Probabilistic Perspective of the Optimal Distributed Generation Integration on a Distribution System

Juan M. Lujano-Rojas[a], Rodolfo Dufo-López[b], José L. Bernal-Agustín[b], José A. Domínguez-Navarro[b], João P. S. Catalão[a,c,d,*]

[a]INESC-ID, Instituto Superior Técnico, University of Lisbon, 1049-001 Lisbon, Portugal
[b]Department of Electrical Engineering, Universidad de Zaragoza, Calle María de Luna 3, 50018 Zaragoza, Spain
[c]INESC TEC and the Faculty of Engineering of the University of Porto, 4200-465 Porto, Portugal
[d]C-MAST, University of Beira Interior, 6201-001 Covilhã, Portugal

* Corresponding author at: INESC TEC and the Faculty of Engineering of the University of Porto, 4200-465 Porto, Portugal. E-mail: catalao@ubi.pt (J.P.S. Catalão).

Abstract
The effects of optimal dimensioning and integration of distributed generation (DG) on an electricity distribution system (DS) from a probabilistic viewpoint is presented in this paper, as a new contribution to earlier studies. The proposed methodology pays special attention to preventing reverse power flow at substation as a consequence of excessive integration of renewable energy based DG. As the analysis of large amounts of data typically measured on an annual basis could be exhausting from a computational perspective, a methodology based on estimating the potential of wind and solar resources is applied; from this procedure, those months of highest renewable potential are selected so that indirectly those situations with probability of reverse power flow at substation are considered. After this, time series of load demand per node and phase are generated using typical profiles and the corresponding peak-load expected. Finally, all this information is introduced on an optimization algorithm based on a genetic algorithm in order to minimize the net present cost over the project lifetime, obtaining the type and number of photovoltaic (PV) panels and wind turbines (WTs) to be installed. This approach allows integrating detailed mathematical models of DG related to PV and wind generation, including specific factors frequently reported by the manufacturers such as temperature coefficients, nominal operating cell temperature, particular WT power curves, and variable efficiency of power converter, among other characteristics. The proposed method is illustrated by studying a DS supposed to be located in Zaragoza, Spain, with 35 nodes under unbalanced conditions, with residential as well as small, medium, and large commercial electricity demands. Focusing our attention on the month of February, due to its high renewable potential, the proposed technique was applied resulting in a system mainly based on wind energy of at least 40% of the substation capacity. This model could be used to perform the renewable energy integration analysis on DS, starting from typical load profiles, hourly estimations of solar and wind resources, and data frequently provided by PV panels and WT manufacturers.

Keywords: distribution system; distributed generation; wind energy; solar photovoltaic energy; genetic algorithm.

List of symbols

n	Index for each node of DS (n=1,…,N).
t	Index for each hour of the month (t=1,…,T).
h	Index for each hour of the day (h=1,…,$H = 24$).
m	Index for each branch of DS (m=1,…,M).
i	Index for each WT of the dataset (i=1,…,I).
j	Index for each PV panel of the dataset (j=1,…,J).
q	Index for month of the year (q=1,…,$Q = 12$).
w	Index for GA individual (w=1,…,W).
γ_t	Value of the white noise at time t.
$a_{t,n}$	Correlated random variable at time t and node n.
$b_{t,n}$	Correlated random variable with the effect of daily profile at time t and node n.
$c_{t,n}$	Transformed and correlated variable with the effect of daily profile at time t and node n.
$d^a_{t,n}$, $d^b_{t,n}$, $d^c_{t,n}$	Normalized load of phases a, b, and c at time t and node n, respectively.
$d^a_{MAX,n}$, $d^b_{MAX,n}$, $d^c_{MAX,n}$	Maximum load of phases a, b, and c at node n, respectively (kVA).
$\mu_{HD,h}$	Average normalized hourly demand at time h.
$\sigma_{HD,h}$	Standard deviation of normalized hourly demand at time h.
μ_{DD}	Average normalized daily demand.
σ_{DD}	Standard deviation of normalized daily demand.
α_{BD}, β_{BD}	Parameters of beta distribution.
$\rho_{HD,n}$	Autocorrelation of load demand time series at node n.
F_{HD}	Cumulative distribution of hourly load demand.
F_{BD}^{-1}	Cumulative distribution of beta distribution.
φ_n	Power factor of the load connected to node n.
$P^a_{HD,t,n}$, $P^b_{HD,t,n}$, $P^c_{HD,t,n}$	Active power at time t, node n, and phases a, b, and c (kW).
$Q^a_{HD,t,n}$, $Q^b_{HD,t,n}$, $Q^c_{HD,t,n}$	Reactive power at time t, node n, and phases a, b, and c (kW).
Z_m^+, Z_m^-, Z_m^0	Positive, negative, and zero sequence impedance (Ω/1000ft).
Z_m	Impedance matrix of branch m (Ω/1000ft).
R_m	Resistance of the phase conductor (Ω/1000ft).
k_m	Parameter of branch m.
GMD_m	Geometric mean distance of phase conductor of branch m.
GMR_m	Geometric mean radius of phase conductor of branch m.
Z_m^{SF}	Self-impedance of phase conductor at branch m (Ω/1000ft).
$Z_{N,m}^{SF}$	Self-impedance of neutral conductor at branch m (Ω/1000ft).
Z_m^{MT}	Mutual impedance between phase conductors of branch m (Ω/1000ft).
$Z_{N,m}^{MT}$	Mutual impedance between phase conductors and the neutral (Ω/1000ft).
f_{DS}	Frequency of the system (Hz).
ρ_{DS}	Earth resistivity (Ωm).
R_e	Resistance of earth return path (Ω/1000ft).
D_e	Equivalent depth of earth (ft).
$P^a_{SUB,t}$, $P^b_{SUB,t}$, $P^c_{SUB,t}$	Active power of substation at time t, and phases a, b, and c (kW).

$U_{WT,i}^{IN}$, $U_{WT,i}^{NOM}$, $U_{WT,i}^{OUT}$	Cut-in, rated, and cut-out speed of wind turbine i (m/s).
$\alpha_{WT,i}$, $\beta_{WT,i}$, $\theta_{WT,i}$	Parameters of WT power curve i.
$P_{WT,t,n}$	Power production of a single WT at time t and node n (kW).
$P_{WF,t,n}^{a}$, $P_{WF,t,n}^{b}$, $P_{WF,t,n}^{c}$	Active power of WF at time t, node n, and phases a, b, and c (kW).
$Q_{WF,t,n}^{a}$, $Q_{WF,t,n}^{b}$, $Q_{WF,t,n}^{c}$	Reactive power of WF at time t, node n, and phases a, b, and c (kVAr).
N_{WT}	Number of WTs in the farm.
G_t	Solar radiation at time t (W/m^2).
$T_{AMB,t}$	Ambient temperature at time t (°C).
$T_{PV,t}$	PV cell temperature at time t (°C).
$NOCT_j$	Nominal operating cell temperature of panel j (°C).
$\eta_{PV,j}$	Efficiency of PV panel j.
$\alpha_{PV,j}$	Temperature coefficient of PV panel j (%/°C).
$k_{PV,j}$	Parameter PV cell model of type j.
$V_{PV,t}^{e}$	Thermal voltage (V).
m_{PV}	PV cell ideality factor.
k_B	Boltzmann constant (J/K).
e_{PV}	Electron charge (C).
V_{DC}	Direct current voltage of PV generator (V).
$V_{PV,t}$, $I_{PV,t}$, $P_{PV,t}$	PV cell voltage (V), current (A), and power (W) at time t.
$V_{PV,j}^{MAX}$, $I_{PV,j}^{MAX}$	Voltage and current at maximum power production (V).
$V_{PV,t}^{OC}$	PV cell open circuit voltage at time t (V).
$V_{PV,STC,j}^{OC}$	PV cell open circuit voltage under STC for panel j (V).
$v_{PV,t}^{OC}$	Relative PV cell open circuit voltage at time t.
$N_{PV,j}^{CS}$	Number of cells connected in serial for panel j.
$N_{PV,j}^{CP}$	Number of cells connected in parallel for panel j.
$FF_{PV,t}^{o}$	Maximum fill factor at time t.
$FF_{PV,j}$	Fill factor of PV panel j.
$I_{PV,t}^{SC}$	Short-circuit current at time t (A).
$I_{PV,STC,j}^{SC}$	Short-circuit current under STC at time t (A).
$R_{PV,t}^{S}$	PV cell resistance at time t (Ω).
$P_{PVF,t,n}^{a}$, $P_{PVF,t,n}^{b}$, $P_{PVF,t,n}^{c}$	Active power of PVF at time t, node n, and phases a, b, and c (kW).
$Q_{PVF,t,n}^{a}$, $Q_{PVF,t,n}^{b}$, $Q_{PVF,t,n}^{c}$	Reactive power of PVF at time t, node n, and phases a, b, and c (kVAr).
$P_{INV,j,n}$	Rated power of converter connected to PVF j and node n (kW).
$p_{INV,t}$	Power flowing through the inverter at time t (kW).
$\alpha_{INV,j}$, $\beta_{INV,j}$, $\theta_{INV,j}$	Parameters of inverter efficiency model.
$\eta_{INV,t}$	Inverter efficiency at time t.
P_{THD}^{MAX}	Maximum total hourly load (kW).
$P_{THD,t}$	Total hourly load at time t (kW).
E_{SG}^{MAX}	Maximum electricity price (€/kWh).
$E_{SG,t}$	Electricity price at time t (€/kWh).
ρ_{AD}	Density of the air (kg/m^3).
T_q	Number of hours of the month q (h).

$T_{SH,q}$	Number of sunshine hours of the month q (h).
wp_q	Wind power potential for month q (W/m^2).
pvp_q	Solar power potential for month q (W/m^2).
GP	GA population.
$\overrightarrow{gp_w}$	Individual w.
gp_1^w, gp_2^w, gp_3^w, gp_4^w	Elements of the individual w.
NPC	Net present cost (€).
ACC, ARC, AMC	Annualized capital, replacement, and maintenance costs (€).
CRF	Capital recovery factor.
f_w	Fitness of the individual w.
$P_r\{\cdot\}$	Calculation of the probability.
$INT(\cdot)$	Calculation of integer value.

1. Introduction

Renewable energies are a critical piece in the infrastructure required to build a sustainable society, but human activities related to technological innovation and economic growth have impacted the ecosystem in a negative manner. Thus, new industrial methods of production and an environmentally friendly lifestyle have to be defined in order to guarantee the existence of animal diversity, ecosystem equilibrium, and our own survival [1].

