k-MILP: A novel clustering approach to select typical and extreme days for multi-energy systems design optimization

# k-MILP: a novel clustering approach to select typical and extreme days for multi-energy systems design optimization

*Matteo Zatti [a,b,*], Marco Gabba [a], Marco Freschini [b], Michele Rossi [c], Agostino Gambarotta [d], Mirko Morini [d], and Emanuele Martelli [b,**]*

[a] *Laboratorio Energia & Ambiente Piacenza, Via Nino Bixio 27/c, 29121, Piacenza, Italy*
[b] *Politecnico di Milano, Department of Energy, Via Lambruschini 4, 20154, Milano, Italy*
[c] *SIRAM, Via Anna Maria Mozzoni 12, 20152, Milano, Italy*
[d] *Università degli Studi di Parma, Centro Interdipartimentale per l'Energia e l'Ambiente, Viale delle Scienze 42, 43124, Parma, Italy*

**Abstract:**

When optimizing the design of multi-energy systems, the operation strategy and the part-load behavior of the units must be considered in the optimization model, which therefore must be formulated as a two-stage problem. In order to guarantee computational tractability, the operation problem is solved for a limited set of typical and extreme periods. The selection of these periods is an important aspect of the design methodology, as the selection and sizing of the units is carried out on the basis of their optimal operation in the selected periods. This work proposes a novel Mixed Integer Linear Program clustering model, named k-MILP, devised to find at the same time the most representative days of the year and the extreme days. k-MILP allows controlling the features of the selected typical and extreme days and setting a maximum deviation tolerance on the integral of the load duration curves. The novel approach is tested on the design of two different multi-energy systems (a multiple-site university Campus and a single building) and compared with the two well-known clustering techniques k-means and k-medoids. Results show that k-MILP leads to a better representation of both typical and extreme operating conditions guiding towards more efficient and reliable designs.

**Keywords:**

Multi-energy Systems, District energy systems, Typical days, Extreme days, Design optimization.

## 1. Introduction

Recently, the Multi-energy Systems (MES) and District Energy Systems (DES) paradigms have been attracting the interest of both private and public institutions. Indeed, they appear as promising solutions to lower the costs and the environmental impact of the energy production, distribution and use in urban areas [1]. The key element for their success is the integration of the energy networks – i.e. electricity, heating and cooling – operating in residential, commercial and industrial districts, so as to maximize the possible synergies among them. In addition, this energy integration approach is considered as a valuable mean to increase the share of renewable energy sources into the production mix [2]. An underlying result of the analysis of the MES and DES is that, when dealing with their design optimization, the development of efficient and reliable models is challenged by at least three main aspects: (i) their multi-location scale, i.e. the spatial distribution; (ii) their multi-temporal scale, i.e. the design and operation variables are part of the same problem; and (iii) the uncertainty of the input data (design and operation choices are made today on the basis of forecasts about the future). As for aspect (i), the optimal design may entail a mix of centralized and distributed generation solutions, thus a design optimization model has to accommodate for the possibility of installing multiple energy conversion units of variable size in many different locations of the district. As for aspect (ii), the optimal design has to take into account the part-load performances and operational limitations (e.g., start-up/shut-down time and costs, ramp-up rates, etc.) that characterize the

[*] Corresponding author: matteo.zatti@polimi.it

[**] Co-corresponding author: emanuele.martelli@polimi.it

operation of the energy conversion units [3], since it has severe implications on the design choices. Thus a reliable and efficient MES and/or DES design methodology needs to take into account the conditions that the energy system (the installed units and networks) would face in the real-world operation, which requires the formulation of the synthesis problem as a two-stage optimization problem [7]. In order to guarantee the computational tractability of such kind of problem, the operation problem is solved for a limited set of typical periods (e.g., days or weeks), for which the hourly profiles of energy demands, energy prices and energy production from Renewable Energy Sources (RES) are usually considered. As for aspect (iii), it requires advanced modeling and optimization strategies, like the stochastic programming approaches proposed by Zatti et al. in [4] and Mitra et al. in [5], respectively, or the advanced approach to account for data uncertainty developed by Gabrielli et al. in [6], which shows how the daily value of energy demands give indications on the robustness of the design.

Among the three above-listed issues, the selection of typical and extreme periods (e.g., days or weeks) plays an important role, as the selection and sizing of the units is carried out on the basis of their performances in the considered periods. Many different methods for the selection of the typical periods have been presented and compared for energy systems design optimization problems. The majority uses clustering approaches (CA) to divide periods with "similar" profiles into clusters and then define a representative period for each cluster. Several clustering approaches are possible, depending on the criteria used to measure the "similarity" between the data profiles of the periods and to select the representative period of the cluster. The main CAs used so far in the field of energy systems design optimization include k-means [7] [8] [10] [11] [12], k-medians [11], k-medoids [8] [11] [12] [12] [15], k-centers [11], hierarchical [8] [12], down-sampling resolution [14], averaging [8], dynamic time warping barycenter averaging [12] and k-shape [12].

Schutz et al. [11] compare k-means, k-medians, k-medoids, k-centers with fixed time based representations using synthetic load profiles of an apartment building and a single-family house. They conclude that all clustering methods are able to determine close-to-optimal designs and k-medoids is able to approximate the operating costs with highest accuracy.

Kotzur et al. [8] compare averaging periods, k-means, k-medoids, and hierarchical clustering using three case studies with different peculiarities. They found that the averaging periods approach leads to inaccurate results while none of the other aggregation methods outperforms all the others for every case study. In general, selecting the medoid of the cluster performs better than the centroid because the averaged time-series of the k-means centroid underestimates the real system cost. Another important conclusion concerns the effects of the peculiarities of the case studies: centralized grid-connected cases are easier to represent with typical periods compared to islanded systems with a higher share of renewable production.

Teichgraeber et al. [12] compare k-means, k-medoids and hierarchical clustering with shape-based clustering approaches (dynamic time warping barycenter averaging and k-shape) for the approximation of electricity price series. They use two different power plants designs as references for the comparison finding that centroid-based clustering (e.g., k-means) can represent the operational domain more predictably than medoid-based approaches. Dynamic time warping barycenter does not perform well while k-shape performs very well on the test problem featuring a battery for electricity arbitrage. In such problem k-shape can represent the full period with just two days that capture the arbitrage potential between hours.

Elsido et al. [10] tackle the optimal design problem for the MES serving a medium size district heating network. The authors propose a "weighted" version of the k-means algorithm in which different weights are assigned to the relative errors of electricity demand, heat demand and ambient temperature. Such weights are related to the influence that each data profile has on the total yearly operating costs. For such purpose, the authors propose a systematic approach in which the relative effect of each data profile on the yearly total operating cost is evaluated for different designs. Results for an industrial case study show that for all designs, the influence/weight of the relative error of heat demand is appreciably higher (almost double) than that of electricity. This result depends on the fact that electricity can be imported/exported from/to the grid, while the heat demand must be necessarily met by the installed CHP units becoming a strong constraint on their scheduling/operation. The use of the proper weights allows decreasing the error in assessing the total operating costs from 10% to 3%.

Bahl et al. [16] propose an iterative time-series aggregation approach aimed at finding under-estimators of energy demand of the clusters. This allows finding a lower bound for the design optimization problem that can be used within a rigorous decomposition method to assess the need of increasing the number of representative periods. The main limit of the methodology is the loss of the chronology of time steps which makes it not suitable for problems with storage systems and/or dynamic constraints (e.g., ramp-up and minimum up-time constraints of units).

Since the typical periods are the most representative profiles of the clusters, extreme periods are not included and need to be added so as to enforce the operational feasibility of the system throughout the year. There is no general consensus on the best approach to select the extreme periods; nonetheless, it seems that the selection of the periods in which the peak values are reached is one of the most successful [9]. Some authors have developed algorithms for the optimization of the design that iteratively add the extreme periods to the input data set on the basis of the feasibility of the design solutions, like in [7]. Nevertheless, this last approach is expected to be rather computationally intensive for large size and multi-site case studies, as the actual extreme period might be found in the last iteration.

Differently from clustering-based selection approaches, some researchers have proposed systematic approaches to select typical periods which do not cluster the yearly data into clusters, targeting directly the selection of the most representative days on the basis of pre-defined quality indexes. For example, Poncelet et al. [17] have proposed an optimization-based method that selects the set of days that minimizes the sum of differences between the original and the approximated load duration curves, ramp duration curves (required for reproducing short-term dynamic issues), and correlation between different data series. The load duration curves need to be discretized into bins and, similarly to clustering-based approaches, extreme periods selection is not included in the optimization problem.

