

# A probabilistic methodology for distributed generation location in isolated electrical service area

H.M. Khodr, Marco R. Silva, Zita Vale, Carlos Ramos

## ABSTRACT

Distributed generation unlike centralized electrical generation aims to generate electrical energy on small scale as near as possible to load centers, interchanging electric power with the network. This work presents a probabilistic methodology conceived to assist the electric system planning engineers in the selection of the distributed generation location, taking into account the hourly load changes or the daily load cycle. The hourly load centers, for each of the different hourly load scenarios, are calculated deterministically. These location points, properly weighted according to their load magnitude, are used to calculate the best fit probability distribution. This distribution is used to determine the maximum likelihood perimeter of the area where each source distributed generation point should preferably be located by the planning engineers. This takes into account, for example, the availability and the cost of the landlots, which are factors of special relevance in urban areas, as well as several obstacles important for the final selection of the candidates of the distributed generation points. The proposed methodology has been applied to a real case, assuming three different bivariate probability distributions: the Gaussian distribution, a bivariate version of Freund's exponential distribution and the Weibull probability distribution. The methodology algorithm has been programmed in MATLAB. Results are presented and discussed for the application of the methodology to a realistic case and demonstrate the ability of the proposed methodology for efficiently handling the determination of the best location of the distributed generation and their corresponding distribution networks.

## Keywords

Distributed generation, Probabilistic methodology, Source location

## 1. Introduction

Distributed generation (DG) represents a change in the paradigm of electrical energy generation. The emergence of new technological alternatives (photovoltaic systems, wind power, cogeneration, etc.) allows to generate part of the required energy closer to the places of consumption, improving quality levels and minimizing the investments costs associated with transmission and distribution systems [1]. In general terms, the development of DG plants aims at improving environmental aspects and quality of energy (uninterrupted provision of electrical energy with suitable voltage level, current, frequency, amongst others) [1–3]. These generation types can benefit from a decision-support tool able to assist decision making in the scope of a project of installation of DG plants. In a first stage, a programming tool determines the location of the DG point's candidates. After this, we can proceed to optimize the capacities, number of units and technology to be implemented.

The objective of this research work is determining the region of higher probability for location of DG plants that will feed the loads under study. These loads have been previously estimated [4–6].

In order to determine the region of greater probability for location of DG plants, a probabilistic methodology, previously used for the optimal location of electric substations, will be adapted [7,8]. To achieve this objective, three distribution probability functions are used: normal, Freund's bivariate exponential distribution and Weibull. The obtained solution must fulfil the parameters of reliability and electrical power quality for the study developed in this work. It should be noticed that this methodology considers the fact that the load center for a group of consumers varies with time [7].

The decision of the location of DG plants is a crucial aspect in the model, with a strong impact on the operation and investment costs.

In [9] an algorithm using primal-dual interior point method for solving non-linear optimal power flow problems was proposed. The main purpose is to optimize location and sizing of DG on distributed systems for solving the problem of line loss reduction. Most of the benefits of employing DG in existing distribution networks have both economic and technical implications and they are interrelated.



In [10], a general approach is presented to quantify the technical benefits of DG. A set of indices is proposed to quantify some of the technical benefits of DG. They are:

- voltage profile improvement index;
- line loss reduction index;
- environment impact reduction index;
- DG benefit index.

The basic idea behind the above proposed indexes' is to compare the voltage profile, the total line losses, and the emission of a particular pollutant with and without the employment of DG.

In [11], a multi-objective formulation for the sitting and sizing of DG resources into existing distribution networks has been proposed. This method permits the planner engineers to decide the best compromise between costs of network upgrading, power losses and energy not supplied. A genetic algorithm and  $\epsilon$ -constraint method have been proposed that allows obtaining a set of non-dominated solutions.

In [12] a multi-objective performance index-based size and location of DG in distribution systems with different load models has been presented. A constant real and reactive power load model is assumed. In this reference, it is shown that the load models can significantly affect the optimal location and sizing of DG resources in distribution systems.

In [13] an analytical method has been presented to determine the optimal location to place DG in radial as well as networked systems to minimize the power loss of the system. In [14] the analytical method is only proposed for sizing and location of DG in radial distribution systems. In this study, a loss sensitivity factor, based on the equivalent current injection, is formulated for distribution systems. The formulated sensitivity factor is employed for the determination of the optimum size and location of DG so as to minimize total losses by the analytical method without use of admittance matrix, inverse of admittance matrix or Jacobian matrix. It is shown that, the proposed method is in close agreement with the classical grid search algorithm based in successive load flows.

Ref. [15] presents a genetic algorithm-based method to determine optimal location and size of the DG to be placed in radial, as well as networked, systems with an objective to minimize the power losses. Several simulation studies have been conducted on radial feeders, as well as networked systems, considering single-distributed generation and multiple-distributed generations separately to minimize the power loss of the system subjected to no voltage violation at any of the distribution network buses.

The location of the DG site depends upon several factors, such as the voltage level, voltage regulation, DG cost and power losses level. In general, some considerations should be taken into account for the selection of the candidate points where the DG will be located, such as [4,16]:

- The DG system must be as close as possible to the load center of its service area, in order to reduce the voltage drop, power losses and the product of the load and the distance from the DG source.
- Proper voltage regulation should be obtainable without taking extensive measures.
- Proper access for the future interconnection with the distribution or transmission electrical networks should be provided, considering future growth.
- The DG fixing project must comply with land regulations, local ordinances and neighbors.
- The DG installation project should reduce the number of customers affected by any unavailability of electricity supply.

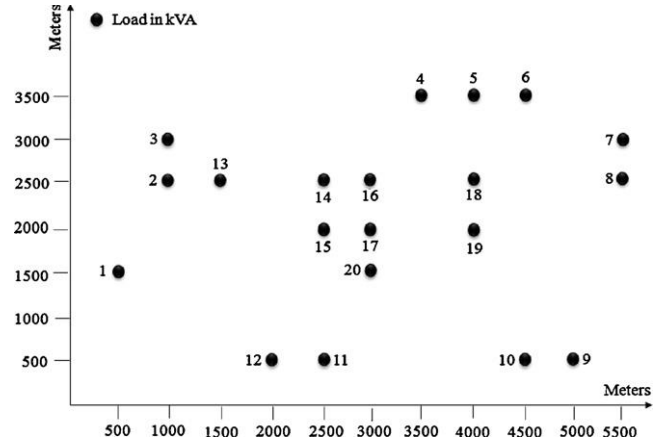


Fig. 1. Geographical locations on the X-Y plane of the loads.

- Other considerations, such as adaptability, emergency, etc., should also be taken into account.

This paper proposes a new probabilistic methodology conceived to assist the electric system planning engineers in the selection of the DG location, taking into account the hourly load changes or the daily load cycle. These location points, properly weighted according to their load magnitude, are used to calculate the best fit probability distribution. This distribution is used to determine the maximum likelihood perimeter of the area where each DG source point should preferably be located by the planning engineers. When all the load centers are determinate, then the planning engineers can better calculate the load nodes that can be feed by each source or DG by the simple calculation of the product ( $\text{kVA} \times \text{distance}$ ) minimum. The algorithm is stopped when the sum of the load nodes is lower or equal than the capacity of the source or DG, taking into account a certain reserve for the future grows of the load.

This paper is organized in the following way. Section 2 expands the probabilistic methodology used to locate the DG and the proposed statistical distributions. After that, Section 3 presents the test system, the application of the three distributions probabilities to a realistic case and the main results of the best location of the DG and their distribution networks. Finally, Section 4 states the conclusions.

