UNSUPERVISED AND SUPERVISED MACHINE LEARNING TECHNIQUES FOR TIMESERIES AGGREGATION IN THE DESIGN AND OPERATION OPTIMIZATION OF MULTI-ENERGY SYSTEMS

Gabriele Volpato[1]*, Edoardo Bregolin[1], Enrico Dal Cin[1], Gianluca Carraro[1], Piero Danieli[2], Andrea Lazzaretto[1]

[1]University of Padova, Industrial Engineering Department
Via Venezia 1, 35131 Padova, Italy
[2]Department of Management and Engineering, University of Padova, Stradella S. Nicola 3, 36100 Vicenza (Italy)
*Corresponding Author: gabriele.volpato.1@phd.unipd.it

ABSTRACT

The optimization of the sizes and operation of energy conversion and storage units in multi-energy systems with time horizons of years is strongly dependent on the variable timeseries of the grid electricity price. Thus, the optimization could be carried out considering representative timeseries obtained by unsupervised (e.g., clustering) or supervised (e.g., artificial neural networks) techniques. However, these techniques have mainly been combined to improve the accuracy of the timeseries prediction without then using these timeseries in an optimization problem. The objective is to evaluate whether unsupervised techniques can be properly combined with supervised techniques to obtain a representative set of typical days of the grid electricity price, the impact of which is assessed on the optimal total cost of a multi-energy system. This paper proposes a novel and preliminary hybrid approach that combines multi-year clustering and a feedforward Backpropagation artificial Neural Network (BPNN), and then compares it with a state-of-the-art multi-year clustering. Both approaches are fairly applied to the same "past" dataset (2005-2014), where the multi-year clustering identifies clusters containing similar timeseries, while the hybrid approach is based on training the BPNN with the timeseries labelled according to the representative clusters found by the multi-year clustering. The multi-year clustering approach finds a representative set of typical days in the "past" dataset and considers it in a "future" dataset (2015-2020), while the hybrid approach finds a representative set by classifying the days of the "future" dataset according to the training of the BPNN in the "past". Subsequently, a stochastic programming model is used to optimize the design-operation of the system by minimizing its life cycle (investment and operational) costs in the "future" dataset, using separately the two different representative sets of typical days of the electricity price. The optimal life cycle costs based on the typical days of the multi-year clustering and hybrid approaches show errors of 3% and 5%, respectively, compared to "perfect knowledge" solutions based on data really occurred. Preliminary results show the validity of the proposed hybrid approach and point to further improvements.

Keywords: multi-energy system, design-operation optimization, clustering, artificial neural network, stochastic programming

1 INTRODUCTION

1.1 Literature review

Nowadays the pressing climatic and environmental issues require the optimal design of sustainable energy systems that can provide affordable renewable energy to different end users (EU, 2021). In this context, Multi-Energy Systems (MES) driven mainly by renewable energy can meet the various energy demands of different end-users. A MES consists of a set of energy conversion and storage units that exploit the interaction between multiple energy vectors (e.g., electricity, heating, cooling, fuels, etc.), at different geographical scales (e.g., neighbourhood, district, municipality, city, etc.), to find the best match between demand and generation (Mancarella, 2014, Guelpa *et al.*, 2019). Mancò *et al.* (2023) showed how nonlinearities of model constraints, the type of optimization (i.e., synthesis, design and operation), the uncertainty in input parameters and the flexibility strategies (e.g., storage units, demand response programs) affect the formulation of the optimization model of MES.

https://doi.org/10.52202/077185-0005

The search for the optimal sizes and operation of energy conversion and storage unit of a MES requires a design-operation optimization problem with a time horizon of at least one year with hourly resolution, assuming that the year is representative of the whole lifetime of the system (Zheng *et al.*, 2021). However, solving the optimization based on all input daily timeseries in one year leads to a high computational burden, which could be avoided by an optimization based on representative timeseries of the year (Kannengießer *et al.*, 2019), obtained by applying Machine Learning (ML) techniques such as unsupervised clustering and supervised Artificial Neural Network (ANN) models.

Clustering approaches are usually applied to annual historical datasets to group similar timeseries into the same clusters based on the similarity of their features (i.e., weather data, energy demands, energy prices, etc.), thus obtaining a representative timeseries for each cluster (Kotzur *et al.*, 2018). Hoffmann *et al.* (2020) reviewed the main clustering techniques to identify representative timeseries. Hoffmann *et al.* (2021) found that the type of the optimization problem (i.e., operation or design-operation) affects the choice between the typical time steps (e.g., hours) and the typical periods (e.g., days or weeks) in the optimization of MES. Some works carried out the design and operation optimization of MES considering typical days of years obtained by clustering. Fazlollahi *et al.* (2014) carried out a one-year multi-objective optimization based on 7 typical days of different parameters (e.g., solar irradiance, energy demands, etc.), obtained by a centroid-based clustering algorithm (Fazlollahi *et al.*, 2012), to achieve the optimal design and operation of a MES consisting of thermal storage, cogeneration units and a solar thermal plant. Bahl *et al.* (2018) and Bahl *et al.* (2017) selected the best sets of typical days among alternative sets generated by K-medoids and K-means clustering, respectively, for a design-operation optimization of a MES by finding the minimum error between the optimal objective function based on typical days and that based on the full annual data.

