Applied Soft Computing 53 (2017) 323–335

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Load estimation for microgrid planning based on a self-organizing map methodology

Jacqueline Llanos [a], Raúl Morales [a], Alfredo Núñez [b,*], Doris Sáez [a], Matías Lacalle [a], Luis Gabriel Marín [a], Roberto Hernández [c], Fernando Lanas [a]

[a] Electrical Engineering Department, University of Chile, Santiago, Chile
[b] Section of Railway Engineering, Delft University of Technology, Delft, The Netherlands
[c] Energy Center, Faculty of Physical and Mathematical Sciences, University of Chile, Santiago, Chile

ARTICLE INFO

Article history:
Received 4 April 2016
Received in revised form 9 December 2016
Accepted 29 December 2016
Available online 3 January 2017

Keywords:
Planning
Microgrid
Self-organizing map (SOM)
Load profile

ABSTRACT

This study presents a novel load estimation method for isolated communities that do not receive energy or only receive it for a limited time each day. These profiles have been used to determine the installed capacity of generating units for microgrid electrification projects. The social characteristics and lifestyles of isolated communities differ from those in urban areas; therefore, the load profiles of microgrids are sensitive to minor variations in generation and/or consumption. The proposed methodology for obtaining the residential profiles is based on clustering algorithms such as k-means, a self-organizing map (SOM) or others. In this work, SOM clustering is considered because it allows a better interpretation of results that can be contrasted with social aspects. The proposed methodology includes the following components. First, the inputs are processed based on surveys of residents that live in each socio-economic level of housing and the community. Second, family types are clustered using an SOM, from which relevant information is derived that distinguishes one family from another. Third, the load profiles of each cluster are selected from a database. Additionally, social aspects and relevant energy supply information from communities with similar characteristics are used to generate the required database. The SOM for the clustering of families of the community with available energy measurements is used as an initial guess for the clustering of the families in the community with unknown energy measurements.

The methodology is applied and tested in the community of El Romeral, Chile, where a microgrid will be installed. The SOM technique compares favorably with a benchmark method that uses the average load profile of a community; furthermore, the SOM clustering algorithm for the methodology is favorably compared with the k-means algorithm because the results obtained by SOM are consistent with the social aspects.

1. Introduction

When designing and developing renewable energy projects that can provide power to an area, information should be obtained on the available energy resources and required power supply. Because of the uncertainties surrounding the availability of resources and their consumption, ensuring a sufficient capacity and availability of electricity to supply peak demand and daily energy consumption levels must be prioritized [1].

The planning and operation of traditional low-voltage electrical networks require the use of load models. Most power companies implement systems that can automatically read electricity consumption (AMR, automatic meter reading) and determine consumption profiles. Records of these measurements have been used to determine electricity consumption classes and behavioral patterns of energy consumers and provide significant improvements in electricity demand forecasting. However, there are a number of nAMR customers (users without automated meters) for whom the consumption profile is not known [2]. This study focuses on consumers that live in isolated communities without an energy supply or only a partial supply and for whom historical records of total consumption or housing are not available to use as references when measuring microgrid generating units and increasing the efficiency of providing electricity to these areas.

Residential demand accounts for most of the system load in isolated electrical systems. These loads are currently modeled with

* Corresponding author.
E-mail address: a.a.nunezvicencio@tudelft.nl (A. Núñez).

http://dx.doi.org/10.1016/j.asoc.2016.12.054

generalized profiles defined by statistical distributions, such as load profiles based on Gaussian functions that capture residential customer behaviors, which have been traditionally assumed to be homogeneous [3]. In small systems, simply turning on and off several appliances may generate significant disturbances to the overall power consumption profile, and several projects have focused on residential demand and proposed algorithms that track the behavior of small loads indicative of changes in the profile through the use of Bayesian change points to identify loads that may appear unpredictable [4].

Demand profiles have traditionally been generated according to consumption measurements, although techniques have been used to identify characteristic patterns of electrical appliance use, particularly in Canadian households [5]. Similarly, the energy consumption profiles for one or more families can be generated by combining the electricity demand of each appliance with a probabilistic approach [6]. In Dickert and Schegner [7], a load curve model based on a probabilistic time series was presented along with measurements of different types of apparatuses used to determine individual load curves and analyze the sequence and timing of operations to generate probabilistic profiles for each appliance according to the appliance power, use frequency, ignition time, and operation times to obtain a load curve per customer or group.

Surveys are also useful tools for generating electrical profiles in domestic buildings as shown by work recently conducted in the UK and reported in Ref. [8]. Simulation profiles have also achieved a good approximation of electrical energy usage based on measurements at a substation [9]. Estimating electricity demand is an insufficient approach in cogeneration systems; thus, per-hour thermal profiles must be generated to optimize electricity usage [10].

Generating load profiles without measurements is a more difficult task, and limited developments have been achieved in this area. In Ref. [2], the authors proposed a method for generating TLPs (typical load profiles) for smart grids by using AMR customer data to analyze loads and generate a virtual load profile (VLP) for nAMR customers, with the data subsequently clustered and classified.

A number of studies have considered stages of classification for load profiles generated in their models, and these stages include clustering residential customers according to their appliances, identifying customer groups based on the number of residents [7], and classifying users according to the type of electronics they own and times at which the electronics are operated [6]. In Kim et al. [2], the authors evaluated several classification techniques for classifying AMR user profiles, such as k-means and fuzzy c-means, which were later used to generate nAMR user profiles.

Self-organizing map (SOM) methodologies have been proposed in several studies. For short-term load forecasting, [11] proposed the use of an input data classifier based on Kohonen neural networks. In Valero et al. [12], two methods were proposed for short-term load forecasting using SOMs for classifying and memorizing historical data. According to Ref. [13], one of the major advantages of using SOM for short-term load demand forecasting is its ability to display an intuitive visualization to compare similar data. In Ref. [14], SOMs were used to automatically classify electricity customers based on their domestic energy consumption demand patterns using a measurement database. In Ref. [15], SOMs were used for segmentation and demand pattern classification for electrical customers. For short-term load forecasting, a neural model containing up to two hierarchical SOMs was proposed in Ref. [16]. In Ref. [17] SOMs were used to cluster the data and support vector machine (SVM) to fit the testing data for predicting the daily peak load for mid-term load forecasting purposes. In Ref. [18] a soft computing system was proposed for day-ahead electricity price based on SOMs, SVM and particle swarm optimization (PSO), improving the forecasting accuracy. In Ref. [19], the authors presented three of the most used clustering methods, k-means, k-medoid and SOMs, for clustering domestic electricity load profile using smart metering data, in Ireland. SOM proved to be the most suitable and was therefore used to segment the data; a Davies-Bouldin (DB) validity index was used to identify the most suitable clustering method and an appropriate number of clusters.

However, these methods are applicable to only traditional power systems. In the case of microgrids, generating profile results is more difficult because of the high variation and uncertainty of load behavior with regard to domestic energy consumption. A load profile generation method for isolated microgrid projects was presented in Ref. [20], in which the information is obtained from a socio-economic survey that is conducted in a community, and an SOM classification stage is used to generate a characteristic load profile for each class. However, the load profiles are based on limited measurements from other grid-connected communities, whose load behavior could differ from that of an isolated community, such as not including electricity consumption measurements.

In this paper, an SOM algorithm is used as a clustering method for generating both the clusters and a representative for each cluster. Several clustering techniques exist; a notable example is the k-means technique, in which each cluster is represented by the most centrally located object in the cluster [21]. The k-means algorithm has also been used for power systems applications such as in Ref. [22] for identifying similar types of profiles of a practical system for demand variation analysis and energy loss estimation, and in Ref. [23] for classifying and recognizing the voltage sag from the measured historical data of a large-scale grid in China.

According to Ref. [21], there are three main approaches for clustering times series: raw-data-based, feature-based, and model-based. Among them, SOM and k-means are raw-data-based methods that allow the user to work with raw data directly. In this paper, a methodology for obtaining the residential profiles is proposed considering clustering algorithms such as k-means, SOMs, or others. In particular, SOM clustering is considered because it allows a better interpretation of results that can be contrasted with social aspects. Thus, the implementation of an SOM is described for estimating load profiles that can be used to plan microgrids according to the unit sizes. The proposed methodology includes the socio-economic characteristics of the grid users (by surveys) as well as the effect of the consumption behavior of the entire community. In addition, this methodology allows new load profiles with the features of each family to be added. The load profiles are obtained from another community with similar characteristics and with available energy supply measurements. The characteristics of this community are clustering by another SOM, which is used to estimate the electrical demand of the families in the community that lacks measurements. In this study, SOMs are employed because they are suitable for representing the utilized surveys. Using SOMs, the survey properties are visualized and analyzed. They correctly represent the similarities among families, which corresponds with expectations from reality and a practical point of view. Thus, a SOM enables a suitable interpretation of the inputs and results and identifies the similarities and differences in the prototypes. Unlike the K-means, the SOMs do not require the cluster number.

The proposed methodology is applied to the community of El Romeral, which is located 21 kilometers from La Serena, Chile. This area lacks basic electricity services and potable drinking water [24], and a microgrid is currently being planned for the energy supply; therefore, an estimate of the load profile is required. The required profiles of a similar community are collected from measurements of Huatacondo village, which is located 230 kilometers from Iquique, Chile, and has a microgrid that operates in standalone mode [25]. The remainder of this paper is organized as follows. Section 2 describes the proposed load estimation method that is based on an SOM, Section 3 provides the case study of the El Romeral community, and Section 4 presents the conclusions and suggestions for further research.

