HOW CLUSTERING APPROACHES AFFECT THE OPTIMAL DESIGN OF FUTURE MULTI-ENERGY SYSTEMS

Alessandro Pampado[1], Gabriele Volpato[1]*, Andrea Lazzaretto[1]

[1]*University of Padova, Industrial Engineering Department*
Via Venezia 1, 35131 Padova, Italy
*Corresponding Author: gabriele.volpato.1@phd.unipd.it

ABSTRACT

This paper focuses on the correct assessment of the total cost and sizes of the energy conversion and storage units of multi-energy systems operating under uncertain energy market conditions. The objectives are to: *i)* find the best set of representative days of uncertain electricity and natural gas prices, global solar irradiance and air temperature, with a special focus on prices, influenced by unpredictable socio-economic events; *ii)* assess the impact of the representative days, including the extreme ones, on the optimal total cost and unit sizes of a multi-energy system. An available historical dataset (2010-2022) of time series is divided by shifting the "present" moment back to 2018 to use the previous period as "past" training dataset (2010-2018) and the subsequent period as two (independent) "future" testing datasets, featured with different price stability, corresponding to a pre- (2019-2020) and a post- (2021-2022) Russia-Ukraine war scenario, respectively. The novel methodology consists in comparing the "*annual*" and "*seasonal*" *clustering* approaches to obtain, respectively, sets of representative days of the uncertain variables in the entire training dataset or in the training dataset divided into seasons, also considering different criteria to evaluate the extreme days of electricity price and thermal demand. A Mixed-Integer Linear Programming model of the system is used to optimize the sizes and operation of its units, minimizing the total investment and operational costs in the training dataset. Subsequently, the optimal sizes are fixed to optimize the operation in the two testing datasets. The optimal total cost and sizes are compared with "perfect knowledge" solutions obtained considering all the time series really occurred in the two testing datasets. Key results highlight that the error of the optimal total cost with respect to the "perfect knowledge" solutions is about 2.5-4% with clustering, compared to 9-17% with the "state-of-the-art" hourly profiles averaged on months or seasons, respectively. This error is higher using *seasonal clustering* than *annual clustering* for a number of representative days higher than 12. Moreover, the extreme days of electricity price do not bring a relevant gain in the solution accuracy because they are very similar to the typical ones in the pre-war scenario and their weight is small in the post-war scenario, given the higher values of the real prices in the period 2021-2022. In contrast, the extreme days of thermal demand are necessary to guarantee a feasible solution.

Keywords: Multi-energy systems, Clustering, Design-operation optimization

1. INTRODUCTION

1.1 Literature review

Uncertainty can be defined as "any deviation from the unachievable ideal of completely deterministic knowledge of the relevant system" (Walker *et al*., 2003). Clearly, several aspects of an energy system are affected by uncertainty: climatic conditions, prices of energy carriers, energy demand, technical performances and investment costs, as suggested by Zhang *et al*. (2019) and Mavromatidis *et al*. (2018a). Several approaches are developed to deal with the optimization of energy systems under uncertainty. Zhang *et al*. (2019) carried out a sensitivity analysis to study the impact of twelve uncertain parameters on the convenience of a Distributed Energy system (DES) with respect to a conventional energy system (CES), where the grid is the only energy source. Mavromatidis *et al*. (2018a) investigated which uncertain parameters are the most relevant for the design and operation of Distributed Energy Systems (DES), performing global sensitivity analysis by using Morris Sobol methods. Gomes *et al.* (2021) developed a model for management and operation planning of a microgrid, modelling it as a two-stage stochastic problem.

https://doi.org/10.52202/077185-0127

Ghaemi *et al*. (2021) used a two-stage stochastic programming model to optimize a district energy system. Golombek *et al.* (2022) built a multi-stage stochastic model to see the role of transmission and energy storage in Europe in 2050. Chen *et al*. (2019) used a two-stage robust optimization model to optimize an energy hub under uncertainties. Franke *et al.* (2020) compared deterministic and stochastic solutions obtained from a multi-objective optimization of a distributed energy system under uncertainty, finding that stochastic solutions are more robust. However, according to Zhou *et al*. (2013), stochastic programming does not lead to significant solution improvements compared to a deterministic optimization approach, so a deterministic model for the design is preferred due to the computational efficiency. In the previous cases, the complexity of the developed models for the design and operation of energy systems is crucial. In fact, computational burdens limit the optimization process, forcing to aggregate input data and, consequently, losing quality on the optimal solution (Kotzur *et al*., 2018). Such a solution must be accurate with respect to a reference optimal one, without entailing an unaffordable computational time. A way to address this issue is to carry out the optimization based on the representative periods (in this paper, these are days) instead of the full-scale optimization, i.e., considering all days of the input datasets. As highlighted by Mavromatidis *et al*. (2018b), solving the optimization based on the representative periods corresponds to a deterministic solution approach.
One of the most common ways to generate such representative days, excluding Monte Carlo simulation (Ghaemi *et al*., 2021) or discrete approximation from probability density functions (Liu *et al.,* 2022), is to use clustering techniques, which allow reducing the computational effort by aggregating similar days. A perfect clustering process should reduce the computational time by orders of magnitude while representing all information of the original dataset (Teichgraeber and Brandt, 2019). Different clustering algorithms could be used, as presented by Paparrizos and Gravano (2015), Merrick (2008) or Zatti *et al*. (2019), who developed k-MILP, a clustering approach derived from k-medoids to generate typical and a-typical days. However, k-means is one of the most commonly used algorithms: Fazlollahi *et al.* (2014) used k-means in the study of multi-objective optimization of distributed energy systems, while De Lima *et al*. (2021) chose it to generate scenarios for two-stage stochastic models. Despite this, the literature lacks a definition of the most appropriate algorithms for the design of energy systems, although comparisons are presented in Teichgraeber and Brandt (2019) or Pfenninger *et al*. (2017). Furthermore, authors adopted different approaches to represent the input dataset, including extreme scenarios or seasonal clusters (Kotzur *et al*., 2018), without an exhaustive explanation of why they do so. On the other hand, this paper focuses on the design and operation optimization of a multi-energy system under the uncertainty of unpredictable variables, highlighting the impact of the different clustering approaches on the optimal design, i.e., the sizes of components, of future multi-energy systems.