An Artificial Neural Network (ANN) is a computational model consisting of interconnected nodes that can learn complex relationships between input and output data to perform classification or regression tasks (Fausett, 2006). Some ANN models were mainly used to forecast the energy demands of different users, taking multi-year historical datasets as input. Runge *et al.* (2019) reviewed ANN models in terms of selection of their architecture (e.g., the number of neurons for each layer of the ANN) and algorithms used to train an ANN. Del Real *et al.* (2020) developed a deep neural network, combining a convolutional neural network and a feedforward Backpropagation ANN (BPNN), to forecast the national electricity demand in France. Petrucci *et al.* (2022) optimized the operation of the energy storage units of an Energy Community (EC), which is an aggregation of users sharing renewable electricity (Bartolini *et al.*, 2020), using a day-ahead electricity demand of the EC predicted by an ANN. Some works combined different ML techniques. Giannuzzo *et al.* (2024) presented a methodology to estimate the shared energy among the members of a Renewable Energy Community (REC). They first applied K-means clustering to identify the typical profiles of the total electricity demand, and then used a Random Forest algorithm to assign these profiles to the REC members according to their monthly consumption data. Few works specifically used both clustering and ANN approaches to obtain accurate forecasts of different variables. Erilli *et al.* (2011) proposed a BPNN that takes, as input, data classified according to the results of a fuzzy clustering algorithm, and gives, as output, the optimal number of clusters. Ząbkowski *et al.* (2023) applied hierarchical clustering to obtain typical demand patterns of commercial users, and then used these patterns as input to a BPNN to forecast the total electricity demand of the Polish power system. They found that the higher the number of clusters, the higher the accuracy in estimating the total demand of the country. Luo (2020) proposed a methodology to accurately predict the day-ahead cooling demand of buildings, depending on typical patterns of weather data and building operating schedules. They implemented K-means clustering to identify different groups of weather timeseries in an annual dataset, and then used each group separately as input in the training of an ANN to predict the day-ahead cooling demand. Du *et al.* (2022) predicted the thermal performance of solar collectors by first applying a hierarchical clustering to identify outliers in experimental data of weather and fluid parameters, and then using the filtered data to train a BPNN.

It should be noted that most of the ANN models were mainly used to forecast energy demands and not energy prices, the variability of which can strongly influence the total cost (investment and operational costs) of a MES. On the other hand, this paper evaluates different ML techniques (i.e., clustering and ANN) to obtain typical days of the grid electricity price. These typical days are used in the design and operation optimization of a MES meeting the energy demand of an EC.

https://doi.org/10.52202/077185-0005

1.2 Goals and novelty

Unsupervised clustering techniques have mainly been integrated with supervised ANNs to improve the accuracy of predicting typical days of different timeseries (e.g., energy demands), without investigating whether these typical days can lead to a good accuracy of the optimal design-operation solution of a MES compared to a "perfect knowledge" solution based on the knowledge of all timeseries. The objective is to evaluate whether a novel and preliminary hybrid approach, which combines multi-year clustering and a feedforward Backpropagation ANN (BPNN), is successful in identifying a representative set of typical days of the grid electricity price, the impact of which is assessed on the optimal total cost of a MES. In addition, the proposed hybrid approach is compared with a state-of-the-art approach based only on multi-year clustering.

The multi-year clustering approach finds a representative set of typical days of prices in a "past" dataset (2005-2014) and considers the same set in a "future" dataset (2015-2020), whereas the proposed hybrid approach finds this set in the "future" by using a BPNN that is trained in the "past" according to the clusters found by the multi-year clustering approach. Subsequently, the two different sets of typical days are used separately as input to a Stochastic Programming (SP) model (Teichgraeber et al., 2020) to optimize the design and operation of the MES and to evaluate their impact on the optimal total cost of the system. The paper is organized as follows. Section 2 presents the methodology. Section 3 presents the results. Section 4 summarizes the main findings and conclusions.

2 METHODOLOGY

Figure 1 shows that an available historical dataset (2005-2020) is divided into a "past" dataset (2005-2014) and a "future" dataset (2015-2020) by shifting the "present" moment back to 2015. This assumption is required to fairly apply the multi-year clustering and hybrid approaches to the same "past" dataset, making the two approaches comparable to each other, since the BPNN of the hybrid approach is trained in the "past" based on the classification made by the multi-year clustering (which is applied in the "past"). The design of the system can take place in any year after 2020, based on the optimal design found in each year of the "future" dataset with the typical days obtained by the two approaches. The multi-year clustering approach is applied to the whole "past" dataset to identify a representative set of typical days of prices, assuming then the same set in the "future" dataset. On the other hand, the hybrid approach uses the days in the "past" dataset for the training, validation and testing of the BPNN, where these days are labelled according to the clusters of the representative set of typical days found by the multi-year clustering approach. The representative set of typical days of prices according to the hybrid approach is identified by testing the BPNN in the "future" dataset. Subsequently, a Stochastic Programming (SP) (Infanger, 1992) model is used to optimize the design and operation of the Multi-Energy System (MES) for each year of the "future" dataset by minimizing its life cycle cost (investment and operational costs), using separately as input the two different representative sets of typical days. To assess their accuracy, the optimal life cycle costs are compared with those of a Mixed-Integer Linear Programming (MILP) optimization, based on the "perfect knowledge" of all timeseries in each year of the "future" dataset. Sections 2.1, 2.2, 2.3 and 2.4 present the optimization model of the system, the multi-year clustering approach, the proposed hybrid approach and the input data, respectively.

2.1 Optimization model of the multi-energy system

Figure 2 shows the Multi-Energy System (MES) that meets the energy demands of a Renewable Energy Community (REC) including residential (Res) and public (Pub) prosumers, and commercial (Com) and tertiary (Ter) consumers. The energy conversion and storage units that can be installed are solar Photovoltaic plants (PV), Heat Pumps (HP), Gas Boilers (GB), Electrical Energy Storage (EES) and Thermal Energy Storage (TES). A Stochastic Programming (SP) model and a Mixed-Integer Linear Programming (MILP) model are used to optimize the sizes and operation of the units, based on a set of typical days and the "perfect knowledge" of all 365 days of a year, respectively. The typical days of electricity prices are obtained by the multi-year clustering approach or the proposed hybrid approach (Sections 2.2 and 2.3, respectively). The time horizon of the optimization is one year, assuming that the system operation is the same for each year of its life cycle. The decision variables and constraints refer to consumers c or prosumers p in hour t of day d.

https://doi.org/10.52202/077185-0005

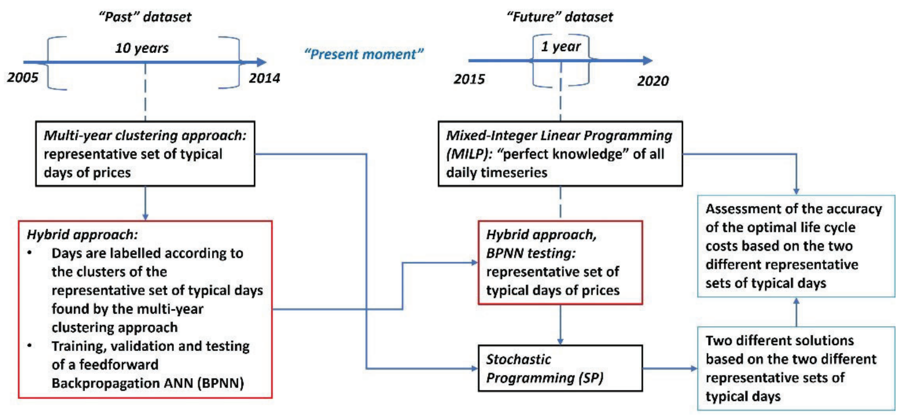

Figure 1: Application of the multi-year clustering approach and the proposed hybrid approach (red boxes) to obtain two different representative sets of typical days of prices, the impact of which is evaluated on the optimal life cycle cost of the system.