## 2. Proposed statistical distribution

The proposed calculation procedure starts with the application of a deterministic methodology to a given set of load curves given in 1 h periods, associated to a set of points in the plane (as presented in Fig. 1).

The set of hourly load centers is calculated using expressions (1) and (2) where  $X_i$  and  $Y_i$  are the coordinates of the load center;  $S_{ij}$  is the apparent power of the load curve  $j$  for hour  $i$ ;  $nh$  is the number of hours for the discretization of the load curves;  $nc$  is the number of load curves.

For hour  $i$ , we have:

$$X_i = \frac{\sum_{j=1}^{nc} S_{ij} X_j}{\sum_{j=1}^{nc} S_{ij}}, \quad \text{for } i = 1, \dots, nh \quad (1)$$

$$Y_i = \frac{\sum_{j=1}^{nc} S_{ij} Y_j}{\sum_{j=1}^{nc} S_{ij}}, \quad i = 1, \dots, nh \quad (2)$$

Any other methodology may be applied at this stage to define the best candidate of source location point for each hourly load, without loss of generality. In order to apply the proposed methodology, the

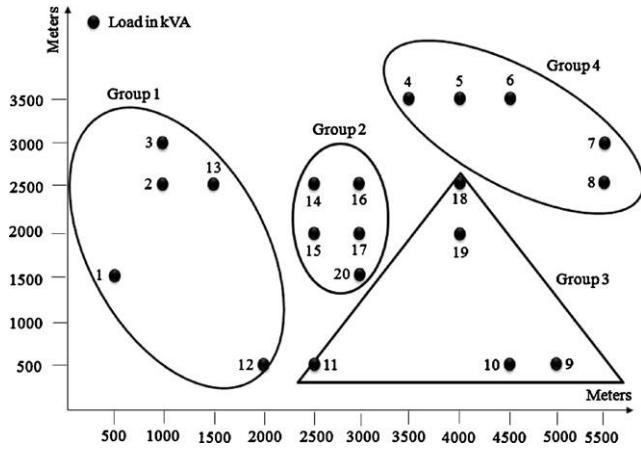


Fig. 2. Representation of the load groups.

vectors  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  corresponding to the coordinates of the hourly load centers of the area to be served are initially assumed as independent, identically distributed observations from a bivariate probability distribution having a density  $f_{e0}$  in a parametric family  $F = \{f_e : e \in \mathcal{E}\}$ , ( $\mathcal{E} \in \mathbb{R}^k$ , or some integer  $k$ ) [17]. It will be assumed that the densities in  $F$  are unimodal, i.e. they have a global maximum which is the only local maximum.

For some relevant families of distributions, the densities are symmetric around the mode in the sense that if the mode of the density occurs at  $(x_0, y_0)$ , then for every vector  $v = (v_1, v_2)$ , the function  $t \rightarrow f_e((x_0, y_0) + tv)$ ,  $t \in \mathbb{R}$ , is an even function of  $t$ . When this symmetry condition holds, the mode coincides with the expected value  $E(X_i, Y_i)$ . Such is the case for the Gaussian family of distributions and the bivariate version of Freund's exponential distribution described below. It is also assumed that the  $k$ -dimensional parameter  $e_0$  is unknown and must be estimated from the sample.

Once an estimate  $e_e$  of  $e_0$  has been found, the estimate for the optimal location of the sources candidate points is the mode of  $(X_i, Y_i)$  with respect to the density  $f_{ee}$ . For the estimated density  $f_{ee}$ , the high probability region at level  $c > 0$  is the subset of the  $X$ - $Y$  plane defined by

$$H(c, \theta_e) = \{(x, y) \in \mathbb{R}^k : f_{ee}(x, y) \geq c\} \quad (3)$$

Due to the unimodality assumption, the high probability regions will be connected sets, having the estimated mode  $(x_0, y_0)$ , of  $f_{ee}$  in their interior. The methodological proposal consists in locating the sources candidate points within the smallest high-probability region (which corresponds to the largest value of  $c$ ) where its construction is feasible.

Three different alternatives were studied for the parametric family  $F$ . For each alternative, the form of the high-probability regions as well as their estimation, when applied to the same real case presented in Fig. 1, are discussed, illustrating the flexibility of the proposed methodology. For easier application of the methodology, the loads have been grouped taking into account the similarity and distance from their possible load center, forming four groups. These groups are presented in Fig. 2. The daily load curves of each load point, numbered from 1 to 20, are shown in Fig. 3.

### 2.1. Gaussian distribution

Let us consider first the Gaussian bivariate family, for which the densities are given by

$$f_{\mu, \Sigma}(X) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp(-(X - \mu)\Sigma^{-1}(X - \mu)^t) \quad (4)$$

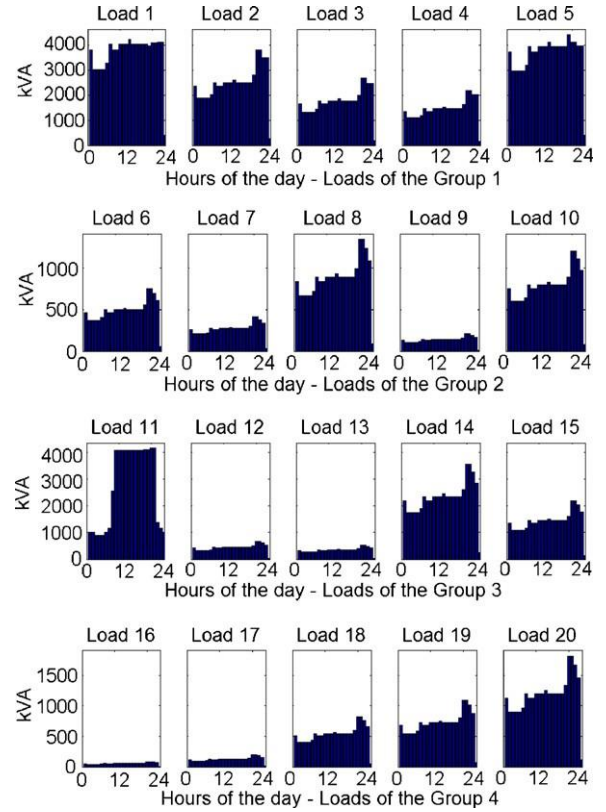


Fig. 3. Daily load curve of each load point-daily load cycle.

with

$$X = \begin{pmatrix} X_i \\ Y_i \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \quad (5)$$

where  $\mu$  is the mean or expected value and the covariance matrix;  $\Sigma$  is a real symmetric  $2 \times 2$  positive definite matrix. This family of distributions has been widely studied in the literature [17].

