Expert Systems With Applications 235 (2024) 121179

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Microgrid planning based on computational intelligence methods for rural communities: A case study in the José Painecura Mapuche community, Chile

Raúl Morales [a], Luis G. Marín [b], Tomislav Roje [a], Víctor Caquilpan [c], Doris Sáez [a,e], Alfredo Nuñez [d,*]

[a] Department of Electrical Engineering, University of Chile, Santiago 8370451, Chile
[b] Department of Electronics Engineering, Pontificia Universidad Javeriana, Bogotá 110321, Colombia
[c] School of Computer Science, The University of Adelaide, Adelaide 5005, Australia
[d] Section of Railway Engineering, Department of Engineering Structures, Delft University of Technology, Delft 2628CN, The Netherlands
[e] Instituto Sistemas Complejos de Ingeniería, Santiago 8370451, Chile

ARTICLE INFO

Keywords:
Microgrids
Planning
Self-organizing map (SOM)
Fuzzy c-means
Markov chains
Prediction intervals

ABSTRACT

Microgrids (MGs) are sustainable solutions for rural zone electrification that use local renewable resources. However, only careful planning at the start of an MG project can ensure its future optimal operation. In this paper, a novel methodology for MG planning by using the uncertainty characterization of renewable resources and demand is presented. Additionally, a model of electricity consumption is proposed and applied in an isolated rural community. In such communities, consumption patterns typically need to be derived as model inputs because consumption measurements are not available for the planning stage. To obtain these inputs, clustering algorithms based on self-organizing maps (SOMs) and fuzzy c-means are used to classify the families of the community given sociodemographic information obtained via surveys. Subsequently, Markov chains (MCs) are employed to generate consumption patterns based on consumption measurements in some dwellings and surveys applied to the community. The nonlinearities and uncertainties associated with renewable resources and consumption are modeled by using prediction interval (PI) models. These PI models provide the required consumption and generation scenarios for deriving the optimal sizing and topological information to address the MG planning problem. The results of the robust planning approach based on scenarios are useful at the feasibility and design phases of an MG project. The proposed methodology is successfully applied to MG planning for a rural Mapuche community, where a conservative criterion was considered to minimize the investment risk. This criterion corresponds to the worst-case scenario in which the demand increases by 19.9% compared to that of the baseline scenario and a lower energy cost is obtained. However, the net present cost and operational costs increase by 14% and 11.75% compared to those of the baseline scenario, respectively.

1. Introduction

1.1. Motivation

Electrification in rural zones influences local development by reducing rural-to-urban migration and by benefiting productive development, access to essential services such as health care and education, and the eradication of poverty (Ferrer-Martí et al., 2012; Niez, 2010; Pereira et al., 2010). Using nonconventional renewable energies (NCRE) as energy sources has a low environmental impact and supports the sustainability of rural areas (Ferrer-Martí et al., 2012; Leary et al., 2012). NCRE sources also have lower operation and maintenance costs than those of traditional electrification projects, making them an attractive solution for isolated rural communities (Ubilla et al., 2014). Furthermore, microgrids (MGs) are low voltage controllable systems that integrate NCRE resources with small, distributed energy resources and manage generation, loads and energy storage (Lasseter, 2002). Although there is vast international experience in installing MGs in urban and rural areas (Del Carpio Huayllas et al., 2010; Vallvé, 2010), ensuring optimal operation during the whole life cycle of MGs remains a challenge.

One aspect of that challenge is the planning process. Planning is a relevant part of the design of an MG solution. A complete analysis of

* Corresponding author.
E-mail addresses: ramorale@ing.uchile.cl (R. Morales), luis_marinc@javeriana.edu.co (L.G. Marín), troje@ug.uchile.cl (T. Roje), victor.caquilpanparra@adelaide.edu.au (V. Caquilpan), dsaez@ing.uchile.cl (D. Sáez), a.a.nunezvicencio@tudelft.nl (A. Nuñez).

https://doi.org/10.1016/j.eswa.2023.121179
Received 13 April 2023; Received in revised form 20 July 2023; Accepted 9 August 2023
Available online 12 August 2023

recent and future consumption is needed, as well as an evaluation of various options for energy generation, where the costs should be widely considered. Thus, an efficient planning model is needed. For MG design, a deterministic, stochastic or robust optimization approach can be taken. The deterministic approach considers the average dynamics or a single realization of the stochastic processes. However, performance is guaranteed only at the operation point with small variations, making the MG design quite conservative and not optimal for a broad range of instances. The stochastic and robust approaches include uncertainty in the optimization; this guarantees good performance under a broader set of instances. In this paper, a robust planning approach based on scenarios is considered. The scenarios are generated from the uncertainty characterization based on fuzzy and neural prediction intervals of consumption profiles and renewable resources that are required for MG planning. Thus, this paper contributes to improving the feasibility and robustness of MG planning based on scenarios.

1.2. Literature review

In that context, the main task of MG planning is to determine the optimal size of generation resources with the objective of meeting the demand, of minimizing the cost, of increasing the reliability, the efficiency, the autonomy and of reducing greenhouse gases, among other objectives (Mumtaz & Bayram, 2017). Note that diverse planning methods, such as optimizing the sizing of energy storage systems to minimize maintenance and operation costs, have been applied, proving the effect of variations in fuel prices and aligning the use of different energy sources based on available resources, as well as the cost of energy that comes from the electrical grid (Mohammadi et al., 2012). In Yon et al. (2021), an MG planning methodology is presented, which considers the distributed grid design, optimal location and sizing of the PV system and battery storage for rural communities. Li et al. (2020) presents a study to demonstrate the techno-economic feasibility of an off-grid hybrid renewable energy system for a rural village in West China. An analysis of combinations of PV panels, wind turbines and biogas is modeled and optimized using Homer. The selected configuration was determined while ensuring a reliable power supply to the demand of the village, including the productive demands. The authors in Kharrich et al. (2018) presented the planning of an isolated microgrid based on PV generation, wind turbine, diesel generator, and a battery bank. To obtain the size of the generators that could compose the microgrid, two objectives were proposed: net present cost (NPC) and emission reduction benefit cost (ERBC). Multiobjective particle swarm optimization was used to size the microgrid, and the optimal solution of the Pareto set at the knee point was chosen.

The above studies are MG planning methodologies based on a deterministic approach to optimize the size and location of resources available in various study cases. The size of the generators and energy storage system of the microgrids are obtained to achieve the minimum price with the highest reliability and lowest environmental emission, among other proposed objectives. However, in those studies, the uncertainty in the demand and renewable resources is not considered in the planning process.

On the other hand, Wang et al. (2017) explore the impact of MG expansion to meet increasing load demand over time. Thus, expansion planning should be considered when designing an MG project to simultaneously increase the economic benefits and to improve the reliability of MGs. Moreover, technical features should consider integrating potential new units of generation in the future if needed. In this way, various authors have created planning models considering the uncertainties induced by both intermittent distributed generators and demand Mumtaz and Bayram (2017). For instance, the authors in Khodaei et al. (2015) presented a robust planning approach considering uncertainty in load, renewable generation, and market prices. The proposed approach was divided into an investment master problem and an operation subproblem. The optimal planning obtained by the master problem is used by the operating subproblem to determine the optimal operation under the worst-case scenario. In Borghei and Ghassemi (2021), the planning of microgrids to maximize the resiliency of distribution networks was presented, and the uncertainty of the energy resources was addressed by robust optimization based on the worst-case scenario. The performance of the proposed approach was evaluated by the IEEE 37- and IEEE 123-bus test systems under several severe fault scenarios. Chen et al. (2022) presented a planning model for a grid-connected microgrid accounting for renewable generation uncertainties. The scenarios are obtained based on the deep convolutional generative adversarial network and improved k-medoids clustering algorithm. The size of the renewable resources of the microgrid in the planning process is obtained by maximizing renewable energy utilization efficiency and minimizing the economic cost and emissions. Furthermore, a robust model under budgeted uncertainty provides a pool of MG solutions that depend on load and generation profiles (Levorato et al., 2022). Regarding the use of stochastic optimization, Khayatian et al. (2018) proposed a two-stage stochastic planning optimization to consider the microgrid expansion problem. The proposed approach aids microgrid companies in deciding whether they should invest in microgrid installation. The objective of stochastic programming is to maximize the expected revenue while ensuring the cost-effectiveness and reliability of the power system under uncertainty of load growth and variability of renewable resources.