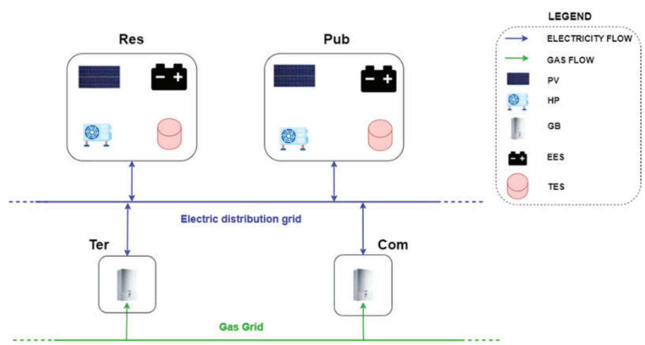

Figure 2: MES of the REC with the residential (Res), public (Pub), tertiary (Ter) and commercial (Com) members.

The design decision variables are the capacities of the PV (cap_p^{PV}), HP (cap_p^{HP}), EES and TES (cap_p^{ES} for a general storage unit). The capacity of the GB of a consumer c (cap_c^{GB}) is directly calculated from its input heating demand. The operational decision variables are: the heating power generated by the HP ($Q_{p,t,d}^{HP}$) and the binary variable indicating its on-off operational state ($\delta_{p,t,d}^{HP}$); the energy stored by the Energy Storage (ES) unit ($E_{p,t,d}^{ES}$), its charging/discharging power ($P_{p,t,d}^{ES,+}/P_{p,t,d}^{ES,-}$) and the binary variable indicating its charging/discharging state ($\delta_{p,t,d}^{ES}$); the energy imported/exported from/to the electrical grid ($E_{p,t,d}^{imp}/E_{p,t,d}^{exp}$); the shifted electricity demands of users ($E_{c,t,d}^{el,shift}$ and $E_{p,t,d}^{el,shift}$). The constraints associated with each prosumer p are reported below. Energy and power variables have, respectively, [kWh] and [kW] as units of measurement.

The electricity balance of a prosumer p is:

$$E_{p,t,d}^{imp} - E_{p,t,d}^{exp} + \left(P_{p,t,d}^{PV} + P_{p,t,d}^{EES,-} - P_{p,t,d}^{EES,+} - P_{p,t,d}^{HP}\right) \cdot \Delta t - E_{p,t,d}^{el,shift} = 0 \quad (1)$$

where $E_{p,t,d}^{imp}/E_{p,t,d}^{exp}$, $P_{p,t,d}^{PV}$, $P_{p,t,d}^{EES,-}/P_{p,t,d}^{EES,+}$, $P_{p,t,d}^{HP}$, Δt and $E_{p,t,d}^{el,shift}$ are the energy imported/exported from/to the electrical grid, the electrical power generated by PV, the power discharged/charged from/into EES, the electrical power consumed by HP, the step of one hour in the optimization and the shifted electricity demand of prosumer p, respectively.

The heating balance of a prosumer p is:

$$\left(Q_{p,t,d}^{HP} + P_{p,t,d}^{TES,-} - P_{p,t,d}^{TES,+}\right) \cdot \Delta t - E_{p,t,d}^{th} = 0 \quad (2)$$

where $Q_{p,t,d}^{HP}$, $P_{p,t,d}^{TES,-}/P_{p,t,d}^{TES,+}$ and $E_{p,t,d}^{th}$ are the heating power generated by HP, the power discharged/charged from/into TES and the heating demand of prosumer p.

https://doi.org/10.52202/077185-0005

The energy balance of an ES (i.e., EES or TES) of a prosumer *p* is:

$$E_{p,t,d}^{ES} = E_{p,t-1,d}^{ES} \cdot (1 - SD) + \left(P_{p,t,d}^{ES,+} \cdot \eta^{ES,+} - P_{p,t,d}^{ES,-} / \eta^{ES,-}\right) \cdot \Delta t \tag{3}$$

where $E_{p,t,d}^{ES}$, SD, $P_{p,t,d}^{ES,+}/P_{p,t,d}^{ES,-}$ and $\eta^{ES,+}/\eta^{ES,-}$ are the state of charge [% of capacity], the self-discharge [% of the state of charge in each hour], the charging/discharging power and the charging/discharging efficiency [-] of ES, respectively. Other constraints of ES are:

$$\theta_{p,t,d}^{ES,+} \leq \delta_{p,t,d}^{ES} \cdot M \tag{4}$$

$$0 \leq cap_p^{ES} - \theta_{p,t,d}^{ES,+} \leq \left(1 - \delta_{p,t,d}^{ES}\right) \cdot M \tag{5}$$

$$\theta_{p,t,d}^{ES,-} \leq \left(1 - \delta_{p,t,d}^{ES}\right) \cdot M \tag{6}$$

$$0 \leq cap_p^{ES} - \theta_{p,t,d}^{ES,-} \leq \delta_{p,t,d}^{ES} \cdot M \tag{7}$$

$$E_{p,t,d}^{ES} \leq cap_p^{ES} \tag{8}$$

$$P_{p,t,d}^{ES,+} \leq C^{ES,+} \cdot \theta_{p,t,d}^{ES,+} \tag{9}$$

$$P_{p,t,d}^{ES,-} \leq C^{ES,-} \cdot \theta_{p,t,d}^{ES,-} \tag{10}$$

$$E_{p,t=1,d}^{ES} = E_{p,t=24,d}^{ES} \tag{11}$$

where cap_p^{ES} [kWh] is the capacity, $\delta_{p,t,d}^{ES}$ is the binary variable associated with the charging (1) or discharging (0) state of ES, $C^{ES,+}$ and $C^{ES,-}$ [kW/kWh] are the specific input and output capacity. The auxiliary variables $\theta_{p,t,d}^{ES,+}$ and $\theta_{p,t,d}^{ES,-}$ and the *M* parameter are used to avoid bilinear constraints.

