Evaluation of methods to select representative days for the optimization of polygeneration systems

Edwin S. Pinto, Luis M. Serra, Ana Lázaro

GITSE I3A, Department of mechanical engineering, University of Zaragoza, Spain

Abstract

The optimization of polygeneration systems considering hourly periods throughout one year is a computationally demanding task, and, therefore, methods for the selection of representative days are employed to reproduce reasonably the entire year. However, the suitability of a method strongly depends on the variability of the time series involved in the system. This work compares the methods Averaging, k-Medoids and OPT for the selection of representative days by carrying out the optimization of grid-connected and standalone polygeneration systems for a building in two different locations. The suitability of the representative days obtained with each method were assessed regarding the optimization of the polygeneration systems. Sizing errors under 5% were achieved by using 14 representative days, and the computational time, with respect to the entire year data, was reduced from hours to a few seconds. The results demonstrated that the Averaging method is suitable when there is low variability in the time series data; but, when the time series presents high stochastic variability (e.g., consideration of wind energy), the OPT method presented better performance. Also, a new method has been developed for the selection of representative days by combining the k-Medoids and OPT methods, although its implementation requires additional computational effort.

Keywords: Representative days, Renewable energy, polygeneration systems, MILP.

1. Introduction

Nowadays it is imperative to search for strategies for the rational use of energy. In this sense, polygeneration systems are a suitable alternative because of the more efficient use of available natural resources, reduction in CO_2 emissions and economic savings in comparison to separate production [1]. To this end, the synthesis of the system configuration (installed technologies and capacities) and operation planning (operation strategy of equipment) must be considered in the design of polygeneration systems [2,3].

The optimal design of polygeneration systems for buildings is a challenging task, due to the wide variety of energy resources, available technologies, and significant diurnal and seasonal fluctuations in energy demands and tariffs [4]. The incorporation of renewable energy technologies can further increase the complexity of the system design. Some renewable energy technologies, such as wind turbines, photovoltaic panels, and solar thermal collectors, present intermittent behavior due to the non-manageable nature of the availability of the energy resource. Moreover, very often there is non-simultaneity between production and consumption. The integration of energy storage (electrical and/or thermal) enables the decoupling of production from consumption and further increases the complexity of the system design.

Mixed Integer Linear Programming (MILP) is widely utilized for the design of polygeneration systems [2,5,6]. Ideally, the optimization of an energy system should consider hourly or sub-hourly periods throughout one or more years. Nonetheless, the solution of MILP problems can easily become intractable as the computational effort increases with the size of the problem, and more specifically with the number of binary variables, which are widely used in synthesis problems and to model the performance of components. The size of the problem is proportional to the number of time series (e.g. number of energy demands or renewable energy production considered, among others), and consequently the system complexity increases as more time series are included.

1

A viable approach to reduce the complexity of the optimization problem is to reduce the number of time periods by employing of typical or representative days, which enables the reasonable reproduction of the behavior of the system throughout long time periods such as months, or one or several years, reducing the computational effort significantly. There are several methodologies for the selection of representative days, such as graphical methods [7], statistical methods [8], aggregation methods [9,10], and the OPT method [11]. Aggregation methods such as the Averaging, k-Means, k-Medoids, and hierarchical clustering, could be considered the most common alternatives for the design of energy systems [9]. For instance, an aggregation method has been utilized along with a first-order Markov model, to generate synthetic daily horizontal irradiation to design photovoltaic systems [12,13]. Within the aggregation methods, k-Medoids is considered as the most reliable [9,10]. However, the variability of the original time series is smoothed when aggregation methods are employed to obtain representative days [10]. Despite not being an exact method, the OPT method has also been applied to select representative days for a system that encompasses time series with high variability such as electricity demand, photovoltaic and wind energy production [11]. Aggregations methods have been previously compared [9,10], but non-exact methods, such as the OPT, have not been included to date in the comparison of methods for the selection of representative days.

In this sense, the objective of the work presented herein is to show the advantages and disadvantages of some methodologies for the selection of representative days, namely the Averaging, k-Medoids [14] and OPT [11] methods. Considering that tackling a high number of attributes is challenging when optimizing polygeneration systems, this work has considered up to seven attributes: three hourly energy demands (heating, cooling and electricity), three renewable energy production technologies (photovoltaic PV, solar thermal ST and wind turbine WT), and hourly CO_2 emissions associated with the electricity from the grid. Some of these attributes, present stochastic behavior, e.g., wind turbine production, which affects the selection of the method employed to obtain representative days for the design of polygeneration systems. This study aims to provide general guidelines for the selection of a suitable method for the selection of representative days, to address the design of polygeneration systems appropriately, according to the nature of the attributes considered. Moreover, a procedure that combines the k-Medoids and OPT methods is presented as an alternative to overcome some limitations of these methods.

Zaragoza and Gran Canaria were selected to evaluate the suitability of each method regarding the optimization of a polygeneration system for a residential building. These Spanish locations present very different climatic conditions and available local energy resources, leading to the consideration of different scenarios with richer variability, and the obtainment of more general conclusions than those obtained for a single location. Zaragoza presents significant heating demands which can result in the installation of a natural gas-fuelled cogeneration system (manageable resource). Gran Canaria, in turn, presents high wind energy resources, which are non-manageable and present a significant stochastic behavior. Therefore, in this case, the installation of wind turbines can be an attractive solution in the optimal configuration. In both locations, two systems with important different features were considered: grid-connected and standalone. Similarly to previous works [9,10], twelve representative days were selected from each method to represent the entire year. Polygeneration systems were optimized using the results obtained from each method, which were compared to those obtained using whole hourly year data. The Quality Function Deployment (QFD) was applied to evaluate the suitability of each method in different systems.

2. Materials and Methods

2.1 Methods for the selection of representative days

This section presents the Averaging, k-Medoids and OPT methods. Also, a combination of the k-Medoids and OPT methods is proposed. Time series henceforth are referred to as attributes. Values for each attribute c corresponding to a time period t (1-8760 hours) or a day d (1-365 days) and hour h (1-24 hours) can be represented as $x'_{c,t}$ or $x'_{c,d,h}$.

2

2.1.1 Averaging method

This method has been employed previously in several works [9], usually by averaging hourly-data for each month. The model can be described as:

$$\overline{x}_{c,m,h} = \frac{\Sigma_{d=d_0}^{d_f} x'_{c,d,h}}{d_m} \tag{1}$$

where:

$\overline{x}_{c,m,h}$: average value of the attribute c for each month m at each hour h.

$x'_{c,d,h}$: value of the element in the time series x' corresponding to the attribute c in day d and hour h for each month m.

d_0: starting day of month m.

d_f: final day of month m.

d_m: number of days in each month m, hence, it is the weight of the respective representative day.

An advantage of this method is that the typical days obtained have a clear order. Nonetheless, the aggregation is based on the original sequence of the days and not on the similarity between days [10].

2.1.2 k-Medoids method

This method aims to group the days of the year into clusters so that the cluster members are as similar as possible. The cluster is then represented by a single day, which is the medoid in this case. Fig. 1 shows the graphic representation of this method.
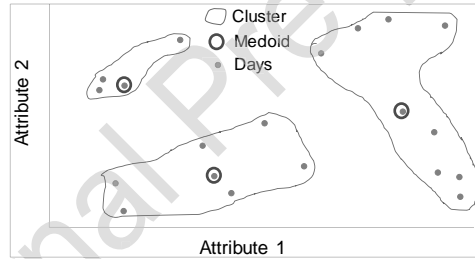

Fig. 1. Graphic representation of the *k*-Medoids method

When different attributes with different scales are taken into account, the input time series must be normalized, so it is evaluated on the same scale [10].

$$x_{c,t} = \frac{x'_{c,t} - \min x'_c}{\max x'_c - \min x'_c}, x_{c,t} \in [0,1] \tag{2}$$

The *k*-Medoids method for the selection of representative days is based on the optimal plant location problem. In the location problem, *k* plants must be located in *n* cities in such a way as to minimize the total distance from the plants to the *n* cities they supply. In the *k*-Medoids method, the location of a plant is interpreted as the selection of a medoid, and the distance between each city and the nearest plant is interpreted as the dissimilarity *dis* between an object and the representative object of the cluster to which it belongs [14]. Therefore, the first step is to define the matrix Ψ that contains all the attributes, in which the number of columns is defined by the product of the number of time steps N_h (usually 24 hours for each attribute) and number of attributes N_c, and the number of rows corresponds to the number of periods N_i (usually 365 days).

$$\Psi = \begin{bmatrix} x_{1,1,1} & \cdots & x_{1,N_h,1} & x_{2,1,1} & \cdots & x_{N_c,N_h,1} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_{1,1,N_i} & \cdots & x_{1,N_h,N_i} & x_{2,1,N_i} & \cdots & x_{N_c,N_h,N_i} \end{bmatrix} \tag{3}$$

The second step is to define the dissimilarity matrix D, which is composed of the distance (dissimilarity) between two elements p and q of the matrix Ψ, and every element is calculated by the Euclidean distance:

$$dis(p,q) = \left(\sum_{h=1}^{N_h}|\Psi_{p,h} - \Psi_{q,h}|^2\right)^{1/2} \tag{4}$$

In this way, the matrix D is defined as:

$$D = \begin{bmatrix} 0 & dis(1,2) & \cdots & dis(1,N_i) \\ \vdots & 0 & & \vdots \\ sim & \cdots & & 0 \end{bmatrix} \tag{5}$$

The MILP model for the k-Medoids method is:

$$\textbf{Minimize } \sum_{i=1}^{N_i}\sum_{j=1}^{N_i}dis(i,j) \cdot z_{i,j} \tag{6}$$

Subject to:

$$\sum_{i=1}^{N_i}z_{i,j} = 1, \forall \; j \in 1,2,..,N_i \tag{7}$$

$$z_{i,j} \leq u_i, \; \forall \; i \in 1,2,..,N_i \tag{8}$$

$$\sum_{i=0}^{N_i}u_i = N_k \tag{9}$$

$$u_i, z_{i,j} \in \{0,1\} \tag{10}$$

$z_{i,j}$ is a binary variable equal to 1 if and only if object j is assigned to the cluster of which i is the representative object (medoid). The number of clusters N_k is defined by the user. The representative day is selected when $u_i = 1$, and its weight ω_i corresponds to the number of days in each cluster k calculated by:

$$\omega_i = \sum_{j=1}^{N_i}z_{i,j} \tag{11}$$

$$\sum_{i=1}^{N_i}\omega_i = N_i \tag{12}$$

The reproduction of the series from representative days does not necessarily preserve the original days, and, therefore, a scale factor is calculated *a posteriori*.

$$\mu_i^c = \frac{\sum_{j=1}^{N_i}\sum_{g=1}^{N_g}x_{j,g}^c}{\omega_i\sum_{g=1}^{N_g}x_{i,g}^c} \tag{13}$$

The scale factor is applied considering that the extreme values of the original time series are not exceeded.

$$x_{scaled,i,g}^c = min[\mu_i^c \cdot x_{i,g}^c, \max(X^c)], x_{i,g}^c \in X^c \tag{14}$$

The aforementioned methods are types of aggregation methods that lead to smooth representative days, which underestimate the variability of the original time series [10].

