ANFIS Synthesis by Clustering for Microgrids EMS Design

Stefano Leonori, Alessio Martino, Antonello Rizzi and Fabio Massimo Frattale Mascioli
Department of Information Engineering, Electronics and Telecommunications,
University of Rome "La Sapienza", Via Eudossiana 18, 00184 Rome, Italy
{*stefano.leonori, alessio.martino, antonello.rizzi, fabiomassimo.frattalemascioli*}*@uniroma1.it*

Keywords: Smart Grids, Microgrids, Energy Management System, ANFIS, Data Clustering, Decision Making System.

Abstract: Microgrids (MGs) play a crucial role for the development of Smart Grids. They are conceived to intelligently integrate the generation from Distributed Energy Resources, to improve Demand Response (DR) services, to reduce pollutant emissions and curtail power losses, assuring the continuity of services to the loads as well. In this work it is proposed a novel synthesis procedure for modelling an Adaptive Neuro-Fuzzy Inference System (ANFIS) featured by multivariate Gaussian Membership Functions (MFs) and first order Takagi-Sugeno rules. The Fuzzy Rule Base is the core inference engine of an Energy Management System (EMS) for a grid-connected MG equipped with a photovoltaic power plant, an aggregated load and an Energy Storage System (ESS). The EMS is designed to operate in real time by defining the ESS energy flow in order to maximize the revenues generated by the energy trade with the distribution grid. The ANFIS EMS is synthesized through a data driven approach that relies on a clustering algorithm which defines the MFs and the rule consequent hyperplanes. Moreover, three clustering algorithms are investigated. Results show that the adoption of *k*-medoids based on Mahalanobis (dis)similarity measure is more efficient with respect to the *k*-means, although affected by some variety in clusters composition.

1 INTRODUCTION

A Microgrid (MG) is an electric grid able to intelligently manage and control local electric power systems affected by stochastic and intermittent behaviours, such as electric generation from renewable energy sources, electric vehicles charging and deferrable and shiftable loads. MGs are the best candidates for the transition to Smart Grids, since they allow a bottom-up approach for building and developing reliable and smart distribution systems, relying on the concept of territorial granulation. Each MG is provided of a suitable Energy Management System (EMS) to intelligently manage local power flows inside MG and with the main grid (Patterson, 2012; Dragicevic et al., 2014). The MG infrastructure relies on power converters connecting power systems and electric loads to the main bus in order to locally route and manage the MG power flows and the power exchanges with the connected grid. To this end, MGs must be equipped with a communication infrastructure able to monitor and supervise the state of all the MG components.
Usually, MGs are supported by Energy Storage Systems (ESSs) able to guarantee both the quality of service and the electric stability, ensuring some energetic autonomy to the system when it is disconnected to the grid (*i.e.* islanded mode). The implementation of a suitable Demand Side Management (DSM) EMS allows to apply Demand Response (DR) services to the costumer, which is more appropriate to refer to as *prosumer* whether equipped with a power generation system.
In (Deng et al., 2015) are well summarized all the DR main services (*i.e.* valley filling, load shifting, peak shaving operations). These services, together with Vehicle-2-Grid (V2G) operations and the intelligent use of the ESS, allow to reduce the stress caused by the MG to the connected distribution grid in order to get incentives, avoid penalties, reduce both the consumptions and the operational costs, which strictly depend on the energy price policies adopted by the distribution grid. Concerning this topic, in (Kirschen, 2003; Amer et al., 2014) are discussed the development of new energy policies which will involve the costumer to assume an active role in the energy market by means of the application of DR services.
In this work, a procedure based on computational intelligence techniques for the data driven synthesis of an EMS is proposed. The EMS must define in real time the energy flow exchanged with the grid in order to maximize the MG profit by considering a Time

Leonori, S., Martino, A., Rizzi, A. and Mascioli, F.
ANFIS Synthesis by Clustering for Microgrids EMS Design.
In *Proceedings of the 9th International Joint Conference on Computational Intelligence (IJCCI 2017)*, pages 328-337
ISBN: 978-989-758-274-5

Of Use (TOU) energy policy. The MG EMS is based on an Adaptive Neuro-Fuzzy Inference System (ANFIS) that is efficiently modelled by a clustering algorithm. In addiction, different clustering algorithms are investigated and compared for the ANFIS modelling, which mainly differ in the (dis)similarity measure adopted. Moreover, it is also investigated the efficacy of considering the EMS output space by the clustering algorithms (*i.e.* joint input-output space) relying on a benchmark solution found through a Mixed-Integer Linear Programming (MILP) problem formulation.
The remainder of the paper is organized as follows. In Sec. 2 is introduced the MG problem formulation. The EMS design and the modelling procedure are described in Sec. 3, where both the Objective Function (OF) formulation and the set of clustering algorithms will be introduced. In Sec. 5 are reported the simulations settings, whereas the achieved results are in Sec. 6, followed by the conclusions discussed in Sec. 7.