The characteristic curve of the HP of a prosumer *p* is:

$$P_{p,t,d}^{HP} = \frac{a_0 \cdot Q_{p,t,d}^{HP} + a_1 \cdot \delta_{p,t,d}^{HP}}{COP_{p,t,d}} \tag{12}$$

where $P_{p,t,d}^{HP}$, $Q_{p,t,d}^{HP}$, $\delta_{p,t,d}^{HP}$, a_0 and a_1, and $COP_{p,t,d}$ are the electrical power consumed, the heating power generated, the binary variable indicating the on-off operational state of the HP, the constant coefficients [-] that linearize the characteristic curve and the coefficient of performance in ideal conditions (Carnot equation) (Dal Cin *et al.*, 2023), respectively. Other constraints of the HP are:

$$\theta_{p,t,d}^{HP} \leq \delta_{p,t,d}^{HP} \cdot M \tag{13}$$

$$0 < cap_p^{HP} - \theta_{p,t,d}^{HP} \leq \left(1 - \delta_{p,t,d}^{HP}\right) \cdot M \tag{14}$$

$$min_{HP} \cdot \theta_{p,t,d}^{HP} \leq Q_{p,t,d}^{HP} \leq \theta_{p,t,d}^{HP} \tag{15}$$

where cap_p^{HP} [kW] is the capacity and min_{HP} [% of the capacity] is the minimum part load of the HP. The auxiliary variable $\theta_{p,t,d}^{HP}$ and the *M* parameter are used to avoid bilinear constraints.

For a prosumer *p*, the electrical power generated by PV is:

$$P_{p,t,d}^{PV} = cap_p^{PV} \cdot I_{t,d} \tag{16}$$

where $P_{p,t,d}^{PV}$ is the power generated, cap_p^{PV} [m^2] is the capacity (considering also the efficiency of PV) and $I_{t,d}$ [kW/m^2] is the global solar irradiance (on a tilted surface).

The hourly electricity demand of each member *i* of the EC can be shifted as follows:

$$\sum_{t=1}^{24} E_{i,t,d}^{el} = \sum_{t=1}^{24} E_{i,t,d}^{el,shift} \tag{17}$$

$$E_{i,d}^{el,min} \leq E_{i,t,d}^{el,shift} \leq E_{i,d}^{el,max} \tag{18}$$

$$(1 - D^{var}) \cdot E_{i,t,d}^{el} \leq E_{i,t,d}^{el,shift} \leq (1 + D^{var}) \cdot E_{i,t,d}^{el} \tag{19}$$

where $E_{i,t,d}^{el}$ [kWh], $E_{i,t,d}^{el,shift}$ [kWh], $E_{i,d}^{el,min}$ [kWh] and $E_{i,d}^{el,max}$ [kWh], and D^{var} are, respectively, the input electricity demand, the shifted electricity demand in hour *t* of day *d*, the minimum and maximum of the input electricity demand in day *d*, and the hourly maximum fraction of the load that can be shifted.

The objective function to be minimized is the life cycle cost of the system referring to one year:

$$c_{life\ cycle} = c_{design} + c_{operation} \tag{20}$$

where c_{design} and $c_{operation}$ are the investment and operational costs, with the investment cost actualized to one year of operation.

The investment cost of the system is:

https://doi.org/10.52202/077185-0005

$$c_{design} = \sum_{u \in U} \left(\left(\frac{a \cdot (1+a)^{lt_u}}{(1+a)^{lt_u} - 1} + O\&M_{fix,u} \right) \cdot c_{inv,u} \cdot \sum_{i=1}^{N} cap_i^u \right) \quad (21)$$

where *u* identifies a specific energy technology (*U* is the set of technologies), *a* [%] is the interest rate, *ltu* [years] is the lifetime of the technology *u*, $O\&M_{fix,u}$ [% of the investment cost] is the operation and maintenance cost of the technology u, $c_{inv,u}$ [€/kW or €/kWh] is the investment cost, cap_i^u is the capacity [kW or kWh] of the technology *u* owned by member *i* and *N* is the number of EC members. The annual operational cost, based on a set of typical days (in the SP model) or 365 days (in the MILP model) of one year is:

$$c_{operation} = \sum_{d=1}^{K} w_d \cdot \sum_{t=1}^{24} \left(\sum_{c \in C} \left(F_{c,t,d}^{GB} \cdot c_{gas} \right) + E_{t,d}^{imp} \cdot c_{t,d}^{imp} - E_{t,d}^{exp} \cdot c_{t,d}^{exp} - E_{s,t,d} \cdot inc_{REC} \right) \quad (22)$$

where *K*, w_d, $F_{c,t,d}^{GB}$ [kWh], c_{gas} [€/kWh], $c_{t,d}^{imp}$ and $c_{t,d}^{exp}$ [€/kWh], $E_{s,t,d}$ [kWh] and inc_{REC} [€/kWh] are the number of days considered in one year (i.e., typical days in the SP model or 365 days in the MILP model), the weight of day *d* (i.e., the number of days in the year represented by the typical day *d* in the SP model or 1/365 in the MILP model), the energy consumed by GB of consumers *c* (*C* is the set representing the consumers), the price of natural gas, the grid purchase and sale prices, the shared energy and the incentive of the REC, respectively. The shared energy is calculated as the hourly minimum between the total net energy withdrawn from and injected to the grid by the REC according to the Italian legislation (ARERA, 2022, ARERA, 2023). The first term of the summation in Eq. (22) represents the cost of natural gas consumed by the boilers. The second and third terms are the cost for the electricity imported from the grid and the revenue for the electricity exported to the grid. The last term is the revenue associated with the shared energy.