2.1.3 OPT method

This method was proposed by K. Poncelet [11] and consists of fitting the data duration curve obtained from representative periods (DC_{rep}) to the duration curve of the original time series (DC). The procedure for selecting the representative days is: i) Similarly to the k-Medoids method, input time series must be normalized to be evaluated on the same scale; ii) Normalized duration curves (NDC) are computed from the normalized time series; iii) Normalized duration curves are divided in b bins in the ordinate, in our case a set of 10 bins of length 0.1, each corresponding to an interval s. A parameter Λ that expresses if data from the original time series c in day i belong to a specific interval, is defined, and iv) each interval is approximated by the expression $\sum_{i=1}^{N_i}\frac{\omega_i}{N_i} \cdot \Lambda_{c,s,i}$, where ω_i is the weight of the representative day i (in other words, ω_i is the number of times that the day i must be repeated to approach the original NDC); N_i is 365 in the case of representative days for a year (Fig. 2).
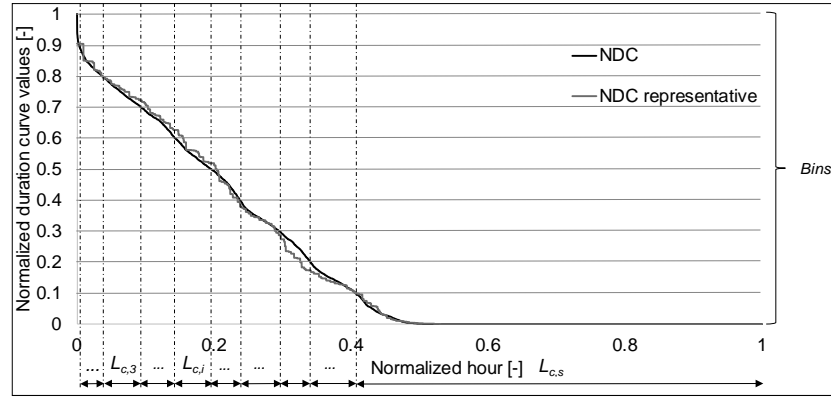
Fig. 2. Graphical description of OPT method

The difference between the length L_s and $\sum_{i=1}^{N_i} \frac{\omega_i}{N_i} \cdot \Lambda_{c,s,i}$, in each interval s is taken as an error metric ($error_{c,s}$). The optimization model minimizes the sum of the error terms, for all attributes c considered in every interval s, by selecting a single set of representative periods with their corresponding weights ω_i. The MILP model for this method is written as:

$$\textbf{\textit{Minimize}} \sum_c \sum_s error_{c,s} \tag{15}$$

Subject to:

$$error_{c,s} = \left| L_{c,s} - \sum_{i=1}^{N_i} \sum_{h=1}^{N_g} \frac{\omega_i}{N_i} \cdot \Lambda_{c,s,i,h} \right| \tag{16}$$

$$\sum_{i=1}^{N_i} u_i = N_k \tag{17}$$

$$\omega_i \leq u_i \cdot N_i \tag{18}$$

$$\sum_{i=1}^{N_i} \omega_i = N_i \tag{19}$$

$$u_i \in \{0,1\}; \; \omega_i \in \mathbb{R}_0^+ \tag{20}$$

One of the drawbacks of the OPT method is that the global optimum is challenging to reach. The work of Poncelet [11] managed three attributes (photovoltaic production, wind production, and electricity load) to select two representative days. The optimization process was stopped after six hours, obtaining good results. However, when dealing with six or more attributes it is more difficult to reach good results in a reasonable time, especially to preserve the area under the curve of each attributes. A scale factor, similar to the procedure described for k–Medoids method, has been applied to address this issue. All runnings that applied the OPT method (results are presented in section 3), were stopped after six hours as well.

2.1.4 Mix kM-OPT method

As already explained, aggregation methods lead to smoothed typical periods that underestimate the variability of the original time series [10], and the OPT method has the disadvantage that the global optimum is difficult to reach. Therefore, a combination of the k-Medoids and OPT methods is proposed to tackle these downsides, reducing the smoothing of typical periods and reaching the global optimum, improving the optimization results for optimization of polygeneration. The new procedure is described as follows:

The k-Medoids and OPT models are combined in a single model Mix kM-OPT. The function to be minimized in the OPT method is converted into a variable $errorOPT$, defined as:

$$errorOPT = \sum_c \sum_s error_{c,s} \tag{21}$$

The model is optimized as k-Medoid method. This means minimizing the distance of each component of the cluster with respect to the medoid. Simultaneously, $errorOPT$ is calculated, obtaining an $errorkM$-

OPT value, which indicates how close the representative duration curves are to the original curves, by applying the original *k*-Medoid method.

The lower the *errorOPT* value is, the closer the duration curves from representative days are to the original curves. Therefore, the model can be optimized again, with a new restriction:

$$errorOPT \leq LimOPT \tag{22}$$

Where *LimOPT* is a value defined by the user, lower than *errorkM-OPT*. The lower the *LimOPT*, the global optimum becomes more difficult to reach. A trade-off value must be found to improve the original models without entailing in excessive computational time. Herein, *LimOPT* is approximately 90% of *errorOPT*.

2.1.5 Metric for the evaluation of representativeness

The accuracy of using representative days can be evaluated by defining a metric to estimate how close the set of representative periods are with respect to the reference case, in which hourly (8760 periods) entire year data were considered. In this work, the Root-Mean-Square Error (RMSE) was selected to evaluate the set of representative days obtained from different methods. RMSE expresses the similarity of the distribution of values and their frequency of occurrence for every representative duration curve with respect to the original ones. The lower the RMSE, there is a better fit with the original representative duration curve.

$$RMSE_c = \frac{\sqrt{\frac{1}{8760} \cdot \Sigma_{h=1}^{8760}(DC_{c,h} - DCrep_{c,h})^2}}{\max{(DC_c)} - \min{(DC_c)}} \tag{23}$$

2.2 Case studies

A multifamily building composed of 40 dwellings, 102.4 m^2 each with average occupancy of three people per dwelling, is studied for two locations in Spain with different natural resources and weather conditions: Zaragoza and Gran Canaria. In both locations, two different configurations grid-connected and standalone systems have been studied. Energy demands and renewable energy production are the attributes considered for the optimization of the polygeneration system. In the grid-connected system, the CO_2 emissions from the grid are considered as an additional attribute to estimate the environmental impact associated with the purchase of electricity from the grid.

2.2.1 Energy demands

Space heating and cooling demands are estimated from annual demand data [15], daily demand values are estimated from the degree-days method, and hourly values are obtained by applying hourly profiles [16]. Regarding the application of the degree-days method, base temperatures for space heating and cooling were set to 15ºC and 21ºC, respectively. Domestic Hot Water (DHW) is calculated considering a reference temperature of 60ºC and the mean monthly temperature of the net water [17]; a monthly consumption factor was applied to obtain monthly distribution of DHW demands [18]. It is assumed that every day of each month presents the same consumption and a hourly profile [16] is applied to obtain the hourly DHW demand. In the case of electricity, the annual electricity demand for appliances [19] is distributed by applying a monthly distribution factor, which is divided by the days of the month and distributed following an hourly distribution function [20], which considers different hourly consumption for each season. The procedures briefly described above provide the hourly demand data series of heating Q_d, cooling R_d and electricity E_d, where the heating demand consists of space heating plus domestic hot water. In Gran Canaria there is no space heating demand, and therefore the heating demand corresponds only to DHW.

2.2.2 Renewable energy production

The hourly photovoltaic energy production per square meter, E_{PV}, is calculated following the procedure described by Duffie and Beckman [21], as a function of solar radiation [22] on the PV module [23], tilted at 36º for Zaragoza and 26º for Gran Canaria, with azimuth angle 0º.

The hourly solar thermal energy production per square meter, E_{ST}, is calculated as a function of the solar radiation over a tilted surface (36º for Zaragoza and 26º for Gran Canaria) with azimuth angle 0º [22], mean difference temperature between the collector temperature and ambient temperature, and parameters of the collector (optical efficiency η_o, 1st and 2nd order heat loss coefficients a_1 and a_2) [24].

The electrical production of a wind turbine E_W, is calculated based on the production curve of the 30 kW nominal capacity turbine [25], as a function of the wind speed [22]. Terrain surface roughness, hub turbine height, and density correction factor were also considered [26].

2.2.3 CO_2 emissions from the grid

For the grid-connected system, hourly operational CO_2 emissions from the grid were collected for the entire year 2018 [27]. These data are available on a time scale of 10 minutes; however, as the analysis is carried out on an hourly basis, average hourly values were utilized herein.

