



HAL
open science

Self-adaptable hierarchical clustering analysis and differential evolution for optimal integration of renewable distributed generation

Rodrigo Mena, Martin Hennebel, Yan-Fu Li, Enrico Zio

► **To cite this version:**

Rodrigo Mena, Martin Hennebel, Yan-Fu Li, Enrico Zio. Self-adaptable hierarchical clustering analysis and differential evolution for optimal integration of renewable distributed generation. *Applied Energy*, 2014, 133, pp.388-402. 10.1016/j.apenergy.2014.07.086 . hal-01090342

HAL Id: hal-01090342

<https://centralesupelec.hal.science/hal-01090342v1>

Submitted on 3 Dec 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Self-adaptable Hierarchical Clustering Analysis and Differential Evolution for Optimal Integration of Renewable Distributed Generation

Rodrigo Mena^a

rodrigo.mena@ecp.fr
Tel. (33) 1 141131307

Martin Hennebel^c

martin.hennebel@supelec.fr

Yan-Fu Li^a

yanfu.li@ecp.fr
yanfu.li@supelec.fr

Enrico Zio^{ab*}

enrico.zio@ecp.fr
enrico.zio@supelec.fr
enrico.zio@polimi.it

^aChair on Systems Science and the Energetic Challenge, European Foundation for New Energy-Electricité de France, at École Centrale Paris – SUPELEC
Grande Voie des Vignes, F-92295 Châtenay-Malabry Cedex
France

^bPolitecnico di Milano
Energy Department
Via Ponzio 34/3, 20133 Milano
Milano
Italia

^cSUPELEC
Department of Power & Energy Systems
3, Rue Joliot Curie, 91190 Gif Sur Yvette
France

Abstract

In a previous paper, we have introduced a simulation and optimization framework for the integration of renewable generators into an electrical distribution network. The framework searches for the optimal size and location of the distributed renewable generation units (DG). Uncertainties in renewable resources availability, components failure and repair events, loads and grid power supply are incorporated. A Monte Carlo simulation – optimal power flow (MCS-OPF) computational model is used to generate scenarios of the uncertain variables and evaluate the network electric performance with respect to the expected value of the global cost (*ECG*). The framework is quite general and complete, but at the expenses of large computational times for the analysis of real systems. In this respect, the work of the present paper addresses the issue and introduces a purposely tailored, original technique for reducing the computational efforts of the analysis. The originality of the proposed approach lies in the development of a new search engine for performing the minimization of the *ECG*, which embeds hierarchical clustering analysis (HCA) within a differential evolution (DE) search scheme to identify groups of similar individuals in the DE population and, then, *ECG* is calculated for selected representative individuals of the groups only, thus reducing the number of objective function evaluations. For exemplification, the framework is applied to a distribution network derived from the IEEE 13 nodes test feeder. The results show that the newly proposed hierarchical clustering differential evolution (HCDE) MCS-OPF framework is effective in finding optimal DG-integrated network configurations with reduced computational efforts.

34 *Keywords:*

35 distributed renewable generation, uncertainty, simulation, optimization, differential evolution, hierarchical
 36 clustering analysis

37 ACRONYMS

DE	Differential evolution	MCS	Monte Carlo simulation
DG	Distribution generation	MS	Main supply
EA	Evolutionary algorithm	OPF	Optimal power flow
EV	Electric vehicle	PSO	Particle swarm optimization
GA	Genetic algorithm	PV	Photovoltaic
HCA	Hierarchical clustering analysis	ST	Storage device
HCDE	Hierarchical clustering analysis differential evolution	W	Wind turbine

38 NOMENCLATURE

$A_{i,i'}^{FD}$	the ampacity of the feeder (i,i') (A)	p_j^+	hourly probability distribution of EV discharging state per day
$B_{i,i'}$	susceptance of the feeder (i,i') ($1/\Omega$)	$Pa_{i,j}^{PS}$	available power in power source of type j allocated at node
BGT	available DG integration budget (\$)	$p_{d\%tile}$	linkage distances percentile
CCC	cophenetic correlation coefficient	$P_{i,j}^{MS}$	power supply of MS type j at node i (kW)
CCC_{th}	cophenetic correlation coefficient threshold	$P_{cap,j}^{MS}$	maximum capacity of the MS type j (kW)
CG	global cost (\$/h)	$P_{R_j}^{EV}$	rated power of EV technology type j (kW)
Ci	total fixed investment and operation cost (\$)	$P_{R_j}^{ST}$	rated power of ST technology type j (kW)
ci_j	investment cost of the DG technology type j (\$)	$P_{R_j}^W$	rated power of W technology type j (kW)
Co	operating costs of power generation and distribution (\$/h)	$Pu_{i,j}^{PS}$	used power from the power source type j at node i
Coc	crossover coefficient $\in [0,1]$	POP	population
Cop	opportunity cost for kWh not supplied (\$/kWh)	PS	set of all types of power sources
Cov_j^{PS}	variable operating cost of the power source j	PV	set of solar photovoltaic technologies
$Cov_{i,i'}^{FD}$	variable operating cost of the feeder (i,i')	ps	number of all types of available power generation technologies
D^{sp}	matrix of linkage distances between groups at step sp	$Q_{i,j}^{ST}$	level of charge in ST type j at node i (kJ)
\bar{D}^{sp}	average of D^{sp}	s_i	solar irradiance at node $i \in [0,1]$
$d_{p,q}^{sp}$	linkage distance between groups p and q	SE_j^{ST}	specific energy of the active chemical in ST type j (kJ/kg)
d_{CO}	cut off linkage distances	ST	set of storage devices technologies
DG	set of available types of distribution generation technologies	T_{a_i}	ambient temperature at node i ($^{\circ}C$)
dg	number of types of available distribution generation technologies	t_d	hour of the day (h)
d_{min}	minimum linkage distance	th	lifetime of the project (h)
$d_{NC=4}$	linkage distances to form at least four clusters	TL	total demand of power in the distribution network (kW)
ECG	expected global cost (\$/h)	TL_h	highest total demand of power in the distribution network (kW)

ECC_{min}	minimum expected global cost (\$/h)	$t_{op,i,j}^{EV}$	time of residence in the operating state $op_{i,j}^{EV}$ of EV type j at node i (h)
ep	energy price (\$/kWh)	$t_{R,i,j}^{ST}$	upper bound of the discharging time interval of ST type j at node i (h)
ep_h	energy price at highest total demand (\$/kWh)	V_{oc_j}	open circuit voltage (V)
EV	set of available types of EV	V_{MPP_j}	voltage at maximum power point (V)
F	differential variation amplification factor $\in [0,2]$	V^{NET}	voltage of the distribution network (kV)
FD	set of feeders	W	set of wind turbines technologies
G	generations count index	w_{a_j}	average wind speed of W type j (m/s)
G_{max}	maximum number of generations	$w_{S_{ci_j}}$	cut-in wind speed of W type j (m/s)
H	matrix of HCA resultant linkage distances	$w_{S_{co_j}}$	cut-out wind speed of W type j (m/s)
\bar{H}	average of H	w_{S_i}	wind speed at node i
$h_{p,q}$	HCA resultant linkage distance between groups p and q	$X_{i,i'}^{FD}$	reactance of feeder (i,i') (Ω/km)
I_{MPP_j}	current at maximum power point (A)		
I_{sc_j}	short circuit current (A)		
k_{T_j}	current0 temperature coefficient ($\text{mA}/^\circ\text{C}$)	α_i^{PV}	shape parameter of the Beta probability density function of the solar irradiance at node i
k_{V_j}	voltage temperature coefficient ($\text{mV}/^\circ\text{C}$)	β_i^{PV}	shape parameter of the Beta probability density function of the solar irradiance at node i
L_i	power demand at node i (kW)	δ_i	voltage angle at node i
$l_{i,i'}$	length of feeder (i,i') (km)	ϑ	operating scenario
LS_i	load shedding at node i (kW)	λ_j^F	failure rate of power source type j (1/h)
$mc_{i,j}$	mechanical state of PS type j at node i	$\lambda_{i,i'}^F$	failure rate of feeder (i,i') (1/h)
$mc_{i,i'}$	mechanical stated of feeder (i,i')	λ_j^R	repair rate of power source type j (1/h)
MS	set of types of MS spots	$\lambda_{i,i'}^R$	repair rate of feeder (i,i') (1/h)
ms	number of types of MS spots	μ_i^L	mean of the normal distribution of the power load at node i (kW)
$M_{T_j}^{ST}$	mass of active chemical in the battery type j at node i (kg)	μ_j^{MS}	normal distribution mean of the MS type j at node i (kW)
N	set of nodes in the distribution network	Ξ	configuration matrix of DG-integrated network
NS	number of operating scenarios ϑ	Ξ^{DG}	DG part of configuration matrix of DG-integrated network
n	number of nodes in the distribution network	Ξ^{MS}	MS part of configuration matrix of DG-integrated network
NFE	number of objective function evaluations	$\xi_{i,j}$	number of units of MS spots or DG technology j that are allocated at a node i
$N_{\sigma T_j}$	nominal cell operation temperature ($^\circ\text{C}$)	σ_i^L	standard deviation of the normal distribution of the power load at node i
NP	population size	σ_j^{MS}	normal distribution standard deviation of the MS type j at node i (kW)
$op_{i,j}^{EV}$	operating state of EV type j at node i	σ_i^W	scale parameter of the Rayleigh distribution function of the wind speed at node i
P_j^-	hourly probability distribution of EV charging state per day	τ_j	maximum number of units of DG technology type j available for integration
P_j^o	hourly probability distribution of EV disconnected state per day	Υ	set of operating scenarios ϑ

GREEK SYMBOLS

39 1 INTRODUCTION

40 Renewable distribution generation (DG) requires the selection of the different available technologies, and their
41 sizing and allocation onto the power distribution network, considering the specific economic, operational and
42 technical constraints [1-5]. This can become a complex optimization problem, depending on the size of the
43 distribution network and the number of renewable DG technologies available, that can lead to combinatorial
44 explosion [1, 3, 6-9]. Furthermore, for each renewable DG plan considered, the power flow problem needs to be
45 solved to assess the response of the distribution network in terms of power and voltage profiles, available power
46 usage, power demand satisfaction, economic performances, etc., with possibly significant computation times.

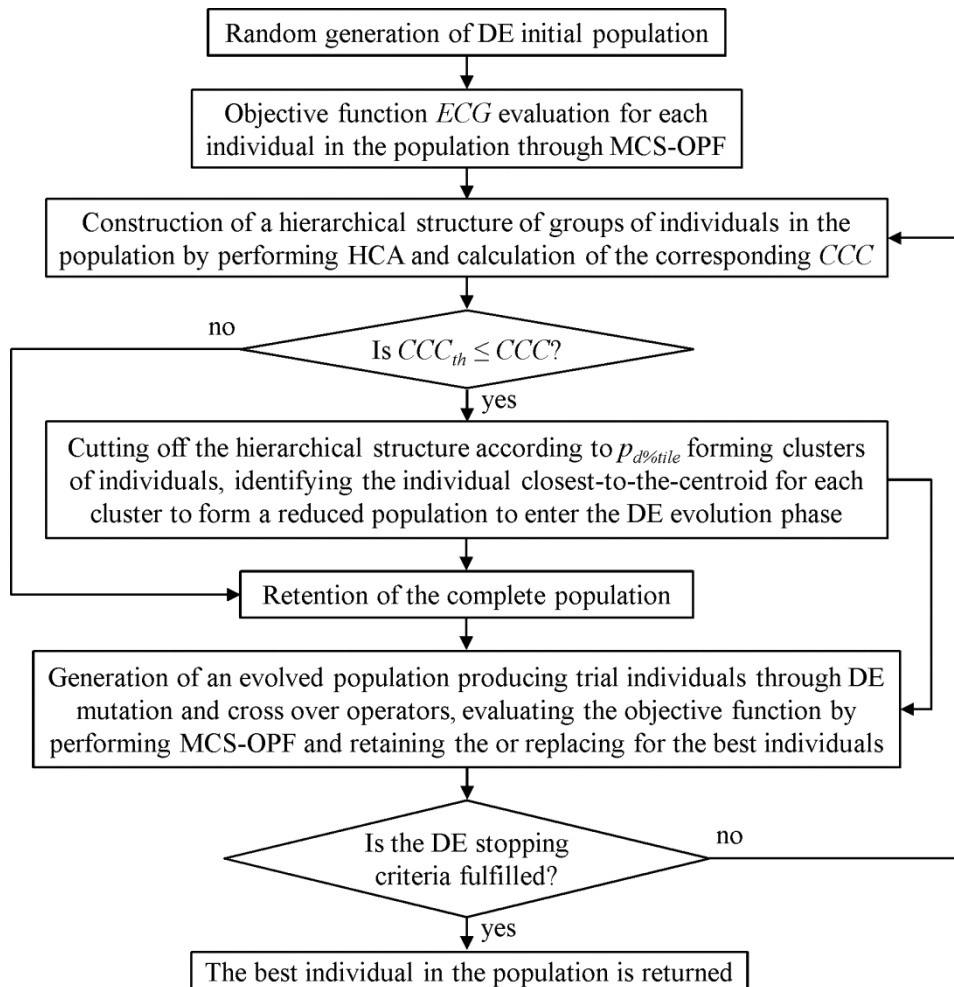
47 Heuristic optimization techniques belonging to the class of Evolutionary Algorithms (EAs), like honey bee mating
48 [10], particle swarm optimization (PSO) [9, 11-13], differential evolution (DE) [14, 15] and genetic algorithms
49 (GA) [2, 3, 16, 17], have been considered for the solution to this problem, since they can deal straightforwardly
50 with non-convex combinatorial problems, discontinuous search spaces and non-differentiable objective functions
51 [1, 9].