2. Load estimation methodology based on SOM

This section describes the proposed methodology in detail. The problem statement is explained, the basic concepts of the SOM are described, and the load estimation method that is based on SOMs is presented. Finally, the methodology is formulated for a specific community.

2.1. Problem statement

One of the main concerns associated with the design and planning of microgrids for isolated communities, where energy is not always available, is how to determine the load for sizing the generation units. Classical load estimation for bulk power systems is not directly applied to microgrids. The large consumption of cities, where the electricity demand is traditionally measured, cannot be scaled to a microgrid because the way of life, economic activity, consumer habits, and appliances vary between isolated communities and large cities. Unlike bulk power systems, the microgrid load has greater variability, and small changes in consumption can significantly affect the electric demand profile. Based on these concerns, methodological tools to support the estimation of the demand for isolated localities are needed because measurements are not available.

The proposed methodology based on computational intelligence is flexible and intuitive for demand estimation in isolated communities that do not have an electricity supply or partial electricity supply. The methodology seeks to solve the problem of the lack of consumption measurements in the study community.

The methodology requires structured surveys for each dwelling and general information from the study community. Electricity measurements from other similar communities are needed. Thus, this methodology is useful for electrification projects, specifically for the design and sizing of the generating units of microgrids that operate in island mode.

The problem statement includes the main objective of estimating the load profiles for families without permanent electricity supply and meters. In the proposed methodology, the measurements and surveys from a similar community that has permanent electricity supply are employed to generate the load profile for other communities without electricity supply.

Thus, a methodology-based SOM for estimating load profiles that can be used to plan microgrids according to the unit size is described.

2.2. Self-organizing map (SOM)

SOMs were originally proposed by Kohonen [26,27], and the principal characteristic of the SOM applied here is its ability to recognize patterns in complex sets via an unsupervised methodology [12]. In this method, a measure is used to determine the distribution of an input space V_i over an output space V_o (generally of a lower dimension). This measure is defined by a group of neurons distributed over a line, rectangular, or hexagonal plane, thus preserving the properties of the patterns in the input space. In Fig. 1, the input space is represented by a vector of inputs x_i, while the output space is given by the values of the diagram of $X \times Y$ nodes that are activated with different colors (from white to red) depending on the input.

The most important feature of SOMs is the possibility of comparing clusters that summarize data. The self-organized network must extract the patterns, regularities, correlations, categories, or important features of each observation and assign them to a cluster,

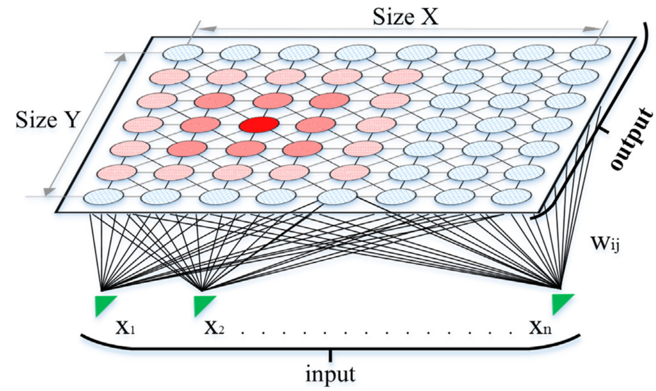

Fig. 1. SOM Diagram.

which is then projected onto a node of the output map. The projections derived from different observations are then compared to estimate the proximity of their respective clusters because similar observations are projected on the same node.

Conversely, dissimilarity increases with the distance between two projections. Therefore, the space cluster is identified with the map so that the projections can be used to simultaneously interpret cluster space (output) and observe space (input) [28], as illustrated in Fig. 1.

The basic SOM training algorithm employs the following steps [29]:

1. Network weights are initialized, which is usually performed at random, although other methods can be employed, such as random entry selection. The weight vector of the neuron j is defined as $\vec{w_j} = \left(\omega_{j1}, \ldots.., \omega_{jn}\right)$, where the weights ω_{ji} is related to the input x_i.

2. Input vector $\vec{x}$ is considered.

3. Active neurons are determined that have weights closest (Euclidean distances) to the vector $\vec{x}$.

4. Weight vectors of the active neuron and those of the neighboring neurons are modified using the following equation:

$$\vec{\omega_j}(t+1) = \vec{\omega_j}(t) + l_r h_{jv}\left(\vec{x} - \vec{\omega_j}(t)\right) \quad (1)$$

where $\vec{\omega_j}$ is the weight vector of the neuron j, l_r is the learning rate, and $h_{jv}()$ is a neighboring function.

The neighborhood size and learning rate are changed (or updated) dynamically during the learning process according to the following equation:

$$l_r(t) = \frac{l_{r0}}{\left(1 + \frac{c \cdot t}{n}\right)} \quad (2)$$

where l_{r0} is the learning rate at the beginning of the iterations; c is a constant, which is usually equal to 0.2; t is the iteration counter; and$n$ is the number of neurons in the network.

5. The procedure is repeated from step 2 with new input vectors $\vec{x}$ until the total number of determined iterations is completed.

After performing the learning process, an input vector $\vec{x} = (x_1, \ldots.., x_n)$ activates neuron j of the output space if the weight vector $\vec{w_j} = \left(\omega_{j1}, \ldots.., \omega_{jn}\right)$ has the least distance from the input vector $\vec{x}$. Thus, each neuron $\vec{w_j}$ corresponds to a prototype vector (an average) of the region of input vectors that trigger neuron j. Thus, two vectors with similar inputs, according to the relationships defined in V_i, activate the same neuron (or two different but nearby neurons) in the output space.

Visualizing SOMs is difficult (see Fig. 1) because the clustering process can be conducted in high-dimensional spaces; however, one of the most popular methods utilizes a unified distance matrix

U that provides a global view of proximity relationships of the reference vectors in the SOM [30].

2.3. Load estimation method

This paper proposes a methodology for estimating the load profiles of residential electricity demand in isolated communities where energy is not always available. In this method, the electricity demand profiles are calculated using information that is obtained from a socio-economic survey that is conducted in the community and do not include measurements of past electricity demand.

In this study, the data are derived from two sources and are employed as the inputs of the clustering algorithms:

- Data 1 includes the survey information for all dwellings of the study communities, i.e., families without permanent electrical supply and unknown past consumption and families with electricity supply. The individual surveys (Table 1) conducted in each community household focus on obtaining information, including household size, age, occupation and income, as well as the number and type of electrical appliances in the household and the number of hours for which they are operated.
- Data 2 is obtained from the electrical meters of families with permanent electrical supply, including the electricity measurements for each family.

The proposed method for obtaining the residential load profile is divided into five main components (Fig. 2). The first **Input Module** includes information about each family in the studied community that is obtained through surveys and site visits to their homes. The second **Clustering Module** of the electricity user incorporates the survey information into a first SOM (named $\mathbf{SOM}_1$) with the following categories: numbers of clusters, elements of each cluster, types of families, and features that differentiate each cluster. The **Update Database Module** that is based on a second SOM (named $\mathbf{SOM}_2$) allows new load profiles to be added with their socio-economic features, which are obtained from both measurements and surveys of other communities that have uninterruptible supplies of electrical energy. The latter component includes a **Database** and a **Search Profile Module**. The **Database** contains profiles of each family type's typical consumption as well as characteristics that differentiate each cluster (these are described in more detail in the next section). The **Search Profile Module** uses a heuristic search method that analyzes the characteristics of each cluster from the previous module and identifies similarities in the database to generate a profile for each cluster that permits the load profile to be estimated. Each module is explained in detail below.

2.3.1. Input module

This module includes relevant information from the community obtained through well-structured surveys, with the information collected as followed: an analysis is performed of magazines, documents, and statistical data from various sources, such as population and housing censuses, websites, and reports in libraries or publications by governmental organizations; and field interviews are performed of representatives of relevant town organizations and members of the community [31].

In this module, Data 1 is employed as inputs for the clustering algorithm and considered as a matrix of the following vectors: number of members, age of each member, activity of each member and income of each member. A quantification is performed for the activity of the members by assigning the following codes to each activity: 1 retired, 2 farmer, 3 housewife, 4 student, 5 working day, 6 working half-day (day), 7 working half-day (afternoon), 8 night work, 9 home trade, and none of these activities.

The individual surveys (Table 1) conducted in each community household focus on obtaining information that includes household size, age, occupation and income as well as the number and type of electrical appliance in the household and hours that each appliance is used. In this module, data processing is performed, and erroneous data are removed, missing numerical data are estimated, qualitative variables obtained from surveys are assigned, and data are normalized. In this case, the input normalizations are performed to control the variance of the vector components, that is, a linear transformation that scales the values such that their variance=1. This method is a convenient use of the Mahalanobis distance measure without changing the distance calculation procedure.

Table 1
Information Obtained from the Survey.

Type	Characteristics
Specifics	Number of Members Ages Activities Economic Incomes
Appliances	Existing Appliances Future Electrical Equipment Hours of Use of Current Equipment Hours of Use of Future Equipment
Increase in the Demand	Number of Members that Increase on Holidays

2.3.2. Clustering module (SOM1)

The primary objective of this module is to obtain the clusters for the different types of homes in the community according to a priori criteria, such as the number of family members as well as their occupations and income and number of household electrical appliances (Data 1). The module also provides information on the number of families in each cluster. The clustering of the families is obtained by an $\mathbf{SOM}_1$ in which the neighboring neurons react more strongly to similar input patterns. Here, the Euclidean distance was selected as a measure of similarity. Home location is not included in the inputs; however, the methodology is flexible and this input could be considered in the further research.