Tables 1 and 2 present a summary of the attributes considered in each location, with respective annual and peak values.

Table 1. Annual and peak values for each attribute in Zaragoza

Attribute	Annual Value		Peak Value	
Heating demand (Q_d)	166298	kWht	218.8	kWt
Cooling demand (R_d)	46694	kWht	234.3	kWt
Electricity demand (E_d)	117555	kWh	24	kW
Photovoltaic production (E_{PV})	285	kWh/m²	0.16	kW/m²
Wind energy production (E_W)	53991	kWh/ud	39	kW/ud
Solar Thermal Production (E_{ST})	995	kWht/m²	0.79	kWt/m²
CO_2 emissions	1815	kgCO₂eq/kWh	0.38	kgCO₂eq/kWh

Table 2. Annual and peak value for each attribute in Gran Canaria

Attribute	Annual Value		Peak Value	
Heating demand (Q_d)	14336	kWht	18.4	kWt
Cooling demand (R_d)	45466	kWht	183.1	kWt
Electricity demand (E_d)	117555	kWh	24.0	kW
Photovoltaic production (E_{PV})	313	kWh/m²	0.1	kW/m²
Wind energy production (E_W)	126677	kWh/ud	39.9	kW/ud
Solar Thermal Production (E_{ST})	1177	kWht/m²	0.73	kWt/m²
CO_2 emissions	5627	kgCO₂eq/kWh	0.8	kgCO₂eq/kWh

2.2.4 Polygeneration system

The superstructure depicted in Fig. 3, considers the candidate technologies and the available energy conversions route. For the grid-connected systems, the model enables the selection of the optimal contracted power from the grid. A natural gas network is available in Zaragoza but not in Gran Canaria, which employs gasoil instead. Candidate technologies can be divided into three groups: i) renewable energy; ii) technologies to produce electricity, heating, or cooling, and iii) energy storage.

Renewable energy consists of photovoltaic modules PV, wind turbine WT, and solar thermal collectors ST. Regarding electricity, heating, or cooling technologies, these are technologies that produce electricity from fossil fuels, such as an electric generator GE or a cogeneration module CM, which produces heat as well. There are also, heating and cooling technologies, such as a conventional boiler GB that consumes fossil fuel to produce heat, a single-effect absorption chiller ACH that uses heat and a small quantity of electricity to produce cooling, and a reversible heat pump HP that converts electricity into thermal energy (heating or cooling). Regarding energy storage, there is thermal energy storage for

heating TSQ and cooling TSR, and Ion lithium (Ion-Li) batteries Bat to store electricity. Equipment can modulate up to nominal load, except for non-manageable loads (PV, ST and WT) and batteries. The reversible heat pump operates in heating mode assuming a constant coefficient of performance COP, and in cooling mode assuming a constant Energy Efficiency Ratio EER, with constant cooling/heating capacity ratio β. The performance of the conventional boiler is considered constant. Concerning thermal energy storage tanks, the energy stored is calculated in each time step taking into account the energy losses by applying a constant factor λ. In the case of Ion-li batteries, the round trip efficiency η_{rt}, determines the energy losses during the charging and discharging process in each time step. A maximum depth of discharge DOD is defined for batteries to avoid premature failures. During battery lifetime operation, the number of charge-discharge cycles must be lower than the maximum number of cycles that cause the failure $N_{c,failure}$ [28]. Battery capacity models are applied to calculate their dynamic
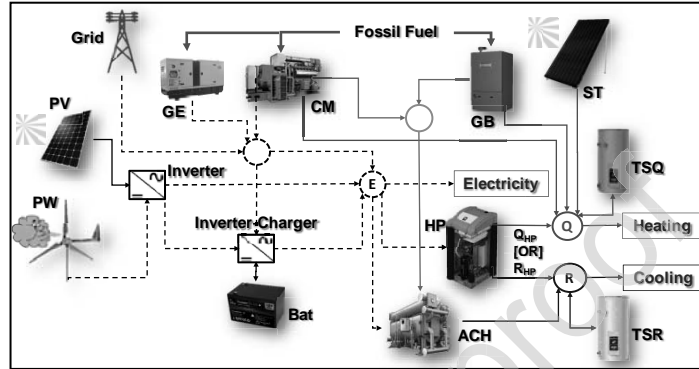

Fig. 3. Superstructure which consider candidate technologies for the polygeneration system. Nodes are represented by circles.

behavior [29].

2.2.5 Optimization model

Grid-connected and standalone systems are optimized by using different sets of representative days (obtained from aforementioned methods). The defined objective function is to minimize the total annual cost:

$$Total\ annual\ cost = CIA + C_{op} \qquad (24)$$

$$CIA = (1 + VAT) \cdot (1 + F_{ind}) \cdot CRF \cdot \sum_{j = component} Cu_j \cdot Cap_j \cdot (1 + FNPV_j)(1 + F_{mj}) \qquad (25)$$

$$C_{op} = C_E + C_g \qquad (26)$$

which is composed of the investment annual cost, *CIA*, of the equipment and the operational annual cost, C_{op}. The investment cost of each component is calculated from the unit cost Cu and the installed capacity Cap. Installation and maintenance costs are considered by applying the factor F_m. Calculation of the fixed annual cost considers a Capital Recovery Factor *CRF*=0.082 yr^{-1}, which is applied based on a lifetime of 20 years for the installation, and an interest rate of 5%. However, components can present different lifetimes n_r, hence, a net present value factor *FNPV* is calculated for each component to consider the substitutions carried out throughout the lifetime of the installation. The indirect costs encompass the application of a factor $F_{ind} = 0.2$. The *VAT* (Value-Added Tax) is applied to each component. The operational annual cost is the sum of electricity costs C_E (only for the grid-connected system) and the fuel consumption cost C_g. Electricity costs are a function of the contracted power from the grid (there are three possible values of contracted power to select from the grid, Pct$_{1,2,3}$, which depends on the tariff in use) and electricity consumption, plus different taxes and fixed costs. In the grid-connected system, C_g is the cost associated with the natural gas (composed of a fixed part related to the annual natural gas consumption and a variable part proportional to the natural gas consumption), plus *VAT* and meter equipment rental costs. For standalone systems, C_g only refers to gasoil

consumption plus *VAT*.

The objective function is subject to capacity and efficiency limits, production restrictions, and balance equations, among others. Furthermore, the environmental cost constituted of fixed CO_2 emissions (embodied in the equipment) and operational CO_2 emissions (associated with fuel combustion and electricity from the grid) are also calculated. Appendix A presents the detailed description of the optimization model and Appendix B presents technical, economic and environmental data.

3. Results and discussion

This work evaluates the suitability of using representative days obtained from different methods to optimize polygeneration systems. Unlike other works, which have studied, among other things, the feasible number of representative days for the optimization of polygeneration systems [10,14], herein a fixed number of representative days is considered to compare different methods for the selection of representative days.

Some considerations must be made to guarantee the suitability of the representative days obtained. Averaging method, usually one representative day for each month of the year is considered [2]. In the case of *k*-Medoids, eight to 12 representative days are sufficient and have provided good results in previous works [10,14]. In the case of the OPT method, previous works have selected two representative days and obtained good results [11], therefore, 12 representative days are considered herein (and sufficient to obtain adequate results as well). Two additional days corresponding to peak heating and cooling demands are considered to guarantee that energy demands are always met, with the corresponding impact on the annual investment cost, but not on the annual operation cost.

After application of these methods for the selection of representative days, the polygeneration system is optimized. Optimal results of design and configuration obtained with different sets of representative days are compared with the reference case, which considers entire year data. The Quality Function Deployment (QFD) methodology is applied to evaluate the suitability of each method.

3.1 Set of representative days, Metrics and Duration Curves

Tables 3 and 4 present the set of representative days obtained from each method, along with the respective weights for Zaragoza and Gran Canaria. As aforementioned, one representative day for each month was obtained from the averaging method, and different days were obtained from the other methods, which do not necessarily match each month.

Table 3. Set of representative days *i* obtained from each method, with respective weights for grid-connected and standalone systems in Zaragoza

		Grid-Connected						**Standalone**					
		k-Medoids		**OPT**		**Mix kM-OPT**		**k-Medoids**		**OPT**		**Mix kM-OPT**	
Item		*i*	*ω*	*i*	*ω*	*i*	*ω*	*i*	*ω*	*i*	*ω*	*i*	*ω*
1		37	38	1	39.0	37	36	9	23	1	18.1	21	6
2		62	25	12	4.4	62	27	31	34	19	32.9	37	47
3		112	50	32	35.6	112	44	38	51	25	36.6	116	21
4		116	22	48	29.7	116	24	116	21	58	33.9	136	16
5		175	33	161	19.4	175	37	136	17	123	32.7	147	51
6		208	24	167	56.1	220	22	141	33	166	42.7	158	62
7		221	11	170	2.2	221	14	147	37	182	17.5	166	15
8		241	58	179	11.6	241	50	165	70	186	44.5	175	34
9		287	22	262	32.7	287	25	166	15	231	10.3	240	15
10		291	34	279	67.6	291	37	240	15	290	54.8	300	41
11		339	36	332	54.4	339	35	276	33	310	31.8	339	48
12		352	12	360	12.2	352	14	346	16	362	9.2	352	9

Table 4. Set of representative days *i* obtained from each method with respective weights for grid-connected and standalone systems in Gran Canaria

	Grid-connected						Standalone					
	k-Medoids		OPT		Mix *k*M-OPT		*k*-Medoids		OPT		Mix *k*M-OPT	
Item	*i*	*ω*	*i*	*ω*	*i*	*ω*	*i*	*ω*	*i*	*ω*	*i*	*ω*
1	40	24	31	28.5	40	24	32	21	1	43.1	32	21
2	52	21	45	63.9	52	21	122	22	65	41.5	122	23
3	58	26	67	0.0	58	26	124	36	70	0.0	138	17
4	120	26	89	4.4	120	26	138	17	85	0.2	147	25
5	127	44	121	12.8	127	44	147	25	100	15.7	149	24
6	138	30	122	58.1	138	24	149	23	102	25.0	176	41
7	186	23	152	35.0	186	29	176	42	173	48.0	189	31
8	205	29	219	33.5	226	29	189	31	197	43.0	190	31
9	264	38	227	53.6	264	38	190	31	229	31.0	266	30
10	266	29	289	51.1	266	29	266	30	288	24.4	326	34
11	281	28	362	20.6	281	28	351	28	319	56.8	351	29
12	360	47	364	3.6	360	47	359	59	352	36.2	359	59

Based on the representative days obtained, the corresponding duration curves for each attribute were built, followed by the calculation of the respective RMSE and *ErrorOPT* values.