2 MG PROBLEM FORMULATION

In this paper it is considered a prosumer grid-connected MG equipped with a DSM EMS. It is in charge of efficiently manage the MG components represented as aggregated systems grouped in renewable sources power generators, electric loads and ESSs. Their energy flows are managed in real time by an EMS that acts as decision making system. It must efficiently redistribute the prosumer energy balance (*i.e* the overall energy produced net of the energy demand at the given time slot) between the grid and the ESS by maximizing the profit given by the energy trade with the main grid.
This work is based on several hypotheses which help defining the correct level of abstraction to properly focus the problem under analysis as made in previous studies (Leonori et al., 2016a; Leonori et al., 2016b). The power value of the MG components has been considered constant within each 15 minutes time slot. Low level operations such as voltage and reactive power control are not considered. The power transmission losses within the MG are considered negligible. The on-line control module ensures that the power balance is achieved during the real-time operation. The EMS has a sample time equal to the time slot duration which is considerably greater than the characteristic time of the ESS power control, therefore the ESS inner loop has been neglected. The power converters which connect the MG sub-components to each other, included the one allowing the MG-grid connection, are neglected in terms of power losses and characteristic time of control.
The MG aggregated energy generation, aggregated load request, energy exchanged with the ESS and energy exchanged with the grid during the n^{th} time slot are denoted with E_n^L, E_n^G, E_n^S and E_n^N, respectively. In figure 1 it is represented a schematic diagram of the MG where the power lines are drawn in black and the signal wires in red. In Figure is also represented the Battery Management System (BMS). It monitors the ESS and estimates its State Of Charge (*SoC*) which is used as an input of the EMS.

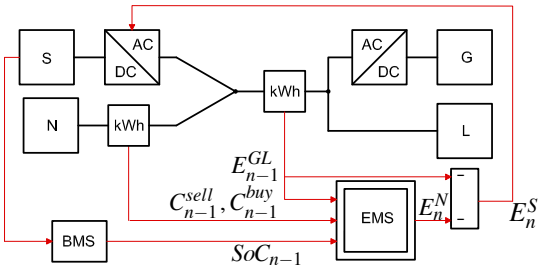

Figure 1: MG architecture. Signal wires in red, power lines in black.

By assuming that the prosumer energy production E_n^G has the priority to meet the prosumer energy demand E_n^L, the prosumer energy balance E_n^{GL} can be defined as

$$E_n^{GL} = E_n^G + E_n^L, \qquad n = 1,2,... \tag{1}$$

in each time slot.
Moreover, in this work it is assumed that the prosumer energy balance E_n^{GL} is a known quantity read in real time by an electric meter. In each time slot, E_n^{GL} must be exchanged with the main grid and the ESS by fulfilling the following energy balance relation

$$E_n^S + E_n^N + E_n^{GL} = 0, \qquad n = 1,2,... \tag{2}$$

The energy E_n^S is assumed positive or negative when the ESS is discharged or recharged, respectively. Similarly, the energy E_n^N is considered positive (negative) when the network is selling (buying) energy to the MG. Considering a TOU price policy, it is possible to formulate the profit P generated by the energy trade with the main grid in a time period composed by N_{slot} time slots as

$$P = \sum_{n=1}^{N_{slot}} P_n \qquad \text{where} \qquad P_n = \begin{cases} E_n^N \cdot C_n^{buy} & \text{if } E_n^N > 0 \\ E_n^N \cdot C_n^{sell} & \text{if } E_n^N \leq 0 \end{cases} \tag{3}$$

where C_n^{buy} and C_n^{sell} define the energy prices in purchase and sale during the n^{th} time slot. According with (Leonori et al., 2017), it is assumed that during the n^{th} time slot the MG cannot exchange with the grid an amount of energy greater than the current

energy balance E_n^{GL}. In other words, in case of over-production (over-demand) (*i.e.* $E^{GL} > 0$ ($E^{GL} < 0$)) the ESS can be only charged (discharged).
In this work the EMS is assumed to be able to efficiently estimate in real time the energy E_n^N and E_n^S to be exchanged during the n^{th} time slot with the connected grid and the ESS, respectively (see figure 1). The EMS is supposed to be fed by the input vector $\mathbf{u}$ constituted by 4 variables, namely, the current energy balance E_{n-1}^{GL}, the current energy prices in sale and purchase, C_{n-1}^{sell} and C_{n-1}^{buy} and the *SoC* value Soc_{n-1}. It should be noted that whereas E_{n-1}^{GL}, C_{n-1}^{sell} and C_{n-1}^{buy} are instantaneous quantities read by proper meters, SoC_{n-1} is a status variable depending on the previous ESS history. Before entering the EMS, the E^{GL}, C^{sell}, C^{buy} inputs, must be normalized in the range $[0,1]$, whilst the *SoC* belongs to $[0,1]$ by its own definition. In the following, the normalized input vector will be referred to as $\bar{\mathbf{u}}$.

3 PROBLEM STATEMENT AND MODELLING APPROACH

The use of computational intelligence techniques and, especially, Fuzzy Logic and Fuzzy Inference Systems (FISs), is often mentioned in literature for solving the EMS real time decision making system design. In such cases, the inferential process (*i.e.* the rule based system) can be realized relying on expert operator(s) with the support of heuristics, such as Genetic Algorithms (GAs). Moreover, Mamdani FIS types based on grid partitioning, are the most commonly used due to their simplicity and effectiveness. For example in (Arcos-Aviles et al., 2016) is proposed a GA-FIS model in order to minimize the power peaks and the fluctuations of the energy exchange with the connected-grid, while keeping the battery *SoC* within certain security limits; in (Leonori et al., 2016b) a rule base system designed by an expert operator has been optimized by a GA in order to maximize the profit generated by the energy trade with the main grid assuming a TOU energy price. In these works, the application of heuristics, specifically GAs, has been motivated by the fact that the synthesis problem has been casted as an unsupervised one.
In this paper it is proposed an EMS synthesis procedure in a supervised fashion. To this end, the EMS model has been synthesized through an ANFIS supported by clustering. The proposed paradigm is well introduced in (Panella et al., 2001).
The supervised problem formulation needs to rely on a ground-truth output values solution, namely the desired output E_n^N. In this regard, a benchmark solution has been evaluated through a MILP formulation.
The adoption of such synthesis procedure with a supervised formulation allows to avoid both expert operator(s) and heuristics, at least for a preliminary study. Conversely to Mamdani FIS type, the proposed model, based on Takagi-Sugeno formulation, is not sensitive to the MF resolution or, in other words, the spatial granularity. Moreover, it does not need an a-priori analysis of the GA complexity and efficiency, that is strictly related to the FIS number of parameters to be tuned.