1.2 Goals and novelty
The following gaps have been identified in the literature:

- A clear explanation of the benefits of using a specific clustering approach in the optimal design of a Multi-Energy System (MES). Often, results do not highlight the need of checking accuracy of obtained solutions by using really occurred data.
- Possible benefits of applying clustering to seasonal datasets, i.e., generating seasonal clusters. Indeed, some authors (De Lima *et al*., 2021) used them without motivating this choice.
- A clear explanation of the advantages of including extreme days in the set of representative days found by clustering.

This paper aims to fill in the above-mentioned gaps in the design and operation optimization of a local MES that meets an aggregated residential energy demand. The goals are to *i)* find the best set of representative daily time series of uncertain electricity and natural gas prices, global solar irradiance and air temperature, and *ii)* assess the impact of these time series, including the extreme ones, on the optimal total cost and unit sizes of a MES.
The methodology is based on splitting an available historical dataset (2010-2022) by shifting the "present" moment back to 2018 to use the previous period as "past" training dataset (2010-2018) and the subsequent period as two independent "future" testing datasets, corresponding to a pre- (2019-2020) and a post- (2021-2022) Russia-Ukraine war scenario, respectively. The *annual* and *seasonal clustering* approaches are applied to obtain representative days of electricity and natural gas prices, global solar

https://doi.org/10.52202/077185-0127

irradiance and air temperature in the training dataset. The *annual* approach identifies representative days using the entire training dataset, whereas the *seasonal* approach first divides the training dataset into seasons and then identifies representative days for each season. A Mixed-Integer Linear Programming (MILP) model of the MES, solved for the representative days found by clustering, is used to optimize the sizes and operation of its units by minimizing the total investment and operational costs in the training dataset. Subsequently, the optimal design configurations of the system are tested (i.e., fixed) in the pre- and post-war scenarios. The optimal total cost found in the two scenarios and sizes of the units are compared with "perfect knowledge" solutions, obtained using all the time series occurred in these scenarios. The advantages of the seasonal clustering approach and the criteria to include extreme days are also discussed. Figure 1 shows the steps of the methodology proposed.

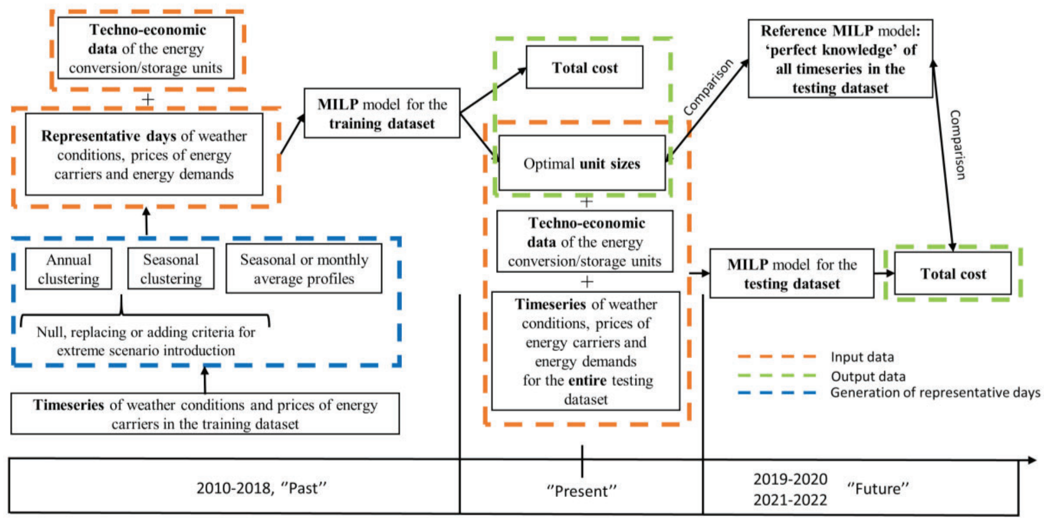

Figure 1: Flowchart of the analysis of the MES.

The novelties of this work can be summarized as follows:

1. The application of the *annual* and *seasonal clustering* approaches to obtain different sets of representative days. Their impact on the optimal cost and unit sizes of a MES is assessed in comparison to the state-of-the-art approach (which considers average seasonal or monthly profiles). Contrary to the main literature, which presents an analysis of several clustering algorithms (e.g., k-means, k-medoids, etc. (Pfenninger *et al*., 2017)), this paper focuses on the comparison between the *annual* and *seasonal clustering* approaches. Extreme days of electricity price and thermal demand are taken into account, as proposed by Kotzur *et al*. (2018), by i) *null* (extreme days are neglected), ii) *replacing* (extreme days replace the representative days of their belonging clusters), or iii) *adding* (extreme days become the representative days of additional clusters) criteria.
2. The evaluation of the accuracy of the optimal system cost and unit sizes in two different testing datasets, reflecting the increased variability of prices in Italy from the period before the Russia-Ukraine war (i.e., 2019-2020) to the subsequent period (i.e., 2021-2022).

2. METHODOLOGY

2.1 Optimization model of the residential multi-energy system

Figure 2 shows the residential Multi-Energy System (MES) under analysis. The MES includes a photovoltaic (PV) plant, an Electrical Energy Storage (EES) unit, a cogeneration (CHP) unit, a natural gas-fired boiler and a Thermal Energy Storage (TES).

The optimization aims at minimising the total cost (investment and operational costs) of the system with an hourly resolution. The optimization model is formulated as a Mixed-Integer Linear Programming (MILP) and includes the following decision variables: design continuous and binary variables representing the sizes of the units and their inclusion/exclusion in the design configuration,

https://doi.org/10.52202/077185-0127

respectively; operational continuous and binary variables representing the energy flows of dispatchable conversion units/storage units and their on/off status, respectively.

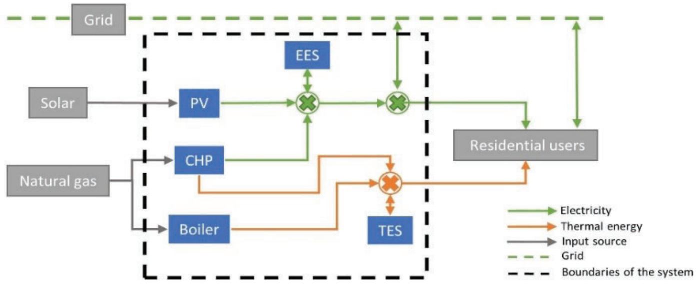

Figure 2: Residential multi-energy system under analysis.