52 To improve the performance of EAs for the complex optimization problem of DG planning, we consider the
53 integration of clustering [18-23]. This can be directed to the enhancement of the global and/or local searching
54 ability of the algorithm, and amounts to identifying groups of similar individuals and applying different evolution
55 operators to those of a same cluster (group) [18, 20-22], e.g. for random generation of new individuals in the
56 neighborhood of cluster centroids [23], or multi-parents crossover over new randomly generated individuals spread
57 in the global feasible space [19]. Even if convergence is improved, some of these methodologies increase
58 temporarily the overall size of the population and, thus, the computational effort. In addition, the accuracy of the
59 clusters structures in representing the distribution of individuals must be controlled for performing clustering
60 conveniently.

61 The main original contribution of the work here presented, lies in the development of the clustering strategy in a
62 controlled manner. The implementation of such clustering strategy is done within a Monte Carlo simulation and
63 optimal power flow (MCS-OPF) model and differential evolution (DE) optimization framework [24] previously
64 developed by the authors for the integration of renewable generators into an electrical distribution network: the
65 framework searches for the optimal size and location of the distributed renewable generation units (DG) [25].
66 Optimality of the DG plan is sought with respect to the expected global cost (*ECG*). The introduction of the
67 clustering is hierarchically (i.e., hierarchical clustering analysis, HCA, [26]) by a controlled way of reducing the
68 number of individuals to be evaluated during the DE search, therefore, improving the computational efficiency.
69 Henceforth, we call our method hierarchical clustering differential evolution (HCDE).

70 HCA is introduced to build a hierarchical structure of grouping individuals of the population that present closeness
71 under the control of a specific linkage criterion based on defined distance metrics [26]. The HCA outcomes are the
72 linkage distances at which the grouping actions take place, defining the different levels in the hierarchical structure.

73 Two control parameters are introduced in the HCA, the cophenetic correlation coefficient (CCC) and a percentile
 74 of the set of linkage distances in the hierarchical structure of the groups ($p_{d\%tile}$). The CCC is a similarity coefficient
 75 that measures how representative is the proposed grouping structure by comparing their linkage distances with the
 76 original distances between all the individuals in the population. In the hierarchical structure, the linkage distance
 77 given by $p_{d\%tile}$ sets the level at which the groups formed below it are considered to be ‘close enough’ to constitute
 78 independent clusters. The two parameters allow HCDE to adapt itself in each generation of the search, ‘deciding’
 79 whether to perform clustering if the CCC is greater than or equal to a preset threshold (CCC_{th}) and cutting the
 80 hierarchical structure in independent clusters according to the linkage distance given by $p_{d\%tile}$. Then, the individual
 81 closest to the centroid of each cluster is taken as the feasible representative solution in the population that enters the
 82 evolution phase of the HCDE algorithm. Figure 1 summarizes schematically the structure of the proposed
 83 framework.



84

85

Figure 1. HCDE framework schema

86 We test the approach on a case study based on the IEEE 13 nodes test feeder distribution network [27], completing
 87 the study with a sensitivity analysis to investigate the effects of the parameters controlling the clustering, namely
 88 CCC and $p_{d\%tile}$.

89 For practical ease of the presentation of the approach, in the next section we provide the basic elements of the
 90 model of the distribution network considered as case study and we briefly summarize the MCS-OPF model taken
 91 from [25]. In Section 3, we embed this in the HCDE for renewable DG selection, sizing and allocation. Finally, in
 92 Section 4 we present the numerical results of the case study and in Section 5 we draw some conclusions on the
 93 work performed.

94 2 RENEWABLE DG-INTEGRATED NETWORK MODEL

95 The operation of the renewable DG-integrated network is considered to be dictated by the location and magnitude
 96 of the power available in the different sources, the loads and the operating states of the components. Uncertainty is
 97 present in the states of operation of the components, due to stochasticity of degradation and failures, and in the
 98 behavior of the renewable energy sources. These uncertainties have a direct impact on the power available (from
 99 the DG units, main supply spots and/or feeders) to satisfy power demands, which are, in turn, also subject to
 100 fluctuations. Furthermore, if the distribution network is considered as a ‘price taker’ entity, the uncertain behavior
 101 of the power demand impacts directly over the energy price [4, 5, 28]. Consequently, an attentive modeling of the
 102 uncertainties in renewable DG planning is imperative for well-supported decision-making.

103 Monte Carlo simulation (MCS) has already been used to emulate the stochastic operating conditions and evaluate
 104 the performance of power distribution networks [19, 28, 29, 32]. In the present paper, non-sequential MCS is used
 105 to randomly sample the modeled uncertain variables for a specific renewable DG plan, without dependence on
 106 previous operating conditions, characterizing the network operation in terms of location and magnitudes of power
 107 available and loads. Then, the performance of the DG-integrated network is evaluated through the optimal power
 108 flow model.

109 2.1 Monte Carlo and Optimal Power Flow Simulation

110 In the proposed framework, the renewable DG technologies considered are of four types: solar photovoltaic (PV),
 111 wind turbines (W), electric vehicles (EV) and storage devices (ST); these are represented by the set DG that
 112 contains all the dg types of technologies. As for main power supply spots or transformers (MS), the set MS
 113 indicates the ms different types of MS considered in the network.

114 The DG-integrated network deployment is represented by the location and capacity size of the power sources, as
 115 indicated in matrix form in equation (1) below, where $\xi_{i,j}$ indicates the number of units of main supply spots or DG
 116 technology j that are allocated at a node i :

$$\Xi = \begin{bmatrix} \xi_{1,1} & \cdots & \xi_{1,j} & \cdots & \xi_{1,ms} & \vdots & \xi_{1,ms+1} & \cdots & \xi_{1,ms+j} & \cdots & \xi_{1,ms+dg} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \xi_{i,1} & \cdots & \xi_{i,j} & \cdots & \xi_{i,ms} & \vdots & \xi_{i,ms+1} & \cdots & \xi_{i,ms+j} & \cdots & \xi_{i,ms+dg} \\ \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \xi_{n,1} & \cdots & \xi_{n,j} & \cdots & \xi_{n,ms} & \vdots & \xi_{n,ms+1} & \cdots & \xi_{n,ms+j} & \cdots & \xi_{n,ms+dg} \end{bmatrix} = [\Xi^{MS} \mid \Xi^{DG}] \quad \forall \xi_{i,j} \in \mathbb{Z}^*, i \in N, j \in PS \quad (1)$$

117 where, N and $PS = \{MS \cup DG\}$ are the set of nodes in the network and the set of all power sources, whose
 118 cardinalities are n and $ps = ms + dg$, respectively.

119 The set of feeders FD is defined by all the pairs of nodes (i, i') connected by a distribution line $\forall (i, i') \in N \times N$.

120 The considered uncertain conditions that determine the operation of the DG-integrated network are accounted for
 121 using different stochastic models, as summarized in Table 1. The interested reader can consult [25] for further
 122 details.

123 Table 1. Uncertain conditions models in the DG-integrated network operation

Variable	Nomenclature	States and Units	Model	Parameters
Hour of the day	t_d	(h)	Discrete uniform distribution	$[1, 24]$
Mechanical state	$mc_{i,j}$ $mc_{i,i'}$	(0): <i>under repair</i> (1): <i>operating</i>	Two-state Markov	λ_j^F, λ_j^R $\lambda_{i,i'}^F, \lambda_{i,i'}^R$
Main power supply	$P_{i,j}^{MS}$	(kW)	Truncated normal distribution $0 \leq P_{i,j}^{MS} \leq P_{cap,j}^{MS}$	$\mu_j^{MS}, \sigma_j^{MS}$ $P_{cap,j}^{MS}$
Solar irradiance	s_i	$[0, I]$	Beta distribution	$\alpha_i^{PV}, \beta_i^{PV}$
Wind speed	ws_i	(m/s)	Rayleigh distribution	σ_i^W
EV operating state	$op_{i,j}^{EV}$	(-1): <i>charging</i> (0): <i>disconnected</i> (1): <i>discharging</i>	'Block groups' Hourly probability distribution of EV operating states per day	t_d
ST level of charge	$Q_{i,j}^{ST}$	(kJ)	Uniform distribution	$[0, SE_j^{ST} \times M_{T_{i,j}}^{ST}]$
Nodal power demand	L_i	(kW)	Daily nodal load profiles, hourly normally distributed load. Truncated normal distribution $0 \leq L_i \leq \infty$	$\mu_i^L(t_d), \sigma_i^L(t_d)$

124 where $\forall i, i' \in N, j \in PS, (i, i') \in FD$, λ_j^F and λ_j^R (1/h) are the failure and repair rates of the power source j ,
 125 respectively, $\lambda_{i,i'}^F$ and $\lambda_{i,i'}^R$ (1/h) are the failure and repair rates of the feeder (i, i') , respectively, μ_j^{MS} and σ_j^{MS} are
 126 the normal distribution mean and standard deviation associated to the main supply j at node i , $P_{cap,j}^{MS}$ is the
 127 maximum capacity of the transformer j (kW), α_i^{PV} and β_i^{PV} are the parameters of the Beta probability density
 128 function of the solar irradiance at node i , σ_i^W is the scale parameter of the Rayleigh distribution function of the wind
 129 speed at node i , SE_j^{ST} (kJ/kg) is the specific energy of the active chemical in the battery type j , $M_{T_{i,j}}^{ST}$ (kg) is the
 130 mass of active chemical in the battery type j at node i , $\mu_i^L(t_d)$ and $\sigma_i^L(t_d)$ are the hourly mean and standard
 131 deviation of the normal distribution of the power load at node i .

132 Concerning the hour of the day t_d (h), sampled from a discrete uniform distribution $U(1,24)$, the night interval is
 133 defined between 22.00 and 06.00 hours. If the value of t_d falls in the night interval, there is no solar irradiation.

134 The resulting realization of one operational scenario of duration ts (h), for the given DG plan denoted by $\{FD, \Xi\}$,
 135 consists in the random sampling of each uncertain variable (Table 1), here indicated by the vector ϑ below:

$$\mathfrak{S} = [t_d, mc_{i,j}, mc_{i,j'}, L_i, P_{i,j}^{MS}, s_i, ws_i, op_{i,j}^{EV}, Q_{i,j}^{ST}] \quad (2)$$

136 To evaluate the performance of the distribution network the OPF model receives as input the location and
 137 magnitude of the available power in the power sources and demanded at the loads, which are set by the operating
 138 conditions defined by $\{FD, \Xi\}$ and \mathfrak{S} . The nodal power loads L_i are directly sampled, whereas the available power
 139 in the power sources (MS and DG) depends on the uncertain variables that represent the behavior of the energy
 140 sources, the specific technical characteristics of each type of technology and the mechanical states. The available
 141 power in each type of power source considered is modeled by the functions summarized in Table 2, for a given
 142 configuration $\{FD, \Xi\}$, operating scenario \mathfrak{S} and a generic node i .