3.1 ANFIS EMS Structure

ANFIS models are one of the most popular type of fuzzy artificial neural networks. They are composed by 7 layers. As well described in (Jang, 1993), ANFISs implement FISs by means of a suitable set of first order Takagi-Sugeno rules. The generic j^{th} rule has the form:

$$\text{if } x_1 \text{ is } \Phi_1^{(j)} \text{ and } \ldots x_m \text{ is } \Phi_m^{(j)} \text{ then } y = \sum_{i=1}^{m} \theta_i^{(j)} x_i + \theta_0^{(j)} \tag{4}$$

where $\mathbf{x} = [x_1, \ldots, x_m]$ is a generic crisp input vector. Each x_i is evaluated by the respective rule antecedent term set, defined by the Fuzzy Set MF Φ_i. The second term y is the output associated to the j^{th} rule. It is estimated through the calculation of the associated rule consequent hyperplane, defined by the coefficients θ_i. In this work it has been decided to define every rule antecedent by a unique MF. Therefore, the ANFIS MFs are modelled by means of multivariate Gaussian functions which assure the coverage of the entire fuzzy domain regardless of the number of employed MFs. The generic MF $\Phi(\bar{\mathbf{u}})$, where the input vector $\bar{\mathbf{u}}$ has been introduced in Sec. 2, is defined as follows:

$$\Phi(\bar{\mathbf{u}}) = e^{-\frac{1}{2}(\bar{\mathbf{u}}-\boldsymbol{\mu})\cdot\mathbf{C}^{-1}\cdot(\bar{\mathbf{u}}^T-\boldsymbol{\mu}^T)} \tag{5}$$

where $\boldsymbol{\mu}$ and $\mathbf{C}$ are the mean vector value and the covariance matrix of the multivariate Gaussian function, respectively. The consequent fuzzy rule is modelled as in (4). The rule consequent outputs E^N, the energy exchanged with the grid, that is evaluated by means of a suitable hyperplane defined as follows:

$$E^N = \theta_0 + \theta_1 E^{GL} + \theta_2 C^{sell} + \theta_3 C^{buy} + \theta_4 SoC \tag{6}$$

The overall output of the ANFIS is computed by adopting a Winner Takes All strategy. All rule weights are fixed to unitary values.

3.2 Benchmark Solutions and Objective Function Formulation

Algorithms based on MILP, along with Dynamic Programming and Linear Programming, namely methods able to find an optimal solution through a deterministic (or sub-optimal since MILP is supported by heuristics) approach, are suitable for the determination of a benchmark solution (Sundstrom and Guzzella, 2009) useful to validate and support the EMS modelling.
In this case, it supports the EMS modelling by casting the problem from unsupervised to supervised learning. Specifically, the ANFIS model is trained on a given dataset together with its respective benchmark solution found through a MILP formulation of the problem, by re-adapting the approach proposed in (Palma-Behnke et al., 2013). By defining P_{upper} and P_{lower} as the MILP optimal solution obtained with and without the ESS, respectively, the OF in ((3)) can be rewritten as:

$$\bar{P} = \frac{P_{upper} - P}{P_{upper} - P_{lower}} \tag{7}$$

Such profit normalization allows to estimate how much the EMS performances are close to the optimal solution considering a given dataset. The respective upper benchmark solution ESS *SoC* profile and E^N profile are named SOC_{opt} and E^N_{opt}, respectively. These are used for the EMS synthesis procedure.

3.3 ANFIS Synthesis by Clustering

In literature, several techniques to train an ANFIS architecture have been analyzed. For example, backpropagation-based and clustering-based training have been proposed in (Jang, 1993) and (Rizzi et al., 1999), respectively. In this work, the latter technique is adopted. It exploits a clustering algorithm in order to build the ANFIS architecture, in particular the MFs' shape and the rule based system.
Specifically, three different clustering algorithms are introduced and successively compared for the EMS synthesis problem. By starting from the widely-known k-means algorithm (MacQueen, 1967; Lloyd, 1982), the others are mainly re-adaptations and/or extensions of it (still, well-known in literature) in order to explore and implement different (dis)similarity measures. Moreover, it is discussed how to take advantage of the output space (*i.e.* the E^N_{opt} benchmark solution) in case of clustering in the so-called *joint input-output* space, as explained in (Panella et al., 2001).

3.3.1 k-means

k-means is an hard partitional clustering algorithm which, given a dataset $S = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{N_P}\}$, returns k non-overlapping groups (clusters), *i.e.* $S = \{S_1, ..., S_k\}$, such that $S_i \cap S_j = \emptyset$ if $i \neq j$ and $\cup_{i=1}^{k} S_i = S$, such that objects in the same cluster are more similar to each other than to those in other clusters. In order to find such clusters, k-means aims at minimizing the following objective function, namely the Within-Cluster Sum-of-Squares (WCCS):

$$WCSS = \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \mathbf{r}(i)\|_2^2 \tag{8}$$

where $\|\mathbf{x} - \mathbf{r}(i)\|_2^2$ is the squared Euclidean distance between pattern $\mathbf{x}$ and the i^{th} cluster representative $\mathbf{r}(i)$, usually known as *centroid*, defined as the component-wise mean amongst patterns in cluster S_i. Minimizing (8) is, however, an NP-hard problem (Aloise et al., 2009) and what is commonly known as k-means is actually an heuristic which, as such, does not guarantee to find an optimal solution. k-means is based on the Voronoi iteration or, equivalently, an Expectation-Maximization algorithm which works as follows:

i Select k initial centroids according to some heuristics (*e.g.* randomly);

ii Assignment (Expectation) Step: assign each pattern to nearest cluster (closest centroid);

iii Update (Maximization) Step: update clusters' centroids;

iv Loop ii–iii until a given stopping criterion is met (e.g. maximum number of iterations is reached or centroids' update is below a given threshold).

