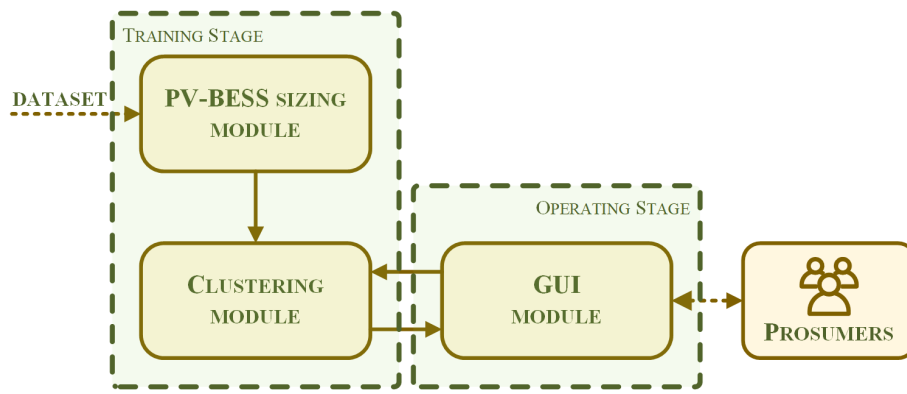# Graphical Abstract

An Online Energy Management Tool for Sizing Integrated PV-BESS Systems for Residential Prosumers

Saman Korjani, Fabio Casu, Alfonso Damiano, Virginia Pilloni, Alessandro Serpi

Highlights

An Online Energy Management Tool for Sizing Integrated PV-BESS Systems for Residential Prosumers

Saman Korjani, Fabio Casu, Alfonso Damiano, Virginia Pilloni, Alessandro Serpi

- Hourly-based electricity generation and consumption profiles of 128 residential prosumers have been analyzed

- Self-sufficiency maps with different combinations of sizes of hybrid Photovoltaic-Battery Energy Storage System (PV-BESS) have been found

- Similar self-sufficiency maps have been clustered by identifying the most significant features of their prosumers

- Results show that Heat Ventilation Air Conditioning (HVAC) systems, water heaters, and peak electricity consumption represent the most important features influencing clustering

- An online energy management tool that recommends PV-BESS sizes according to the answers to a few simple questions related to user's energy consumption habits has been proposed

An Online Energy Management Tool for Sizing Integrated PV-BESS Systems for Residential Prosumers

Saman Korjani, Fabio Casu, Alfonso Damiano, Virginia Pilloni, Alessandro Serpi

[a]*University of Cagliari, Department of Electrical and Electronic Engineering, Via Marengo, 2, Cagliari, 09123, Sardenia, Italy*

Abstract

This paper presents an online energy management tool that suggests the most suitable size of a hybrid photovoltaic-battery energy storage system (PV-BESS) to residential prosumers based on their self-sufficiency expectations. An offline analysis of electricity generation and consumption expected from 128 residential prosumers has been carried out at first in order to find out their self-sufficiency map with different sizes of PV and BESS; this is carried out by the genetic algorithm based energy management (GA) presented in a previous work. Subsequently, a number of clusters have been defined, each of which groups prosumers that share similar self-efficiency maps; particularly, clustering has been carried out and refined by identifying the most significant features of prosumers belonging to the same cluster, as well as those that differentiate prosumers belonging to different clusters. As a result, it has been revealed that the habit of usage of some appliances, such as Heat Ventilation Air Conditioning system (HVAC) and water heater, and peak electricity consumption represent the most important features influencing clustering. Based on these outcomes, the proposed online energy management tool is able to assign a prosumer to the most suitable cluster just based on the answers to a few simple questions related to their energy consumption habits, providing the corresponding self-efficiency map almost immediately. The results achieved by the proposed tool, which is currently running on-line, are promising and show that significant self-sufficiency increases can be obtained, allowing the proper choice of PV-BESS depending on specific prosumer's needs and expectations.

Keywords: Battery energy storage systems, Clustering, Energy management, Energy self-sufficiency, Photovoltaic power plants, Sizing

---

1. Introduction

United Nations goals, particularly the Paris Agreement of 2015 for the mitigation of climate change, pointed out the fundamental role of Renewable Energy Sources (RESs) to decrease the amount of CO_2 produced by traditional thermal power plants [1]. Since residential energy consumption accounts for 35% of the total energy consumed worldwide [2], in recent years many energy users have been encouraged to install photovoltaic (PV) power plants [3]. However, PV generation is intermittent and, thus, weakly programmable, resulting in frequent imbalances between production and consumption that lead to a number of potential challenges, such as voltage instability [4], frequency deviations [5], and reverse power flow phenomena [6]. Two solutions are generally suggested for solving these imbalances: i) demand side management, namely consumption is shifted to the times that better match production [7]; ii) the use of Energy Storage Systems (ESSs), where excess energy can be stored and delivered back when consumption exceeds production [8]. Both solutions are already being applied since they increase self-consumption capabilities, improving the time-balancing between consumption and production [3, 9, 10]. However, demand side management has limited applicability as matching consumption with production in real-time is usually complex to achieve [11]. Consequently, the integration of PV with Battery ESS (BESS) is becoming a popular solution for increasing self-consumption of residential prosumers [12].

Self-consumption provides several advantages for both prosumers and Distribution System Operators (DSOs) [13], such as economic profit for prosumers [14, 15], RES integration into the grid [16], peer-to-peer energy trading by maximizing the reserve capacity of energy prosumers [17], power quality improvement [18], preventing or mitigating grid over-voltage and overloads due to Electrical Vehicle (EV) charging [19, 20, 21], peak shaving [22], demand-response programming [23], grid losses and CO_2 emission reduction [13, 24], and reduced network investments [25].Various operation strategies for residential PV-BESS self-consumption are studied in different publications. In [3], the PV self-consumption in buildings obtained by BESS is higher than by demand side management. In [26], BESS contribution to increase PV self-consumption is investigated; the corresponding results show

2

that PV-BESS systems increase the total consumption of PV energy at the Point of Common Coupling (PCC). Also, the increased PV self-consumption rate in residential prosumers by using charge-discharge control strategy of Electric Vehicle (EV) has been considered [19].

To get full benefit from self-consumption, the optimal sizing of PV-BESS system is an important factor [27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39], which should account for the different consumption habits of prosumers, even during the day. For instance, yearly operation of BESS in presence of PV power plants has been optimised in [33] by considering market condition and BESS lifetime at the aim of obtaining optimal sizes of PV-BESS systems. An iterative optimization method to obtain the optimal size of a PV-BESS system in a microgrid for minimising annual costs is presented in [34]. The authors of [35] develop an analytical and economical model based on Levelized Cost of Energy (LCOE) to determine the best size of PV-BESS systems in a interconnected grid, which increase the self-consumption through PV-BESS and PV by 75% and 48%, respectively. In [36], a PV-BESS sizing approach based on a Linear Programming (LP) optimization is proposed for distribution networks by referring to different scenarios and targets, among which grid self-sufficiency. Meanwhile, [36] analyses a BESS siting strategy for the various distribution networks for the goal of voltage stability. On the other had, [37] presents an optimal sizing of BESS for Net-Zero-Energy homes with PV to minimize the annual net electricity and BESS costs. Similarly, [28] and [38] present a BESS management and planning strategy for reducing electricity bills and achieving Net-Zero-Energy homes equipped with PV. Considering grid-connected residential PV-BESS systems, [40] presents different self-consumption maximisation strategies, particularly BESS size is optimsed according to economic criteria through a web-based application, which is designed by using a neural network model. Differently, an optimal planning and management method for BESS integrated with existing PV is presented in [39] with the aim of minimising PV power curtailments.

Optimal sizing of PV-BESS system is pursued also for purposes different from self-consumption, such as economic benefits and/or power system resiliency. In this regard, the optimal size of a PV-BESS system that maximises the prosumer's profit is determined in [41] by LP, which finds out the optimal sizing parameters of the system. In [42], a sizing strategy of a PV-BESS system is proposed to maximise economic and resilience benefits, which is based on a stochastic optimization methodology and Monte Carlo simulation considering storm-related outage of electricity. The authors of [43]

3

propose a combined sizing and scheduling approach of PV-BESS systems aimed at increasing prosumers' profit, as well as improving voltage quality of the distribution system.

Despite many studies on PV-BESS sizing, this is still a challenging task [44, 45] and, thus, worth of investigation. In this regard, the need for a simple but effective sizing tool seems required as prosumers are not generally fully-aware of their own electricity consumption and/or production profiles and, anyway, they are not generally able to provide them in real-time. In addition, the optimal PV-BESS sizing procedures proposed in the literature generally provide just a single outcome, i.e. the best PV-BESS combination, whereas multiple solutions should be provided as as far as multi-objective optimisation is concerned.

In this context, an online energy management tool is proposed in this paper, which aims at increasing prosumer's self-sufficiency, namely the gross energy exchanged with the electric grid, by recommending the combinations of PV and BESS sizes that best suit their energy consumption profiles. Differently from many other solutions presented in the literature, the proposed online energy management tool does not provide a single outcome, e.g. the best PV-BESS size, but a self-sufficiency map, each element of which represents the increased self-sufficiency that can be achieved by the prosumer with a given PV-BESS size. In order to enable the tool to work without the need for users to provide their own energy consumption profiles, self-sufficiency maps are computed offline using the consumption profiles available from 128 residential prosumers as a reference [46], together with PV expected production profiles; these data have been processed through the genetic algorithm based energy management (GA) presented in [47] for determining the maximum self-efficiency achievable for each (PV, BESS) combination. The achieved self-efficiency maps are thus processed further by the Hierarchical Clustering (HC) method to identify few representative clusters, each of which groups prosumers with similar self-efficiency capabilities [45]. Differently from previous works [45, 47], prosumers' profiles have been analyzed in detail to assess which common features identify/differentiate each cluster. Indeed, some features, such as usage of Heat Ventilation Air Conditioning system (HVAC) and water heater, and the peak consumption reveal different impacts on the hourly energy consumption profile of prosumers of specific clusters, differentiating them from the others. Such features thus represent the fingerprint of the clusters that enable the proposed online energy management tool to easily and quickly associate unknown on-line users to a cluster after mak-

4

ing them some simple but relevant questions; consequently, the proposed tool does not require precise information on prosumers' consumption profile, which is generally hard to achieve and/or provide, especially in real-time. As a result, any potential prosumer who wants to investigate the possibility of installing a PV-BESS system is associated to the most suitable cluster almost instantaneously, and the corresponding self-efficiency map, which is stored together with those of the other clusters into a suitable database, is retrieved and showed on the platform, by letting him/her choose the most suitable size in accordance with his/her expected self-sufficiency increase and system cost.

It should be noted that, to the best of the authors' knowledge, none of the papers already published in the literature dealing with PV-BESS and self-sufficiency proposes an online energy management tool that provides a comprehensive PV-BESS sizing resulting in a self-sufficiency map (not in a single optimal PV-BESS sizing) based on prosumer's consumption habit of some appliances. Another main allotment of this paper is the online management tool; differently from commercial tools developed by PV companies, the proposed tool optimises BESS and PV sizes simultaneously. Also, other tools provided by research agencies like IRENA, NREL, Google, etc., do not take into account consumption habit of users and/or provide a single optimal sizing outcome. Consequently, the proposed energy management tool could fill a gap in existing commercial and research-grade products.