As found by different authors ([14] [8] [12], no aggregation method outperforms all the others for every case study and the best CA varies depending on the intra-day variability of profiles and problem constraints (e.g., operating range and ramp limitations of the units, size of the energy storage systems, islanded or grid-connected system). Another important conclusion of comparison analyses [8][11] is that assessing the quality of representative days on the basis of performance indicators may be misleading. Ideally, the effects of the approximations caused by the use of aggregated time series in the design should be evaluated with respect to the exact design solution considering the original yearly time series, as done in [8]. If this is not practicable because the design optimization problem becomes intractable, an alternative approach is to evaluate the operation of the optimized design across the whole year (using the original time series), as done by [11]. This provides a more accurate estimate of the actual total annual costs of the optimized systems and it can be used to compare the different time series aggregation approaches.

From a more general perspective, while the selection of typical periods has been investigated by several authors, very little attention has received the selection of extreme periods (days or weeks). Their relevance is expected to be considerable especially for islanded systems (e.g., rural microgrids) and all those applications requiring high reliability in which outages need to be avoided or minimized (e.g., MES serving public buildings like hospitals and schools).

In this work, we propose an optimization-based approach for the systematic selection of both typical and extreme periods (e.g., days or weeks). The approach is based on a MILP model that divides the real periods into similar clusters while automatically identifying the extreme periods. The MILP model allows including essential requirements on the selected typical and extreme periods. The novel clustering method is tested against k-means and k-medoids on the design optimization of the district energy system for the University of Parma Campus and a single building, both in Northern Italy. The district energy system is made of five district heating networks and features electricity, heating and cooling demands. The single building features electricity and heating demands. The optimal design problem is linearized and formulated as a large-scale MILP and solved with state-of-the-art MILP solvers.

## 2. Formulation of the design optimization problem

The design optimization problem considered in this work has been formulated as a two-stage MILP problem, involving investment decisions (first stage) and operation decisions (second stage). This two-stage structure is represented by the following compact formulation:

$$\min_{x_u^{(1)}, x_{u,t}^{(2)}} TAC = \sum_{u \in U} C^{INV} \cdot x_u^{(1)} + \sum_{u \in U} \sum_{t \in T} C^{OP} \cdot x_{u,t}^{(2)} \quad (1)$$

Subject to

$$A^{(1)} x_u^{(1)} + A^{(1)} y_u^{(1)} = b_u^{(1)} \quad \forall\, u \quad (2)$$

$$A^{(2)} x_{u,t}^{(2)} + A^{(2)} y_{u,t}^{(2)} = b_t^{(2)} \quad \forall\, u, \forall\, t \quad (3)$$

$$A^{(1)} x_u^{(1)} + A^{(1)} y_u^{(1)} + A^{(2)} x_{u,t}^{(2)} + A^{(2)} y_{u,t}^{(2)} = b_{u,t}^{(1,2)} \quad \forall\, u, \forall\, t \quad (4)$$

$$x_u^{(1)} \in \mathbb{R}, x_{u,t}^{(2)} \in \mathbb{R}, y_u^{(1)} \in \{0,1\}, y_{u,t}^{(2)} \in \{0,1\} \quad (5)$$

Where $TAC = CAPEX + OPEX$ is the total annualized costs; $u \in U$ is the sets of energy conversion and storage units and $t \in T$ is the set of time steps considered in the operation. $x^{(1)}$ and $y^{(1)}$ are, respectively, the continuous (unit sizes, storage capacities, network branches capacities) and binary (unit/storage/network branch selection and installation) investment variables, and $x^{(2)}$ and $y^{(2)}$ are the continuous (unit load, storage level, network branch power flow) and binary (unit on/off status) operation variables. As one can see in Eq. ( 5 ) , there are both real and binary variables in both stages making the problem structurally similar to a two-stage stochastic problem with integer recourse [18]. There are constraints that accounts only for the investment stage, as in Eq. ( 2 ), as can be the minimum and maximum unit sizes or the available locations for units' installation. There are also constraints referring only to the operation stage, as in Eq. ( 3 ), e.g., the balances between the energy production and demand. Finally, there are constraints that bind the first and the second stage variables, as in Eq. ( 4 ), e.g., the performance maps of the units which depend on their size. Indeed, in the model used in this study, the part load-load operation of the units is represented with the convex hull formulation [19] and the strategy proposed in [3] is adopted to linearize the size effects on units' performance.

To sum up, Problem ( 1 ) - ( 5 ) can be stated as follows:

Given:

- Topology of the district
- Hourly profiles in the selected representative periods: energy demands, Renewable Energy Sources production, energy prices, weather parameters
- Catalogue data of the considered technologies in terms of: CAPEX [€/kW], OPEX [€/kWh], performance maps, operational constraints.

Determine:

- Selection and sizes of units to be installed in each site
- Hourly profiles in the representative periods of: on/off and loads of the energy conversion units, management strategy of the storage systems, energy import/export profiles of the district, district network energy flows.

So as to minimize the Total Annualized Cost

# 3. Clustering techniques for design optimization

The ideal design problem would involve the optimization of the operation in the whole set of time steps, yet this way the problem would be, nowadays, not solvable in reasonable amount of time. This becomes even more relevant when Problem ( 1 ) – ( 5 ) is applied to DES with distributed multiple energy demands and distributed multi-energy production systems, bringing about a further set, namely the one involving the sites in which a DES can be divided, as we already considered in [4]. To enable the reduction of the complexity of such problems, time series aggregation is usually used in literature to find a few representative operating profiles [9],[10] or to group the binary operational variables [20] and reduce the problem size. The required feature of such aggregation is to be representative of the original time series, in such a way that: 1) the operational feasibility is preserved and 2) the operational costs appearing in the objective function resemble properly the actual operational costs. For the first task, extreme periods are usually added to the input data set, while for the second one, clustering techniques are commonly used.

As thoroughly explained in [8], the aim of time series aggregation techniques is to gather a set of periods $j \in \{1, \dots, N_d\}$ (e.g., $N_d$ are the 365 days of a year), each consisting of the same number of time steps $h \in \{1, \dots, N_h\}$ (e.g., $N_h$ are the 24 hours), with $N_a$ attributes (e.g., heating demand, irradiance, etc.), into a pre-defined $N_k$ number of groups such that the group members are as similar as possible. Usually, the aggregation is performed by minimizing a distance measure of the attributes between each group member. The groups are then represented by a single period. The selection of the representative period, often called *typical periods*, depends on the specific techniques.

In this work, the attributes considered for the clustering are: heating, electricity and cooling demands, irradiance, ambient temperature and electricity prices. Moreover, based on the periodicity of the energy demand profiles, of the physical phenomena (sun irradiance) and the typical usage of the storage systems, the time step basis we have chosen for the typical periods is 24 hours, that is we have considered *typical days*.

## 3.1. Traditional clustering techniques

### 3.1.1 k-means clustering

Given a fixed number of clusters (i.e., number of typical days), the k-means clustering algorithm creates the clusters by minimizing the squared error between the empirical mean of a cluster and all the candidates in the cluster. This defines a mixed-integer nonlinear program (MINLP) [8], which can be written as in Eq. ( 6 ):

$$\min \sum_{k=1}^{N_k} \sum_{j=1}^{N_d} \left[ \sum_{h=1}^{N_h} \sum_{a=1}^{N_a} \left( x_{a,h,j} - \mu_{a,h,k} \right)^2 \right] \cdot z_{k,j} \qquad (6)$$

where $x_{a,h,j}$ is the value of the attribute (e.g., the heating demand) in the available dataset, $\mu_{a,h,k}$ is the mean value, $z_{k,j}$ is a binary variable that is equal to 1 if the candidate $j$ is assigned to cluster $k$, zero otherwise. In order to make sure that each candidate is assigned exactly to a cluster, constraint ( 7 ) is added.

$$\sum_{k=1}^{N_k} z_{k,j} = 1 \qquad \forall\, j \in \{1, \dots, N_d\} \qquad (7)$$

The MINLP is solved by a greedy algorithm [21] that converges to a local minimum.

The major advantage of such an approach is that, being each representative period calculated as the mean profile of the cluster that it represents, the repetition of the $k$ representative periods, according to the size of the clusters, features the same total value as the original time series for each attribute. On the other hand, its major limit relies on the "smoothness" of the representative period: being averaged profiles they smooth sudden hourly variations with the risk that the shape of the resulting aggregated load duration curve (LDC), i.e. the LDC made by the repetition of the representative days, is quite different from the original LDC, i.e. the LDC made by the real data set. Owing to this smoothing effect of the energy demand profiles, the design optimization problem may be affected by the following limitations: 1) the operational costs tend to be underestimated compared to the real ones; 2) the required installed capacity of generation units, networks and storages may be underestimated (set only by the extreme days); 3) on/off costs and flexibility requirements of the units may be underestimated favoring energy technologies with limited flexibility.