3.3.2 k-medians

A commonly used variant of the k-means algorithm consists in changing the (dis)similarity measure from squared Euclidean distance to 1-norm (also known as Manhattan, TaxiCab or CityBlock distance), leading to the so-called k-medians problem (Bradley et al., 1997).
The 1-norm (dis)similarity measure implies to consider the Assignment Step and the OF (9) with no squares involved, but considering the absolute value only. Therefore, the k-medians OF shall be referred to as, more generally, the Within-Clusters Sum-of-Distances (WCSD):

$$WCSD = \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \mathbf{r}(i)\|_1 \tag{9}$$

k-medians still works by the Expectation-Maximization steps introduced in Sec. 3.3.1. However, in this case the cluster's representative is the *median*, evaluated by taking the component-wise median rather than the mean amongst patterns in clusters. Due to the minimization of the 1-norm rather than squared 2-norm, k-medians is more robust to noise and outliers with respect to k-means; indeed, the median is not (so-much) skewed in presence of (few) very low or very high values.

3.3.3 k-medoids

In k-medoids (Kaufman and Rousseeuw, 1987) the cluster's representative (known as *medoid* or *MinSOD*[1]) is the cluster datapoint which minimizes the sum of distances within the cluster itself. Conversely to k-means and k-median, in k-medoids clusters' representatives are actual members of the dataset at hand by definition. In this work, the k-medoids problem has been solved by means of the implementation proposed in (Park and Jun, 2009), which is based to the same Voronoi iterations at the basis of k-means and k-medians. k-medoids, due to the representatives' definition, can ideally deal with any (dis)similarity measures. Therefore, its objective function can generally be defined as

$$WCSD = \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} D(\mathbf{x} - \mathbf{r}(i)) \tag{10}$$

where $D(\cdot,\cdot)$ is the (dis)similarity measure. In this paper, the adopted (dis)similarity measure for k-medoids is the Mahalanobis distance (Mahalanobis, 1936), defined as following:

$$d(\mathbf{x},\mathbf{r}(i)) = \sqrt{(\mathbf{x}-\mathbf{r}(i))^T \cdot \mathbf{C}_i^{-1} \cdot (\mathbf{x}-\mathbf{r}(i))} \tag{11}$$

where $\mathbf{C}_i$ is the covariance matrix for the i^{th} cluster and $\mathbf{r}(i)$ is its representative (*i.e.* the medoid).
Since the k-medoids algorithm minimizes the sum of pairwise distances rather than the sum of squares, it is more robust to noise and outliers with respect to k-means.

3.3.4 ANFIS Synthesis by Joint Input-output Space Clustering

Albeit clustering is an unsupervised problem by definition, the dataset at hand consists in labelled patterns; thus, it has the form $S = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), ..., (\mathbf{x}_{N_P}, y_{N_P})\}$ where $\mathbf{x}_i \in \mathbb{R}^{N_F}$ and $y_i \in \mathbb{R}$, for $i = 1, ..., N_P$, which we refer to as *input* and *output* space(s), respectively. Since in this work the clustering problem in the joint input-output space will be considered, the cluster's representative(s) must be re-defined. Indeed, if one has to work in the input space (*i.e.* with unlabelled patterns), the clusters' representatives as defined in Secs. 3.3.1–3.3.3 suffice. Conversely, as concerns joint input-output spaces, each cluster will be described by:

- its original representative $\mathbf{r}$ (either mean, median or medoid – depending on the algorithm at hand)
- its covariance matrix $\mathbf{C}$
- a set of $N_F + 1$ coefficients $\boldsymbol{\theta}$

The former two quantities will be used in order to build the ANFIS MF according to (5), whereas the latter will be used in order to define the hyperplane which locally approximates the input-output mapping according to (6). Specifically, it can be evaluated using the Least Mean Squares (LMSs) estimator:

$$\boldsymbol{\theta}_i = \left(\mathbf{X}_i^T \mathbf{X}_i\right)^{-1} \mathbf{X}_i^T Y_i \tag{12}$$

where $\mathbf{X}_i$ is the set of input patterns lying in the i^{th} cluster and Y_i is the set of corresponding output values (ground-truth). It is worth noticing that patterns in $\mathbf{X}_i$ will be augmented by appending a heading 1 such that their dimension is $N_F + 1$: in this manner $\boldsymbol{\theta}_i \in \mathbb{R}^{N_F+1}$, as it also considers the hyperplane's intercept (cf. (6)).
In order to fully consider both the input and output spaces, the (dis)similarity measure has been readapted, regardless of the specific adopted clustering algorithm. The pattern-to-cluster (dis)similarity measure is defined as a convex linear combination between the point-to-representative distance (input space) and the approximation error given by the interpolating hyperplane (output space):

$$\hat{d}(\mathbf{x}, \langle \mathbf{r}, \mathbf{C}, \boldsymbol{\theta} \rangle) = \varepsilon \cdot d(\mathbf{x},\mathbf{r}) + (1-\varepsilon) \left(y - \boldsymbol{\theta}^T \cdot \mathbf{x}\right)^2 \tag{13}$$

where $d(\cdot,\cdot)$ is one of the given (dis)similarity measures (either squared Euclidean, Manhattan or Mahalanobis), the triad $\langle \mathbf{r}, \mathbf{C}, \boldsymbol{\theta} \rangle$, as introduced, defines the fuzzy rule and, finally, $\varepsilon \in [0,1]$ is a trade-off parameter which tunes the linear convex combination. It is worth noticing that if $\varepsilon = 1$ the rightmost term in (13) will not be considered, thus collapsing into a standard clustering problem.
The rationale behind (13) is that the algorithm aims at minimizing the approximation error due to the hyperplane (rightmost term) and, at the same time, at discovering well-formed[2] clusters in the input space (leftmost term).