The paper is organized as follows: Sections 2 and 3 introduce a general overview of the proposed methodology and methods, respectively. Section 4 describes the cases of study and data; Section 5 describes and analyses the results; conclusion is given in Section 6.

2. General overview of the proposed recommendation tool

The schematic architecture of the proposed PV-BESS recommendation tool is shown in Fig. 1. Two main stages can be distinguished, namely the training and operating stages, which are carried out offline and in real-time, respectively. During the training stage, several prosumer consumption and PV production profiles are processed by the PV-BESS sizing module, which has the task of computing the corresponding self-sufficiency maps through a GA-based energy management algorithm already presented in [47]. These profiles include data collected over at least one year to account for seasonality that generally affects both consumption and PV production profiles.

5

Furthermore, the higher the number of similar prosumers monitored, the higher the resulting accuracy of the following clustering procedure.
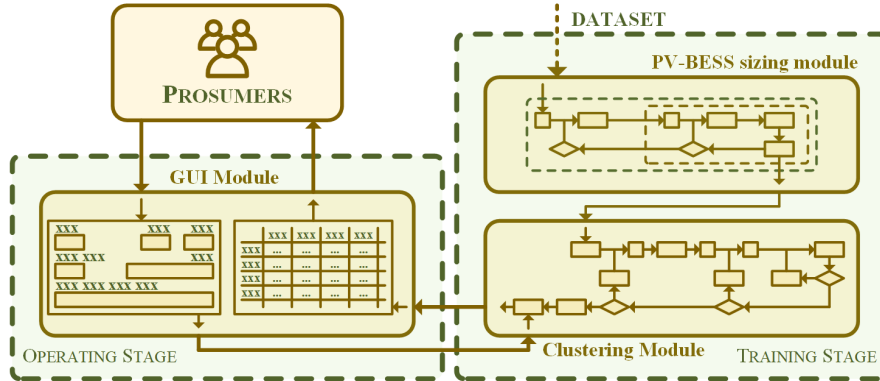

Figure 1: Overall architecture of the proposed on-line energy management tool.

Once all the maps have been computed, they are sent to the Clustering module to be grouped in accordance with their degree of similarity. In this regard, the clustering module operates in accordance with the most significant features that characterize each cluster, i.e. the common features of all its elements and that mostly affect them; these features are identified based on an accurate analysis and post-processing of the outcomes of the PV-BESS sizing module, but also considering that these features should be collected in real-time from any prosumer to assign him/her to the most suitable cluster. Consequently, too much technical information and/or accurate data (e.g. yearly consumption profiles) have been avoided to broaden the catchment area of the proposed tool. Each resulting cluster is then associated with an average self-sufficiency map, which best fits the prosumers' profiles included in that cluster.

Considering the operating stage, it consists of a Graphical User Interface (GUI) module, i.e. an online form that enables each user to fill in some basic information, such as the average monthly energy consumption, the number of inhabitants, the presence of some specific appliances (e.g. an electric water heater), and so on. These information are sent to the Clustering Module, which assigns the user to the most suitable cluster, giving back the corresponding self-sufficiency map; this process can be accomplished almost instantaneously so that the user can view its self-sufficiency map in real-time.

Based on this outcome, the user can thus choose the PV-BESS sizing most suitable for him/her, e.g. a trade-off between increased self-sufficiency and installation costs.

In conclusion, since the PV-BESS sizing module has been already detailed in [47], it is just briefly summarized in the following section, which instead focuses mostly on the Clustering and GUI modules.

3. Methods

3.1. PV-BESS sizing module

The flowchart representing the PV-BESS sizing module, which has been already proposed in [47], is shown in Fig. 2. The procedure starts from a user's yearly consumption (P_c) and PV production profiles (P_g), which can be respectively achieved from a suitable dataset [46, 48] and determined based on user's geographical location. Subsequently, the PV production profile is scaled through a suitable coefficient (S_n) for emulating different PV sizes by varying n within $[1, N]$, and it is subtracted from P_c to achieve the user's residual power profile (P_r). Afterwards, different BESS sizes are considered by introducing a sizing coefficient (K_m) that amplifies both BESS energy and power base capability (E_0 and P_0, respectively). Consequently, for each (S_n,K_m) pairs of values, the BESS is driven to maximize the user's self-sufficiency through a Genetic Algorithm Multi Period Power Flow approach (GA-MPOPF) [49, 50], by complying with all the system constraints. The GA-MPOPF ends when both n and m reach the corresponding maximum values (N and M, respectively, which should be set in accordance with typical maximum sizes for residential users), delivering the user's self-sufficiency map to the Clustering module.

3.2. Clustering module

The flowchart of the Clustering module is shown in Fig. 3, in which d and l denote the generic distance and linkage criterion, respectively, which can vary within $[1, D]$ and $[1, L]$, respectively. In addition, N_c is the generic number of clusters, and W is the overall number of prosumers. The overall flowchart can be split into four main steps, namely feature extraction (step 1), distance metric criterion (step 2), clustering method and linkage criterion (step 3), and defining clusters (step 4), each of which is described in the following subsections.
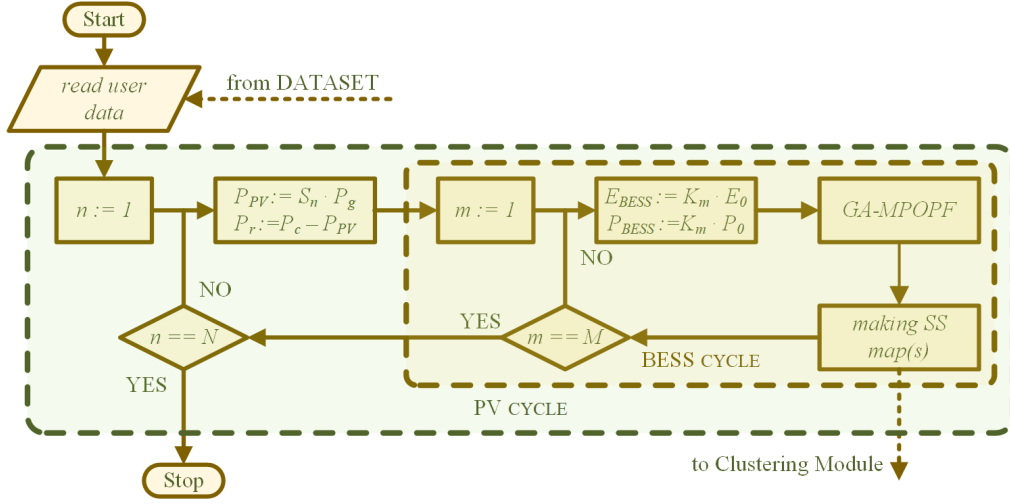
7

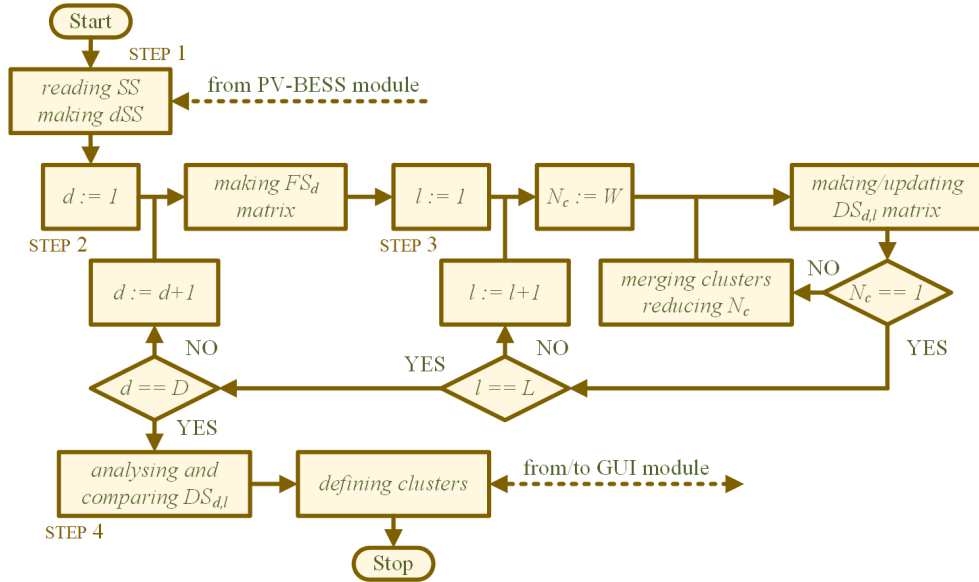Figure 2: Flowchart of the PV-BESS sizing module.

Figure 3: Flowchart of the proposed Clustering module.

8

3.2.1. Step 1: feature extraction

The first step of the proposed Clustering module consists of extracting users' most distinctive features from the self-sufficiency matrix (SS) previously determined by the PV-BESS sizing module. In particular, each element ss_{ij} represents the self-sufficiency achievable by the user through the employment of a combined PV-BESS system of size S_n and K_m, respectively. This matrix is elaborated further in order to achieve a per-unit differential SS (dSS) as:

$$dSS = \frac{SS_{00} - SS}{ss_{00}} \tag{1}$$

where SS_{00} is a matrix of the same size of SS but with all elements equal to ss_{00}; the latter represents the self-sufficiency of the user with no PV-BESS system, namely his/her starting yearly gross energy exchange with the grid. Equation 2 shows an example of dSS, in which positive values mean saved yearly gross energy exchange and, thus, increased self-sufficiency. On the other hand, negative values mean increased energy exchange between the user and the grid, which may occur due, for example, to an oversized PV.

$$dSS = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0.12 & 0.13 & 0.13 & 0.13 \\ 0.23 & 0.28 & 0.36 & 0.44 \\ -0.031 & 0.07 & 0.11 & 0.31 \end{pmatrix} \tag{2}$$

In addition to dSS, additional features could be employed in accordance with the feature extraction criterion, as pointed out in Section 5.2. Consequently, users' features are more generally denoted by the object vector F in the following.

3.2.2. Step 2: distance metric criterion

Given F of all W prosumers, a feature similarity matrix (FS) has been produced at first by using correlation and distance among all the different F. To this purpose, as no general rule applies, various distance and correlation metrics have been evaluated ($d = 1..D$), such as Pearson, Spearman rank and Kendall Tau rank correlations, as well as Minkowski and Euclidean distances, each of which are explained briefly in the following:

- The Pearson correlation ($FS^{(P)}$) is a measure of the linear correlation between two data sets [51], particularly FS^P of two feature sets $F^{(1)}$ and $F^{(2)}$ is defined by the following equation (3):

$$FS_{1,2}^{(P)} = \frac{\sum_{i=1}^{Nv}(F_i^{(1)} - \bar{F}^{(1)})(F_i^{(2)} - \bar{F}^{(2)})}{\sqrt{\sum_{i=1}^{Nv}(F_i^{(1)} - \bar{F}^{(1)})^2}\sqrt{\sum_{i=1}^{Nv}(F_i^{(2)} - \bar{F}^{(2)})^2}} \tag{3}$$

where Nv^x is the number of elements of $F^{(x)}$, and $\bar{F}^{(x)}$ represents the average value of $F^{(x)}$, namely $\bar{F}^{(x)} = \frac{1}{Nv}\sum_{i=1}^{Nv} F_i^{(x)}$.