The estimates  $\hat{X}$  and  $\hat{\Sigma}$  are calculated using (6) or (7) and (8):

$$\hat{X} = \frac{1}{nh} \sum_{i=1}^{nh} \begin{pmatrix} X_i \\ Y_i \end{pmatrix}$$

for  $\mu$  if the total apparent power is not considered for the hour  $i$

(6)

or

$$\hat{X} = \frac{1}{nh} \sum_{i=1}^{nh} \left( \sum_{j=1}^{nc} S_{ij} \right) \frac{\begin{pmatrix} X_i \\ Y_i \end{pmatrix}}{\sum_{i=1}^{nh} \sum_{j=1}^{nc} S_{ij}}$$

for  $\mu$  if the total apparent power is considered for the hour  $i$  (7)

if the total apparent powers are considered, and

$$\hat{\Sigma} = \frac{1}{nh-1} \sum_{i=1}^{nh} \left( \begin{pmatrix} X_i \\ Y_i \end{pmatrix} - \hat{X} \right) \left( \begin{pmatrix} X_i \\ Y_i \end{pmatrix} - \hat{X} \right)^t \quad \text{for } \Sigma \quad (8)$$

for a given probability value  $P$ , the probability regions:

$$h_{(P)} = \left\{ \begin{pmatrix} X \\ Y \end{pmatrix} : \left( \begin{pmatrix} X \\ Y \end{pmatrix} - \hat{X} \right) \hat{\Sigma}^{-1} \left( \begin{pmatrix} X \\ Y \end{pmatrix} - \hat{X} \right)^t \leq P^2 \right\} \quad (9)$$

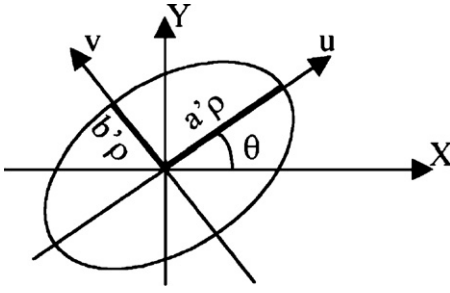


Fig. 4. Rotated ellipse in the new coordinate system.

describe rotated ellipses in the  $X$ - $Y$  plane. If

$$\hat{S}^{-1} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \hat{X} = \begin{pmatrix} c \\ d \end{pmatrix} \quad (10)$$

the following expression is found for the ellipse:

$$a_{11}(X-c)^2 + 2a_{12}(X-c)(Y-d) + a_{22}(Y-d)^2 \leq \rho^2 \quad (11)$$

A new coordinate system, described by Fig. 4, is defined:

$$\begin{aligned} X &= u \cos(\theta) - v \sin(\theta); & Y &= u \sin(\theta) + v \cos(\theta) \\ \text{where } \theta &= \frac{1}{2} \arctan\left(\frac{a_{11} - a_{22}}{2a_{12}}\right) \end{aligned} \quad (12)$$

The modified equation that describes the high probability region is

$$\frac{u^2}{a'^2} + \frac{v^2}{b'^2} = \rho^2 \quad (13)$$

where

$$a' = \frac{1}{\sqrt{a_{11} \cos^2(\theta) + 2a_{12} \sin(\theta) \cos(\theta) + a_{22} \sin^2(\theta)}} \quad (14)$$

$$b' = \frac{1}{\sqrt{a_{11} \sin^2(\theta) + 2a_{12} \sin(\theta) \cos(\theta) + a_{22} \cos^2(\theta)}} \quad (15)$$

and the parameter  $p$  depends upon the chosen value for the probability  $P$ :

$$\begin{aligned} P(\rho) &= 1 - \exp(-\rho^2) = P \Rightarrow \rho = \sqrt{\ln\left(\frac{1}{1-P}\right)}; \\ \text{finally the ellipse area is } A &= \pi \cdot a' \cdot b' \cdot \rho \end{aligned} \quad (16)$$

## 2.2. Freund's generalized exponential distribution

Another option for the parametric family  $F$  is a generalization of Freund's bivariate exponential distribution [18], with densities given by

$$f_{\theta}(X) = \begin{cases} \alpha\beta' \exp(-\beta|Y-d| - (\alpha+\beta-\beta')|X-c|), & |X-c| < |Y-d| \\ \alpha\beta' \exp(-\alpha|X-c| - (\alpha+\beta-\alpha')|Y-d|), & |X-c| \geq |Y-d| \end{cases} \quad (17)$$

with

$$X = \begin{pmatrix} X_i \\ Y_i \end{pmatrix}, \quad \theta = (\alpha, \beta, \alpha', \beta', c, d) \quad (18)$$

where  $e$  is a five-dimensional parameter, since the restriction  $\alpha' = \alpha$  is applied to guarantee continuity of the density;  $c$  and  $d$  are real numbers; and  $\alpha, \beta, \alpha', \beta'$  are positive constants. All of the parameters can be estimated using the following expressions:

$$\alpha = \frac{m}{\sum_{R_1} |X_i - \hat{c}| + \sum_{R_2} |Y_i - \hat{d}|} \quad (19)$$

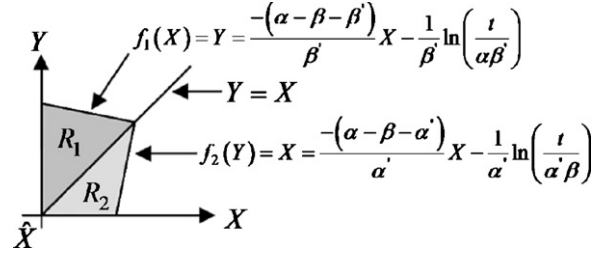


Fig. 5. First quadrant of the high-probability region of the exponential distribution.

$$\beta = \frac{nh - m}{\sum_{R_1} |X_i - \hat{c}| + \sum_{R_2} |Y_i - \hat{d}|} \quad (20)$$

$$\alpha' = \frac{nh - m}{\sum_{R_2} (|X_i - \hat{c}| - |Y_i - \hat{d}|)} \quad (21)$$

$$\beta' = \frac{m}{\sum_{R_1} (-|X_i - \hat{c}| + |Y_i - \hat{d}|)} \quad (22)$$

The mean  $\hat{X} = (\hat{c}, \hat{d})$  is calculated using (6) or (7). Expressions (19)-(22) result after maximizing the natural logarithm of the product  $f_e(X_i)$  for all the loads. Regions  $R_1$  and  $R_2$  are described by

$$R_1: |X - c| < |Y - d| \quad (X - c \geq 0 \text{ and } Y - d \geq 0) \quad (23)$$

$$R_2: |X - c| \geq |Y - d| \quad (X - c \geq 0 \text{ and } Y - d \geq 0) \quad (24)$$

The part of the high-probability region corresponding to the first quadrant ( $X, Y > 0$ ) is determined by two straight lines, as shown in Fig. 5.

After integrating, the following probability function is obtained:

$$P(t) = 1 - \frac{t}{\alpha + \beta} \left( \frac{1}{\beta'} \left( 1 + \frac{\beta}{\alpha} \right) - \left( \frac{1}{\alpha'} + \frac{1}{\beta'} \right) \ln \left( \frac{t}{\alpha\beta'} \right) \right) \quad (25)$$

The value of the parameter that corresponds to a given probability value, has been obtained numerically.

Finally, to calculate the area of the polygon, it is necessary to calculate the area corresponding to the first quadrant and multiplies by 4, we obtain:

For the region  $R_1$  of the Fig. 5 the area is

$$\begin{aligned} \text{Area } R_1 &= \int_0^{X_0} \int_X^{f_1(X)} dY \cdot dX \\ &= \left( \frac{1}{\beta'} - \left( \frac{\alpha + \beta - \beta'}{\beta'} + 1 \right) \cdot \frac{1}{2(\alpha + \beta)} \right) \cdot \frac{(\ln(t/\alpha\beta'))^2}{\alpha + \beta} \end{aligned} \quad (26)$$

For the region  $R_2$  of the Fig. 5 the area is

$$\begin{aligned} \text{Area } R_2 &= \int_0^{Y_0} \int_Y^{f_2(Y)} dX \cdot dY = \left( \frac{1}{\alpha'} \cdot \ln \left( \frac{t}{\alpha'\beta'} \right) \right. \\ &\quad \left. - \left( \frac{\alpha + \beta - \alpha'}{\alpha'} + 1 \right) \cdot \frac{\ln(t/\alpha\beta')}{2(\alpha + \beta)} \right) \cdot \frac{\ln(t/\alpha\beta')}{\alpha + \beta} \end{aligned} \quad (27)$$

Assuming continuity of the probability distribution function can be written as

$$\begin{aligned} \text{Area } R_2 &= \int_0^{Y_0} \int_Y^{f_2(Y)} dX \cdot dY \\ &= \left( \frac{1}{\alpha'} - \left( \frac{\alpha + \beta - \alpha'}{\alpha'} + 1 \right) \cdot \frac{1}{2(\alpha + \beta)} \right) \cdot \frac{(\ln(t/\alpha\beta'))^2}{\alpha + \beta} \end{aligned} \quad (28)$$

Finally the total area of the polygon described in the  $X$ - $Y$  plane by this distribution function is

$$\text{Total Area} = 4(\text{Area } R_1 + \text{Area } R_2) \quad (29)$$

### 2.3. Weibull distribution

The use of the Weibull distribution offers more flexibility in terms of the shape of the high probability region, since the distribution is not symmetrical in the sense described before. However, its application is limited to the identification of a natural zero in the

$X$ - $Y$  plane, so that all of the coordinates  $X$ ,  $Y$  of the hourly load centers are positive. Also, the perimeter of the high-probability region can only be determined numerically and not algebraically, as in cases  $A$  and  $B$ .