$\mathbf{SOM}_1$ does not require labels for each cluster; however, to reduce the complexity of the visualization process, the data can be labeled with the names of family members to make the results easier to understand.

For the families in communities with and without an energy supply, clusters from $\mathbf{SOM}_1$ are established with similar characteristics to organize families in clusters in accordance with their properties.

2.3.3. Update of the database (SOM2)

This module generates load profiles for the required database. Fig. 3 shows the procedure, which requires another community with a 24-h energy supply. The first step is to install electrical meters in each house of this community. The inputs for $\mathbf{SOM}_2$ are represented as a matrix defined by the electricity measurement vectors. In this study, measurements are taken every 15 min, that is, 96 measurements each day. The obtained measurements (Data 2) are clustered using an $\mathbf{SOM}_2$ to obtain a certain number of clusters with corresponding load profiles. Using this information, the average of the profiles is calculated as a representative load profile. The socio-economic features of the cluster are then assigned for each representative load profile. The features are obtained from surveys of the community (Data 1, refer to Table 1); this information (both the classes and the corresponding profiles) is saved in the database.

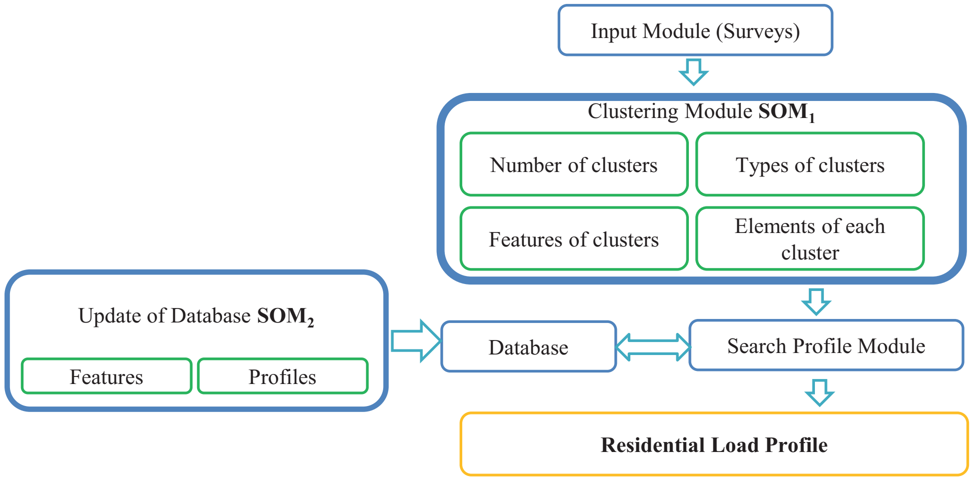

Fig. 2. Method for Residential Load Estimation Based on SOMs.

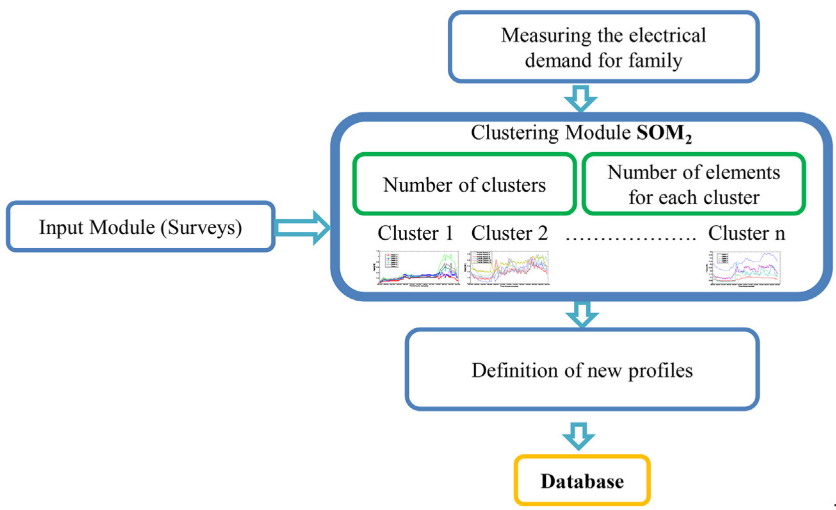

Fig. 3. Update of Database (SOM2).

Thus, load demand information clusters from **SOM**$_2$ are established for families with similar load profiles; this information is employed as a database.

2.3.4. *Database*

All of the profiles that are associated with the characteristics of the families are contained in a database that can accommodate additional family types as they are identified and additional profiles using the **Database Update Module**. The database may be designed to be highly generalized for different types of isolated communities, and it may also include a certain number of pre-defined profiles. The clusters included in the database are as follows: elderly couples, elderly individuals living alone, elderly individuals living with an adult, adults living alone, adult couples, adult couples living with a child, adult couples living with a teenager, adult couples living with two young children, adult couples living with two teenagers, adult couples living with more than three children, and a living arrangement that does not correspond to any of the clusters mentioned.

The database that is generated by the **Database Update Module** is an important element of the proposed method because it attempts to include the largest number of possible patterns.

2.3.5. *Search profile module*

A heuristic selection technique is used for the searching process. The module considers the characteristics of all clusters and then searches for the most similar cluster within the database to select an input, at which point the corresponding load profile is assigned.

2.3.6. *Residential load profile*

The total residential demand d_r is obtained by adding the product of the number of elements in each cluster by the daily load profile assigned to the cluster:

$$d_r = \sum_{c=1}^{C} pc_c \cdot ne_c \tag{3}$$

where c is the cluster, p_c is the profile that is assigned to cluster c, which is obtained by the **Search Profile Module**, and ne_c is the number of elements of cluster c.

The residential demand of community d_C is determined as follows:

$$d_C = d_r \cdot inc_f + d_s + d_{cs} \tag{4}$$

where d_r is the residential daily demand multiplied by a constant inc_f, which represents a multiplier rate of population growth for holidays; d_s represents the characteristic load profile of the schools;

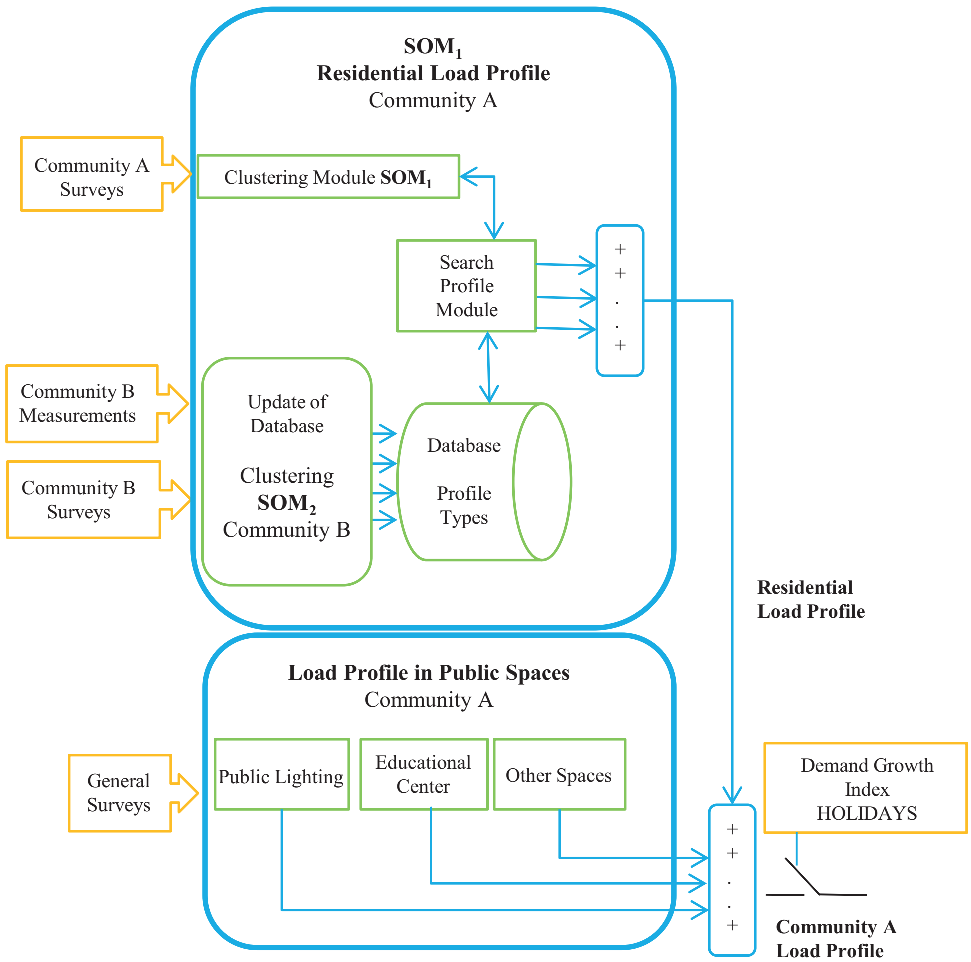

Fig. 4. SOM methodology for Load Profile Generation in an Isolated Community.

and d_{cs} represents the daily demand profile for commonly used centers such as hospitals. In turn, d_s is calculated according to the product of the profile pattern multiplied by the number of students and number of centers.

The daily demand patterns of institutions and common areas, such as schools, medical centers, churches, camping areas, and street lighting, are added to the residential demand profile, and the sum of all of these profiles generates a community profile. The proposed method is flexible and can also incorporate the profiles of centers that are commonly used. Increases in community power consumption that occur on holidays must also be considered and are managed by multiplying the total consumption profile of the community by a factor that represents the increased number of individual consumers on holidays. Note that this generated load profile is vital for sizing the microgrid distributed generation units during the project design stage.