For MG planning, understanding electrical consumption is crucial, especially when accurate data are not available. Overall, rural communities do not have consumption measurements in all dwellings or do not have them at all; therefore, estimating load profiles becomes essential in designing an MG system. For instance, Dominguez et al. (2021) presents a method that combines data from surveys, climate, and satellite imagery to estimate the hourly load profiles in East African rural households. We focus on using a limited number of smart meters and on implementing computational intelligence methods to estimate the load profile of the entire community. For this task, a combination of clustering algorithms and Markov chains (MCs) is considered. Numerous clustering algorithms have been proposed in the literature for various purposes. Unsupervised learning models include classic clustering techniques such as k-means, hierarchical clustering, kernel k-means and self-organizing maps (SOMs) (Alloghani et al., 2020). In this work, we compare two of the most common clustering methods: fuzzy c-means and SOM. The first method allows one piece of data to belong to one or more clusters with a certain degree of membership. Thus, fuzzy c-means has been useful in load profile clustering, where a load profile can belong to several groups simultaneously. For instance, in Prahastono et al. (2008), this method is applied for clustering electricity load profiles in Indonesia, considering pattern consumption. In addition, Anuar and Zakaria (2012) use fuzzy c-means to classify customers based on load patterns by measuring load profiles from feeders connected to different customer types. On the other hand, SOMs have been widely used due to the ease of visualization of clusters from high dimension input into a lower dimension output space. SOMs have been used to automatically classify electricity customers based on their consumption patterns (McLoughlin et al., 2015; Sanchez et al., 2009) and to study load profiles from known sociodemographic features (Richardson et al., 2010). In the case of MGs, the uncertainty in load behavior complicates the generation of consumption profiles. For example, in Llanos et al. (2017), SOMs were used to classify customers given their socioeconomic characteristics; then, the method assigned deterministic load profiles to different groups of customers to generate the load profiles required for sizing the MG. However, this approach considered only one MG sizing and design scenario.

Given the consumer classification by a clustering algorithm, it is pertinent to consider data with sufficient variability to prevent overfitting in the MG planning process and design. In cases with a limited amount of measured consumption data, the generation of synthetic consumption profiles is helpful to obtain more data with variability.

MCs are used in this work, as they have certain features that make them useful in this context.

MCs are stochastic processes that can capture stationary processes (Taylor & Karlin, 1984). Approaches based on MCs have been used to construct occupancy and activity profiles for household residents and to generate associated load profiles (Richardson et al., 2010; Widén & Wäckelgård, 2010). One drawback of MC approaches, however, is the considerable amount of detailed information about the activities of inhabitants required to build the profiles. In many cases, this detailed information is unavailable, and it may be considered invasive to even ask about it from the perspective of user privacy. One alternative is to use consumption measurements, as in Roje et al. (2017), in which a set of MCs was adjusted to generate stochastic load profiles characterizing the consumption of a whole community. However, this approach does not create variations for independent dwellings, which are helpful when planning a community MG.

When planning community grids, especially for small communities, having the most accurate estimate of consumption profiles possible is essential to define the dimension of the energy resources needed. It has been found that the smaller a community is, the more difficult it is to plan the size of the MG because variations are proportionally more substantial (Llanos et al., 2012).

We propose combining clustering algorithms and MCs to generate a more accurate method for small-scale MG planning. This method aims to identify the optimal size and topology of the MG for applications under various situations and community scenarios. A Mapuche rural community in southern Chile was selected for this purpose. The Mapuche are Indigenous people who represent 84.4% of the total population of indigenous people in Chile (Ministerio de Desarrollo Social, 2013). The harmony between humans, the natural world and spirits is part of the Mapuche cosmovision, and MG projects that use NCRE should fit within their cosmovision worldview.

To generate consumption profiles with added variability, the dwellings in this small Mapuche community were grouped by using sociodemographic data. For each group, consumption measurements from a representative dwelling were obtained. The characterization was performed with MCs, enabling the simulation of a year of consumption, while solar and wind potential in the community were determined with measured data from a weather station in the same community and mesoscale models. On the other hand, prediction interval models were identified to characterize the data uncertainty (Khosravi et al., 2010). A prediction interval gives a value range around the estimated output of the model, which represents the uncertainty of the demand and renewable generation. The objective for the construction of prediction intervals is to quantify the uncertainty in the point prediction to generate multiple scenarios for the best and worst conditions of the system (Alcántara et al., 2022; Cartagena et al., 2021; Serrano-Guerrero et al., 2021). With the characterization of the generation and consumption of the community, the MG units were sized under various scenarios, considering technical and economic criteria. These criteria generate significant impacts on the cost of energy (CoE) for the community, which is government subsidized. For a better comparison with previous works, in Table 1, we show the methodology of each reference for load profile generation and for uncertainty characterization of the load and renewable resources. Additionally, the comparison presented in this table is based on the planning methodology of the MGs: deterministic, stochastic or robust.

1.3. Contributions

Compared to previous works, we propose a robust microgrid planning methodology based on scenarios that (i) generate consumption profiles with a clustering method and MCs, (ii) the scenarios required for microgrid planning are obtained with a prediction interval method that characterizes the uncertainty of the demand and renewable resources, and (iii) the microgrid sizing approach considers a criterion of social and economic optimization. In view of the presented literature review, the main contributions of this work are:

- To the best of the authors' knowledge, this is the first paper that combines clustering methods (fuzzy c-means and SOMs) with MCs to generate consumption profiles with added variability required for MG planning.
- Using clustering methods enables a reduction in instrumentation for consumption measurements as they are clustered, thereby reducing the initial cost of the design compared to that of installing smart meters in every dwelling.
- The use of fuzzy and neural prediction intervals based on scenarios contributes to generating study scenarios of consumption and generation that improve the feasibility and robustness of an MG planned to be implemented in the studied community.
- This proposed robust planning approach considers a criterion of social-economic optimization that considers the global welfare of the community and society.

This paper is organized as follows: Section 2 presents the novel methodology for load profile generation with clustering algorithms (fuzzy c-means and SOMs) and MCs. Section 3 derives prediction intervals for modeling the various scenarios of the energy resources and demand of the MG, and the section includes the MG planning methodology-based social/economic criteria under these scenarios. Section 4 shows the application of the described methods in the case study. Finally, Section 5 presents the main conclusions and future work that may be done to extend and to verify this project for other applications.

2. Load profile generation based on cluster algorithms and Markov chains

This section presents the proposed method for the generation of load profiles for sizing MGs, considering the variability in consumption. First, the dwellings were classified into clusters according to sociodemographic information obtained through the use of short surveys of each family group in the community. In this work, two unsupervised methods were evaluated: fuzzy c-means and SOM (Greene et al., 2008; Llanos et al., 2017). The goal of using an unsupervised model is to group dwellings into a set of clusters based on the similarity of their features. These clusters are then useful to the generation of load profiles in the community. As each model produces different groups, these are analyzed to choose the more appropriate solution in the next step.