In this work, we consider the following approach based on the k-means clustering technique: days reaching the peak value of each attribute are selected as extreme days (six) while the remaining ones are clustered with a Matlab® implementation of the k-means algorithm [21] to identify the typical periods.

### 3.1.2 k-medoids clustering

In the k-medoids method, instead of using the mean of the cluster as representative period, a real period, the medoid, is chosen among the elements of the cluster. The problem can be stated as a Mixed-Integer Linear Program (MILP) [8]. First, the Euclidean distance between each candidate is calculated as in Eq. ( 8 ):

$$d_{i,j} = \sqrt{\sum_{h=1}^{N_h}\sum_{a=1}^{N_a}\left(x_{a,h,i} - x_{a,h,j}\right)^2} \quad \forall\, i,j \in \{1,\dots,N_d\} \tag{8}$$

Then, the MILP problem ( 9 ) – ( 12 ) can be formulated.

$$\min \sum_{i=1}^{N_d}\sum_{j=1}^{N_d} d_{i,j} \cdot z_{i,j} \tag{9}$$

Subject to:

$$\sum_{j=1}^{N_d} z_{i,j} = 1 \quad \forall\, i \in \{1,\dots,N_d\} \tag{10}$$

$$z_{i,j} \le y_i \quad \forall\, i,j \in \{1,\dots,N_d\} \tag{11}$$

$$\sum_{i=1}^{N_d} y_i = N_k \tag{12}$$

Where $z_{i,j}$ is equal to 1 if candidate $j$ is represented by candidate $i$ and 0 otherwise; $y_i$ is equal to 1 if the candidate $i$ is selected as representative for its cluster and 0 otherwise. Constraint ( 10 ) imposes that each day of the year is associated to one typical day. Constraint ( 11 ) imposes that each day $j$ can be assigned to day $i$ if this last is a representative day. Constraint ( 12 ) ensures that exactly $N_k$ days are selected as typical days. This optimization problem can be solved to global optimality, but the large number of binary variables may lead to an excessive computational time. We used the greedy built-in function in Matlab® that approximates the solution of this problem [21].

The most important advantage of the k-medoids approach relies on the fact that the representative periods are actual values extracted from the original time series, in principle less smoothed than in the k-means case [8]. Moreover, the seasonal and daily correlation between the many attributes is preserved. On the other hand, the crucial limit of the k-medoids approach rests on the fact that, generally, the repetition of the representative periods, according to the size of the clusters, may lead to a total monthly or yearly value, for any attribute, that is quite different from the one calculated for the original data set, especially for small numbers of typical days. This means that the operational costs in the optimization problem can be considered a reliable measure of the operational costs only if a sufficiently large number of clusters are generated. For complex applications featuring multiple networks and sites (e.g., buildings), this may hinder the computational tractability of the design optimization problem.

In this work, we consider the following approach based on the k-medoids clustering technique: the days reaching the peak value of each attribute (six), like in k-means A, are selected as extreme days while the remaining ones are clustered with the k-medoids algorithm [22] to identify the typical days.

## 3.2. k-MILP model for the automatic identification of extreme and typical days

We have developed a MILP model for simultaneously selecting the typical and the extreme days while setting the maximum allowed violation of the yearly total value of the energy demands (heating, cooling and electricity). It is essentially a modification of the tradition k-medoids.

First, in order to allow the model for automatically excluding particular days from the clustering, constraint ( 10 ) of the k-medoids MILP problem needs to be modified into constraint ( 13 ).

$$\sum_{j=1}^{N_d} z_{i,j} \leq 1 \qquad \forall\, i \in \{1, \dots, N_d\} \tag{13}$$

where $z_{i,j}$ is equal to 1 if candidate $j$ is represented by candidate $i$ and 0 otherwise. This way, it is not imposed anymore that each day of the year is associated to one typical day. Indeed, the extreme days are the objects $e$ that are not clustered (i.e., $\sum_{i=1}^{N_d} z_{i,e} = 0$). In order to exclude the trivial solution where all days are classified as extreme, constraint ( 14 ) needs to be added.

$$\sum_{i=1}^{N_d} \sum_{j=1}^{N_d} z_{i,j} = N_d - N_{ED} \tag{14}$$

where $N_{ED}$ is the number of extreme days, a parameter to be fixed a priori.

The MILP model can therefore automatically identify as extreme days the most “atypical” days of the year (i.e., real days not well represented by the representative days of the clusters). These “atypical” days are expected to contain operating conditions for the energy systems very different from those of the typical days.

Moreover, by adding the set of constraints ( 15 ), that bounds, for some selected attribute $a \in A^{LDC}$ (e.g., the energy demands), the difference between the sum over the original entire data set $\left(\sum_{j=1}^{N_d} x_{a,h,j}\right)$ and the sum over the repetition of the representative periods$\left(\sum_{i=1}^{N_d} x_{a,h,i}\right)$, the model can ensure that the LDC for each attribute is properly preserved.

$$\sum_{j=1}^{N_d} \sum_{i=1}^{N_d} \sum_{h=1}^{N_h} \left| x_{a,h,j} - x_{a,h,i} \right| \cdot z_{i,j} \leq \lambda_a^{LDC} \cdot \sum_{j=1}^{N_d} \sum_{h=1}^{N_h} x_{a,h,j} \qquad \forall\, a \in A^{LDC} \tag{15}$$

Clearly, in order to preserve the linearity, the absolute value is formulated within the MILP using two complementary constraints. For example, if we consider the heating demand, constraint ( 15 ) ensures that the difference between the aggregated LDC ~~(i.e., corresponding to the repetition of the representative days)~~ and the original LDC ~~(i.e., corresponding to the real data set)~~ is smaller than a determined tolerance, expressed as a fraction of the original LDC, $\lambda_a^{LDC}$ (e.g., 2-5%).

In addition, we have added constraint ( 16 ) which guarantees that for some selected attributes $a \in A^{peak}$ (e.g., the energy demands), in our case the energy demands, at least one of the extreme periods chosen by the algorithm contains the highest peak value (or close-to) in the data set:

$$\sum_{j=1}^{N_d} m_{a,j} \cdot \max_{\mathrm{h}} x_{a,h,j} \geq \lambda_a^{PEAK} \cdot \max_{d,h} x_{a,h,d} \qquad \forall\, a \in A^{peak} \tag{16}$$

where $h \in \{1..N_h\}$, $d \in \{1..N_d\}$ and $\lambda_a^{PEAK}$ is a real value expressing the fraction of the peak to be represented with the extreme days (e.g., 90-100%). $m_{a,j}$ is a binary variable equal to 1 when day $j$ is selected as extreme day for attribute $a$ and 0 otherwise. This feature is enforced by constraints ( 17 ) and ( 18 ):

$$\sum_{j=1}^{N_d} m_{a,j} = 1 \quad \forall\, a \tag{17}$$

$$m_{a,j} \leq 1 - \sum_{i}^{N_d} z_{i,j} \quad .\forall\, a, \forall\, j \tag{18}$$

# 4. Real-world case studies

## 4.1. University Campus

The first case study concerns the Campus of University of Parma (Italy), featuring electricity, heating and cooling demands. The Campus features in total 21 buildings spread over an area of approximately 0.77 $km^2$. Currently a centralized layout is adopted: the thermal power is provided by five boilers and the cooling demand is supplied by four refrigeration units, all arranged in a dedicated building. From this central site, the heating and cooling networks reach all the other buildings through four independent pipe loops, as shown in Fig. 1. The objective of this study is to determine the retrofit design of the energy supply system that minimizes the total annual cost (annualized capital costs + operating costs). In order to limit the capital cost of the retrofit, the layout of the heating and cooling networks is kept fixed.

The technologies considered in the case study are: cogeneration internal combustion engines (ICE), natural gas boilers, compression refrigerators, heat pumps (HP), heat storage, photovoltaic (PV) panels and solar heating (SH) panels. According to the site characteristics, it has been assumed that internal combustion engines, natural gas boilers and compression refrigerators can only be located in the central site. The number of units that can be installed in each site and the surface available for the installation of PV and SH panels are also parameters of the model. The techno-economic parameters of the model are available in [23].

Within the MILP design optimization model described in Section 2, the four heating and cooling loops (serving several buildings each) have been modeled as sites with given energy demands, given choice of possible generation and storage units, and maximum area available for hosting solar PV and SH panels.