- The Spearman rank correlation $(FS^{(S)})$ assesses monotonic relationships between two variable sets [51], whether linear or not, and it is defined as the Pearson correlation between the rankings of two data sets. Particularly, a ranking is the ordering labels of given data sets according to specific criterion, e.g. ascending or descending order. Consequently, equation (4) holds:

$$FS_{1,2}^{(S)} = \frac{\sum_{i=1}^{Nv^1}(R(F_i^{(1)}) - \bar{R}(F^{(1)}))(R(F_i^{(2)}) - \bar{R}(F^{(2)}))}{\sqrt{\sum_{i=1}^{Nv^1}(R(F_i^{(1)}) - \bar{R}(F^{(1)}))^2}\sqrt{\sum_{i=1}^{Nv^2}(R(F_i^{(2)}) - \bar{R}F^{(2)}))^2}} \tag{4}$$

where $R(F^{(1)})$ and $R(F^{(2)})$ are the ranks of $F^{(1)}$ and $F^{(2)}$, respectively, while $\bar{R}(F^{(1)})$ and $\bar{R}(F^{(2)})$ are their corresponding average values.

- The Kendall Tau rank correlation $(FS^{(K)})$ is a measure of rank correlation between two data sets based on the similarity of their rankings [52], as defined by equation (5):

$$FS_{1,2}^{(K)} = \frac{(H - L)}{\sqrt{((H + L + T) \times (H + L + U))}} \tag{5}$$

where H is the number of concordant pairs, L the number of discordant pairs, while T and U are the number of $F^{(1)}$ and $F^{(2)}$ exclusive ties, respectively. In particular, a concordant pair is a pair of observations, e.g. (X_1, Y_1) and (X_2, Y_2), having the following property:

$$\mathrm{sgn}\,(X_2 - X_1) = \mathrm{sgn}\,(Y_2 - Y_1) \tag{6}$$

where *sgn* is the signum function. If (6) is not valid, a discordant pair occurs. On the other hand, an exclusive tied pairs is neither concordant

nor discordant . Consequently, if a tie occurs for the same pair in both $F^{(1)}$ and $F^{(2)}$, a tie pair is not added to neither T nor U.

- The Minkowski distance $(FS^{(M)})$ of two vectors consists of the norm of their difference [53], as shown in equation (7):

$$FS_{1,2}^{(M)} = \left(\sum_{i=1}^{Nv} |F_i^{(1)} - F_i^{(2)}|^p \right)^{\frac{1}{p}} \tag{7}$$

where p can be set in accordance with the positive infinity Chebyshev distance [54], as described by equation (8):

$$\lim_{p \to \infty} \left(\sum_{i=1}^{Nv} |F_i^{(1)} - F_i^{(2)}|^p \right)^{\frac{1}{p}} = \max_{i=1}^{Nv} |F_i^{(1)} - F_i^{(2)}| \tag{8}$$

- Euclidean distance $(FS^{(E)})$ of two feature sets can be obtained by equation (7) by considering $p = 2$ [53].

Therefore, since none of these correlations is inherently the best for this study, all of them are used alternatively to produce FS, leading to different clustering results; these are then compared to each other to select the most reliable distance metric criterion for the proposed Clustering module, as detailed in Section 5.

3.2.3. Step 3: clustering method and linkage criterion

Once FS has been computed, a clustering method has to be chosen. In this regard, the Hierarchical Clustering method (HC) has been selected for this study; particularly, in data mining and statistics, HC is an algorithm that groups similar objects into clusters; the endpoint is a set of clusters, each of which is distinct from the others so that the objects belonging to each cluster are broadly similar to each other [55]. Therefore, HC groups objects based on the principle of correlation-distance-based clustering, according to which closer objects are grouped in the same cluster. Consequently, HC has been chosen for this study because it gives a deep insight of each step of the clustering process, creating a dendogram that helps to figure out the most suitable cluster combination and, thus, the final outcome of the clustering procedure [55].

11

Two different types of HC can be used, namely Agglomerative and Divisive; the former means that each object is considered as a single cluster at first, sequentially merging similar objects to each other. On the other hand, Divisive HC initially groups all objects into one cluster, successively exploding it into multiple clusters. Agglomerative HC has been chosen for this study, but it is worth mentioning that both the HC types lead to the same final result.

HC computes distances based on two criteria, i.e. metric and linkage, leading to a distance matrix ($DS_{d,l}$). In particular, the distance metric criteria have been already defined at Step 2, while the linkage criterion is related to inter-cluster distance, i.e. the distance between two generic clusters C_i and C_j. As the choice of the linkage criterion is not constrained by any rule, various linkages criterion can be used alternatively ($l = 1..L$), namely single, ward, complete, average, and centred linkage, to determine from where distances should be calculated [55], as briefly described in the following:

- Single linkage (nearest neighbor, d_S) [55]: the distance (dissimilarity) between two clusters C_i and C_j is measured as the distance between the two most similar (closest) points belonging to the two different clusters:

$$d_S\left(C_i, C_j\right) = \min_{x \in C_i, y \in C_j}\{d(x, y)\} \tag{9}$$

- Complete linkage (furthest neighbor, d_C) [55]: the distance between two clusters C_i and C_j is measured as the distance between the two least similar (furthest) points belonging to different clusters:

$$d_C\left(C_i, C_j\right) = \max_{x \in C_i, y \in C_j}\{d(x, y)\} \tag{10}$$

- Average linkage (unweighted pair-group average, d_A) [55]: the distance between two clusters C_i and C_j is the average value of the pairwise distances between points of the two clusters:

$$d_A\left(C_i, C_j\right) = \text{avg}_{x \in C_i, y \in C_j}\{d(x, y)\} \tag{11}$$

- Centroid linkage (d_{Ce}) [55]: the distance between two clusters C_i and C_j is the distance between their centroids c_i and c_j:

$$d_{Ce}\left(C_i, C_j\right) = d\left(c_i, c_j\right) \tag{12}$$

in which c_i is defined by equation (13) [55]:

$$c_i = \left(\frac{1}{n_{C_i}} \sum_{n=1}^{n_{C_i}} C_i^n\right) \tag{13}$$

where n_{C_i} is the number of the objects C_i^n of the cluster C_i.

- Ward linkage (d_W) [56]: the distance between two clusters C_i and C_j are defined by comparing the squared distances from the centroid of the "merged" cluster with those from the centroids of the original clusters, as defined by the following equation:

$$d_W(C_i, C_j) = \sum_{p \in C_i \cup C_j} \left\|\vec{x}_p - \vec{m}_{C_i \cup C_j}\right\|^2 - \sum_{p \in C_i} \left\|\vec{x}_p - \vec{m}_{C_i}\right\|^2$$

$$- \sum_{p \in C_j} \left\|\vec{x}_p - \vec{m}_{C_j}\right\|^2 = \frac{n_{C_i} n_{C_j}}{n_{C_i} + n_{C_j}} \left\|\vec{m}_{C_i} - \vec{m}_{C_j}\right\|^2 \tag{14}$$

where $\vec{m}_{C_i}$ is the centroid of cluster C_i, while n_{C_i} is the number of its objects.

Once the distance metric and linkage criterion have been chosen, the distance matrix $(DS_{d,l})$ can be achieved for each (d, l) pair of values. As a result, all the $DS_{d,l}$ can be analysed and compared to each other to determine the optimal number of clusters in accordance with the Silhouette-Score and Elbow methods, as detailed in the following subsection.

3.2.4. Step 4: defining clusters

The optimal number of clusters is commonly achieved based on the Silhouette Score (S_s), which estimates how close each point in a generic cluster is to the points of the neighboring clusters by the following equation (15) [45, 57]:

$$S_s = \frac{(b-a)}{max(a, b)} \tag{15}$$

where a is the average intra-cluster distance and b is the average nearest-cluster distance, i.e. b is the average distance between points of a cluster and those of the nearest cluster. Therefore, S_s ranges from -1 to $+1$, where values near $+1$ mean that objects well match their own cluster and poorly match neighbouring clusters. On the other hand, S_s values near 0 mean that objects are on or very close to the decision boundary between neighboring clusters, while negative values reveal that the objects might belong to the wrong cluster. Consequently, achieving the highest value of S_s corresponds to optimizing the overall number of clusters.

Alternatively to S_s, clustering optimization could be achieved through the Elbow method (EL) [58], which is represented by the following equation (16):

$$EL_k = \sum_{i=1}^{N_c} \frac{1}{n_{C_i}} D_{C_i} \qquad (16)$$

where N_c is the overall number of clusters, n_{C_i} is the number of objects in cluster C_i, and D_{C_i} is the sum of distances (e.g. Euclidean) among all the objects in cluster C_i. In particular, EL represents a cut-off point beyond which increased returns do not further worth the additional costs. Therefore, for this study, this means that adding another cluster does not improve clustering significantly.

In conclusion, by applying both methods, if the obtained numbers of clusters are the same, clustering results are considered consistent. Otherwise, the second and, more in general, the next highest S_s would be considered until both methods converge to the same optimisation result.

3.3. GUI module

The results of the Clustering module are used to setup the questions to ask to prosumers through an online form, an example of which is shown in Fig. 4. In this regard, it is worth noting that the proposed online tool is already running and available at [59].

More specifically, the questions are determined based on the most distinctive features shared by prosumers belonging to the same cluster. Therefore, according to prosumer's answers, the GUI module selects his/her most fitting cluster through the support of the Clustering module, giving back the corresponding results; this operation occurs almost instantaneously as all massive computations are carried out offline, which is one of the main goal of the

Figure 4: A screenshot of the online form provided by the GUI module to prosumers.

proposed tool. In particular, the results consist of four tables, in which rows and columns correspond to different PV and BESS sizes, respectively;

- the average per-unit differential self-sufficiency map of the cluster to which the prosumer belong (dSS), which represents the amount of saved gross energy exchange with the grid;

- the amount of saved CO_2 emissions, which is determined based on both PV production and saved energy drawn from the grid;

- the number of trees that would have absorbed the saved CO_2 emissions;

- the cost corresponding to each PV-BESS combinations, including installation, inverter, battery and maintenance.

Based on these outcomes, the prosumer can select the PV-BESS combination most suitable for him/her depending on its needs, such as a target self-sufficiency value and/or the masimum budget available.