2.4. Load estimation for a community that lacks an energy supply

Fig. 4 shows the application of the proposed methodology to the load estimation of a community A (which lacks an energy supply). This estimation is required for optimal microgrid planning. The methodology requires measurements and surveys from another community B, which shares similar characteristics to Community A but has an available electrical energy 24/7.

The first block of Fig. 4 is used to determine the residential load profile. The results are analyzed by the **Search Profile Module** algorithm to assign characteristic profiles to each cluster, and the residential load profile is then obtained as the sum of the identified load profiles.

The surveys include information that is obtained from families of communities A and B (Table 1). In addition, general surveys that are based on general inquiries and field interviews and that include the characteristics of the community, such as the population and the number of inhabited houses, schools, and communal areas, are conducted in both communities.

The clustering module **SOM**$_1$ of electrical demand uses the survey information of community A as an input and groups the types of families. **SOM**$_2$ is used in the **Update Database** module of community B, and the electricity consumption measurements from the corresponding surveys are clustered.

This algorithm uses a database that can be expanded based on the measurements of the other community (B). The database update clustering **SOM**$_2$ of the electrical demand of community B uses the measurements of the electricity consumption of each house as an input. The clustering is performed for community B, and representative load profiles (as the average of the corresponding electricity load profiles for each cluster) are associated with the socio-economic features based on the surveys of the community. Both the clusters and measurements from community B are stored in the database.

In Fig. 4, the residential load profile is considered together with the profile of load in public spaces such as the load required by the public lighting, educational centers and other public spaces. The proposed method also includes a non-clustered load profile for any family that does not belong to a cluster.

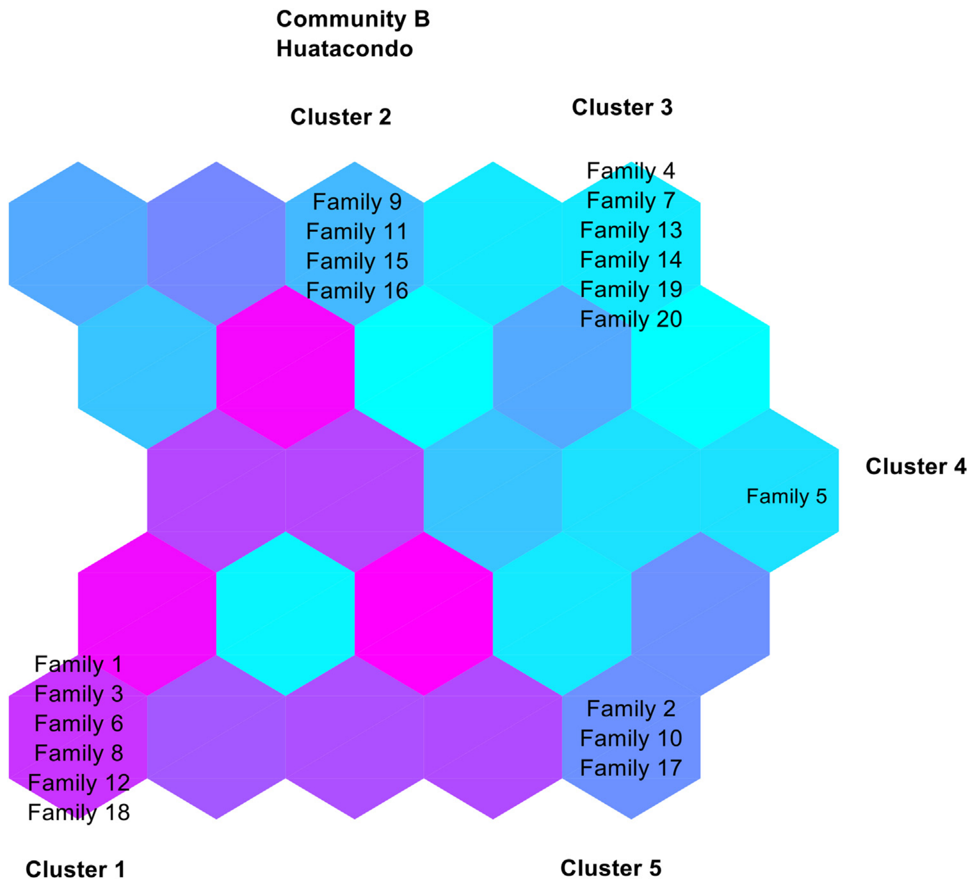

Fig. 5. Huatacondo Consumption Clusters, including Measurements of Consumption for 20 Families.

Finally, the sum of the profiles is completed for the community profile. The output includes a selector with the power to obtain the maximum consumption profile when there are visitors on holidays, which is achieved by multiplying the traditional consumption profile of the community with an index of the growth rate of the population on holidays.

3. Case study

This section shows the results that were obtained when the methodology was applied to the community of El Romeral (community A), where a microgrid will be installed in the near future, using a database that was composed of the measurements and characteristics from another community (Huatacondo; community B), where a microgrid currently operates. The electrically isolated community El Romeral has suitable characteristics for the installation of a microgrid as defined in Ref. [32], where an intelligent self-managed microgrid based on renewable energy, including solar and wind resources are going to be installed. This community is located 21.1 kilometers north of the city of La Serena, Region of Coquimbo, Chile (29° 42′ 53.6″ S, 71° 12′ 59.24″ W) and 70 families living in the area did not have the basic services of electricity supply, potable water, or sewers, which is similar to the current situation. In its first phase, the electrification plan through the microgrid will provide power to approximately half of the population as well as the school, square, neighborhood council, and other common areas. Numerical results of the proposed methodology are presented. A comparison of the k-means and SOM clustering algorithms is also included.

3.1. Load estimation method results

The methodology begins with the database update procedure (Fig. 2). In this work, measurements and surveys of the microgrid that currently operates in Huatacondo (community B) are available and can be used to determine the per-family consumption and its features. The installed microgrid in Huatacondo is composed of two photovoltaic systems ($P^{S}_{max} = 24$ kW), a wind turbine ($\left(P^{W}_{max} = 5\text{ kW}\right)$, the existing diesel generator in the village, which is typical of isolated grids, an energy storage system that is composed of a lead-acid battery bank that is connected to the grid through a bidirectional inverter, a water pump, and the loads ($L_{max} = 28$ kW) [33,34].

The database was updated using the profiles from the Huatacondo measurements by determining all of the measurements per family and subjecting them to the clustering stage (**SOM**$_2$) to obtain profile clusters that were then associated with each family based on the cluster features that were obtained from the surveys to obtain average profiles per cluster. Based on the database update module (see Fig. 3), the measurements of Huatacondo (community B) are gathered and then clustered using clustering module **SOM**$_2$ of the electrical demand. Five clusters of consumer profiles were identified using **SOM**$_2$ (Fig. 5), and the results were studied to establish the coherence. To determine the representative pattern of each cluster, the profiles of each family in every cluster were averaged and then integrated into the database. Fig. 6 shows the profiles of each family that correspond to class 1. This process was used for all of the clusters of identified profiles. Fig. 7 shows the characteristic profiles of each cluster in the database. Note that a profile of the families that could not be surveyed is required and corresponds to

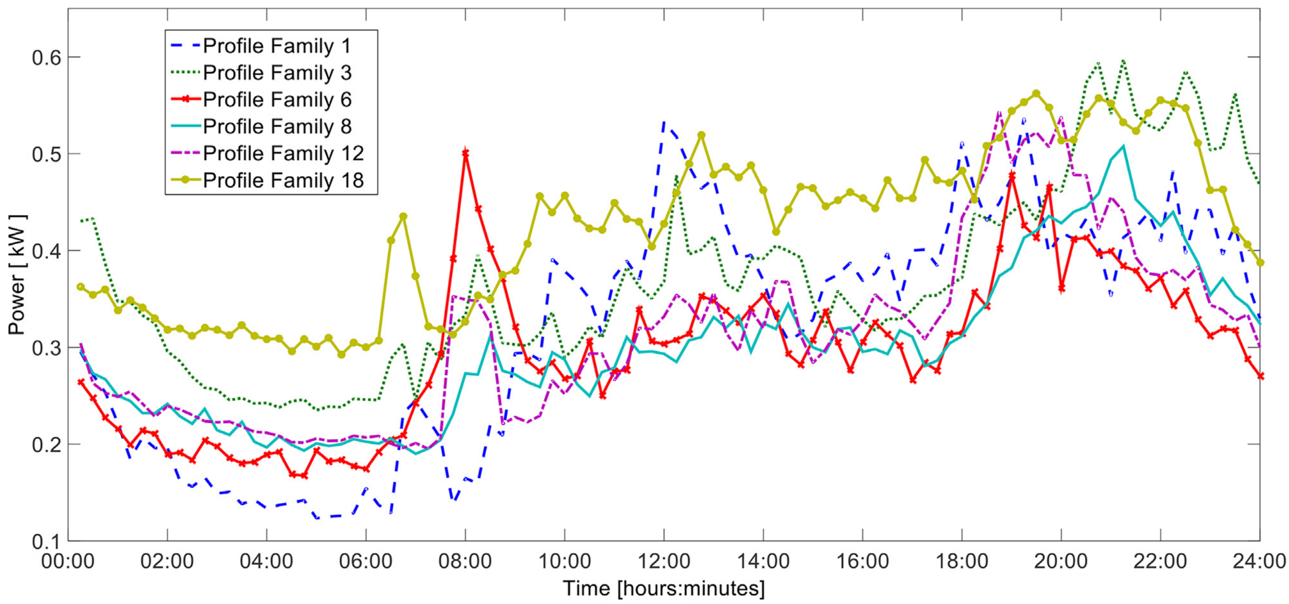

Fig. 6. Load Profiles of Cluster 1.