In this case, if  $X$ ,  $Y$  are modelled as independent Weibull variables, their joint density is given by

$$f_{\theta}(X, Y) = abcdX^{c-1}Y^{d-1} \exp(-aX^c - bY^d); \quad X, Y > 0 \quad (30)$$

and the parameter  $e = (a, b, c, d)$  for positive  $a, b, c, d$ .

If both  $c, d$  are greater than zero, the density will have a unique mode with both coordinates greater than zero.

By independence, we have

$$f_{\theta}(X, Y) = f_X(X)f_Y(Y) \quad (31)$$

where

$$f_X(X) = aX^{c-1} \exp(-aX^c); \quad X > 0 \quad (32)$$

$$f_Y(Y) = bY^{d-1} \exp(-bY^d); \quad Y > 0 \quad (33)$$

Moment estimation equations for the estimation of the four parameters can be numerically solved, two parameters at a time, solving for  $a$  and  $c$  using  $X_1, \dots, X_{nh}$  and for  $b$  and  $d$  using  $Y_1, \dots, Y_{nh}$ . In order to calculate the parameters, a transformation is performed:

$$X_i = \exp(-\omega x_i), \quad i = 1, \dots, nh \Rightarrow \omega x_i = -\ln(-X_i); \quad i = 1, \dots, nh \quad (34)$$

$$Y_i = \exp(-\omega y_i), \quad i = 1, \dots, nh \Rightarrow \omega y_i = -\ln(-Y_i); \quad i = 1, \dots, nh \quad (35)$$

The probability densities for  $\omega X$  and  $\omega Y$  are

$$f_{\omega X}(\omega x) = a \cdot c \exp(-c \cdot \omega x) \exp[-a \exp(-c \cdot \omega x)] \quad (36)$$

$$f_{\omega Y}(\omega y) = b \cdot d \exp(-d \cdot \omega y) \exp[-b \exp(-d \cdot \omega y)] \quad (37)$$

Introducing a new set of parameters by

$$a = \exp\left(\frac{mx}{\beta x}\right); \quad c = \frac{1}{\beta x} \quad (38)$$

$$b = \exp\left(\frac{my}{\beta y}\right); \quad d = \frac{1}{\beta y} \quad (39)$$

we have that the vector  $(mx, my)$  is the center or location of the probability distribution of  $(\omega x, \omega y)$ , while  $\beta x$  and  $\beta y$  are scale parameters.

The expectation or mean of  $\omega x$  (or  $\omega y$ ) is given by

$$\mu = E\omega_i = \int_{\mathbb{R}} \omega f_{\omega}(\omega) d\omega; \quad \omega \in \mathbb{R} \quad (40)$$

and the variance or squared deviation is

$$\sigma^2 = \text{Var } \omega_i = \int_{\mathbb{R}} (\omega - \mu)^2 f_{\omega}(\omega) d\omega; \quad \omega \in \mathbb{R} \quad (41)$$

The estimates for  $(\mu y a^2)$  are

$$\hat{\mu} = \hat{\omega} = \frac{1}{nh} \sum_{i=1}^{nh} \omega_i$$

for  $\mu$  if the total apparent power is not considered for the hour  $i$

or

$$\hat{\mu} = \hat{\omega} = \frac{\sum_{i=1}^{nh} (\sum_{j=1}^{nc} S_{ij}) \omega_i}{\sum_{i=1}^{nh} \sum_{j=1}^{nc} S_{ij}}$$

for  $\mu$  if the total apparent power is considered for the hour  $i$  (42)

if the total apparent power is considered, and

$$\hat{\sigma}_{\omega}^2 = \frac{1}{nh-1} \sum_{i=1}^{nh} (\omega_i - \hat{\omega})^2 \quad \text{for } \sigma^2 \quad (43)$$

$\alpha$  and  $\beta$  are related via:

$$\sigma^2 = \frac{1}{6} \pi^2 \beta^2 \quad (44)$$

from where we get the following estimate for  $\beta$ :

$$\hat{\beta} = \frac{\sqrt{6}\hat{\sigma}}{\pi} \quad (45)$$

while  $m$  depends on  $\mu$  and  $\beta$  through

$$m = \mu - \gamma \beta \quad (46)$$

where  $\gamma$  is Euler's constant: 0.57721566. This gives the following estimate for  $m$ :

$$\hat{m} = \hat{\omega} - \frac{0.57722 \sqrt{6} \hat{\sigma}}{\pi} \quad (47)$$

The parameters of the Weibull distribution are evaluated using (39) and (40), using the  $X$  coordinates for  $m_x$  and  $\beta_x$  and the  $Y$  coordinates for  $m_y$  and  $\beta_y$ .

The high-probability regions can be found numerically from the inequality:

$$acX^{c-1} \exp(-aX^c) b d Y^{d-1} \exp(-bY^d) \geq \rho \quad (48)$$

### 3. Results

The optimal location and sizing of DG and their distribution networks is a complex mathematical combinatorial optimization problem [19]. This optimization problem is a non-linear one, it contains binary variables (0, 1) that are the decision to install lines and generators and continuous positive variables that represents the power circulating on the lines [20]. This problem can be solved by deterministic or meta-heuristic mathematical optimization technique. Moreover, if the number of decision binary variables is high, then the execution time is an exponential function during computer calculation process.

The meta-heuristic technique can solve this problem with a rational execution time, but its difficulties consist in the tuning of the parameters of the algorithm, and each case is different from the other one and therefore, the treatment and adjustment of the parameters is also different. For this reason, the authors have thought in developing a new methodology that can solve this problem taking into account the hourly load changes or the daily load cycle. In this methodology the hourly load centers, for each of the different hourly load scenarios, are calculated deterministically. These location points, properly weighted according to their load magnitude, are used to calculate the best fit probability distribution. This distribution is used to determine the maximum likelihood



**Table 1**

Load curve of each load point in kVA/hour.