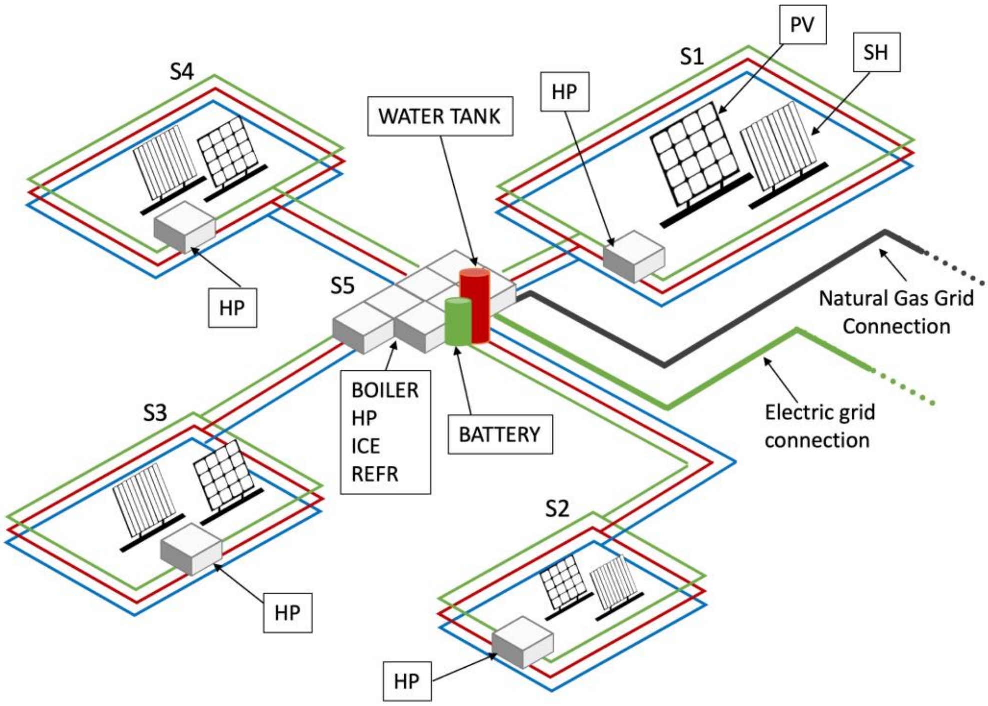



*Fig. 1. The energy district corresponding to the University of Parma Campus studied in this work. Networks: blue = cooling, red = heating, green = electricity. The district is connected to the national electricity grid (green box). There is a central site, where the major part of units can be installed (i.e., batteries, hot water tank for heat storage, boilers, heat pumps, refrigeration cycles) and four sites where only heat pumps, and Photovoltaic (PV) and Solar heating (SH) panels can be installed. HP = heat pumps, ICE = Internal combustion engine, REFR = compression refrigeration cycle*

Hourly values for one year of heating, cooling and electricity demand have been used as input of the model. As concerns heating and cooling, data were made available for each building in the Campus as results of both data collection and physical models, as described in [24]. As regards electricity, only measures of the total monthly demands of the Campus were available. They have been allocated to the different buildings on the basis of the following considerations:

- the characteristic daily and weekly profiles for schools [25];
- the electricity demand during evening hours, weekends and holidays is a very low and constant since the buildings are closed.

In addition to the energy demand profiles, hourly values for ambient temperature and solar irradiation have been retrieved and used for the identification of the typical and extreme days. Global horizontal irradiance and beam horizontal irradiance data have been used to calculate the global irradiance on a tilted surface (angle = 35°, commonly used values at these latitudes), which has been assumed to be directed towards south, so as to maximize the daily and yearly production. Finally, hourly electricity prices have been collected from the Italian power exchange website [26]. The one-year time series for the relevant attributes (normalized between 0 and 1 as explained in Section 5.1.1) are reported in Fig. 2. Three major remarks can be done on these time series. The total heat and electricity demand profiles are quite smooth (compared to the single building) due to the averaging effect of the large number of buildings and users simultaneously connected. The Electricity Demand presents a peak during the summer due to the fact several buildings/offices feature also air conditioning units (absorbing electric power) that are not accounted for in the Cooling Demand but in the Electricity Demand. The Global Tilted Irradiance shows relatively constant values along the year (i.e., it features a weak-seasonal profile) because it represents the global irradiance on a tilted surface whose inclination and orientation has been set so as to maximize the sun energy harvesting at the University Campus location (as previously described), therefore the profile is different from the one expected for the horizontal irradiance.

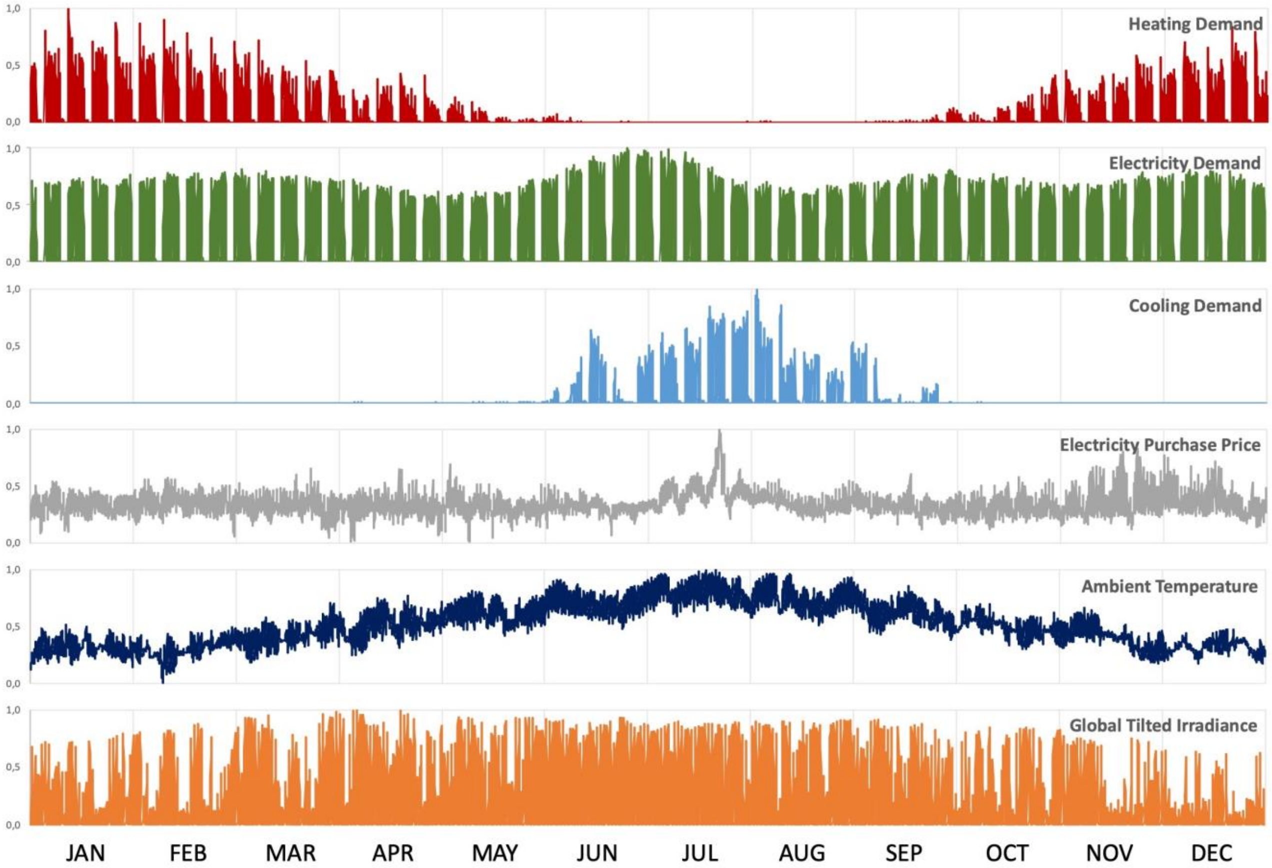



*Fig. 2. University Campus – one-year original (normalized) data sets of heating, cooling and electricity demands (the profiles report the total of the five sites), electricity purchase price, ambient temperature, and the irradiance on a tilted surface (35°) oriented towards the south.*

## 4.2. Single building

The second case study concerns a single university building consisting mainly of offices. Currently, the whole thermal demand is provided by a single boiler, while the electricity demand is fulfilled by the national grid. The cooling demand is negligible. The delivery temperature of the hot water heating system of the building is set at 95 °C and the return temperature at 65 °C. There is a maximum installation capacity for the heat storage equal to 35 $m^3$. The surface available for the installation of panels coincides with the roof of the building, which is equal to 200 $m^2$.

Like for the previous system, the objective of the study is to determine the optimal energy supply system design which minimizes the total annual cost. With respect to the university Campus, here we have also included absorption refrigeration cycles, biomass-fired Organic Rankine Cycles as well as cogeneration micro-ICEs.