In the following equations, index *t* represents the total time, including the hour *h*, day *d* and year *y*. The constraints of the MILP model include the characteristic curves of the energy conversion units, the energy balances and the techno-economic constraints. For reasons of space, only the most relevant ones are presented. The electrical power generated by PV is:

$$P_{PV,t} = A_{PV} * G_{tot,t} * \eta_{PV,t} * \eta_{BOS} \tag{1}$$

where A_{PV} is the area of the PV system, $G_{tot,t}$ is the global solar irradiance on the tilted plane, $\eta_{PV,t}$ is the efficiency of the PV system, which depends on the air temperature and η_{BOS} is the balance of system efficiency. The characteristic curves of CHP are obtained through interpolation from data given by manufacturers:

$$F_{CHP,t} = 0{,}2722 * P_{CHP,t} + 3{,}7125 * \delta_{CHP,t} \tag{2}$$

$$Q_{CHP,t} = 1{,}7246 * P_{CHP,t} + 9{,}3109 * \delta_{CHP,t} \tag{3}$$

$$\delta_{CHP,t} \leq \delta_{CHP,inv} \tag{4}$$

where $P_{CHP,t}$ and $Q_{CHP,t}$ are the power and heat generated, while $F_{CHP,t}$ is the fuel consumed. $\delta_{CHP,t}$ and $\delta_{CHP,inv}$ are binary variables representing the on/off status of the CHP and its inclusion/exclusion in the design configuration, respectively.
The fuel consumption of the boiler is:

$$F_{boiler,t} = Q_{boiler,t}/\eta_{boiler} \tag{5}$$

where $Q_{boiler,t}$ is the heat produced and η_{boiler} is the efficiency.
The constraints of the EES are:

$$E_{EES,t} = E_{EES,t-1} * (1 - E_{selfdischarge}) + \eta_{EES} * P^{+}_{EES,t} - \frac{1}{\eta_{EES}} P^{-}_{EES,t} \tag{6}$$

$$SOC_{min} E_{EES,max} \leq E_{EES,t} \leq SOC_{max} E_{EES,max} \tag{7}$$

$$E_{EES,23,d,y} = 0.5 * E_{EES,max} \quad \forall\, d \in days, y \in years \tag{8}$$

$$0 \leq P^{+}_{EES,t}\ (P^{-}_{EES,t}) * \Delta t \leq const * E_{max,EES} \tag{9}$$

In Eq. (6) $E_{EES,t}$ is the energy stored at time *t*, $E_{EES,t-1}$ is the energy at the previous time *t-1*, $E_{selfdischarge}$ is the self-discharge capacity of the system, $P^{-}_{EES,t}$ ($P^{+}_{EES,t}$) is the energy discharged (charged) and η_{EES} is the EES round trip efficiency. In constraints (7), SOC_{min} and SOC_{max} are the minimum and maximum state of charge, while $E_{EES,max}$ represents the size of the EES. Eq. (8) indicates that the energy level $E_{EES,23,d,y}$ at the end of each day is set, implying only intra-day storage. This is required because each representative day is independent from each other. Alternatively, other approaches (Gabrielli *et al*., 2018) could be adopted. Constraints (9) bound the minimum and maximum charging (discharging) power, with Δt corresponding to the hourly resolution. The big M method is adopted to prevent the EES to charge and discharge simultaneously. The constraints of the Thermal Energy Storage (TES) are similar to those of the EES.

$$Q_{TES,t} = Q_{TES,t-1} * (1 - E_{selfdischarge}) + \eta_{TES} * Q^{+}_{TES,t} - \frac{1}{\eta_{TES}} Q^{-}_{TES,t} \tag{10}$$

https://doi.org/10.52202/077185-0127

The big M method is used to prevent the system to buy/sell energy from/to the grid simultaneously. The total electric energy balance is:

$$P_{PV,t} + P_{CHP,t} + P_{grid,purchase,t} + P_{EES,t}^{-} = P_{EES,t}^{+} + P_{grid,sell,t} + P_{load,t} \tag{11}$$

The total thermal energy balance is:

$$Q_{CHP,t} + Q_{boiler,t} + Q_{TES,t}^{-} = Q_{TES,t}^{+} + Q_{load,t} \tag{12}$$

The total cost, evaluated as sum between the investment cost (Eq. (14)) and the operational cost (Eq. (15)), is the objective function to minimize. For a full-scale optimization problem, it can be written as:

$$min\, f = min\left(Cost_{des} + \sum_{y=y_0}^{Y} \sum_{d=0}^{D} \sum_{h=0}^{H} Cost_{op,t} \right) \tag{13}$$

$$Cost_{des} = \frac{1}{\alpha} \sum_{u=1}^{U} (c_{fixed,u} * \delta_{inv,u} + c_{var,u} * C_u) \tag{14}$$

$$Cost_{op,t} = (F_{CHP,t} + F_{boiler,t}) * c_{naturalgas,t} + c_{grid,purchase,t} * P_{grid,purchase,t} - c_{grid,sell} * P_{grid,sell,t} \tag{15}$$

In Eq. (14), u refers to a specific unit and $1/\alpha$ is the amortization factor, assuming the system lifetime equal to 20 years. r is the interest rate, equal to 4%. $c_{fixed,u}$ and $c_{var,u}$ represent the fixed and the variable costs for a given unit, respectively. $\delta_{inv,u}$ is the binary variable describing the inclusion/exclusion of the unit, while C_u is its size. In Eq. (15), $(F_{CHP,t} + F_{boiler,t}) * c_{naturalgas,t}$ indicates the operational costs of CHP and boiler, where $c_{naturalgas}$ is the price of natural gas, $c_{grid,purchase,t} * P_{grid,purchase,t}$ is the cost for purchasing electricity from the grid and $c_{grid,sell} * P_{grid,sell,t}$ is the revenue obtained by selling energy to the grid.

In case N representative days are used in the optimization, the objective function is:

$$min\, f = min\left(Cost_{des} + \sum_{d=0}^{N} w_d * \sum_{h=0}^{H} Cost_{op,t} \right) \tag{16}$$

where w_d is the weight of the representative day d in the entire training dataset.