Table 1: Derived profiles from each prosumer in accordance with appliance usage

Prosumer Type Appliances	Type 1	Type 2	Type 3	Type 4
HVAC	✓	✓	✗	✗
Water heater	✓	✗	✓	✗
Interior-Exterior lighting	✓	✓	✓	✓
Ventilation Fans	✓	✓	✓	✓
Furniture (TV, refrigerator, $\cdots$)	✓	✓	✓	✓
Other (Computer, mobile, $\cdots$)	✓	✓	✓	✓

4. Case of Study and Data Set

The proposed online energy management tool has been trained by referring to several prosumer's annual consumption and PV annual production profiles [46, 48]. In particular, one-year time horizon has been chosen to account for seasonal effects properly. Residential prosumers' annual consumption profiles are obtained from [46], and they are sampled with one-hour time step, from January 1, 2004 to December 31, 2004. These consumption profiles correspond to different residential users from various US states, so they account also for different climate zones (very cold, cold, mixed-humid, mixed-dry, hot-dry, hot-humid, and marine) [60]. Totally, 128 user's profiles are considered, from each of which four different derived profiles have been achieved according to the use of different appliances, i.e. from Type 1 to Type 4, as summarized in Table 1. The PV production profiles are obtained from [48] and they are sampled hourly, from January 1, 2006 to December 31, 2006. In particular, various PV production profiles are available for each US state, with different latitude and longitude; consequently, the PV data have been selected and associated to each prosumer based on his/her geographical location. In this regard, although both prosumer's annual consumption and PV production come from relatively old datasets, this does not impair the general usefulness of the proposed online energy management tool, which can be trained with any dataset on the condition that the usage of different appliances is clearly distinguishable.

Given both prosumer's consumption and PV production profiles, the corresponding self-sufficiency maps have been determined by the PV-BESS sizing module, whose operating principle is detailed in [47]. In this regard, PV and BESS sizing steps and ranges have been defined in accordance with peak

16

Table 2: Sizing range of PV-BESS

	$P_c^{(peak)} \leq 4$	$4 < P_c^{(peak)} \leq 12$	$P_c^{(peak)} > 12$
PV size range (kW_p)	0 - 6.4	0 - 12.8	0 - 19.2
PV size steps (kW_p)	0.8	1.6	2.4
BESS size range (kWh)	0 - 32	0 - 64	0 - 96
BESS steps (kWh)	4	8	12

prosumer's consumption ($P_c^{(peak)}$), as highlighted in Table 2; it is worth mentioning that although this value should be difficult to be provided by a generic prosumer, it can be reasonably assumed equal to his/her maximum contractual power subscribed with the DSO. In addition, BESS has been always assumed charging and discharging at 1 C, so its rated power always equals its capacity. In addition, charging and discharging efficiencies have been set both at 0.95, leading to an overall round-trip efficiency of approximately 0.9.

A self-sufficiency matrix requires approximately 6 hours of simulation on Matlab by using 40 CPUs simultaneously of DELL-Workstation which main characteristics are Intel(R) Xeon(R) CPU E5-2698 v3 @ 2.30 GHz, CPU MHz: 3236.761, CPU(s): 64,On-line CPU(s) list:0-63, Core(s) per socket:16, Socket(s): 2, RAM 125 GB. Once the PV-BESS sizing module had determined the self-sufficiency matrix of all the prosumers, the Clustering module was run, whose results are presented and discussed in the following Section.

5. Results and Discussion

5.1. Clustering module, case 1: F consisting of dSS only

The HC method is firstly applied by considering the feature set (F) consisting of the per-unit differential self-sufficiency matrix (dSS) only. Therefore, various feature similarity matrices (FS) are produced according to different distance metrics and linkage criteria, as pointed out in Section 3; the Silhouette-Score (S_s) and Elbow methods (EL) are then applied to each FS, leading to the results shown in Figs. 5 and 6.

Considering Fig. 5a at first, it can be seen that the highest S_s value is obtained by Pearson correlation ($FS^{(P)}$) and Ward linkage (d_W); this leads to an optimal number of clusters equal to 3, as highlighted in Fig. 5b.

Fig. 6 illustrates the outcomes of the Elbow method in case of $FS^{(P)}$ and d_W; particularly, the optimal number of clusters is achieved when the 2nd

derivative of distances with the number of clusters becomes negative. Consequently, the Elbow method leads to the same optimal number of clusters
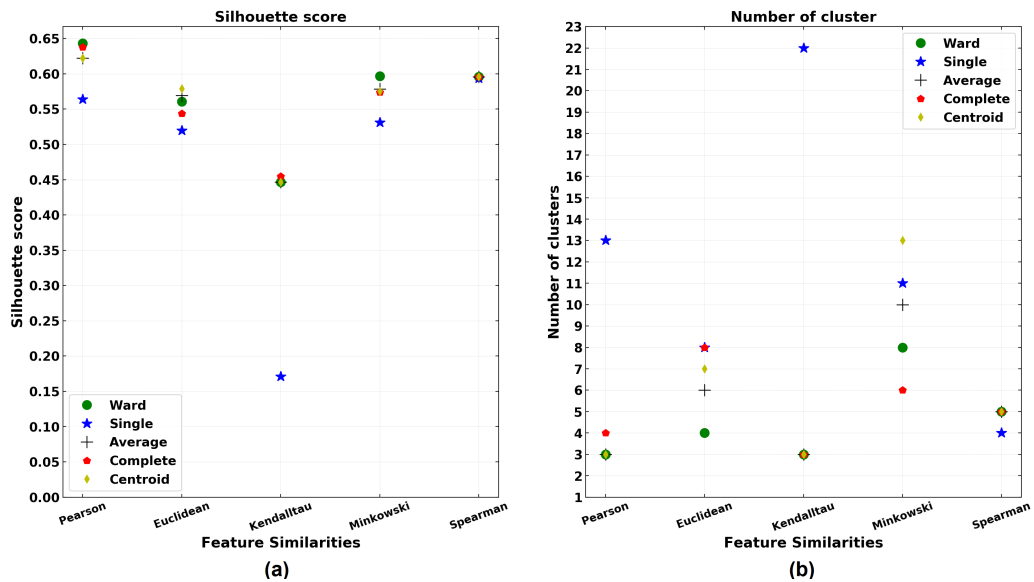

Figure 5: S_s (a) and optimal number of clusters (b) achieved with various distance metric and linkage criteria (case 1).
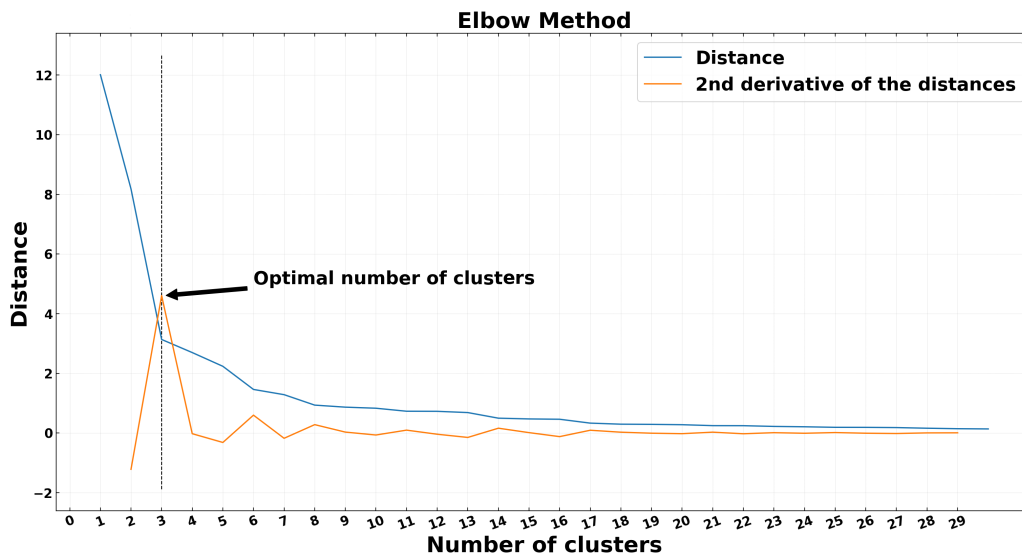

Figure 6: Elbow method outcomes in terms of optimal number of clusters (case 1).

of the S_s method, proving the consistency of the clustering results.

Clustering map of case 1 is shown in Fig. 7. This figure shows the distance between the prosumers obtained by using $FS^{(P)}$ and d_W. Fig. 8 shows the type of prosumers in the each cluster. Particularly, Cluster 1 is dominated by Type 1 and few Type 2 prosumers, Cluster 2 by Type 4 and some Type 2 and Type 1 prosumers, while Cluster 3 is dominated by some Type 2 and most of Type 3 prosumers. At a first glance, clustering results are not convincing as prosumers with similar consumption pattern should belong to the same cluster, i.e. Cluster 2 is in contradiction with the goal of the proposed methodology. After a detailed data and clustering analysis, it has been found out that such an unsuitable clustering outcome is because $P_c^{(peak)}$ of Type 1 prosumers belonging to Cluster 2 is different from that of Type 1 prosumers belonging to Cluster 1, namely $4 < P_c^{(peak)} \leq 12$ instead of $4 < P_c^{(peak)}$. This mismatch makes dSS of Type 1 prosumers with a relative high $P_c^{(peak)}$ to be similar to that of Type 4 prosumers. Once identified, this issue has been addressed by considering additional prosumers' features to dSS, as detailed in the following subsection.
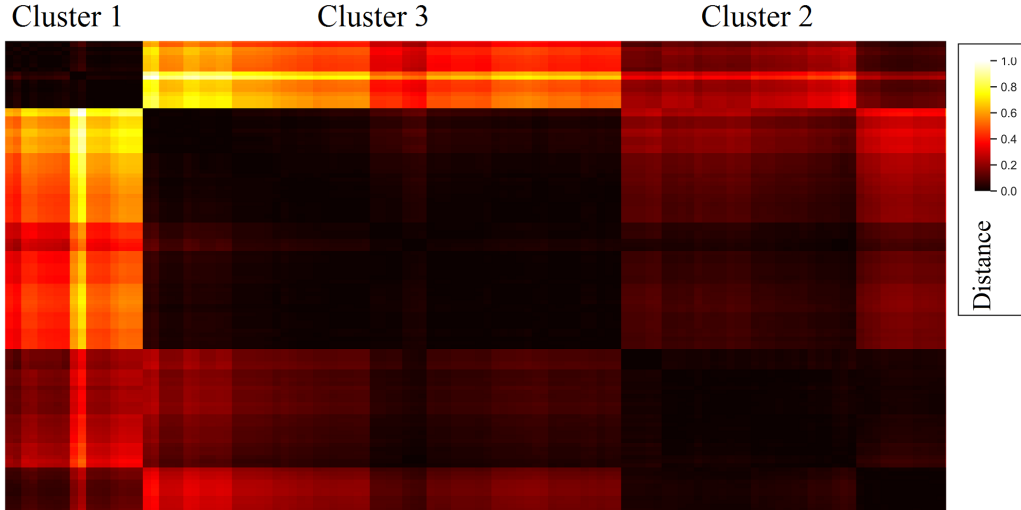

Figure 7: Cluster map (case 1).

5.2. Clustering module, case 2: F based on dSS maps and additional features

The unsuitable clustering results achieved in case 1 have been overcome by expanding the feature set F with two additional variables, namely the sizing

Table 3: Sizing index of PV and BESS (I_{size})

	$P_c^{(peak)} \leq 4$	$4 < P_c^{(peak)} \leq 12$	$P_c^{(peak)} > 12$
I_{size}	0	1	2

index of PV and BESS (I_{size}), and the normalised peak power consumption ($P_{c,n}^{(peak)}$). In particular, I_{size} depends on $P_c^{(peak)}$ as shown in Table 3, while ($P_{c,n}^{(peak)}$) is computed by getting all prosumers' $P_c^{(peak)}$ and normalising them between 0 and 1 using the following equation (17)

$$P_{c,n}^{(peak)} = \frac{A^{(p)} - A_{min}^{(p)}}{A_{max}^{(p)} - A_{min}^{(p)}} \tag{17}$$

where $A^{(p)}$ represents the set of all prosumers' $P_c^{(peak)}$, while $A_{min}^{(p)}$ and $A_{max}^{(p)}$ are the minimum and maximum value of $A^{(p)}$, respectively.

