimization, since it can be used to establish concepts of local, global and directional searches. These concepts can be used to improve the space exploitation, and consequently increase the algorithm efficiency.

B. Mutation Operator

The network distance metric, described previously, has been employed for building an operator which can generate networks that are at a prescribed distance from an initial network. This operator allows direct implementations of distance-based AIS algorithms (which are usually conceived for optimization in continuous-variable spaces) without structural changes, just replacing the real-space norm by the network distance metric.

A mutation operator that is used in AIS algorithms for continuous problems in $\mathbb{R}^n$ should generate a new candidate solution ω_{k+1} from a current candidate solution ω_k such that $\omega_{k+1} = \omega_k + \rho$, with ρ being a random vector with a random norm r generated with pre-defined mean. An analogous operation can be defined as the basis of a mutation operator for dealing with networks, based on the metric (14).

Let r be the mutation radius, generated randomly by the algorithm. A new network is to be generated randomly at that distance r, in relation to the initial network N_0. The proposed operator works by increasing the distance by steps, with incremental modifications in N_0 that are denoted by $N_0^1, N_0^2, \ldots, N_0^k$, such that $dist(N_0, N_0^k) \approx r$. Each step is composed of two substeps.

1) A new branch is randomly inserted in the current network N_0^i

2) Another branch is removed to maintain the radiality constraint, leading to the next network N_0^{i+1}. This branch is chosen randomly from the ones that compose the "closed loop" that was introduced in the network by the previous step.

It is straightforward to notice that except in the rare situation in which a random change is reversed in a next step, it holds that $dist(N_0, N_0^i) < dist(N_0, N_0^{i+1})$. This sequence of operations is performed until $|dist(N_0, N_0^k) - r| < \epsilon_1$ for a given rough tolerance ϵ_1. When such tolerance is reached, the branches which are common in both networks have their type changed, in order to make a "fine tuning", obtaining a distance that is closer to r, up to a tighter tolerance ϵ_2.

C. Description of the CSA-DN

The clonal selection principle states that those defense cells[5] (candidate solutions) with higher affinity to the antigen (with better objective function values) have greater proliferation rates. The cells are submitted to a maturation affinity process, that starts with the generation of several clones of each cell, followed by the mutation of such clones. The mutation distance of each cell clone is generated randomly, from a Gaussian distribution, with both the mean and standard deviation of such distribution inversely proportional to the affinity of that cell. This means that the best cells have their clones probably mutated into nearer new cells, in this way performing mainly a local search, while the worst cells have their clones probably mutated into farther new cells, in a mechanism that performs a global search. Hence, the maturation process can be viewed as a global search with local refinement mechanism, which provides defense cells that bind the antigen (minimize the objective function) more efficiently.

[5] *Antibodies*, in a biological analogy.

An off-line population named memory population has been associated to the CSA-DN to improve the mapping of suboptimal solutions (this memory population is usually employed in immune network algorithms [16]). The cells in this memory population are compared to each other, for verifying the degree of similarity or recognition among them. In typical immune algorithms for real-valued functions, a similarity function for binary strings (binary coding) or the Euclidean distance for real vectors (real coding) can be employed. For network optimization the distance metric in (14) is used for determining the degree of recognition between two cells. The CSA-DN employs a suppression mechanism over the memory population in order to eliminate redundancy in this set: only the best cell is maintained, from a group of similar cells (those with $dist(N_A, N_B)$ under a given threshold σ); the other ones are eliminated.

Based on such principles a simple optimization algorithm is proposed as follows.

1) A population of N cells is generated.

2) An affinity measure for each cell is evaluated by the objective function.

3) A maturation affinity process is started, what generates more clones for the best cells.

4) Following the maturation affinity process, a mutation is applied to each clone, keeping higher mutation distances for the clones of the worst cells.

5) The worst cells are replaced by new randomly generated ones, in order to maintain diversity.

6) The solutions of current population of cells are stored in a memory population and the suppression mechanism is applied.

7) Go to step 2 until a desired stop criteria is met.