The accuracy of the optimal design solution found in the training dataset (2010-2018) is assessed in two distinct testing datasets, representing two scenarios of the "future": a predictable one with prices before the Russia-Ukraine war (2019-2020), and a more unpredictable one with post-war prices (2021-2022). Details are already indicated in Section 1.2. In summary, the following MILP optimization models are considered:

1. "Perfect knowledge" models solved in the testing datasets. A full-scale optimization (Eq. (13)) is carried out with all input time series really occurred in the testing dataset. Therefore, the optimized solution is the best one for such a period.
2. Deterministic models solved in the training dataset. The optimization is based on representative days (Eq. (16)) obtained from the training dataset. The different ways to generate representative days are explained in detail in the next Section 2.2.
3. Deterministic models solved in the testing dataset. The sizes of the units are fixed according to the optimized solution found in the training dataset, thus only a full-scale optimization of the operation of the system is performed.

2.2 Clustering approaches for the generation of representative days

Clustering approaches aggregate similar elements (i.e., temporal periods), identifying representative ones that can be used in an optimization problem to achieve a good compromise between solution accuracy and acceptable computational effort. In this work, periods correspond to days.

The main steps of a clustering approach are data normalization (to take into account variables that have different units of measurement), application of a clustering algorithm and selection of representative days, including the extreme ones (Teichgraeber and Brandt, 2019).

The k-means algorithm is applied to the "past" training dataset (2010-2018) to generate clusters and identify representative days of global solar irradiance, air temperature, electricity and natural gas prices. Demand profiles are not considered in the clustering process because the electrical one does not change from year to year and the thermal demand is already represented by the temperature (Section 2.3). The

https://doi.org/10.52202/077185-0127

representative day of each cluster is chosen as the one with the lowest value of the Euclidean distance from the centroid, i.e., the arithmetic mean of all periods assigned to a cluster, to avoid average profiles that could not correctly represent strong daily fluctuations. The weight of each representative day is the number of elements belonging to its cluster.

Another crucial aspect is the addition of extreme days, which could strongly affect the optimal design of energy systems. In this work, the extreme days are represented by the days with the highest and lowest daily sum of the hourly electricity prices, as well as by the day with the highest thermal demand to guarantee the feasibility of the system, so fulfilling all load requirements. To include these extreme periods in the representative set of days, the *null*, *replacing* and *adding* criteria (Kotzur *et al*., 2018) are considered. In the *null* criterion, extreme days are neglected. In the *replacing* criterion, they become the representative elements of the clusters they are assigned to. This criterion usually overestimates their weight in the objective function of the optimization problem. In the *adding* criterion, the extreme days become new centres of additional clusters after an iterative process that reassigns all the elements to the clusters. The influence of these extreme periods on the optimization results will be small, as only a few elements will be assigned to these new clusters, implying a lower weight in the objective function compared to the *replacing* criterion. Furthermore, three different approaches are considered to generate representative days. The *annual clustering* approach is applied to the entire training dataset (using all criteria to evaluate extreme days), varying the number of representative days from 3 to 28. The *seasonal clustering* approach differs from the previous one by first dividing the training dataset into seasons. The *state-of-the-art* approach, involving the generation of average seasonal or monthly profiles, is also applied. The extreme scenario of thermal demand is still added to guarantee the feasibility of the solution. The time required to generate the clusters is always less than one minute for all the approaches. Figure 3 shows four representative days obtained by the *seasonal* approach (one for each season) and three extreme days, obtained by the *adding* criterion, associated with the highest and lowest profiles of the electricity price (green and blue curves) and the highest thermal demand (red curve).

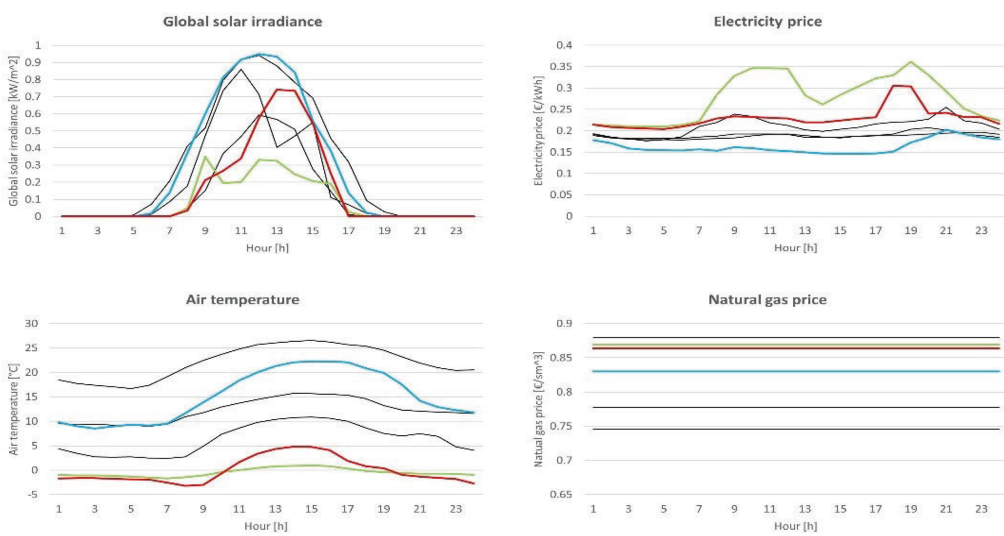

Figure 3: Four representative days of global solar irradiance, air temperature, electricity and natural gas prices and three extreme days of electricity price (green and blue) and thermal demand (red).

2.3 Input data

The weather data are global solar irradiance and air temperature in the period 2010-2020 for a PV system located in Padova (Italy) with the best inclination and azimuth angles. The weather time series in the testing dataset with more unpredictable prices (i.e., 2021-2022) are assumed to be the same as in the other testing dataset (i.e., 2019-2020), which is acceptable given the low variability of climatic conditions over the analysed years and the given demand curves, as explained in the following.

The electricity price, available in the period 2010-2022 (Terna, 2024) is calculated as:

https://doi.org/10.52202/077185-0127

$$c_{purchase,h,d,y} = PUN_{h,d,y} + T\&M_{d,y} + Charges_{d,y} + Taxes_{d,y} + constant\ [c€/kWh] \quad (17)$$

where $PUN_{h,d,y}$, $T\&M_{d,y}$, $Charges_{d,y}$, $Taxes_{d,y}$ and $constant$ are the day-ahead market price in Italy, the transport and management cost, the system charges, the taxes and a term to fit the prices to those of typical Italian families in such a period, respectively. This data vary quarter to quarter.