[1] Minimum Sum Of Distances

[2] either compact or homogeneous, depending on the (dis)similarity measure and, by extension, on the clustering algorithm objective function

4 EMS MODELLING PROCEDURE

In this section is explained in details how the ANFIS EMS is efficiently modelled through the exploitation of a clustering algorithm. The ANFIS optimization process is explained from a generic point of view, valid for each clustering algorithm proposed in the previous section, including their variants (*i.e.* along with the output space). The clustering algorithm relies on a given dataset composed by E^{GL}, C^{buy} and C^{sell} time series.

The whole dataset has been partitioned in TrS, VlS and TsS (*i.e.* Training Set, Validation Set and Test Set, respectively). These are used to train the ANFISs, to select the best for a given number of MFs and to measure the performances of the optimum one, respectively. The dataset partition among TsS, TrS and VlS is made on a daily base, namely all time slots associated with the same day will belong to a single set. More precisely, the whole dataset is firstly divided in two subsets having the same cardinality. The first subset constitutes the TsS whereas the second one is partitioned in 5 different ways in order to constitute 5 different TrS and VlS pairs.

Algorithm 1 EMS Training Procedure

procedure EMS DESIGN
  TsS and $\langle TrSs, VlSs\rangle$ pairs partitioning
  for $j = 1$ to 5 **do** ▷ for each TrS_j
    $\{E^N_{opt}, SoC_{opt}\}_j := \text{MILP}(TrS_j)$
      ▷ evaluation of the optimal solution
    $\Gamma^{cl}_j := \{TrS_j, SoC_{opt}, E^N_{opt}\}$
      ▷ Clustering input dataset
  end for

  for $k = 2$ to 25 **do** ▷ for each value of k
    for $j = 1$ to 5 **do** ▷ for each TrS_j
      $\{\mu_{kj}, \mathbf{C}_{kj}, \boldsymbol{\theta}_{kj}\} := \text{clustering}(k, \Gamma^{cl}_j)$
      $\boldsymbol{\Phi}_{kj} := \{\mu_{kj}, \mathbf{C}_{kj}\}$ ▷ MF evaluation
      $ANFIS_{kj} := \{\boldsymbol{\Phi}_{kj}, \boldsymbol{\theta}_{kj}\}$
      simulation of $ANFIS_{kj}$ on $VlS_j \rightarrow \bar{P}_{kj}$
    end for
    selection of $ANFIS^{best}_k$ according with $\bar{P}_{kj}$
    simulation of $ANFIS^{best}_k$ on the $TsS \rightarrow P_k$
  end for
end procedure

The clustering procedure exploits the generic TrS_j dataset, namely the TrS of the j^{th} TrS-VlS pair, together with the corresponding optimum profiles of $(SoC_{opt})_j$ and $(E^N_{opt})_j$ found through the MILP optimization. For the sake of ease, the dataset read by the clustering algorithm is defined by Γ^{cl}_j. It is clear that if the only EMS input space is considered, $(E^N_{opt})_j$ will not take part as a feature for clustering and will be only used for the estimation of the hyperplanes coefficients defined by $\boldsymbol{\theta}_j$, once they are defined (see (12)). Conversely, if the clustering procedure is applied on the joint input-output space, the coefficients $\boldsymbol{\theta}_j$ are updated after each Voronoi iteration together with the clusters. The optimization process is ran for a number of clusters k ranging from $k = 2$ to $k = 25$ on the j^{th} dataset Γ^{cl}_j.
For each k, the clustering procedure is repeated for each Γ^{cl}_j (*i.e.* 5 times), returning the respective k representative vectors, covariance matrices and hyperplanes coefficients. These are used to model the ANFIS multivariate Gaussian MFs $\boldsymbol{\Phi}_{kj}$ (*i.e.* rule antecedent set) and the Sugeno hyperplanes (*i.e.* rule consequent set) represented by the $\boldsymbol{\theta}_{kj}$ coefficient sets, respectively, according to (5) and (6), as introduced in Sec.3.3.4. The previously described procedure results in 24×5 different ANFIS synthesis. More precisely, for each k value there exist 5 different ANFIS corresponding to the 5 different TrSs. Each generated ANFIS is then simulated on its respective VlS and its performance is evaluated according to (7). For each value of k, the best ANFIS among the 5 TrS-VlS pairs, named $ANFIS^{best}_k$, is selected whereas the remaining 4 are discarded. Finally, for each $ANFIS^{best}_k$ the MG EMS is simulated on the TsS in order to evaluate the generalization capability. The overall procedure of the EMS ANFIS design and optimization is illustrated in Algorithm 1.