There are many different ways of implementing those basic ideas and the implementation may depend on the characteristics of the specific problem being solved. Here, the CSA-DN is presented as an adaptation of the general algorithm described above. The distance metric shown in Section IV-A has been used.

Steps 1 and 2 are trivial. For implementing step 3, it is necessary to introduce a parameter B that defines the fraction of the set of cells to be selected for cloning ($0 \leq B \leq 1$). An affinity measure of each cell is computed using (1), and all cells are sorted from the best to the worst one. The $B \cdot N$ best cells are selected for cloning. The number of clones for each cell is given by

$$C_i = \text{round}\left(\frac{\beta \cdot N}{i}\right), \quad i = 1, \ldots, (B \cdot N) \tag{15}$$

where N is the size of the population, and β is a constant ($0 \leq \beta \leq 1$). As an example, for $B = 0.5$, $\beta = 1$, and $N = 50$, the best cell ($i = 1$) will receive 50 clones, the second best cell ($i = 2$) will receive 25, and so forth until $i = 25$.

Once the number of clones for each cell becomes determined, C_i new cells (candidate networks) are generated using the mutation operator described before. For the clones of the best original cells, the mutations produce new cells that are closer to the original cell, and for the clones of the worst cells, the new cells are generated farther from the original one. Finally, all cells not selected for cloning are replaced by new randomly generated networks. It should be noticed that this algorithm performs global search, through the diversity generation mechanism and also local search, through the maturation affinity process. Since CSA-DN does not rely on combination operators, like the crossover in genetic algorithms, the population is maintained disperse through the search space and is not forced to concentrate around an attraction basin. Therefore, CSA-DN is capable of providing several different local solutions at the final of the search process. CSA-DN has still other advantages: it is very simple to implement and presents few parameters to adjust, namely, N, B, and β.

D. *Constraint Handling*

The constraint handling in artificial immunology based algorithms can be performed using the same strategies usually adopted in genetic algorithms. In the present problem, three different strategies have been employed to handle with the four constraints discussed in Section II.

- Graph connectivity and radiality: the mutation operator employed here always generates feasible solutions (tree connected graphs). Therefore, these constraints are never violated.
- Line capacity: if one of the network branches cannot comply with the power demand, then this branch is replaced by another one, with higher capacity. If this replacement is not possible (the branch with highest capacity cannot supply the demand for that structure) then the solution affinity receives a penalty factor.
- The voltage level in load buses is analyzed after the line capacity, with the branches already fixed in that step. If the voltage level is bellow (or above) the standards, the solution affinity receives a penalty factor.

It is important to notice that these constraint handling strategies are employed only within the optimization process. At the end of the optimization procedure performed by the CSA-DN, all delivered solutions are feasible under the mean scenario. In the multiobjective sensitivity analysis, the solutions that could not comply with those constraints in some scenarios are just considered infeasible in such scenarios; the fixing process is not used in this stage.

V. Results

A. *Benchmark System*

The first case analyzed has the only purpose of establishing the ability of the CSA-DN for competitively finding the optimal solutions of electric distribution networks, still in the case of no system uncertainty. This is performed via a comparison with an example extracted from a previous work [7]. The system is a 23 node system, and the encoding has been based on the possible connections considered in [7], resulting in a problem with 35 variables (Fig. 3). The design time considered has been of 20 years and two possible conductors have been considered in the design.

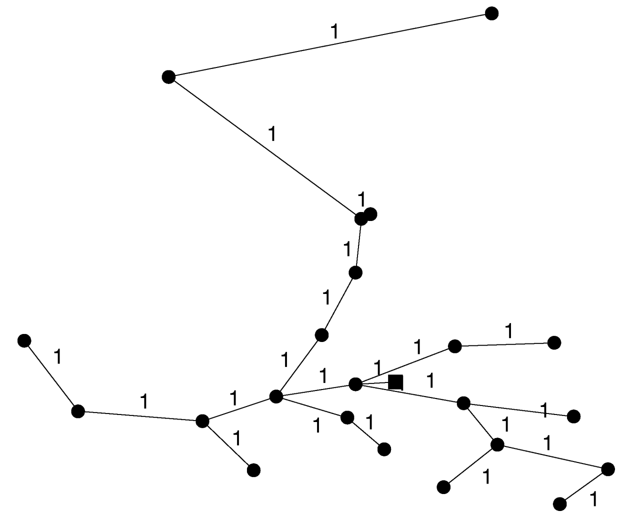

Fig. 2. Benchmark system: best solution.