The price of natural gas is known for each quarter of the period 2010-2022 (Arera, 2024) and is calculated as:

$$c_{natural_{gas},h,d,y} = Raw_{material_{d,y}} + Infrastructure_{d,y} + Selling_{d,y} + Taxes_{d,y}\ \left[\frac{c€}{kWh}\right] \quad (18)$$

where $Raw_{material_{d,y}}$, $Infrastructure_{d,y}$, $Selling_{d,y}$ and $Taxes_{d,y}$ are the costs of raw materials, infrastructure, distribution and the taxes, respectively. The electrical and thermal demands, taken from an internal database, refer to a group of residential users. The electrical demand is assumed constant over the years, while the thermal demand varies over the years depending only on the external temperature.

Table 1: Techno-economic data of the energy conversion and storage units (DEA-Danish Energy Agency *(2023)*).

Technology	Quantity	Unit		Technology	Quantity	Unit	
PV	$c_{var,PV}$	€/kWp	1250	**EES**	$c_{var,EES}$	€/kWh	880,28
	$c_{fixed,PV}$	€	0		$c_{fixed,EES}$	€	3494,44
	A_{PV} / C_{PV}	m2/kWp	9,7		η_{EES}	.	0,87
	$\eta_{PV,std}$	-	0,136		$E_{selfdischarge}$	$\%E_{max,EES}/h$	0,04
	η_{BOS}	-	0,85		SOC_{min}	$\%E_{max,EES}$	0
CHP	$c_{var,CHP}$	€/kWel	1738,26	**TES**	$c_{var,TES}$	€/kWh	244,05
	$c_{fixed,CHP}$	€	32046		$c_{fixed,TES}$	€	968,52
Boiler	$c_{var,boiler}$	€/kWth	64,86		η_{TES}	.	0,7
	$c_{fixed,boiler}$	€	1622		$Q_{selfdischarge}$	$\%Q_{max,TES}/h$	2,1
	η_{boiler}	.	0,97				

3. RESULTS

3.1 Annual clustering

Figure 4 shows the relative errors of the optimal total costs found by solving the deterministic model in the pre- and post-war scenarios (with fixed unit sizes according to the optimal values found in the training dataset), compared to the “perfect knowledge” solutions (section 2.1), for the different criteria to consider extreme days (i.e., *null*, *replacing*, *adding*). Focusing on the *annual clustering* approach, these errors decrease rapidly when the number *N* of representative days (Eq. (16)) increases. The trend presents oscillating values without being strictly monotonic, which is coherent to the chosen clustering algorithm (Kotzur *et al*., 2018). Figure 5 shows the relative errors of the unit sizes, based on the representative days found by the *annual clustering* approach with the *adding* criterion in the training dataset, compared to the “perfect knowledge” solutions of the pre- and post-war scenarios.

Starting from the pre-war scenario (2019-2020), according to the “perfect knowledge” solution the sizes of PV, CHP, boiler and TES are 17.18, 41.05, 54.55 [kW] and 66.87 [kWh], respectively, while the EES is not installed. The optimal total cost is equal to 79.87 [k€] (21.68 [k€] for the design and 58.19 [k€] for the operation). Figure 4 highlights that it is not trivial to choose the best criterion to consider the extreme days. For example, the *adding* criterion should be considered the best as it gives the lowest error of the optimal total cost for 11 out of the 26 cases (i.e., each characterized by a number of representative days ranging from 3 to 28), but it also gives the highest error in 8 cases. Similarly, the *replacing* criterion should be considered the worst as it leads to the highest error for 11 out of 26 cases, but it also presents the lowest error in 5 cases. Furthermore, as the number of representative days increases, the error reaches values below 3%, so differences between the extreme day criteria become negligible. These small errors are justified by the repeatability of climatic conditions, the predictable prices and the inelastic electricity demand profiles. Moreover, the possibility of using only intra-day energy storage also reduces the error. As the representative days are independent, seasonal storage is not allowed and, to ensure a fair comparison, it is also neglected in the full-scale optimization.

https://doi.org/10.52202/077185-0127

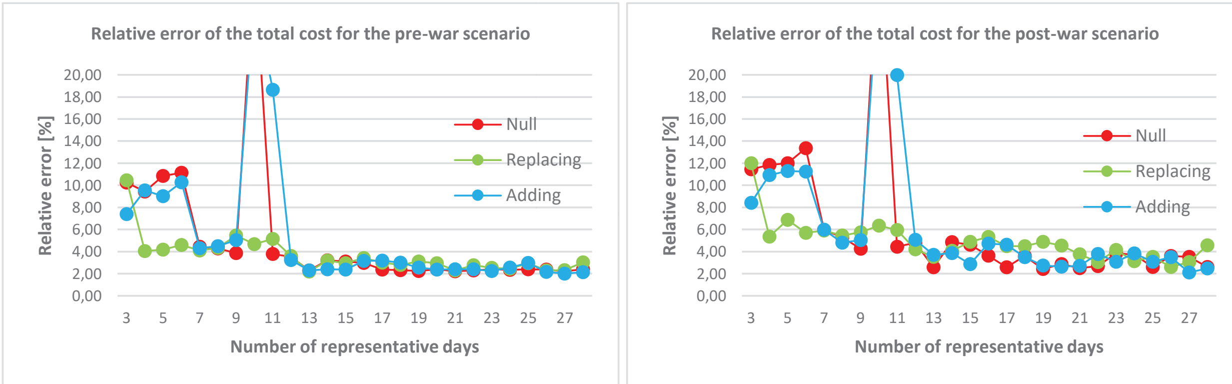

Figure 4: Relative errors of the optimal total cost in the pre-war scenario (on the left) and in the post-war scenario (on the right) for the *annual clustering* approach. The time required to get an optimal solution with a gap of 3% is less than one minute and around 540s for 3 and 28 representative days, respectively, for all the criteria.