5 SIMULATION SETTINGS

In this work it has been considered a MG composed by the following energy systems: a PV generator of 19 kW, an aggregated load with a peak power around 8 kW, and ESS with an energy capacity of 24 kWh. For the ESS modelling it has been taken into consideration the Toshiba ESS SCiB module having a rated voltage of 300 V, a current rate of 8 C-Rate and a capacity of about 80 Ah.
The dataset used in this work has been provided by *AReti S.p.A.*, the electricity distribution company in Rome. The energy prices are the same used in previous works, (Leonori et al., 2016a; Leonori et al., 2016b).
The considered dataset covers an overall period of 20 days, sampled with a 15 minutes frequency. The overall dataset is shown in figure 2, together with the energy prices both in sale (positive) and purchase (negative).

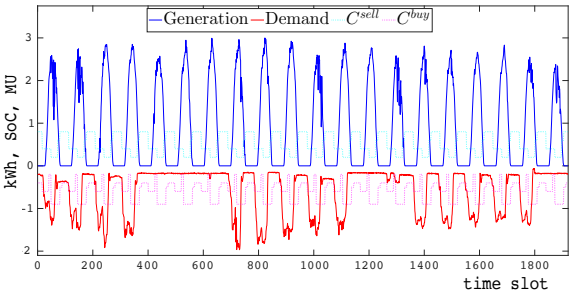

Figure 2: MG power production and demand of the overall dataset and TOU energy prices.

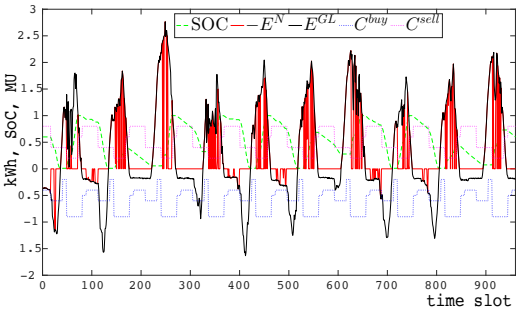

Figure 3: MG optimum solution energy flows, ESS *SoC* and energy prices computed on the TsS by MILP approach.

The even days are assigned to the TsS, whereas the odd days are partitioned with a random selection between the TrS and VlS in order to form 5 different TrS-VlS partitions (see Sec. 4). It has been chosen to assign the 70% of the odd days to the TrS and the remaining 30% to the VlS.

The ANFIS optimization process is executed for each clustering algorithm introduced in Sec. 3.3 and by varying the ε values, namely the local approximation error influence, as follows:

- k-means for ε equal to $1, 0.75, 0.5$ and 0.25.
- k-medians for ε equal to $1, 0.75, 0.5$ and 0.25.
- k-medoids for ε equal to $1, 0.75, 0.5$ and 0.25.

The clustering algorithms have been set with a maximum number of iterations equal to 50 and every execution of the algorithm (*i.e.* number of replicates) is repeated 20 times with a new, random, initial representatives selection.

The solution chosen by the clustering algorithm after each run is the one which minimizes its respective objective function (see Sec. 3.3), namely the WCCS defined in (8) for k-means, (9) for k-medians and (10) for k-medoids.

6 RESULTS

The simulation results have been studied by computing the OF $\bar{P}$ defined in (7) on the TsS, whose optimal solution is shown in figure 3.

The results on the TsS given by the procedure described in Algorithm 1 are shown in figure 4, as a function of k, by reporting the respective value of $\bar{P}$ for each $ANFIS_k^{best}$ generated. These are grouped by the different ε coefficient values, which range between 1 and 0.25 as defined in Sec. 5. It is possible to observe that k-medoids in most cases presents the best results, being its OF curves (in green) lower with respect to the other two competitors (k-means in red and k-medians in blue). Moreover, for lower values of ε, especially for $\varepsilon = 0.25$ (*i.e.* when the approximation error due to hyperplane prevails (see (13)), the k-medoids results appear to be more stable as k increases with respect to the other two algorithms (see figure 4-d). k-means and k-medians show a smoother behaviour of $\bar{P}$, especially for $\varepsilon = 1$ (figure 4-a), meaning that these are more stable when considering the input space only. In all cases their profit results are around 80% of the optimum solution.

In order to focus on the results reliability and sensitivity, the carried out tests have been repeated 10 times. In this study, for the sake of clarity, it has been decided to select only $ANFIS_k^{best}$ associated with the best OF result on the TsS. The selected solutions are displayed through a boxplot representation in figure 5 after being grouped by ε coefficient and clustering algorithm. Specifically, in figure 5-a are reported the $ANFIS_k^{best}$ selected solution OF ($\bar{P}$) values; in figure 5-b their respective number of clusters and, finally, in figure 5-c are reported the Davies-Bouldin Index (DBI) values (Davies and Bouldin, 1979), which measures both intra-clusters compactness and inter-clusters separation.

As shown in figure 5-a, best results are confirmed to be prevalently given by k-medoids. These are followed by the k-medians. On the other hand, k-medoids solutions present high level of variability both on the value of k and the DBI, as pointed by higher extension of the boxes (see figure 5-b and c). Nevertheless, for $\varepsilon = 0.5$ the boxplots show a lower variance for $\bar{P}$, k and DBI. In k-means and k-medians solutions, considering the joint input-output space term seems to be scarcely significant. Only for k-means the joint input-output space term in OF yields a relevant improvement in terms of reduction of the DBI and the number of clusters. More into details, as far as k-medoids is concerned, by looking at the DBI it is possible to see that the clustering