Subsequently, one dwelling per cluster was selected based on the availability of occupants. Then, based on the electrical consumption information of the representative household, first-order discrete finite MCs were used to generate load profiles. Finally, together with the electrical consumption of common spaces and productive demand, a representative stochastic community-consumption data set was obtained for a year.

2.1. Load pattern classification based on clustering algorithms

Fuzzy c-means follows an iterative process to minimize an objective function. This process considers numerical values from different features to find cluster center points, and in each iteration, the center is updated. The iterations consider the distance from all data to a cluster center, weighted by the degree of membership (Nayak et al., 2015; Ruspini et al., 2019). In contrast, SOM is a type of artificial neural network that is trained to produce a low-dimensional (typically, two dimensional) discretized representation of the input (Jovanović & Hikawa, 2022). This is a positive advantage of SOM, providing easily interpretable results. Additionally, this approach has gained popularity in many areas, and some authors have implemented this method for load profile generation purposes (Asan & Ercan, 2012). In addition, SOMs have the capacity to classify complex patterns (Kohonen, 1990; Kohonen et al., 2000). This grouping is performed by the projection

Table 1
Comparison between existing planning methods with our proposed approach.

Ref	Load profile generation	Uncertainty characterization	Planning methodology
Yon et al. (2021)	Assume that the load profile is known	Does not consider uncertainties	Deterministic
Li et al. (2020)	Estimate data from data of a similar rural area	Does not consider uncertainties	Deterministic
Kharrich et al. (2018)	Data is given	Does not consider uncertainty	Deterministic
Wang et al. (2017)	Given hourly data	Uncertainty of load is considered using the standard deviation ($\sigma = \pm 10\%$) to generate two scenarios	Robust
Khodaei et al. (2015)	Given hourly data	The load and renewable generation uncertainty is included considering forecast errors of $\pm 10\%$ and $\pm 20\%$	Robust
Borghei and Ghassemi (2021)	Data is given	A worst-case estimation of renewable generation is considered using historical data (generation and climatic)	Robust
Chen et al. (2022)	Data is given	Uncertainty of renewable generation using generative adversarial network based on historical data	Robust
Khayatian et al. (2018)	Data is given	Uncertainties in demand and component outages are considered using the latin hypercube Sampling method	Stochastic
Levorato et al. (2022)	Given hourly data	Uncertainty in demand and renewable generation considering a budget uncertainty set to control the level of conservatism	Robust
Our proposed approach	SOM and Markov Chains	Including prediction intervals with NN and fuzzy models	Robust

Table 2
Relevant features for consumption characterization by family.

Feature	Characteristic	Description
Number of youths	Age Range	Less than 18
Number of adults	Age Range	Between 18 and 60
Number of seniors	Age Range	More than 60
Number of students	Activities	Students
Number of workers	Activities	Fulltime workers
Number of farmers	Activities	Dedicated to agriculture
Number of housekeepers	Activities	Housekeeping activities
Number of retired or unemployed	Activities	Without a job

of an input space V_i on the output space V_o (generally of a lower dimension), defined as a set of neurons over a line or rectangular or hexagonal plane. The more dissimilar the input observations are, the more separated their projections in the output space. By using this method, different characteristics are compared for each cluster, and data with similar behavior are grouped.

After the training process described in Kangas and Kohonen (1996) and given an input vector $\vec{x} = (x_1, \ldots, x_n)$, SOMs activate neuron j in the output space if the weight vector $\vec{\omega}_j = (\omega_{j1}, \ldots, \omega_{jn})$ corresponding to that neuron has the shortest (Euclidean) distance to the input vector $\vec{x}$. To visualize the resulting classification, the unified distance matrix (U-matrix) is the most popular method (Ultsch & Siemon, 1990), where long and short distances between neurons are represented with dark and light colors, respectively.

To group dwellings, information from their sociodemographic characteristics is needed. According to Llanos et al. (2017), Richardson et al. (2010), Yamaguchi et al. (2011), three characteristics are directly related to consumption patterns: *members*, as the number of people living in the dwelling affects the consumption; *age range* because the behavior of the inhabitants is related to age; and *activities*, which provide an indication of occupancy in the dwelling. By using these characteristics, eight relevant features (Table 2) are retrieved from surveys and expressed for each dwelling in the input vector $\vec{x} = (x_1, \ldots, x_8)$.

Fuzzy c-means and SOM algorithms were used to group the dwellings into clusters. Then, a meter was installed in one dwelling of each group, and the measurements taken from these houses made it possible to estimate various load profiles with MCs, as explained in the next section. As we have two different clustering algorithms, we can evaluate the results given by these two models and select which gives the best approximation to the real monthly consumption given by the utility company by using the normalized mean absolute error (*NMAE*).

2.2. Load generation based on Markov chains

Once the clustering is obtained, as presented in the previous section, the measured consumption data of each of the clusters are used to generate representative synthetic demand profiles, which consider the variability of the data to be used for MG planning. For this work, MCs are used since they correspond to stochastic models that generate a sequence of events with different transition probabilities that are obtained based on the measured consumption, ensuring (with a certain degree of confidence) correlated behavior in the original data.

According to Levin and Peres (2017), MCs are processes that change their values between elements in a finite set ω. Each of these elements is known as a *state* that represents a particular variable. The transition between elements is determined by p_{ij}, that is, the transition probability of moving from state i to state j. The probabilities $\hat{p}_{ij}$ are calculated with the maximum likelihood estimator as (Anderson & Goodman, 1957):

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i} \quad \forall i = 1 \ldots m, j = 1 \ldots m, \tag{1}$$

where n_{ij} is the amount of data that has transitioned from a state i to a state j, while $n_i = \sum_{j=1}^{m} n_{ij}$. Then, the transition probabilities $\hat{p}_{ij}$ are stored in the transition matrix of dimension $m \times m$, with m being the number of states of the chain.

To validate that the MCs model the consumption within each hourly block (consisting of the set S_q, explained later), it is necessary to prove that their transition probabilities are constant within each set S_q; that is, $\hat{p}_{ij} = \hat{p}_{ij}(t) \quad \forall t$.

First, to determine the m number of states of each MC, Markov's inequality is used according to the procedure in Navarrete (2014), where an upper bound of m is obtained. The proof of this process can be found in Appendix A.

Starting with a high number of states m, we verified whether $c(n_i) \leq p^* \quad \forall i$, where p^* and t are design parameters (with p^* being

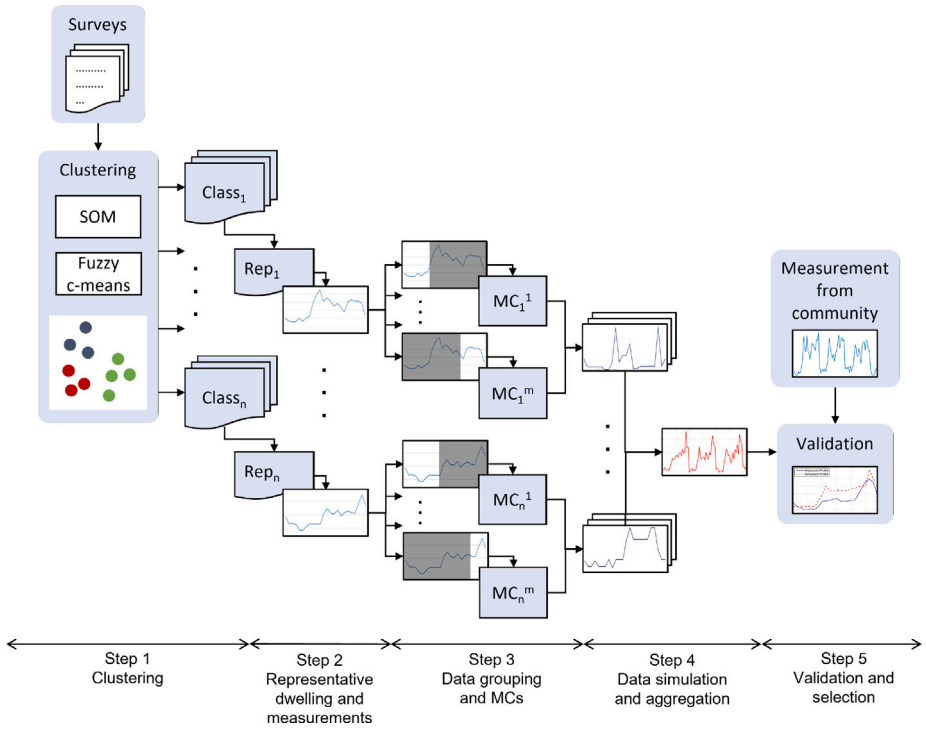

Fig. 1. Diagram of consumption generation.