Hourly values for one year of heating and electricity demands, ambient temperature, solar irradiation, and electricity price have been gathered and used as inputs of the clustering approaches. The one-year time series for the relevant attributes (normalized between 0 and 1 as explained in Section 5.1.1) are reported in Fig. 3.

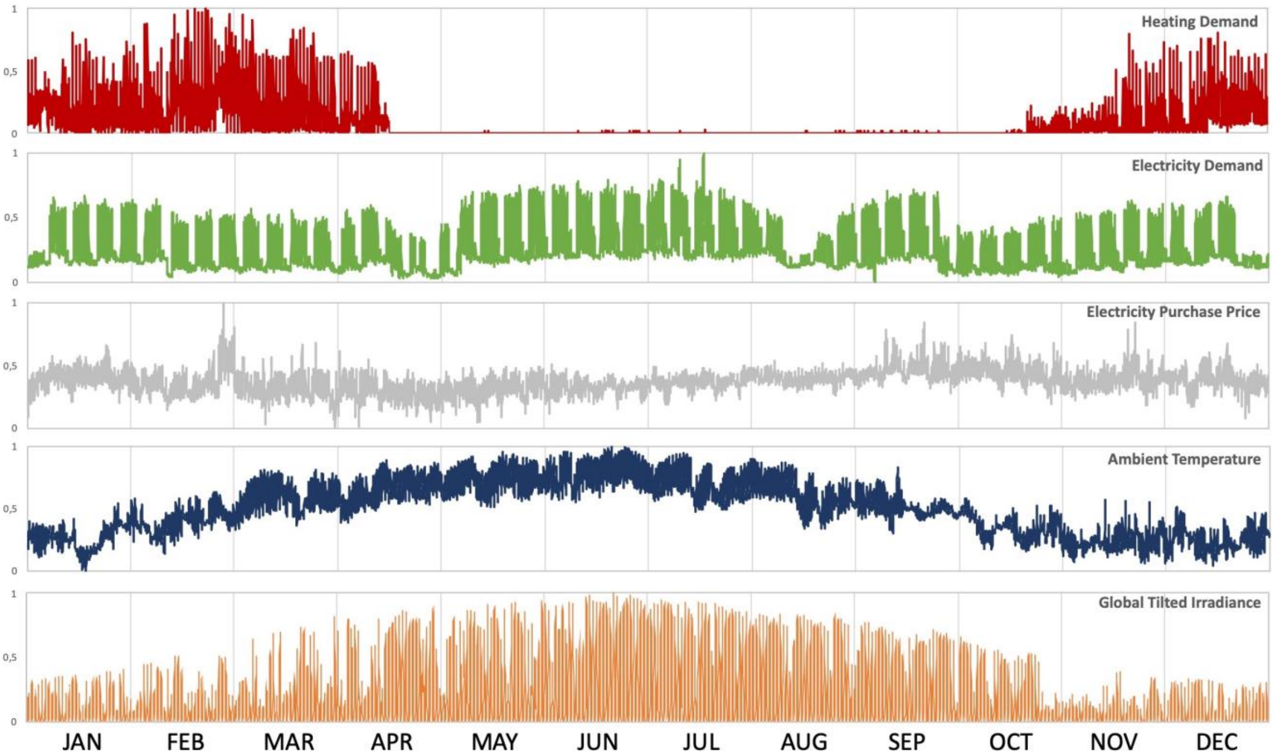



*Fig. 3 Single building – one-year original (normalized) data sets of heating and electricity demands, electricity purchase price, ambient temperature, and the irradiance on a tilted surface (35°) oriented towards the south.*

# 5. Results

The use of the clustering methods analyzed in this work requires the normalization of the data sets, in order to evaluate all the time series on the same scale. We have calculated the normalized time series as:

$$x_{a,h,i} = \frac{\tilde{x}_{a,h,i} - \min \tilde{x}_{a,h,i}}{\max \tilde{x}_{a,h,i} - \min \tilde{x}_{a,h,i}} \tag{19}$$

Where $\tilde{x}_{a,h,i}$ is the original value and $x_{a,h,i}$ is the normalized value. For each attribute $a$, we have normalized the time series with respect to the max and min values within the data set. This is quite straightforward as long as the so-called general attributes are considered, that is the irradiance, the electricity price and the ambient temperature. For the university Campus, the normalization of the attributes that are specific per each site (i.e., the heating, cooling and electricity demands of each site) requires particular attention. In this work, good results have been obtained by normalizing the profiles with respect to the peak value among the sites.

In order to obtain a computationally tractable MILP design problem, we have considered 6 typical days and 6 extreme days. We assumed 100 iterations for the k-means and k-medoids algorithms and a MIP gap of 0.05% for the solver (CPLEX [27]) of the MILP clustering model.

## 5.1. University Campus

### 5.1.1 Clustering results

Fig. 4 shows the normalized plots of the typical and extreme days selected by the three different approaches. The shape of the profiles of the typical days identified by the k-means is smoother than that of the k-medoids and k-MILP clustering method, as expected. The extreme days selected by the k-means and the k-medoids methods are those in which the peaks of the heating demand (ED1), the sun irradiation (ED2), the electricity demand (ED3), the ambient temperature (ED4), the electricity purchase price (ED5) and the cooling demand (ED6) are reached.

Concerning the extreme days, the analysis of the systematic selection made by k-MILP indicates that:

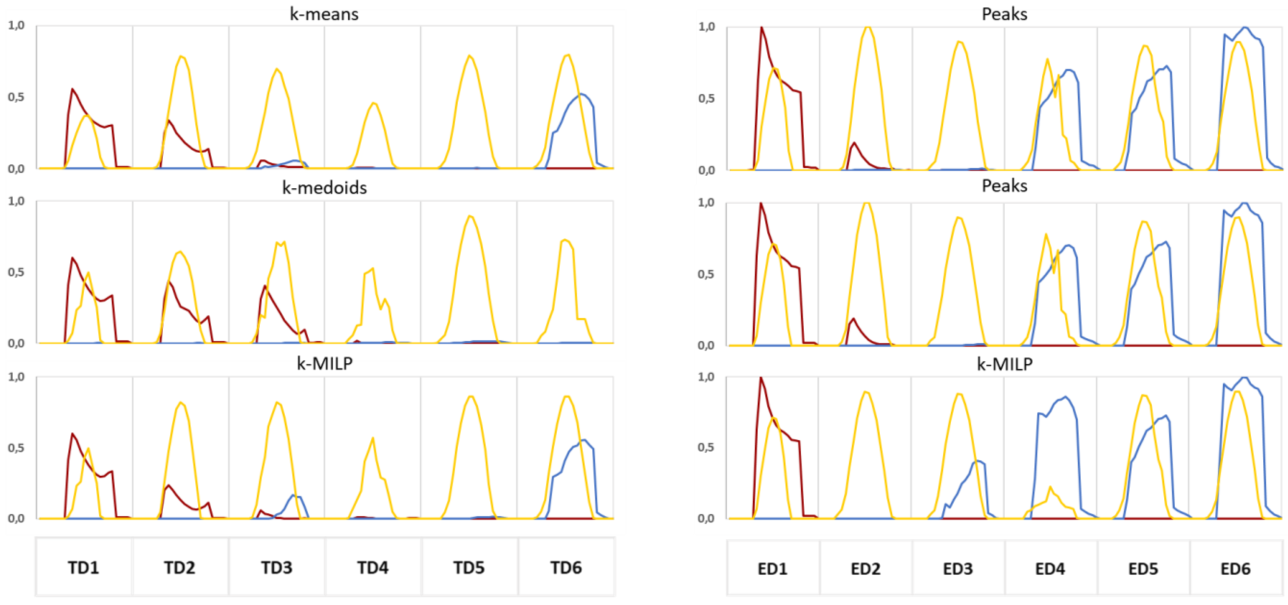
- ED1 features the peak heating demand;
- ED2 is identified as “a-typical” (not similar to the typical days) because with respect to typical day 5 – which is similar concerning the high sun irradiation and null cooling demand – it features a higher electricity purchase price (not visible in the plot);
- ED3 features the peak electricity demand (not visible in the plot);
- ED4 is identified as “a-typical” because it features a high cooling demand but a low irradiation (indeed it is unusual for summer days and it could be critical day for a design employing solar PV to provide electricity to refrigeration cycles);
- ED5 identified as “a-typical” because of its considerably higher average cooling demand with respect to the typical day 6;
- ED6 for the peak cooling demand;

It is worth noting that 4 out of 6 EDs selected by the k-MILP clustering feature high cooling demand, while “peaks” method includes at least a day with some heating demand besides the peak one.