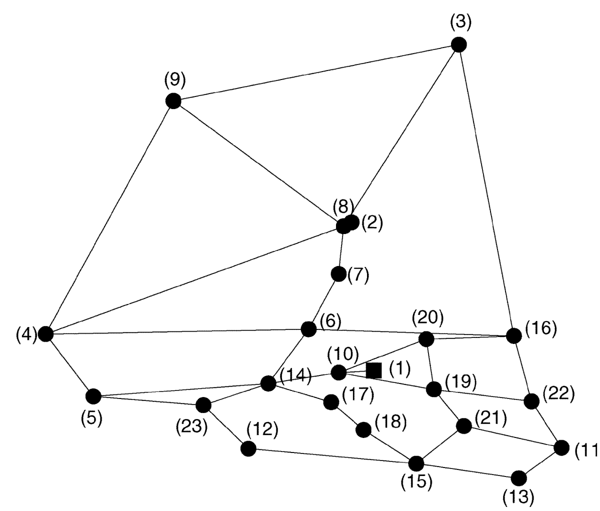

Fig. 3. Benchmark system: encoding.

The CSA-DN has been executed 50 times to evaluate the convergence of algorithm. The relevant issue to be shown is that, in all runnings, it has converged to a solution which is better than the best one found in [7] (Fig. 2). This solution has a cost of $171 698.03 and has been met using 16 909 function evaluations in average.

B. *The 21-Node System*

In the second example, the proposed algorithm has been applied to the optimization of a test problem composed of a 21-node system (62 variables), considering now load uncertainty and energy tax uncertainty. This system has been used in [5], in the context of a multiobjective optimization procedure.

The time horizon considered here has been of ten years and nine possible conductor types have been allowed in design. The whole data set related to this case can be found in [17].

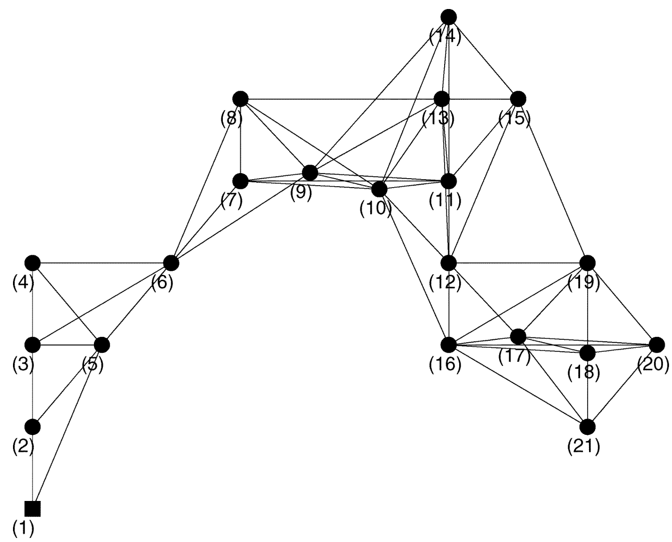

Fig. 4. The 21-node system: encoding.

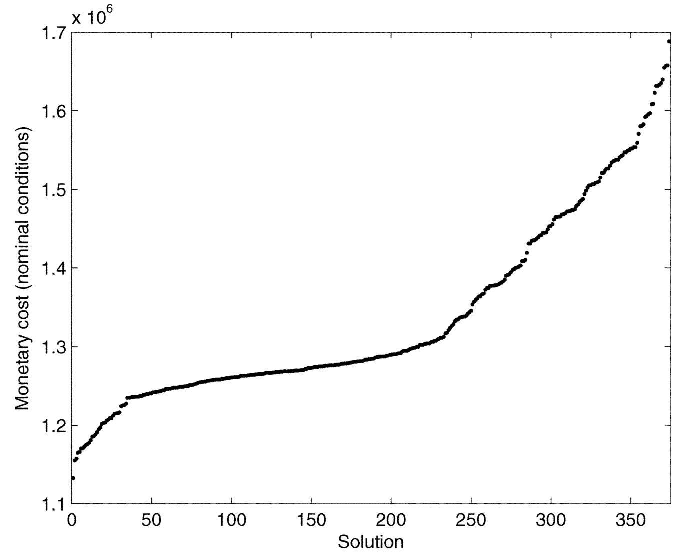

Fig. 6. The 21-node system: CSA-DN solutions.