Tables 5 and 6 present the RMSE values for each attribute and the *ErrorOPT* values obtained for each method. The RMSE values obtained from the Averaging method are higher than those obtained from other methods. The wind energy production attribute, E_W, presented the highest RMSE value and therefore, the Averaging method presents the weakest approach to the original values. The lowest RMSE values, in general, were obtained from the OPT method. *ErrorOPT* represents how close the representative duration curves are to the original ones, and RMSE expresses the similarity between the distribution of values and their frequency of occurrence for every representative duration curve with respect to the original curves. The values obtained from the Mix *k*M-OPT method are usually intermediate values between those obtained from the *k*-Medoids and OPT methods. This does not occur necessarily for all attributes, because *ErrorOPT* is an absolute value that considers all attributes simultaneously (fitting consider the entire set and not individual values).

Table 5. *ErrorOPT* values obtained for each method and RMSE values for each attribute corresponding to the different methods for grid-connected and standalone systems in Zaragoza.

	Grid-Connected							
	RMSE							*ErrorOPT*
Method	Qd	Rd	Ed	E_{PV}	E_W	E_{ST}	CO_2	
Averaging	3.2%	2.3%	0.0%	3.7%	20.6%	7.2%	7.2%	-
k-Medoids	2.8%	1.8%	1.0%	1.9%	3.7%	1.6%	4.4%	0.83
OPT	1.6%	1.6%	0.6%	1.5%	1.6%	1.6%	1.6%	0.36
Mix *k*M-OPT	2.7%	1.5%	0.9%	1.6%	3.7%	1.1%	4.2%	0.74
	Standalone							
	RMSE							*ErrorOPT*
Method	Qd	Rd	Ed	E_{PV}	E_W	E_{ST}	CO_2	
Averaging	3.2%	2.3%	0.0%	3.7%	20.6%	7.2%	-	-
k-Medoids	1.6%	1.7%	1.2%	2.1%	2.9%	1.7%	-	0.55
OPT	1.3%	1.5%	1.1%	1.0%	1.7%	1.0%	-	0.16
Mix *k*M-OPT	2.7%	1.6%	1.1%	2.1%	2.5%	1.6%	-	0.50

10

Table 6. *ErrorOPT* values obtained for each method and RMSE values for each attribute corresponding to the different methods for grid-connected and standalone systems in Gran Canaria.

Grid Connected								
	RMSE							*ErrorOPT*
Method	Qd	Rd	Ed	E_{PV}	E_W	E_{ST}	CO_2	
Averaging	0.0%	1.6%	0.0%	3.0%	16.0%	8.9%	8.9%	-
k-Medoids	0.9%	1.3%	0.5%	1.6%	1.7%	2.0%	2.9%	0.62
OPT	1.3%	2.1%	0.6%	2.0%	2.0%	2.2%	1.3%	0.29
Mix *k*M-OPT	0.8%	1.4%	0.7%	1.6%	1.6%	2.1%	2.7%	0.56
Standalone								
	RMSE							*ErrorOPT*
Method	Qd	Rd	Ed	E_{PV}	E_W	E_{ST}	CO_2	
Averaging	0.0%	1.6%	0.00%	3.0%	16.0%	8.9%	-	-
k-Medoids	1.0%	1.6%	1.2%	2.0%	2.1%	2.1%	-	0.45
OPT	0.6%	1.4%	0.0%	1.7%	1.3%	2.4%	-	0.16
Mix *k*M-OPT	1.0%	1.6%	1.2%	1.5%	2.1%	1.6%	-	0.41

Fig. 4 and Fig. 5 depict some duration curves that aid in the visualization of the accuracy of each method to reproduce the reference DC (original duration curve). In agreement with the values presented in the previous Tables, it is observed that E_W and CO_2 emissions are not well reproduced by the Averaging method because these attributes present the highest variability.

Fig. 4. Comparison of the original duration curve (Reference) and the duration curves obtained from the analyzed methods for the selection of representative days for Q_d, E_{PV}, E_W and CO_2 emissions attributes in Zaragoza.
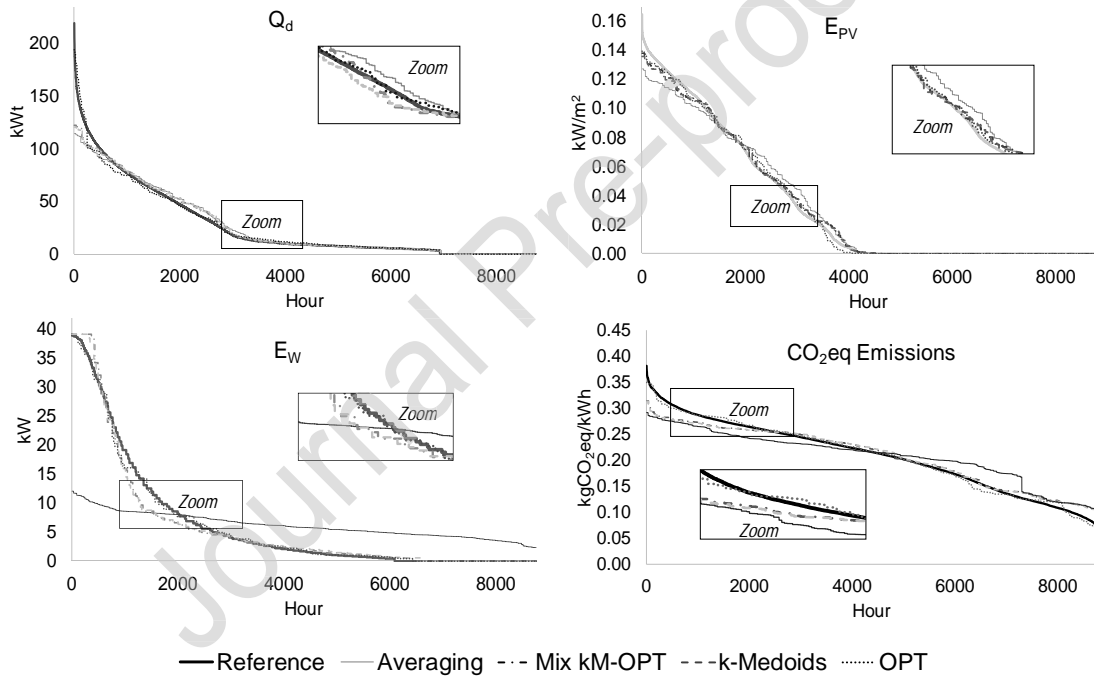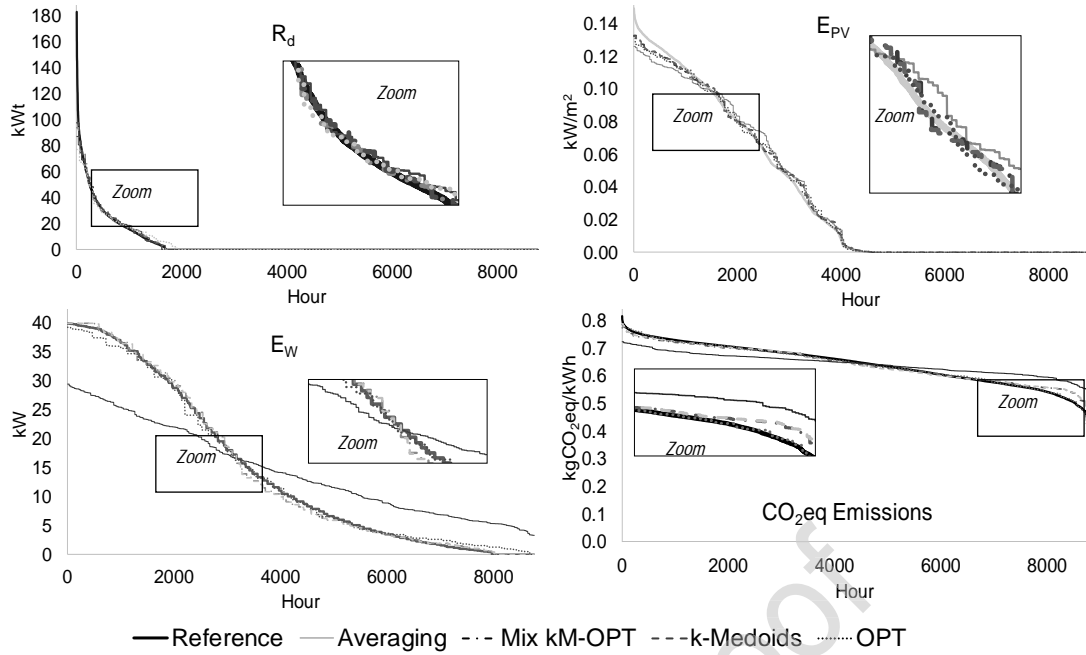
11

Fig. 5. Comparison of the original duration curve (Reference) and the duration curves obtained from the analyzed methods for the selection of representative days for R_d, E_{PV}, E_W and CO_2 emissions attributes in Gran Canaria.

3.2 Optimization of polygeneration systems

The optimization of the polygeneration systems minimized the total annual cost and used different sets of representative days (obtained from different methods). These results were compared to a reference case that considered hourly data throughout the entire year (8760 hours). Optimization was carried out for grid-connected and standalone systems, yielding an optimal configuration along with its optimal operation. Herein, synthesis refers to the selection of technologies to install, design is about of sizing of the installed components and operation refers to electricity/fuel consumption.