This shows that the k-MILP clustering method has the capability of automatically identifying a-typical days that are potentially extreme for the energy systems providing energy to the district of buildings. It is worth noting that classic approach to select extreme days does not find such potentially critical days (indeed their extreme days feature high cooling demand and high solar radiation). The dimensions of the clusters and the selection of the extreme days are reported in **Errore. L'origine riferimento non è stata trovata.**.

Fig. 5 reports the comparison between the original LDC of the real time series (original LDC) and those estimated with the typical and extreme days (aggregated LDC). The corresponding errors and root mean square deviations (RMSD) are reported in **Errore. L'origine riferimento non è stata trovata.**. The heat LDC is well approximated by all approaches but the k-medoids, whose aggregated LDC features a total heating demand 30% larger than the original one, with substantial deviations at mid-low loads. The k-medoids overestimates the electricity demand – the aggregated LDC has an error of about 11% with respect to the original LDC – and considerably underestimates the cooling demand, indeed the total yearly value is even 12% of the actual one. The k-MILP clustering approach closely resembles the LDCs of all attributes and, for the selected ones (electricity, cooling and heating demands), features a deviation of the total yearly value limited to 2%, as imposed by constraint ( 15 ).

We have repeated the identification of the typical days doubling the number of representative days (i.e., 12). The computational time of the k-MILP clustering method remains limited to 60 seconds. The differences between the typical days identified by the three approaches tend to decrease and all them can closely match the original LDCs. On the other hand, the computational time required to solve the MILP design problem increase exponentially showing that considering 12 typical days becomes impracticable for plants featuring more than 5 sites. Thus, for multi-site systems, achieving a close representation of the whole yearly time series with a few typical days is essential.




(a) (b)

*Fig. 4. University Campus – Normalized profiles of (a) typical days and (b) extreme days for the three clustering techniques. The profiles are: red = heating demand, yellow = irradiance and light blue = cooling demand.*

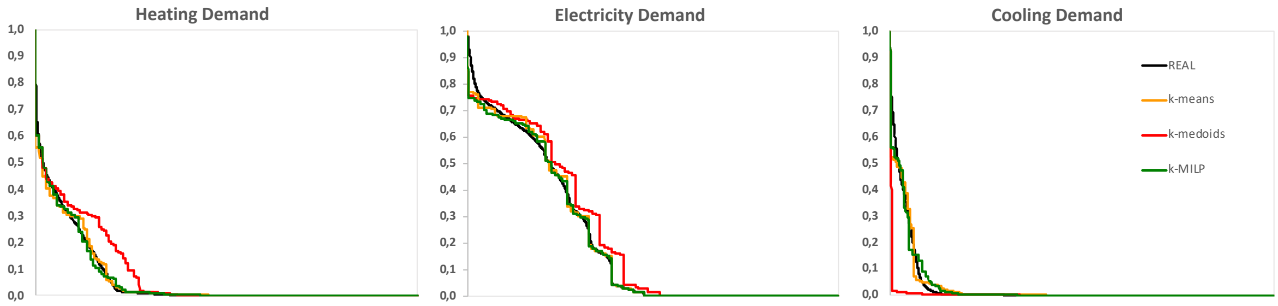



*Fig. 5. University Campus – load duration curves of the three energy demands.*

### 5.1.2 Design optimization results

The MILP design optimization problem has been solved considering the typical and extreme days returned by the three analyzed approaches. Four scenarios have been considered:

- Scenario 1UC (UC stands for University Campus): outages are not allowed (the heat and cooling demand must always be satisfied by the installed units).
- Scenario 2UC: if the energy systems supplying heat and cooling power cannot meet the users' demands in a certain hour of the day, a fixed fee of 5000 Euro/hour must be paid plus a variable

fee proportional to the load shedding (100 and 200 Euro per MWh of, respectively, heating and cooling demand not met).

- Scenario 3UC: in case of outages, only a variable fee of 100 and 200 Euro per MWh of, respectively, heating and cooling demand not met is paid.
- Scenario 4UC: outages are not allowed, and a limit is imposed on the net fossil $CO_2$ emissions of the DES, equal to 75% of the value reported for Scenario 1UC.

Scenarios 1UC, 2UC and 3UC have been devised to investigate: (i) how different outages policies affect the design choices and the corresponding annual costs; (ii) how the clustering methods affect the design accuracy and feasibility when outages are allowed in the design process. The fees that have been assumed are typical values considered in the operation of energy systems serving buildings. Scenarios 4UC has been devised to investigate the impact of the clustering methods on the Renewable Energy Sources integration into the DES design.

In light of the results obtained from the comparison of the clustering approaches in the first three scenarios, only k-means and k-MILP typical and extreme days are considered for Scenario 4UC calculations. **Errore. L'origine riferimento non è stata trovata.** reports the details of the energy conversion and storage systems selected in each scenario for the considered clustering approaches and their normalized sizes, when optimizing the total annual costs. The normalization has been performed with respect to: (i) the peak heating demand for boilers, heat pumps, and the heat capacity of internal combustion engines, SH and thermal storage; (ii) the peak cooling demand for refrigeration cycles; (iii) the peak electricity demand for PV and the electrical capacity of internal combustion engines. As it can be seen in the table, the optimal design of the MES for the Campus is essentially centralized: energy conversion units are installed only in sites S2 and S5. **Errore. L'origine riferimento non è stata trovata.** reports the annualized capital costs, the operating costs estimated in the design MILP problem on the basis of the selected typical and extreme days, the actual operating costs assessed with a yearly optimization of the system operation using the original data set, the total annual cost (TAC) estimated by the MILP problem using the typical/extreme days (TAC TDs), and the actual TAC, assessed with a yearly optimization of the system operation using the original data set (TAC 365 days).

In scenario 1UC, the k-means and k-medoids indicate that the optimal design must feature 3 boilers (a large one covering approximately 54-59% of the peak heating demand, two small size ones to be used during the mid-season when space heating is considerably lower), a heat pump in site 2 (covering around 14% of the peak heating demand) a large ICE (covering between 40-70% of the peak electricity demand) and 3 refrigeration cycles (two large ones covering 45-50% of the demand and a smaller one). According to the k-MILP extreme/typical days, increasing the capacity of one of the two small boilers is more profitable than installing a heat pump in site 2. Looking at the actual TAC (365 days operation) in **Errore. L'origine riferimento non è stata trovata.**, from an economic point of view this design choice does not bring about a significant difference (1% TAC decrease is within the MIP gap of the solver).

Results reported in **Errore. L'origine riferimento non è stata trovata.** show that the designs found with the three different sets of typical days feature some differences in the capital costs and operating costs but, in the end, they achieve the same total annual cost and guarantee no outages. A common feature of all three designs is to help the large size boiler to meet the peak of heating demand by using the heat storage system up to the maximum size allowed. The storage system is also used during the typical days to allow a more efficient and profitable management of the internal combustion engine.

In scenarios 2UC and 3UC, the optimal designs exploit the possibility of making outages to save some capital costs. For all the three sets of typical and extreme days, the optimal design reduces the capacity of installed boilers. In scenario 2UC, the designed system is not able to meet about 0.5-0.9% of the requested heat demand, while in scenarios 3UC, due to the lower outage fees, this fraction rises to 2.8-4.4%. Also, in these two scenarios the optimal solution essentially resembles the centralized paradigm to exploit economies of scale. Even if outages are allowed and not expensive, renewable technologies (solar PV and solar thermal panels) are not selected in most approaches because of their higher costs compared to conventional fossil-fired technologies (boilers and ICEs).

In scenario 3UC the set of typical days selected with the k-medoids leads to a very poor design which does not feature any refrigeration cycle to meet the cooling demand. This is due to the fact that the selected typical days underestimate the cooling demand (since it occurs only in a few months of the year), so the design MILP decides not to install any refrigeration cycle and to pay the outage fees during the extreme days. Actually, the outage occurs during the whole summer and this leads to very high outage costs. This issue is overcome by using the proposed MILP clustering approach, as shown in **Errore. L'origine riferimento non è stata trovata.**.

In scenario 4UC, in order to comply with the limit on the net $CO_2$ fossil emissions, the DES features for both the k-means and the k-MILP a very large share of PV to cover as much electricity demand as possible during the sunny days. Moreover, if compared to Scenario 1UC, a smaller fraction of heating demand is covered by the boilers (capacity 47-54% of the peak demand) in favor of a second internal combustion engine that enables a larger cogeneration along the year. These different designs entail a much larger capital investment cost, with respect to Scenario 1UC (more than doubled), which is not sufficiently counterbalanced by the lower operational expenditures (lower fuel consumption). As far as the comparison between k-means and k-MILP is concerned, both are able to approximate the actual yearly operating costs with good accuracy, while also avoiding outages. In terms of costs, for this last scenario (4UC), k-means performs slightly better than k-MILP.