the probability that $|p_{ij} - \hat{p}_{ij}|$ is greater than or equal to a threshold t), n_i is the amount of data that belongs to state i, and $c(n_i)$ is defined as:

$$c(n_i) = \min\left\{1, 2e^{-2n_i t^2}, \frac{1}{4n_i t^2}\right\} \tag{2}$$

If $c(n_i) \leq p^*$ is false, the maximum number of states m is reduced by one. When the condition is true $\forall i$, the maximum number of states m^* that makes the chain valid is obtained.

A collection of MCs is chosen for each identified class given by the clustering method following the procedure in Roje et al. (2017), seeking to explain the consumption patterns throughout the day. Each MC (where the states are power consumption, in [kW]) is valid within a time window that minimizes the intragroup quadratic error, defined by:

$$\min_{l_1 \ldots l_s}(J) = \sum_{q=1}^{s} \sum_{x \in S_q} \left\| x - u_q \right\|^2, \text{where} \quad u_q = \frac{1}{N_q} \sum_{x \in S_q} x \tag{3}$$

where x is the data to be grouped, S_q is the data set that defines the time windows with $q \in \{1 \ldots s\}$, s is the number of time windows selected, $l_1 \ldots l_s$ is the obtained hourly limits defined by S_q, N_q is the number of elements in the set $x \in S_q$, and u_q is the average of the data x in the set S_q.

To determine the power values of these states, the *k-means* algorithm (MacQueen, 1967) with m^* clusters over the power consumption within the time window was performed, selecting the centroids as the values for the states.

Although the proposed use of MCs does not represent seasonal variations or changes in demand habits, it does generate profiles with a representative variability of the measured data. This proposal enables us to obtain different consumption levels and scenarios, which is useful for MG planning. Fig. 1 summarizes the steps used to generate the synthetic load profiles.

First, data collected from surveys are used to create groups of dwellings (classes) by using a clustering method (SOM or fuzzy c-means). In step 2, a representative dwelling is selected for each of the identified classes, and its consumption is measured for a certain period. For each of these measurements, the data are divided into sets of consecutive time groups, and for each one, the number of states of the MC, such that it is statistically representative, is determined (step 3).

Once this process is finished, in step 4, synthetic consumption profiles with MCs are generated for 365 days for each of the classes identified through clustering. These profiles are associated with each of the dwellings present in the community and then aggregated to obtain a community demand profile, which is validated (step 5) and used for MG planning. The following section presents the prediction interval models for generating scenarios and the proposed optimization used in the MG planning process.

3. Microgrid planning based on modeling of uncertainty of renewable resources and load

This section presents an optimization approach for MG planning based on scenarios. The prediction interval (PI) models ensure, with a certain coverage probability, that the trajectories of the renewable resources (wind and solar) and demand consumption (the latter explained in Section 2) are included in the interval defined by the upper and lower bounds; thus, the scenarios were defined by these bounds. The PI models developed in this paper were adopted from the study presented in Marín et al. (2019) and are included here for self-containment. For this approach, we use the concept of fuzzy numbers to obtain PI models based on fuzzy systems and a neural network, which are explained next.

3.1. Prediction interval models based on Fuzzy numbers

When an affine linear model is used to obtain a prediction model by using a finite quantity of measured data, the prediction output $\hat{y}(k)$ at time k is defined as follows:

$$\hat{y}(k) = \theta_o + \theta_1 z_1(k) + \cdots + \theta_p z_p(k), \tag{4}$$

where θ_i are the regression coefficients and $z_i(k)$ $(i = 0, 1, \ldots, p)$ are the input measurement data at time step k. To account for uncertainty, the coefficients θ_i are defined as interval fuzzy numbers (Lee, 2005;

Mendel, 2017). Uncertainty is related to the error between the prediction $\hat{y}(k)$ and the actual output $y(k)$, and it is defined by the interval $[\hat{y}_L, \hat{y}_U]$ in which the predicted value may lie.

The parameters θ_i are defined by interval fuzzy numbers, and therefore, they are characterized by the mean (m) and spread (s). The uncertainty of the expected value is characterized by using different spread values, i.e., $\theta_i = [m_i - \underline{s}_i, m_i + \overline{s}_i]$. The lower bound ($\hat{y}_L$) and upper bound ($\hat{y}_U$) that define the PI are obtained by the theorem of the affine combination of type-1 interval fuzzy numbers (Karnik & Mendel, 2001; Mendel, 2017), as follows:

$$\hat{y}_L(k) = \sum_{i=1}^{p} m_i z_i(k) + m_0 - \sum_{i=1}^{p} |z_i(k)| \underline{s}_i \tag{5}$$

$$\hat{y}_U(k) = \sum_{i=1}^{p} m_i z_i(k) + m_0 + \sum_{i=1}^{p} |z_i(k)| \overline{s}_i \tag{6}$$

The expected value is characterized by the mean (m_i), as shown in (5) and (6), and the uncertainty of the PI is characterized by the parameters $(\overline{s}_i, \underline{s}_i)$ of the last term in both equations.

In this PI modeling approach, the mean values of the parameters (m_i) associated with providing the expected value $\hat{y}(k)$ are obtained by using an appropriate model identification procedure, and the parameters determine which prediction model is used: fuzzy system or neural network. The parameters $\overline{s}_i$ and $\underline{s}_i$ are tuned to guarantee the desired coverage probability $(1-\alpha)\%$ with the smallest interval width. The proposed method for identifying these parameters (spreads) is described in Marín et al. (2019). Finally, the PI provides the expected value $\hat{y}(k)$ and the values of the upper ($\hat{y}_U$) and lower ($\hat{y}_L$) bounds according to the desired coverage probability. In the next section, both fuzzy and neural network PI models based on fuzzy numbers are presented.

3.1.1. Fuzzy prediction interval modeling

In the fuzzy PI model based on fuzzy numbers, the parameters of the consequences (θ_i^j) of each rule ($j = 1, \ldots, M$) are defined by fuzzy numbers. Thus, the lower ($\hat{y}_L^j$) and upper ($\hat{y}_U^j$) output of the fuzzy interval for each rule (j) are given by the following bounds:

$$\hat{y}_L^j(k) = \sum_{i=1}^{p} m_i^j z_i(k) + m_0^j - \sum_{i=1}^{p} |z_i(k)| \underline{s}_i^j \tag{7}$$

$$\hat{y}_U^j(k) = \sum_{i=1}^{p} m_i^j z_i(k) + m_0^j + \sum_{i=1}^{p} |z_i(k)| \overline{s}_i^j \tag{8}$$

In (7)–(8), m_i^j are the mean values, and $\overline{s}_i^j$, $\underline{s}_i^j$ are the spread values of the parameter θ_i^j of each rule (j) associated with a set of p inputs $(z_1(k) \in Z_1, \ldots, z_p(k) \in Z_p)$ at time k. Then, the global bounds of the fuzzy PI are obtained as follows:

$$\hat{y}_L(k) = \sum_{j=1}^{M} \beta^j(Z(k)) \hat{y}_L^j(k) \tag{9}$$

$$\hat{y}_U(k) = \sum_{j=1}^{M} \beta^j(Z(k)) \hat{y}_U^j(k) \tag{10}$$

where $\beta^j(Z(k))$ is the normalized activation degree. In this approach, a fuzzy clustering method is considered to define the rule numbers and the parameters (center and standard deviation) of the Gaussian membership functions of the antecedents. The means (m_i^j) of the consequences are estimated by the minimum least-squares optimization method (Cartagena et al., 2021). Finally, the spreads $(\overline{s}_i^j, \underline{s}_i^j)$ are obtained by solving the unconstrained optimization problem presented in Marín et al. (2019).