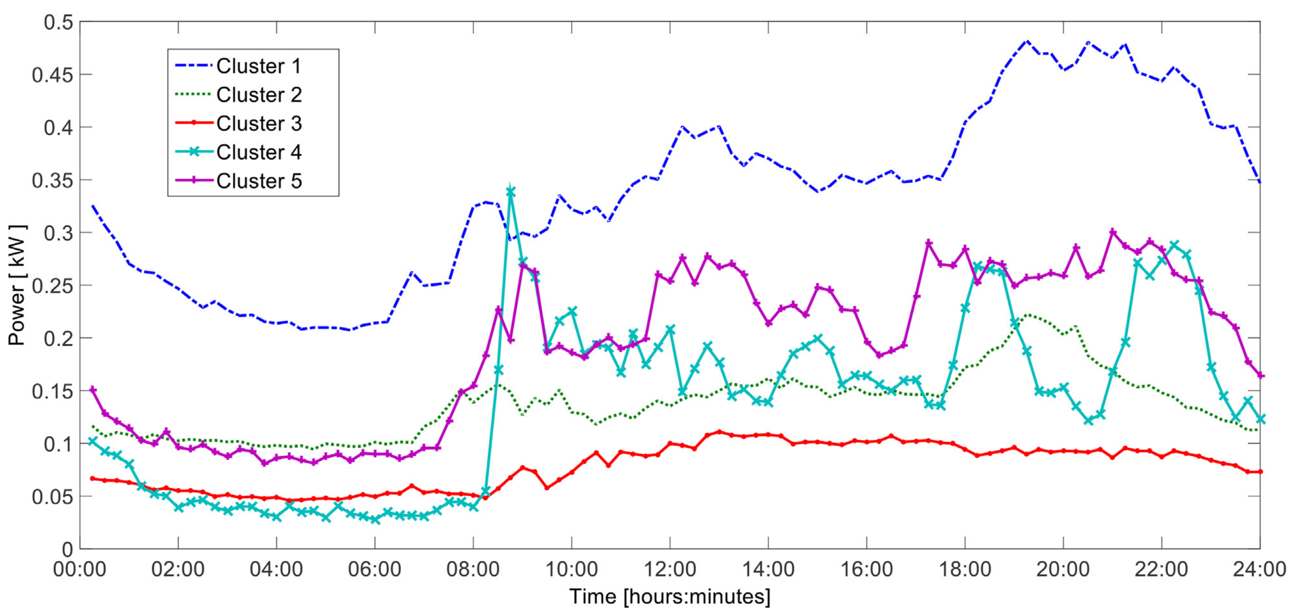

Fig. 7. Electricity Demand Profiles for Each Type of Family assigned in the Database.

the profile with the highest peak demand in the database, which is class 1 in this case (Fig. 7).

The profiles are linked to determine the socio-economic characteristics of the families that are obtained by surveys of the houses in which the meters were installed.

After the database was updated with the new profiles that were obtained from Huatacondo, the load estimation that is based on SOM_1 is applied to El Romeral (community A). In the *Input Module* (Fig. 2), 150 inhabitants from 70 families were identified, and 39 of these families (80 inhabitants) received benefits from the microgrid. Of these 39 families, 17 have been surveyed to generate electricity consumption profiles. In addition, the community currently does not have streetlights; therefore, based on the geographical distribution of the houses, usage of 15 lights has been estimated. In the *Clustering Module* of electricity users that is based on SOM_1 for El Romeral, eight types of families (Fig. 8) were identified, including retired elderly individuals who live alone, retired elderly couples, families with six members, families with three adults, and adult couples.

Because there are eight clusters, eight profile patterns have been identified by the Search Profile Module in the database. The clusters 3, 4, 6, 7 and 8 generated for El Romeral community (see Fig. 8) are similar to the clusters 1–5 of the Huatacondo community (see Fig. 5), respectively. Therefore, the corresponding load profiles for the El Romeral community are assigned as the load profiles of the Huatacondo families. The load profiles for the other classes of the El Romeral community (clusters 1, 2 and 5 of Fig. 8), which are not represented in the database, are associated with the load profile of class 1 of the Huatacondo database.

These load patterns have been multiplied by the number of families corresponding to each cluster to generate the *Residential Load Profile*. After adding this profile to the other profiles of communal

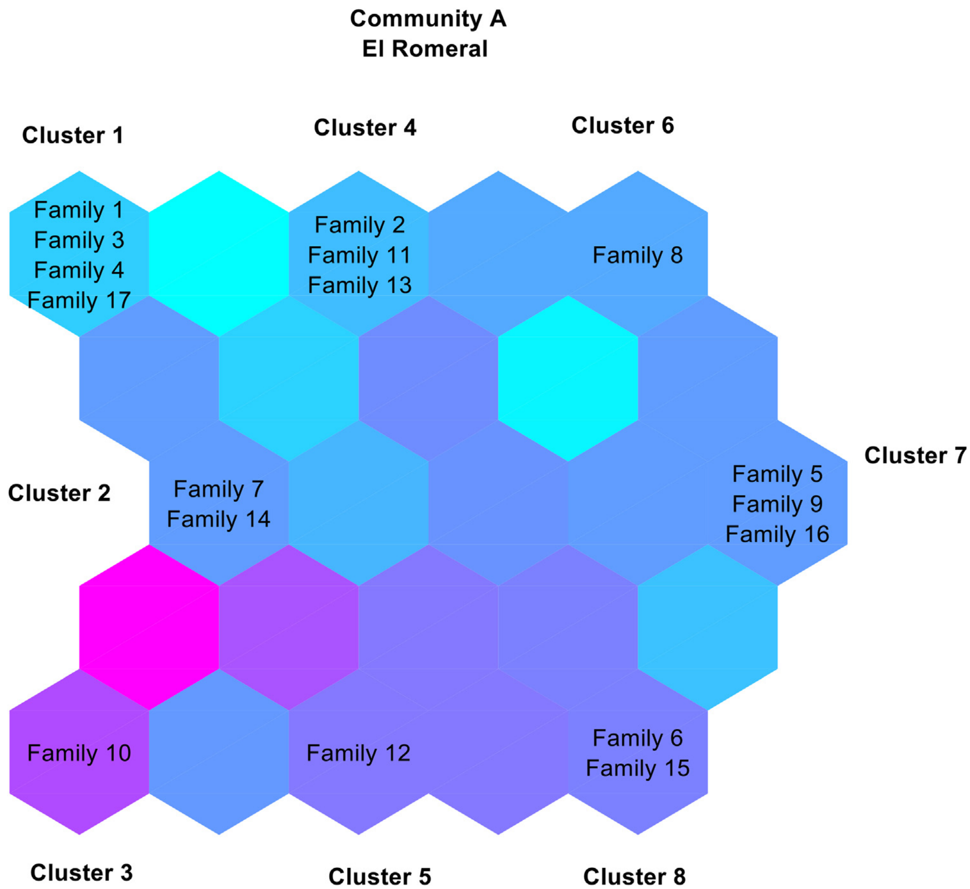

Fig. 8. Clustering of Families Types in the El Romeral.

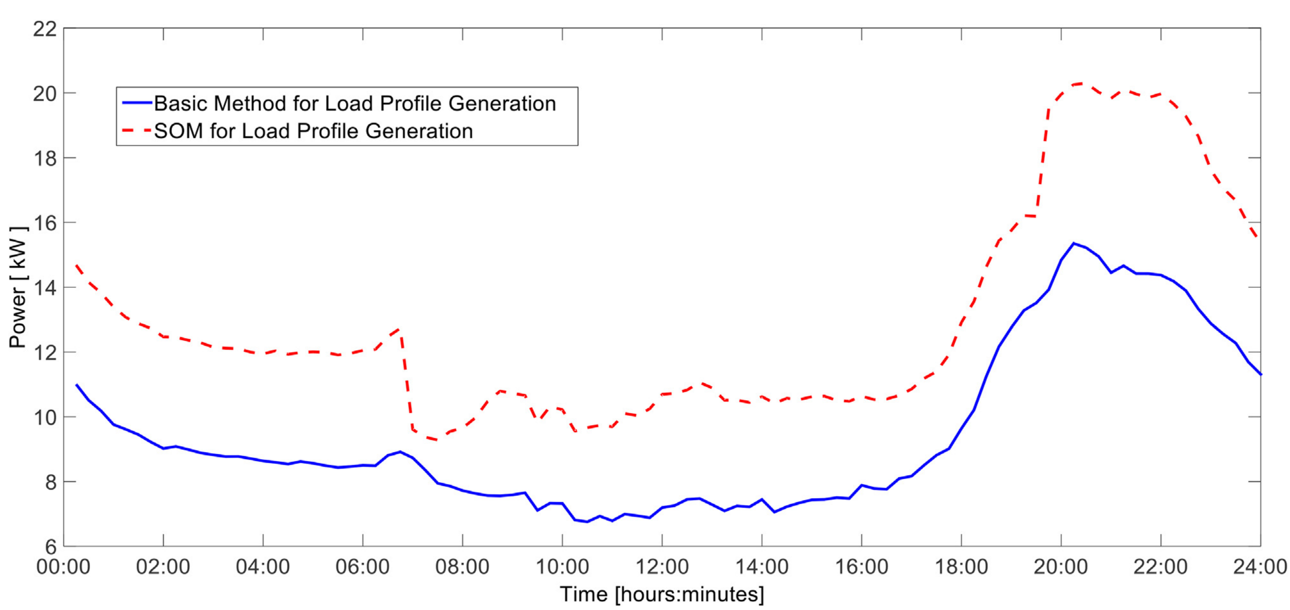

Fig. 9. Load Profile Generated by the SOM Method versus the Basic Method in the El Romeral Community.

locations, the Community Load Profile is obtained. Fig. 9 shows the Community Load Profile of El Romeral generated by applying this methodology; however, it has not been validated with actual data because the microgrid has not been implemented. Nevertheless, it is used to size the generation units. However, this methodology can be verified with a basic technique that consists of obtaining the average consumption of Huatacondo and scaling that consumption based on the number of inhabitants to be supplied with energy in El Romeral. This technique does not consider the increased demand on weekends or holidays, which is typical of isolated communities.