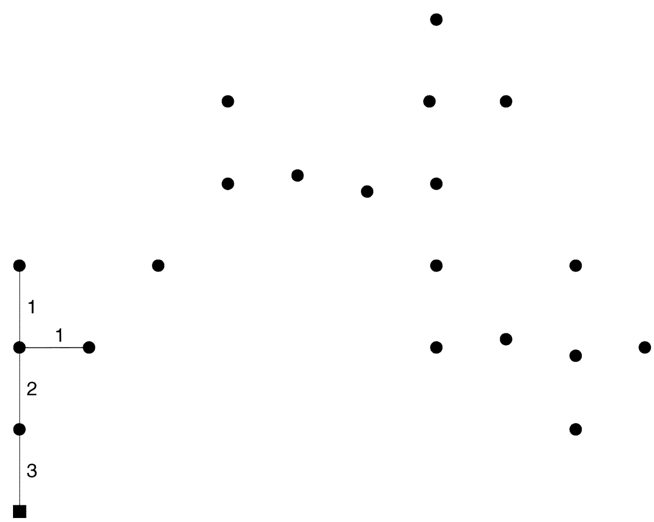

Fig. 5. The 21-node system: initial network.

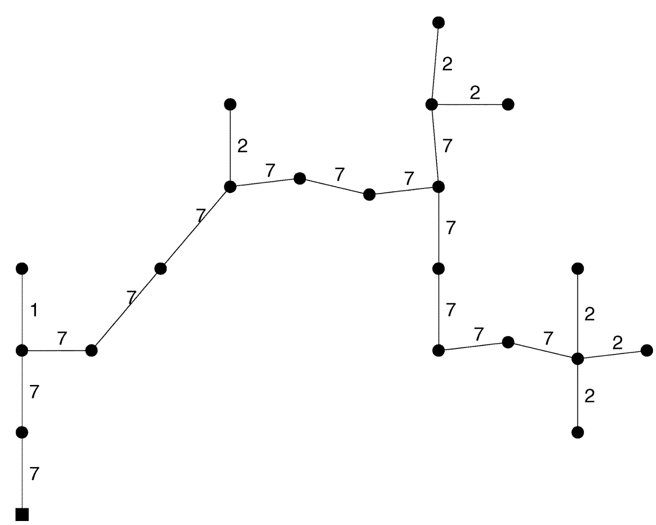

Fig. 7. The 21-node system: best solution for nominal conditions ($1a$).

Here, the encoding scheme has been found using the *Controlled-Greedy Encoding*, proposed in [5]. The resulting encoding is shown in Fig. 4 and contains 62 possible connections. In this case an initial system has been considered along the optimization process; it is shown in Fig. 5.

The CSA-DN algorithm, with memory population, has mapped 374 solutions as shown in Fig. 6. Fig. 7 shows the best solution achieved for nominal conditions (this solution will be referred to as $1a$), that has a cost of \$1 132 799.55.

C. Multiobjective Sensitivity Analysis

The 374 solutions achieved by CSA-DN have been submitted to a Monte Carlo Simulation performed for 2000 scenarios. These scenarios have been generated according to Gaussian probability distributions. The results of Monte Carlo Simulation have been used to evaluate each solution following the criteria shown in Section III. A Pareto-set has been built with the nondominated solutions. The parameters of probability distributions are shown in Table I. The mean load growth rates and energy tax variation have been considered equal for all nodes. This has been adopted for simplicity; as discussed in Section III, different probability distributions could be used for each node.