2.2 Multi-year clustering approach

The implemented multi-year clustering approach is based on the K-means algorithm (Hoffmann *et al.*, 2020), which is applied to the entire "past" dataset to obtain typical days of electricity prices, increasing the number of clusters generated (containing similar timeseries) from 2 to 30. Thus, 29 alternative sets of typical days (i.e., the representative of the clusters) are generated, each consisting of 2 up to 30 typical days. The typical day of each cluster is selected as the real day with the lowest value of the Euclidean distance from the centroid of the cluster (Zatti *et al.*, 2019). However, this procedure could neglect the extreme days of the electricity prices. Thus, following the "additional cluster center" criterion (Kotzur *et al.*, 2018), the two extreme days characterized by the minimum and maximum daily sums of the hourly grid electricity prices are set as typical days of two new clusters added in each of the 29 alternative sets. Subsequently, all daily timeseries of electricity prices are assigned again to the different clusters, as the 29 alternative sets now contain 4 up to 32 clusters, and the typical days of all clusters are updated. Each typical day is considered in the SP model (Section 2.1) with a weight corresponding to the frequency of its cluster divided by the 10 years of the "past" dataset (i.e., the number of days in a year represented by that typical day). The daily timeseries of solar irradiance, ambient temperature, electricity and heating demands are considered on the same days of the typical days of electricity prices to preserve the chronological correlation between the different daily timeseries.
The representative set of typical days of electricity prices is found by searching for the lowest Root Mean Squared Error (RMSE) of the optimal life cycle costs (Eq. (20), Section 2.1) obtained by solving the SP and MILP models over the years of the "past" dataset (2005-2014). The RMSE is (Mosavi *et al.*, 2019):

$$RMSE = \sqrt{\frac{\sum_{y=1}^{Y} \left(c_{SP,y,k} - c_{MILP,y} \right)^2}{Y}} \quad (23)$$

where $c_{SP,y,k}$ and $c_{MILP,y}$, and Y are, respectively, the optimal life cycle costs found by solving the SP model for year y with k typical days (varying from 4 to 32) and the MILP model for year y, and the

https://doi.org/10.52202/077185-0005

number of years considered (10), respectively. This representative set of typical days in the "past" dataset (Figure 1) is then used to solve the SP model for each year of the "future" dataset (2015-2020).

2.3 Hybrid approach: multi-year clustering combined with an artificial neural network

The novel hybrid approach consists of a feedforward Backpropagation Neural Network (BPNN) (Runge *et al.*, 2019) that uses the representative clusters obtained by the multi-year clustering approach (Section 2.2) in the "past" (2005-2014) to find a representative set of typical days of grid electricity price directly in the "future" (2015-2020). The architecture of the BPNN includes an input layer, hidden layers and an output layer, the latter consisting of nodes representing the output classes (e.g., a class represents days characterized by similar profiles of the grid electricity price). Each node *i* of a general hidden layer j is associated with an output $z_{i,j}$, which is calculated as:

$$z_{ij} = \varphi(z) \cdot \sum_{i=1}^{n} (w_{i,j} \cdot z_{i,j-1}) + b_j \qquad (24)$$

where $\varphi(z)$, n, $w_{i,j}$, $z_{i,j-1}$ and b_j are the activation function (i.e., representing the non-linear relationship between input and output data), the number of nodes in the layer j, the weights of the connections between the nodes of different layers, the output of the previous hidden layer j-1 and the bias function (representing the noise) in layer j, respectively. Figure 3 shows the chosen BPNN, which consists of an input layer with 24 variables (e.g., the hourly grid electricity prices in a day), 3 hidden layers with 100 nodes each, and an output layer with 16 nodes. The number of hidden layers and their neurons was chosen by trial-and-error procedures according to the dimension of the "past" training dataset. The nodes of the output layer represent the output classes to which the daily timeseries of grid electricity price can be assigned to, and their number was chosen equal to the number of representative clusters found from the multi-year clustering approach.

The BPNN is trained with a dataset of daily timeseries of the grid electricity price, which are labelled according to input classes, i.e., the representative clusters of the multi-year clustering approach. The training of the BPNN updates iteratively the weights and biases by minimizing an error function between the computed output class (i.e., z_{ij} in Eq. (24)) and the corresponding input class for each daily timeseries. In each training step, the BPNN calculates the probabilities of belonging to the different classes for each daily timeseries, which is assigned to the class with the highest probability.

The training of the BPNN is composed of two phases. In one phase, the daily timeseries in the first 60% part of the "past" dataset are used to train the BPNN, while in the other phase the consequent 20% and the remaining 20% of the "past" dataset are used to validate and test the BPNN, respectively. The two phases are performed simultaneously to prevent overfitting and ensure a high generalization ability of the network. Subsequently, the BPNN is tested to predict the classes of the daily timeseries of prices in the "future" dataset. According to the proposed hybrid approach (Figure 1), the representative set of typical days of electricity prices includes real days in the "future" dataset with the lowest Euclidean distance from the centroid of each group of days with the same output class. This representative set of typical days is then used to solve the SP model for each year of the "future" dataset.

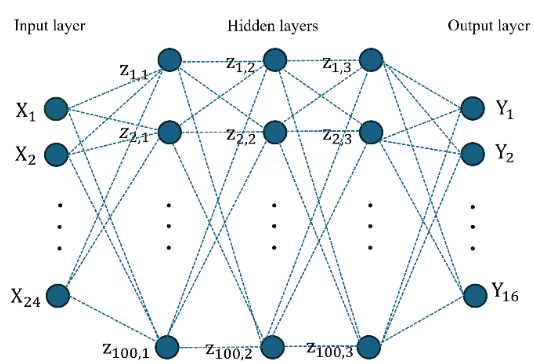

Figure 3: Architecture of the chosen BPNN with the input layer (24 nodes), the three hidden layers (100 nodes each) and the output layer (16 nodes).

https://doi.org/10.52202/077185-0005

2.4 Input data

Table 1 shows the values of the techno-economic parameters ((Sterchele *et al.*, 2020), DEA (2023)). The interest rate a of the investments and the lifetime lt_u of each technology u (Eq. (21)) are assumed to be 0.05 [-] and 20 years, respectively. Solar irradiance and ambient temperature refer to the location of Padova (Italy) and are taken from (PVGIS, 2022). The daily timeseries of the grid electricity price are based on a historical dataset (2005-2020) of the day-ahead market price in Italy (GME, 2022), where the grid sale price is assumed to be half of the day-ahead market price and the grid purchase price is calculated as the grid sale price plus 0.2 €/kWh. The electricity and heating demands of different users are taken from (DOE, 2023). Moreover, the maximum hourly fraction of the load that can be shifted (constraints (19) in Section 2.1) is equal to 0.1. The price of natural gas is equal to 0.098 €/kWh. The incentive for shared energy is 0.12 €/kWh.