143 Table 2. Available power functions of the power sources (PS) [25, 29, 30]

PS type j	Parameters	Available power function (kW)
MS	-	$Pa_{i,j}^{MS;\mathfrak{S}} = \xi_{i,j} mc_{i,j}^{\mathfrak{S}} P_{i,j}^{MS;\mathfrak{S}}$ (3)
PV	T_{a_i}	$Pa_{i,j}^{PV;\mathfrak{S}} = \xi_{i,j} mc_{i,j}^{\mathfrak{S}} FF_j V_{i,j}^{\mathfrak{S}} I_{i,j}^{\mathfrak{S}} / 1,000$ (4)
	N_{oT_j}	$T_{c_i,j}^{\mathfrak{S}} = T_{a_i} + s_{i,j}^{\mathfrak{S}} (N_{oT_j} - 20) / 0.8$
	I_{sc_j}	$I_{i,j}^{\mathfrak{S}} = s_{i,j}^{\mathfrak{S}} (I_{sc_j} + k_{I_j} (T_{c_i,j}^{\mathfrak{S}} - 25))$
	V_{oc_j}	$V_{i,j}^{\mathfrak{S}} = V_{oc_j} + k_{V_j} T_{c_i,j}^{\mathfrak{S}}$
	k_{V_j}, k_{I_j}	$FF_j = (V_{MPP_j} I_{MPP_j}) / (V_{oc_j} I_{sc_j})$
W	ws_{ci_j}	$Pa_{i,j}^{W;\mathfrak{S}}(ws_i^{\mathfrak{S}}) = \xi_{i,j} mc_{i,j}^{\mathfrak{S}} \times \begin{cases} P_{R_j}^W \frac{ws_i^{\mathfrak{S}} - ws_{ci_j}}{ws_{a_j} - ws_{ci_j}} & \text{if } ws_{ci_j} \leq ws_i^{\mathfrak{S}} < ws_{a_j} \\ P_{R_j}^W & \text{if } ws_{a_j} \leq ws_i^{\mathfrak{S}} < ws_{co_j} \\ 0 & \text{otherwise} \end{cases}$ (5)
	ws_{a_j}	
	ws_{co_j}	
	$P_{R_j}^W$	
EV	$t_{op_{i,j}^{EV;\mathfrak{S}}}, P_{R_j}^{EV}$	$Pa_{i,j}^{EV;\mathfrak{S}}(op_{i,j}^{EV;\mathfrak{S}}, t) = \xi_{i,j} mc_{i,j}^{\mathfrak{S}} op_{i,j}^{EV;\mathfrak{S}} P_{R_j}^{EV} \quad \forall t \in [0, t_{op_{i,j}^{EV;\mathfrak{S}}}]$ (6)
ST	$P_{R_j}^{ST}$	$Pa_{i,j}^{ST;\mathfrak{S}}(t) = \xi_{i,j} mc_{i,j}^{\mathfrak{S}} P_{R_j}^{ST} \quad \forall t \in [0, t_R]$ (7)
		$t_{R,j}^{ST;\mathfrak{S}}(Q_{i,j}^{ST;\mathfrak{S}}) = Q_{i,j}^{ST;\mathfrak{S}} / P_{R_j}^{ST}$

144 In Table 2, $Pa_{i,j}^{PS;\mathfrak{S}}$ (kW), $\xi_{i,j}$ and $mc_{i,j}^{\mathfrak{S}}$ denote the available power, the units and the mechanical state of the power
 145 source of type j allocated at node i . For solar photovoltaic technologies $j \in PV$, the parameter T_{a_i} ($^{\circ}C$) is the
 146 ambient temperature at node i , N_{oT_j} ($^{\circ}C$) is the nominal cell operation temperature, I_{sc_j} (A) is the short circuit
 147 current, V_{oc_j} (V) is the open circuit voltage, k_{V_j} (mV/ $^{\circ}C$) is the voltage temperature coefficient, k_{I_j} (mA/ $^{\circ}C$) is the
 148 current temperature coefficients and V_{MPP_j} (V) and I_{MPP_j} (A) are the voltage and current at maximum power point,
 149 respectively. For wind turbines of types $j \in W$, ws_{ci_j} , ws_{a_j} and ws_{co_j} (m/s) are the cut-in, rated and cut-out wind
 150 speeds, respectively, and $P_{R_j}^W$ (kW) is the rated power of the turbine. For electric vehicles $j \in EV$, $t_{op_{i,j}^{EV;\mathfrak{S}}}$ (h) is the

151 time of residence in the operating state $op_{i,j}^{EV:\vartheta}$ and $P_{R_j}^{EV}$ (kW) is the rated power. For storage devices $j \in ST$, $t_{R_i,j}^{\vartheta}$ (h)
 152 is the upper bound of the discharging time interval and $P_{R_j}^{ST}$ (kW) is the rated power.

153 Under the operating conditions set forth, the given configuration of the renewable DG-integrated network $\{FD, \Xi\}$
 154 and the scenario ϑ , the OPF objective is the minimization of the operating cost associated to the generation and
 155 distribution of power, considering the revenues per kWh sold. Power flow analysis is performed by DC modeling,
 156 neglecting power losses and assuming the voltage throughout the network as constant, linearizing the classic non-
 157 linear power flow formulation by accounting solely for active power flows [31, 32]. The present formulation of the
 158 DC optimal power flow problem is:

$$\min Co^{\vartheta}(Pu, \Delta\delta) = \sum_{i \in N} \sum_{j \in PS} (Cov_j^{PS} - ep^{\vartheta}) Pu_{i,j} + \sum_{(i,i') \in FD} Cov_{i,i'}^{FD} |B_{i,i'}(\delta_i - \delta_{i'})| + (Cop + ep^{\vartheta}) \sum_{i \in N} LS_i \quad (8)$$

159 s.t.

$$L_i^{\vartheta} - LS_i - \sum_{j \in PS} Pu_{i,j} - \sum_{i' \in N} mc_{i,i'}^{\vartheta} B_{i,i'}(\delta_i - \delta_{i'}) = 0 \quad (9)$$

$$0 \leq Pu_{i,j} \leq Pa_{i,j}^{PS:\vartheta} \quad (10)$$

$$|B_{i,i'}(\delta_i - \delta_{i'})| \leq V^{NET} A_{i,i'}^{FD} \quad (11)$$

160 where $\forall i, i' \in N, j \in PS, (i, i') \in FD$ and the operating scenario ϑ , Co^{ϑ} (\$/h) is the operating cost of the total power
 161 supply and distribution, Cov_j^{PS} (\$/kWh) is the variable operating cost of the power source j , ep^{ϑ} (\$/kWh) is the
 162 energy price, $Pu_{i,j}$ (kW) is the used power from the source of type j at node i , $Cov_{i,i'}^{FD}$ (\$/kWh) and $B_{i,i'}$ ($1/\Omega$) are the
 163 variable operating cost and the susceptance of the feeder (i, i') , respectively, δ_i is the voltage angle at node i , Cop
 164 (\$/kWh) is the opportunity cost for kWh not supplied, V^{NET} (kV) is the nominal voltage of the network and $A_{i,i'}$ (A)
 165 is the ampacity of the feeder (i, i') . The load shedding LS_i (kW) is defined as the amount of load disconnected at
 166 node i to alleviate congestions in the feeders and/or balance the demand of power with the available power supply.

167 The distribution network is considered as a ‘price taker’ entity, assuming a correlation between the total demand of
 168 power and the energy price ep (\$/kWh). Then, the energy price is calculated from an intermediate correlation
 169 proposed by [4, 5, 28]:

$$ep(TL) = ep_h \left(-0.38 \left(\frac{TL(t_d)}{TL_h} \right)^2 + 1.38 \frac{TL(t_d)}{TL_h} \right) \quad (12)$$

170 where, ep_h is the energy price corresponding to the highest value of total demand considered TL_h . The total demand
 171 of power $TL(t_d)$ at the hour of the day t_d is the summation of all the nodal loads $L_i(t_d)$ (Table 1).

172 The constraint given by the equation (9) corresponds to the power balance equation at node i , whereas equations
 173 (10) and (11) represent the bounds of the power generation and technical limits of the feeders, respectively.

174 One realization of the MCS-OPF consists of the sampling of NS operating scenarios \mathfrak{Y} regarded as the set
 175 $\Upsilon = \{\mathfrak{Y}_1, \dots, \mathfrak{Y}_h, \dots, \mathfrak{Y}_{NS}\}$ for each of which the optimal power flow problem is solved, giving in output the values of
 176 the minimum operating cost of the total power supply and distribution $Co^\Upsilon = \{Co^{\mathfrak{Y}_1}, \dots, Co^{\mathfrak{Y}_h}, \dots, Co^{\mathfrak{Y}_{NS}}\}$.

177 2.2 Expected Global Cost ECG

178 The proposed renewable DG-integrated network solutions are evaluated with respect to the expected global cost
 179 ECG . The global cost CG is composed by two terms: the fixed investment and operation (maintenance) costs Ci (\$),
 180 which are prorated hourly over the life of the project th (h), and the operating costs Co^Υ (\$/h) that is the outcome of
 181 the MCS-OPF (equation (8)) described in the precedent Section 2.1. Thus, the global cost function for a scenario \mathfrak{Y}
 182 is given by:

$$CG^\mathfrak{Y} = Ci + Co^\mathfrak{Y} \quad \forall \mathfrak{Y} \in \Upsilon \quad (13)$$

$$Ci = \frac{1}{th} \sum_{i \in N} \sum_{j \in DG} \xi_{i,j} ci_j \quad (14)$$

183 where, ci_j (\$) is the investment cost of the DG technology type j .

184 Then, the global cost $CG^\Upsilon = \{CG^{\mathfrak{Y}_1}, \dots, CG^{\mathfrak{Y}_h}, \dots, CG^{\mathfrak{Y}_{NS}}\}$ is considered as realizations of the probability mass
 185 function of CG , and from multiple realizations the expected value ECG^Υ can be obtained.

186 3 RENEWABLE DG SELECTION, SIZING AND ALLOCATION

187 The aim of the proposed simulation and optimization framework is to find the optimal plan of integration of
 188 renewable DG in terms of selection, sizing and allocation of generation units from different technologies available
 189 (PV, W, EV and ST). The corresponding decision variables are contained in Ξ^{DG} of the configuration matrix Ξ
 190 defined in equation (1).

191 3.1 Optimization Problem Formulation

192 Considering a network configuration (FD, Ξ) and a set of randomly generated scenarios Υ , the optimization
 193 problem is formulated as follows:

$$\min ECG^\Upsilon \quad (15)$$

194 s.t.

$$\forall \xi_{i,j} \in \mathbb{Z}^* \quad (1)$$

$$\sum_{i \in N} \sum_{j \in DG} \xi_{i,j} ci_{i,j} \leq BGT \quad (16)$$

$$\sum_{i \in N} \xi_{i,j} \leq \tau_j \quad (17)$$

$$\text{MCS-OPF}((FD, \Xi), \Upsilon) \quad (18)$$

195 The meaning of each constraint $\forall i, i' \in N, j \in PS, (i, i') \in FD, \tau_j \in \mathbb{Z}^+$ is:

- 196 ▪ (1): the decision variable $\xi_{i,j}$ is a non-negative integer number.
- 197 ▪ (16): the total investment and fixed operation and maintenance costs must be less than or equal to the available
- 198 budget BGT .
- 199 ▪ (17): the total number of renewable DG units of each technology j to be allocated must be less than or equal to
- 200 the maximum number of units available for integration τ_j .
- 201 ▪ (18): all the equations (8)-(11) of MCS-OPF must be satisfied.

202 3.2 Hierarchical Clustering Differential Evolution (HCDE)

203 The complex combinatorial optimization problem of DG planning under uncertainties described above is solved by
 204 integrating DE with HCA to reduce computational efforts, whereby the evaluation of the objective function is
 205 performed by the MCS-OPF presented in Section 3.

206 DE is a population-based and parallel, direct search method, shown to be one of the most efficient evolutionary
 207 algorithms to solve complex optimization problems [19, 21, 24]. The implementation of the original version of DE
 208 involves two main phases: initialization and evolution, summarized below for completeness of the paper [24]:

209 Initialization

- 210 ▪ Set values of parameters:
 - 211 • NP : population size
 - 212 • G_{max} : maximum number of generations
 - 213 • Coc : cross over coefficient $\in [0,1]$
 - 214 • F : differential variation amplification factor $\in [0,2]$
- 215 ▪ Generate randomly NP individuals X (decision vectors) within the feasible space, to form the initial
 216 population $POP^0 = \{X_1^0, \dots, X_k^0, \dots, X_{NP}^0\}$.
- 217 ▪ Evaluate the objective function $f(X) = y$ for each individual

218 Evolution loop

- 219 ▪ Set generations count index $G = 1$
- 220 ▪ Set $POP^G = POP^0$
- 221 ▪ While $G \leq G_{max}$ (stopping criterion)

222 Trial loop

223 For each individual X_k^G in POP^G , $\forall k \in \{1, \dots, NP\}$:

- 224 • Sample from the uniform distribution three integer indexes r_1, r_2, r_3 with $k \neq r_1 \neq r_2 \neq r_3$ and choose the
 225 corresponding three individuals $X_{r_1}^G, X_{r_2}^G, X_{r_3}^G$
- 226 • Mutation: generate a mutant individual V_k^G according to:

$$V_k^G = X_{r_1}^G + F(X_{r_2}^G - X_{r_3}^G) \quad (19)$$

- 227 • Crossover: initialize a randomly generated vector U_k^G , whose dimensionality dim is the same as that of
 228 X_k^G and each coordinate $u_{k,i}^G$ follows a uniform distribution with outcome in $[0,1] \forall i \in \{1, \dots, dim\}$. In
 229 addition, generate randomly an integer index $ri \in \{1, \dots, dim\}$ from a uniform distribution to ensure that at

230 least one coordinate from V_k^G is exchanged to form a trial individual XT_k^G , whose coordinates are
 231 defined as follows:

$$x_{k,i}^G = \begin{cases} v_{k,i}^G & \text{if } u_{k,i}^G \leq Cco \text{ or } i = ri \\ x_{k,i}^G & \text{if } u_{k,i}^G > Cco \text{ and } i \neq ri \end{cases} \quad (20)$$

- 232 • Selection: evaluate the objective function for the trial individual $f(XT_k^G)$; if $f(XT_k^G) < f(X_k^G)$
 233 (minimization), then XT_k^G replaces X_k^G in the population POP^G , otherwise X_k^G is retained
 234 ▪ Set $G = G + 1$
 235 ▪ Once the stopping criterion is reached, sort the individuals in $POP^{G_{max}}$ in descending order according to their
 236 values of the objective function and return $X_I^{G_{max}}$.