TABLE I
PROBABILITY DISTRIBUTION PARAMETERS

Load growth rate:	**01 year**	**10 years**
mean:	0.050	0.629
std:	0.025	0.280
Energy tax variation:	**01 year**	**10 years**
mean:	0.000	0.000
std:	0.050	0.629

For these parameters, 41 of the 374 solutions have not been dominated by any other solution. Table II shows the performance of the efficient solutions in each one of the analysis criteria. Additional information about these solutions, including

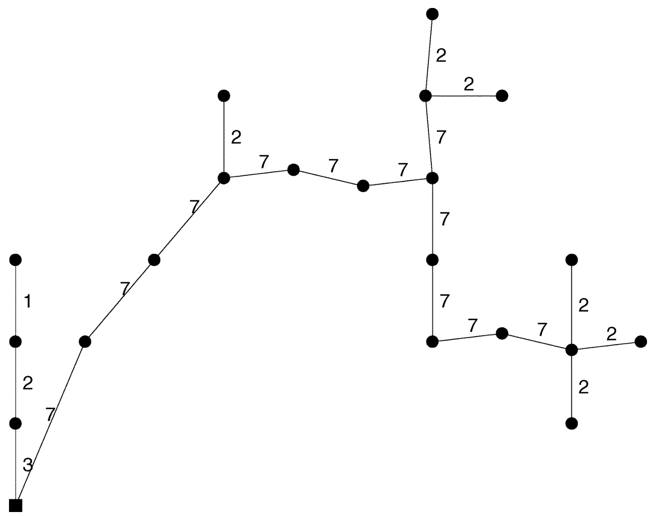

Fig. 8. The 21-node system – solution $3a$.

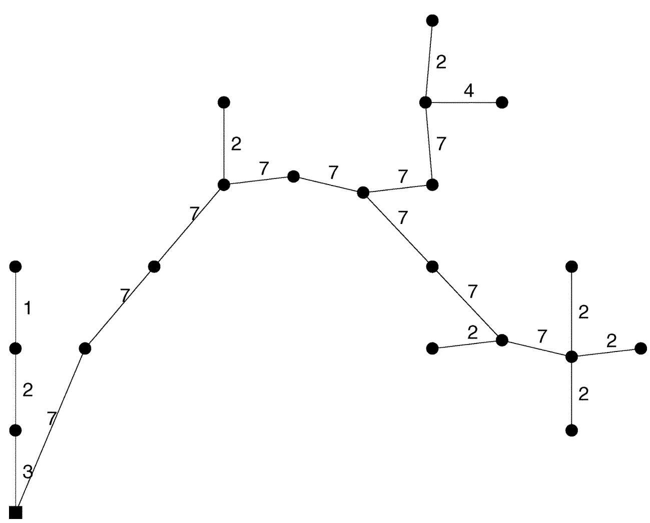

Fig. 9. The 21-node system – solution $17a$.