Fig. 9 compares the load profiles of El Romeral that were generated by this methodology and a benchmark method. The benchmark method scales the average consumption of Huatacondo by the number of inhabitants that will be supplied by the planned

Table 2
Clustering performance for Update of Database using K-means and SOM_2, Huatacondo Community.

	Cluster error k-means [W]	Cluster error SOM_2 [W]	Elements per cluster k-means	Elements per cluster SOM_2
Cluster 1	4.782	6.161	3	6
Cluster 2	6.159	2.079	6	4
Cluster 3	3.499	3.013	2	6
Cluster 4	4.472	0	3	1
Cluster 5	2.314	4.783	6	3

microgrid in El Romeral. Note that the load profile that is based on the new methodology has a higher consumption than that from the benchmark method because the new method considers uninhabited houses (without surveys) that will be inhabited as well as socio-economic aspects. Thus, the new method can be applied to any community that does not have an electricity supply and where measurements are not available.

A greater influx of people occurs in these communities during holidays; therefore, this work includes an index of population increases for festival dates that scales the generating units according to the maximum community consumption where the microgrid will be installed (Fig. 4). In Fig. 10, the profiles generated in El Romeral for holidays between the two methods are compared.

An analysis of the results presented in Figs. 9 and 10 show that it is necessary to consider holidays to ensure sufficient power and energy that can manage an occasional influx of visitors.

3.2. Comparison between SOM and K-means

As explained in the Section 2, two stages of clustering based on SOM are proposed in the methodology. However, there are other clustering algorithms in the literature that can be used at these stages. In this section, a comparison with a well-known clustering method, k-means, is considered for both stages instead of SOM. Other clustering algorithm can be easily tested as well.

To compare the algorithms, a performance index e_k is defined for each cluster k as follows:

$$e_k = \frac{1}{\Omega_H}\sqrt{\sum_{i \in H}\left(x_{i,k} - x_{c,k}\right)^2} \tag{5}$$

where x_i is the element of the cluster k, x_c is the centroid of the cluster k, and Ω_H is the number of elements of the set H containing the entire data set.

In Table 2, the performance index e_k is shown for both k-means- and $\mathbf{SOM_2}$-based methodologies at the stage of update of the database using measurements from the Huatacondo village (community B). It can be observed that the performance of the methodologies are similar, with $\mathbf{SOM_2}$ providing a slightly better result. The $\mathbf{SOM_2}$-based methodology for cluster 4 has e_k equal to 0 because the cluster contains only a single element. In Table 3, the elements for each cluster are shown. The elements of cluster 1 by the k-means algorithm are the same as those of cluster 5 for SOM (Families 2, 10, and 17). Thus, their performance index and number of elements are the same (Table 2). SOM_2 grouped cluster 2 and 4 of the k-means output in cluster 1. Fig. 6 shows 6 profiles (Families 1, 3, 6, 8, 12, and 18) that correspond to cluster 1, as output by $\mathbf{SOM_2}$. Otherwise, using k-means, the same 6 families are clustered in two clusters: cluster 2 (Fig. 11) and cluster 4 (Fig. 12). By analyzing the socio-economic characteristics of these 6 families belonging to cluster 1 of $\mathbf{SOM_2}$, their features are similar, mainly considering their economical incomes and daily activities. Thus, $\mathbf{SOM_2}$ manages to cluster the families in a more coherent way; the use of the 6 elements in a single class is preferred for describing the data.

Table 3
Members of Clusters for Update of Database Classifier K-means SOM_2, Huatacondo Community.

	Members k-means	Members SOM_2
Cluster 1	Family 2, Family 10, Family 17	Family 1, Family 3, Family 6 Family 8, Family 12, Family 18
Cluster 2	Family 1, Family 18	Family 9, Family 11, Family 15 Family 16
Cluster 3	Family 4, Family 5, Family 9 Family 11, Family 15, Family 16	Family 4, Family 7, Family 13 Family 14, Family 19, Family 20
Cluster 4	Family 3, Family 6, Family 8 Family 12	Family 5
Cluster 5	Family 7, Family 13, Family 14 Family 19, Family 20	Family 2, Family 10, Family 17

Table 3 shows the members of clusters obtained for the update of the database using k-means and $\mathbf{SOM_2}$ applied to the Huatacondo measurements. From the table, using $\mathbf{SOM_2}$, cluster 4 contains only family 5, while cluster 3 obtained by k-means contains families 4, 5, 9, 11, 15 and 16. In Fig. 13, the families of cluster 3 obtained by k-means are presented; note that family 5 shows a different profile from the others mainly during the periods between 8:00 to 10:00 and 18:00 to 23:00. This finding validates the $\mathbf{SOM_2}$ method, under which family 5 is assigned as the only element of cluster 4.

Table 4 shows the performance at the clustering stage for El Romeral (community A) using both clustering methods, $\mathbf{SOM_1}$ and k-means, using the surveys for clustering the families. It can be observed that the performances of the methodologies are similar. In Table 5, the elements for each cluster are shown. Three clusters coincide; family 12 is obtained in both cluster 4 with k-means and cluster 5 with $\mathbf{SOM_1}$; families 1, 3, 4 and 17 with cluster 5 k-means and cluster 1 $\mathbf{SOM_1}$; and family 8 with cluster 8 k-means and cluster 6 $\mathbf{SOM_1}$. Thus, their performance indexes are identical, as shown in Table 5.

From Table 5 using $\mathbf{SOM_1}$, family 10 is the only element of cluster 3, while the k-means approach places this family in cluster 2 with families 6, 7, 14 and 15. By analyzing the surveys, it can be

Table 4
Clustering Performance for K-means and SOM_1, El Romeral Community.

	Cluster error k-means [W]	Cluster error SOM_1 [W]	Elements per cluster k-means	Elements per cluster SOM_1
Cluster 1	0.097	0.053	2	4
Cluster 2	0.219	0.173	5	2
Cluster 3	0.093	0	2	1
Cluster 4	0	0.149	1	3
Cluster 5	0.053	0	4	1
Cluster 6	0	0	1	1
Cluster 7	0	0.281	1	3
Cluster 8	0	0.108	1	2

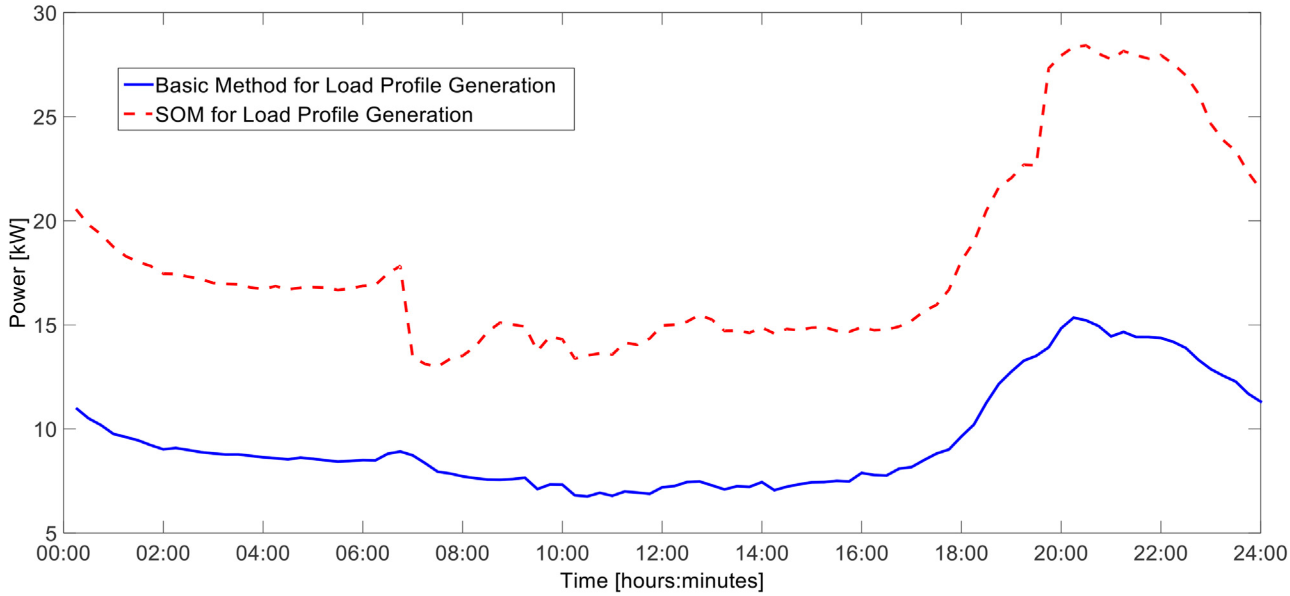

Fig. 10. Load Profile Generated by the SOM Method versus the Basic Method in the El Romeral Community for a Holiday Season.