The integration of solar photovoltaic (PV) generation in the urban environment has been highly promoted; however, important barriers have limited its widespread adoption. Some operational problems are the apparition of reverse power flows, deficit of reactive power and consequently the depression of voltage profile, the unexpected operation of protecting devices, and the reduction of power quality. These problems could be prevented by carrying out a detailed planning study taking into account the effects of distributed generation (DG) in terms of active and reactive power, settings of protection devices, short-circuit currents, and power quality indicators. Implementation of optimal power flow (OPF) considering several objectives such as power losses or operating costs has been traditionally suggested. Similarly, incorporation of heuristics optimization approaches and artificial intelligence techniques (and their combination) also have been widely analyzed [2]. In this regard, Lin *et al.* [3] developed a control method specifically designed to prevent the distortion of voltage profile by curtailing the excess of power generation from PV farms (PVFs) during hours of high solar irradiation. Agalgaonkar *et al.* [4] have paid special attention to the effects of forecasting error of PV generation on system performance, specifically in the operation of tap changers and voltage regulators. Peng *et al.* [5] implemented crisscross algorithm

to determine the optimal placement of DG, whereas the uncertainty introduced by renewable generation and its impact on a distribution system (DS) was incorporated by using the Monte Carlo simulation (MCS) approach embedded in a probabilistic power flow. El Batawy and Morsi [6] proposed an optimization model for designing secondary DSs provided with PV generation. The methodology aims to determine the amount of customers who should be connected to a determined distribution transformer, taking into account the PV penetration level in order to minimize the total cost on an annual basis. Similarly, Pillai *et al.* [7] designed a methodology to analyse the influence of DS on power production of PVFs installed at the residential level. The methodology was developed to estimate the PV integration level at which renewable power has to be curtailed as well as the economic performance of PVFs and their dependence on the connection points and generation profiles. Ahmadigorji *et al.* [8] proposed a procedure based on a two-stage evolutionary optimization able to incorporate DG behaviour into the planning of DS.

Most of these initiatives and ideas have been adopted to create computational tools specifically designed to support technicians in the process of integrating DG under determined environmental and operational conditions. From these efforts, several programs such as OpenDSS [9] and GridLAB-D [10] have been created, as a result of the collaboration established between industry and academia. OpenDSS has a wide list of capabilities. It is able to model components related to power delivery, as well as control and protection devices, among other elements. This software performs temporal simulations in sequential manner, so that DS can be analysed on daily and yearly bases, during a determined peak day or under a specific duty cycle, which provides results that are very useful for analysing the behaviour of DS under the influence of DG. In a similar manner, GridLAB-D considers a wide variety of DS elements such as transformers, voltage regulators, capacitor banks, energy storage systems, underground and overhead distribution lines, and DG, among other elements. It is capable of performing time series simulations including environmental information useful to carrying out studies related to DG integration through load flow calculation as well as load control and peak demand management [11]. Using these tools as a starting point, several authors have extended and enhanced their capabilities. Martinez-Velasco and Guerra [12] developed a computational model based on MCS approach, which was implemented by combining MATLAB and OpenDSS. Li *et al.* [13] proposed an optimization model for optimal sizing and placing of DG based on solar PV generation.

Optimal sizing is carried out by using OpenDSS; then, taking advantage of the radial structure of DS, second order cone programming is applied in order to estimate a suitable installation place for DG along a distribution feeder. Hariri and Faruque [14] combined an electromagnetic transient simulation tool with OpenDSS.

In this paper, a probabilistic perspective of the analysis of DSs provided with DG is given, as a new contribution to earlier studies. On the one hand, the analysis of energy systems on an annual basis could be overwhelming due to the amount of data to be studied. On the other hand, estimation errors could be significant if too short time-periods are considered due to the seasonal characteristics of wind and solar resources. To deal with this dilemma, a procedure to choose the proper month of the year to perform a long-term analysis has been used; this procedure is based on estimating the potential of wind and solar resources on a monthly basis, so that those months with highest renewable resources are identified. In this way, those events with high probability of resulting in reverse power flow at substation are indirectly considered. Once data of interest related to the renewable resources have been identified, load demand time series based on typical profiles and expected peak-load values are synthetically generated using probability theory. Finally, the optimization problem is formulated to minimize net present cost (NPC) over the project lifetime, maintaining the probability of observing reverse power flow at substation under a determined significance level; this problem is efficiently solved by means of an integer-coded GA able to consider the type and number of PV panels and wind turbines (WTs) to be installed. The probability of observing reverse power flow at substation is estimated from the results obtained from a stochastic analysis based on power flow.

The paper is organized as follows: Section 2 describes the mathematical model of the energy system including natural resources, DS, DG, and load demand. Section 3 carefully describes the proposed approach including the long-term simulation and optimization processes. Proposed methodology in this paper is illustrated in section 4 by analysing a hypothetical case study of a system to be installed in Zaragoza, Spain. Finally, conclusions are presented in section 5.

2. Energy system modelling

The concept of a smart grid results from the necessity of improving the actual energy system to face the problems related to environmental pollution, energy efficiency,

integration of DG, and improving economic performance of electricity markets [15]. Allowing the interaction between the utilities and consumers through an advanced metering infrastructure (AMI), management strategies such as load-side control and peak-shaving enhance the accommodation and consumption of energy obtained from renewable sources [16]. In addition, this idea can be modified to incorporate thermal and gas systems, extending the dimension at which energy resources can be optimally used in urban environments [17]. In this section, important parts of our probabilistic model for urban and rural energy systems are deeply described, including the renewable resources, ambient temperature, DS, and DG based on wind and solar energy.

2.1 Natural resources and energy consumption models

Describing the behaviour of renewable energy sources such as wind speed, solar radiation, and ambient temperature requires the installation of measurement equipment in the place of interest [18], in order to guaranty a reliable estimation of power generation from DG. However, in many cases this is a procedure difficult to carry out because the magnitude of the dataset to be measured becomes too large, so that information from wind and solar atlases is frequently used to estimate the typical meteorological year (TMY). In the case of wind energy, this is frequently represented by a time series following a Weibull distribution [19] with some degree of correlation, with a specific dynamic diurnal profile; these characteristics can be simulated by using first or second order Markov processes, autoregressive moving average (ARMA) models, and evolutionary algorithms, among other techniques [20]. Other factors regarding the specific conditions of a wind farm (WF) such as the installation height and roughness length are commonly considered in the calculation of the energy production. Regarding solar energy, it is frequently represented by a time series built from information typically presented in solar atlases in terms of the global horizontal insolation or by means of diffuse solar radiation models [21,22]. Additional parameters related to the installation of a PVF such as tilt angle and azimuth are included in procedures carried out to estimate energy production. Ambient temperature is a relevant parameter that impacts the performance of renewable DG and DS parameters. TMY of ambient temperature could be estimated by using monthly averages or by means of artificial neural networks [23].

In this work, TMY of wind speed, solar radiation, and ambient temperature are assumed to be previously known on an hourly basis; nevertheless, these variables can be

modelled from the references aforementioned or by applying classical approaches used by Poggi *et al.* [24] for wind speed simulation, by Graham and Hollands [25] for the simulation of solar radiation, and by Erbs *et al.* [26] for the simulation of ambient temperature. The corresponding input data for each of these models can be found in the weather database of National Aeronautics and Space Administration (NASA) [27]. Regarding the electricity consumption, this is a specific variable that strongly depends on the system and conditions under analysis, so a comprehensive method to represent it is described in this paper. As a starting point, the type of load and its variation, as well as its maximum value, autocorrelation and power factor, estimated from field measurements through AMI or by using technical-personnel experience, are supposed to be known for each node and phase of DS. Then, according to the type of load (residential, small commercial, medium commercial, and large commercial), diurnal load profile ($\mu_{HD,h}\ \forall\ h = 1, \dots, H$) is assumed following the curves shown in Fig. 1 [28], whereas the hourly variation ($\sigma_{HD,h}\ \forall\ h = 1, \dots, H$) is expressed as a percentage of the hourly mean ($\mu_{HD,t}$). This variation ($\sigma_{HD,h}$) could be set by a user according to his or her experience and knowledge.

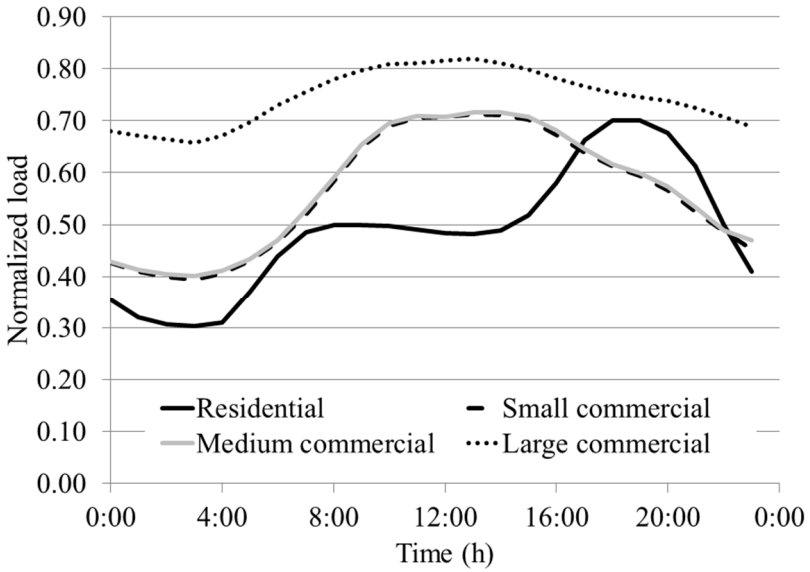

Fig. 1: Typical daily load profiles.

To generate a time series with T elements representing the load connected to the node n of DS, it is necessary to create a random vector ($a_{t,n}\ \forall\ t = 1, \dots, T$) with a determined autocorrelation coefficient ($\rho_{HD,n}$). This is carried out by following (1), where the

random variable ($\gamma_t \; \forall \; t = 1, \dots, T$) has a Gaussian distribution with mean zero and standard deviation equal to $\sqrt{1 - {\rho_{HD,n}}^2}$.

$$a_{t,n} = a_{t-1,n} + \gamma_t \; \forall \; t = 1, \dots, T. \tag{1}$$

Then, the mean ($\mu_{HD,h}$) and standard deviation ($\sigma_{HD,h}$) of hourly values are introduced in the simulation process by following (2),

$$b_{t,n} = \mu_{HD,h} + a_{t,n}\sigma_{HD,h} \; \forall \; t = 1, \dots, T. \tag{2}$$

In the next step, the mean (μ_{DD}) and standard deviation (σ_{DD}) of the daily load profile (Fig. 1) are calculated using (3) and (4); note that these values (μ_{DD} and σ_{DD}) are the mean and standard deviation of the 24 values of the daily profile, whereas the variables ($\mu_{HD,h}$ and $\sigma_{HD,h} \; \forall \; h = 1, \dots, H = 24$) represent the mean and standard deviation of the load profile at the corresponding time h. On the one hand, the series ($\mu_{HD,h}$ and $\sigma_{HD,h}$) are periodical functions; it means $\mu_{HD,t=25} = \mu_{HD,h=1}$ and $\sigma_{HD,t=25} = \sigma_{HD,h=1}$; $\mu_{HD,t=26} = \mu_{HD,h=2}$ and $\sigma_{HD,t=26} = \sigma_{HD,h=2}$;…; $\mu_{HD,t=T} = \mu_{HD,h=H}$ and $\sigma_{HD,t=T} = \sigma_{HD,h=H}$. On the other hand, as the number of hours of a corresponding month under analysis (T) is proportional to the number of hours of the day (H=24h), equation (2) can be successfully evaluated.