Table 7 shows the optimization results obtained for the grid-connected polygeneration system located in Zaragoza. Regarding the synthesis, the same technologies (PV, CM, HP, TSQ and TSR) are present in the optimal solutions obtained with different sets of representative days. The value of the contracted power from the grid is different when the k-Medoids set of representative days was applied. The OPT set performed better as only the TSR sizing error was higher than 5% (but still lower than 20%). The other sets of representative days presented sizing errors under 20% except for the TSR obtained with the k-Medoids set.

Regarding electricity consumption from the grid, error is below 10% except for the k-Medoids set, which was approximately 14%. The error associated with natural gas consumption was under 2% for all sets of representative days. The errors associated with the total annual cost and CO_2 emissions were below 2% for all sets of representative days.

Table 7. Results of the optimization of the grid-connected polygeneration system in Zaragoza, corresponding to the sets of representative days shown in Table 3. The reference case considers 365 days.

Technology	Reference	Averaging	k-Medoids	OPT	Mix kM-OPT
Pct [kW]	$34.6_{1,2,3}$	$34.6_{1,2,3}$	$31.2_{1,2,3}$	$34.6_{1,2,3}$	$34.6_{1,2,3}$
CM [kWe]	15.2	14.5	15.7	15.7	15.4
PV [kW]	27.4	29.6	30.3	27.6	29.9
WT [kWe]	0	0	0	0	0
ST [m²]	0	0	0	0	0
HP [kWt]	231.3	232.8	224.1	233.7	237.5

12

GB [kWt]	61.4	65.1	69.6	59.2	60.7
ACH [kWt]	0	0	0	0	0
TSQ [kWht]	2.3	2.3	2.3	2.3	2.3
TSR [kWht]	32.7	30.0	46.4	28.3	21.2
Bat [kWh]	0	0	0	0	0
Electricity [kWh]	36630	34028	31649	34133	33076
Natural gas CM [kWh]	250287	241932	248526	258462	247067
Natural Gas GB [kWh]	57611	60544	55213	50943	56352
Total cost [€/yr]	60086	59523	59599	59781	59593
CO_2 emissions [kgCO_2eq/yr]	74969	73449	73404	74973	73527

Table 8 shows the results of the standalone polygeneration system located in Zaragoza. In this case, the set of representative days obtained from the Averaging, k-Medoids, and OPT methods resulted in the same optimal configuration composed of CM, PV, WT, HP, GB, TSQ, TSR, and Bat. ST and ACH were not included. The utilization of the representative days from the Mix kM-OPT method results in the previous configuration but includes ACH (although it could be neglected).

Regarding design, PV and WT are not well-sized by using the Averaging set of representative days. The sizing errors obtained for PV and WT are approximately 6% and 33% for the k-Medoids set, 11% and 20% for the OPT set, and 6% and 20% for the Mix kM-OPT set. HP and GB sizing errors are below 10% for all sets of representative days (for the Mix kM-OPT set it was under 5%).

Concerning energy storage, TSQ and Bat are barely considered (and therefore could be disconsidered). TSR presented sizing errors below 20% for all sets of representative days. For gasoil consumption, the error was approximately 29% for Averaging, 2% for k-Medoids, 5% for OPT and 1% for Mix kM-OPT set.

In terms of total annual cost, all sets of representative days presented errors of approximately 1% except for the Averaging set, which presented an error of about 4%. For the CO_2 emissions, the Averaging set presented an error of approximately 27%, k-Medoids about 2%, OPT about 4% and Mix kM-OPT approximately 0.2%.

Table 8. Results of the optimization of the standalone polygeneration system in Zaragoza corresponding to the sets of representative days shown in Table 3. The reference case considers 365 days.

	Reference	Averaging	k-Medoids	OPT	Mix kM-OPT
CM [kWe]	23.3	18.9	23	23	23.3
PV [kW]	43	25.1	40.3	38.2	45.5
WT [kWe]	8.6	43.7	5.7	6.9	6.7
ST [m²]	0	0	0	0	0
HP [kWt]	167.4	177.4	155.4	153.2	168
GB [kWt]	120.8	126.4	122.6	122.4	125.5
ACH [kWt]	0	0	0	0	1.4
TSQ [kWht]	2.9	1.8	2.8	1.1	5.8
TSR [kWht]	153.2	134.4	175.8	180	170.9
Bat [kWh]	1.3	16.5	1.9	2.3	1.5
Gasoil CM [kWh]	230721	91380	233120	234681	227990
Gasoil GB [kWh]	139568	169887	144578	154499	140700
Total cost [€/yr]	81243	77743	80304	81137	82157
CO_2 emissions [kgCO_2eq/yr]	115160	84130	117031	120326	114913

Results of the optimization for Gran Canaria are presented in Tables 9 and 10 for grid-connected and standalone systems, respectively. For the grid-connected system, all sets of representative days provided the same optimal configuration, composed of PV, WT, HP, GB, and TSR. Unlike the reference system, each set included ST in the optimal configuration. Regarding contracted power, only the OPT set

resulted in the same Pct as the reference case. In terms of design, the Averaging set resulted in sizing errors are much higher than 5%, for most technologies, especially for renewable energy based. For the *k*-Medoids, OPT and *k*M-OPT sets, the sizing errors obtained were below or approximately 20% for all technologies. Despite the high sizing errors obtained from Averaging, the total annual cost error was only about 2%. This result confirms that, similarly to Zaragoza, the total annual cost error is not sufficient to evaluate the suitability of a method. For the other analyzed sets, the total annual cost error was about 1%.

For electricity from the grid, the obtained error was below 5% for all sets of representative days, except for the Averaging set, which presented an error above 20%. The gasoil consumption errors for the Averaging, *k*-Medoids, OPT, and Mix *k*M-OPT sets were approximately 18%, 23%, 7% and 16% respectively. In terms of total CO_2 emissions errors, for Averaging the error was approximately -21%, for OPT approximately -3% and for *k*-Medoids and *k*M-OPT, about 1%.

Table 9. Optimization results for grid-connected polygeneration system in Gran Canaria, corresponding to the sets of representative days shown in Table 4. The reference case considers 365 days.

	Reference	Averaging	*k*-Medoids	OPT	Mix *k*M-OPT
Pct [kW]	$31.2_{1,2}$-17.3_3	$34.6_{1,2}$-17.3_3	$34.6_{1,2}$-17.3_3	$31.2_{1,2}$-17.3_3	$34.6_{1,2}$-17.3_3
GE[kWe]	0	0	0	0	0
PV [kW]	23.0	14.7	19.1	24.6	21.8
WT [kWe]	8.4	16.2	8.6	7.2	7.8
ST [m²]	0	7	8	7	6
HP [kWt]	158.5	165.6	165.3	159.8	169.1
GB [kWt]	15.5	13.8	12.8	14.1	12.9
ACH [kWt]	0	0	0	0	0
TSQ [kWht]	0	0	0	0	0
TSR [kWht]	61.4	48.0	48.6	59.1	53.4
Bat [kWh]	0	0	0	0	0
Electricity [kWh]	68361	51843	71557	66270	70587
Gasoil GB [kWh]	23458	19276	17950	21865	19734
Total cost [€/yr]	33020	32298	32835	32837	33372
CO_2 emissions [kgCO₂eq/yr]	56030	44128	56652	54412	56629

For the standalone system in Gran Canaria, the set of representative days obtained from the Averaging, OPT and *k*M-OPT methods result in the same optimal configuration, which included GE, PV, WT, HP, GB, TSQ and TSR. A similar configuration was obtained by using the *k*-Medoids set of representative days, but TSQ was not included. Regarding design, the TSQ included was so insignificant it could be disconsidered.

When using the Averaging set of representative days, the technology sizing errors were much higher than 5%, especially for WT technology, similar to the result obtained for the grid-connected system. Using the other sets of representative days, sizing errors obtained were below 20% for all technologies, except for the OPT set where the sizing error for WT was about 24%. For gasoil, the errors were below 2%, except for the Averaging set, which an error of approximately 49%. In terms of total annual cost, the obtained error was approximately 1% except for the Averaging set, which presented an error of about 15%. In terms of CO_2 emissions, the errors obtained were approximately 1%, except for the Averaging set, with an error of approximately -45%.

Table 10. Optimization results for the standalone polygeneration system in Gran Canaria, corresponding to the sets of representative days shown in Table 4. The reference case considers 365 days.

	Reference	Averaging	*k*-Medoids	OPT	Mix *k*M-OPT
GE[kWe]	25.6	25.8	25.5	25.3	25.5
PV [kW]	30.5	14.0	26.9	24.4	27.1
WT [kWe]	18.7	38.2	19.9	23.2	20.0
ST [m²]	0	0	0	0	0

HP [kWt]	158.5	163.2	164.5	164.5	165.1
GB [kWt]	13.9	12.7	15.4	13.1	14.5
ACH [kWt]	0	0	0	0	0
TSQ [kWht]	1.6	2.9	0.0	2.3	0.9
TSR [kWht]	61.4	52.7	50.2	50.2	49.1
Bat [kWh]	0	0	0	0	0
Gasoil GE [kWh]	143270	54827	144171	147134	144685
Gasoil GB [kWh]	45306	42070	46231	44646	46037
Total cost [€/yr]	48166	41086	47861	48783	47985
CO_2 emissions [kgCO_2eq/yr]	60643	33601	60954	61371	61070

3.3 Qualitative evaluation of methods for the selection of representative days

The QFD technique [30] was applied to evaluate the suitability of each method for the synthesis and design of polygeneration systems, estimating which method is useful for each case. To this end, the criteria shown in Table 11 (based on the engineering criteria practice) were applied. A score is applied to the sizing errors E_{sz} of each component. Capacities with a deviation above 20% are not considered acceptable and, therefore receive a score of 0. A deviation between 5% and 20% is considered reasonably accurate, receiving a score of 1, and a deviation below 5% is considered good accuracy and receives a score of 2. The maximum score that can be reached for each group of technologies is 6 for renewable energy technologies, 8 for electricity/heating/cooling production technologies, and 6 for energy storage technologies.