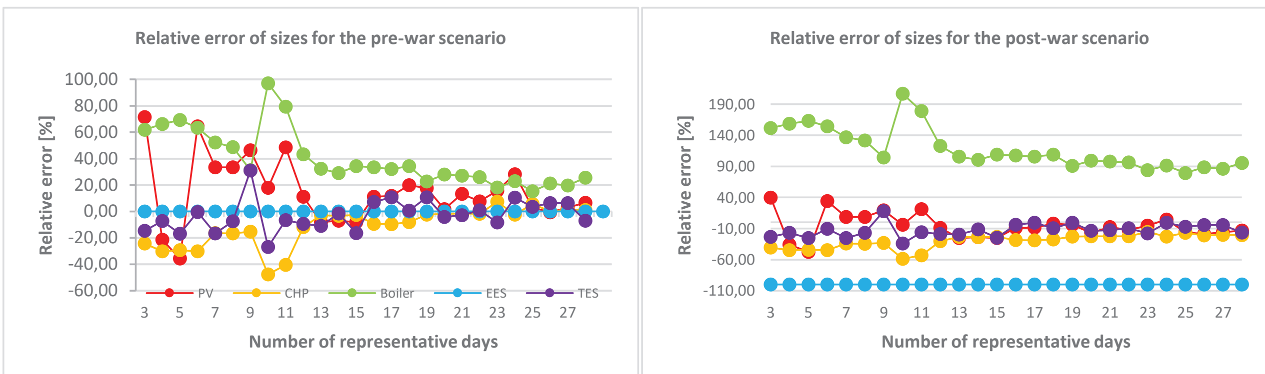

Figure 5: Relative errors of unit sizes in the pre-war scenario (on the left) and in the post-war scenario (on the right) for the *annual clustering* approach with the *adding* criterion.

Regarding the sizes (Figure 5), the error in the PV oscillates and does not have a clear trend when the number of representative days increases, for all the three criteria used to evaluate the extreme days. In general, the *adding* and *null* criteria overestimate the PV size, while the *replacing* criterion underestimates it. The optimal size of the CHP increases with the number of representative days, resulting in errors oscillating around +1.3%, +7.7% and +3% for the *null*, *replacing* and *adding* criteria, respectively. On the contrary, the optimal size of the boiler decreases with the number of representative days, resulting in an error in the range +20-40% for the three criteria. No optimal solution envisages the installation of the EES due to its high investment cost, which does not guarantee a relevant advantage throughout the operation, especially considering only intra-day storage. Finally, the average error of the TES size is around +4% for the three criteria, without a precise trend.

Focusing on the post-war scenario (2021-2022), the "perfect knowledge" solution has a total cost of 110.63 [k€] (29.26 [k€] for the design and 81.37 [k€] for the operation), which is 38.5% more expensive compared to the pre-war scenario. The optimal sizes are 21.05 [kW] for the PV (+22.38%), 52.11 [kW] for the CHP (+26.79%), 35.05 [kW] for the boiler (-35.69%), 26.56 [kWh] for the EES and 74.43 [kWh] for the TES (+11.2%). The inclusion of the EES (resulting in a higher total investment cost) is justified by the improved flexibility of the system, which reduces operating costs. The increase in the sizes of PV and CHP is related to the higher electricity prices, while the size of the boiler decreases as it is used only when the heat generated by the CHP (the size of which increases) is not sufficient to cover load requirements.

The trend in the relative error of the optimal total cost is similar to that found in the pre-war scenario (Figure 4). However, in the post-war scenario, the optimal total cost (error of +4% compared to the "perfect knowledge" solution) is on average 1.5% higher compared to the pre-war scenario. Although

https://doi.org/10.52202/077185-0127

the difference may seem small, it would be likely much higher if elastic electricity demands and a longer (than two years) optimization period could be considered.
The optimal PV size is strongly underestimated for the three criteria, especially for the *replace* and *adding* ones (-18,66% and -16,03% on average), because they underestimate the use of PV, while this post-war scenario would require its higher use to face the increase in electricity prices. The size of CHP is also underestimated (-25% on average for the three criteria), while the error for the boiler is higher than 100% for *replacing* (the most conservative criterion) and around 90% for *null* and *adding* criteria. The EES is never included as in the optimal deterministic solution of the pre-war scenario, whereas the TES size is always underestimated, especially for the *replacing* criterion (-15% on average).
The main outcome is that the introduction of the extreme days of electricity price does not guarantee an increase in the accuracy of the optimal solution, since such periods are extreme in the training dataset, but strongly lower compared to prices reached in the period 2021-2022.

3.2 Seasonal clustering and state-of-the-art approach

The use of the *seasonal clustering* approach implies the generation of *M* days for each season, resulting in a total of $N=4*M$ representative days.
In the pre-war scenario, Figure 6 shows that the relative error in the total cost with respect to the "perfect knowledge" solution is lower (compared to the *annual clustering* approach) only with N=8 or 12, N=8 and N=4 or 8 for the *null*, *replacing* and *adding* criteria, respectively. When the *annual clustering* approach leads to errors around 2.5%, *seasonal clustering* approach gives errors of 3.23%, 4.02% and 3.5% on average for the *null*, *replacing* and *adding* criteria. The reason for this is that seasons are artificial periods and, thus, their consideration does not bring an evident advantage in terms of higher solution accuracy, unless only climatic conditions are evaluated by clustering. All the three extreme day criteria overestimate the PV size, especially the *replacing* one (+26% on average). The CHP size is underestimated, with an error of -14% on average for all the three criteria, while the boiler size is overestimated (38% on average).
Considering the post-war scenario, as shown in Figure 6, the relative errors in the total cost according to the *annual* and *seasonal clustering* approaches are similar (around +4%). Considering all the three extreme day criteria, the PV size is close to the "perfect knowledge" value, but the CHP size is still underestimated, and the boiler size is overestimated (-32% for the CHP and +120% for the boiler, on average for all criteria). In both the pre- and post-war scenarios, the EES is never included in the optimal design configuration, while the error in the TES size fluctuates without a clear trend.