Table 1: Input techno-economic parameters of the optimization models (Sterchele *et al.*, 2020, DEA, 2023).

Technology *"u"*	Parameter	Value
GB	η_{GB}[-]	0.97
	$c_{inv,u}$ [€/kW$_{th}$]	300
	$O\&M_{fix,u}$ [% of $c_{inv,u}$]	4.9
PV	$c_{inv,u}$ [€/kW$_{el}$]	1250
	$O\&M_{fix,u}$ [% of $c_{inv,u}$]	1.1
HP	a_0, a_1 [-]	1.7961, 2.6527
	min_{HP} [% of capacity]	50
	$c_{inv,u}$ [€/kW$_{th}$]	1500
	$O\&M_{fix,u}$ [% of $c_{inv,u}$]	2.8
EES(TES)	SD [% of the state of charge in each hour]	0.04(2.1)
	$\eta^{ES,+}, \eta^{ES,-}$ [-]	0.95(0.99), 0.95(0.99)
	$C^{ES,+}, C^{ES,-}$ [kW/kWh]	0.5(0.7), 3(0.7)
	$c_{inv,u}$ [€/kWh]	1500(400)
	$O\&M_{fix,u}$ [% of $c_{inv,u}$]	1(4)

3 RESULTS

Figure 4(a) and (b) show the representative sets of 16 and 15 typical days of the grid electricity (sale) price obtained by the multi-year clustering approach in the "past" dataset (2005-2014) and by the hybrid approach in the "future" dataset (2015-2020), respectively. Although the hybrid approach uses the 16 representative clusters of the multi-year clustering approach as input classes to the BPNN to predict the class of each daily timeseries in the "future" (Section 2.3), no timeseries is assigned to the class represented by the extreme profile with the highest prices (shown in red in Figure 4(a)). Consequently, the daily timeseries of prices in the "future" dataset are assigned to 15 classes, resulting in 15 typical days (Figure 4(b)). The two sets of typical days are used separately to solve the SP model for each year of the "future" dataset, leading to different optimal life cycle costs that are compared with those found by the MILP model with "perfect knowledge" (Section 2).

The representative set of typical days according to the multi-year clustering approach is found by searching for the minimum RMSE of the optimal life cycle costs obtained by solving the SP model and the MILP model based on typical days and all 365 days, respectively. Figure 5 shows the RMSE for 29 different sets of typical days found by the multi-year clustering approach, with a number ranging from 4 to 32 (Section 2.2). The calculated RMSE is characterized by large deviations around 50 [k€], with several local minima for 10, 16, 23 and 29 typical days. Among these sets, that with 16 typical days (Figure 4(a)) is chosen as the most representative, as it leads to the lowest average relative error in the

https://doi.org/10.52202/077185-0005

optimal life cost in the "future" dataset (2015-2020). Table 2(a) shows the optimal life cycle costs obtained by solving the SP model for the representative set of 16 typical days, found with the multi-year clustering approach, and the MILP model for all days over the years of the "future" dataset. The average relative error in the optimal life cycle cost obtained by solving the SP model compared to the MILP model is 3.07%.

Table 2(b) shows the optimal life cycle costs obtained by solving the SP model for the representative set of 15 typical days, found with the hybrid approach, and the MILP model for all days over the years of the "future" dataset. The average relative error in the optimal life cycle cost obtained by solving the SP model compared to the MILP model is 5.21%, which is acceptable but slightly higher than that of the multi-year clustering approach. This result could have two main reasons. First, the hybrid approach is based on training the BPNN on the "past" dataset. This process introduces an error related to the ability of the BPNN to classify all the daily timeseries of the "past" dataset according to the given input classes. This error, in turn, affects the prediction of the classes of the daily timeseries in the "future" dataset. Second, the representative set of 15 typical days found by the hybrid approach (Figure 4(b)) does not contain a real extreme daily timeseries of the grid electricity price. In fact, the BPNN fails in finding daily timeseries of the "future" dataset that can be classified according to the extreme profile with the highest prices of the "past" dataset (which is instead included in the representative set found by the multi-year clustering approach).

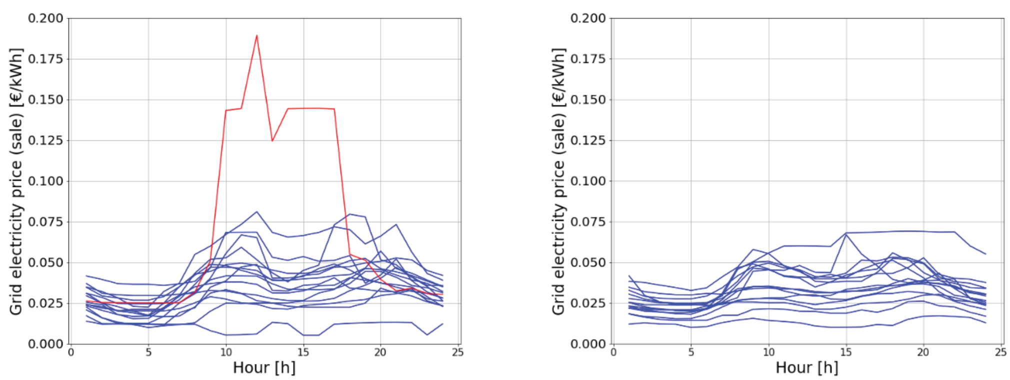

Figure 4: Representative set of *a)* 16 typical days (extreme profile is red) and *b)* 15 typical days of the grid electricity price, obtained by the multi-year clustering and hybrid approaches, respectively.