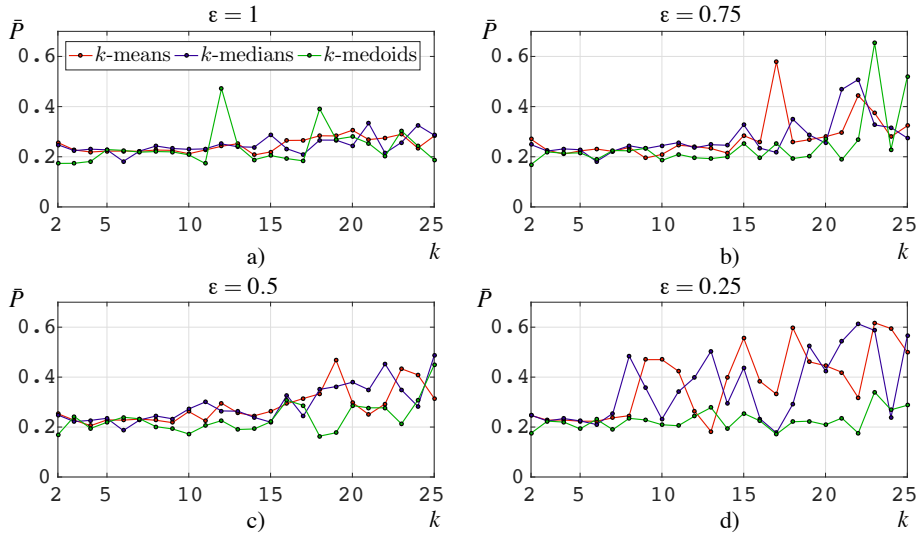

Figure 4: $ANFIS_k^{best}$ normalized profit values evaluated on TsS as a function of k. Solutions are grouped according to ε.

problem is hard to solve[3], whereas the k-means and k-medians lead to better solutions. However, in terms of OF (see (7)) k-medoids overperforms the other two competitors. This is mainly due to the Mahalanobis distance which, by considering up to the second-order statistics (*i.e.* the covariance matrix), is aware of the clusters' shapes as well. Indeed, by using the Mahalanobis distance k-medoids effectively updates the covariance matrix at each iteration, conversely to the other two clustering-based ANFIS synthesis procedures, where the covariance matrix is evaluated once, at the end of the clustering procedure.

7 CONCLUSION

In this work it has been investigated a procedure for the synthesis of a MG EMS based on computational intelligence techniques. It performs a cluster analysis in order to find automatically the number and the location on the input space of the rules composing an ANFIS, as the core inference engine of the EMS.
The ANFIS synthesis is casted as a supervised machine learning problem, where patterns are input-output pairs, taking as ground-truth output values the ones coming from the benchmark solution obtained by a MILP procedure.
In particular, three considered clustering algorithms have been compared, which differ for their respective (dis)similarity measures and the way clusters' representatives are evaluated. Besides, a clustering procedure in the joint input-output space has been investigated, evaluating its effectiveness when the weighting coefficient in the OF changes. Results show that adopting the Mahalanobis distance on the joint input-output space leads to profits way superior than k-means. This improvements, however, are paid with higher values of both the average number of clusters (*i.e.* model complexity) and its variance (*i.e.* algorithm robustness).
On the other hand, the adoption of the k-medians seems a good compromise in terms of both robustness and effectiveness of OF performance and partitions quality in terms of DBI. Starting from these findings, further experiments can be done. First, it is possible to adopt a hierarchical clustering algorithm in order to better deal with larger datasets, improving both speed and accuracy in clusters discovery. Moreover, a genetic algorithm can be applied to tune the ANFIS parameters (De Santis et al., 2017) (*e.g.* rule weights and MF shape) once it is synthesised by the clustering algorithm.

[3]Recall: the lower DBI, the better the partition. Indeed, a low DBI means low intra-variance (high compactness) and high inter-variance (high separation).

REFERENCES

Aloise, D., Deshpande, A., Hansen, P., and Popat, P. (2009). Np-hardness of euclidean sum-of-squares clustering. *Machine learning*, 75(2):245–248.

Amer, M., Naaman, A., M'Sirdi, N. K., and El-Zonkoly, A. M. (2014). Smart home energy management systems survey. In *International Conference on Renewable Energies for Developing Countries 2014*, pages 167–173.

Arcos-Aviles, D., Pascual, J., Marroyo, L., Sanchis, P., and Guinjoan, F. (2016). Fuzzy logic-based en-

ergy management system design for residential grid-connected microgrids. *IEEE Transactions on Smart Grid*, PP(99):1–1.

Bradley, P. S., Mangasarian, O. L., and Street, W. N. (1997). Clustering via concave minimization. In *Advances in neural information processing systems*, pages 368–374.

Davies, D. L. and Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227.

De Santis, E., Rizzi, A., and Sadeghian, A. (2017). Hierarchical genetic optimization of a fuzzy logic system for energy flows management in microgrids. *Applied Soft Computing*, 60:135 – 149.

Deng, R., Yang, Z., Chow, M. Y., and Chen, J. (2015). A survey on demand response in smart grids: Mathematical models and approaches. *IEEE Transactions on Industrial Informatics*, 11(3):570–582.

Dragicevic, T., Vasquez, J. C., Guerrero, J. M., and Skrlec, D. (2014). Advanced lvdc electrical power architectures and microgrids: A step toward a new generation of power distribution networks. *IEEE Electrification Magazine*, 2(1):54–65.

Jang, J.-S. (1993). Anfis: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics*, 23(3):665–685.

Kaufman, L. and Rousseeuw, P. (1987). Clustering by means of medoids. *Statistical data analysis based on the L1-norm and related methods*.

Kirschen, D. S. (2003). Demand-side view of electricity markets. *IEEE Transactions on Power Systems*, 18(2):520–527.

Leonori, S., De Santis, E., Rizzi, A., and Frattale Mascioli, F. M. (2016a). Multi objective optimization of a fuzzy logic controller for energy management in microgrids. In *2016 IEEE Congress on Evolutionary Computation (CEC)*, pages 319–326.