Hour	Group									
	1					2				
	Load 1	Load 2	Load 3	Load 12	Load 13	Load 14	Load 15	Load 16	Load 17	Load 20
1	3815	2364	1673	3724	1370	465	833	256	129	747
2	3040	1884	1333	2967	1092	371	663	204	102	595
3	3040	1884	1333	2967	1092	371	663	204	102	595
4	3040	1884	1333	2967	1092	371	663	204	102	595
5	3040	1884	1333	2967	1092	371	663	204	102	595
6	3279	2031	1437	3200	1177	400	716	220	111	642
7	4054	2511	1777	3957	1456	494	885	272	137	794
8	3815	2364	1673	3724	1370	465	833	256	129	747
9	3815	2364	1673	3724	1370	465	833	256	129	747
10	4054	2511	1777	3957	1456	494	885	272	137	794
11	4054	2511	1777	3957	1456	494	885	272	137	794
12	4054	2511	1777	3957	1456	494	885	272	137	794
13	4232	2622	1856	4131	1520	516	924	284	143	829
14	4054	2511	1777	3957	1456	494	885	272	137	794
15	4054	2511	1777	3957	1456	494	885	272	137	794
16	4054	2511	1777	3957	1456	494	885	272	137	794
17	4054	2511	1777	3957	1456	494	885	272	137	794
18	4054	2511	1777	3957	1456	494	885	272	137	794
19	3987	2807	1986	4422	1627	553	989	304	153	887
20	4100	3804	2692	4119	2205	749	1340	412	207	1202
21	4100	3804	2692	4119	2205	749	1340	412	207	1202
22	4125	3509	2483	3984	2034	691	1236	380	191	1109
23	4125	3509	2483	3984	2034	604	1080	332	167	969
24	417	259	183	407	150	51	91	28	14	82
Hour	Group									
	3					4				
	Load 9	Load 10	Load 11	Load 18	Load 19	Load 4	Load 5	Load 6	Load 7	Load 8
1	334	421	1013	1366	2209	55	124	510	681	1127
2	266	336	1013	1088	1760	44	99	407	543	898
3	266	336	908	1088	1760	44	99	407	543	898
4	266	336	908	1088	1760	44	99	407	543	898
5	266	336	908	1088	1760	44	99	407	543	898
6	287	362	1013	1174	1898	48	106	438	585	968
7	355	448	1176	1451	2347	59	131	542	724	1197
8	334	421	2554	1366	2209	55	124	510	681	1127
9	334	421	4100	1366	2209	55	124	510	681	1127
10	355	448	4100	1451	2347	59	131	542	724	1197
11	355	448	4100	1451	2347	59	131	542	724	1197
12	355	448	4100	1451	2347	59	131	542	724	1197
13	371	467	4100	1515	2451	61	137	566	755	1250
14	355	448	4100	1451	2347	59	131	542	724	1197
15	355	448	4100	1451	2347	59	131	542	724	1197
16	355	448	4100	1451	2347	59	131	542	724	1197
17	355	448	4100	1451	2347	59	131	542	724	1197
18	355	448	4123	1451	2347	59	131	542	724	1197
19	397	500	4123	1622	2623	66	147	606	809	1338
20	538	678	4158	2198	3555	89	199	821	1096	1813
21	538	678	4158	2198	3555	89	199	821	1096	1813
22	496	625	1378	2027	3279	82	184	757	1011	1672
23	434	546	1176	1771	2865	72	160	662	883	1461
24	37	46	1013	149	242	6	14	56	74	123

perimeter of the area where each DG source point should preferably be located by the planning engineers. When all the load centers are computed, then the planning engineers can better calculate the load nodes that can be fed by each DG source by the simple calculation of the minimal of the function ( $\text{kVA} \times \text{distance}$ ). The algorithm is stopped when the sum of the load nodes is lower or equal than the capacity of DG source, taking into account a certain reserve for the future load growth.

The proposed methodology has been applied to a realistic case that has been extracted from reference [21]. This real case study is a region in development that belongs to the Oil of Venezuela Company. The concentrated charges placement on the  $X$ - $Y$  plane of the area and their values has been provided by the company.

The geographical locations on the  $X$ - $Y$  plane as well as the hourly load cycles of each of the 20 loads and their coordinates are presented in Figs. 1 and 3, Table 1 and Table 2, respectively.

The service area of each DG source point could be any geometrical form (see Fig. 2); in this work, each service area has been represented by different geometrical form, but the ideal geometrical form is the circle because all the load points are equidistant from the DG source.

The criterion used for load grouping is by similitude of their load curve and by distance (see Fig. 2). The size of DG generators to be used for feeding the isolated electrical service area is normally conditioned by the load of the area, the power losses and the reserve that have to be taken into account. Then, if the choice of load group

**Table 2**  
Coordinates  $X$  and  $Y$  of each load point.

Group	Load number	Coordinate	
		$X$ (m)	$Y$ (m)
1	Load 1	500	1500
	Load 2	1000	2500
	Load 3	1000	3000
	Load 12	2000	500
	Load 13	1500	2500
2	Load 14	2500	2500
	Load 15	2500	2000
	Load 16	3000	2500
	Load 17	3000	2000
	Load 20	3000	1500
3	Load 9	5000	500
	Load 10	4500	500
	Load 11	2500	500
	Load 18	4000	2500
	Load 19	4000	2000
4	Load 4	3500	3500
	Load 5	4000	3500
	Load 6	4500	3500
	Load 7	5500	3000
	Load 8	5500	2500

is different, the size of the generator that should feed these nodes could be different.

In general terms, the generators capacity choice might be greater or smaller depending on its service area, but the sum of the capacity should not exceed the sum of load nodes taking into account the power losses and a determined reserve.

The main core of the proposed methodology is the fact of the source points' determination. These points are the candidates where the DG will be installed. Once these points are determined, we can decide each corresponding circuits resulting by the application of the minimal function ( $kVA \times \text{distance}$ ) taking into account the DG capacities and their reserve.

The determination of the source candidates points by the application of the methodology are presented for the three different probability distributions that were selected.

Fig. 6 depicts the results when the Gaussian bivariate distribution was assumed. The rotated ellipses indicate the perimeter of the 90% and 95% probability areas for the load groups 1, 2, 3 and 4, respectively.

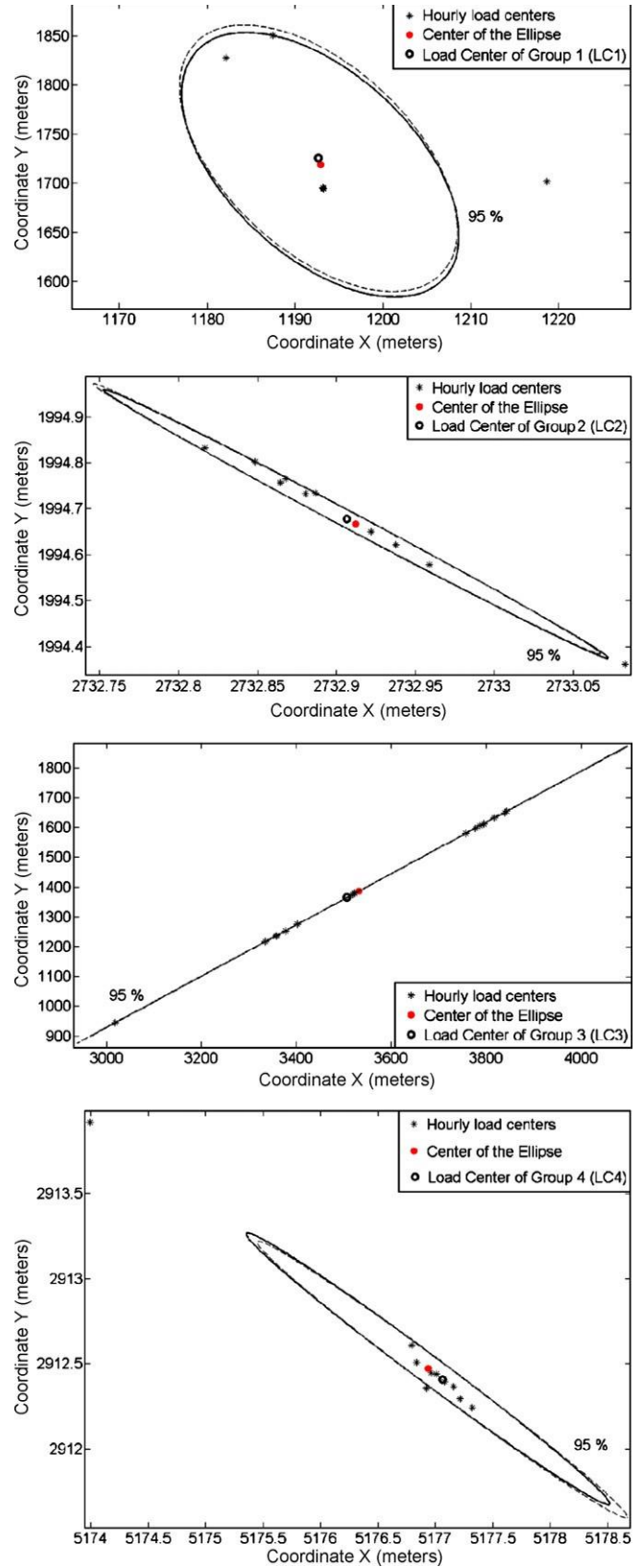
Fig. 7 presents the resultant perimeter for the 90% and 95% probability areas, assuming the generalization of Freund's exponential distribution for the load groups 1, 2, 3 and 4, respectively. In this case, the particular characteristics of the problem due to the absence of  $(X, Y)$  pairs in the  $R_1$  region in the four quadrants of the  $X$ - $Y$  plane, lead to mathematical inconsistencies in the calculation of the parameters of the estimated distribution.