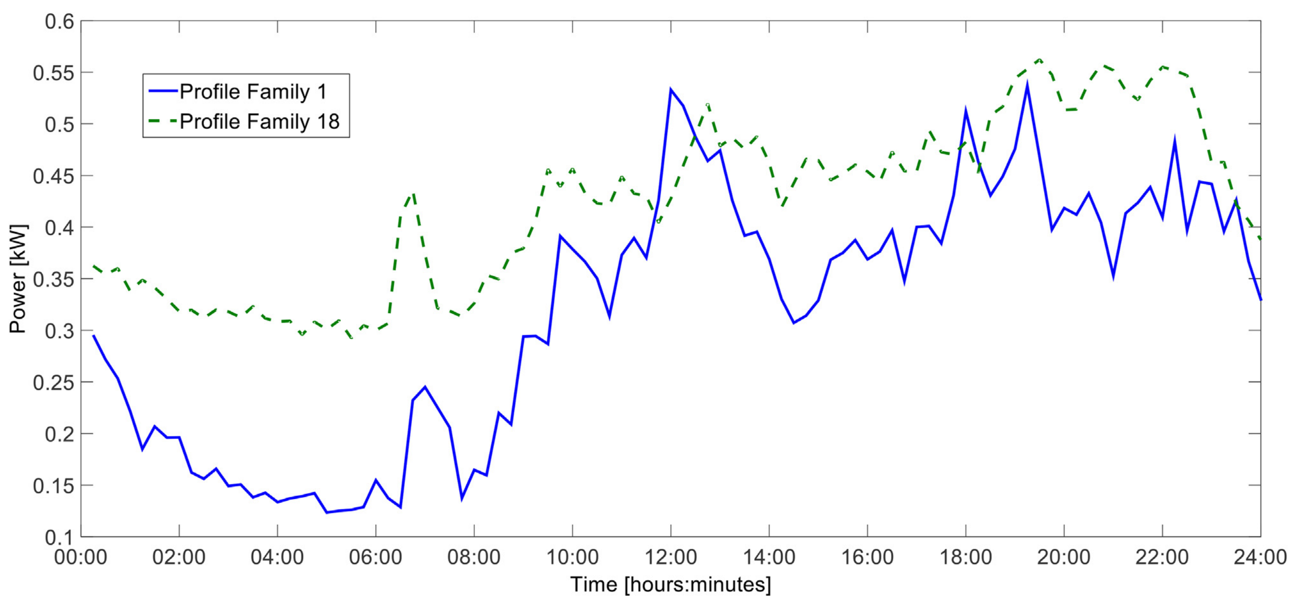

Fig. 11. Load Profiles of Cluster 2 using K-means.

Table 5
Members of Clusters for K-means and SOM_1, El Romeral Community.

	Members k-means	Members SOM_1
Cluster 1	Family 9, Family 11	Family 1, Family 3, Family 4 Family 17
Cluster 2	Family 6, Family 7, Family 10 Family 14, Family 15	Family 7, Family 14
Cluster 3	Family 2, Family 13	Family 10
Cluster 4	Family 12	Family 2, Family 11, Family 13
Cluster 5	Family 1, Family 3, Family 4 Family 17	Family 12
Cluster 6	Family 5	Family 8
Cluster 7	Family 16	Family 5, Family 9, Family 16
Cluster 8	Family 8	Family 6, Family 15

validated that family 10 is properly clustered with $\mathbf{SOM}_1$ because it is composed of 6 members, while the others families (considered by k-means) have 3 members each. In another example, cluster 7 obtained by $\mathbf{SOM}_1$ is composed of families 5, 9 and 16, while k-means places family 5 in cluster 6 and family 16 in cluster 7. By analyzing the socio-economic characteristics of these three families, their activities, ages and incomes are very similar; thus, $\mathbf{SOM}_1$ is validated here as well.

4. Conclusions

This paper reports on the use of computational intelligence techniques for the planning of microgrids in small and isolated communities that have not been measured for their electricity consumption.

The main contribution is the proposed methodology based on clustering algorithms that utilize information about similar com-

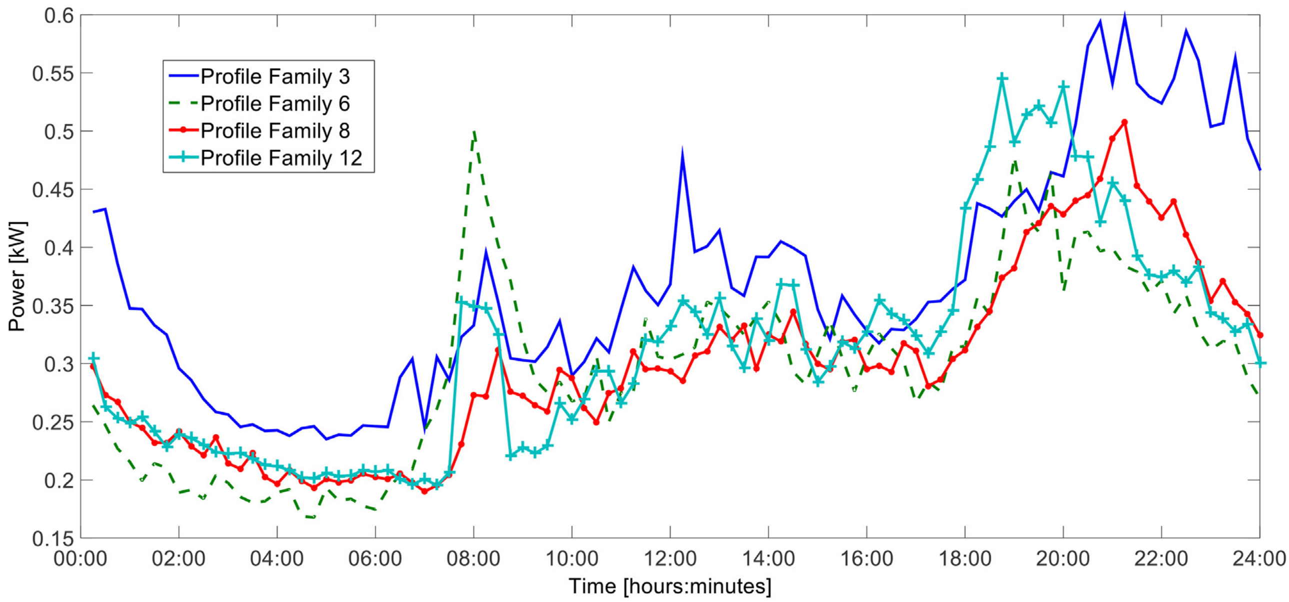

Fig. 12. Load Profiles of Cluster 4 using K-means.

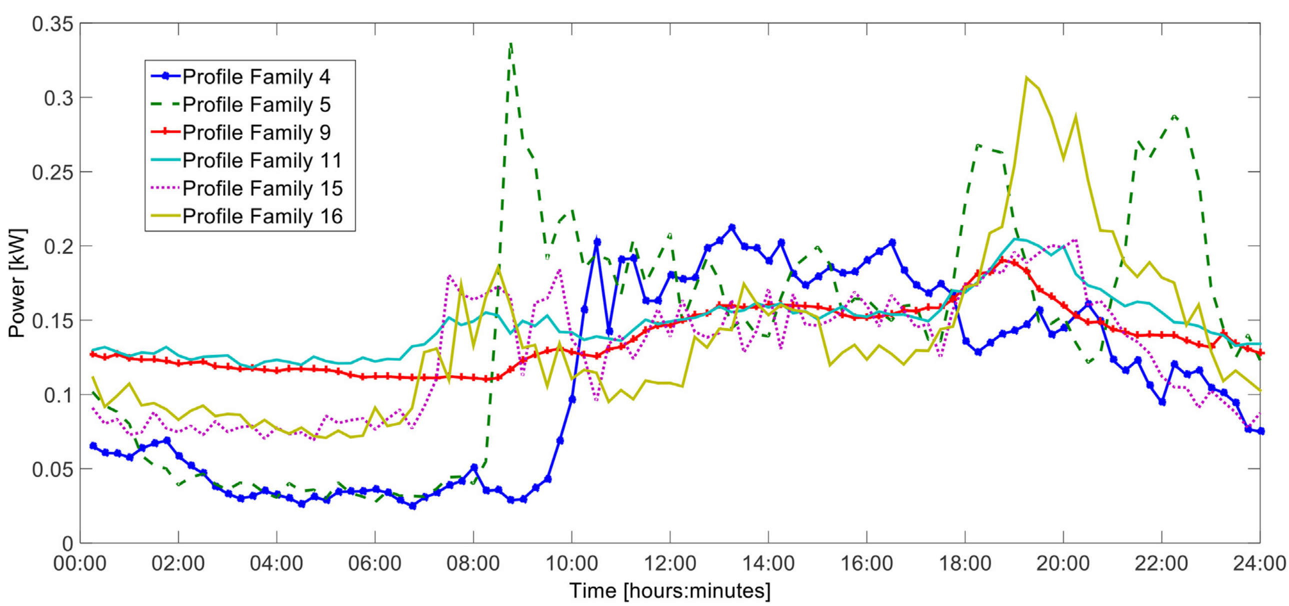

Fig. 13. Load Profiles of Cluster 3 using K-means.

munities that have a permanent electricity supply to estimate the future load profiles of families without a current permanent supply. The methodology employs two clustering steps (for communities A and B) and two clustering algorithms. In this case, SOM and k-means are employed; the latter is primarily employed for comparison. An SOM is recommended because it is especially suitable for the survey data required in this methodology due to its prominent visualization properties. An SOM enables the automatic presentation of a map in which an intuitive description of the similarities among the data can be observed and the distance between two neighborhoods can be calculated. Unlike K-means, the cluster number does not have to be defined for an SOM; in an SOM, the prototypes that do not represent the number of clusters are defined. However, for k-means, a sensitivity step can be subsequently performed to determine the appropriate cluster number, which requires greater computational effort.