Leonori, S., De Santis, E., Rizzi, A., and Frattale Mascioli, F. M. (2016b). Optimization of a microgrid energy management system based on a fuzzy logic controller. In *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, pages 6615–6620.

Leonori, S., Paschero, M., Rizzi, A., and Frattale Mascioli, F. M. (2017). An optimized microgrid energy management system based on fis-mo-ga paradigm. In *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6.

Lloyd, S. (1982). Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA.

Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India, 1936*, pages 49–55.

Palma-Behnke, R., Benavides, C., Lanas, F., Severino, B., Reyes, L., Llanos, J., and Sez, D. (2013). A microgrid energy management system based on the rolling horizon strategy. *IEEE Transactions on Smart Grid*, 4(2):996–1006.

Panella, M., Rizzi, A., Frattale Mascioli, F. M., and Martinelli, G. (2001). Anfis synthesis by hyperplane clustering. In *Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569)*, volume 1, pages 340–345 vol.1.

Park, H.-S. and Jun, C.-H. (2009). A simple and fast algorithm for k-medoids clustering. *Expert systems with applications*, 36(2):3336–3341.

Patterson, B. T. (2012). Dc, come home: Dc microgrids and the birth of the "enernet". *IEEE Power and Energy Magazine*, 10(6):60–69.

Rizzi, A., Frattale Mascioli, F. M., and Martinelli, G. (1999). Automatic training of anfis networks. In *Fuzzy Systems Conference Proceedings, 1999. FUZZ-IEEE '99. 1999 IEEE International*, volume 3, pages 1655–1660 vol.3.

Sundstrom, O. and Guzzella, L. (2009). A generic dynamic programming matlab function. In *2009 IEEE Control Applications, (CCA) Intelligent Control, (ISIC)*, pages 1625–1630.

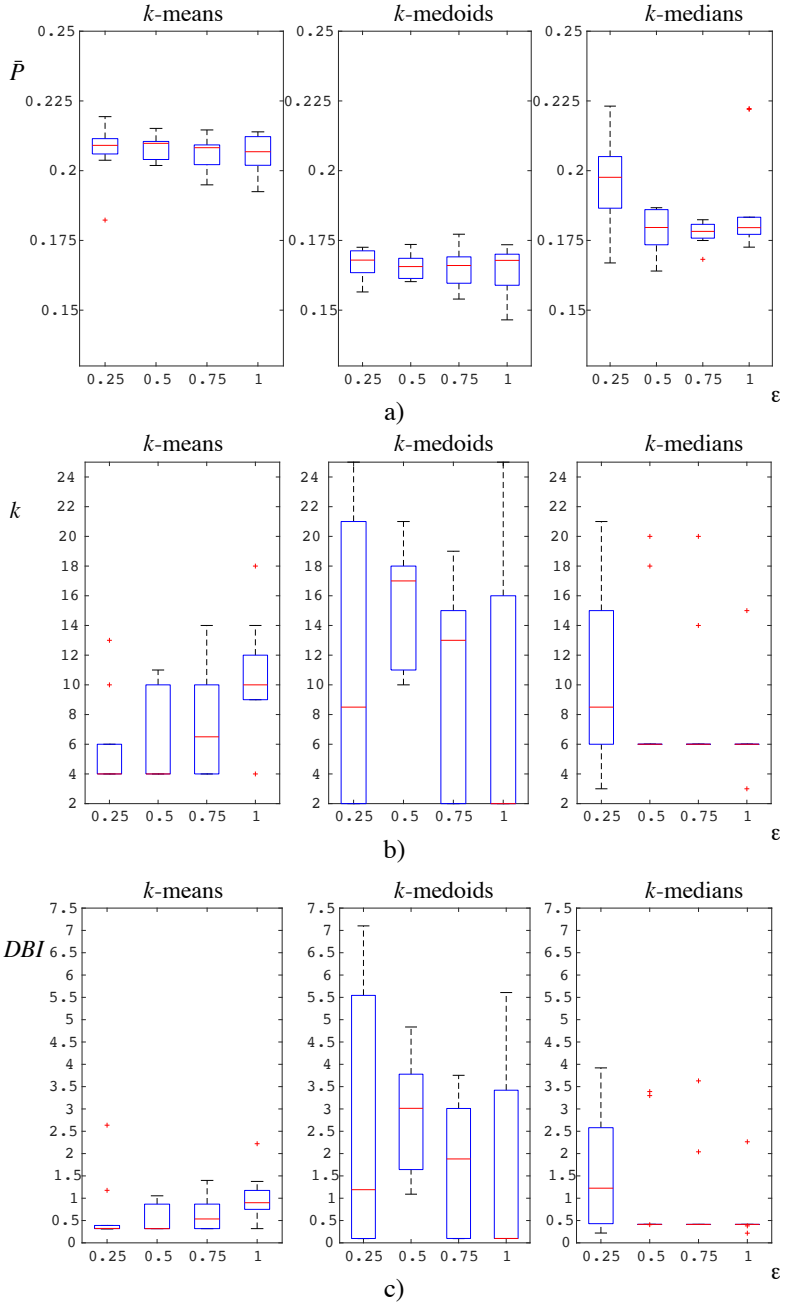

Figure 5: Simulation results on TsS considering the best solution of each run illustrated through a box-plot representation. Top and bottom box sides correspond to first and third quantiles, respectively, whereas the red dash corresponds to the median. Top and bottom whiskers extremities correspond to the maximum and minimum points not considered as outliers, respectively. These latter are marked with a red + symbol. In (a) are shown the OF values; in (b) the number of MFs; in (c) the DBI associated to the clustering solution which models each $ANFIS$.