3.1.2. Neural network prediction interval modeling

Similar to the approach followed for the fuzzy PI presented above, this section describes a PI based on fuzzy numbers by using neural networks. In this approach, the network weights are modeled as interval fuzzy numbers.

The output of the neural network at time k is defined as follows:

$$\hat{y}(k) = \sum_{j=1}^{L} w_j^0 \left(\tanh\left(\sum_{i=1}^{p} w_{j,i}^h z_i(k) + b_j^h \right) \right) + b^0 \tag{11}$$

where $z_i(k)$ is the set of p inputs $(z_1(k) \in Z_1, \ldots, z_p(k) \in Z_p)$ and $j = 1, \ldots, L$, and L is the number of hidden layer units. The neural network in (11) can be written as follows:

$$\hat{y}(k) = \sum_{j=1}^{L} w_j^0 \tilde{Z}_j(k) + b^0 \tag{12}$$

where:

$$\tilde{Z}_j(k) = \tanh\left(\sum_{i=1}^{p} w_{j,i}^h z_i(k) + b_j^h \right) \tag{13}$$

The PI based on a neural network is developed such that the output weights (w_j^0) are considered interval fuzzy numbers with means (m_j) and spreads $(\underline{s}_j, \overline{s}_j)$. Therefore, the lower ($\hat{y}_L$) and upper ($\hat{y}_U$) bounds of the interval are calculated as follows:

$$\hat{y}_L(k) = \sum_{j=1}^{L} m_j \tilde{Z}_j(k) + b^0 - \sum_{j=1}^{L} \left| \tilde{Z}_j(k) \right| \underline{s}_j \tag{14}$$

$$\hat{y}_U(k) = \sum_{j=1}^{L} m_j \tilde{Z}_j(k) + b^0 + \sum_{j=1}^{L} \left| \tilde{Z}_j(k) \right| \overline{s}_j \tag{15}$$

To train this PI model, two identification routines must be executed. The first training procedure is responsible for identifying mean (m_j) parameters and consists of a traditional setup for neural network regression consisting of least means square optimization through stochastic gradient descent via the backpropagation algorithm (Cartagena et al., 2021). The second procedure obtains the spreads $(\underline{s}^j, \overline{s}^j)$ by solving the unconstrained optimization problem presented in Marín et al. (2019).

The PI models presented in this section are used to predict the expected values and to characterize the uncertainty of the renewable resources and load. The intervals are used to generate several scenarios (for example, a worst-case scenario, given by the PI upper bound of consumption and PI lower bound of generation), which can be used in an optimization problem to define the appropriate microgrid topologies. Next, the optimization problem for MG planning is presented.

3.2. Microgrid planning

The MG operation, in general, should perform optimally in all its objectives: economic, technical, environmental, among others (Guo et al., 2016; Hafez & Bhattacharya, 2012; Khodaei et al., 2015; Su et al., 2010; Ubilla et al., 2014). This criterion is adjusted to the methodologies used by the National Public Investment System of Chile and, in particular, to the Methodology for the Formulation and Evaluation of Rural Electrification Projects provided by its Ministry of Social Development and Family (Ministerio de Desarrollo Social y Familia, 2022), which allows determining the dimensioning of that MG alternative among the various feasible energy supply solutions. The latter provides the supply in the most efficient way to the population of the locality in which the project is implemented. Thus, the various MG alternatives must be evaluated by using a cost-efficiency approach since Chilean public policy has defined the need and duty of the state to provide energy to all rural sectors.

Regarding the technical feasibility, all the MG alternatives must cover the recent and projected demand of the population with the energy resources available in the study locality considering the useful life and constraints and technical characteristics of each piece and equipment of the MG, among other design considerations.

Regarding social feasibility, it is important to determine the degree of acceptance of each technically feasible alternative by the community; therefore, citizen participation activities are essential.

Regarding the economic feasibility, we must consider the investment, operation (including fuel), maintenance and replacement costs of

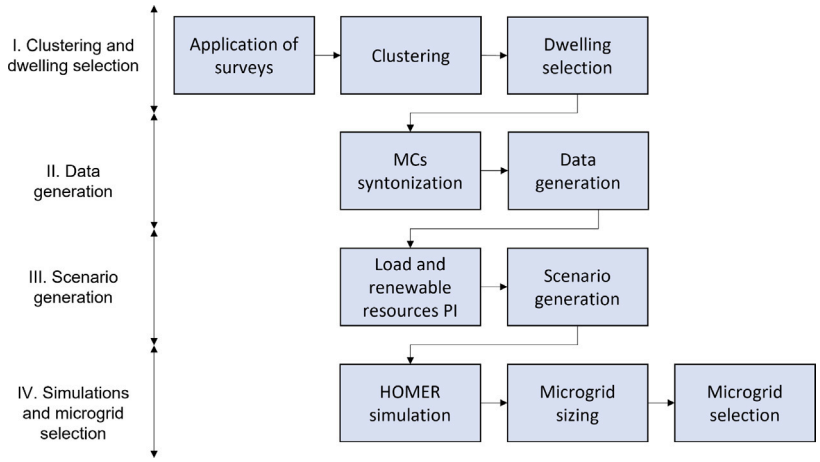

Fig. 2. Diagram of consumption generation.

each piece of equipment and component of the MG, in addition to the positive and negative externalities caused by its implementation (for example, reduction of CO2), and the discount rate, which reflects the opportunity cost of its financing. All costs must be valued at the social price through the procedure and corrections indicated by the Ministry of Social Development and Family of Chile (Ministerio de Desarrollo Social y Familia, 2022) to ensure that the values used reflect the true cost for society of using economic resources (whether public or private) during the execution and operation of the project. Finally, the decision criterion to determine the most convenient alternative corresponds to the one that presents the lowest net present cost (NPC).

$$min\left(NPC_i\right) \tag{16a}$$

$$NPC_i = I_i + \sum_{t=1}^{n} \frac{C_{i,t}}{(1+r)^t} \tag{16b}$$

where i are the feasible alternatives; NPC_i is the net present cost of i; I_i is the initial investment of i; n is the evaluation horizon; $C_{i,t}$ is the operation, maintenance and replacement costs of i at t; and r is the discount rate.

Different scenarios were obtained with the approach presented in this study to analyze the feasibility of the implementation of the MG. The baseline scenario corresponds to annual measured profiles of solar irradiance, wind speed and consumption demand obtained as explained in Section 2. Then, several scenarios are obtained based on the prediction interval models for renewable resources and consumption for different desired coverage probabilities. The last step in this method corresponds to selecting the final MG design, relying on the optimization software HOMER for distributed systems (National Renewable Energy Laboratory NREL, 2005).