The clustering results achieved with this new feature sets (case 2) are shown from Fig. 9 to Fig. 11. Considering Fig. 9a at first, it can be seen that the highest S_s value is still achieved by $FS^{(P)}$ and d_W, and it is greater than that achieved in case 1. However, given $FS^{(P)}$, any linkage criterion leads to the same S_s value, whereas different performance are achieved in case 1 (Fig. 5a). Similar considerations apply to the optimal number of clusters (Fig. 9b), namely 4 clusters is the unique value achieved by $FS^{(P)}$ in case 2, while different values occur in case 1, namely 3, 4, and 13 clusters

Figure 8: Prosumers' type in the each cluster (case 1).

have been achieved depending on the linkage criterion (Fig. 5b). This unique optimal number of cluster is corroborated further by the Elbow method, as highlighted in Fig. 10.
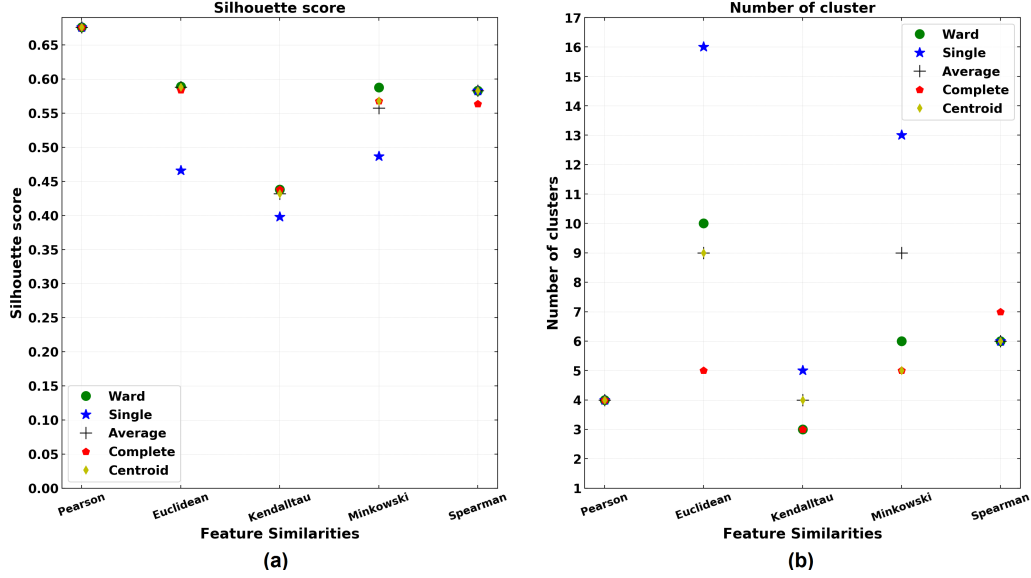

Figure 9: S_s (a) and optimal number of clusters (b) with various distance metric and linkage criteria (case 2).

Clustering map obtained by using $FS^{(P)}$ and d_W are shown in Fig. 11. The type of prosumers in the each cluster is shown in Fig. 12. As it can be seen from Fig. 12 the Cluster 1 is dominated only by Type 1 prosumers with $P_c^{(peak)} < 4$, whereas it includes two Type 3 prosumers in case 1 (Fig. 7). Similarly, Cluster 2 is dominated by Type 1 and some Type 2 prosumers with $4 < P_c^{(peak)} \leq 12$, whereas it includes also all Type 4 prosumers in case 1 (Fig. 8); the latter are now grouped on their own in Cluster 4, while Cluster 3 groups all Type 3 and some Type 2 prosumers with $P_c^{(peak)} < 4$.

Table 4 summarizes the main features of each cluster achieved in case 2. Particularly, the Type 1 prosumers belonging to Cluster 2 are low consumers of water heater and high consumers of HVAC. However, reduced water heater consumption is more than compensated by increased HVAC consumption, which makes these prosumers belonging to Cluster 2 instead of to Cluster 1.

Fig. 13 shows the dSS map of two prosumers that are selected randomly from each cluster, while Fig. 14 shows the average dSS map of each cluster that has been achieved by averaging the dSS maps of all their own prosumers.

21

These figures confirm the validity of the proposed clustering approach as randomly-selected prosumers show dSS maps very similar to the average dSS map of their own cluster, while differing from those of the other clusters.

For further assessing the effectiveness of the clustering results, the cophe-
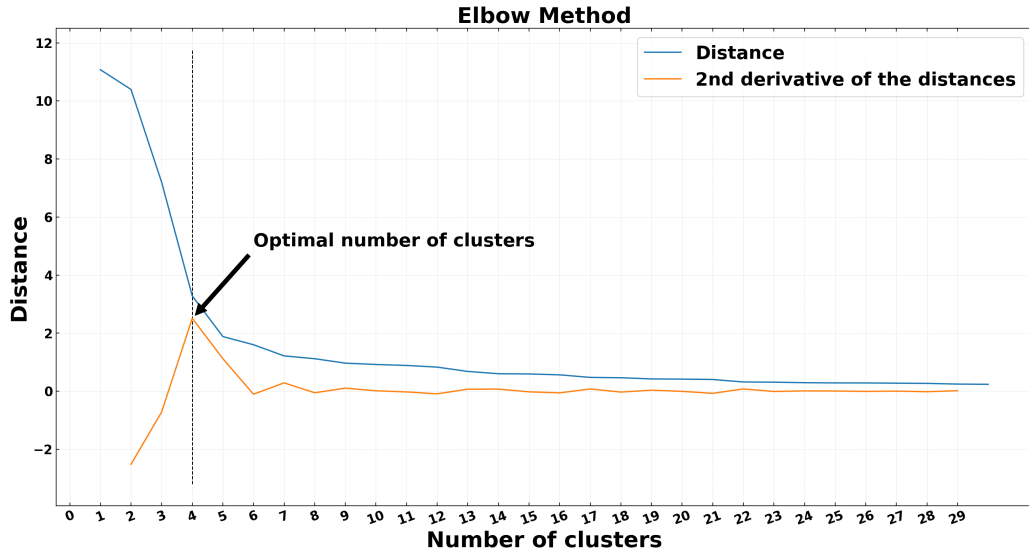

Figure 10: Elbow method outcomes in terms of optimal number of clusters (case 2).
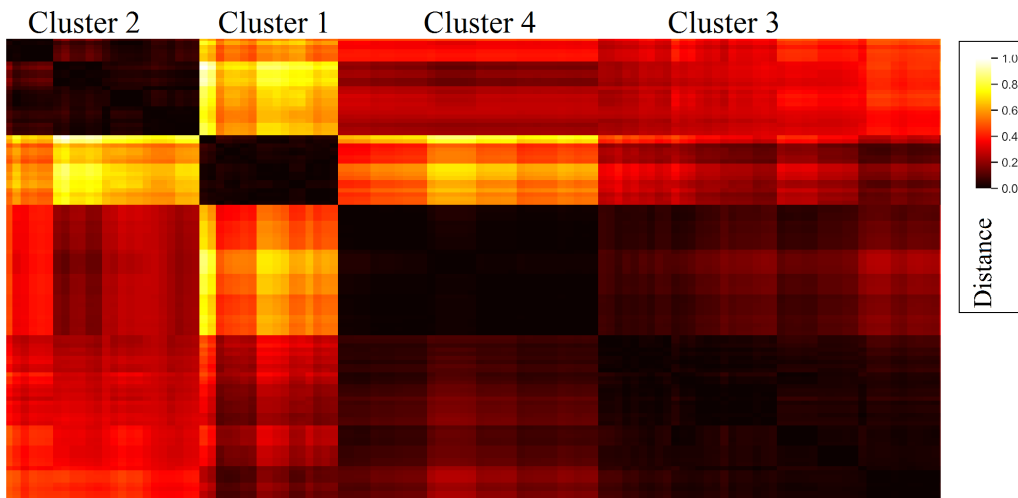

Figure 11: Cluster map (case 2).

Table 4: Main features of clusters achieved in case 2

	Cluster 1	Cluster 2		Cluster 4	Cluster 3	
$P_c^{(peak)}$ (kW)	≤ 4	> 4 & ≤ 12		≤ 4	≤ 4	
I_{size}	0	1		0	0	
Prosumer Type	1	1	2	4	3	2
Water heater	Medium or High	Low	-	-	Medium or High	-
HVAC	Medium or Low	High		-	-	Medium or Low

netic coefficient (ρ) has been computed [61], which measures the quality of clustering, and which is defined by the following equation (18):

$$\rho = \frac{\sum_{i<j}(x(i,j) - \bar{x})(z(i,j) - \bar{z})}{\sqrt{\left[\sum_{i<j}(x(i,j) - \bar{x})^2\right]\left[\sum_{i<j}(z(i,j) - \bar{z})^2\right]}} \tag{18}$$

where $x(i,j)$ is the Euclidean distance between observations i and j, $z(i,j)$ is the dendrogrammatic distance that shows the hierarchical relationship between objects i and j, while $\bar{x}$ and $\bar{z}$ are the average values of $x(i,j)$ and $z(i,j)$, respectively [61]. As a result, ρ ranges from 0 to 1, in which 1 means
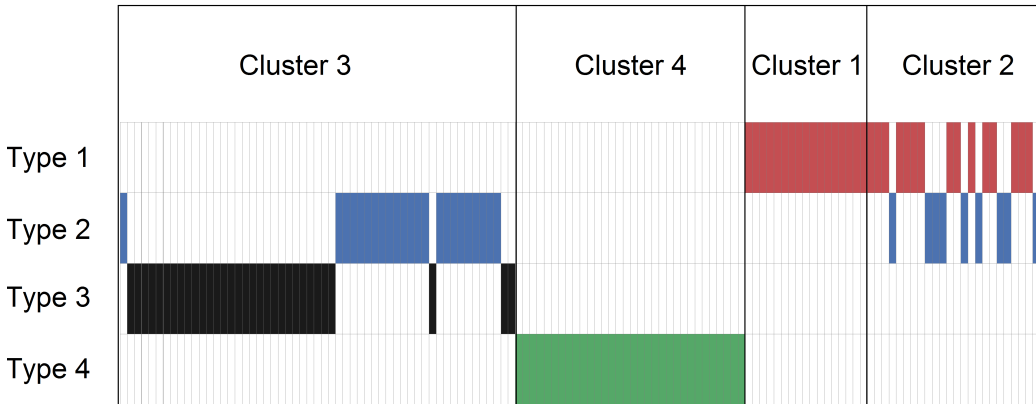

Figure 12: Prosumers' type in the each cluster (case 2).