237 The original version of DE keeps the population size NP constant, making the computational performance
 238 dependent mainly on the number of objective function evaluations carried out during the evolution phase of the
 239 algorithm. Then, the integration of HCA into DE is aimed at the reduction of the number of individuals that enter
 240 the evolution loop in each generation so as to decrease the number of objective function evaluations.

241 HCA links individuals or groups of individuals which are similar with respect to a specific property, translated into
 242 a metric of distance, obtaining a hierarchical structure. In practice, we use an agglomerative procedure which in sp
 243 $= NP - 1$ steps fuses the closest pair or individuals or groups of individuals through a linkage function, e.g. single
 244 linkage (nearest neighbor distance), complete linkage (furthest neighbor), average linkage, among others, until the
 245 complete hierarchical structure is built. The base hierarchical clustering algorithm used in this study can be
 246 expressed as follows [26]:

247 **Step 0:** Given a population $POP = \{X_1, \dots, X_k, \dots, X_{NP}\}$, form the set of singleton groups
 248 $O = \{O_p = \{X_k\}\}$, $\forall p = k \in \{1, \dots, NP\}$ and calculate the linkage distances between all the NP groups using
 249 the average as linkage function and the Euclidean distance as metric:

$$D^1 = \begin{bmatrix} d_{1,2}^1 & \dots & d_{1,q}^1 & \dots & d_{1,NP}^1 \\ & \ddots & \vdots & \ddots & \vdots \\ & & d_{p,q}^1 & \dots & d_{p,NP}^1 \\ & & & \ddots & \vdots \\ & & & & d_{NP-1,NP}^1 \end{bmatrix} \quad \text{with } d_{p,q}^1 = \frac{\sum_{X_{kp} \in O_p} \sum_{X_{kq} \in O_q} \sqrt{(X_{kp} - X_{kq})^2}}{|O_p| |O_q|} \quad (21)$$

$\forall p, q \in \{1, \dots, NP\}, kp, kq \in \{1, \dots, NP\}$

250 where, $d_{p,q}^1$ is the average of the Euclidean distances between all the individuals X_k belonging to the groups
 251 O_p and O_q , respectively.

252 **Step 1:** Fuse the first pair of groups $O_{p'}$ and $O_{q'}$, for which $d_{p',q'}^1$ is the minimum distance $\min(D^1)$ and form a new
 253 group $O_{NP+1} = \{O_{p'} \cup O_{q'}\}$.

254 Update the set of groups O replacing $O_{p'}$ and $O_{q'}$ by O_{NP+1} , and calculate the linkage distances D^2 between
 255 all the $NP - 1$ groups in O using (21).

256 **Step 2:** Fuse the second pair of groups $O_{p'}$ and $O_{q'}$ for which $d_{p',q'}^{2}$ is the minimum distance $\min(D^2)$, and form a
 257 new group $O_{NP+2} = \{O_{p'} \cup O_{q'}\}$.
 258 As in the preceding step, update the set of groups O and calculate the linkage distances D^3 between all the
 259 $NP-2$ groups in O using (21).

⋮

260 **Step NP-1:** Fuse the last pair of groups with linkage distance $d_{p',q'}^{NP-1}$, forming the last group $O_{2NP-1} = \{O_{p'} \cup O_{q'}\}$
 261 that contains all the individuals X .

262 The outcoming hierarchical (or tree) structure can be reported as a sorted table containing the $NP-1$ linkage
 263 distances relative to each pairing action of individuals/groups and be graphically illustrated as a dendrogram. Table
 264 3 and Figure 2 present, respectively, the resultant linkage distances and dendrogram obtained from an example set
 265 of $NP = 8$ two-dimensional individuals X using the above introduced HCA algorithm.

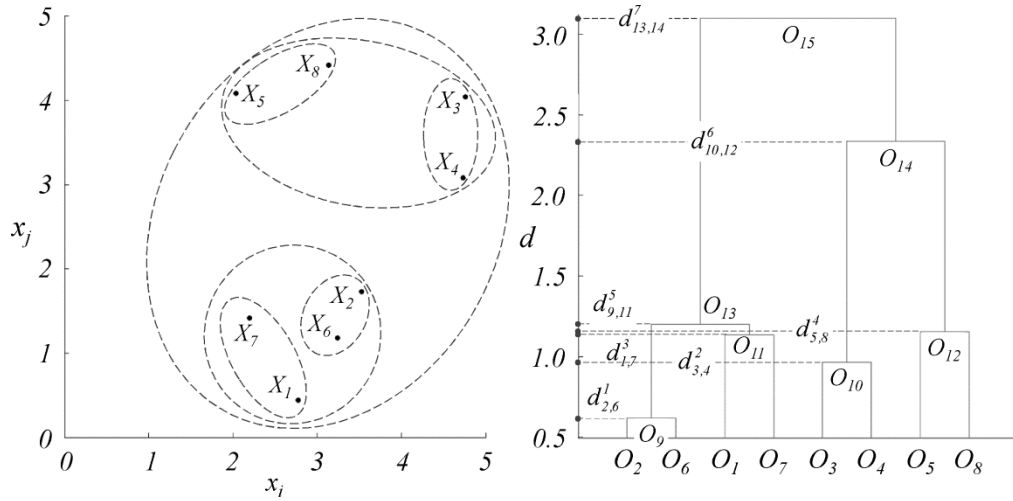


Figure 2. Example dendrogram for average linkage HCA

Table 3. Example hierarchical structure outcome

Step sp	Group	Groups linked	Linkage distance $d_{p',q'}^{sp}$
1	O_9	$\{O_2 \cup O_6\} = \{\{X_2\} \cup \{X_6\}\}$	$d_{2,6}^1$
2	O_{10}	$\{O_3 \cup O_4\} = \{\{X_3\} \cup \{X_4\}\}$	$d_{3,4}^2$
3	O_{11}	$\{O_1 \cup O_7\} = \{\{X_1\} \cup \{X_7\}\}$	$d_{1,7}^3$
4	O_{12}	$\{O_5 \cup O_8\} = \{\{X_5\} \cup \{X_8\}\}$	$d_{5,8}^4$
5	O_{13}	$\{O_9 \cup O_{11}\} = \{\{X_2, X_6\} \cup \{X_1, X_7\}\}$	$d_{9,11}^5$
6	O_{14}	$\{O_{10} \cup O_{12}\} = \{\{X_3, X_4\} \cup \{X_5, X_8\}\}$	$d_{10,12}^6$
7	O_{15}	$\{O_{13} \cup O_{14}\} = \{\{X_1, X_2, X_6, X_7\} \cup \{X_3, X_4, X_5, X_8\}\}$	$d_{13,14}^7$

269 As stated above, HCA builds the hierarchical structure through a linkage function introducing in each grouping
 270 action a larger or smaller degree of distortion with respect to the original distances between (ungrouped)
 271 individuals. The measurement of this distortion is important and the cophenetic correlation coefficient (CCC) is

272 introduced to evaluate how representative is the hierarchical structure proposed by the HCA. The CCC can be
 273 obtained from equations (22) and (23) below [26].

$$CCC = \frac{\sum_{p < q} (d_{p,q}^I - \bar{D}^I)(h_{p,q} - \bar{H})}{\sqrt{\sum_{p < q} (d_{p,q}^I - \bar{D}^I)^2 \sum_{p < q} (h_{p,q} - \bar{H})^2}} \quad \forall p, q \in \{1, \dots, NP\} \quad (22)$$

$$H = \begin{bmatrix} h_{1,2} & \cdots & h_{1,q} & \cdots & h_{1,NP} \\ & \ddots & \vdots & \ddots & \vdots \\ & & h_{p,q} & \cdots & h_{p,NP} \\ & & & \ddots & \vdots \\ & & & & h_{NP-1,NP} \end{bmatrix} \quad \begin{array}{l} \text{with } h_{p,q} = d_{p',q'}^{sp*} \quad sp* = \{ \min(sp) / X_p \wedge X_q \in O_{NP+sp} \} \\ \forall p, q \in \{1, \dots, NP\}, p', q' \in \{1, \dots, 2NP - 1\}, sp \in \{1, \dots, NP - 1\} \end{array} \quad (23)$$

274 where, \bar{D}^I is the mean of the original Euclidean distances $d_{p,q}^I$ between all the individuals, $h_{p,q}$ is the linkage
 275 distance $d_{p',q'}^{sp*}$ where the pair of individuals X_p and X_q become members of the same group and \bar{H} is the mean of
 276 the resultant linkage distances $h_{p,q}$ between all the individuals.

277 Recalling that the aim of nesting HCA into DE is to increase the computational performance by decreasing the
 278 number of individuals to be evaluated in each generation G , the presetting of a threshold CCC_{th} for the CCC value
 279 allows defining the level of representativeness required to the hierarchical structure proposed. If the CCC^G obtained
 280 from applying HCA over the corresponding population POP^G is higher than or equal to the threshold CCC_{th} , the
 281 built hierarchical structure is considered an acceptable representation of the original distances amongst the
 282 individuals and the selection of a particular partition of the sets of groups can be performed, i.e., the determination
 283 of a specific number of clusters. Conversely, if CCC^G is less than CCC_{th} , the hierarchical structure is considered not
 284 representative enough since it introduces unacceptable distortion that may affect the global searching process in the
 285 HCDE.

286 Whether the hierarchical structure is accepted, the clustering process itself takes place. As before stated, the HCA
 287 outcome linkage distances $d_{p',q'}^{sp}$ define each level (height) at which a pairing action takes place. If the hierarchical
 288 structure is 'cut off' at a specific linkage distance d_{CO} , all the groups that are formed below the level d_{CO} become
 289 independent clusters. In each generation G of HCDE, a d_{CO} relative to the HCA outcome linkage distances for the
 290 corresponding POP^G , is determined from a preset percentile $p_{d\%tile}$ of the linkage distances between the minimum
 291 $d_{p',q'}^{sp}$ that correspond to the first pairing action and the distance to form at least four clusters needed to perform the
 292 mutation process in the HCDE. Thus, d_{CO} can be obtained from equations (24) and (25). Figure 3 shows the cutoff
 293 distance representation for the example aforementioned, for which the formed clusters are $\{O_2, O_6\}$, $\{O_1\}$, $\{O_7\}$,
 294 $\{O_3, O_4\}$, $\{O_5\}$ and $\{O_8\}$.

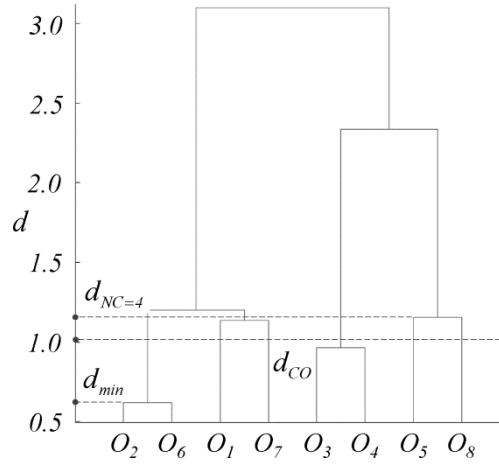


Figure 3. Example of cutoff distance calculation

$$d_{CO} = d_{min} + p_{d\%tile} (d_{NC=4} - d_{min}) \quad (24)$$

$$d_{NC=4} = d_{\left(1 - \frac{4}{NP}\right)\%tile} \quad (25)$$

295

296

297 The integration of HCA into DE and the definition of the parameters CCC_{th} and $p_{d\%tile}$ allow HCDE adaptation at
 298 each generation, i.e., deciding whether to perform HCA and determining the clusters to be taken. Then, the
 299 individuals closest to the centroids of the formed clusters are considered as the representatives of the group which
 300 they belong to and are taken in a reduced population that enters the evolution phase of the HCDE. The proposed
 301 HCDE algorithm is summarized schematically in the flowchart of Figure 4.