TABLE II
CSA-DN – NONOMINATED SOLUTIONS

Sol.	f_1	f_2	f_3	f_4
$1a$	1132799.55	0.7286	1244216.18	1831968.35
$2a$	1155002.59	0.6992	1211798.82	1726380.49
$3a$	1157128.93	0.0035	1226846.20	1736293.99
$4a$	1164948.03	0.6967	1221791.14	1718656.74
$5a$	1165814.14	0.6952	1222324.53	1691850.80
$6a$	1170425.18	0.6856	1224558.65	1641068.99
$10a$	1175613.79	0.6856	1229825.58	1587254.26
$17a$	1194965.91	0.0035	1261480.42	1689429.56
$19a$	1201459.76	0.7291	1263077.08	1551298.81
$25a$	1208758.56	0.7423	1273264.87	1549835.29
$34a$	1227517.01	0.7403	1291989.34	1550149.19
$50a$	1240780.79	0.7787	1310098.86	1298258.54
$51a$	1241676.00	0.7787	1310947.16	1298215.04
$65a$	1247486.58	0.7681	1314482.80	1209464.15
$66a$	1247595.64	0.8324	1300151.79	1379186.02
$67a$	1248171.33	0.0025	1311512.42	1683388.93
$69a$	1248872.96	0.7661	1315515.42	1176593.07
$73a$	1250052.95	0.7651	1316742.28	1176784.06
$104a$	1261099.73	0.2699	1338725.00	1585951.92
$117a$	1264890.36	0.7023	1321124.30	1558351.26
$132a$	1268127.01	0.6992	1323606.86	1500580.70
$134a$	1268529.06	0.7651	1335090.79	1176736.84
$139a$	1269184.29	0.5575	1337490.17	1567388.26
$141a$	1269425.84	0.6992	1325327.51	1448208.43
$151a$	1273493.81	0.7078	1331363.83	1438844.68
$206a$	1291271.35	0.8770	1354175.85	1133528.79
$210a$	1294811.74	0.8704	1356947.20	1046831.37
$232a$	1311185.53	0.7590	1366745.50	1316450.89
$233a$	1312146.27	0.7590	1367752.68	1296170.80
$237a$	1323571.48	0.7635	1381889.55	1183375.30
$238a$	1326310.50	0.7332	1385793.22	1406853.41
$241a$	1334298.56	0.7119	1390992.27	1359490.32
$250a$	1345556.52	0.2587	1417873.43	1150903.27
$252a$	1356879.30	0.2734	1431309.07	1139807.09
$253a$	1359283.93	0.2724	1432832.30	1139571.50
$262a$	1377050.62	0.2527	1454658.83	1674244.40
$265a$	1377741.42	0.3605	1455127.69	1115444.35
$266a$	1378220.24	0.2527	1455891.12	1652436.58
$293a$	1441345.88	0.2162	1511650.22	1454593.24
$301a$	1455851.51	0.1808	1520748.97	1511240.00
$309a$	1469515.97	0.8223	1531425.87	818023.63

total losses, maximum voltage drop and total length of cables, can be found in [17].

D. Analysis of Results

As discussed in Section III, the solutions with high infeasibility rates should not be considered, since they are not robust enough to comply with most of possible scenarios. One should note that the network with minimum cost for nominal conditions ($1a$) has an infeasibility rate greater than 70%, and should be considered nonrobust.

Taking an infeasibility rate lower than 20% as the robustness threshold that would be still acceptable, the set of candidate solutions becomes reduced to four networks only: $3a$, $17a$, $67a$ and $301a$. Since this number of networks is quite small, they can be easily analyzed by an experienced designer. The authors believe that the solutions $3a$ and $17a$ (Figs. 8 and 9, respectively) are good choices for this system. These solutions have a nominal cost slightly greater than $1a$ (about 2% and 5% respectively) and outperform $1a$ in all other criteria.

The data in Table II shows an interesting pattern of conflict between the objective of network feasibility and the objective of reliability: all the four indicated solutions, which have high feasibility, have also high cost of reliability. In particular, if the system reliability is very important, then solution $301a$ could be chosen instead of solution $3a$ or $17a$. Also, if the infeasibility rate could be relaxed to 26%, solution $250a$ could be chosen, with a reliability cost about 30% better than solutions $3a$ and $17a$, but at a financial cost about 20% greater.

E. Population Diversity – GA Comparison

The GA proposed in [1] has been used in the same problem, to establish a comparison with the CSA-DN. The final population of GA, after the employment of the suppression mechanism[6], has been defined as the set of candidate solutions. Then this set, which has 13 solutions, has been submitted to the same multi-objective sensitivity analysis described earlier in this paper, for

[6]This is the same suppression mechanism employed in memory population of CSA-DN.

TABLE III
GA – NONDOMINATED SOLUTIONS

Sol.	f_1	f_2	f_3	f_4
$1b$	1132799.55	0.7280	1244439.22	18328346.32
$2b$	1155002.59	0.6985	1211967.69	17269902.72
$3b$	1170425.18	0.6850	1224697.18	16416035.96
$6b$	1177197.19	0.7015	1234459.51	16098496.83
$10b$	1317687.58	0.7635	1373407.11	12962783.36
$11b$	1337903.39	0.2965	1412298.59	19579966.26
$12b$	1377400.40	0.0040	1453732.57	18880288.24

the same probability distribution parameters. The efficient set achieved is composed of seven solutions, as shown in Table III[7].