the highest clustering quality. By applying this criterion to both case 1 and case 2, it has been found that ρ_1 and ρ_2 are 0.836 and 0.878, respectively,
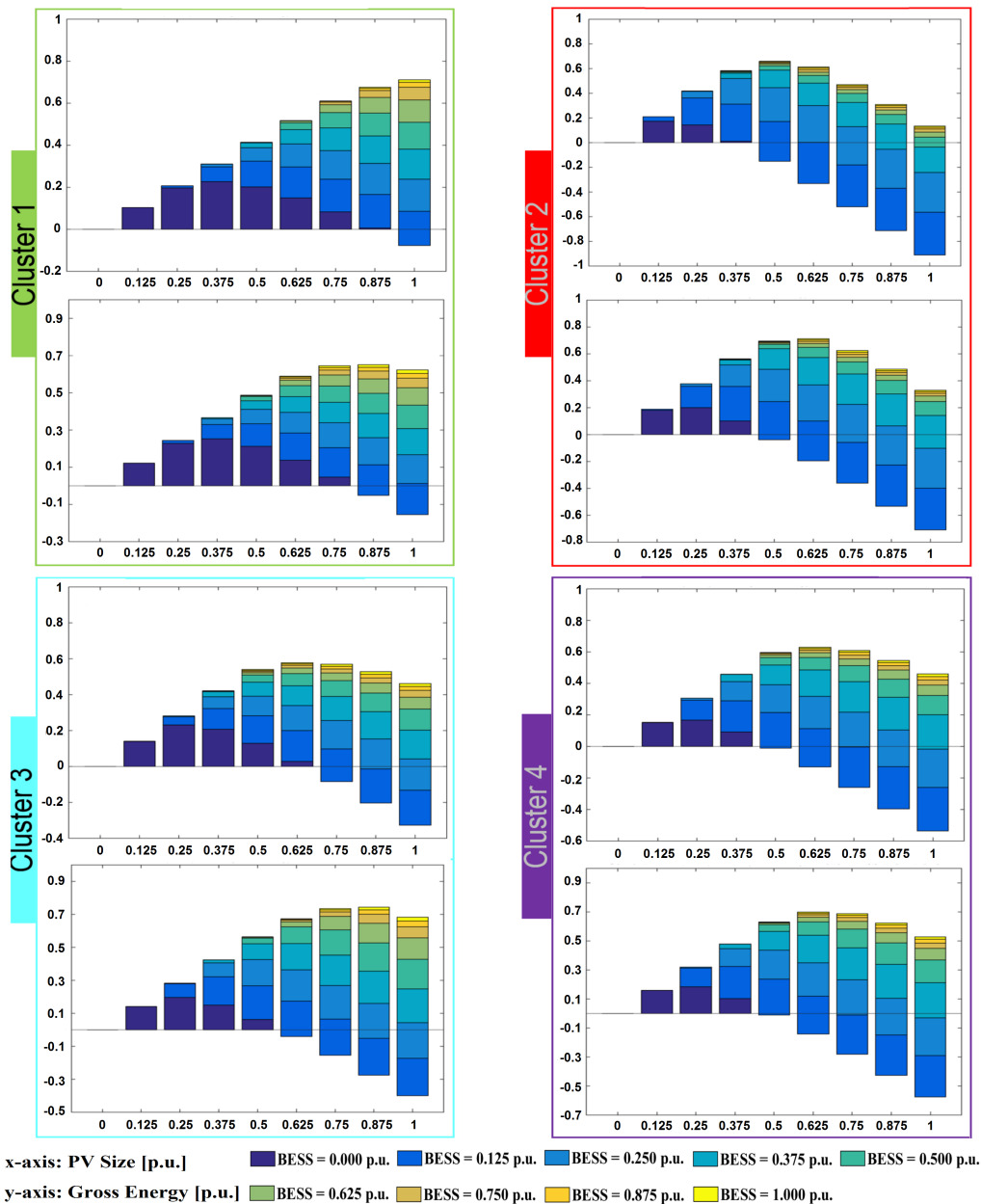
Figure 13: Two randomly-selected dSS maps of each cluster (case 2).

thus proving the enhanced clustering results achieved in case 2.

5.3. GUI module

Based on the clustering results, and in particular on Table 4, the questions that ensure assigning any prosumer to the right cluster are defined as follows:

- the cost of the monthly electricity bill, based on which it is possible to assess the peak consumption value ($P_c^{(peak)}$) in accordance with Italian electricity tariffs;

- the size of the PV system already installed (if any);

- the presence of an electric water heater and/or an HVAC, and their indicative usage (low, medium, high);

Figure 14: Average dSS map of each cluster (case 2).

25

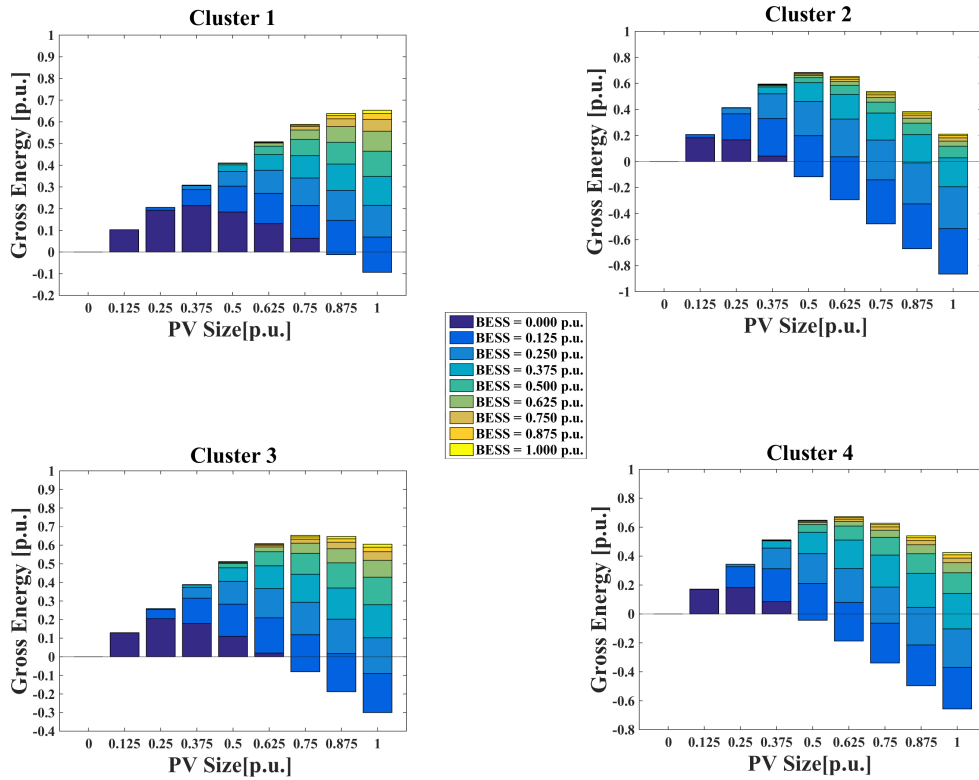| Savings on Gross Energy Exchanged	BESS size [KWh]								
PV size [KW]	0	8	16	24	32	40	48	56	64
1.6	18.25 %	20.75 %	20.75 %	20.75 %	20.75 %	20.75 %	20.75 %	20.75 %	20.75 %
3.2	16.62 %	36.62 %	41.13 %	41.38 %	41.4 %	41.4 %	41.4 %	41.4 %	41.4 %
4.8	4.2 %	33.02 %	51.98 %	57.29 %	58.35 %	58.89 %	59.22 %	59.41 %	59.53 %
6.4	-11.75 %	19.8 %	45.98 %	60.58 %	64.36 %	66.06 %	67.15 %	67.85 %	68.36 %
8	-29.37 %	3.6 %	32.59 %	51.58 %	58.55 %	61.42 %	63.18 %	64.41 %	65.35 %
9.6	-47.92 %	-14.09 %	16.49 %	37.23 %	45.65 %	49.09 %	51.23 %	52.68 %	53.78 %
11.2	-67.04 %	-32.59 %	-1.1 %	20.67 %	29.49 %	33.3 %	35.64 %	37.23 %	38.37 %
12.8	-86.51 %	-51.57 %	-19.46 %	2.83 %	11.8 %	15.67 %	18.15 %	19.87 %	21.0 %

Figure 15: Percentage savings on gross energy exchanged for the reference example

- the geographical location of the prosumer to evaluate the potential of PV production.

It is worth noting that the GUI form also requests the number of occupants, and their highest and lowest age, as shown in Fig. 4. As pointed out in the next Section, these information are not used yet, but they will be likely used in future extensions of the tool to further refine clustering results based on the new data collected through the online platform.

As already disclosed in Section 3.3, the GUI module assigns any prosumer to his/her most suitable cluster almost instantaneously, by providing four tables, in which rows and columns represent different PV and BESS size, respectively:

- the per-unit differential self-sufficiency map of the cluster to which the prosumer belong (dSS, Fig. 15)

- the amount of saved CO_2 emissions (Fig. 16);

- the number of trees that would have absorbed the saved CO_2 emissions (Fig. 17);

- installation, inverter, battery and maintenance costs of the PV-BESS system (Fig. 18).

In particular, the results shown from Fig. 15 to Fig. 18 refer to a potential prosumer located in Cagliari (Italy), with a monthly electric bill of 70

CO2 [Kg]	BESS size [KWh]								
PV size [KW]	0	8	16	24	32	40	48	56	64
1.6	955.36	926.6	921.82	918.61	916.7	915.41	914.54	914.16	913.55
3.2	1910.73	1855.29	1841.24	1838.76	1837.18	1835.43	1834.96	1834.41	1833.83
4.8	2866.09	2798.17	2758.43	2746.63	2743.76	2742.27	2741.25	2740.42	2739.88
6.4	3821.46	3749.06	3698.58	3669.58	3661.52	3657.97	3655.87	3654.59	3653.43
8	4776.82	4701.67	4646.8	4610.53	4595.63	4589.73	4586.08	4583.86	4582.23
9.6	5732.18	5655.06	5597.69	5558.06	5540.43	5533.08	5528.62	5525.99	5523.92
11.2	6687.55	6609.18	6549.87	6508.68	6489.7	6481.46	6476.77	6473.6	6471.21
12.8	7642.91	7563.25	7503.28	7460.9	7441.22	7432.39	7427.32	7423.62	7421.09

Figure 16: Savings on CO_2 kg for the reference example

Trees [#]	BESS size [KWh]								
PV size [KW]	0	8	16	24	32	40	48	56	64
1.6	38	37	36	36	36	36	36	36	36
3.2	76	74	73	73	73	73	73	73	73
4.8	114	111	110	109	109	109	109	109	109
6.4	152	149	147	146	146	146	146	146	146
8	191	188	185	184	183	183	183	183	183
9.6	229	226	223	222	221	221	221	221	220
11.2	267	264	261	260	259	259	259	258	258
12.8	305	302	300	298	297	297	297	296	296

Figure 17: Number of trees that would absorb the saved CO_2 amount for the reference example

Cost [€]	BESS size [KWh]								
PV size [KW]	0	8	16	24	32	40	48	56	64
1.6	2400.00	6400.00	10400.00	14400.00	18400.00	22400.00	26400.00	30400.00	34400.00
3.2	4800.00	8800.00	12800.00	16800.00	20800.00	24800.00	28800.00	32800.00	36800.00
4.8	7200.00	11200.00	15200.00	19200.00	23200.00	27200.00	31200.00	35200.00	39200.00
6.4	9600.00	13600.00	17600.00	21600.00	25600.00	29600.00	33600.00	37600.00	41600.00
8	12000.00	16000.00	20000.00	24000.00	28000.00	32000.00	36000.00	40000.00	44000.00
9.6	14400.00	18400.00	22400.00	26400.00	30400.00	34400.00	38400.00	42400.00	46400.00
11.2	16800.00	20800.00	24800.00	28800.00	32800.00	36800.00	40800.00	44800.00	48800.00
12.8	19200.00	23200.00	27200.00	31200.00	35200.00	39200.00	43200.00	47200.00	51200.00

Figure 18: Capital costs of the PV-BESS system for the reference example

€, no PV previously installed, low usage of electric water heater and high usage of HVAC. Note that this is just a reference example of the results that a prospective GUI user would get by answering the questions of the online form shown in Fig. 4. It is possible to notice that the self-sufficiency increase can be as high as 68% for a 6.4 kW PV with a 64 kWh BESS, which corresponds to more than 3600 kg of saved CO_2 and almost 150 trees. In this regard, the amount of saved CO_2 emissions is computed based on the Italian emission factors of electricity production, which were 296.5 gCO_2/kWh in 2018 [62]. Similarly, the number of trees is computed by using the average annual CO_2 offsetting rate, which is about 25 kg CO_2/tree [63]. However, such a high-performance system would cost more than 40 k€, which would be unfeasible/inconvenient for most of the users. A good trade-off between increased self-sufficiency and cost is a 4.8 kW PV with a 16 kWh BESS, which ensures more than 50% increase in self-sufficiency with a reasonable cost of 15.2 k€ and more than 2700 kg of CO_2 and 110 trees saved.