In general, for the University Campus case study, k-means and k-MILP have very similar performance in representing the operating costs of the whole year. In this large-size case study, the averaging effect of k-means does not affect its representation accuracy since the input demand profiles are quite smooth. Nevertheless, the extreme days selected by the k-MILP allow decreasing the number and extent of the outages.

## 5.2. Single Building

Since the single building does not feature the cooling demand, that is covered by single electric compression chillers, only five Extreme Days have been considered for this case study: heating demand, electricity demand, ambient temperature, electricity prices, and sun irradiation. In light of the poor performances reported by the k-medoids in the previous case study, only k-means and k-MILP have been considered in this one.

### 5.2.1 Clustering results

Fig. 6 shows the normalized plots of the typical and extreme days selected by the two approaches.

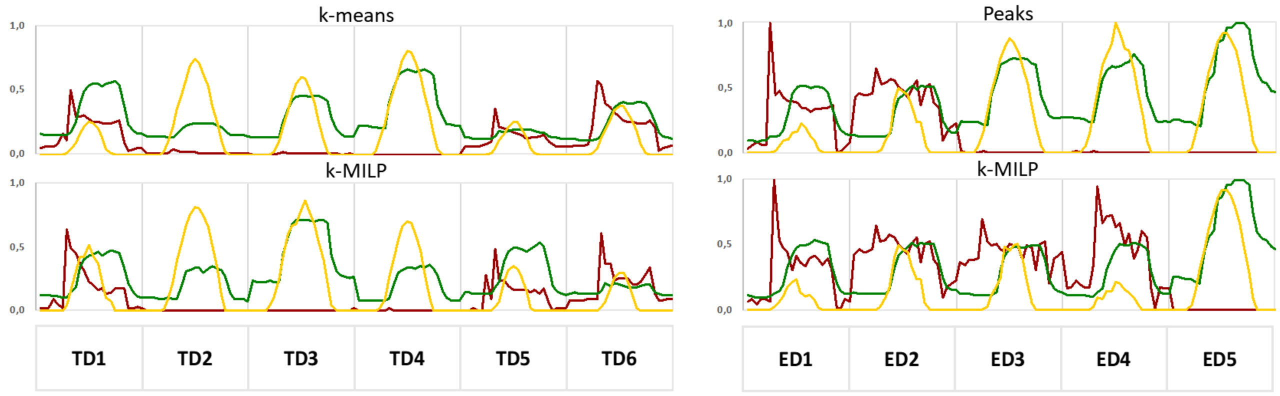
The extreme days selected by the k-means method are those in which the peaks of the heating demand (ED1), the electricity price (ED2), the sun irradiance (ED3), the ambient temperature (ED4), and the electricity demand (ED5) are reached. Concerning the extreme days selected by the k-MILP clustering method, the analysis of the results indicates that:

- ED1 features the peak heating demand;
- ED2 features a high electricity purchase price (not visible in the plot);
- ED3 is identified as “a-typical” because the heating demand does not get close to zero in the evening hours, unlike in the other typical and extreme days;

- ED4 features a very high heating demand together with a high average value throughout the day: such extreme days would turn out to be useful in avoiding the under-sizing of the thermal units (energy conversion and storage systems together);
- ED5 features a peak electricity demand.

It is worth noting that four out of five EDs selected by the k-MILP clustering method features high or relatively high heating demand, while the k-means extreme days include three days with null heating demand.

Fig. 7 reports the comparison between the original and aggregated LDCs. The corresponding errors and RMSD are reported in **Errore. L'origine riferimento non è stata trovata.**. As it can be seen, the two methods have similar performances: k-means is generally characterized by a lower percentage error (by definition, its typical days feature average profiles) but for the heating demand and the electricity purchase price the RMSD of the k-MILP is smaller. The only attribute for which we can observe an appreciable error is the global tilted radiation, where the k-MILP overestimates by 8% the available irradiance. This could be decreased by setting a lower bound on the approximation error of the radiation LDC, but it was not considered relevant for the case study due to the limited area available for solar heating and photovoltaics panels.




(a) (b)

*Fig. 6 Office building – Normalized profiles of (a) typical days and (b) extreme days for the two clustering techniques. The profiles are: red = heating demand, yellow = irradiance and green = electricity demand.*

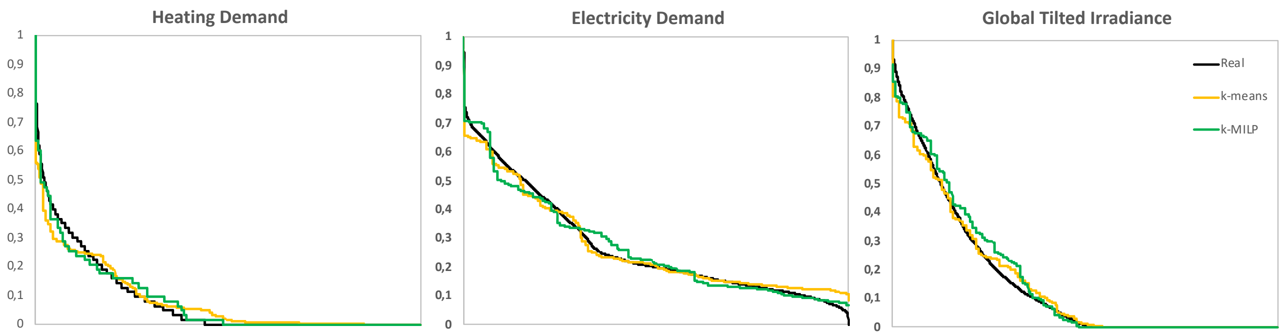


*Fig. 7 Office Building – load duration curves of the two energy demands.*

### 5.2.2 Design optimization results

The MILP design problem has been solved considering the typical and extreme days returned by the two clustering methods. Three scenarios have been considered:

- Scenario 1OB (OB stands for Office Building): outages are not allowed (the heat and cooling demand must always be satisfied by the installed units).
- Scenario 2OB: outages not allowed yet it is not possible to install cogeneration units (e.g., neither ICE nor micro-ICEs).
- Scenario 3OB: outages not allowed, cogeneration units can be installed (as in 1OB) and a limit is imposed on the net fossil $CO_2$ emissions of the DES, equal to 75% of the value reported for Scenario 1OB.

**Errore. L'origine riferimento non è stata trovata.** reports the details of the energy conversion and storage systems selected in each scenario when optimizing the total annual costs using typical and extreme days from the two clustering approaches. Like before, sizes have been normalized with respect to: (i) the peak heating demand for boilers, the thermal capacity of internal combustion engines and biomass-fired Organic Rankine Cycles (ORCs), SH, and thermal storage, and (ii) the peak electricity demand for the electric capacity of internal combustion engines and biomass-fired ORCs. **Errore. L'origine riferimento non è stata trovata.**, instead, reports annualized capital costs, operating costs estimated in the design MILP problem on the basis of the selected typical and extreme days, actual operating costs assessed with a yearly optimization of the system operation, total annual cost (TAC) estimated by the MILP problem using the typical/extreme days, and actual TAC (assessed with a yearly optimization of the system operation).

Both in scenario 1OB and 2OB, the use of the typical and extreme days coming from the k-means and the k-MILP leads to the selection of the same types of units yet featuring different sizes. In particular, in scenario 2OB, 43-58% of the heating demand is satisfied by two boilers (a larger one and a smaller one), which are helped, during the peak hours, by a heat storage system. Whereas, in scenario 1OB, for both clustering methods, the system takes advantage of cogeneration by reducing the size of the larger boilers (10-23% of the peak, instead of 33-49%) and installing a micro-ICE, that covers around 30% of the peak.

Looking at the actual TAC (365 days operation) in **Errore. L'origine riferimento non è stata trovata.**, from an economic point of view, the design choices made when using the k-MILP typical and extreme days brings about a significant difference with respect to the k-means: -20% in Scenario 2OB and -30% in Scenario 1OB. This is mainly due to the extreme days selected by the k-MILP, which turns out to be more "challenging" in terms of heating demand with respect to those of the k-means. Indeed, ED4 of k-MILP (featuring a high heat demand for several hours) guides the design optimization towards bigger capacities of the boiler and internal combustion engine and smaller heat storage systems. The under-sized ICE and boilers coming from the use of the k-means extreme days would turn out to be insufficient to cover the demand causing outages and less efficient operation throughout the year. Another advantage of k-MILP over k-means is the preservation of the intra-hour variability of the input profiles within the typical days, which is smoothed by the averaging effect of k-means.

The aforementioned over-estimation of the global tilted radiation associated to the k-MILP typical days leads to the installation of SH panels both in scenario 1OB and 2OB when its typical and extreme days are considered; however, as already seen, such choice does not penalize the economic figures appreciably.