$$\mu_{DD} = \frac{1}{24}\sum_{h=1}^{24} \mu_{HD,h}\,; \tag{3}$$

$$\sigma_{DD} = \sqrt{\frac{1}{23}\sum_{l=1}^{24} \left|\mu_{HD,l} - \mu_{DD}\right|^2}. \tag{4}$$

At this stage of the simulation process, a probability transformation [29] is introduced to convert the probability distribution of the time series ($b_{t,n}$) to a Beta distribution, which is a flexible function reasonably able to represent many conditions. Equations (5) and (6) establish the relationship between Gaussian and Beta distributions [30].

$$\alpha_{BD} = \frac{(1 - \mu_{DD}){\mu_{DD}}^2}{{\sigma_{DD}}^2} - \mu_{DD}; \tag{5}$$

$$\beta_{BD} = \left[\frac{1 - \mu_{DD}}{\mu_{DD}}\right]\alpha_{BD}. \tag{6}$$

The aforementioned probability transformation consists of evaluating the series ($b_{t,n}$) in terms of its own cumulative distribution (F_{HD}) and then in terms of the inverse cumulative distribution of the Beta distribution (F_{BD}^{-1}) with parameters (α_{BD}) and (β_{BD}); this process is presented in (7),

$$c_{t,n} = F_{BD}^{-1}\left(F_{HD}(b_{t,n})\right) \forall t = 1, \dots, T. \quad (7)$$

The time series obtained in (7) is limited to the interval [0,1] following the daily profiles described in Fig. 1 with some degree of correlation. Now, this series is scaled according to the maximum demand to be observed per node and phase ($d_{MAX,n}^{a}$; $d_{MAX,n}^{b}$; and $d_{MAX,n}^{c}$). Results are computationally stored in different matrices ($d_{t,n}^{a}$; $d_{t,n}^{b}$; and $d_{t,n}^{c}$) with T rows and N columns, as shown in (8-10).

$$d_{t,n}^{a} = c_{t,n}(d_{MAX,n}^{a}) \forall t = 1, \dots, T; \quad (8)$$

$$d_{t,n}^{b} = c_{t,n}(d_{MAX,n}^{b}) \forall t = 1, \dots, T; \quad (9)$$

$$d_{t,n}^{c} = c_{t,n}(d_{MAX,n}^{c}) \forall t = 1, \dots, T. \quad (10)$$

Active and reactive power is calculated per phase and node using (11)-(16), according to a pre-defined power factor (φ_n).

$$P_{HD,t,n}^{a} = d_{t,n}^{a}(\varphi_n) \; \forall \; t = 1, \dots, T; \quad (11)$$

$$P_{HD,t,n}^{b} = d_{t,n}^{b}(\varphi_n) \; \forall \; t = 1, \dots, T; \quad (12)$$

$$P_{HD,t,n}^{c} = d_{t,n}^{c}(\varphi_n) \; \forall \; t = 1, \dots, T; \quad (13)$$

$$Q_{HD,t,n}^{a} = \left(d_{t,n}^{a}\right) sin\left(cos^{-1}(\varphi_n)\right) \; \forall \; t = 1, \dots, T; \quad (14)$$

$$Q_{HD,t,n}^{b} = \left(d_{t,n}^{b}\right) sin\left(cos^{-1}(\varphi_n)\right) \; \forall \; t = 1, \dots, T; \quad (15)$$

$$Q_{HD,t,n}^{c} = \left(d_{t,n}^{c}\right) sin\left(cos^{-1}(\varphi_n)\right) \; \forall \; t = 1, \dots, T. \quad (16)$$

Once the renewable resources and load demand have been represented, a mathematical model of DS and renewable DG will be carefully described in the next subsections.

2.2 Distribution system modelling

A general-purposes DS model [28] has been used in this paper to illustrate the capabilities of the methodology developed for the integration of DG. Based on the

sequence analysis of DS, general expressions for self and mutual impedances can be derived and used to build and approximate the impedance matrix for each branch of DS. This reasoning allows us to describe typical systems with three or four wires, both unigrounded and multigrounded, using the mathematical model briefly presented in (17-24). In (17), positive impedance is presented, and the influence of earth resistivity and system frequency is shown in (18). Depending on the type of system (a three-wire, four-wire multigrounded, or four-wire unigrounded system), zero-sequence impedance undergoes some changes: equation (19) presents zero-sequence impedance for a three-wire system, equation (20) shows zero-sequence impedance for a four-wire multigrounded system, and equation (21) describes zero-sequence impedance for a four-wire unigrounded system. Finally, the relationship between self and mutual impedances with the impedance matrix is shown in (22)-(24).

$$Z_m^+ = Z_m^- = R_m + \left(\sqrt{-1}\right)(k_m) log_{10}\left(\frac{GMD_m}{GMR_m}\right); \tag{17}$$

$$D_e = 2160\sqrt{\frac{\rho_{DS}}{f_{DS}}}; \tag{18}$$

$$Z_m^0 = R_m + 3R_e + \left(\sqrt{-1}\right)(k_m) log_{10}\left(\frac{D_e}{\sqrt[3]{GMR_m (GMD_m)^2}}\right); \tag{19}$$

$$Z_m^0 = R_m + 3R_e + \left(\sqrt{-1}\right)3(k_m) log_{10}\left(\frac{D_e}{\sqrt[3]{GMR_m (GMD_m)^2}}\right) - 3\left(\frac{{Z_{N,m}^{SF}}^2}{Z_{N,m}^{MT}}\right); \tag{20}$$

$$Z_m^0 = R_m + 3R_{N,m} + \left(\sqrt{-1}\right)\left\{R_{N,m} + 3(k_m) log_{10}\left(\frac{{GMD_{N,m}}^2}{GMR_{N,m}\sqrt[3]{GMR_m (GMD_m)^2}}\right)\right\}; \tag{21}$$

$$Z_m^{SF} = \frac{Z_m^0 + 2Z_m^+}{3}; \tag{22}$$

$$Z_m^{MT} = \frac{Z_m^0 + 2Z_m^+}{3}; \tag{23}$$

$$Z_m = \begin{bmatrix} Z_m^{SF} & Z_m^{MT} & Z_m^{MT} \\ Z_m^{MT} & Z_m^{SF} & Z_m^{MT} \\ Z_m^{MT} & Z_m^{MT} & Z_m^{SF} \end{bmatrix}. \tag{24}$$

2.3 Distributed generation modelling

The structure of DG analysed in this paper is based on renewable energies; Fig. 2 briefly shows some of the components taken into account in our study. WF has been represented by a single WT, the power production of which has been scaled according to the number of turbines; however, models of higher complexity could also be integrated by considering the power curve of a determined manufacturer and the specific geographical conditions and topography. Regarding the PV generator, a detailed model of the PV panel was scaled according to the number of panels in the farm, the effects of maximum power point tracking (MPPT) algorithm on power production, and the variable efficiency of the power converter has been incorporated. The influence of the interconnection transformer on DG performance has not been included, as the efficiency of such transformers can be considered as close to unity.

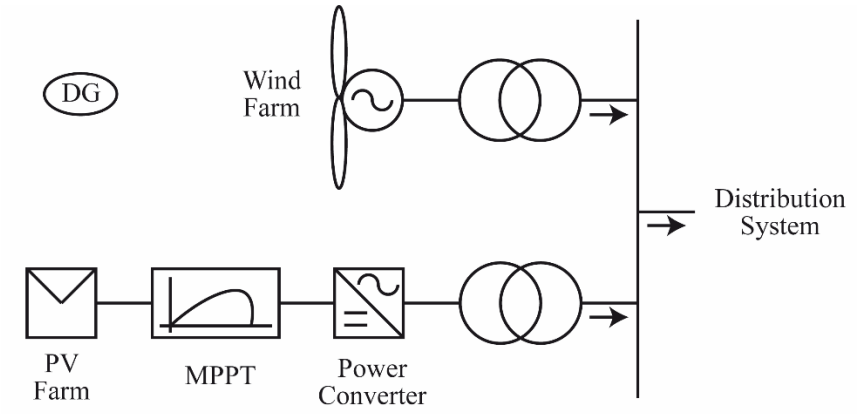

Fig. 2: Scheme of a DG based on wind and solar energies.

The model of this conversion system is carefully described in the following subsections.

2.3.1 Wind farm power generation

Currently, WT generation is estimated using its power curve, which expresses the power production ($P_{WT,t,n}$) as a function of the wind speed (u_t). This expression can be found easily in the WT datasheets offered by manufacturers. Based on these datasheets, analytical models have been developed to support researchers and technicians. In (25)-(28) [31] the model used in this paper is presented, in which only the cut-in, rated, and cut-out speed, and the rated power of WT, are required.

$$P_{WT,t,n} = \begin{cases} 0; 0 \le u_t < U_{WT,i}^{IN}, u_t > U_{WT,i}^{OUT} \\ (\alpha_{WT,i}) + (\beta_{WT,i})(u_t) + (\theta_{WT,i})(u_t)^2;\ U_{WT,i}^{IN} \le u_t \le U_{WT,i}^{NOM}; \\ P_{WT,i}^{NOM};\ U_{WT,i}^{NOM} \le u_t \le U_{WT,i}^{OUT} \end{cases} \quad (25)$$

$$\alpha_{WT,i} = \frac{1}{\left(U_{WT,i}^{IN} - U_{WT,i}^{NOM}\right)^2}\left[U_{in}^{IN}\left(U_{WT,i}^{IN} + U_{WT,i}^{NOM}\right) - 4U_{WT,i}^{IN}U_{WT,i}^{NOM}\left(\frac{U_{WT,i}^{IN} + U_{WT,i}^{NOM}}{2U_{WT,i}^{NOM}}\right)^3\right]; \quad (26)$$

$$\beta_{WT,i} = \frac{1}{\left(U_{WT,i}^{IN} - U_{WT,i}^{NOM}\right)^2}\left[4\left(U_{WT,i}^{IN} + U_{WT,i}^{NOM}\right)\left(\frac{U_{WT,i}^{IN} + U_{WT,i}^{NOM}}{2U_{WT,i}^{NOM}}\right)^3 - \left(3U_{WT,i}^{IN} + U_{WT,i}^{NOM}\right)\right]; \quad (27)$$

$$\theta_{WT,i} = \frac{1}{\left(U_{WT,i}^{IN} - U_{WT,i}^{NOM}\right)^2}\left[2 - 4\left(\frac{U_{WT,i}^{IN} + U_{WT,i}^{NOM}}{2U_{WT,i}^{NOM}}\right)^3\right]. \quad (28)$$

The active power to be injected into the DS is modelled by scaling the power production of a single WT according to the number of turbines on the farm (N_{WT}). This idea is shown in (29).

$$P_{WF,t,n}^{a} = P_{WF,t,n}^{b} = P_{WF,t,n}^{c} = -\left(P_{WT,t,n}\right)N_{WT}. \quad (29)$$

On the other hand, it is assumed that WF works at a power factor close to unity, which means that no reactive power is injected into DS. This assumption is presented in (30).

$$Q_{WF,t,n}^{a} = Q_{WF,t,n}^{b} = Q_{WF,t,n}^{c} = 0. \quad (30)$$

2.3.2 Photovoltaic farm power generation

The building mathematical model of PVF starts with the description of a typical solar cell.