In conclusion, it is worth mentioning that negative values appearing in Fig. 15 correspond to PV oversizing and a relatively small BESS; this means that these configurations should be avoided because they increase the energy exchanged with the grid and installation costs, without bringing any significant benefit to the prosumer. However, any prosumer is of course free to select the configuration most suitable to their needs/expectations based on dSS and on the other maps, revealing the flexibility of the proposed online energy management tool. For the reader to have an idea of the benefit determined by the use of the GUI, Table 5 reports the ranges of results that it

Table 5: Overall simulation outcomes for different clusters

	Cluster	Energy savings [%]	PV size [kW]	BESS size [kWh]	CO2 [kg]	tree #	Cost [k€]
Highest saving percentage	1	65.3	4.8	32	2181	87	23.200
	2	67.3	4.0	32	1821	72	22.000
	3	65.3	6.4	32	2923	116	25.600
	4	68.4	6.4	64	3653	146	41.600
Saving percentage closer to 50%	1	50.2	3.2	16	1461	58	12.800
	2	50.5	2.4	12	1097	43	09.600
	3	50.2	4.0	20	1833	73	16.000
	4	52.0	4.8	16	2758	110	15.200

is possible to get considering all the clusters, for the highest possible saving percentage, and for the saving percentage that is closer to 50%. As it can be noticed, even in the best-case scenario savings higher than 68% are not expected. Indeed, in order to have higher savings, it is not sufficient to install a new PV-BESS system, but prosumers should also change their habits so as to fit their consumption profile to the PV-BESS system production profile as much as possible.

6. Conclusion

This paper has presented an online energy management tool aimed at supporting a prosumer in choosing the size of a Photovoltaic Power Plant-Battery Energy Storage System (PV-BESS) most suited to his/her needs and/or expectations. The proposed tool is based on an off-line PV-BESS sizing module, which computes the self-efficiency maps of potential prosumers in accordance with their electricity consumption and expected PV production profiles, and an off-line Clustering module, which processes the self-efficiency maps and some additional features to find out the optimal number and distinctive features of prosumers' clusters. The proposed clustering approach has been validated by numerical simulations, which regard 128 prosumers from different geographical locations and with different consumption habits. The results highlight that each prosumer can be clustered successfully in accordance with few basic and easily recoverable information, such as the kind of appliances available/used and the maximum contractual power. As

a result, based on such off-line computations, the proposed online tool enables any prosumer to achieve his/her most suited self-efficiency map and related information almost instantaneously, by filling in an online form available through a graphical-user-interface module. The proposed tool is free and already available at [59]. Future works will further refine and extend the tool, by including additional information for the clustering procedure, as well as asking other questions that will enable the refinement of results based on data collected from the users. Furthermore, other types of building, such as tertiary buildings (e.g. schools, shops), and prosumer (e.g. community microgrids) will be considered to broaden the catchment area of the proposed tool. The PV-BESS sizing module will be also extended, by considering other renewable energy sources, such as micro-wind turbines.

Acknowledgement

References

[1] J. Rogelj, M. Den Elzen, N. Höhne, T. Fransen, H. Fekete, H. Winkler, R. Schaeffer, F. Sha, K. Riahi, M. Meinshausen, Paris Agreement climate proposals need a boost to keep warming well below 2 °c (jun 2016). doi:10.1038/nature18307.

[2] Global Alliance for Buildings and Construction 2020 GLOBAL STATUS REPORT FOR BUILDINGS AND CONSTRUCTION Towards a zero-emissions, efficient and resilient buildings and construction sector EXECUTIVE SUMMARY 2020 GLOBAL STATUS REPORT FOR BUILDINGS AND CONSTRUCTION Towards a zero-emissions, efficient and resilient buildings and construction sector, Tech. rep. (2020). URL http://www.un.org/Depts/

[3] R. Luthander, J. Widén, D. Nilsson, J. Palm, Photovoltaic self-consumption in buildings: A review (mar 2015). doi:10.1016/j.apenergy.2014.12.028.
URL http://dx.doi.org/10.1016/j.apenergy.2014.12.028

[4] D. Cheng, B. A. Mather, R. Seguin, J. Hambrick, R. P. Broadwater, Photovoltaic (pv) impact assessment for very high penetration levels, IEEE Journal of Photovoltaics 6 (1) (2016) 295–300. `doi:10.1109/JPHOTOV.2015.2481605`.

[5] Y. Wang, V. Silva, M. Lopez-Botet-Zulueta, Impact of high penetration of variable renewable generation on frequency dynamics in the continental Europe interconnected system`doi:10.1049/iet-rpg.2015.0141`.
URL `www.ietdl.org`

[6] H. Mortazavi, H. Mehrjerdi, M. Saad, S. Lefebvre, D. Asber, L. Lenoir, A monitoring technique for reversed power flow detection with high pv penetration level, IEEE Transactions on Smart Grid 6 (5) (2015) 2221–2232. `doi:10.1109/TSG.2015.2397887`.

[7] B. Drysdale, J. Wu, N. Jenkins, Flexible demand in the GB domestic electricity sector in 2030, Applied Energy 139 (2015) 281–290. `doi:10.1016/j.apenergy.2014.11.013`.
URL `http://creativecommons.org/licenses/by/3.0/`

[8] G. de Oliveira e Silva, P. Hendrick, Lead–acid batteries coupled with photovoltaics for increased electricity self-sufficiency in households, Applied Energy 178 (2016) 856–867. `doi:10.1016/j.apenergy.2016.06.003`.

[9] T. Boström, B. Babar, J. B. Hansen, C. Good, The pure PV-EV energy system – A conceptual study of a nationwide energy system based solely on photovoltaics and electric vehicles, Smart Energy 1 (2021) 100001. `doi:10.1016/j.segy.2021.100001`.

[10] U. G. Mulleriyawage, W. X. Shen, Impact of demand side management on optimal sizing of residential battery energy storage system, Renewable Energy 172 (2021) 1250–1266. `doi:10.1016/j.renene.2021.03.122`.
URL `https://doi.org/10.1016/j.renene.2021.03.122`

[11] B. Dupont, K. Dietrich, C. De Jonghe, A. Ramos, R. Belmans, Impact of residential demand response on power system operation: A Belgian case study, Applied Energy 122 (2014) 1–10. `doi:10.1016/j.apenergy.`

2014.02.022.
URL http://dx.doi.org/10.1016/j.apenergy.2014.02.022

[12] F. M. Vieira, P. S. Moura, A. T. de Almeida, Energy storage system for self-consumption of photovoltaic energy in residential zero energy buildings, Renewable Energy 103 (2017) 308–320. doi:10.1016/j.renene.2016.11.048.

[13] W. Marańda, Analysis of self-consumption of energy from grid-connected photovoltaic system for various load scenarios with short-term buffering, SN Applied Sciences 1 (5) (2019) 406. doi:10.1007/s42452-019-0432-5.
URL https://doi.org/10.1007/s42452-019-0432-5

[14] R. Alvaro-Hermana, J. Merino, J. Fraile-Ardanuy, S. Castaño-Solis, D. Jiménez, Shared self-consumption economic analysis for a residential energy community, in: 2019 International Conference on Smart Energy Systems and Technologies (SEST), 2019, pp. 1–6. doi:10.1109/SEST.2019.8849101.

[15] L. Zhou, Y. Zhang, X. Lin, C. Li, Z. Cai, P. Yang, Optimal sizing of PV and bess for a smart household considering different price mechanisms, IEEE Access 6 (2018) 41050–41059. doi:10.1109/ACCESS.2018.2845900.

[16] D. Masa-Bote, M. Castillo-Cagigal, E. Matallanas, E. Caamaño-Martín, A. Gutiérrez, F. Monasterio-Huelín, J. Jiménez-Leube, Improving photovoltaics grid integration through short time forecasting and self-consumption, Applied Energy 125 (2014) 103–113. doi:https://doi.org/10.1016/j.apenergy.2014.03.045.
URL https://www.sciencedirect.com/science/article/pii/S0306261914002761

[17] D. Neves, I. Scott, C. A. Silva, Peer-to-peer energy trading potential: An assessment for the residential sector under different technology and tariff availabilities, Energy 205 (2020) 118023. doi:https://doi.org/10.1016/j.energy.2020.118023.
URL https://www.sciencedirect.com/science/article/pii/S0360544220311300

[18] F. Lamberti, V. Calderaro, V. Galdi, G. Graditi, Massive data analysis to assess PV/ESS integration in residential unbalanced LV networks to support voltage profiles, Electric Power Systems Research 143 (2017) 206–214. doi:10.1016/j.epsr.2016.10.037.
URL http://dx.doi.org/10.1016/j.epsr.2016.10.037

[19] I. Ranaweera, O.-M. Midtgård, Optimization of operational cost for a grid-supporting pv system with battery storage, Renewable Energy 88 (2016) 262–272. doi:https://doi.org/10.1016/j.renene.2015.11.044.
URL https://www.sciencedirect.com/science/article/pii/S0960148115304651

[20] M. van der Kam, W. van Sark, Smart charging of electric vehicles with photovoltaic power and vehicle-to-grid technology in a microgrid; a case study, Applied Energy 152 (2015) 20–30. doi:https://doi.org/10.1016/j.apenergy.2015.04.092.
URL https://www.sciencedirect.com/science/article/pii/S0306261915005553

[21] N. Liu, Q. Chen, X. Lu, J. Liu, J. Zhang, A charging strategy for pv-based battery switch stations considering service availability and self-consumption of pv energy, IEEE Transactions on Industrial Electronics 62 (8) (2015) 4878–4889. doi:10.1109/TIE.2015.2404316.