The complete process for the proposed robust microgrid planning can be visualized in Fig. 2, where it starts with the application of a characterization survey to the families of the community, from which the clustering is performed, as described in Section 2.1, that allows the grouping of similar dwellings and from there, selecting a representative dwelling for the measuring of electricity consumption data (step I from the figure). From the obtained data, MCs are tuned to generate consumption data representative of the dwellings, which can be generated for the required period (step II), according to Section 2.2. The consumption data, together with those of variable renewable resources, prediction intervals (PI) are obtained as described in Section 3.1, to generate various scenarios of interest (step III), from which simulations are performed that allow sizing and selecting the most appropriate microgrid, according to previously defined evaluation and selection criteria (step IV).

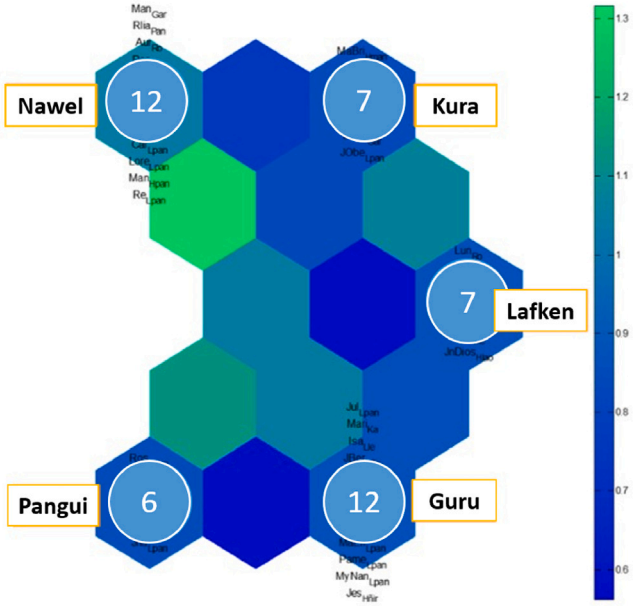

Fig. 3. Grouping of family types in the community using SOM.

4. Case study

The case study involved an Indigenous Mapuche community, the José Painecura Hueñalihuen, La Araucanía Region, Chile. Although the community is connected to the main grid, the quality of service is poor, with many unscheduled outages.

4.1. Load generation based on clustering and Markov chain methods

The steps described in Section 2 are applied to the survey results from Table 2 of the 44 dwellings in the community, with these data used in the SOM and fuzzy c-means configurations. For this case, an SOM hexagonal plane in 3 rows and 2 columns (6 total neurons) is obtained by using a U-matrix representation, while fuzzy c-means creates five clusters (which can be visualized using a PCA algorithm). The results of both methods are shown in Figs. 3 and 4, specifying the number of families that belong to each group. Five groups were obtained by using both clustering methods, each named in the Mapuche language and sharing common characteristics.

For each case (groups given by fuzzy c-means and SOM), a dwelling was selected for each of the previously obtained groups. In these

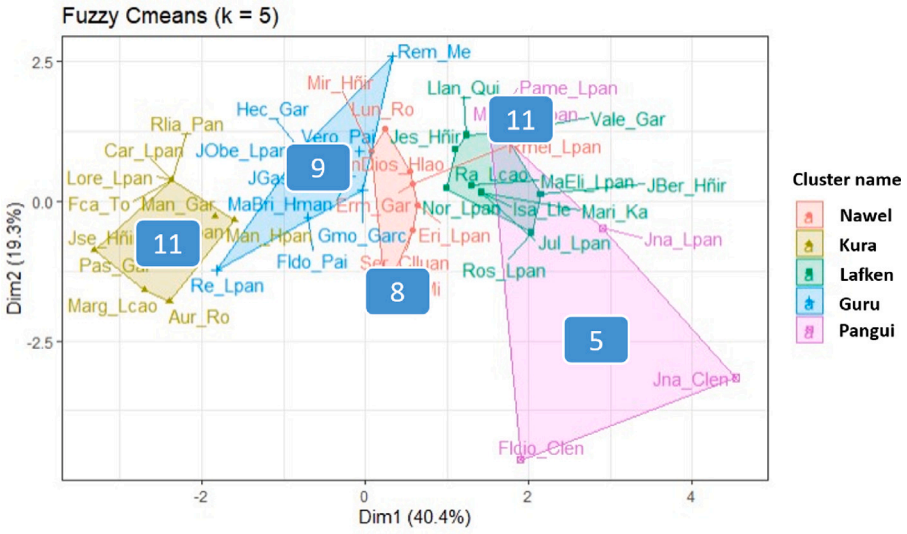

Fig. 4. Grouping of family types in the community using fuzzy c-means.

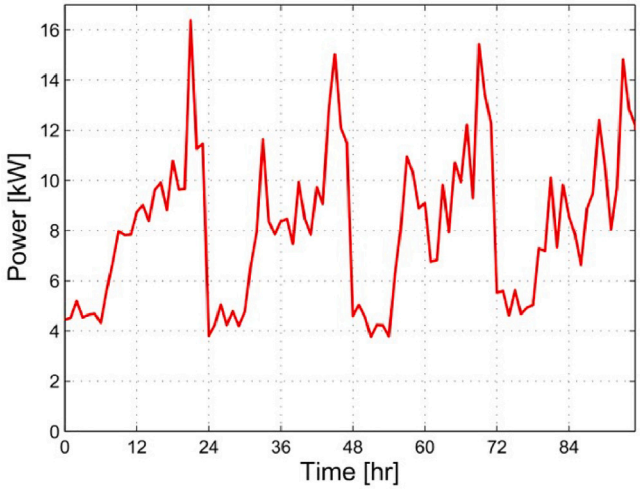

Fig. 5. Load profile for the community for 4 days of SOM-MC simulation.

dwellings, measurements were obtained with a sample time of 1 h for a period of approximately 60 days by using smart meters. These records were used in MCs. A heuristic analysis of the measured data showed between two and three levels of consumption (roughly low, medium, and high consumption) throughout the day, depending on the dwelling class. These consumption levels are obtained by the application of (3), which minimizes the intragroup error. Generally, high, medium, and low consumption are observed in the evening, afternoon, and early morning, respectively. Each identified consumption level from each group has a particular MC, and the range of time in which this MC is valid depends on the minimization of (3). Next, the maximum quantity of valid states for each MC was determined with the procedure of (2), setting the design parameters to $r = 0.25$ and $p^* = 0.25$ and the initial maximum quantity of states $m = 8$ for each selected time window. The results of the application of this procedure are shown in Table 3 for each group, where m_1, m_2, and m_3 are the numbers of time windows and the values are the numbers of states for each.

The resulting MCs were defined with their transition matrices ($\hat{p}_{ij}$) and centroids u_i representing the power values for each state obtained by the *fuzzy c-means* algorithm. With these parameters, several realizations of the MC were performed, obtaining different load profiles for each of the identified classes, based on which the consumption for each dwelling was simulated.

Table 3
Number of states of the obtained MCs.

Group name	No. of Time windows	No. of states		
		m_1	m_2	m_3
Nawel	2	3	1	–
Kura	3	7	3	3
Lafken	3	2	3	2
Guru	3	3	3	4
Pangui	2	8	8	–

In addition to residential consumption, there was community consumption corresponding to the school, which was highly predictable and differed between workdays and weekends, with a peak power of 216 [W] and 16 [W], respectively. Thus, the average consumption of the school was added to the residential consumption to obtain the overall community consumption. As the consumption varied throughout the year, the load profiles were weighted according to the energy consumption at various times of the year by using data obtained from the electric utility. With these considerations, a simulation for 365 days was performed for both clustering cases. In this part, the *NMAE* between the simulation and the measured data of two months given by the utility company was used to select the best clustering algorithm. An *NMAE* of 24.8 for SOM and 25.9 for fuzzy c-means was obtained. Thus, for this case study, the SOM algorithm was selected as the clustering method. Fig. 5 shows four days of simulated consumption by using SOM. Fig. 6 shows the averaged data for the simulation (SOM case) and from the readings of the utility company. The model overestimates the average load profile, which is desirable for sizing purposes. Moreover, as the MG scales up, the simulated results should have a lower *NMAE* as the aggregated consumption should have lower uncertainty. The obtained simulation (when using SOM clustering) has a mean consumption of 195.92 [kWh/day], a mean power of 8.16 [kW] and a peak power of 19.51 [kW]. The MC output was used as input for the identification of the prediction interval model of consumption due to the larger amount of training data with respect to the measured consumption to avoid overfitting.