Considering the manufacturer j during time instant t, ambient temperature ($T_{AMB,t}$) is combined with the solar radiation (G_t), nominal operating cell temperature (NOCT), cell efficiency ($\eta_{PV,j}$), and temperature coefficient ($\alpha_{PV,j}$) to estimate the PV cell temperature according to (31) [32].

$$T_{PV,t} = \frac{T_{AMB,t} + G_t\left(\frac{NOCT_j - 20\ ^{o}C}{800\ W/m^2}\right)\left[1 - \frac{\eta_{PV,j}\left(1 - \alpha_{PV,j}25\ ^{o}C\right)}{k_{PV}}\right]}{1 + \left(NOCT_j - 20\ ^{o}C\right)\left(\frac{G_t}{800\ W/m^2}\right)\left(\frac{\alpha_{PV,j}\eta_{PV,j}}{k_{PV,j}}\right)}. \quad (31)$$

The parameter $k_{PV,j}$ is related to the solar transmittance and absorptance, and it trends to 0.9 ($k_{PV,j}\rightarrow 0.9$) [32]. Using information reported by the manufacturers measured under the standard test condition (STC), such as open-circuit voltage and short-circuit current, equations (32)-(40) [33] can be used to approximate cell voltage, current, and power, following specific meteorological conditions.

$$V_{PV,t}^{e} = \frac{m_{PV} k_B T_{PV,t}}{e_{PV}}; \tag{32}$$

$$V_{PV,t}^{OC} = \frac{V_{PV,STC,j}^{OC}}{N_{PV,j}^{CS}} + \alpha_{PV,j}\left(T_{PV,t} - 25\ ^{\circ}C\right) + V_{PV,t}^{e} log\left(\frac{G_t}{1000\ W/m^2}\right); \tag{33}$$

$$v_{PV,t}^{OC} = \frac{V_{PV,t}^{OC}}{V_{PV,t}^{e}}; \tag{34}$$

$$FF_{PV,t}^{o} = \frac{v_{PV,t}^{OC} - log\left(v_{PV,t}^{OC} + 0.72\right)}{v_{PV,t}^{OC} + 1}; \tag{35}$$

$$I_{PV,t}^{SC} = \frac{I_{PV,STC,j}^{SC}}{N_{PV,j}^{CP}}\left(\frac{G_t}{1000\ W/m^2}\right); \tag{36}$$

$$FF_{PV,j} = \frac{V_{PV,j}^{MAX} I_{PV,j}^{MAX}}{V_{PV,STC,j}^{OC} I_{PV,STC,j}^{SC}}; \tag{37}$$

$$R_{PV,t}^{S} = \left(1 - \frac{FF_{PV,j}}{FF_{PV,t}^{o}}\right)\left(\frac{V_{PV,t}^{OC}}{I_{PV,t}^{SC}}\right); \tag{38}$$

$$I_{PV,t} = I_{PV,t}^{SC}\left[1 - exp\left(\frac{V_{PV,t} - V_{PV,t}^{OC} + I_{PV,t} R_{PV,t}^{S}}{V_{PV,t}^{e}}\right)\right]; \tag{39}$$

$$P_{PV,t} = I_{PV,t}\left[V_{PV,t}^{OC} - I_{PV,t} R_{PV,t}^{S} + V_{PV,t}^{e} log\left(1 - \frac{I_{PV,t}}{I_{PV,t}^{SC}}\right)\right]. \tag{40}$$

Once cell power has been analytically approximated through (40), the effects of the MPPT algorithm on power production has been modelled using the approach recently proposed in [34] based on the golden section search algorithm. Regarding the power converter, it has been dimensioned according to the peak-power of the panel under simulation (j=1,…,J), and it has later been scaled by following the size of the farm. It is expressed in (41).

$$P_{INV,j,n} = \left(I_{PV,j}^{MAX} V_{PV,j}^{MAX}\right)\left(N_{PV,j}^{PS} N_{PV,j}^{PP}\right). \quad (41)$$

The efficiency of the power converter reduces as the solar power production reduces; this fact is simulated through (42) and (43) [35].

$$p_{INV,t} = \frac{P_{PV,t}\left(N_{PV,j}^{CS} N_{PV,j}^{CP}\right)\left(N_{PV,j}^{PS} N_{PV,j}^{PP}\right)}{P_{INV,j,n}}; \quad (42)$$

$$\eta_{INV,t} = \frac{p_{INV,t}}{p_{INV,t} + \alpha_{INV,j} + \beta_{INV,j}\left(p_{INV,t}\right) + \theta_{INV,j}\left(p_{INV,t}\right)^2}. \quad (43)$$

Finally, three-phase power injection from the solar source is shown in (44) and (45), assuming a power factor close to unity in order to take advantage of all active power produced [36].

$$P_{PVF,t,n}^{a} = P_{PVF,t,n}^{b} = P_{PVF,t,n}^{c} = -\left(P_{PV,t}\eta_{INV,t}\right)\left(N_{PV,j}^{PS} N_{PV,j}^{PP}\right)\left(N_{PV,j}^{CS} N_{PV,j}^{CP}\right); \quad (44)$$

$$Q_{PVF,t,n}^{a} = Q_{PVF,t,n}^{b} = Q_{PVF,t,n}^{c} = 0. \quad (45)$$

2.3.3 Smart grid real-time pricing

The concept of smart grid involves the application of dynamic hourly prices resulting from the wholesale electricity market operation in order to incentivize end-users to respond to this signal and reduce their energy consumption at peak hours. In our study, the real-time price signal is modelled according to (46)-(48), under the assumption that electricity prices are highly correlated to the total hourly demand.

$$P_{THD,t} = P_{HD,t,n}^{a} + P_{HD,t,n}^{b} + P_{HD,t,n}^{c} \; \forall \; t = 1, \ldots, T; \quad (46)$$

$$P_{THD}^{MAX} = max\left(P_{THD,t} \; \forall \; t = 1, \ldots, T\right); \quad (47)$$

$$E_{SG,t} = E_{SG}^{MAX}\left(\frac{P_{THD,t}}{P_{THD}^{MAX}}\right). \quad (48)$$

3. Proposed methodology

The method presented in this paper aims to determine the appropriate capacity to be installed in a determined node of DS to avoid the problems related to reverse power flow at the substation. The integration of DG strongly depends on decisions of private investors, which are not concerned about the technical limitations of DS. In this regard, the proposed methodology can be used to define the technical limitations and

consequences of integrating renewable DG on a determined node of DS. In other words, the computational model developed in this work is a tool to be used by the utility in order to establish the proper amount of DG could be technically affordable. This defines the highest capacity to be installed by the private investors.

From a computational perspective, evaluating every hour of the year, considering a three-phase unbalanced power flow in a DS with a large number of nodes, could be a very difficult task.

As stated before, our analysis is focused on finding an economically suitable solution that prevents the generation of a reverse power flow at the substation as a result of excessive integration of renewable power sources. In order to reduce the computational complexity of the problem, the months with the highest renewable potential are selected. Equation (49) [37] could be used to estimate the wind power potential (wp_q) of a determined month (q) for the location under analysis. The application of this concept gives us a qualitative description of wind potential on a monthly basis. In the case of solar resources, for a determined month (q), the average radiation in the daytime (pvp_q) could be used to qualitatively evaluate the solar potential; this reasoning is presented in (50). Finally, the addition of both results ($wp_q + pvp_q \; \forall \; q = 1, \dots, Q = 12$) offers a global perspective on those months with the highest renewable potential. However, other methodologies with higher accuracy could also be considered, such as the methodology recently proposed in [38] to improve the quality and reliability of the obtained design.

$$wp_q = \frac{\rho_{AD}}{2T_q} \sum_{l=1}^{T_q} (u_l)^3 \; \forall \; q = 1, \dots, Q; \tag{49}$$

$$pvp_q = \frac{1}{T_{SH,q}} \sum_{l=1}^{T_{SH,q}} \{G_l |\; G_l > 0\} \; \forall \; q = 1, \dots, Q. \tag{50}$$

Then, depending on the computational resources available, several months with high renewable potential are selected, and in this way, the number of hours to be evaluated in our model (T) is defined by the number of hours of the chosen months. If only one month is selected, it could be described as (51).

$$T = T_l |\; wp_l + pvp_l = max\{wp_q + pvp_q \; \forall \; q = 1, \dots, Q\}. \tag{51}$$

GA has been extensively discussed for the analysis of energy systems [39]; in the proposed implementation, each individual (w=1,…W) is described by the vector shown in (52):

$$\overrightarrow{gp}_w = [gp_1^w \quad gp_2^w \quad gp_3^w \quad gp_4^w]; \tag{52}$$

where each element gp_1^w, gp_2^w, gp_3^w, and gp_4^w is an integer number. The first element gp_1^w represents the manufacturer of the PV panel ($gp_1^w = j$). The second element gp_2^w is the number of PV strings (groups of PV panels connected in parallel) built using the panels of manufacturer (j). The third element gp_3^w is the manufacturer of WT ($gp_3^w = i$), while the fourth element gp_4^w is the number of WTs of type (i) connected to complete the corresponding WF size. The maximum number of PV strings (pvs_n^{MAX}) and WTs (wt_n^{MAX}) to be connected to a determined node (n) could be determined by measuring the available geographical space and layout and the local conditions and laws. To sum up, the first element is defined by the index $1 \leq gp_1^w \leq J$, the second element is defined according to the condition $1 \leq gp_2^w \leq pvs_n^{MAX}$, the third element is defined by the index $1 \leq gp_3^w \leq I$, while the fourth element is defined by the condition $1 \leq gp_4^w \leq wt_n^{MAX}$.

The population of GA is formed by (W) individuals grouped in the matrix (GP), as shown in (53).

$$GP = \begin{bmatrix} \overrightarrow{gp_1} \\ \vdots \\ \overrightarrow{gp_w} \\ \vdots \\ \overrightarrow{gp_W} \end{bmatrix}. \tag{53}$$

The proposed GA-based method consists of the steps described as follows:

Step 1: Define the number of generations (Y), the population size (W), the crossover and mutation rates of GA, the significance level of the probabilistic analysis, the maximum number of PV strings (pvs_n^{MAX}), the maximum number of WTs (wt_n^{MAX}), and the size of WT and PV-panel datasets (I and J). Additionally, select the month or months of the year under study using (49)-(51).