[22] R. Luthander, J. Widén, J. Munkhammar, D. Lingfors, Self-consumption enhancement and peak shaving of residential photovoltaics using storage and curtailment, Energy 112 (2016) 221–231. doi:10.1016/j.energy.2016.06.039.

[23] A. C. Duman, H. S. Erden, Ö. Gönül, Ö. Güler, A home energy management system with an integrated smart thermostat for demand response in smart grids, Sustainable Cities and Society 65 (2021) 102639. doi:10.1016/j.scs.2020.102639.

[24] A. Bartolini, G. Comodi, D. Salvi, P. A. Østergaard, Renewables self-consumption potential in districts with high penetration of electric vehicles, Energy 213 (2020) 118653. doi:https://doi.org/10.1016/j.energy.2020.118653.

URL https://www.sciencedirect.com/science/article/pii/S0360544220317618

[25] Z. Wang, C. Gu, F. Li, P. Bale, H. Sun, Active demand response using shared energy storage for household energy management, IEEE Transactions on Smart Grid 4 (4) (2013) 1888–1897. doi:10.1109/TSG.2013.2258046.

[26] A. Jossen, D. Magnor, K. Büdenbender, M. Braun, Photovoltaic self-consumption in germany using lithium-ion storage to increase self-consumed photovoltaic energy, 2009.

[27] J. Hernández, F. Sanchez-Sutil, F. Muñoz-Rodríguez, C. Baier, Optimal sizing and management strategy for pv household-prosumers with self-consumption/sufficiency enhancement and provision of frequency containment reserve, Applied Energy 277 (2020) 115529. doi:https://doi.org/10.1016/j.apenergy.2020.115529.
URL https://www.sciencedirect.com/science/article/pii/S0306261920310412

[28] R. Hemmati, H. Saboori, Stochastic optimal battery storage sizing and scheduling in home energy management systems equipped with solar photovoltaic panels, Energy and Buildings 152 (2017) 290–300. doi:10.1016/j.enbuild.2017.07.043.

[29] K. Lebedeva, A. Krumins, A. Tamane, E. Dzelzitis, Analysis of latvian households' potential participation in the energy market as prosumers, Clean Technologies 3 (2) (2021) 437–449. doi:10.3390/cleantechnol3020025.
URL https://www.mdpi.com/2571-8797/3/2/25

[30] A. Maleki, M. A. Nazari, F. Pourfayaz, Harmony search optimization for optimum sizing of hybrid solar schemes based on battery storage unit, Energy Reports 6 (2020) 102–111. doi:10.1016/j.egyr.2020.03.014.

[31] J. Cervantes, F. Choobineh, Optimal sizing of a nonutility-scale solar power system and its battery storage, Applied Energy 216 (2018) 105–115. doi:10.1016/j.apenergy.2018.02.013.

[32] M. Mehrtash, F. Capitanescu, P. K. Heiselberg, T. Gibon, A. Bertrand, An enhanced optimal pv and battery sizing model for zero energy buildings considering environmental impacts, IEEE Transactions on Industry Applications 56 (6) (2020) 6846–6856. `doi:10.1109/TIA.2020.3022742`.

[33] Y. Yoo, G. Jang, S. Jung, A Study on Sizing of Substation for PV with Optimized Operation of BESS, IEEE Access 8 (2020) 214577–214585. `doi:10.1109/ACCESS.2020.3040646`.

[34] A. S. da Silveira, A. da Rosa Abaide, L. N. F. da Silva, F. C. Lucchese, B. K. Hammerschmitt, Optimal sizing of a pv-bess grid-connected microgrid in southern region of brazil, in: 2020 6th International Conference on Electric Power and Energy Conversion Systems (EPECS), 2020, pp. 52–57. `doi:10.1109/EPECS48981.2020.9304963`.

[35] Y. Khawaja, D. Giaouris, H. Patsios, M. Dahidah, Optimal cost-based model for sizing grid-connected PV and battery energy system, 2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies, AEECT 2017 2018-January (2017) 1–6. `doi:10.1109/AEECT.2017.8257779`.

[36] T. Weckesser, D. F. Dominković, E. M. Blomgren, A. Schledorn, H. Madsen, Renewable Energy Communities: Optimal sizing and distribution grid impact of photo-voltaics and battery storage, Applied Energy 301 (April) (2021) 117408. `doi:10.1016/j.apenergy.2021.117408`. URL `https://doi.org/10.1016/j.apenergy.2021.117408`

[37] V. Sharma, M. H. Haque, S. M. Aziz, Energy cost minimization for net zero energy homes through optimal sizing of battery storage system, Renewable Energy 141 (2019) 278–286. `doi:10.1016/J.RENENE.2019.03.144`.

[38] X. Wu, X. Hu, X. Yin, C. Zhang, S. Qian, Optimal battery sizing of smart home via convex programming, Energy 140 (2017) 444–453. `doi:10.1016/j.energy.2017.08.097`. URL `https://doi.org/10.1016/j.energy.2017.08.097`

[39] A. Colmenar-Santos, M. Monteagudo-Mencucci, E. Rosales-Asensio, M. de Simón-Martín, C. Pérez-Molina, Optimized design method for

storage systems in photovoltaic plants with delivery limitation, Solar Energy 180 (January) (2019) 468–488. doi:10.1016/j.solener.2019.01.046.
URL https://doi.org/10.1016/j.solener.2019.01.046

[40] B. D. Olaszi, J. Ladanyi, Comparison of different discharge strategies of grid-connected residential PV systems with energy storage in perspective of optimal battery energy storage system sizing, Renewable and Sustainable Energy Reviews 75 (November 2016) (2017) 710–718. doi:10.1016/j.rser.2016.11.046.
URL http://dx.doi.org/10.1016/j.rser.2016.11.046

[41] F. Rukavina, M. Vasak, Optimal parameterization of a PV and a battery system add-on for a consumer, 2020 IEEE 11th International Symposium on Power Electronics for Distributed Generation Systems, PEDG 2020 (2020) 621–626doi:10.1109/PEDG48541.2020.9244462.

[42] E. Chatterji, M. D. Bazilian, Battery storage for resilient homes, IEEE Access 8 (2020) 184497–184511. doi:10.1109/ACCESS.2020.3029989.

[43] J. T. Liao, Y. S. Chuang, H. T. Yang, M. S. Tsai, BESS-Sizing Optimization for Solar PV System Integration in Distribution Grid, IFAC-PapersOnLine 51 (28) (2018) 85–90. doi:10.1016/j.ifacol.2018.11.682.
URL https://doi.org/10.1016/j.ifacol.2018.11.682

[44] D. Kolokotsa, D. Rovas, E. Kosmatopoulos, K. Kalaitzakis, A roadmap towards intelligent net zero- and positive-energy buildings (dec 2011). doi:10.1016/j.solener.2010.09.001.

[45] F. Casu, S. Korjani, V. Pilloni, A. Serpi, A. Damiano, D. Giusto, Designing nearly zero energy buildings: a recommendation tool for optimal sizing of renewable energy source systems, in: 2020 2nd IEEE International Conference on Industrial Electronics for Sustainable Energy Systems (IESES), Vol. 1, 2020, pp. 305–310. doi:10.1109/IESES45645.2020.9210701.

[46] Open ei dataset.
URL http://en.openei.org/datasets/files/961/pub/

[47] S. Korjani, A. Serpi, A. Damiano, A genetic algorithm approach for sizing integrated pv-bess systems for prosumers, in: 2020 2nd IEEE International Conference on Industrial Electronics for Sustainable Energy Systems (IESES), Vol. 1, 2020, pp. 151–156. `doi:10.1109/IESES45645.2020.9210700`.

[48] Solar Power Data for Integration Studies — Grid Modernization — NREL.
URL `https://www.nrel.gov/grid/solar-power-data.html`

[49] S. Korjani, A. Facchini, M. Mureddu, A. Rubino, A. Damiano, Battery management for energy communities—Economic evaluation of an artificial intelligence-led system, Journal of Cleaner Production 314 (2021) 128017. `doi:10.1016/j.jclepro.2021.128017`.

[50] S. Korjani, M. Mureddu, A. Facchini, A. Damiano, Aging cost optimization for planning and management of energy storage systems, Energies 10 (11) (2017). `doi:10.3390/en10111916`.

[51] J. Hauke, Tomasz Kossowski, Bogucki Wydawnictwo Naukowe, QUAESTIONES GEOGRAPHICAE 30 (2) (2011) 87–93. `doi:10.2478/v10117-011-0021-1`.

[52] M. G. Kendall, A new measure of rank correlation, Biometrika 30 (1/2) (1938) 81–93.
URL `http://www.jstor.org/stable/2332226`

[53] M. Ichino, H. Yaguchi, Generalized minkowski metrics for mixed feature-type data analysis, IEEE Transactions on Systems, Man, and Cybernetics 24 (4) (1994) 698–708. `doi:10.1109/21.286391`.

[54] S.-H. Cha, Comprehensive survey on distance/similarity measures between probability density functions, City 1 (2) (2007) 1.

[55] S. C. Johnson, Hierarchical clustering schemes, Psychometrika 32 (3) (1967) 241–254. `doi:10.1007/BF02289588`.
URL `https://link.springer.com/article/10.1007/BF02289588`

[56] S. Miyamoto, R. Abe, Y. Endo, J.-i. Takeshita, Ward method of hierarchical clustering for non-euclidean similarity measures, in: 2015 7th

International Conference of Soft Computing and Pattern Recognition (SoCPaR), 2015, pp. 60–63. `doi:10.1109/SOCPAR.2015.7492784`.

[57] P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, Journal of Computational and Applied Mathematics 20 (C) (1987) 53–65. `doi:10.1016/0377-0427(87) 90125-7`.

[58] C. Shi, B. Wei, S. Wei, W. Wang, H. Liu, J. Liu, A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm, Eurasip Journal on Wireless Communications and Networking 2021 (1) (2021) 31. `doi:10.1186/s13638-021-01910-w`.

[59] TDM Simulator.
URL `http://tdm-project2020.herokuapp.com/`

[60] Commercial and Residential Hourly Load Profiles for all TMY3 Locations in the United States - Building Characteristics for Residential Hourly Load Data - OpenEI Datasets.
URL `https://openei.org/datasets/dataset/commercial-and-residential-hourly-load-profiles-for-all-tmy3-locations-in-the-united-states/resource/cd6704ba-3f53-4632-8d08-c9597842fde3`

[61] S. Saraçli, N. Doğan, I. Doğan, Comparison of hierarchical cluster analysis methods by cophenetic correlation, Journal of Inequalities and Applications 2013 (1) (2013) 203. `doi:10.1186/1029-242X-2013-203`.
URL `https://journalofinequalitiesandapplications.springeropen.com/articles/10.1186/1029-242X-2013-203`

[62] ISPRA, Fattori di emissione atmosferica di gas a effetto serra nel settore elettrico nazionale e nei principali paesi europei, `https://www.isprambiente.gov.it/files2020/pubblicazioni/rapporti/Rapporto317_2020.pdf`, last accessed: 01/07/2021 (2020).

[63] ENCON, Calculation of co2 offsetting by trees, `https://www.encon.be/en/calculation-co2-offsetting-trees`, last accessed: 01/07/2021.