Table 11. Criteria for the evaluation of methods for the selection of representative days by QFD.

Sizing Error (E_{sz})	Score
$E_{sz} > \lvert 20\% \rvert$	0
$\lvert 5\% \rvert < E_{sz} <= \lvert 20\% \rvert$	1
$E_{sz} <= \lvert 5\% \rvert$	2
Group of technologies	**Max Score**
Renewable Energy (RE)	6
Electricity/Heating/Cooling Production Technologies (E/H/C)	8
Energy Storage (ES)	6
Polygeneration system (PS)	*20*

The QFD results for each system and location are presented in Tables 12-15. For grid-connected systems, the maximum score was 34, obtained from the OPT method, followed by Mix *k*M-OPT and *k*-Medoid methods, and for standalone systems, the maximum score was 26, obtained from Mix *k*M-OPT method followed by *k*-Medoids and OPT methods.

Table 12. QFD results for the grid-connected system in Zaragoza.

	Renewable Energy				E/H/C Production Technologies					Energy Storage				Polygeneration system
Method	**PV**	**WT**	**ST**	***Score RE***	**CM**	**HP**	**GB**	**ACH**	***Score E/H/C***	**TSQ**	**TSR**	**Bat**	***Score ES***	***Score***
Averaging	1	2	2	*5*	2	2	1	2	*7*	2	1	2	*5*	*17*
***k*-Medoids**	1	2	2	*5*	2	2	1	2	*7*	2	0	2	*4*	*16*
OPT	2	2	2	*6*	2	2	2	2	*8*	2	1	2	*5*	*19*
***k*M-OPT**	1	2	2	*5*	2	2	2	2	*8*	2	0	2	*4*	*17*

Table 13. QFD results for the standalone system in Zaragoza.

| Method | Renewable Energy | | | | E/H/C Production Technologies | | | | | Energy Storage | | | | Polygeneration system |
	PV	WT	ST	*Score RE*	CM	HP	GB	ACH	*Score E/H/C*	TSQ	TSR	Bat	*Score ES*	*Score*
Averaging	0	0	2	*2*	1	1	2	0	*4*	0	1	0	*1*	*7*
k-**Medoids**	1	0	2	*3*	2	1	2	0	*5*	2	1	0	*3*	*11*
OPT	1	1	2	*4*	2	1	2	0	*5*	0	1	0	*1*	*10*
*k***M-OPT**	1	0	2	*3*	2	2	2	0	*6*	0	1	1	*2*	*11*

Table 14. QFD results for the grid-connected system in Gran Canaria.

| Method | Renewable Energy | | | | E/H/C Production Technologies | | | | | Energy Storage | | | | Polygeneration system |
	PV	WT	ST	*Score RE*	GE	HP	GB	ACH	*Score E/H/C*	TSQ	TSR	Bat	*Score ES*	*Score*
Averaging	0	0	0	*0*	2	2	1	2	*7*	2	0	2	*4*	*11*
k-**Medoids**	1	2	0	*3*	2	2	1	2	*7*	2	0	2	*4*	*14*
OPT	1	1	0	*2*	2	2	1	2	*7*	2	2	2	*6*	*15*
*k***M-OPT**	2	1	0	*3*	2	1	1	2	*6*	2	1	2	*5*	*14*

Table 15. QFD results for the standalone system in Gran Canaria.

| Method | Renewable Energy | | | | E/H/C Production Technologies | | | | | Energy Storage | | | | Polygeneration system |
	PV	WT	ST	*Score RE*	GE	HP	GB	ACH	*Score E/H/C*	TSQ	TSR	Bat	*Score ES*	*Score*
Averaging	0	0	2	*2*	2	2	1	2	*7*	0	1	2	*3*	*12*
k-**Medoids**	1	1	2	*4*	2	2	1	2	*7*	0	1	2	*3*	*14*
OPT	1	0	2	*3*	2	2	1	2	*7*	0	1	2	*3*	*13*
*k***M-OPT**	1	1	2	*4*	2	2	2	2	*8*	0	1	2	*3*	*15*

3.4 Reduction of computational effort by using representative days

The purpose of using representative days instead of entire year data is to reduce the computational time required for the optimization of polygeneration systems. Table 16 presents data related to the number of variables, constraints, and elapsed time for the optimization of a grid-connected system using entire year data and using representative days. All runs were performed on an Intel Core i5-6200 CPU @ 2.3 GHz, with a memory of 8 GB and 64-bit system. Similar results were also obtained for standalone systems. It must be highlighted that the global optimum reach using 8760 hours was achieved because the optimization model did not consider many technical details such as partial load of equipment. Otherwise, calculation time could increase up to several days. However, these results demonstrated that there is a remarkable reduction in runtime, of about three orders of magnitude, which is significant when a detailed design is required.

Table 16. Number of variables and runtime for the optimization of polygeneration system using entire year data and sets of representative days.

| Number of days | Variables | | Constraints | Runtime | | |
	Integer	Total		Hours	Minutes	Seconds
365	21	928763	1270363	1	27	6
14	21	35824	48888	0	0	8

4. Conclusions

The performance of three different methods were compared for the selection of representative days for the optimization of polygeneration systems. Two different systems were considered - standalone and grid-connected located in two different locations in Spain. The locations presented different climatic

conditions, natural resources and energy demands, covering a wide range of possibilities that enable the extension of the results to other case studies.

In general terms, the use of representative days in the optimization of polygeneration systems was more accurate for grid-connected than standalone systems. The OPT method could be considered the best option for grid-connected systems. In the case of standalone systems, the methods yielded similar results. For instance, in Gran Canaria the Mix kM-OPT method performed best, but in Zaragoza the k-Medoids and Mix kM-OPT methods presented the same score.

The stochastic variability of the attributes has strong influence and must be taken into account when assessing the most suitable method for the selection of representative days. This was the case for wind turbine production. The Averaging method was not adequate to address with attributes with high stochastic variability, such as wind energy; however, it was a good alternative to tackle attributes such as solar energy.

It is important to highlight that the total annual cost was not a determinant factor when evaluating the suitability of a method for the selection of representative days. When the Averaging method was applied to grid-connected systems in Gran Canaria or standalone systems in Zaragoza, the annual cost errors of the optimal systems were below 2% with respect to the corresponding reference cases, but the optimal designs were significantly different than those obtained for the reference systems.

Regarding conventional technologies to produce electricity/heating/cooling, all methods performed well when sizing the components. In the case of energy storage, although there is no connection between the selected representative days to model its continuous dynamic behavior, the sizing errors observed were below 20% in most cases. Herein only thermal energy storage for cooling TSR was feasible in the optimal configurations, and the best results were obtained when applying the OPT method, achieving errors under 5% in Gran Canaria for grid-connected systems. When TSQ and Bat were part of the optimal configuration, sizes were negligible. The energy storage considered herein was a short-term storage.

From the point of view of operation, the electricity and fuel consumption obtained from the application of k-Medoids, OPT and Mix kM-OPT methods presented good fit with the reference system. The results obtained for standalone systems were remarkable, with fuel consumption errors below 5%.

In general, based on the obtained results, the OPT method seems to be the best alternative to select representative days. However, computational effort can be very high in the process to obtain the global optimum and, as a consequence, it is advisable to stop the optimization process before the global optimum is reached.

It has been demonstrated herein that the k-Medoids method results can be improved by combining the k-Medoids and the OPT methods. This leads, however, to increases in computational effort and cost. The importance of reducing RMSE metric values was highlighted, to improve the results of the optimization of polygeneration systems by using representative days. The lowest RMSE values were obtained with the OPT method, and it was observed that the best results for grid-connected systems were also obtained by applying this method. Nonetheless, when this method is applied, it is more difficult to reach the global optimum, which entails an uncertainty associated with the application of the OPT method. The kM-OPT method allows the achievement of a global optimum, which is an advantage with respect to the OPT method, but does not necessarily mean that results were improved.

As a final global conclusion, the synthesis and optimization of polygeneration systems using appropriate sets of representative days provide a good and reasonable approach and pre-design in terms of configuration, sizing and operation. However, due to the complexity of the problem, an appropriate method for the selection of a set of representative days should be applied, as these highly complex polygeneration systems should meet several energy demands (heating, cooling and electricity in the analyzed cases), using manageable (electricity from the grid, fossil fuels, or biomass) and non-manageable (e.g., solar and wind) energy resources considering several candidate energy conversion technologies (cogeneration, heat pumps, wind turbines, solar PV panels, solar thermal collectors,

mechanical and absorption chillers) as well as short-term energy storage.