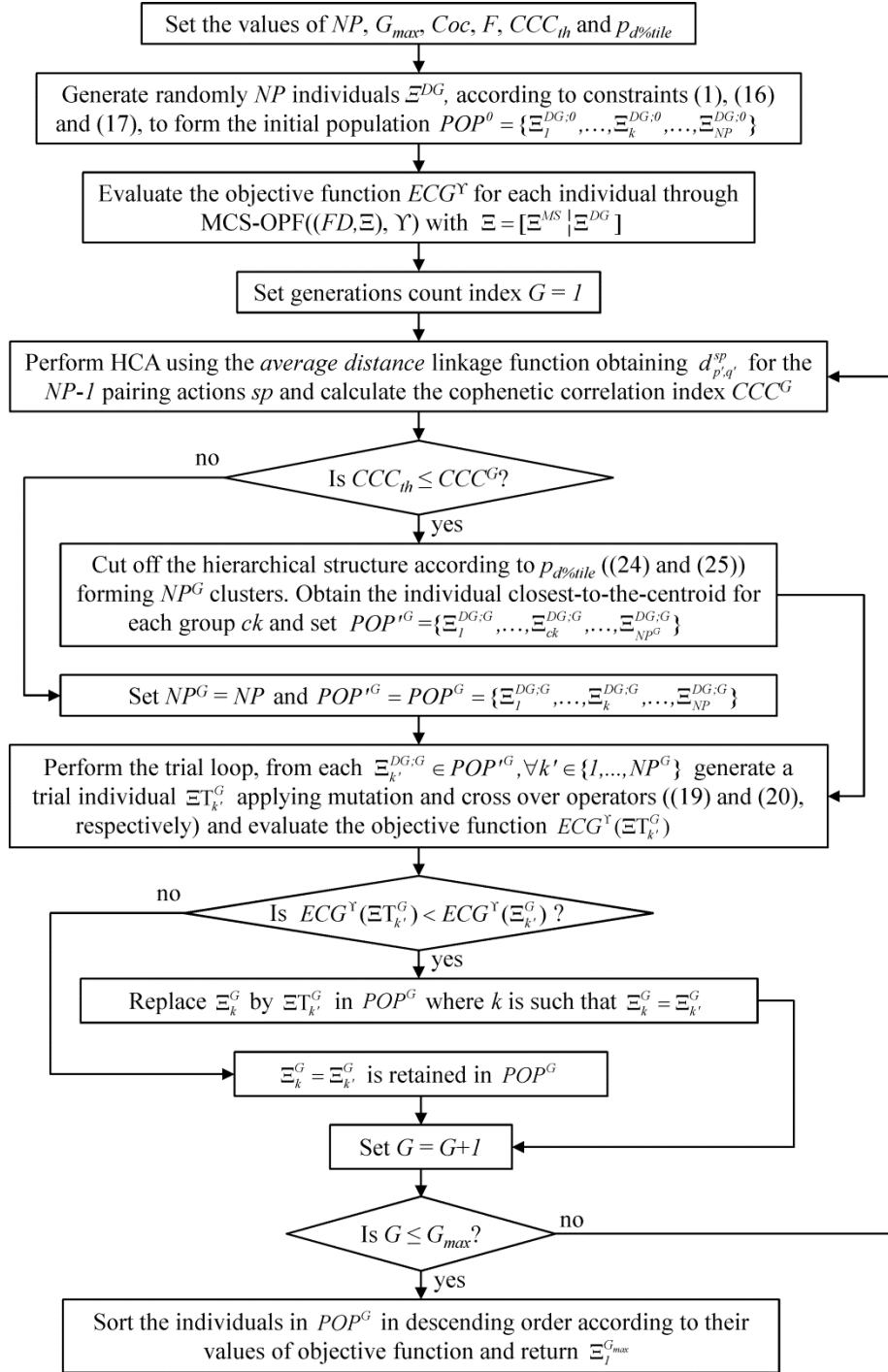


Figure 4. Flowchart of the framework

302

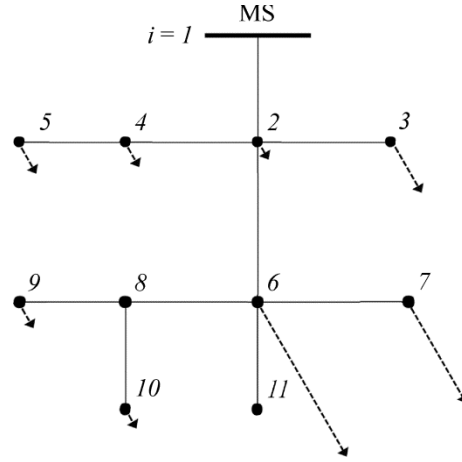
303

304 4 CASE STUDY

305 We consider a modification of the IEEE 13 nodes test feeder distribution network [27] with the original spatial
 306 structure but neglecting the feeders of length zero, the regulator, capacitor and switch. The resulting network has 11
 307 nodes and presents the relevant characteristics of interest for the analysis, e.g. the presence of a main power supply
 308 spot and comparatively low and high spot, and distributed load values [33].

309 4.1 Distribution Network description

310 The distribution network presents a radial structure of $n = 11$ nodes as shown in Fig. 1. The nominal voltage V^{NET} is
 311 4.16 (kV), kept constant for the resolution of the DC optimal power flow problem.



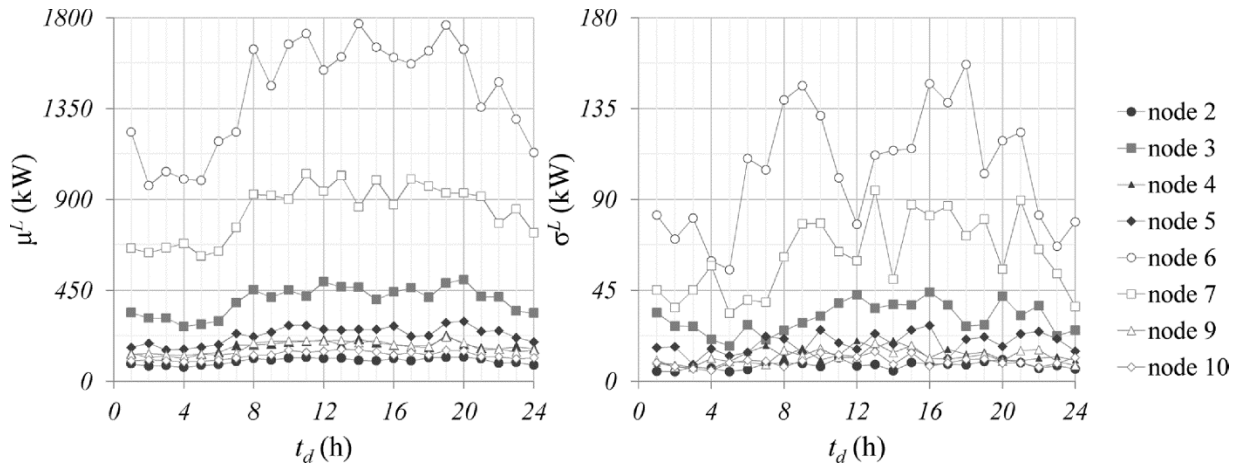
312
 313 Figure 5. Radial 11-nodes distribution network

314 Table 4 contains the technical characteristics of the different types of feeders considered: specifically, the indexes
 315 of the pairs of nodes (i, i') that they connect, their length l , reactance X^{FD} , ampacity A^{FD} and failure and repair rates.

316 Table 4. Feeders characteristic and technical data [11, 27, 28]

Type	node i	node i'	l (km)	X^{FD} (Ω /km)	A^{FD} (A)	λ^F (1/h)	λ^R (1/h)	Cov (\$)
T1	1	2	0.610	0.371	730	3.333e-04	0.198	1.970e-02
T2	2	3	0.152	0.472	340	4.050e-04	0.162	9.173e-03
T3	2	4	0.152	0.555	230	3.552e-04	0.185	6.205e-03
T1	2	6	0.610	0.371	730	3.333e-04	0.198	6.205e-03
T3	4	5	0.091	0.555	230	3.552e-04	0.185	6.205e-03
T6	6	7	0.152	0.252	329	4.048e-04	0.164	8.904e-03
T4	6	8	0.091	0.555	230	3.552e-04	0.185	1.970e-02
T1	6	11	0.305	0.371	730	3.333e-04	0.198	1.970e-02
T5	8	9	0.091	0.555	230	3.552e-04	0.185	9.173e-03
T7	8	10	0.244	0.318	175	3.552e-04	0.185	6.205e-03

317 The nodal power demands are built from the load data given in [27] and reported in Figure 6 as daily profiles,
 318 normally distributed on each hour t_d with mean μ^L and standard deviation σ^L [29, 34].



319

320

Figure 6. Mean and standard deviation values of normally distributed nodal power demand daily profiles

321

322

323

324

325

326

The technical parameters, failure and repair rates and costs of the MS and the four different types of DG technologies (PV, W, EV and ST) available to be integrated into the distribution network are given in Table 5. For the present case study, the distribution region is such that the solar irradiation and wind speed conditions are assumed uniform in the whole network, i.e., the values of the parameters of the corresponding Beta and Rayleigh distributions are assumed constant in the whole network.

Table 5. Power sources parameters and technical data [13, 17, 28, 29, 35-37]

Type	Technical parameters	Distributions parameters, failure and repair rates	Costs
MS	$P_{cap}^{MS} = 4250$ (kW)	$\mu^{MS} = 4000$ (kW) $\sigma^{MS} = 125$ (kW) $\lambda^F = 4.00e-04$ (1/h) $\lambda^R = 1.30e-02$ (1/h)	$Cov = 0.145$ (\$/kWh)
PV	$T_a = 30.00$ (C) $N_{oT} = 43.00$ (C) $I_{sc} = 1.80$ (A) $k_i = 1.40$ (mA/C) $V_{oc} = 55.50$ (V) $k_v = 194.00$ (mV/C) $V_{MPP} = 38.00$ (V) $I_{MPP} = 1.32$ (A)	$\alpha^{PV} = 0.26$ $\beta^{PV} = 0.73$ $\lambda^F = 5.00e-04$ (1/h) $\lambda^R = 1.30e-02$ (1/h)	$Ci = 48$ (\$) $Cov = 3.76e-05$ (\$/kWh)
W	$ws_{ci} = 3.80$ (m/s) $ws_a = 9.50$ (m/s) $ws_{co} = 23.80$ (m/s) $P_R^W = 50.00$ (kW)	$\sigma^W = 7.96$ $\lambda^F = 6.0e-04$ (1/h) $\lambda^R = 1.3e-02$ (1/h)	$Ci = 113,750$ (\$) $Cov = 0.039$ (\$/kWh)
EV	$P_R^{EV} = 6.30$ (kW)	$\lambda^F = 2.0e-04$ (1/h) $\lambda^R = 9.7e-02$ (1/h)	$Ci = 17,000$ (\$) $Cov = 0.022$ (\$/kWh)
ST	$P_R^{ST} = 0.28$ (kW/kg) $SE = 0.04$ (kJ/kg)	$\lambda^F = 3.0e-04$ (1/h) $\lambda^R = 7.3e-02$ (1/h)	$Ci = 135.15$ (\$) $Cov = 4.62e-05$ (\$/kWh)

327

328

The hourly per day operating state probability profiles of the EV are presented in Figure 7: p^0 , p^- and p^+ correspond to the profiles of disconnected, charging and discharging states, respectively.