Table 4 provides a general overview of the main features of each group obtained by the SOM algorithm. Groups obtained by SOM follow a clear split based on the main features of households, as expected.

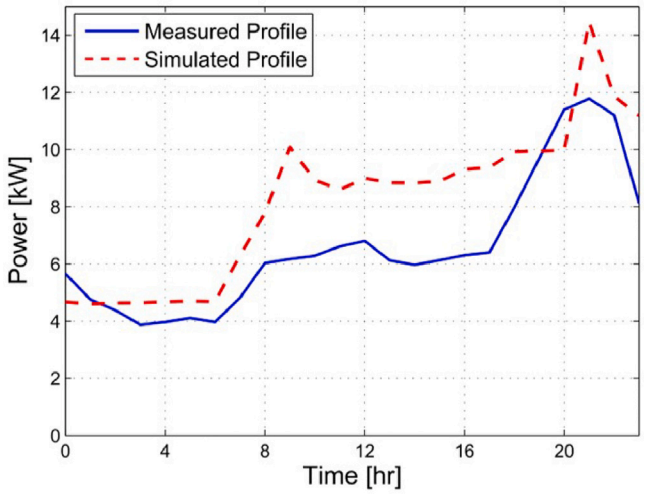

Fig. 6. Averaged load profile for the community.

Table 4
SOM classification of family types.

Group name	No. of dwellings	No. of family members	Economic activities
Nawel (*Tiger*)	12	4 or less	Retired seniors
Kura (*Rock*)	7	2 or 3 adults	Agriculture
Lafken (*Sea*)	7	3 (no more than 2 youths)	Agriculture or housekeeping
Guru (*Fox*)	12	4 (at least 2 youths)	Housekeeping, agriculture and studies
Pangui (*Lion*)	6	5 or more (at least 3 adults)	Agriculture or retired

4.2. Prediction interval models for the electrical demand and renewable generation

The scenarios required for the MG planning process were obtained by using the PI described in Section 3.1. Thus, linear, fuzzy and neural network PI models were employed to evaluate the accuracy of the models. The accuracy corresponds to the PI model that achieves the desired coverage probability with the sharpest interval because a narrower width of the PI provides more accurate information about the modeled uncertainty. The data used to construct the PI models are as follows:

- **Wind speed**: 365 measured and estimated days at a sampling time of 1 h, in [m/s]. The data were measured for 292 days at 5 [m] height and extrapolated using the *wind profile power law* to 15 [m]. The "*Explorador Eólico*"[1] mesoscale model was used to complete a whole year. The wind speed was 4.87 [m/s] on average, and over 75% of the time, it was more than 3 [m/s], the starting speed of a typical wind turbine. Furthermore, less than 5% of the time, the wind speed was over 12 [m/s], which is the safety threshold for operation.
- **Solar irradiance**: 365 measured and estimated days at a sampling time of 1 h, in [kW/m^2]. Measurements were recorded for 292 days and filled in with the "*Explorador Solar*"[2] mesoscale model to complete a whole year.
- **Demand**: 365 simulated days at a sampling time of 1 h, in [kW]. A detailed description is provided in Section 4.1.

[1] Explorador Eólico https://eolico.minenergia.cl/inicio
[2] Explorador Solar https://solar.minenergia.cl/inicio

The data set was divided into 60% for training, 20% for testing, and 20% for validation. During the identification procedure of the models based on fuzzy systems and neural networks, the relevant inputs (regressors) must be defined, and structure optimization of the models must be performed. For fuzzy systems, structure optimization corresponds to defining the number of rules, and for neural networks, it corresponds to defining the number of hidden layer units. Then, the parameters of the PI are identified (Marín et al., 2019). The results of the identification procedure of the models are presented in Table 5.

The prediction interval coverage probability ($PICP$), which computes the number of measured values that fall within the predicted interval; the prediction interval normalized average width ($PINAW$), which quantifies the width of the interval; and the root mean square error ($RMSE$), which evaluates the accuracy of the model, are used as metrics for evaluating the prediction interval models. $PICP$ and $PINAW$ are described mathematically as follows:

$$PICP = \frac{1}{N}\sum_{k=1}^{N} \delta_k \times 100\% \tag{17}$$

$$PINAW = \frac{1}{N \cdot R}\sum_{k=1}^{N} \left(\hat{y}_U(k) - \hat{y}_L(k)\right) \times 100\% \tag{18}$$

where N is the number of observations in the data set; $\hat{y}_L(k)$ and $\hat{y}_U(k)$ are the lower and upper bounds of the prediction interval; $\delta_k = 1$ if $y(k) \in [\hat{y}_L(k), \hat{y}_U(k)]$; otherwise, $\delta_k = 0$; and R is the range of the output defined by the difference between the maximum and minimum measured values.

Table 6 shows these metrics with the test data set tuning to a desired $PICP$ of 50% and 90% (while the parameters of these models were obtained from the training and validation data sets) for wind speed, solar irradiance and load by using the linear, fuzzy and neural network models. In this study, PI models with 50% and 90% of the desired $PICP$ for demand and renewable generation were used for MG planning.

Table 6 indicates that the lowest $RMSE$ for the wind speed, solar irradiance and demand models was obtained with the PI based on the neural network. Additionally, with this neural network PI, the coverage probability is close to the desired $PICP$ with the test data, and the width (measured by the $PINAW$ index) of the interval is sharpest with respect to that of the other approaches: linear and fuzzy systems.

Because the objective of a PI is to obtain as narrow upper and lower bounds as possible while guaranteeing that the interval contains the greatest possible amount of measured data, the neural network models generate the most high-quality intervals. Finally, Table 6 shows that solar irradiance has low variability, measured by the $PINAW$ metric. Therefore, the expected value of solar irradiance is used in the MG planning process directly instead of the lower or upper bound of the PI.

Figs. 7 and 8 show the neural network PI for the wind speed and demand for four days when using the test data set with 90% of the desired $PICP$.

Based on the results of the PI, the upper and lower bounds are generated with neural network models with different desired $PICP$. Then, several scenarios can be generated according to the desired coverage probability in which the PI models are tuning.

4.3. Planning based on scenarios

Given the neural network prediction interval models generated as discussed in the previous section, the optimization process described in 3.2 was conducted to obtain feasible alternatives for the MG, which are compared to the deterministic design process, **Baseline**. The various scenarios are considered as follows:

- **Worst-case scenario:** in which the wind speed equals the lower bound of the PI and the demand equals the upper bound of the PI, both adjusted to 90% of $PICP$.

Table 5
Identification procedure results.

	Fuzzy prediction interval		Neural network prediction interval	
	Number of relevant inputs (regressors)	Number of rules	Number of relevant inputs (regressors)	Number of hidden layer units
Wind Speed	9	3	10	8
Solar Irradiance	10	2	11	5
Demand	13	4	12	10

Table 6
Performance indices for prediction interval models.