The capacity of Freund's distribution to adapt the probability region to the pairs  $(X, Y)$  of hourly load centers may be observed in the shape of the resultant polygon.

Finally, Fig. 8 presents the results for the Weibull distribution when the probability was fixed in 90% and 95%. The asymmetry of the resultant perimeters of the high-probability areas may be observed.

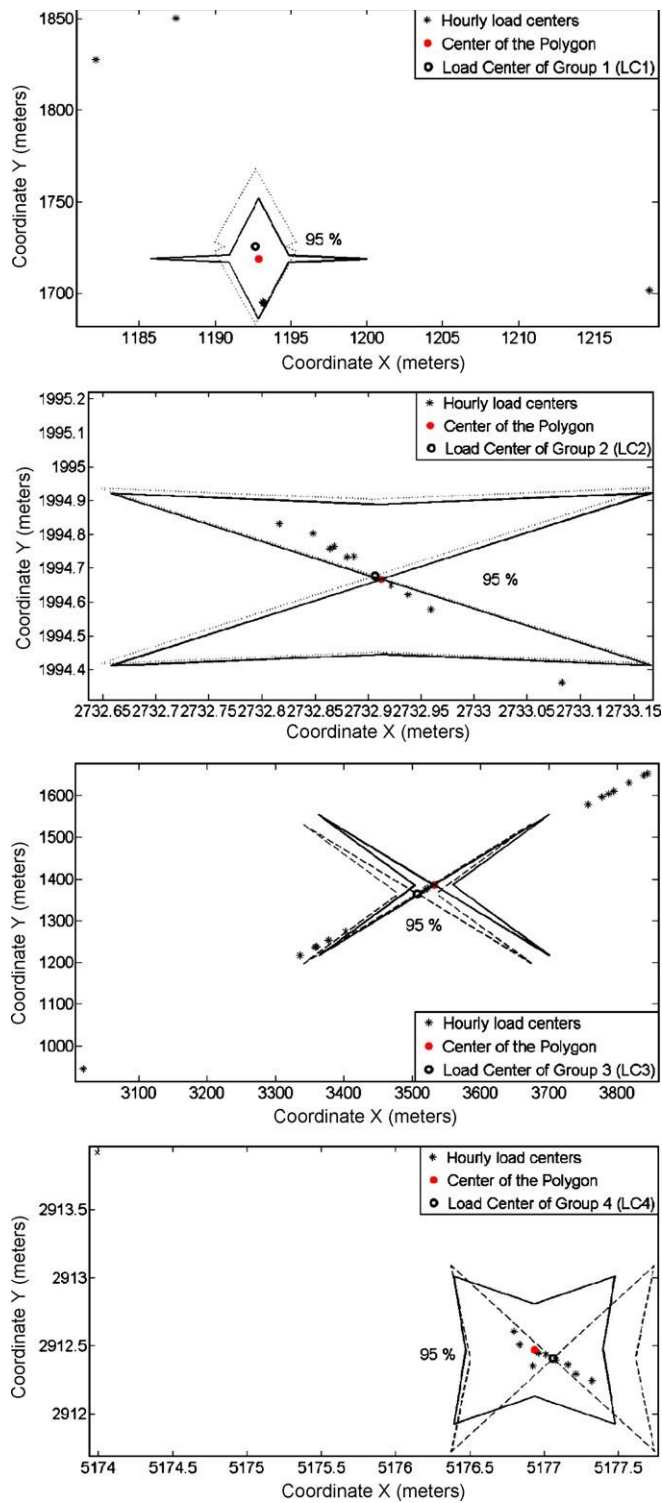
This particular distribution requires the definition by the planning engineers of a natural zero in the  $X$ - $Y$  plane. The selected position for the natural zero modifies the final shape of the probability distribution, especially if the zero is located in the vicinity of the hourly load centers.

This characteristic of the Weibull distribution may be applied to handle natural constraints, such as a sea shoreline, amongst others.



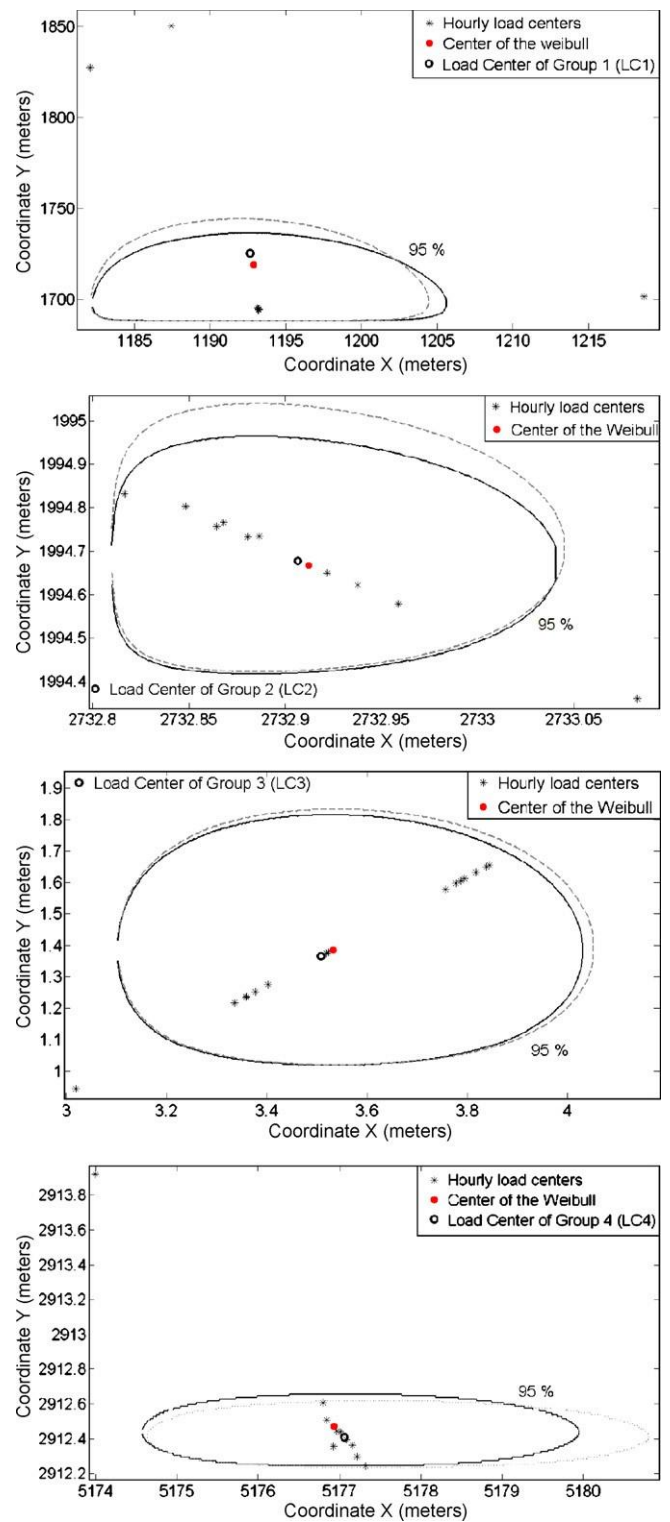
**Fig. 6.** Resultant perimeters for the 90% and 95% probability areas assuming the Gaussian bivariate distribution for the load groups 1, 2, 3 and 4, respectively.