Acknowledgments

Appendix A: Optimization model

$$\textbf{\textit{Minimize}}\ Total\ anual\ cost \tag{A.1}$$

$$Total\ annual\ cost = CIA + C_{op} \tag{A.2}$$

$$CIA = (1 + VAT) \cdot (1 + F_{ind}) \cdot CRF \cdot \sum_{j\ =\ component} Cu_j \cdot Cap_j \cdot (1 + FNPV_j)(1 + F_{mj}) \tag{A.3}$$

$$C_{op} = C_E + C_g \tag{A.4}$$

$$C_E = ((C_{fixe} + C_{ve}) \cdot (1 + Tax_e) + C_{alqe}) \cdot (1 + VAT) \tag{A.5}$$

$$C_v = \sum_{d\ =\ 1}^{N_{rep}} \omega_d \cdot \left(\sum_{h\ =\ 1}^{24} cp_e(h) \cdot Ep(h)\right)_d \tag{A.6}$$

$$C_{fixe} = \sum_{i\ =\ 1}^{3} cPct_i \cdot Pct_i \tag{A.7}$$

$$Pct_{nom} = [Pct_1 ... Pct_n] \tag{A.8}$$

$$Pct = [Pct_{nom} \cdot Y_{Pct1}\ \ Pct_{nom} \cdot Y_{Pct2}\ \ Pct_{nom} \cdot Y_{Pct3}] \tag{A.9}$$

$$Pct \geq Ep \tag{A.10}$$

$$C_g = (C_{fixg} + C_{alqg} + C_{vg})(1 + VAT) \tag{A.11}$$

$$C_{vg} = \sum_{d\ =\ 1}^{N_{rep}} \omega_d \cdot \left(\sum_{h\ =\ 1}^{24} cp_{fuel} \cdot F(h)\right)_d \tag{A.12}$$

Balance equations:
An energy balance is carried out in each node m of the superstructure

$$\sum_m (E_{in}^m - E_{out}^m) = 0 \tag{A.13}$$

Equipment efficiency:

GB: $\eta_{GB} \cdot F_b - Q_b = 0$ (A.14)

HP: $Q_{HP} - E_{HP} \cdot COP = 0$ (A.15)

HP: $R_{HP} - E_{HP} \cdot EER = 0$ (A.16)

GE: $\alpha_w \cdot F_{GE} - W_{GE} = 0$ (A.17)

CM: $\alpha_w \cdot F_{CM} - W_{CM} = 0$ (A.18)

CM: $\alpha_q \cdot F_{CM} - Q_c = 0$ (A.19)

ACH: $R_{ach} = COP_{ach} \cdot Q_{ach}$ (A.20)

For thermal energy storages for heating q and cooling r:

$$S_{q,r}(t) = S_{q,r}(t-1) \cdot \lambda_{q,r} + E_{inq,r} - E_{outq,r} \tag{A.21}$$

Equipment capacities:
For renewable energy production components:

PV: $W_{PV} = E_{PV} \cdot A_{PV}$ (A.22)

ST: $Q_{ST} = E_{ST} \cdot A_{ST}$ (A.23)

WT: $W_W = E_{PW} \cdot N_{WT}$ (A.24)

For each component j, the energy production is equal or lower than its nominal capacity. Thus, for heating Q, cooling R or electricity W production:

$$Q_j \leq Cap_j \tag{A.25}$$

$$R_j \leq Cap_j \tag{A.26}$$

$$W_j \le Cap_j \tag{A.27}$$

Stored energy S is equal or lower to nominal capacity of the energy storage.

$$S \le Cap_{nominal} \tag{A.28}$$

Appendix B. Technical, economic and environmental data

Table B.1. Technical, economic and environmental data of components

Component j	Technical data	Economic data			Environmental data
		Cu [€/*]	Fm	n_r [Years]	CO_2U [$kgCO_2eq$/*]
PV	η_{PV}= 15.66% [23]	113.4 €/m² [31]	0.9	20	161 $kgCO_2eq$/m² [32]
WT	Manufacturer curve [25]	51230 €/ud [33]	0.7	20	21600 $kgCO_2eq$/ud [34]
ST	$\eta_o = 0.801$ $a_1 = 3.188$ W/m²·K $a_2 = 0.011$ W/m²·K² [24]	254 €/ m² [35]	1.5	20	95 $kgCO_2eq$/m² [36]
GB	η_b: 0.96 [37]	80 €/kWt [37]	0.5	15	10 $kgCO_2eq$/kWt [2]
HP	COP=3.0, EER= 4.0, β=0.9 [38]	500 €/kW [38]	0.5	20	160 $kgCO_2eq$/kWt [2,39]
ACH	COP_{ACH}= 0.7 [40]	485 €/kWt [40]	1.5	20	165 $kgCO_2eq$/kWt [2,39]
GE	α_w= 0.28 [41]	600 €/kW [41]	0.2	10	65 $kgCO_2eq$/kWe [2]
CM	α_w= 0.28,α_q= 0.56 [42]	1150 €/kWe [43]	0.7	10	
TSQ	λ= 1% (Estimated)	212 €/kWh [38]	0.1	15	31 $kgCO_2eq$/kWht [39]
TSR	λ= 3% (Estimated)	257 €/kWh [38]			62 $kgCO_2eq$/kWht [39]
Bat	η_{rt}=90%; DOD=90%; $N_{c,failure}$ =2000 [44]	370 €/kWh [44]	0.25	12	160 $kgCO_2eq$/kWh [45]

Table B.2. Economic and environmental data for electricity and fuels

Economic data		
	€/kWh	Reference
Electricity price for Zaragoza	-	[46]
Electricity price for gran Canaria	-	[47]
Natural Gas	-	[48]
Gasoil A	0.1174	[49]
Gasoil for heating	0.0678	[49]
Environmental data		
	$kgCO_2eq$/kWh	Reference
Electricity from grid	-	[27]
Natural Gas	0.204	[50]
Gasoil	0.294	[50]

Nomenclature
Renewable energy
PV: Photovoltaic
ST: Solar thermal
WT: Wind turbine
E_{PV}: Hourly electrical production of a wind turbine
E_W: Hourly photovoltaic energy production per square meter
E_{ST}: Hourly solar thermal energy production per square meter

Methods for selecting representative days
c: Attribute or time serie
t: Time period (1-8760 hours)
d: Day (1-365 days)
h: Hour (1-24 hours)

19

$x^{'}$: Value of the element of the attribute

x: Normalized value of the element of the attributes

dis: Dissimilarity

D: Dissimilarity matrix

Ψ: Matrix which contains normalized values x of all attributes

$z_{i,j}$: binary variable equal to 1 if and only if object j is assigned to the cluster of which i is the representative day.

N_k: number of clusters or representative days defined by the user

u: Binary variable used to select the representative day

ω: weight of each representative day

N_i: Number of periods, usually 365 days

N_h: Number of time steps, usually 24 hours

N_c: Number of attributes

μ: Scale factor

DC: Duration curve

DC$_{rep}$: Duration curve from representative days

NDC: Normalized duration curve

b: Division of the ordinate (bin)

s: Division of the abscissa (interval)

L_s: Length of the interval s

$RMSE$: Root-Mean-Square Error

Candidate technologies

GE: Electric generator

CM: Cogeneration module

GB: Conventional boiler

ACH: Absorption chiller

HP: Reversible heat pump

TSQ: Thermal energy storage for heating

TSR: Thermal energy storage for cooling

Bat: Batteries

Data Cost

CIA: Annual investment cost [€/yr]

C_{op}: Annual operational cost [€/yr]

C_E: Electricity bill cost [€/yr]

C_g: Annual cost of fuel consumption [€/yr]

C_{fix}: Fixed cost bill [€/yr]

C_v: Electricity or fuel consumption cost[€/yr]

C_{alq}: Meter equipment rental cost [€/yr]

CRF: Capital Recovery Factor 0.0802

Cu: Unit Cost [€/*]

cp: Electricity or fuel price [€/kWh]

F_{ind}: Indirect cost factor

$FNPV$: Net Present Value factor

F_m: Installation and maintenance cost factor

$cPct$: Contracted power unit cost [€/kW·yr]

Pct_{nom}: Nominal power from the grid [kW]

Tax_e: Electricity tax 0.05113

VAT: value-added tax, 0.21 for Zaragoza and 0.03 for gran Canaria

Energy fluxes

Ep: Purchased electricity [kWh]

F: Fuel consumption [kWh]

E: Energy [kWh]

W: Electricity production [kWh]

Q: Heating production [kWh]

R: Cooling production [kWh]

Variables

A: Area [m^2]

Cap: Nominal capacity

N_{WT}: Number of wind turbines

S: Store energy [kWh]

Y: Binary variable to select contracted power from the grid

Technical parameters

α_w: Electric efficiency engine

α_q: Thermal efficiency engine

COP: Coefficient of performance

EER: Energy Efficiency ratio

η: Efficiency

λ: Energy losses factor for thermal energy storage

Sub-indexes $_e$ and $_g$ indicates electricity and fuel respectively

References

[1]   P. Mancarella, MES (multi-energy systems): An overview of concepts and evaluation models, Energy. 65 (2014) 1–17. doi:10.1016/J.ENERGY.2013.10.041.

[2]   E.A. Pina, M.A. Lozano, L.M. Serra, A Multicriteria Approach for the Integration of Renewable Energy Technologies and Thermal Energy Storage to Support Building Trigeneration Systems, in: Int. Confenerence Sol. Heat. Cool. Build. Ind., 2017.

[3]   T. Wakui, R. Yokoyama, Optimal structural design of residential cogeneration systems with battery based on improved solution method for mixed-integer linear programming, Energy. 84 (2015) 106–120. doi:10.1016/J.ENERGY.2015.02.056.

[4]   K. Tapia-Ahumada, I.J. Pérez-Arriaga, E.J. Moniz, A methodology for understanding the impacts of large-scale penetration of micro-combined heat and power, Energy Policy. 61 (2013) 496–512. doi:10.1016/J.ENPOL.2013.06.010.

[5]   A. Rong, Y. Su, Polygeneration systems in buildings: A survey on optimization approaches, Energy Build. 151 (2017) 439–454. doi:10.1016/J.ENBUILD.2017.06.077.

[6]   M. Carvalho, M.A. Lozano, L.M. Serra, Multicriteria synthesis of trigeneration systems considering economic and environmental aspects, Appl. Energy. 91 (2012) 245–254. doi:10.1016/J.APENERGY.2011.09.029.

[7]   J. Ortiga, J.C. Bruno, A. Coronas, Selection of typical days for the characterisation of energy demand in cogeneration and trigeneration optimisation models for buildings, Energy Convers. Manag. 52 (2011) 1934–1942. doi:10.1016/J.ENCONMAN.2010.11.022.

[8]   S.G. Chen, Bayesian approach for optimal PV system sizing under climate change, Omega. 41 (2013) 176–185. doi:10.1016/j.omega.2011.07.013.