The proposed methodology includes real measurements that were collected from a community with an operating microgrid in Huatacondo village. The method was applied to the community of El Romeral, where the inhabitants are not currently supplied with electricity. The estimated profiles were used in the planning of a microgrid that is in the design stages; the results can be validated with actual data when the grid becomes operational. Furthermore, the SOM-proposed methodology was compared with a k-means algorithm and delivered more favorable and consistent results according to social aspects of the community.

The proposed methodology based on SOMs is a viable solution for generating load profiles and sizing generating units for microgrids in isolated communities that have either a partial or no power supply. This tool provided information from surveys, and current efforts are focused on enhancing the database with an increased quantity of actual measurements.

Acknowledgements

This study was partially supported by the Solar Energy Research Center SERC-Chile (CONICYT/FONDAP/Project under Grant 15110019), the Complex Engineering Systems Institute, ISCI (ICM-FIC: P05-004-F, CONICYT: FB0816)", FONDECYT Chile Grant No. 1140775 "Design of Robust Predictive Control Strategies for the Operation of Microgrids with High Penetration of Renewable Energy" and FONDEF ID14I10063 project "Design and Implementation of an Experimental Prototype of Microgrid for Mapuche Communities".

References

[1] G. Strbac, Demand side management: benefits and challenges, Energy Policy 36 (December (12)) (2008) 4419–4426.
[2] Y.-I. Kim, J.-M. Ko, S.-H. Choi, Methods for generating TLPs (typical load profiles) for smart grid-based energy programs, 2011 IEEE Symposium on Computational Intelligence Applications In Smart Grid (CIASG) (2011) 1–6.
[3] B. Stephen, A.J. Mutanen, S. Galloway, G. Burt, P. Jarventausta, Enhanced load profiling for residential network customers, Power Deliv. IEEE Trans. 29 (February (1)) (2014) 88–96.
[4] B. Stephen, F.R. Isleifsson, S. Galloway, G.M. Burt, H.W. Bindner, Online AMR domestic load profile characteristic change monitor to support ancillary demand services, Smart Grid IEEE Trans. 5 (2) (2014) 888–895.
[5] M.M. Armstrong, M.C. Swinton, H. Ribberink, I. Beausoleil-Morrison, J. Millette, Synthetically derived profiles for representing occupant-driven electric loads in Canadian housing, J. Build. Perform. Simul. 2 (1) (2009) 15–30.
[6] J.K. Gruber, M. Prodanovic, Residential energy load profile generation using a probabilistic approach, 2012 Sixth UKSim/AMSS European Symposium on Computer Modeling and Simulation (EMS) (2012) 317–322.
[7] J. Dickert, P. Schegner, A time series probabilistic synthetic load curve model for residential customers, IEEE Trondheim, PowerTech (2011) 1–6.
[8] A.M. Ihbal, H.S. Rajamani, R.A. Abd-Alhameed, M.K. Jalboub, Statistical predictions of electric load profiles in the UK domestic buildings, 1st International Conference on Energy, Power and Control (EPC-IQ) (2010) 345–350.
[9] N. Pflugradt, J. Teuscher, B. Platzer, W. Schufft, Analysing low-voltage grids using a behaviour based load profile generator, International Conference on Renewable Energies and Power Quality vol. 11 (2013) p. 5.
[10] M.M. Huamani, A.F. Orlando, Methodology for generating thermal and electric load profiles for designing a cogeneration system, Energy Build. 39 (9) (2007) 1003–1010.
[11] H. Mori, T. Itagaki, A precondition technique with reconstruction of data similarity based classification for short-term load forecasting, IEEE, in: Power Engineering Society General Meeting, vol. 1, 2004, pp. 280–285.
[12] S. Valero, J. Aparicio, C. Senabre, M. Ortiz, J. Sancho, A. Gabaldon, Comparative analysis of self organizing maps vs. multilayer perceptron neural networks for short-term load forecasting, Proceedings of the International Symposium Modern Electric Power Systems (MEPS) (2010) 1–5.
[13] S. Valero, J. Aparicio, C. Senabre, M. Ortiz, J. Sancho, A. Gabaldon, Analysis of different testing parameters in self-organizing maps for short-term load demand forecasting in Spain, Proceedings of the International Symposium Modern Electric Power Systems (MEPS) (2010) 1–6.
[14] I.B. Sanchez, I.D. Espinos, L. Moreno Sarrion, A. Quijano Lopez, I.N. Burgos, Clients segmentation according to their domestic energy consumption by the use of self-organizing maps, EEM 2009. 6th International Conference on the European Energy Market (2009) 1–6.
[15] S.V. Verdu, M.O. Garcia, C. Senabre, A.G. Marin, F.J.G. Franco, Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps, Power Syst. IEEE Trans. 21 (November (4)) (2006) 1672–1682.
[16] O.A. Carpinteiro, A.J. Reis, A.P. da Silva, A hierarchical neural model in short-term load forecasting, Appl. Soft Comput. 4 (4) (2004) 405–412.
[17] J. Nagi, K.S. Yap, F. Nagi, S.K. Tiong, S.K. Ahmed, A computational intelligence scheme for the prediction of the daily peak load, Appl. Soft Comput. 11 (8) (2011) 4773–4788.
[18] D. Niu, D. Liu, D.D. Wu, A soft computing system for day-ahead electricity price forecasting, Appl. Soft Comput. 10 (3) (2010) 868–875.
[19] F. McLoughlin, A. Duffy, M. Conlon, A clustering approach to domestic electricity load profile characterisation using smart metering data, Appl. Energy 141 (March) (2015) 190–199.
[20] J. Llanos, D. Sáez, R. Palma-Behnke, A. Núñez, G. Jiménez-Estévez, Load profile generator and load forecasting for a renewable based microgrid using self organizing maps and neural networks, IEEE World Congress on Computational Intelligence (2012) p. 8.
[21] T. Warren Liao, Clustering of time series data—a survey, Pattern Recognit. 38 (11) (2005) 1857–1874.
[22] D.D. Sharma, S.N. Singh, Electrical load profile analysis and peak load assessment using clustering technique, in: Proc. IEEE PES General Meeting, 27–31 July, Washington, USA, 2014.
[23] R.C. Duan, F.H. Wang, J. Zhang, R.H. Huang, X. Zhang, Data mining & pattern recognition of voltage sag based on K-means clustering algorithm, in: IEEE Power & Energy Society General Meeting, Denver, CO, 2015, 2015, pp. 1–5, http://dx.doi.org/10.1109/PESGM.2015.7286079.
[24] Centro de Energía FCFM, Planta fotovoltaica en escuela El Romeral, 2012 [Online]. Available: http://www.centroenergia.cl/ce-fcfc/?page_id=1922.
[25] D.E. Olivares, A. Mehrizi-Sani, A.H. Etemadi, C. a. Cañizares, R. Iravani, M. Kazerani, A.H. Hajimiragha, O. Gomis-Bellmunt, M. Saeedifard, R. Palma-Behnke, G. a. Jiménez-Estévez, N.D. Hatziargyriou, Trends in microgrid control, IEEE Trans. Smart Grid 5 (4) (2014) 1905–1919.
[26] T. Kohonen, The self-organizing map, Proc. IEEE 78 (September (9)) (1990) 1464–1480.
[27] T. Kohonen, S. Kaski, K. Lagus, J. Salojarvi, J. Honkela, V. Paatero, A. Saarela, Self organization of a massive document collection, Neural Netw. IEEE Trans. 11 (3) (2000) 574–585.
[28] G. Dreyfus, Neural Networks: Methodology and Applications, Springer, 2005.
[29] J. Kangas, T. Kohonen, Developments and applications of the self-organizing map and related algorithms, Math. Comput. Simul. 41 (1–2) (1996) 3–12.
[30] J.C. Patra, E.L. Ang, P.K. Meher, Q. Zhen, A new SOM-based visualization technique for DNA microarray data, International Joint Conference on Neural Networks, IJCNN '06 (2006) 4429–4434.
[31] C. Alvial-Palavicino, N. Garrido-Echeverría, G. Jiménez-Estévez, L. Reyes, R. Palma-Behnke, A methodology for community engagement in the introduction of renewable based smart microgrid, Energy Sustain. Dev. 15 (3) (2011) 314–323.
[32] K. Ubilla, G.A. Jimenez-Estevez, R. Hernadez, L. Reyes-Chamorro, C. Hernandez Irigoyen, B. Severino, R. Palma-Behnke, Smart microgrids as a solution for rural electrification: ensuring long-term sustainability through cadastre and business models, Sustain. Energy IEEE Trans. 5 (4) (2014) 1310–1318.
[33] R. Palma-Behnke, C. Benavides, E. Aranda, J. Llanos, D. Sáez, Energy management system for a renewable based microgrid with a demand side management mechanism, IEEE SSCI 2011 – Symp. Ser. Comput. Intell. – CIASG 2011 IEEE Symp. Comput. Intell. Appl. Smart Grid (2011) 131–138.
[34] R. Palma-Behnke, D. Ortiz, L. Reyes, G. Jimenez-Estevez, N. Garrido, G. Jiménez-Estévez, N. Garrido, A social SCADA approach for a renewable based microgrid – the Huatacondo project, IEEE Power Energy Soc. Gen. Meet. (2011) 1–7.