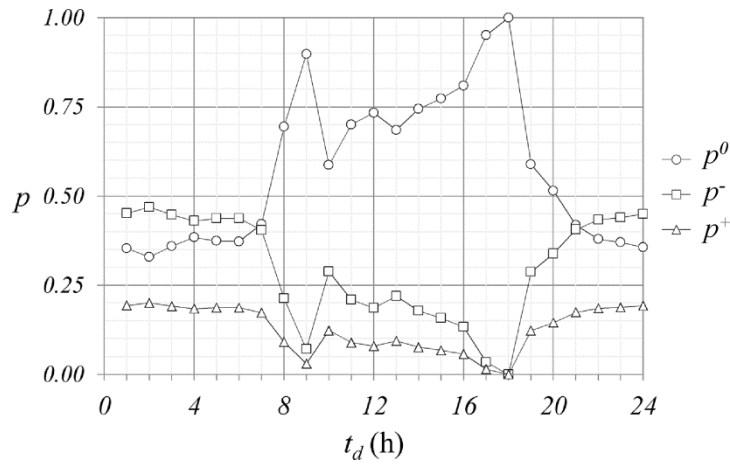


Figure 7. Hourly per day probability data of EV operating states

329

330

331 Coherently with constraints (16) and (17), the budget is set to $BGT = 4,500,000$ (\$) and the limit of units of the
 332 different DG technologies available to be purchased is $\tau = [20000, 8, 250, 10000]$. The maximum value of the
 333 energy price is $ep_h = 0.12$ (\$/kWh) [5] and the highest value of total demand TL_h is set to 4,800 (kW). The
 334 opportunity cost for kWh not supplied Cop is considered as twice of the maximum energy price.

335 A total of $NS = 500$ random scenarios are simulated by the MCS-OPF with time step $ts = 1$ (h), over a horizon of
 336 analysis of 10 years ($th = 87,600$ (h)), in which the investment and fixed costs are prorated hourly.

337 The DE iterations are set to perform $G_{max} = 500$ generations over five different cases of population $NP \in \{10, 20,$
 338 $30, 40, 50\}$. The differential variation amplification factor F is 1 to maintain the integer-valued definition of the
 339 individuals after the mutation, whereas the crossover coefficient Coc is 0.1.

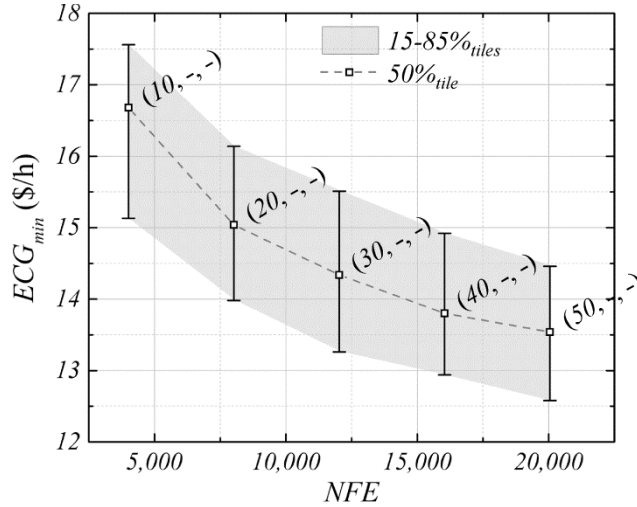
340 HCDE runs are performed under the same conditions set for DE (G_{max} , F and Coc), but for the population size NP
 341 of 50 individuals. A sensitivity analysis is performed over the HCA control parameters, namely the cophenetic
 342 correlation coefficient CCC_{th} and linkage distance percentile $p_{d\%tile}$, for all the nine possible pairs (CCC_{th} , $p_{d\%tile}$)
 343 with $CCC_{th} \in \{0.6, 0.7, 0.8\}$ and $p_{d\%tile} \in \{25\%_{tile}, 50\%_{tile}, 75\%_{tile}\}$. Finally, for each of the five DE and nine
 344 HCDE settings, twenty realizations are carried out.

345 4.2 Results and Discussion

346 The results of the DE MCS-OPF for the different population sizes $NP \in \{10, 20, 30, 40, 50\}$ are shown in Figure 8.
 347 The 50%tile (median) values of the minimum global costs EGC_{min} , obtained from each experiment with fixed
 348 values of NP , are presented as functions of the respective numbers of objective function evaluations NFE ; the error
 349 bars represent the 15 and 85%tiles.

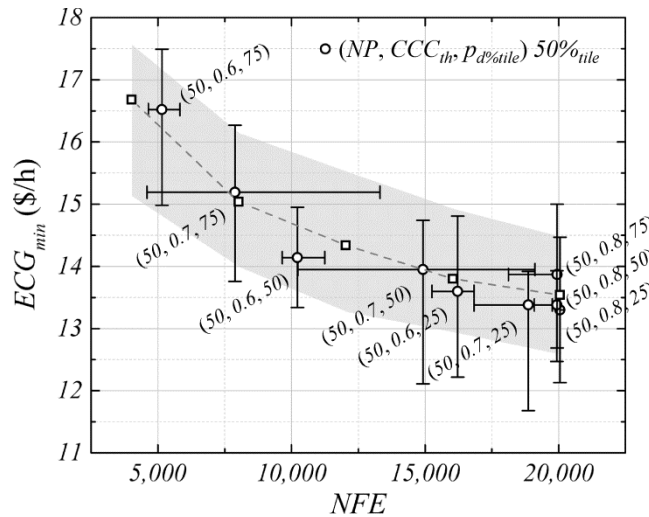
350 As expected, for the same number of generations set in the DE MCS-OPF, the larger the population size considered
 351 the lower the values of EGC_{min} obtained (better ‘quality’ of the minimum). Additionally, we can observe marked
 352 tendencies in the reduction of both median and 15-85%tiles values of EGC_{min} for increasing NFE . Performing a
 353 curve fitting over these values, we get: $EGC_{min;50\%tile} = 49.07NFE^{-0.13}$, $EGC_{min;15\%tile} = 49.07NFE^{-0.115}$ and

354 $ECC_{min,85\%tile} = 49.07NFE^{-0.118}$, with the respective coefficients of determination $R^2_{50\%tile} = 0.994$, $R^2_{15\%tile} = 0.998$
 355 and $R^2_{85\%tile} = 0.998$. The fact that the difference between the values of the 15-85%tiles is constant indicates that
 356 the dispersion in the $ECC_{min}(NFE)$ does not depend on NP and can suggest that the global searching performed by
 357 the DE is performed homogenously in the feasible space that contains multiple local minima.



358
 359 Figure 8. ECC_{min} vs NFE for $NP \in \{10, 20, 30, 40, 50\}$ set in DE

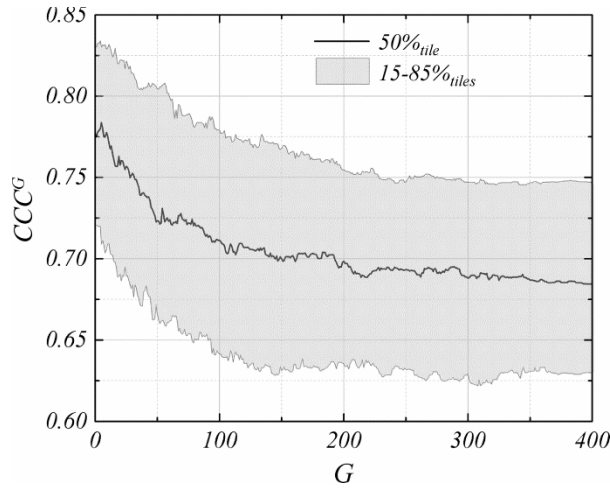
360 Figure 9 reports the median ECC_{min} values corresponding to the HCDE MCS-OPF realizations superposed to the
 361 distribution of the median ECC_{min} and 15-85%tiles values of the base DE experiments represented by the square
 362 markers and shaded area, respectively. The vertical and horizontal error bars account for the 15-85%tiles of the
 363 outcome ECC_{min} and NFE values.



364
 365 Figure 9. ECC_{min} vs NFE for each $(NP, CCC_{th}, P_d\%tile)$ set in HCDE

366 Focusing on CCC_{th} , it can be noticed that for the two extreme cases, $CCC_{th} = 0.6$ and 0.8 , the dispersion of the
 367 number of objective function evaluations is relatively small. On the contrary, the cases with a $CCC_{th} = 0.7$ present
 368 high variability. This can be explained by the behavior of the CCC along each generation G in the evolution loop.

369 Figure 10 shows the median, 15 and 85%tiles CCC values as a function of generation G derived from all HCDE
 370 MCS-OPF realizations. On the one hand, recalling that CCC_{th} is used to control whether it is convenient to perform
 371 HCA, the small NFE dispersion in the case with $CCC_{th} = 0.6$ is because clustering is practically been applied in all
 372 generations ($CCC_{th} \leq CCC^G$), thus disabling any effect generated by passing from populations with original size NP
 373 to reduced populations with $NP^G \leq NP$ and vice versa. On the other hand, the effect is also being avoided in the
 374 case $CCC_{th} = 0.8$ by not applying clustering. Indeed, in Figure 10 it can be observed that after the generation 50 it is
 375 unlikely that by performing HCA the proposed hierarchical grouping structures represent well enough the
 376 population.



377
 378 Figure 10. CCC behavior per generation G

379 Differently, the cases for which $CCC_{th} = 0.7$ present high dispersion in the NFE since the median values of CCC^G
 380 move in the neighborhood of the threshold throughout the major part of the evolution loop in the HCDE. Moreover,
 381 in general terms, the values of CCC^G 15-85%tiles maintain certain symmetry with respect to the median, i.e.,
 382 performing or not HCA are equally likely events, producing high fluctuations in the number of individuals
 383 considered as population and, therefore, affecting in the same way the NFE .

384 The above mentioned insights are noticeable also in Figure 11, which shows the empirical probability density
 385 functions ($pdfs$) of the population size NP^G per generation for each $(NP, CCC_{th}, p_{d\%tile})$ set in HCDE. Indeed, the
 386 average probabilities of performing HCA throughout the evolution cycle for the different values of $CCC_{th} = 0.6, 0.7$
 387 and 0.8 are $0.98, 0.54$ and 0.078 , respectively.

388 Regarding the percentile of the linkage distance $p_{d\%tile}$, in Figure 11 it is possible to identify the three peaks of
 389 reduction in the population size, confirming the role of this control parameter in defining the scale at which the
 390 hierarchical structures proposed are ‘cut off’ when the HCA takes place. In fact, lower values of $p_{d\%tile}$ imply
 391 smaller reduction in the population size because of the higher demand of proximity between individuals or groups
 392 of individuals. In the opposite side, higher values of $p_{d\%tile}$ allow forming clusters from individuals or groups which
 393 are relatively less similar.

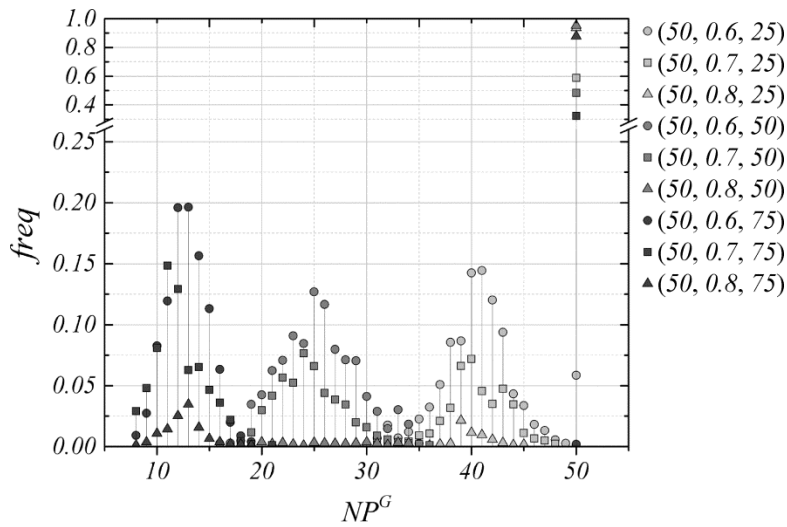


Figure 11. Empirical NP^G pdf for each $(NP, CCC_{th}, p_{d\%tile})$ set in HCDE

394

395

396 From the results obtained for all the different DE and HCDE settings, we look for six representative cases for the
 397 analysis (Figure 9). From the DE runs, we select the settings with extreme and middle population size $NP \in \{10,$
 398 $30, 50\}$, whereas from HCDE we choose the cases $(NP, CCC_{th}, p_{d\%tile})$ set as $(50, 0.6, 25)$, $(50, 0.6, 50)$, $(50, 0.7,$
 399 $50)$ and $(50, 0.7, 75)$. The former $(50, 0.6, 25)$ and $(50, 0.6, 50)$ cases present significant reductions in the number
 400 of NFE , with small dispersion and loss of quality of the minimum ECG obtained, compared to the results obtained
 401 by diminishing directly the fixed NP in DE from 50 to 10. Similarly, the cases $(50, 0.7, 50)$ and $(50, 0.7, 75)$ may
 402 lead to considerable reductions in NFE , with acceptable losses of ECG_{min} , but subject to a high degree of variability
 403 that compromises the performance.