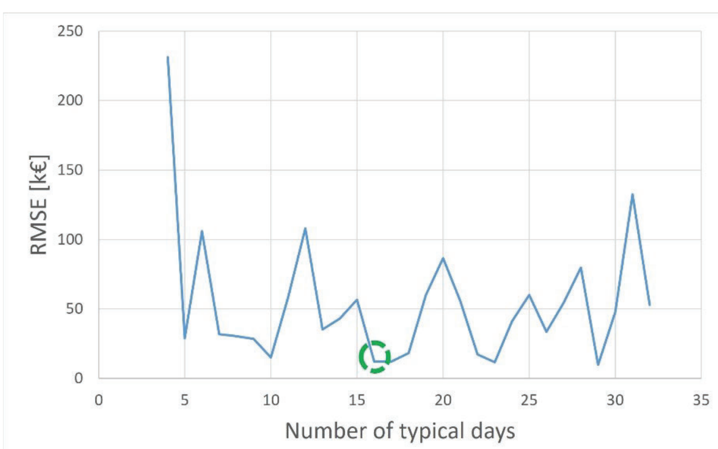

Figure 5: RMSE as the number of typical days in the "past" dataset increases from 4 to 32. The dashed green circle identifies the most representative set of typical days of the grid electricity price according to the multi-year clustering approach.

https://doi.org/10.52202/077185-0005

Table 2: Optimal life cycle costs of the system obtained by solving the SP model for the representative set of *a)* 16 typical days (multi-year clustering approach) and *b)* 15 typical days (hybrid approach), and the MILP model for all days over the years of the "future" dataset.

a)

	Years					
	2015	**2016**	**2017**	**2018**	**2019**	**2020**
SP: life cycle costs [k€]	383.03	387.07	369.60	362.14	392.86	357.87
MILP: life cycle costs [k€]	370.70	377.46	375.75	375.65	374.90	367.15
Relative errors [%]	3.33	2.54	1.64	3.60	4.79	2.53

b)

	Years					
	2015	**2016**	**2017**	**2018**	**2019**	**2020**
SP: life cycle costs [k€]	354.93	358.42	344.12	377.61	355.18	338.62
MILP: life cycle costs [k€]	370.70	377.46	375.75	375.65	374.90	367.15
Relative errors [%]	4.25	5.05	8.42	0.52	5.26	7.77

4 CONCLUSIONS

This paper focuses on finding typical days of the grid electricity price, the variability of which can strongly influence the optimal life cycle cost (investment and operational costs) of a Multi-Energy System (MES) within a Renewable Energy Community (REC). The analysed REC consists of residential and public prosumers who could install solar photovoltaic plants, heat pumps, electrical and thermal energy storage units, and tertiary and commercial users who could install natural gas boilers. The application of a novel hybrid approach, combining multi-year clustering and a feedforward Backpropagation artificial Neural Network (BPNN), is proposed to identify a representative set of typical days of the grid electricity price and assess their impact on the optimal life cycle cost of the system. In addition, the hybrid approach is compared with a state-of-the-art multi-year clustering.

The methodology implemented is based on splitting an available historical dataset (2005-2020) of daily timeseries of grid electricity prices into a "past" dataset (2005-2014) and a "future" dataset (2015-2020) by shifting the "present" moment back to 2015 to achieve a fair comparison between the multi-year clustering and hybrid approaches. Indeed, the multi-year clustering approach applies K-means clustering to the "past" to identify a representative set of typical days, assuming then the same set in the "future". On the other hand, the novel hybrid approach uses the clusters of the representative set of typical days found by the multi-year clustering approach in the "past" to train a BPNN. Then, the representative set of typical days according to the hybrid approach is found in the "future" by testing the BPNN to predict the class of each daily timeseries. Subsequently, the two different representative sets of typical days are used separately in a Stochastic Programming (SP) model to optimize the design and operation of the system by minimizing its life cycle cost for each year of the "future" dataset.

Preliminary results show that the representative sets of typical days of the grid electricity price according to the multi-year clustering and hybrid approaches are accurate. In fact, the optimal life cycle costs of the system obtained by solving the SP model in the "future" show average errors of 3% (multi-year clustering approach) and 5% (hybrid approach), respectively, compared to the solution of a Mixed-Integer Linear Programming (MILP) model based on the "perfect knowledge" of all 365 daily timeseries over years. The main advantage of the proposed hybrid approach over the multi-year clustering approach lies in the selection of a representative set of typical days of prices, which considers the actual values of the daily timeseries in the "future" (where the BPNN is tested) and does not neglect the relationships with those in the "past" (where the BPNN is trained). Although the proposed hybrid approach leads to an acceptable error in the optimal life cycle cost of the system, future work will focus on further reducing this error by improving the accuracy *i)* in the classification of the daily timeseries in the "past" dataset during the training of the BPNN and *ii)* in the selection of extreme daily timeseries of prices in the "future" dataset during the testing of the BPNN.

https://doi.org/10.52202/077185-0005

NOMENCLATURE

ANN	Artificial Neural Network
BPNN	Feedforward Backpropagation Neural Network
EC	Energy Community
MES	Multi-Energy System
MILP	Mixed-Integer Linear Programming
REC	Renewable Energy Community
RMSE	Root Mean Squared Error
SP	Stochastic Programming

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. (2022). "ARERA, Autorità di Regolazione per Energia Reti e Ambiente." Retrieved 11 May 2022, from https://www.arera.it/it/inglese/index.htm.

ARERA. (2023). "Testo integrato autoconsumo diffuso – TIAD." Retrieved 01 August 2023), from https://www.arera.it/allegati/docs/22/727-22alla.pdf.

Bahl, B., A. Kümpel, H. Seele, M. Lampe and A. Bardow (2017). "Time-series aggregation for synthesis problems by bounding error in the objective function." Energy 135: 900-912.

Bahl, B., T. Söhler, M. Hennen and A. Bardow (2018). "Typical periods for two-stage synthesis by time-series aggregation with bounded error in objective function." Frontiers in Energy Research 5: 35.

Bartolini, A., F. Carducci, C. B. Muñoz and G. Comodi (2020). "Energy storage and multi energy systems in local energy communities with high renewable energy penetration." Renewable Energy 159: 595-609.

Dal Cin, E., G. Carraro, A. Lazzaretto, G. Tsatsaronis, G. Volpato and P. Danieli (2023). Integrated design and operation optimization of multi-energy systems including energy networks. Proceedings of ECOS 2023 - The 36th International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems, Las Palmas de Gran Canaria, Spain.