One should note that the solution $12b$ is the only one with infeasibility rate lower than 0.20. It is noticeable that this solution has a poor performance compared to the solutions that have been suggested as reasonable choices among the outcome set of CSA-DN, $3a$ and $17a$. Those networks outperform $12b$ in all analysis criteria: the only acceptable GA solution, $2b$, would be a dominated solution in the set achieved by CSA-DN.

VI. CONCLUSION

This paper has proposed a methodology for designing electric distribution networks that can take into account the load-evolution uncertainty in the considered design time horizon. An immune-based evolutionary optimization algorithm has been developed in order to generate a set of nearly-optimal solutions. These solutions are submitted to a Monte Carlo simulation that are performed under the load-uncertainty probability distribution in each node. The effect of such uncertainty is then evaluated, and the candidate solutions are compared via a multiobjective analysis.

This procedure has been shown to be effective and, particularly, has revealed that the conventional procedures that consider only the mean scenario of load growth can be "fragile" under uncertainty. Such conventional design, when compared to the proposed methodology outcomes, tend to be either more costly (in the case of a conservative design that over-estimates the load growth) or more likely to become infeasible (in the case of a mean-scenario minimal cost design).

REFERENCES

[1] E. G. Carrano, R. H. C. Takahashi, E. P. Cardoso, O. M. Neto, and R. R. Saldanha, "Optimal substation location and energy distribution network design using an hybrid GA-BFGS algorithm," *IEE Proc. Generation, Transmission and Distribution*, vol. 152, no. 6, pp. 919–926, 2005.

[2] V. Miranda, J. V. Ranito, and L. M. Proença, "Genetic algorithms in optimal multistage distribution network planning," *IEEE Trans. Power Syst.*, vol. 9, no. 4, pp. 1927–1933, Nov. 1994.

[3] P. M. S. Carvalho, L. A. F. M. Ferreira, F. G. Lobo, and L. M. F. Barruncho, "Optimal distribution network expansion planning under uncertainty by evolutionary decision convergence," *Elect. Power & Energy Syst.*, vol. 20, no. 2, pp. 125–129, 1998.

[4] H. L. Willis, H. Trans, and L. F. M. V. Engel, "Selecting and applying distribution optimization methods," *IEEE Comput. Appl. Power*, vol. 9, no. 1, pp. 12–17, Jan. 1996.

[5] E. G. Carrano, L. A. E. Soares, R. H. C. Takahashi, R. R. Saldanha, and O. M. Neto, "Electric distribution multiobjective network design using a problem-specific genetic algorithm," *IEEE Trans. Power Del.*, vol. 21, no. 2, pp. 995–1005, Apr. 2006.

[7]The index b indicates the solutions achieved with the GA.

[6] I. J. Ramirez-Rosado and J. L. Bernal-Augustín, "Genetic algorithms applied to the design of large power distribution systems," *IEEE Trans. Power Syst.*, vol. 13, no. 2, pp. 696–703, May 1998.

[7] J. Goméz, H. Khodr, P. Oliveira, L. Ocque, J. Yusta, R. Villasana, and A. Urdaneta, "Ant colony system algorithms for the planning of primary distribution circuits," *IEEE Trans. Power Syst.*, vol. 19, no. 2, pp. 996–1004, May 2004.

[8] G. Duan and Y. Yu, "Power distribution system optimization by an algorithm for capacited Steiner tree problems with complex-flows and arbitrary cost functions," *Elect. Power & Energy Syst.*, vol. 25, pp. 515–523, 2003.

[9] V. Parada, J. A. Ferland, and M. A. K. Daniels, "Optimization of electrical distribution feeders using simulated annealing," *IEEE Trans. Power Del.*, vol. 19, no. 3, pp. 1135–1141, Jul. 2004.

[10] H. L. Willis and J. E. D. Northcode-Green, "Spatial electric load forecasting: A tutorial review," *Proc. IEEE*, vol. 71, no. 2, pp. 232–253, Feb. 1983.