404 As for computational times, running on an Intel® Core™ i7-3740QM (PC) 2.70GHz without performing parallel
 405 computing, the average time to evaluate the objective function is 4.592 (s) for the $NS = 500$ scenarios in the MCS-
 406 OPF; for a fixed population of $NP = 50$ and its corresponding $NFE = 20,050$, the total time for a single run is on
 407 average 25.574 (h). Taking into account this, under commonly used hardware configurations, the reductions in
 408 computational time that can be achieved by using HCDE with $(50, 0.6, 25)$ and $(50, 0.6, 50)$ settings are 19% and
 409 49% for the median, 23% and 51% for the 15%tile, and 16% and 43% for the 85%tile, respectively.

410 The integration of HCA into the DE algorithm introduces a significant time complexity, conditioning the
 411 reductions of computational efforts that can be obtained by applying the proposed HCDE MCS-OPF framework.
 412 Indeed, if performing HCA along all generations of DE and running the MCS-OPF on an eventually reduced
 413 population (depending on CCC_{th} and $p_{d\%tile}$) is computationally heavier than running the MCS-OPF over the
 414 complete population, the effects of the framework can be negligible or even negative.

415 It is possible to formulate the condition to obtain reductions in the computational efforts by the proposed HCDE
 416 MCS-OPF framework, from the asymptotic time complexities of the main algorithms that compose it. Table 6

417 reports the independent asymptotic time complexities as functions of the generic size m of the input to each
 418 algorithm and of the parameters that define the dimensionality of the HCDE MCS-OPF framework [26, 38].

419 Table 6. Asymptotic time complexity of the algorithms

	Algorithm			
	PDIST	HC	MCS	OPF
Time complexity T	$O(dm^2)^*$	$O(m^2 \log(m))$	$O(m)$	$O(\text{size}(A))^{**}$
	$O(nps \times NP^2)$	$O(NP^2 \log(NP))$	$O(NS \times nps)$	$O(NS \times nps^2)$

* Pairwise distance PDIST between all m vectors of size d
 ** The matrix A comes from the canonical form $Ax \leq b$ of the linear programming of the DC OPF problem approximation

420 where, nps represents the size of the DG-integrated network, i.e., the number of nodes n times the number of all the
 421 technologies of power generation available ps , NP is the size of the complete population and NS is the number of
 422 scenarios in the MCS-OPF.

423 Comparing the asymptotic time complexities of the algorithms involved in the realization of the proposed
 424 framework with and without integrating HCA, the following inequalities must be fulfilled in order to obtain a
 425 reduction in the computational time by HCDE:

$$\begin{aligned}
 & T^{\text{PDIST}}(nps, NP) + T^{\text{HC}}(NP) + E[NP^G] \times T^{\text{MCS-OPF}}(NS, nps) < NP \times T^{\text{MCS-OPF}}(NS, nps) \\
 & \Downarrow \\
 & nps \times NP^2 + NP^2 \log(NP) + E[NP^G] \times NS \times nps^2 < NP \times NS \times nps^2 \\
 & \Downarrow \\
 & \kappa = \frac{NP}{NS \times nps} + \frac{NP \log(NP)}{NS \times nps^2} + \varepsilon < 1 \quad \forall n, ps, NP, NS \in \mathbb{Z}^*, \varepsilon = \frac{E[NP^G]}{NP} \in (0, 1] \tag{26}
 \end{aligned}$$

426 where, ε is the expected ratio of the population NP^G evaluated along all generations G of DE to the total population
 427 NP and κ is the ratio of the asymptotic time complexities of HCDE to DE.

428 From equation (26), we can observe that the contribution of the terms related with the complexity of MCS-OPF,
 429 dependent on NS and nps , is considerably large for the fulfilment of the inequality conditions. In fact, when using
 430 DE, it is commonly accepted to set a size of the population NP not greater than ten times the size of the decision
 431 variables, in this case, $10nps$ [24], making the first two terms of κ strongly dependent on the number of scenarios
 432 NS . Moreover, given the complexity of the general problem, higher values of NS lead to a better approximation of
 433 the objective function via MCS-OPF, i.e., the more likely is to fulfill the condition and the greater can be the
 434 reduction of computation time. However, the value of ε depends on the probability of performing clustering in each
 435 generation and at what scale, controlled by CCC_{th} and $p_{d\%tile}$ respectively. In some cases, ε can be close to 1 (as we
 436 inferred from Figure 11) implying negligible benefits. Table 7 shows the values of the ratio κ for each $(NP, CCC_{th},$
 437 $p_{d\%tile})$ set in HCDE considering the dimensionality of the present case study defined by the values of the
 438 parameters $nps = 55, NS = 500, NP = 50$. The value of $1 - \kappa$ can be interpreted as the expected asymptotic relative
 439 time reduction achieved by performing HCDE.

Table 7. Ratio κ for each $(NP, CCC_{th}, p_{d\%tile})$

$(NP, CCC_{th}, p_{d\%tile})$	$\frac{NP}{NS \times nps}$	$\frac{NP \log(NP)}{NS \times nps^2}$	$\varepsilon = \frac{E[NP^G]}{NP}$	κ	$1 - \kappa$
(50, 0.6, 25)	1.818E-03	3.418E-05	0.817	0.819	0.181
(50, 0.7, 25)			0.921	0.923	0.077
(50, 0.8, 25)			0.987	0.989	0.011
(50, 0.6, 50)			0.510	0.512	0.488
(50, 0.7, 50)			0.738	0.740	0.260
(50, 0.8, 50)			0.978	0.979	0.021
(50, 0.6, 75)			0.259	0.261	0.739
(50, 0.7, 75)			0.487	0.488	0.512
(50, 0.8, 75)			0.909	0.911	0.089

441 Figure 12 shows the convergence curves for the DE and HCDE cases selected, for the twenty runs performed for
 442 each $(NP, CCC_{th}, p_{d\%tile})$ setting: no significant differences can be found among the convergence curves except for
 443 the expected behavior of converging to lower values of E_{GC}_{min} for settings which imply a larger population size.

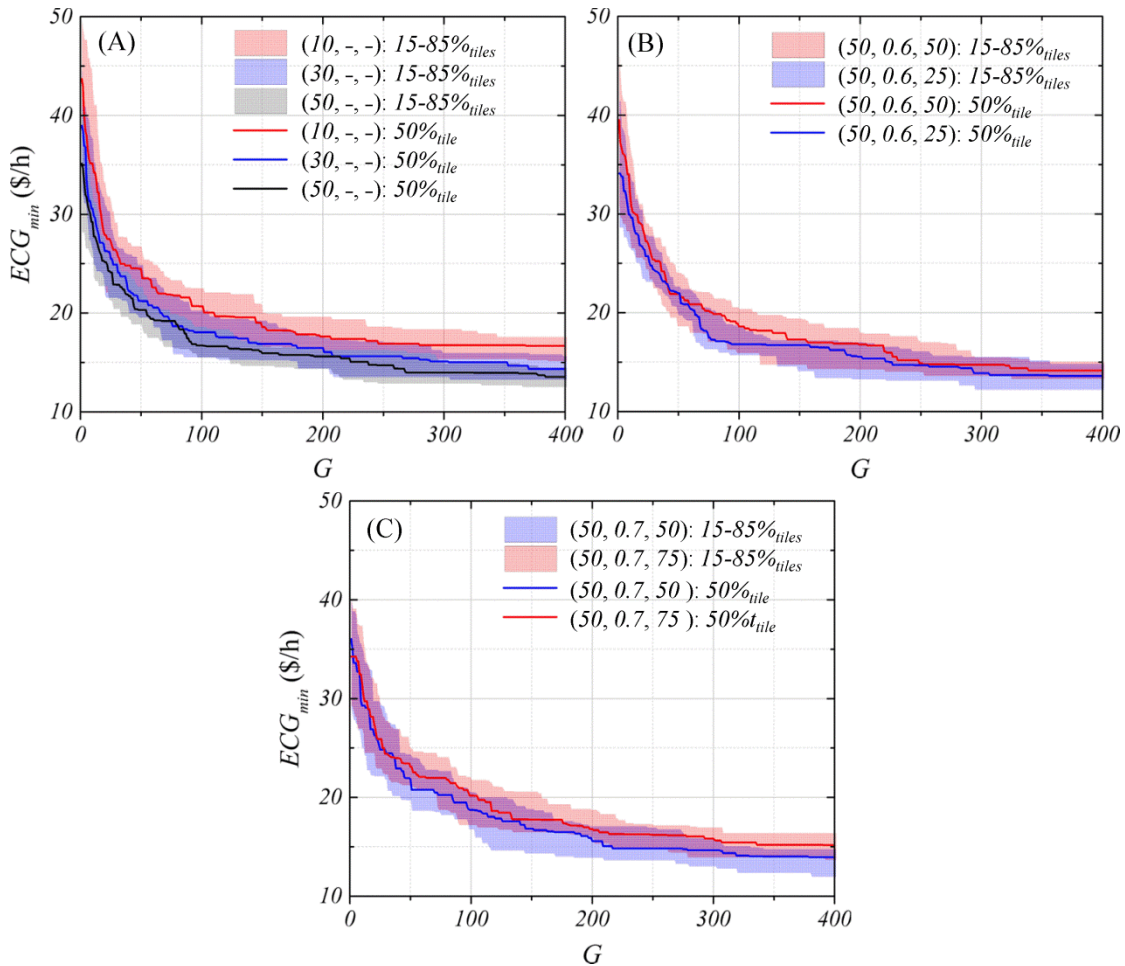
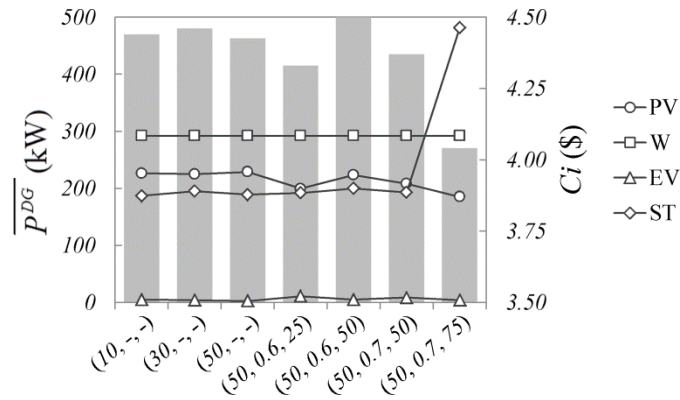


Figure 12. Convergence curves for representative $(NP, CCC_{th}, p_{d\%tile})$ settings

446 Figure 13 shows the average total DG power allocated in the distribution network and the corresponding
 447 investment costs of the DE and HCDE MCS-OPF cases selected, choosing the corresponding optimal DG-

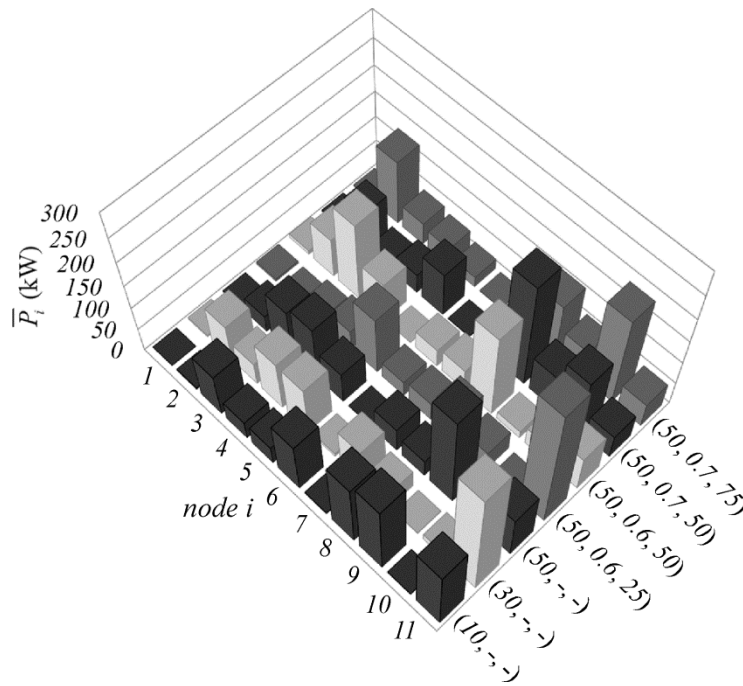
448 integrated plans as the decision matrixes Ξ^{DG} for which their ECC_{min} values are the closest to the median ECC_{min}
 449 value obtained for the twenty runs of each $(NP, CCC_{th}, p_{d\%tile})$ setting. It can be pointed out that in all the cases, the
 450 contribution of EV is practically negligible if compared with the other technologies. This is due to a combination of
 451 two facts: the probability that the EV is in a discharging state is much lower than that of being in the other two
 452 possible operating states, charging and disconnected (see Figure 7) and when EV is charging, the effects are
 453 opposite to those desired, i.e., it is acting as loads.