DEA. (2023). "DEA - Danish Energy Agency, Technology Data." Retrieved 24 September 2023), from https://ens.dk/en/our-services/projections-and-models/technology-data.

Del Real, A. J., F. Dorado and J. Durán (2020). "Energy demand forecasting using deep learning: Applications for the French grid." Energies 13(9): 2242.

DOE. (2023). "U.S. Department of Energy. Open EI." Retrieved 14 March 2023), from https://openei.org/datasets/files/961/pub/.

Du, B., P. D. Lund and J. Wang (2022). "Improving the accuracy of predicting the performance of solar collectors through clustering analysis with artificial neural network models." Energy Reports 8: 3970-3981.

Erilli, N. A., U. Yolcu, E. Eğrioğlu, Ç. H. Aladağ and Y. Öner (2011). "Determining the most proper number of cluster in fuzzy clustering by using artificial neural networks." Expert Systems with Applications 38(3): 2248-2252.

EU. (2021). "European Commission's Communication, 'Fit for 55': delivering the EU's 2030 Climate Target on the way to climate neutrality (COM(2021) 550 final)." Retrieved 11 May 2022, from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021DC0550.

Fausett, L. V. (2006). Fundamentals of neural networks: architectures, algorithms and applications, Pearson Education India.

Fazlollahi, S., G. Becker and F. Maréchal (2014). "Multi-objectives, multi-period optimization of district energy systems: II—Daily thermal storage." Computers & Chemical Engineering 71: 648-662.

https://doi.org/10.52202/077185-0005

Fazlollahi, S., S. L. Bungener, G. Becker and F. Maréchal (2012). Multi-objectives, multi-period optimization of district heating networks using evolutionary algorithms and mixed integer linear programming (milp). Computer Aided Chemical Engineering, Elsevier. 30: 262-266.
Giannuzzo, L., F. D. Minuto, D. S. Schiera and A. Lanzini (2024). "Reconstructing hourly residential electrical load profiles for Renewable Energy Communities using non-intrusive machine learning techniques." Energy and AI 15: 100329.
GME. (2022). "Gestore Mercati Energetici." Retrieved 11 May 2022), from https://www.mercatoelettrico.org/en/default.aspx.
Guelpa, E., A. Bischi, V. Verda, M. Chertkov and H. Lund (2019). "Towards future infrastructures for sustainable multi-energy systems: A review." Energy 184: 2-21.
Hoffmann, M., L. Kotzur, D. Stolten and M. Robinius (2020). "A review on time series aggregation methods for energy system models." Energies 13(3): 641.
Hoffmann, M., J. Priesmann, L. Nolting, A. Praktiknjo, L. Kotzur and D. Stolten (2021). "Typical periods or typical time steps? A multi-model analysis to determine the optimal temporal aggregation for energy system models." Applied Energy 304: 117825.
Infanger, G. (1992). Planning under uncertainty solving large-scale stochastic linear programs, Stanford Univ., CA (United States). Systems Optimization Lab.
Kannengießer, T., M. Hoffmann, L. Kotzur, P. Stenzel, F. Schuetz, K. Peters, S. Nykamp, D. Stolten and M. Robinius (2019). "Reducing computational load for mixed integer linear programming: an example for a district and an Island energy system." Energies 12(14): 2825.
Kotzur, L., P. Markewitz, M. Robinius and D. Stolten (2018). "Impact of different time series aggregation methods on optimal energy system design." Renewable energy 117: 474-487.
Luo, X. (2020). "A novel clustering-enhanced adaptive artificial neural network model for predicting day-ahead building cooling demand." Journal of Building Engineering 32: 101504.
Mancarella, P. (2014). "MES (multi-energy systems): An overview of concepts and evaluation models." Energy 65: 1-17.
Mancò, G., U. Tesio, E. Guelpa and V. Verda (2023). "A review on multi energy systems modelling and optimization." Applied Thermal Engineering: 121871.
Mosavi, A., M. Salimi, S. Faizollahzadeh Ardabili, T. Rabczuk, S. Shamshirband and A. R. Varkonyi-Koczy (2019). "State of the art of machine learning models in energy systems, a systematic review." Energies 12(7): 1301.
Petrucci, A., G. Barone, A. Buonomano and A. Athienitis (2022). "Modelling of a multi-stage energy management control routine for energy demand forecasting, flexibility, and optimization of smart communities using a Recurrent Neural Network." Energy Conversion and Management 268: 115995.
PVGIS. (2022). "Photovoltaic Geographical Information System (PVGIS)." Retrieved 11 May 2022), from https://ec.europa.eu/jrc/en/pvgis.
Runge, J. and R. Zmeureanu (2019). "Forecasting energy use in buildings using artificial neural networks: A review." Energies 12(17): 3254.
Sterchele, P., J. Brandes, J. Heilig, D. Wrede, C. Kost, T. Schlegl, A. Bett and H.-M. Henning (2020). Paths to a Climate-Neutral Energy System. The German Energy Transition in its Social Context. Fraunhofer ISE2020.
Teichgraeber, H. and A. R. Brandt (2020). "Optimal design of an electricity-intensive industrial facility subject to electricity price uncertainty: Stochastic optimization and scenario reduction." Chemical Engineering Research and Design 163: 204-216.
Ząbkowski, T., K. Gajowniczek, G. Matejko, J. Brożyna, G. Mentel, M. Charytanowicz, J. Jarnicka, A. Olwert, W. Radziszewska and J. Verstraete (2023). "Cluster-Based Approach to Estimate Demand in the Polish Power System Using Commercial Customers' Data." Energies 16(24): 8070.
Zatti, M., M. Gabba, M. Freschini, M. Rossi, A. Gambarotta, M. Morini and E. Martelli (2019). "k-MILP: A novel clustering approach to select typical and extreme days for multi-energy systems design optimization." Energy 181: 1051-1063.
Zheng, Z., X. Li, J. Pan and X. Luo (2021). "A multi-year two-stage stochastic programming model for optimal design and operation of residential photovoltaic-battery systems." Energy and Buildings 239: 110835.

 https://doi.org/10.52202/077185-0005