Step 2: Create the initial population of GA. To perform this step, the structure of PV panels and WT datasets has to be considered. The PV panel dataset is built by compiling information about (J) manufacturers regarding the open circuit voltage, short circuit current, and efficiency, etc. The WT dataset, on the other hand, consists of a compilation of power curves or a general perspective on some characteristics such as cut-in, cut-out, and the rated speed of (I) manufacturers. Then, using an algorithm for integer random number generation, the elements $gp_1^w \; \forall \, w = 1, \ldots, W$ are initialized using a random number in the interval [1,J]; the elements $gp_2^w \; \forall \, w = 1, \ldots, W$ are initialized using random numbers in the interval [1,pvs_n^{MAX}]; the elements $gp_3^w \forall \, w = 1, \ldots, W$ are initialized using random numbers in the interval [1,I]; and the elements $gp_4^w \forall \, w = 1, \ldots, W$ are initialized using numbers in the range [1,wt_n^{MAX}].

Step 3: Analyze the first generation by assigning $y \leftarrow 1$.

Step 4: Evaluate each individual of the population ($\overrightarrow{gp_w}$), determining NPC and the probability of reverse power observed at the substation. The number of PV panels to be connected in a series are adjusted according to a determined direct current voltage (V_{DC}) and the PV-panel manufacturer, as shown in (54).

$$N_{PV,j}^{PS} = INT\left(\frac{V_{DC}}{V_{PV,STC,j}^{OC}}\right). \tag{54}$$

The evaluation of the corresponding individual (w) is carried out by performing probabilistic power flow analysis [40-43] using the information (wind speed, solar radiation, ambient temperature, and load demand) for each hour (t=1,…,T) of the month previously selected. Based on this analysis, active power obtained from the utility per phase ($P_{SUB,t}^a$, $P_{SUB,t}^b$, and $P_{SUB,t}^c$) is used to build the discretized distributions of the active power at the substation per phase. These distributions are then used to approximate the probability of observing reverse power flow at the substation ($P_r\{P_{SUB,t}^a < 0\}$, $P_r\{P_{SUB,t}^b < 0\}$, and $P_r\{P_{SUB,t}^c < 0\}$). NPC is approximated through (55) [32], where annualized capital cost (ACC), annualized replacement cost (ARC), and annualized maintenance cost (AMC) are integrated based on the capital recovery factor (CRF) concept. Due to the fact that we are not evaluating the entire year, the estimated

NPC is scaled by the factor (8760/T) to report a better approximation in terms of the order of magnitude.

$$NPC = \frac{ACC + ARC + AMC}{CRF}\left(\frac{8760}{T}\right). \quad (55)$$

The cost related to the electricity obtained from the utility is included in the term (AMC). This is estimated as the product of the electricity price and the power delivered by the substation ($E_{SG,t}P^{a}_{SUB,t} + E_{SG,t}P^{b}_{SUB,t} + E_{SG,t}P^{c}_{SUB,t}\ \forall\ t = 1, \dots, T$). Regarding the value of the objective function, if the maximum probability of the reverse power flow at the substation, expressed in terms of (56),

$$max\left(P_r\left\{P^{a}_{SUB,t} < 0\right\}, P_r\left\{P^{b}_{SUB,t} < 0\right\}, P_r\left\{P^{c}_{SUB,t} < 0\right\}\right); \quad (56)$$

is higher than the half of the significance level previously selected, a value artificially high is assigned to the NPC ($NPC \rightarrow \infty$); in the contrary case, the estimation described in (55) is adopted.

Step 5: Based on the NPC previously obtained, the fitness of each individual is assigned according to its rank; the individual with the lower NPC obtains rank 1, while the individual with the highest NPC obtains rank W. Fitness value ($f_w\ \forall\ w = 1, \dots, W$) is assigned according to (57).

$$f_w = \frac{(W+1) - w}{\sum_{l=1}^{W}[(W+1) - l]}\ \forall\ w = 1, \dots, W. \quad (57)$$

Step 6: Reproduction, crossing, and mutation operators are performed in order to define the individuals that belong to the next generation.

Step 7: If ($y < Y$), assign $y \leftarrow y + 1$ and go to Step 4; otherwise, stop.

4. Case study

The probabilistic optimization algorithm proposed in this work is illustrated by means of a hypothetical DS located in Zaragoza, Spain. Field measurements of wind speed and ambient temperature [44] have been used; however, solar radiation time series have been simulated using information from the NASA database [27] combined with the

general-purpose model proposed by Graham and Hollands [25]. Both operating installations and installations currently undergoing design were assumed to have azimuth equal to 0° and tilt angle equal to 60°. DS has 35 nodes and branches (N=M=35); between branches 1 (m=1) and 15 (m=15), the phase and neutral conductors were assumed to be 1000-kcmil, all-aluminum conductor (AAC) types. Between branches 16 (m=16) and 35 (m=35), the phase and neutral conductors were assumed to be 105.6 kcmil AAC type. This information is combined with the structure shown in Fig. 3, where the assumed disposition of each conductor on the feeder is presented; this allows us to calculate the geometric mean radius and the geometric mean distance.

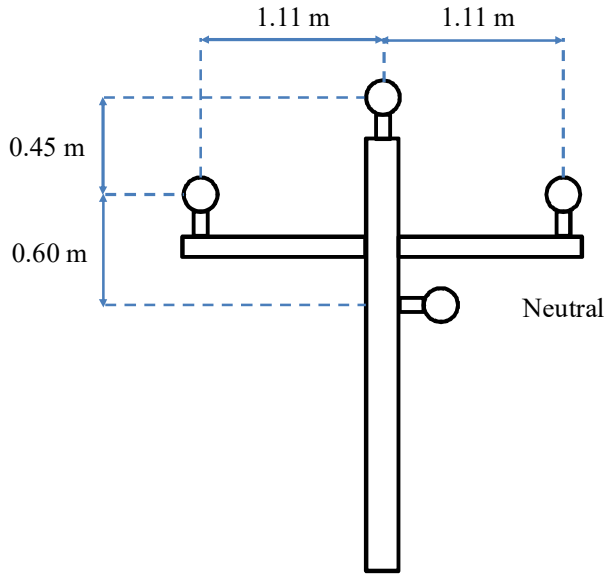

Fig. 3: Structure of the distribution feeder.

The system is a four-wire multigrounded one; its rated voltage is 12.47 kV; the substation capacity is 15000 kVA, its system frequency is 50 Hz, and the earth resistivity is 100 Ωm. The probabilistic power flow was solved by considering a tolerance of 0.000001 for each time step t. Load demand was modeled according to subsection 2.1 using the information presented in Table 1 combined with the profiles shown in Fig. 1; in addition, the hourly load variability was assumed to be 10% of the corresponding mean value ($\sigma_{HD,h} = 0.1\mu_{HD,h}\ \forall\ h = 1, \ldots, 24$). Table 1 presents relevant information related to DS topology and its load type, as well as actual condition; the system currently under operation or already installed is described by the condition *Operating* in column five, while the system under optimal sizing is described by the condition *Designing*, which corresponds to node 32 (n=32).

Table 1: DS topology and load data.

From	To	Length (km)	Node Type	Condition	Load Type
Substation	1	3.3989	Load	-----	Residential
1	2	3.3030	Load	-----	Residential
2	3	3.8243	Load	-----	Residential
3	4	3.6802	Load	-----	Residential
4	5	3.9330	Generation	Operating	-----
5	6	3.6691	Load	-----	Residential
6	7	3.7648	Load	-----	Residential
7	8	3.8022	Load	-----	Small commercial
8	9	3.8136	Generation	Operating	-----
9	10	3.2853	Load	-----	Residential
10	11	4.3452	Generation	Operating	-----
11	12	4.1699	Load	-----	Residential
12	13	4.3158	Load	-----	Medium commercial
13	14	4.4843	Load	-----	Residential
14	15	5.4739	Load	-----	Residential
2	16	3.4486	Load	-----	Residential
16	17	4.0128	Load	-----	Residential
17	18	4.4717	Load	-----	Small commercial
18	19	4.4335	Generation	Operating	-----
19	20	4.4200	Load	-----	Residential
4	21	4.0573	Load	-----	Residential
21	22	4.2988	Load	-----	Residential
6	23	4.2569	Load	-----	Residential
23	24	4.0246	Load	-----	Residential
8	25	4.4762	Generation	Operating	-----
25	26	4.2766	Load	-----	Residential
10	27	4.0660	Load	-----	Residential
27	28	4.4021	Load	-----	Small commercial
27	29	4.2794	Generation	Operating	-----
27	30	4.3483	Load	-----	Residential
12	31	3.8096	Load	-----	Residential
31	32	3.9946	Generation	Designing	-----
32	33	3.7354	Load	-----	Residential
14	34	4.0065	Load	-----	Residential
34	35	4.0941	Load	-----	Large commercial

In Table 2, the power factor ($\varphi_{,n}$), autocorrelation coefficient ($\rho_{HD,n}$), and maximum estimated load per phase ($d^{a}_{MAX,n}$, $d^{b}_{MAX,n}$, and $d^{c}_{MAX,n}$) per node (n=1,…,N=35) and phase are reported. As can be observed, power factor and autocorrelation were assumed to be close to unit, while some residential consumers were modelled by unbalanced loads.

Table 2: Load demand description.

Node	Power factor	Autocorrelation	Maximum load (kVA)		
			Phase a	Phase b	Phase c
1	0.9079	0.8932	150	150	150
2	0.9346	0.8218	25	37.5	25
3	0.9400	0.8351	15	15	15
4	0.9048	0.8104	0	0	25
5	-----	-----	-----	-----	-----
6	0.9385	0.8711	15	0	0
7	0.9262	0.8895	0	0	25
8	0.9426	0.8529	1850	1850	1850
9	-----	-----	-----	-----	-----
10	0.9285	0.8545	0	25	0
11	-----	-----	-----	-----	-----
12	0.9227	0.8714	25	0	0
13	0.9419	0.8193	3250	3250	3250
14	0.9144	0.8629	0	25	0
15	0.9268	0.8366	37.5	37.5	37.5
16	0.9112	0.8601	50	0	0
17	0.9452	0.8884	0	50	0
18	0.9257	0.8070	1700	1700	1700
19	-----	-----	-----	-----	-----
20	0.9070	0.8499	0	0	50
21	0.9251	0.8535	37.5	0	0
22	0.9340	0.8178	0	0	25
23	0.9389	0.8186	50	0	0
24	0.9115	0.8227	0	0	37.5
25	-----	-----	-----	-----	-----
26	0.9362	0.8363	25	0	0
27	0.9179	0.8434	0	50	0
28	0.9048	0.8479	1500	1500	1500
29	-----	-----	-----	-----	-----
30	0.9381	0.8417	25	25	25
31	0.9378	0.8865	0	37.5	0
32	-----	-----	-----	-----	-----
33	0.9293	0.8765	25	0	0
34	0.9370	0.8354	0	25	0
35	0.9434	0.8116	5000	5000	5000

Regarding the optimization process, the WT and PV-panel databases are shown in Table 3 and 4, respectively. In Table 4, the type-one PV panel (j=1) was supposed to have 48 cells connected in serial, while the rest of panels in the database (j=[2,13]) were supposed to have 60 cells. Regarding the cells in parallel, all of the panels were supposed to have only a string of cells under this connection.