**Fig. 7.** Resultant perimeters for the 90% and 95% probability areas assuming the generalization of Freund's exponential distribution for the load groups 1, 2, 3 and 4, respectively.

In order to determine the recommended perimeter for selection of the best location of the DG associated to the high-probability area for a 90% and 95% probability level, three different probability distribution types were successfully tested on a realistic case. The results are presented in Fig. 9, and the exact best point's coordinates LC1, LC2, LC3 and LC4 are presented in Table 3. These points represent valuable information for supporting decision making



**Fig. 8.** Resultant perimeters for the 90% and 95% probability areas assuming the Weibull distribution for the load groups 1, 2, 3 and 4, respectively.

**Table 3**

Best location of the possible points of distributed generation.

Candidate point	Coordinate	
	X (m)	Y (m)
LC1	1193	1694
LC2	2732	1995
LC3	3531	1385
LC4	5177	2912

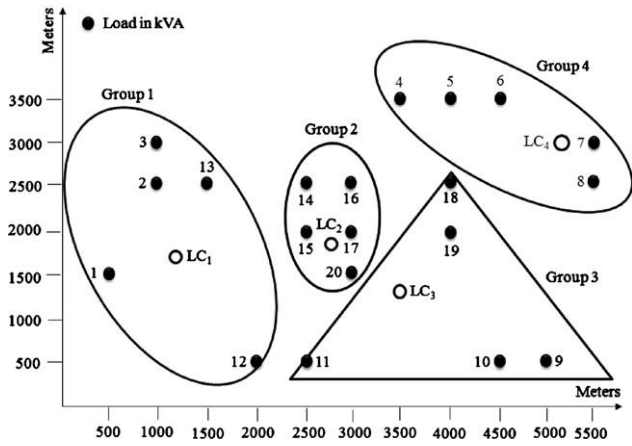


Fig. 9. Determination of load center of each group.

concerning the possible location of DG points, allowing the choice of the capacity and the design of distribution networks that feed all the consumers with high power quality and minimal cost.

The representation of the raised problem is realized on the base of an initial network that must be introduced by the planning engineer. This initial network is presented in Fig. 10.

The distribution network is composed by a set of lines and another set of DG, whose function is to supply energy to a given load zone. These points have a geographical location and their magnitude demands are known. The number of connection lines between the loads and the DG sources can be high. Theoretically, there can be a line between two nodes and in every node a DG source; nevertheless, in the practice, there exist technical norms, obstacles of the area and other factors that limit the quantity of lines connection, the points of DG source and the capacities that have to be evaluated. Fig. 10 shows the network connection scheme that contains 20 load nodes, four DG sources nodes and 52 lines that represent the initial network. To build it, it is necessary to take into account the following:

- The DGs are located on the source nodes that could have demand. The connection lines are represented with only one direction (for obvious reason) from the DG source up to the load nodes that must be supplied.
- There are several defined DG capacities that correspond to norms of the local Electric Company or equivalent entity. In all the points, the proposed capacities are evaluated. The DG can be one-phase,

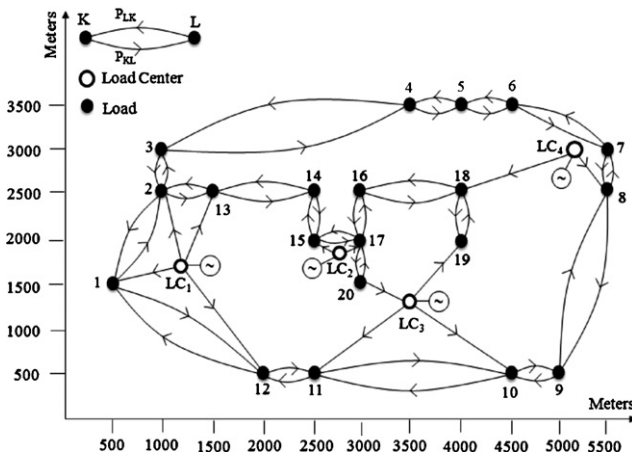


Fig. 10. Initial network introduced by the planning engineers.

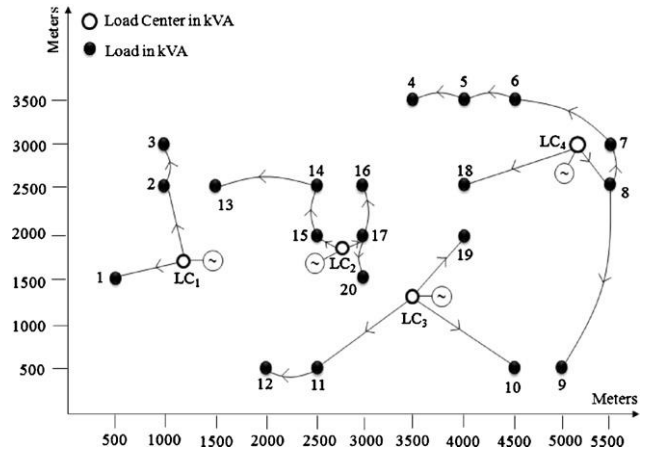


Fig. 11. Resultant networks obtained by proposed methodology.

two-phases or three-phases in accordance with the load requirements.

- The feasibility connection lines of the load nodes are studied. These connection lines must be logical trajectories from technical-economical point of view. The power flow is represented in a single direction only if this circulates exclusively from the source towards the load nodes. If there is a doubt about feeding the load nodes from any existing DG source, a line with two-way direction is used to represent the power flow, so that each node can be supplied from any DG source.

To obtain the distribution network, the minimal function ( $\text{kVA} \times \text{distance}$ ) has been applied taking into account the best location of the DG source; we can assure that the resulting distribution network is a low cost one because the cost is proportional to the length of the network. Consequently, the active power losses are lower, and therefore, the power quality is good due to the same fact.

Fig. 11 shows the distribution networks resulting from the application of the proposed methodology.

Note that the resultants distribution networks are radial due to its economical operation costs. The methodology also allows to obtain a meshed distribution networks under the discretion of planning engineers.

The authors are presently working in the formulation of the optimization problem to determine the optimal capacity of the DG that can be installed in these location points and their corresponding distribution networks. The mathematical model will allow to obtain a solution for the optimal capacity of the DG, where these can be installed (in this realistic case study, these points are LC1, LC2, LC3 and LC4), and the distribution networks topology. It is important to note that the mathematical model can eliminate some points that do not result in the minimization of the total cost subject to all the technical constraints. The results will be reported for publication in the near future.