Scenario 3OB is quite different: it features the installation of renewable energy sources to meet the limits imposed on the fossil $CO_2$ emissions. Namely, for both the k-means and k-MILP typical and extreme days, a biomass-fired Organic Rankine Cycles is installed. Even if the smallest possible size of the units is selected, yet they are over-sized for the heating demand of the users. The selection of the ORCs is due to the fact that the available area on the building roof would not be enough to meet the $CO_2$ emissions limits and solar PV panels would also require the installation of a heat pump to generate the required heat with a considerable increase in cost.

For this scenario, the differences in the costs are limited since the design choices have been forced in both cases by the $CO_2$ limits, but it is worth pointing out that k-MILP case selects a simpler configuration with respect to the k-means, replacing the boiler with a larger heat storage.

## Conclusions

The proposed MILP-based clustering approach significantly improves the accuracy in reproducing the load duration curves and the total yearly values of the relevant attributes compared to the classic k-medoids approach. In addition, it automatically identifies extreme days that cannot be well clustered because atypical compared to the representative days of each cluster. The analysis of the results for two real-world case studies (a university Campus and a single building) show that these days feature either very high/very low attributes or atypical combinations of attributes (e.g., high cooling demand with very low solar radiation) or high values maintained for several hours. All these periods could be critical for the operation of energy systems.

For the University Campus case study, k-means and k-MILP have very similar performance in representing the operating costs of the whole year, while k-medoids considerably underestimates the cooling duration curve. In this large-size case study, the averaging effect of k-means does not affect its representation accuracy since the input demand profiles are quite smooth. Nevertheless, the extreme days selected by the k-MILP allow decreasing the number and extent of the outages. Among the extreme days, k-MILP correctly identifies as a-typical a day featuring a very high cooling demand and low radiation which would be critical for systems relying on solar photovoltaic panels to drive the refrigeration cycles.

For the single building case study, the advantages of k-MILP over k-means become considerable in terms of both costs (up to -30% total annual cost) and reliability of the optimized designs. This is mainly due to the extreme days selected by the k-MILP, which turns out to be more “challenging” in terms of heating demand with respect to those of the k-means. In particular, k-MILP is able to find as extreme day one featuring a high heat demand for several hours which would be critical for systems relying on the use of storage units to meet the peak demand. Another advantage of k-MILP over k-means is the capability of preserving the hourly fluctuations of input profiles in the typical days.

In conclusion, k-MILP appears to be a valid alternative to well-known clustering techniques for its capability of inheriting the advantages of k-medoids (accuracy of the hourly profiles) and k-means (accuracy of the integral of the aggregated load duration curve) while automatically identifying extreme (and a-typical) operating periods which would not be easily identified on the basis of a-priori defined criteria.

## Acknowledgments

The authors acknowledge SIRAM for the useful techno-economic information and the key data of the case study. This work was supported by the “Efficity - Efficient energy systems for smart urban districts” project (CUP E38I16000130007) co-funded by Regione Emilia-Romagna through the European Regional Development Fund 2014-2020.

## References

[1] Weber C., Maréchal F., Favrat D., (2007), Design and optimization of district energy systems, Computer Aided Chemical Engineering, 24, 1127–1132.

[2] Mancarella P., (2014). MES (multi-energy systems): An overview of concepts and evaluation models, Energy, 65, 1–17.

[3] Yokoyama R., Hasegawa Y., Ito K., (2002), A MILP decomposition approach to large scale optimization in structural design of energy supply systems, Energy Conversion and Management, 43, 771–790.

[4] Zatti M., Martelli E., Amaldi E., (2017), A three-stage stochastic optimization model for the design of smart energy districts under uncertainty, In A. Espuña, M. Graells, & L. Puigjaner (Eds.), 27th European Symposium on Computer Aided Process Engineering, 40, 2389–2394

[5] Mitra, S., Pinto, J. M. and Grossmann, I. E. (2014) Optimal multi-scale capacity planning for power-intensive continuous processes under time-sensitive electricity prices and demand uncertainty. Part I: Modeling, Computers & Chemical Engineering. Elsevier Ltd, 65, pp. 89–101. doi: 10.1016/j.compchemeng.2014.01.016.

[6] Gabrielli P., Fürer F., Mavromatidis G., Mazzotti M., (2019), Robust and optimal design of multi-energy systems with seasonal storage through uncertainty analysis, Applied Energy, 238, 1192-210. doi:10.1016/j.apenergy.2019.01.064

[7] Bahl B., Söhler T., Hennen M., Bardow, A., (2018), Typical Periods for Two-Stage Synthesis by Time-Series Aggregation with Bounded Error in Objective Function, Frontiers in Energy Research, 5(January).

[8] Kotzur L., Markewitz P., Robinius, M., Stolten D., (2018), Impact of different time series aggregation methods on optimal energy system design, Renewable Energy, 117, 474–487.

[9] Fazlollahi S., Bungener S. L., Mandel P., Becker G., Maréchal F., (2014), Multi-objectives, multi-period optimization of district energy systems: I. Selection of typical operating periods, Computers & Chemical Engineering, 65, 54–66.

[10] Elsido C., Bischi A., Silva P., Martelli E., (2017), Two-stage MINLP algorithm for the optimal synthesis and design of networks of CHP units, Energy, 121, 403–426.

[11] Schütz T., Schraven MH., Fuchs M., Remmen P., Müller D., Comparison of clustering algorithms for the selection of typical demand days for energy system synthesis. Renew Energy 2018, 129-570, doi:10.1016/j.renene.2018.06.028.

[12] Teichgraeber H, Brodrick PG, Brandt AR. Optimal design and operations of a flexible oxyfuel natural gas plant. Energy 2017;141:506–18. doi:10.1016/j.energy.2017.09.087

[13] Stadler P., Ashouri A., Maréchal F., (2016), Model-based optimization of distributed and renewable energy systems in buildings, Energy and Buildings, 120, 103–113.

[14] Pfenninger S., Dealing with multiple decades of hourly wind and PV time series in energy models: A comparison of methods to reduce time resolution and the planning implications of inter-annual variability. Appl Energy 2017;197:1–13. doi:10.1016/j.apenergy.2017.03.051.

[15] Domínguez-Muñoz F., Cejudo-López J. M., Carrillo-Andrés A., Gallardo-Salazar M., (2011), Selection of typical demand days for CHP optimization, Energy and Buildings, 43, 3036–3043.

[16] Bahl B., Lützow J., Shu D., Hollermann DE., Lampe M., Hennen M., et al. Rigorous synthesis of energy systems by decomposition via time-series aggregation. Computer Chemical Engineering 2018;112:70–81. doi:10.1016/j.compchemeng.2018.01.023

[17] Poncelet K., Hoschle H., Delarue E., Virag A. and Drhaeseleer, W. (2017) Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems, IEEE Transactions on Power Systems, 32(3), pp. 1936–1948. doi: 10.1109/TPWRS.2016.2596803.

[18] Birge J. R., Louveaux F., (2011), Introduction to Stochastic Programming (2nd ed.), Springer Series in Operations Research and Financial Engineering, Springer, New York.

[19] Lahdelma R., Hakonen H., (2003), An efficient linear programming algorithm for combined heat and power production, European Journal of Operational Research, 148, 141–151.

[20] Gabrielli P., Gazzani M., Martelli E., & Mazzotti M. (2017). Optimal design of multi-energy systems with seasonal storage, Applied Energy, in press.

[21] Hastie T., Tibshirani R., Friedman J., (2008), The elements of statistical learning: data mining, inference and prediction. 2 ed. Springer.

[22] Mathworks®, 2017. Statistics and machine learning toolbox, Matlab® v.2017b.

[23] Zatti M., Gabba M., Rossi M., Morini M., Gambarotta A. and Martelli E. (2018), Towards the optimal design and operation of multi-energy systems: the Efficity project, Environmental Engineering and Management Journal, 17(10), pp. 2409–2419.

[24] Gambarotta A., Morini M., Rossi M., Stonfer, M., (2017), A Library for the Simulation of Smart Energy Systems: The Case of the Campus of the University of Parma, Energy Procedia, 105, 1776–1781.

[25] F. Bianchi, M. Altomonte, M. E. Cannata, G. Fasano, (2009), Definizione degli indici e livelli di fabbisogno dei vari centri di consumo energetico degli edifici adibiti a scuole - consumi energetici delle scuole primarie e secondarie, ENEA-Ricerca Sistema Elettrico.

[26] Italian power exchange (GME) website, http://www.mercatoelettrico.org/it/

[27] IBM ILOG, 2017. CPLEX MILP solver, v.12.7.0.0