Table 3: WT database.

Type	Cut-in speed (m/s)	Rated speed (m/s)	Cut-out speed (m/s)	Rated power (kW)
1	3	12	23	30
2	3	12	25	1500
3	4	14	25	100
4	3	14	25	35
5	3	14	25	25
6	3	13	25	1650
7	3	14	20	250
8	3	14	20	85

Table 4: PV panel dataset.

Type	NOCT (°C)	Open−circuit voltage (V)	Short−circuit current (A)	Temperature coefficient (%/°C)	Efficiency (%)	Point of maximum power	
						Voltage (V)	Current (A)
1	47.9	32.49	9.95	−0.377	19.82	27.53	9.3
2	48	39.4	9.97	−0.4	18.3	31.2	9.63
3	46	38.25	9.6	−0.442	17.4	31.65	9.09
4	46	38.15	9.5	−0.442	17.1	31.55	8.98
5	49	39.3	9.8	−0.4	17.6	31.3	9.25
6	49	39.3	9.87	−0.4	18	31.3	9.42
7	48	45.6	9.08	−0.43	16.2	37.2	8.47
8	44	38.5	9.25	−0.41	16.8	31.1	8.84
9	44	38.4	9.18	−0.41	16.5	30.9	8.73
10	46	37.72	9.06	−0.42	16.2	31.04	8.61
11	46	37.45	8.98	−0.42	15.9	30.79	8.52
12	47	37.6	8.88	−0.42	15.5	30.4	8.38
13	47	37.5	8.76	−0.42	15.2	30.3	8.24

Tables 5, 6, and 7 show the parameters and characteristics of the already installed systems (Condition: *Operating* in Table 1). In Table 6, the PVFs connected to nodes 5 and 19 were built by using PV panels with 48 cells connected in serial and only one string of cells connected in parallel; the rest of the PVFs used PV panels with 60 cells in serial and only one string of cells in parallel. Table 7 presents the maximum number of PV panels to be connected in parallel as well as the DC voltage (V_{DC}) of each PVF; this information was used to determine the number of PV panels to be connected in serial according to (54). Similarly, the maximum number of WTs is also shown. For those nodes with DG already installed (Condition: *Operating*), the columns for the maximum PV strings and maximum WTs represent the systems' characteristics, while for the systems under sizing (Condition: *Designing*), this information was used to establish the limits to be considered in the optimization algorithm mathematically expressed by the

variables (pvs_n^{MAX} for PVF and wt_n^{MAX} for WF). For node 32 specifically, $pvs_{n=32}^{MAX}$=25 and $wt_{n=32}^{MAX}$=20.

Table 5: Operating WFs.

Node	Cut-in speed (m/s)	Rated speed (m/s)	Cut-out speed (m/s)	Rated power (kW)
5	4	14	25	100
9	3	14	25	25
11	3	14	25	35
19	3	14	25	25
25	4	14	25	100
29	3	14	20	250

Table 6: Operating PVFs.

Node	NOCT (°C)	Open-circuit voltage (V)	Short-circuit current (A)	Temperature coefficient (%/°C)	Efficiency (%)	Point of maximum power	
						Voltage (V)	Current (A)
5	47.9	32.49	9.95	-0.377	19.82	27.53	9.3
9	48	39.4	9.97	-0.4	18.3	31.2	9.63
11	46	38.25	9.6	-0.442	17.4	31.65	9.09
19	47.9	32.49	9.95	-0.377	19.82	27.53	9.3
25	46	38.15	9.5	-0.442	17.1	31.55	8.98
29	49	39.3	9.8	-0.4	17.6	31.3	9.25

Table 7: Installed and maximum allowed renewable penetration.

Node	Condition	Maximum PV strings	DC Voltage (V)	Maximum wind turbines
5	Operating	5	600	1
9	Operating	15	600	1
11	Operating	20	600	1
19	Operating	20	600	1
25	Operating	25	600	1
29	Operating	25	600	1
32	Designing	25	600	20

As a complement to the previously described information, Fig. 4 shows the topology of DS under analysis, and DG is also presented.

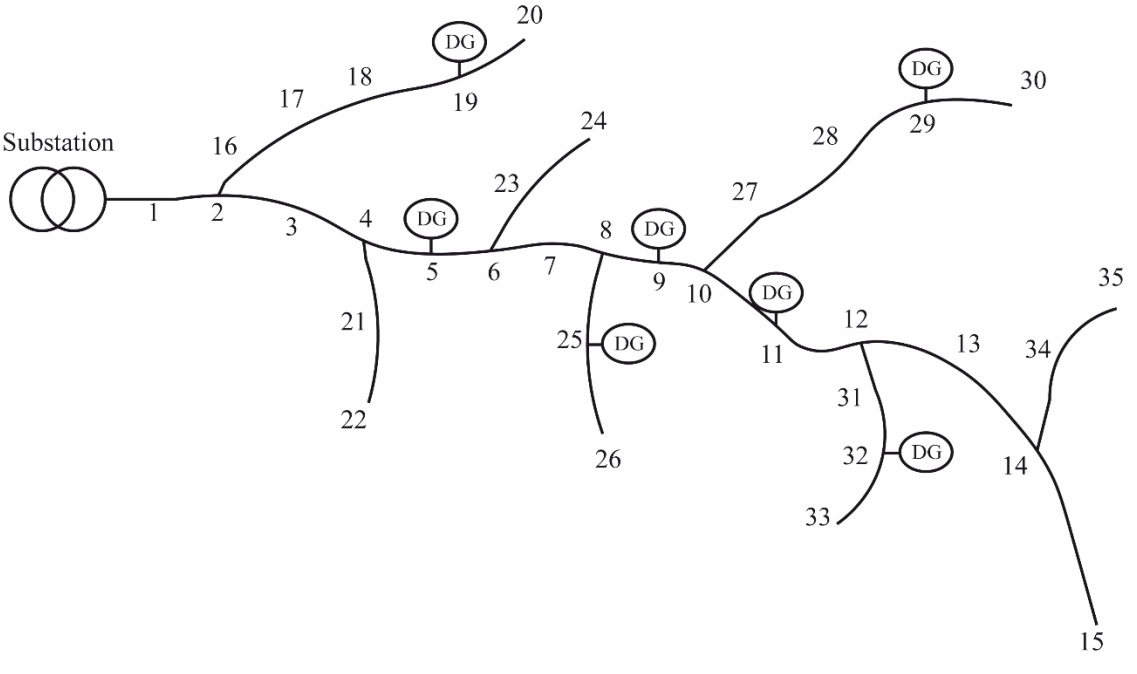

Fig. 4: DS under analysis.

Table 8 as well as Figs. 5, 6, and 7 present the results obtained from the implementation of the procedure described in (49-51); as can be observed in Table 8, the month with highest potential is February, with approximately 869 W/m^2. The corresponding time series for wind speed, solar radiation, and temperature are also presented in the previously mentioned figures. In this way, a total of 672 hours (T=672 h) were analyzed for each individual.

Table 8: Renewable power potential per month.

Month	q	wp_q(W/m^2)	pvp_q(W/m^2)	$wp_q + pvp_q$(W/m^2)
January	1	292.0060	453.2425	745.2485
February	2	336.5908	532.5634	869.1543
March	3	134.1022	571.3557	705.4580
April	4	209.3386	519.3456	728.6843
May	5	145.6036	495.7815	641.3852
June	6	116.1155	448.1411	564.2567
July	7	161.91002	455.5307	617.4408
August	8	204.9154	485.9916	690.9071
September	9	119.83701	548.2797	668.1168
October	10	75.37703	496.9125	572.2895
November	11	150.4374	432.940	583.3777
December	12	147.5994	432.9402	580.5397

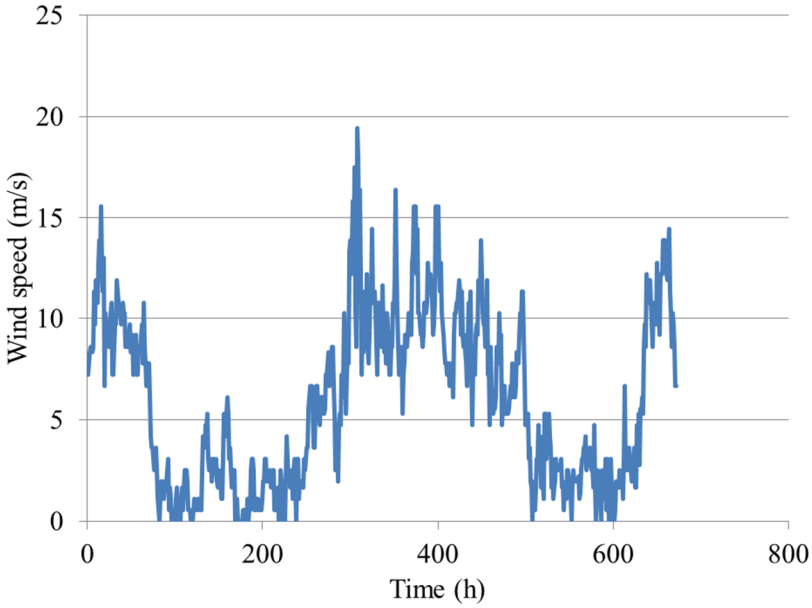

Fig. 5: Wind speed time series (February).

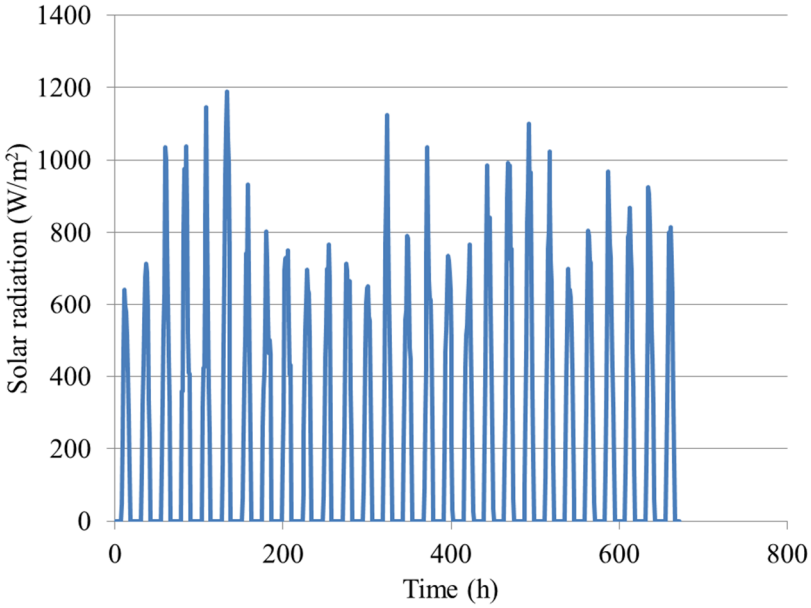

Fig. 6: Solar radiation time series (February).