454

455 Figure 13. Average total DG power allocated and investment cost for representative $(NP, CCC_{th}, p_{d\%tile})$ settings

456 In all generality, both the investment cost C_i and the average power installed by DG is comparable in all the cases,
 457 except for the setting $(50, 0.7, 75)$ for which the scale of clustering determined by $p_{d\%tile} = 75\%$, that translates into
 458 higher reductions of the population size, may lead to less similar local minima than the other settings.



459

460

Figure 14. Nodal average total DG power for representative $(NP, CCC_{th}, p_{d\%tile})$ settings

461 The average total renewable DG power allocated per node is summarized in Figure 14. Even though all the *ECG*
462 optimal decision matrixes Ξ^{DG} show differences, the tendency is to install localized sources of renewable DG
463 power between two identifiable portions of the distribution network, up and downstream the feeder (2,6) (Figure 5),
464 giving preference to the second portion which presents higher and non-stream homogeneous nodal load profiles.

465 5 CONCLUSIONS

466 In a previous paper, we have presented a simulation and optimization framework for the planning of integration of
467 renewable generation into a distribution network. The optimization is considered with respect the objective of
468 minimizing the expected global cost of the system. The inherent uncertain behavior of renewable energy sources,
469 variability in the main power supply and loads, as well as the possibility of failures of network components are
470 included in a Monte Carlo simulation, which samples realizations of the uncertain operational scenarios for the
471 optimal power flow.

472 The framework is quite general and complete in the characteristics of the realistic system scenarios considered.
473 However, this is at the expenses of the computational time required for the overall optimization.

474 In this respect, in the present paper we have addressed the problem of computational efficiency in the resolution of
475 the renewable DG planning optimization problem. We have done so by an original introduction of a controlled
476 clustering strategy, with, the main original contributions being:

- 477 · The integration of differential evolution and hierarchical clustering analysis for grouping similar individuals
478 from a given population and selecting representatives to be evaluated for each group, thus reducing the number
479 of objective function evaluations during the optimization.
- 480 · The introduction of two control parameters, namely the cophenetic correlation coefficient and a percentile of
481 the set of linkage distances, for allowing controlled adaptation during the search process and decision on
482 whether or not to perform clustering and at which level of the hierarchical structure built.

483 A case study has been analyzed derived from the IEEE 13 nodes test feeder. The results obtained show the
484 capability of the framework to identify optimal plans of renewable DG integration. The sensitivity analysis over the
485 control parameters of the hierarchical clustering shows that the efficiency is improved with cophenetic correlation
486 thresholds that allow the clustering in almost all generations along the differential evolution, setting the scale of
487 clustering to no more than the fiftieth percentile of the linkage distances in the hierarchical structure proposed.
488 Indeed, this is shown to lead to acceptable reductions in the number of objective function evaluations, with small
489 dispersion and loss of quality in the minimum global cost obtained.

490 References

491

- 492 [1] A. Alarcon-Rodriguez, G. Ault, and S. Galloway, "Multi-objective planning of distributed energy
493 resources: A review of the state-of-the-art," *Renewable and Sustainable Energy Reviews*, vol. 14, pp. 1353
494 - 1366, 2010.
- 495 [2] C. L. T. Borges, "An overview of reliability models and methods for distribution systems with renewable
496 energy distributed generation," *Renewable and Sustainable Energy Reviews*, vol. 16, pp. 4008-4015, 2012.
- 497 [3] V. F. Martins and C. L. T. Borges, "Active Distribution Network Integrated Planning Incorporating
498 Distributed Generation and Load Response Uncertainties," *Power Systems, IEEE Transactions on*, vol. 26,
499 pp. 2164 -2172, nov. 2011.
- 500 [4] H. Ren and W. Gao, "A MILP model for integrated plan and evaluation of distributed energy systems,"
501 *Applied Energy*, vol. 87, pp. 1001 - 1014, 2010.
- 502 [5] H. Ren, W. Zhou, K. a. t. Nakagami, W. Gao, and Q. Wu, "Multi-objective optimization for the operation
503 of distributed energy systems considering economic and environmental aspects," *Applied Energy*, vol. 87,
504 pp. 3642 - 3651, 2010.
- 505 [6] C. Chen, S. Duan, T. Cai, B. Liu, and G. Hu, "Optimal Allocation and Economic Analysis of Energy
506 Storage System in Microgrids," *Power Electronics, IEEE Transactions on*, vol. 26, pp. 2762 -2773, oct.
507 2011.
- 508 [7] S. Haffner, L. F. A. Pereira, L. A. Pereira, and L. S. Barreto, "Multistage model for distribution expansion
509 planning with distributed generation - Part I: Problem formulation," *Ieee Transactions on Power Delivery*,
510 vol. 23, pp. 915-923, Apr 2008.
- 511 [8] S. Haffner, L. F. A. Pereira, L. A. Pereira, and L. S. Barreto, "Multistage model for distribution expansion
512 planning with distributed generation - Part II: Numerical results," *Ieee Transactions on Power Delivery*,
513 vol. 23, pp. 924-929, Apr 2008.
- 514 [9] M. E. Samper and A. Vargas, "Investment Decisions in Distribution Networks Under Uncertainty With
515 Distributed Generation-Part II: Implementation and Results," *Ieee Transactions on Power Systems*, vol. 28,
516 pp. 2341-2351, Aug 2013.
- 517 [10] T. Niknam, S. I. Taheri, J. Aghaei, S. Tabatabaei, and M. Nayeripour, "A modified honey bee mating
518 optimization algorithm for multiobjective placement of renewable energy resources," *Applied Energy*, vol.
519 88, pp. 4817-4830, Dec 2011.
- 520 [11] S. Ganguly, N. C. Sahoo, and D. Das, "A novel multi-objective PSO for electrical distribution system
521 planning incorporating distributed generation," *Energy Systems*, vol. 1, pp. 291-337, 2010.
- 522 [12] M. Gomez-Gonzalez, A. LÃ³pez, and F. Jurado, "Optimization of distributed generation systems using a
523 new discrete PSO and OPF," *Electric Power Systems Research*, vol. 84, pp. 174 - 180, 2012.
- 524 [13] K. Zou, A. P. Agalgaonkar, K. M. Muttaqi, and S. Perera, "Multi-objective optimisation for distribution
525 system planning with renewable energy resources," in *Energy Conference and Exhibition (EnergyCon),
526 2010 IEEE International*, ed, 2010, pp. 670 -675.
- 527 [14] H. A. Hejazi, A. R. Araghi, B. Vahidi, S. H. Hosseinian, M. Abedi, and H. Mohsenian-Rad, "Independent
528 Distributed Generation Planning to Profit Both Utility and DG Investors," *Ieee Transactions on Power
529 Systems*, vol. 28, pp. 1170-1178, May 2013.
- 530 [15] H. A. Hejazi, M. A. Hejazi, G. B. Gharehpetian, and M. Abedi, "Distributed generation site and size
531 allocation through a techno economical multi-objective Differential Evolution Algorithm," in *Power and
532 Energy (PECon), 2010 IEEE International Conference on*, ed, 2010, pp. 874 -879.
- 533 [16] M. F. Shaaban, Y. M. Atwa, and E. F. El-Saadany, "DG Allocation for Benefit Maximization in
534 Distribution Networks," *Ieee Transactions on Power Systems*, vol. 28, pp. 639-649, May 2013.
- 535 [17] M. Raoofat, "Simultaneous allocation of DGs and remote controllable switches in distribution networks
536 considering multilevel load model," *International Journal of Electrical Power and Energy Systems*, vol.
537 33, pp. 1429 - 1436, 2011.

- 538 [18] Z. H. Cai, W. Y. Gong, C. X. Ling, and H. Zhang, "A clustering-based differential evolution for global
539 optimization," *Applied Soft Computing*, vol. 11, pp. 1363-1379, Jan 2011.
- 540 [19] M.-Y. Cheng, D.-H. Tran, and Y.-W. Wu, "Using a fuzzy clustering chaotic-based differential evolution
541 with serial method to solve resource-constrained project scheduling problems," *Automation in
542 Construction*, vol. 37, pp. 88-97, 1// 2014.
- 543 [20] G. Liu, Y. X. Li, X. Nie, and H. Zheng, "A novel clustering-based differential evolution with 2 multi-
544 parent crossovers for global optimization," *Applied Soft Computing*, vol. 12, pp. 663-681, Feb 2012.
- 545 [21] R. Mukherjee, G. R. Patra, R. Kundu, and S. Das, "Cluster-based differential evolution with Crowding
546 Archive for niching in dynamic environments," *Information Sciences*, 2014.
- 547 [22] S. Song and X. J. Yu, "Multi-peak function optimization using a hierarchical clustering based genetic
548 algorithm," *ISDA 2006: Sixth International Conference on Intelligent Systems Design and Applications,
549 Vol 1*, pp. 425-428, 2006.
- 550 [23] Y. J. Wang, J. S. Zhang, and G. Y. Zhang, "A dynamic clustering based differential evolution algorithm for
551 global optimization," *European Journal of Operational Research*, vol. 183, pp. 56-73, Nov 16 2007.
- 552 [24] R. Storn and K. Price, "Differential evolution - A simple and efficient heuristic for global optimization
553 over continuous spaces," *Journal of Global Optimization*, vol. 11, pp. 341-359, Dec 1997.
- 554 [25] R. Mena, M. Hennebel, Y. Li, C. Ruiz, and E. Zio, "A Risk-Based Simulation and Multi-Objective
555 Optimization Framework for the Integration of Distributed Renewable Generation and Storage,"
556 *Renewable and Sustainable Energy Reviews*, vol. 37, pp. 778-793, 2014.
- 557 [26] B. S. Everitt, S. Landau, M. Leese, and D. Stahl, *Cluster Analysis*: Wiley, 2011.
- 558 [27] IEEE Power and Energy Society. *Distribution Test Feeders*. Available:
559 <http://ewh.ieee.org/soc/pes/dsacom/testfeeders/index.html>
- 560 [28] H. Falaghi, C. Singh, M.-R. Haghifam, and M. Ramezani, "DG integrated multistage distribution system
561 expansion planning," *International Journal of Electrical Power and Energy Systems*, vol. 33, pp. 1489 -
562 1497, 2011.
- 563 [29] Y. M. Atwa, E. F. El-Saadany, M. M. A. Salama, and R. Seethapathy, "Optimal Renewable Resources Mix
564 for Distribution System Energy Loss Minimization," *Power Systems, IEEE Transactions on*, vol. 25, pp.
565 360 -370, feb. 2010.
- 566 [30] Y. Li and E. Zio, "Uncertainty analysis of the adequacy assessment model of a distributed generation
567 system," *Renewable Energy*, vol. 41, pp. 235 - 244, 2012.
- 568 [31] D. V. Hertem, "Usefulness of DC power flow for active power flow analysis with flow controlling
569 devices," *AC and DC Power Transmission, IEEE International Conference on*, 2006.
- 570 [32] K. Purchala and L. Meeus, "Usefulness of DC power flow for active power flow analysis," *Power
571 Engineering and Optimization*, 2005.
- 572 [33] W. H. Kersting, "Radial distribution test feeders," *IEEE Transactions on Power Systems*, vol. 6, pp. 975-
573 985, 1991.
- 574 [34] L. F. Wang and C. Singh, "Multicriteria Design of Hybrid Power Generation Systems Based on a Modified
575 Particle Swarm Optimization Algorithm," *Ieee Transactions on Energy Conversion*, vol. 24, pp. 163-172,
576 Mar 2009.
- 577 [35] Y.-F. Li and E. Zio, "A multi-state model for the reliability assessment of a distributed generation system
578 via universal generating function," *Reliability Engineering & System Safety*, vol. 106, pp. 28-36, 2012.
- 579 [36] F. Pilo, G. Celli, S. Mocci, and G. G. Soma, "Active distribution network evolution in different regulatory
580 environments," in *Power Generation, Transmission, Distribution and Energy Conversion (MedPower
581 2010), 7th Mediterranean Conference and Exhibition on*, ed, 2010, pp. 1 -8.

- 582 [37] R. Webster, "Can the electricity distribution network cope with an influx of electric vehicles?," *Journal of*
583 *Power Sources*, pp. 217-225, 1999.
- 584 [38] B. Korte and J. Vygen, *Combinatorial Optimization: Theory and Algorithms*: Springer Publishing
585 Company, Incorporated, 2007.
- 586
- 587