#### 4. Conclusions

This paper proposes a probabilistic methodology to assist planning engineers in the selection of the best location of the candidate points of distributed generation. The methodology considers the hourly load changes and calculates an estimate probability distribution. This allows to determine a recommended perimeter for the selection of the distributed generation location, associated to the high-probability area for a given probability level.

Three different probability distribution types were successfully tested on a realistic case, and results are presented showing the potential of the proposed approach. The resultant perimeter represents valuable information for the location of the distributed generation points, where the availability and the cost of the land lots as well as other relevant constraints that may be present in the selection of the distributed generation site must be considered, especially in urban areas. The application of the methodology to a real case demonstrates the effectiveness of the method to solve a combinatorial complex problem of the location and sizing of DG and their corresponding distribution networks in an isolated electrical service area.

## Acknowledgments

The authors would like to acknowledge the Portuguese Science and Technology Foundation (FCT), the European Fund for Regional Development (FEDER), the Operational Program for Science Technology and Innovation (POCTI), the Information Society Operational Program (POSI), the Operational Program Science and Innovation (POCI), the Operational Program for the Knowledge Society (POSC) and the Program of Projects in all Scientific Areas (PTDC) for their support to R&D Projects and GECAD Unit.

## References

- [1] J. Dolezal, P. Santarius, J. Thusty, V. Valouch, F. Vybíralík, The effect of dispersed generation on power quality in distribution system, in: Quality and Security of Electric Power Delivery Systems, CIGRE/IEEE PES International Symposium, October 8–10, 2003, pp. 204–207.
- [2] K.A. Nigim, Y.G. Hegazy, Intention islanding of distributed generation for reliability enhancement, in: Power Engineering Society General Meeting, 2003, IEEE, vol. 4, July 13–17, 2003, pp. 2446–2451.
- [3] H. Iyer, S. Ray, R. Ramakumar, Voltage profile improvement with distributed generation, in: IEEE Power Engineering Society, General Meeting, San Francisco, CA, USA, June 12–16, 2005.
- [4] T. Gönen, Electric Power Distribution System Engineering, McGraw-Hill, New York, 1986.
- [5] J. Nazarko, W. Zalewski, The fuzzy regression approach to peak load estimation in power distribution systems, IEEE Transactions on Power Systems 14 (August (3)) (1999) 809–814.
- [6] H.L. Willis, C.L. Brooks, An interactive end-use electric load model for micro-computer implementation, IEEE Transactions on PAS (November) (1983) 3693–3700.
- [7] H.M. Khodr, J.A. Melián, A.J. Quiroz, D.C. Picado, J.M. Yusta, A.J. Urdaneta, A probabilistic methodology for distribution substation location, IEEE Transactions on Power Systems 17 (February (1)) (2003) 388–393.
- [8] M.A. Eldery, E.F. El-Saadany, M.M.A. Salama, Effect of distributed generator on the allocation of D-STATCOM in distribution network, in: IEEE Power Engineering Society, General Meeting, San Francisco, CA, USA, June 12–16, 2005.
- [9] T.Q.D. Khoa, P.T.T. Binh, H.B. Tran, Optimizing location and sizing of distributed generation in distribution systems 2006, IEEE/PES Power Systems Conference and Exposition 1–5 (2006) 725–732.
- [10] P. Chiradeja, R. Ramakumar, An approach to quantify the technical benefits of distributed generation, IEEE Transactions on Energy Conversion 19 (December) (2004) 764–773.
- [11] G. Celli, E. Ghiani, S. Mocci, F. Pilo, A multiobjective evolutionary algorithm for the sizing and siting of distributed generation, IEEE Transactions on Power Systems 20 (May) (2005) 750–757.
- [12] D. Singh, D. Singh, K.S. Verma, Multiobjective optimization for DG planning with load models, IEEE Transactions on Power Systems 24 (February) (2009) 427–436.
- [13] C.S. Wang, M.H. Nehrir, Analytical approaches for optimal placement of distributed generation sources in power systems, IEEE Transactions on Power Systems 19 (November) (2004) 2068–2076.
- [14] T. Gozel, M.H. Hocaoglu, An analytical method for the sizing and siting of distributed generators in radial systems, Electric Power Systems Research 79 (June) (2009) 912–918.
- [15] R.K. Singh, S.K. Goswami, Optimum siting and sizing of distributed generations in radial and networked systems, Electric Power Components and Systems 37 (2009) 127–145.
- [16] IEEE Committee Rep., Proposed definitions of terms for reporting and analyzing outages of electrical transmission and distribution facilities and interruptions, IEEE Transactions on Power Apparatus and Systems PAS-87 (May) (1968) 1318–1323.
- [17] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, 2nd ed., Wiley, New York, 1984.
- [18] S. Kotz, N. Balakrishnan, N.L. Johnson, Continuous multivariate distributions Models and Applications, vol. I, Wiley, New York, 2000.
- [19] H.M. Khodr, J.F. Gomez, L. Barnique, J.H. Vivas, P. Paiva, J.M. Yusta, A.J. Urdaneta, A linear programming methodology for the optimization of electric power-generation schemes, IEEE Transactions on Power Systems 17 (August (3)) (2002) 864–869.
- [20] H.M. Khodr, Z.A. Vale, C. Ramos, A benders decomposition and fuzzy multicriteria approach for distribution networks remuneration considering DG, IEEE Transactions on Power Systems 24 (May (2)) (2009) 1091–1101.
- [21] L.C. Diaz-Golding, Location and optimization of distributed generation considering different technologies, Specialization Thesis, Simón Bolívar University, Caracas, Venezuela, January 2005 (in Spanish).

*H.M. Khodr* received his Ph.D., M.Sc., and Engineer degrees in electrical engineering from the José Antonio Echeverría Higher Polytechnic Institute (ISPJAE), Havana, Cuba, in 1997 and 1993, respectively. He is a former Associate Professor of electrical engineering at Universidad Simón Bolívar, Caracas, Venezuela. He was a researcher at INESC Porto, Porto, Portugal. Presently, he is a researcher at GECAD, Porto. He has participated in a number of projects performed for the local industries. His current research activities are concentrated on planning, operation, and economics of electrical distribution and industrial power systems; electricity market; grounding systems; and optimization.

*Marco R. Silva* received his Engineer degree in Electrical Engineering from the Polytechnic Institute of Porto (ISEP/IPP), Portugal in 2007. He is an Assistant Researcher at GECAD – Knowledge Engineering and Decision-Support Research Center of the Engineering Institute of Porto – Polytechnic Institute of Porto (ISEP/IPP) granted by Fundação para a Ciência e a Tecnologia (FCT). His current research activities are concentrated in future energy distribution networks.

*Zita Vale* is a Coordinator Professor of Power Systems at the Engineering Institute of Porto – Polytechnic Institute of Porto (ISEP/IPP), Portugal. She coordinates the Intelligent Energy Systems research line of GECAD. She received her diploma in Electrical Engineering in 1986 and her Ph.D. in 1993, both from University of Porto. Her main research interests concern Artificial Intelligence (A.I.) applications to Power System operation and control, Electricity Markets and Distributed Generation. She is involved in several R&D projects concerning the application of A.I. and Decision-Support Techniques to Engineering problems.

*Carlos Ramos* received his graduation and the Ph.D. degree in Electrical Engineering from the University of Porto in 1986 and 1993, respectively. He is a Coordinator Professor of Computer Engineering at the Engineering Institute of Porto – Polytechnic Institute of Porto (ISEP/IPP), Portugal. His main R&D interests are Artificial Intelligence and Decision Support Systems.