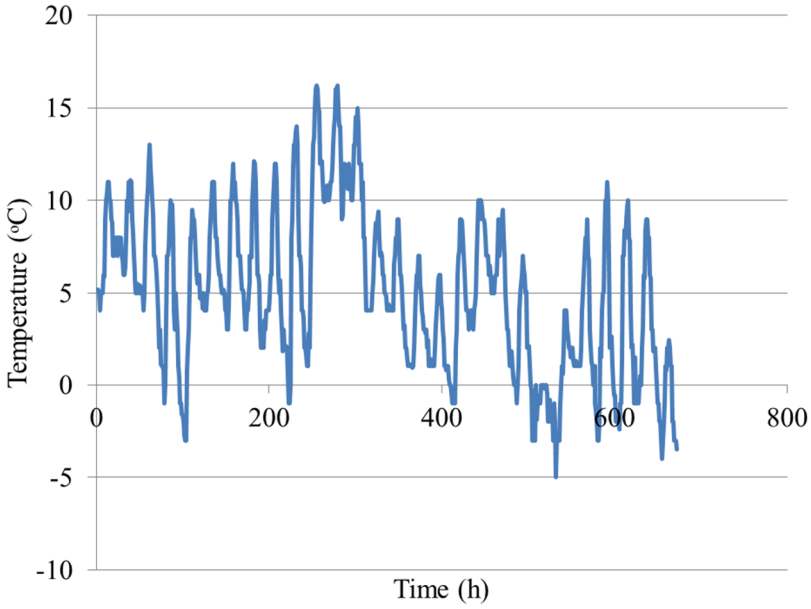

Fig. 7: Ambient temperature (February).

GA was implemented by considering 50 generations (Y=50), a population size of 15 individuals (W=15), and crossover and mutation rates of 0.95 and 0.05, respectively. Under these conditions, Fig. 8 shows the evolution of the implemented optimization process. An optimization algorithm was implemented in MATLAB on a personal computer with an i7-3630QM CPU at 2.40 GHz with 8 GB of memory. The total process required 134.5935 minutes (2.2432 h). Regarding the investment and operating costs, the project was analysed by considering a 35-year lifespan, with discount and inflation rates of 0.04 and 0.02, respectively. The capital costs of the PV panels were assumed to be 1.3 €/W with replacement time of approximately 30 years. Power converter acquisition cost was assumed to be 500 €/kVA with a lifespan of 30 years. The capital costs for WF were supposed to be 850 €/kW, with replacement time of 20 years. Costs related to the energy bought from the utility was estimated by assuming a peak price of 150 €/MWh. After evaluating the 50 generations, the NPC of the suggested configuration was approximately 584 M€, with a maximum probability of observing reverse power equal to 0.148810% or, in other words (58):

$$max\left(P_r\{P^a_{SUB,t} < 0\}, P_r\{P^b_{SUB,t} < 0\}, P_r\{P^c_{SUB,t} < 0\}\ \forall\ t = 1, \dots, 672\right) = 0.001488. \quad (58)$$

The significance level of the analysis was adjusted to 1%.

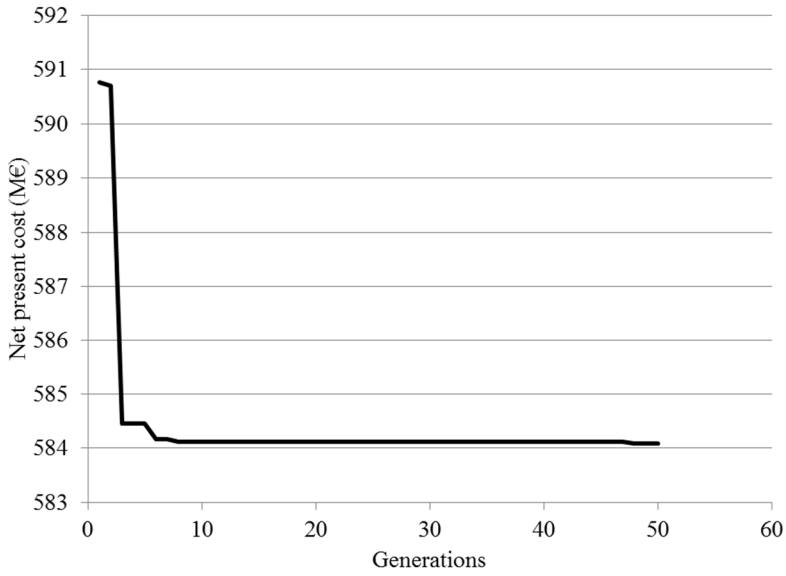

Fig. 8: GA evolution.

The optimal configuration obtained from the proposed methodology consists of a one-string PVF built by using a type-nine (see row 9 in Table 4) PV panel, while wind energy is generated by using four type-two WTs (see row 2 in Table 3). For this configuration, Fig. 9 shows the distribution function of active power flow per phase at the substation level.

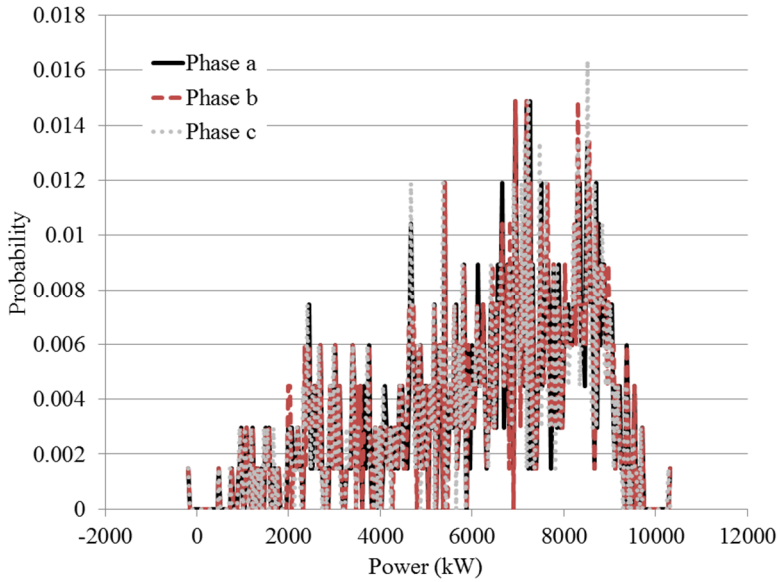

Fig. 9: Probability distribution of active power at the substation.

Finally, Tables 9 and 10 present the average voltage and branch currents as well as the active and reactive power, respectively.

Table 9: Average voltage and current per node and phase.

Node/Branch	Average voltage (kV)			Average current (A)		
	Phase a	Phase b	Phase c	Phase a	Phase b	Phase c
1	12.54	12.425	12.445	325.869	327.051	325.394
2	12.54	12.425	12.445	322.574	323.723	322.075
3	12.536	12.467	12.408	279.218	279.996	278.777
4	12.534	12.467	12.409	278.899	279.674	278.454
5	12.543	12.454	12.413	278.088	279.674	277.372
6	12.543	12.452	12.415	278.764	280.362	278.053
7	12.564	12.439	12.408	277.367	280.362	277.238
8	12.566	12.435	12.41	277.367	280.362	276.696
9	12.573	12.415	12.423	233.74	236.62	233.096
10	12.575	12.41	12.425	234.304	237.2	233.662
11	12.511	12.444	12.455	199.884	201.097	199.211
12	12.51	12.441	12.459	200.541	201.762	199.866
13	12.495	12.454	12.461	258.509	260.175	258.907
14	12.495	12.453	12.462	173.269	174.652	173.43
15	12.473	12.469	12.467	0.848	0.849	0.849
16	12.473	12.429	12.508	44.364	44.496	44.321
17	12.48	12.43	12.499	43.233	44.496	44.321
18	12.478	12.442	12.49	43.233	43.363	44.321
19	12.476	12.455	12.479	1.083	1.085	1.186
20	12.474	12.468	12.469	0	0	1.132
21	12.46	12.481	12.469	0.849	0	0.566
22	12.465	12.478	12.467	0	0	0.566
23	12.45	12.482	12.478	1.133	0	0.848
24	12.461	12.48	12.47	0	0	0.848
25	12.465	12.486	12.459	1.685	1.853	1.857
26	12.465	12.478	12.467	0.566	0	0
27	12.533	12.434	12.444	36.748	37.957	36.818
28	12.486	12.461	12.463	38.996	39.078	39.07
29	12.486	12.461	12.463	3.169	3.175	3.175
30	12.486	12.461	12.463	0.565	0.566	0.566
31	12.486	12.455	12.469	75.212	75.304	75.401
32	12.479	12.46	12.471	75.212	75.465	75.401
33	12.472	12.465	12.473	0.565	0	0
34	12.49	12.454	12.467	172.422	173.24	172.582
35	12.48	12.462	12.468	172.422	172.673	172.582

From the results presented in Table 9, it is possible to observe how a relatively small imbalance level on the system can perturb the voltage profile depending on the phase, the type of load connected, the reverse power flow, and the integration of DG. In addition, according to the average active power reported in Table 10, it is possible to observe how the DG system suggested to be installed in node 32 was able to reduce local power consumption, specifically in the neighborhood of branches 31, 32, and 33. This local influence on the reduction of energy consumption allows us to prevent the generation of reverse power flow at the substation. Another important aspect is that the

DG system is at least 40% of the substation's capacity and is mainly composed of wind generation, due to the assumption about the system's integration cost and lifespan as well as the availability of this resource.

Table 10: Average active and reactive power per node and phase.