[11] D. M. Morsi, N. H. Abbasy, and M. S. Abul Ella, "A hybrid expert system assisting decision making for distribution system load forecasting," in *Proc. 7th Mediterranean Electrotechnical Conf.*, Antalya, Turkey, Apr. 1994, pp. 893–896.

[12] S. Pierre, "Application of artificial intelligence techniques to computer network design," *Eng. Applic. Artific. Intell.*, vol. 6, no. 5, pp. 465–472, 1993.

[13] L. N. de Castro and F. J. Von Zuben, "Learning and optimization using the clonal selection principle," *IEEE Trans. Evol. Comput.*, vol. 6, no. 3, pp. 239–251, Jun. 2002.

[14] L. N. de Castro and J. Timmis, "Artificial immune systems as a novel soft computing paradigm," *Soft Comput.*, vol. 7, no. 8, pp. 526–544, Aug. 2003.

[15] J. L. Bernal-Agustín, "Aplicación de Algoritmos Genéticos al Diseño óptimo de Sistemas de Distribución de Energía Eléctrica," Ph.D. dissertation, Zaragoza Univ., Zaragoza, Spain, 1998.

[16] L. N. de Castro and J. Timmis, "An artificial immune network for multimodal function optimization," in *Proc. IEEE Cong. Evolutionary Computation*, May 2002, vol. 1, pp. 699–704.

[17] E. G. Carrano, F. G. Guimaraes, R. H. C. Takahashi, O. M. Neto, and F. Campelo, "Electric Distribution Network Expansion Under Load Evolution Uncertainty Using an Immune System Inspired Algorithm: Simulation Data," Tech. Rep. Universidade Federal de Minas Gerais, 2007 [Online]. Available: http://www.mat.ufmg.br/~taka/techrep/aisnet01.pdf

Eduardo G. Carrano was born in Brazil in 1981. He received the B.Sc. degree in electrical engineering from the Universidade Federal de São João Del Rei in 2003. Since March 2004, he has been pursuing the Ph.D. degree at the Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.

He is currently performing part of his research as a Visiting Student at the Universidade do Algarve, Faro, Portugal. His main research interests are in network design, evolutionary algorithms, multiobjective combinatorial optimization, and optimization theory.

Frederico G. Guimarães was born in Brazil in 1979. He received the B.Sc. and M.Sc. degrees in electrical engineering in 2003 and 2004, respectively, from the Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, where he has been pursuing the Ph.D. degree since August 2004.

He is currently performing part of his research as a Visiting Student at McGill University, Montreal, QC, Canada. His main research interests are in the field of optimization, including multiobjective optimization and evolutionary algorithms. He is also interested in applied optimization, particularly the design of electromagnetic devices and systems.

Ricardo H. C. Takahashi was born in Brazil in 1965. He received the B.Sc. and M.Sc. degrees in electrical engineering from Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brazil, in 1989 and 1991, respectively, and the Dr. degree in electrical engineering from the University of Campinas (UNICAMP), Campinas, Brazil, in 1998.

He is currently is an Associate Professor with the Department of Mathematics, UFMG. His main research interests are in the fields of optimization theory (including the theory of evolutionary computation) and control theory.

Oriane M. Neto was born in Brazil in 1961. He received the B.Sc. and M.Sc. degrees in electrical engineering from the Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brazil, in 1985 and 1989 respectively, and the Ph.D. degree from Imperial College of Science, Technology and Medicine, London, U.K., in 1996.

He was with the Department of Electrical Engineering, Tokyo Metropolitan University, Tokyo, Japan, from September 2001 to September 2002 as a Visiting Professor. He is currently an Associate Professor with the Department of Electrical Engineering, UFMG. His main research interests are in power system dynamics and power system optimization and control.

Felipe Campelo (S'04) received the B.E.E. degree from the Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brazil, in 2003, and the M.E. degree in 2006 from Hokkaido University, Hokkaido, Japan, where he is currently pursuing the Ph.D. degree at the Graduate School of Information Science and Technology, where he is a Monbusho Scholarship student.