[9]   T. Schütz, M.H. Schraven, M. Fuchs, P. Remmen, D. Müller, Comparison of clustering algorithms for the selection of typical demand days for energy system synthesis, Renew. Energy. 129 (2018) 570–582. doi:10.1016/J.RENENE.2018.06.028.

[10]   L. Kotzur, P. Markewitz, M. Robinius, D. Stolten, Impact of different time series aggregation methods on optimal energy system design, Renew. Energy. 117 (2018) 474–487. doi:10.1016/J.RENENE.2017.10.017.

[11] K. Poncelet, H. H oschle, E. Delarue, A. Virag, W.D'haeseleer, Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems, IEEE Trans. POWER Syst. 32 (2017). doi:10.1109/TPWRS.2016.2596803.

[12] M. Muselli, P. Poggi, G. Notton, A. Louche, Classification of typical meteorological days from global irradiation records and comparison between two Mediterranean coastal sites in Corsica Island, Energy Convers. Manag. 41 (2000) 1043–1063. doi:10.1016/S0196-8904(99)00139-9.

[13] M. Muselli, P. Poggi, G. Notton, A. Louche, First order Markov chain model for generating synthetic "typical days" series of global irradiation in order to design photovoltaic stand alone systems, Energy Convers. Manag. 42 (2001) 675–687. doi:10.1016/S0196-8904(00)00090-X.

[14] F. Domínguez-Muñoz, J.M. Cejudo-López, A. Carrillo-Andrés, M. Gallardo-Salazar, Selection of typical demand days for CHP optimization, Energy Build. 43 (2011) 3036–3043. doi:10.1016/j.enbuild.2011.07.024.

[15] IDAE, Escala de calificación energética para edificios de nueva construcción, 2009. http://www.idae.es/uploads/documentos/documentos_CALENER_07_Escala_Calif_Energetica_A2009_A_5c0316ea.pdf (accessed October 18, 2017).

[16] J. Ramos, Optimization of the design and operation of cogeneration systems for the residential and commercial sector. Ph. D. Thesis, Universidad de Zaragoza, 2012.

[17] AENOR, Instalaciones solares térmicas para producción de agua caliente sanitaria-UNE 94.002, (2005).

[18] A. Viti, DTIE 1.01 Preparación de agua caliente para usos sanitarios, ATECYR, 1996.

[19] IDAE, Consumos del Sector Residencial en España - Resumen de Información Básica, 2011. http://www.idae.es/uploads/documentos/documentos_Documentacion_Basica_Residencial_Unido_c93da537.pdf (accessed November 5, 2017).

[20] J.M. Marín Giménez, Evaluation of alternatives for the energy supply of a residential building in Zaragoza, Universidad de Zaragoza, 2004.

[21] J.A. Duffie, W.A. Beckman, Solar Engineering of Thermal Processes, 4th ed., John Wiley & Sons, 2013.

[22] Meteotest, Meteonorm Software, (2017). http://www.meteonorm.com/ (accessed November 3, 2017).

[23] Atersa, Specifications of photovoltaic module A-255P, (2017). http://www.atersa.com/Common/pdf/atersa/manuales-usuario/modulos-fotovoltaicos/Ficha_Tecnica_A-255P-A-265P_Ultra.pdf (accessed January 6, 2018).

[24] Salvador Escoda S.A, Tarifa de precios, (2017). https://www.salvadorescoda.com/tarifas/Energias_Renovables_Tarifa_PVP_SalvadorEscoda.pdf (accessed January 6, 2018).

[25] Aeolos, Aeolos Wind Turbine 30kW Specification, (2006). http://www.windturbinestar.com/30kwh-aeolos-wind-turbine.html (accessed May 28, 2019).

[26] J.F. Manwell, J.G. McGowan, A.L. Rogers, Wind Energy Explained, 2nd ed., WILEY, 2009.

[27] REE, Demanda y producción en tiempo real, (2019). http://www.ree.es/es/actividades/demanda-y-produccion-en-tiempo-real (accessed May 28, 2019).

[28] R. Dufo-López, J.M. Lujano-Rojas, J.L. Bernal-Agustín, Comparison of different lead–acid battery lifetime prediction models for use in simulation of stand-alone photovoltaic systems, Appl. Energy. 115 (2014) 242–253. doi:10.1016/j.apenergy.2013.11.021.

[29] N. DiOrio, A. Dobos, S. Janzou, A. Nelson, B. Lundstrom, Technoeconomic Modeling of Battery Energy Storage in SAM, NREL Technical Report, 2015.

[30] D.R. Kiran, D.R. Kiran, Quality Function Deployment, Total Qual. Manag. (2017) 425–437. doi:10.1016/B978-0-12-811035-5.00030-1.

[31] R. Fu, D. Feldman, R. Margolis, M. Woodhouse, K. Ardani, U.S. Solar Photovoltaic System Cost Benchmark: Q1 2017, 2017. doi:10.2172/1395932.

[32] P.S. R. Frischknecht, R. Itten, F. Wyss, I. Blanc, G. Heath, M. Raugei, A. Wade, Life Cycle Assessment of Future Photovoltaic Electricity Production from Residential - scale Systems Operated in Europe Life Cycle Assessment of Future Photovoltaic., 2015. doi:10.1016/j.jamda.2016.12.070.

[33] A. Orrell, E. Poehlman, Benchmarking U.S. Small Wind Costs With the Distributed Wind Taxonomy, 2017. https://wind.pnnl.gov/pdf/Benchmarking_US_Small_Wind_Costs_092817_PNNL.pdf (accessed January 15, 2019).

[34] B. Tremeac, F. Meunier, Life cycle analysis of 4.5 MW and 250 W wind turbines, Renew. Sustain. Energy Rev. 13 (2009) 2104–2110. doi:10.1016/J.RSER.2009.01.001.

[35] IDAE, Plan de Energías Renovables (PER) 2011-2020, 2011. http://www.idae.es/tecnologias/energias-renovables/plan-de-energias-renovables-2011-2020 (accessed January 1, 2018).

[36] M. Guadalfajara, Economic and environmental analysis of central solar heating plants with seasonal storage for the residential sector. Ph. D. Thesis, Universidad de Zaragoza, 2016.

[37] BAXI, Catálogo tarifa, (2017). https://www.baxi.es/-/media/websites/baxies/files/catalogo-2017-bajaress.pdf (accessed January 6, 2018).

[38] ENERTRES, Catálogo tarifa 11E, (2017). https://enertres.com/aerotermia/ (accessed May 2, 2019).

[39] M. Beccali, M. Cellura, S. Longo, D. Mugnier, A Simplified LCA Tool for Solar Heating and Cooling Systems, Energy Procedia. 91 (2016) 317–324. doi:10.1016/J.EGYPRO.2016.06.226.

[40] U.S. Department of Energy, Absorption Chillers for CHP Systems, 2017. https://energy.gov/sites/prod/files/2017/06/f35/CHP-Absorption Chiller-compliant.pdf (accessed January 7, 2018).

[41] Ayerbe, Catálogo de grupos electrógenos 2018, (2018). http://www.ayerbe.net/AYERBE-CATALOGO-2018.pdf (accessed October 3, 2018).

[42] Yanmar, Combined Heat & Power, (2017). http://www.yanmar-es.com/products/mchp/ (accessed January 6, 2018).

[43] K. Darrow, R. Tidball, J. Wang, A. Hampson, Catalog of CHP technologies, 2017. https://www.epa.gov/sites/production/files/2015-07/documents/catalog_of_chp_technologies.pdf (accessed January 4, 2018).

[44] IRENA, Electricity storage and renewables: Costs and markets to 2030, 2017. doi:ISBN 978-92-9260-038-9 (PDF).

[45] J.F. Peters, M. Baumann, B. Zimmermann, J. Braun, M. Weil, The environmental impact of Li-Ion batteries and the role of key parameters – A review, Renew. Sustain. Energy Rev. 67 (2017) 491–506. doi:10.1016/J.RSER.2016.08.039.

[46] Endesa, Electricity price for Zaragoza, (2019). https://www.endesaclientes.com/static/iberia/empresas/condiciones/anexo-precio-electricidad-es.pdf (accessed May 29, 2019).

[47] Aura Energía, Electricity price for Gran Canaria, (2019). https://www.aura-energia.com/tarifas-luz-empresa-canarias/ (accessed May 29, 2019).

[48] Endesa, Natural gas prices, (2019). https://www.endesaclientes.com/empresas/tarifa-gas-estable.html.

[49] IDAE, Estudios, informes y estadísticas, (2018). http://www.idae.es/estudios-informes-y-estadisticas (accessed January 30, 2019).

[50] Carbon footprint, 2016 Carbon conversion factors, (2016). https://www.carbonfootprint.com/2016_carbon_conversion_factors.html (accessed February 14, 2019).

CRediT author statement

Edwin S. Pinto: Conceptualization, Methodology, Software, Validation, Formal analysis Investigation, Writing - Original Draft.

Luis M. Serra: Conceptualization, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Supervision.

Ana Lázaro: Conceptualization, Investigation, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

Term	Definition
Conceptualization	Ideas; formulation or evolution of overarching research goals and aims
Methodology	Development or design of methodology; creation of models
Software	Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components
Validation	Verification, whether as a part of the activity or separate, of the overall replication/ reproducibility of results/experiments and other research outputs
Formal analysis	Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data
Investigation	Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection
Resources	Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools
Data Curation	Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse
Writing - Original Draft	Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation)
Writing - Review & Editing	Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre-or postpublication stages
Visualization	Preparation, creation and/or presentation of the published work, specifically visualization/ data presentation
Supervision	Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team
Project administration	Management and coordination responsibility for the research activity planning and execution
Funding acquisition	Acquisition of the financial support for the project leading to this publication

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Highlights

- Comparison of representative days' selection methods for optimization
- Pros and cons of averaging, *k*-Medoids and OPT methods are evaluated
- Guidelines for applying adequate representative days' selection method are provided
- A new method, combination of *k*-Medoids and OPT methods, is proposed