Branch	Average active power (kW)			Average reactive power (kVAr)		
	Phase a	Phase b	Phase c	Phase a	Phase b	Phase c
1	6272.575	6271.847	6242.12	3013.559	3012.03	3004.082
2	6206.024	6205.296	6175.569	2982.841	2981.311	2973.363
3	5310.778	5303.52	5280.424	2606.844	2605.24	2597.146
4	5303.889	5296.631	5273.535	2604.346	2602.741	2594.648
5	5286.935	5296.631	5251.074	2597.388	2602.741	2585.085
6	5310.071	5319.767	5274.21	2597.388	2602.741	2585.085
7	5280.248	5319.767	5257.507	2586.454	2602.741	2577.548
8	5280.248	5319.767	5246.191	2586.454	2602.741	2572.943
9	4328.387	4379.343	4305.768	2235.023	2255.605	2225.806
10	4346.706	4397.662	4324.087	2235.023	2255.605	2225.806
11	3640.634	3657.82	3618.015	1871.747	1878.101	1862.529
12	3664.752	3681.937	3642.132	1871.747	1878.101	1862.529
13	5270.768	5293.389	5270.768	1862.529	1871.74	1862.529
14	3533.098	3555.719	3533.098	1242.947	1252.157	1242.947
15	16.987	16.987	16.987	6.882	6.882	6.882
16	883.829	884.651	883.728	371.652	369.556	371.873
17	861.561	884.651	883.728	361.586	369.556	371.873
18	861.561	861.561	883.728	361.586	361.586	371.873
19	−23.398	−23.398	−1.231	0	0	10.287
20	0	0	22.167	0	0	10.287
21	16.954	0	11.408	6.958	0	4.363
22	0	0	11.408	0	0	4.363
23	22.942	0	16.703	8.404	0	7.536
24	0	0	16.703	0	0	7.536
25	−28.631	−40.068	−40.068	4.294	0	0
26	11.437	0	0	4.294	0	0
27	706.072	728.5	706.072	363.277	372.969	363.277
28	763.139	763.139	763.139	359.046	359.046	359.046
29	−68.528	−68.528	−68.528	0	0	0
30	11.461	11.461	11.461	4.231	4.231	4.231
31	−1617.29	−1611.45	−1628.64	4.51	6.361	0
32	−1617.29	−1628.64	−1628.64	4.51	0	0
33	11.35	0	0	4.51	0	0
34	3516.11	3527.562	3516.11	1236.065	1240.332	1236.065
35	3516.11	3516.11	3516.11	1236.065	1236.065	1236.065

Regarding the practical applicability of the proposed optimization method, it could be used to perform renewable energy integration analysis for a DS, starting with typical load profiles, hourly estimations of solar and wind resources, and data frequently provided by PV panels and WT manufacturers. The methodology was only illustrated

for node 32; however, it could sequentially implemented in other nodes by changing the DG condition in node 32 (Tables 1 and 7) from *Designing* to *Operating*; then, the sizing procedure would be repeated in the new node where DG needs to be added.

5. Conclusions

In this work, an innovative method for optimal sizing and incorporation of DG into a DS based on wind and solar energy has been described, illustrated, and discussed. The proposed methodology integrates heuristic optimization, load flow analysis, typical consumption profiles, and environmental information in a probabilistic theory for sizing DG to prevent the apparition of reverse power flow at distribution substations, as a new contribution to earlier studies. In addition, the presented methodology allows the integration of manufacturer data that are typically found on datasheets related to PV panels and WTs. The results obtained from the analysis of a case study reveal that DG should be integrated to supply energy demand at a local level; in our case, the optimal capacity was found to be approximately 40% of the substation's capacity. This method can be used to carry out an analysis about the incorporation of renewable energy generation into a DS.

Acknowledgment

J.P.S. Catalão acknowledges the support by FEDER funds through COMPETE 2020 and by Portuguese funds through FCT, under Projects SAICT-PAC/0004/2015 - POCI-01-0145-FEDER-016434, POCI-01-0145-FEDER-006961, UID/EEA/50014/2013, UID/CEC/50021/2013, UID/EMS/00151/2013, and 02/SAICT/2017 - POCI-01-0145-FEDER-029803, and also funding from the EU 7th Framework Programme FP7/2007-2013 under GA no. 309048. Moreover, this work was supported by the Ministerio de Economía y Competitividad of the Spanish Government under Project ENE2013-48517-C2-1-R.

References

[1] Annual energy outlook 2018 with projections to 2050. U. S. Energy Information Administration, https://www.eia.gov/outlooks/aeo/; 2018 [accessed 17.07.2018].

[2] Jain S, Kalambe S, Agnihotri G, Mishra A. Distributed generation deployment: State-of-the-art of distribution system planning in sustainable era. Renew Sust Energ Rev 2017;77:363-385.

[3] Lin C-H, Hsieh W-L, Chen C-S, Hsu C-T, Ku T-T. Optimization of photovoltaic penetration in distribution systems considering annual duration curve of solar irradiation. IEEE Trans Power Syst 2012;27(2):1090-1097.

[4] Agalgaonkar YP, Pal BC, Jabr RA. Stochastic distribution system operation considering voltage regulation risks in the presence of PV generation. IEEE Trans Sustain Energy 2015;6(4):1315-1324.

[5] Peng X, Lin L, Zheng W, Liu Y. Crisscross optimization algorithm and Monte Carlo simulation for solving optimal distributed generation allocation problem. Energies 2015;8:13641-13659.

[6] El Batawy SA, Morsi WG. Optimal secondary distribution system design considering rooftop solar photovoltaics. IEEE Trans Sustain Energy 2016;7(4):1662-1671.

[7] Pillai G, Putrus G, Pearsall N, Georgitsioti T. The effect of distribution network on the annual energy yield and economic performance of residential PV systems under high penetration. Renew Energy 2017;108:144-155.

[8] Ahmadigorji M, Amjady N, Dehghan S. A novel two-stage evolutionary optimization method for multiyear expansion planning of distribution systems in presence of distributed generation. Appl Soft Comput 2017;52:1098-1115.

[9] Simulation Tool – OpenDSS. Electrical Power Research Institute (EPRI), http://smartgrid.epri.com/SimulationTool.aspx; 2018 [accessed 17.07.2018].

[10] GridLAB-D. Pacific Northwest National Laboratory (PNNL), http://www.gridlabd.org/; 2018 [accessed 17.07.2018].

[11] Huda ASN, Živanović R. Large-scale integration of distributed generation into distribution networks: Study objectives, review of models and computational tools. Renew Sust Energ Rev 2017;76:974-988.

[12] Martinez-Velasco JA, Guerra G. Reliability analysis of distribution systems with photovoltaic generation using a power flow simulator and a parallel Monte Carlo approach. Energies 2016;9:1-21.

[13] Li Q, Ayyanar R, Vittal V. Convex optimization for DES planning and operation in radial distribution systems with high penetration of photovoltaic resources. IEEE Trans Sustain Energy 2016;7(3):985-995.

[14] Hariri A, Faruque MO. A hybrid simulation tool for the study of PV integration impacts on distribution networks. IEEE Trans Sustain Energy 2017;8(2):648-657.

[15] Farhangi H. Smart grid. In: Elias SA, editor. Reference module in earth systems and environmental sciences, Elsevier Inc.; 2017.

[16] Faruqui A, Arritt K, Sergici S. The impact of advanced metering infrastructure on energy conservation: A case study of two utilities. The Electricity Journal 2017;30:56-63.

[17] Lo Cascio E, Borelli D, Devia F, Schenone C. Future distributed generation: An operational multi-objective optimization model for integrated small scale urban electrical, thermal and gas grids. Energ Convers Manage 2017;143:348-359.

[18] Moreno-Tejera S, Silva-Pérez MA, Lillo-Bravo I, Ramírez-Santigosa L. Solar resource assessment in Seville, Spain. Statistical characterization of solar radiation at different time resolutions. Sol Energy 2016;132:430-441.

[19] Wais P. A review of Weibull functions in wind sector. Renew Sust Energ Rev 2017;70:1099-1107.

[20] Feijóo A, Villanueva D. Assessing wind speed simulation methods. Renew Sust Energ Rev 2016;56:473-483.

[21] Sharifi SS, Rezaverdinejad V, Nourani V. Estimation of daily global solar radiation using wavelet regression, ANN, GEP and empirical models: A comparative study of selected temperature-based approaches. J Atmos Sol–Terr Phys 2016;149:131-145.

[22] Cao F, Li H, Yang T, Li Y, Zhu T, Zhao L. Evaluation of diffuse solar radiation models in Northern China: New model establishment and radiation sources comparison. Renew Energy 2017;103:708-720.

[23] Almonacid F, Pérez-Higueras P, Rodrigo P, Hntoria L. Generation of ambient temperature hourly time series for some Spanish locations by artificial neural networks. Renew Energy 2013;51:285-291.

[24] Poggi P, Muselli M, Notton G, Cristofari C, Louche A. Forecasting and simulating wind speed in Corssica by using an autoregressive model. Energ Convers Manage 2003;44:3177-3196.

[25] Graham VA, Hollands KGT. A method to generate synthetic hourly solar radiation globally. Sol Energy 1990;44:333-341.

[26] Erbs DG, Klein SA, Beckman WA. Estimation of degree-days and ambient temeperature bin data from monthly-average. ASHRAE Journal 1983;25:60-65.

[27] NASA surface meteorology and solar energy. National Aeronautics and Space Administration (NASA), https://power.larc.nasa.gov/data-access-viewer/; 2018 [accessed 17.07.2018].

[28] Short TA. Electric power distribution handbook. CRC Press LLC; 2004.

[29] Rosenblatt M. Remarks on a multivariate transformation. Ann Math Statist 1952;23(3):470-472.

[30] Bludszuweit H, Domínguez-Navarro JA, Llombart A. Statistical analysis of wind power forecast error. IEEE Trans Power Syst 2008;23(3):983-991.

[31] Giorsetto P, Utsurogi KF. Development of a new procedure for reliability modelling of wind turbine generators. IEEE Trans Power App Syst 1983;PAS-102(1):134-143.

[32] Lambert T, Gilman P, Lilienthal P. Micropower system modelling with HOMER. In: Farret FA, Simões MG, editors. Integration of alternative sources of energy. John Wiley & Sons, Inc.; 2006. P. 379-418.

[33] Lorenzo E, Araujo G, Cuevas A, Egido M, Minano J, R Zilles. Solar electricity: engineering of photovoltaic systems. Progensa; 1994.

[34] Kheldoun A, Bradai R, Boukenoui R, Mellit A. A new golden section method-based maximum power point tracking algorithm for photovoltaic systems. Energ Convers Manage 2016;111:125-136.

[35] Rampinelli GA, Krenzinger A, Romero FC. Mathematical models for efficiency of inverters used in grid connected photovoltaic systems. Renew Sust Energ Rev 2014;34:578-587.

[36] Liu E, Bebic J. Distribution system voltage performance analysis for high-penetration photovoltaics. National Renewable Energy Laboratory 2008; NREL/SR-581-42298.

[37] Manwell JF, McGowan JG, Rogers AL. Wind energy explained. Theory design and application. John Wiley & Sons, Inc.; 2002.

[38] Prasad AA, Taylor RA, Kay M. Assessment of solar and wind resource synergy in Australia. Appl Energy 2017;190:354-367.

[39] Al-falahi MDA, Jayasinghe SDG, Enshaei H. A review on recent size optimization methodologies for standalone solar and wind hybrid renewable energy system. Energ Convers Manage 2017;143:252-274.

[40] Teng J-H. A direct approach for distribution system load flow solutions. IEEE Trans Power Del 2003;18(3):882-887.

[41] Ciric RM, Feltrin AP, Ochoa LF. Power flow in four-wire distribution networks – General approach. IEEE Trans Power Syst 2003;18(4):1283-1290.

[42] Teng J-H. Modelling distributed generations in three-phase distribution load flow. IET Gener Transm Dis 2008;2(3):330-340.

[43] Kabir MN, Mishra Y, Bansal RC. Probabilistic load flow for distribution systems with uncertain PV generation. Appl Energy 2016;163:343-351.

[44] Dufo-López R, Bernal-Agustín JL. Design and control strategies of PV-diesel systems using genetic algorithms. Sol Energy 2005;79:33-46.