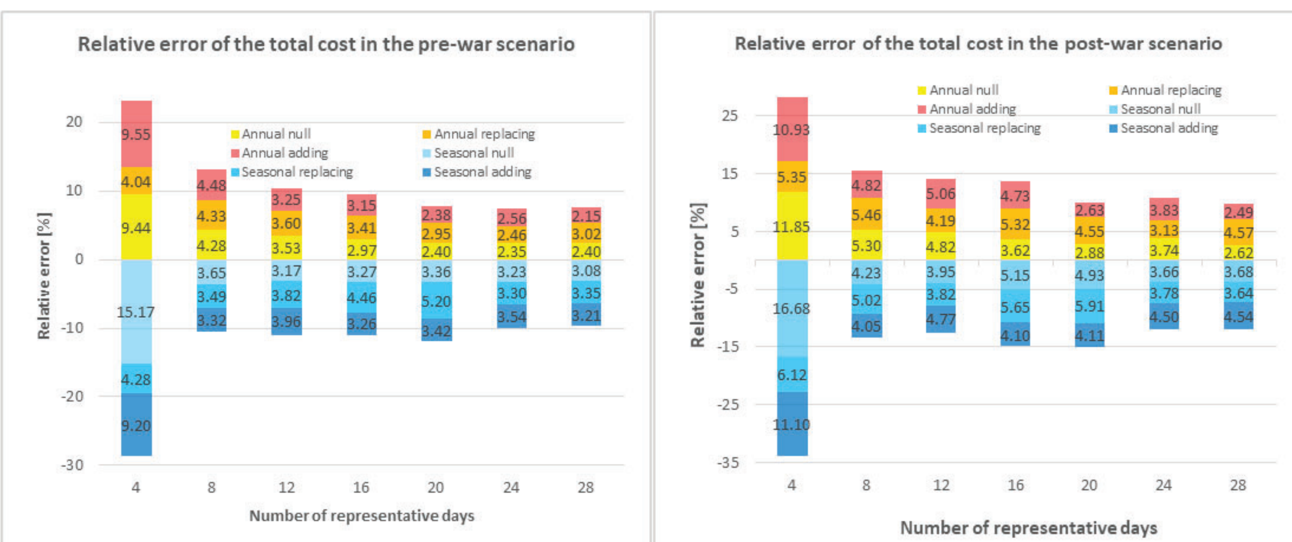

Figure 6: Relative error of the optimal total cost for the *annual* and *seasonal clustering* approaches, considering the *null*, *replacing* and *adding* criteria, in the pre- (left) and post-war (right) scenarios.

https://doi.org/10.52202/077185-0127

The state-of-the-art approach, based on average seasonal or monthly profiles for all the input time series, leads to higher errors in the optimal total costs and sizes compared to the *annual* and *seasonal clustering* approaches, in both pre- and post-war scenarios. In the pre-war scenario, the seasonal and monthly profiles result in relative errors in the optimal total cost of 16.30% and 8.75% compared to the “perfect knowledge” solution, respectively. The mean profiles strongly underestimate the PV size (-25.29% for seasonal and -20.29% for monthly profiles), because they overestimate its power production, and the CHP size (-38.10% for seasonal and -28.88% for monthly profiles). They overestimate the boiler size (90.73% and 64.15%) and underestimate the TES size (-41.61% and -6.50%), without including the EES. In the post-war scenario the errors in the optimal total cost are even higher, i.e., 17.82% for seasonal and 10.12% for monthly profiles. The errors in the sizes are -38.96% and -34.87% for PV, -51.18% and -43.91% for CHP, 196.62% and 155.26% for boiler and -47.52% and -15.55% for TES, according to the seasonal and monthly profiles, respectively.

4. CONCLUSIONS

This paper focuses on the design and operation optimization of a Multi-Energy System (MES) under the uncertainty of unpredictable variables, highlighting the impact of different clustering approaches on the optimal total cost and unit sizes. The objectives are to *i)* find the best set of representative days of uncertain electricity and natural gas prices, global solar irradiance and air temperature, with a special focus on prices, influenced by unpredictable socio-economic events (e.g., a war) and *ii)* assess the impact of such days, including the extreme ones, on the optimal total cost and unit sizes of a MES. A MILP model of the system, based on representative days of the “past” training dataset (2010-2018), is used to optimize the sizes and operation of its units by minimizing the total cost. The representative days are generated by *annual* or *seasonal clustering* approaches, including extreme scenarios with *null, replacing* or *adding* criteria. Subsequently, the optimal design solutions are tested in two distinct “future” datasets, characterized by different prices of the pre- (2019-2020) and post (2021-2022) Russia-Ukraine war scenarios. Eventually, the optimal total costs and sizes are compared with the “perfect knowledge” solutions based on data really occurred in the two testing datasets. The general guidelines obtained from this work are summarized in the following.

- Compared to the average profiles, which present an error about 9-17%, the representative days found by clustering allow obtaining a 2.5-4% error in the total costs with respect to the “perfect knowledge” solutions of the pre- and post-war scenarios. The best solution, in terms of total cost, is given by the *annual clustering* with *adding* criterion and 26 typical days, with a relative error of +2% and +2.11% for the pre- and post-war scenarios, respectively. In the period 2019-2020, the relative errors in the sizes of PV, CHP, boiler and TES are +3%, +1.33%, +19.6% and +6%, respectively. Instead, in the period 2021-2022, errors are higher (-15% for PV and CHP, +86% for the boiler).
- The *seasonal clustering* approach does not bring relevant advantages compared to the *annual clustering* approach, especially for a number of representative days higher than 12. In fact, seasons are artificial distinctions that do not reflect the real distribution of time series, such as the price of electricity or natural gas.
- The introduction of extreme days of electricity price, obtained in the training dataset, does not result in a higher accuracy in terms of cost or unit sizes. This is because such days are similar to the representative ones of the pre-war scenario and therefore their prices are much lower compared to those of the post-war scenario. However, the inclusion of the extreme days of thermal demand in the set of representative days is necessary to guarantee a feasible and optimal design of the system. In fact, since a thermal network is not conceived in this work, no optimal solution would be found if they were neglected.

In summary, the use of clustering techniques allows obtaining accurate solutions in comparison to the use of average profiles and this is proved by testing them with data really occurred. However, the inclusion of extreme days is not necessary for such a small system, regardless the variability of input data, unless feasibility problems arise. Future work will look at the temporal relation between representative days (here assumed independent) to properly model seasonal storage. More complex systems will also be analysed, including units such as heat pumps or hydrogen conversion/storage units.

https://doi.org/10.52202/077185-0127

NOMENCLATURE

Acronyms

CHP	Cogeneration system
EES	Electric energy storage
MES	Multi-energy system
PV	Photovoltaic system
SOC	State of charge
TES	Thermal energy storage

Greek letters

δ_{inv}	Binary variable associated with the inclusion of a given energy conversion unit
$\delta_{unit,t}$	Binary variable associated with on/off status of a given energy conversion unit at time t
η_{BOS}	Efficiency for the balance of system
η_{unit}	Efficiency of a given energy conversion unit

Symbols

A_{PV}	Area of the photovoltaic system $[m^2]$
C_u	Size of an energy conversion (storage) unit [kW] ([kWh])
c_u	Specific cost for an energy conversion (storage) unit [€/kW] ([€/kWh])
$E_{EES,t}$	Energy level for the electric energy storage [kWh]
$F_{unit,t}$	Fuel required by a given component at time t [sm^3/h]
G_{tot}	Global solar irradiance for the tilted plane [kW/m^2]
N	Number of representative days [-]
$P_{unit,t}$	Power produced by an energy conversion unit, or required by the load, at time t [kW]
P_t^+, P_t^-	Power charged in or discharged from the electric energy storage, at time t [kW]
$Q_{unit,t}$	Heat flow produced by an energy conversion unit, or required by the load, at time t [kW]
Q_t^+, Q_t^-	Heat flow charged in or discharged from the thermal energy storage, at time t [kW]
$Q_{TES,t}$	Level of energy in the thermal energy storage at time t [kWh]
w_d	Weight of a given representative day [-]

Subscripts

d	Day of the year
des	Design
fix	Fixed
h	Hour of a day
k	Cluster
op	Operation
t	Total time
var	Variable
y	Year

ACKNOWLEDGEMENT

Funder: Project funded under the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.3 - Call for tender No. 1561 of 11.10.2022 of Ministero dell'Università e della Ricerca (MUR); funded by the Euro-pean Union –NextGenerationEU. Award Number: Project code PE0000021, Concession Decree No. 1561 of 11.10.2022 adopted by Ministero dell'Università e della Ricerca (MUR), CUPC93C22005230007, according to attachment E of Decree No. 1561/2022, Pro-ject title "Network 4 Energy Sustainable Transition – NEST".

REFERENCES

ARERA, Autorità di Regolazione per Energia Reti e Ambiente, https://www.arera.it/it/inglese/index.htm. (Accessed February 2024).

Chen C., Shen X., Guo O., Sun H., 2019, Robust planning operation co-optimization of energy hub considering precise model of batteries' economic efficiency, Energy Procedia, Volume 158, February 2019, Pages 6496-6501.

https://doi.org/10.52202/077185-0127

DEA-Danish Energy Agency. Technology Data, 2023. Available: https://ens.dk/en/our-services/projections-and-models/technology-data

De Lima T.D., Tabares A., Banol Arias N., Franco J.F., 2021, Investment & generation costs vs CO2 emissions in the distribution system expansion planning: A multi-objective stochastic programming approach. International Kournal of Electrical Power and Energy Systems, Volume 131, 106925.

Fazlollahi S., Bungener S.L., Mandel P., Becker G., Marechal F., 2014, Multi-objectives, multi-period optimization of district energy systems: I. Selection of typical operating periods, Computers & Chemical Engineering Volume 65, Pages 54-66.

Franke G., Schneider M., Weitzel T., Rinderknecht S., 2020, Stochastic Optimization Model for Energy Management of a hybrid microgrid using Mixed Integer Linear Programming, IFAC-PapersOnLine, Volume 53, Issue 2, Pages 12948-12955.

Gabrielli P., Gazzani M., Martelli E., Mazzotti M., 2018, Optimal design of multi-energy systems with seasonal storage, Applied Energy 219 408–424.

Ghaemi Z., Tran T., Smith A.D., Sizing optimization of district energy systems considering meteorological, demand, and electricity emission factor uncertainties, 2021, International Mechanical Engineering Congress and Exposition IMECE2021, November 1-5, 2021, Virtual, Online.

Golombek R., Lind A., Ringkjøb H.K., Seljom P., 2022 The role of transmission and energy storage in European decarbonization towards 2050. Energy, Volume 239, Part C, 122159.

Gomes I.L.R., Melicio R., Mendes V.M.F, 2021, A novel microgrid support management system based on stochastic mixed-integer linear programming, Energy, Volume 223, 120030.

Kotzur L., Markewitz P., Robinius M., Stolten D., 2018, Impact of different time series aggregation methods on optimal energy system design. Renewable Energy 117 474e487. Renewable Energy, Volume 117, March 2018, Pages 474-487.

Liu J., Zhou Y., Yang H., Wu H., 2022 Uncertainty energy planning of net-zero energy communities with peer-to-peer energy trading and green vehicle storage considering climate changes by 2050 with machine learning methods. Applied Energy, Volume 321, 119394.

Mavromatidis G., Orehounig K., Carmeliet J., 2018a, Uncertainty and global sensitivity analysis for the optimal design of distributed energy systems. Applied Energy, Volume 214, Pages 219-238.

Mavromatidis G., Orehounig K., Carmeliet J., 2018b, A review of uncertainty characterisation approaches for the optimal design of distributed energy systems. Renewable and Sustainable Energy revoews, Volume 88, Pages 258-277.

Merrick J.H., 2016, On representation of temporal variability in electricity capacity planning models, Energy Econ. ISSN: 01409883 59 261e274.

Paparrizos J, Gravano L. k-shape: efficient and accurate clustering of time series, 2015, Acm Sigmod: 1855–70.

Pfenninger S., 2017, 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, Applied Energy Volume 197, Pages 1-13.

Teichgraeber H., Brandt A., 2019, Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison. Applied Energy, Volume 239, Pages 1283-1293.

Terna SPA, https://www.terna.it/it/sistema-elettrico/dispacciamento (Accessed February 2024).

Walker W., Harremoës P, Rotmans J, Sluijs JP, van der, Asselt MBA, van Janssen P, et al., 2003, Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support, Integrated Assessment 2003, Vol 4, No.1 pp.5–17.

Zatti M., Gabba M., Freschini M., Rossi M., Gambarotta A., Morini M., Martelli E., 2019, k-MILP: A novel clustering approach to select typical and extreme days for multi-energy systems design optimization, Energy 181 (2019) 1051e1063.

Zhang C., Xue X., Du O., Luo Y., Gang W., 2019, Study on the performance of distributed energy systems based on historical loads considering parameter uncertainties for decision making. Energy, Volume 176, Pages 778-791.

Zhou Z., Zhang J., Liu P., Li Z., Georgiadis M.C., Pistikopoulos E.N, 2013, A two-stage stochastic programming model for the optimal design of distributed energy systems, Appl. Energy 103, 135–144.

https://doi.org/10.52202/077185-0127