Desired PICP	Metrics	Lineal model			Fuzzy model			Neural network model		
		Wind Speed	Solar Irradiance	Load	Wind Speed	Solar Irradiance	Load	Wind Speed	Solar Irradiance	Load
50%	RMSE (kW)	1.118	0.056	1.131	1.089	0.055	1.078	1.067	0.052	0.941
	PICP (kW)	49.31	49.08	51.48	49.83	50.13	49.95	49.58	50.55	50.07
	PINAW (%)	10.75	5.91	16.21	9.98	4.72	15.98	9.03	4.55	14.18
90%	RMSE (kW)	1.118	0.056	1.131	1.089	0.055	1.078	1.067	0.052	0.941
	PICP (%)	89.01	89.18	92.18	89.43	88.93	88.72	89.38	88.55	88.67
	PINAW (%)	19.66	10.24	30.29	19.55	7.85	25.18	18.13	7.88	23.08

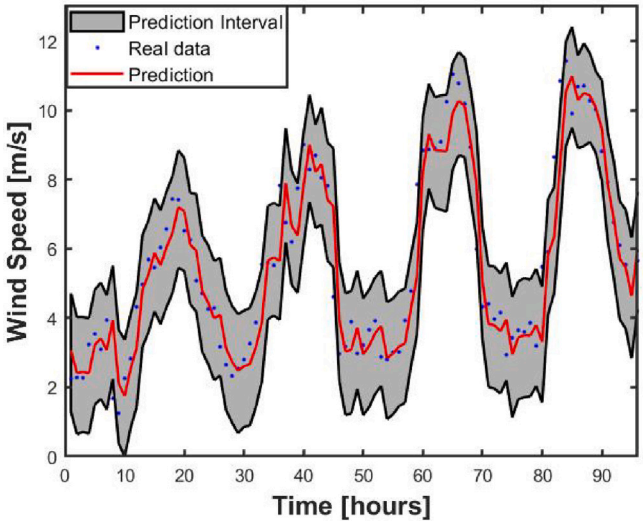

Fig. 7. Neural network prediction interval for wind speed with $PICP = 90\%$.

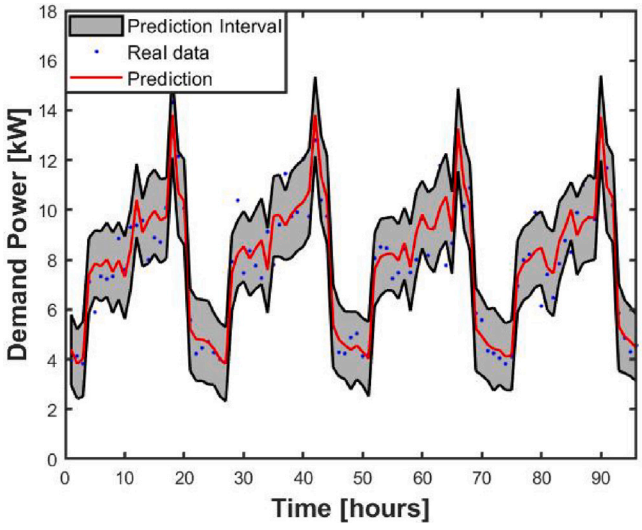

Fig. 8. Neural network prediction interval for demand consumption with $PICP = 90\%$.

- **Bad-case scenario:** in which the wind speed is estimated to be the lower PI and the demand is estimated by the upper PI, both adjusted to 50% of $PICP$.
- **Good-case scenario:** in which the wind speed equals the upper PI and the demand equals the lower PI, both adjusted to 50% of $PICP$.
- **Best-case scenario:** in which the wind speed is estimated to be the upper bound of the PI, and the demand is estimated by the lower bound of the PI, both adjusted to 90% of $PICP$.

In all scenarios, we use the expected value of solar radiation due to the low variability of this resource, and the baseline case corresponds to the expected values of the models.

In this regard, the community expressed the need not to depend on the main grid supply from the electricity company, a social feasibility restriction that is considered in the design. Table 7 presents the optimum MG design, costs, and operation performance under the five scenarios.

The optimization process resulted in a different MG design for each scenario. In general terms, as a more unfavorable scenario is considered (higher demand and lower wind speed), the project's NPC is higher, mainly due to increased operating costs associated with higher fuel consumption. In this case study, it is also possible to determine that, because of the high cost of investment in the installation of wind turbines, only if the real wind conditions are better than the baseline, should the hybrid solar-wind configuration be proposed; otherwise, wind turbine installation should not be recommended.

A conservative criterion was considered to minimize the investment risk, so the design in the worst-case scenario was selected. Compared to that of the baseline, in the worst-case, the load profile demands 11.5% more power from the gensets than does the baseline: this increases the operational costs by USD$1,305/year and the fuel consumption by 15.3%. Because in this scenario a 19.9% higher demand is supplied, a lower energy cost is obtained. Finally, yet importantly, if an investor in the project is less risk-averse, they may consider one of the other designs proposed in Table 7 for a more favorable scenario, which may allow them to make a smaller initial investment.

5. Conclusions

A planning methodology that includes various scenarios in an MG connected to the main grid in a rural Mapuche community was proposed. Four scenarios were generated by using prediction interval models to represent different possible wind speed and load conditions. A simulator of power consumption based on sociodemographic data

Table 7
Microgrid design, associated costs and operation under different scenarios.

Configuration	Unit	Worst-case	Bad-case	Baseline	Good-case	Best-case
Wind turbine of 10 [kW]	[units]	0	0	0	1	1
Photovoltaic array	[kWp]	111.0	92.8	89.7	63.1	38.6
Solar inverter	[kW]	40.0	30.0	30.0	20.0	18.0
Battery bank	[kWh]	162.4	144.3	139.8	117.3	85.7
Battery inverter	[kW]	23.2	20.1	20.4	15.6	13.9
Genset	[kW]	20.0	20.0	20.0	20.0	20.0
Costs	Unit	Worst-case	Bad-case	Baseline	Good-case	Best-case
Net Present Cost	[USD]	$476,212	$433,575	$417,734	$391,240	$336,241
Initial investment	[USD]	$333,914	$296,491	$290,407	$280,069	$229,262
Operating costs	[USD/year]	$12,406	$11,952	$11,101	$9,780	$9,327
Energy costs	[USD/kWh]	$0.491	$0.507	$0.516	$0,523	$0.514
Operation	Unit	Worst-case	Bad-case	Baseline	Good-case	Best-case
Fraction of NCRE	[%]	84.1	85.7	86.4	91,9	91.4
Fuel consumption	[lts/year]	3,768	3,672	3,269	1,822	1,671
Production of the gensets	[kWh/year]	10,977	9,627	9,560	5,286	4,887

was presented by using clustering methods (SOM and fuzzy c-means) and MCs for obtaining load profiles from the dwellings, with sufficient variability to simulate data. The interval models enabled characterization of the best-, good-, bad- and worst-case scenarios of consumption and renewable resources within a defined coverage probability.

Based on the generated scenarios and considering technical-economic criteria, the appropriate MG topologies for implementation were suggested to facilitate decision-making. An appropriate topology for an MG was obtained by using the worst-case scenario of low wind speed and increased demand. Compared to those of the other scenarios, the MG topology was similar, but the investment risk was minimized under the worst-case scenario.

The use of computational intelligence methods (MCs, SOMs and PIs) constitutes a novelty for this type of microgrid planning methodology. MG planning requires data that are not usually available, especially in small, underserved communities. In these situations, the use of clustering methods is useful to group dwellings and obtain measurements from representative selected locations. Grouping dwellings together generates savings in the planning stage, as fewer measurement instruments are needed. Furthermore, grouping reduces the time necessary to locate all the occupants and recover the data from the home-based measurement instruments. Finally, the prediction interval introduced here opens a path to obtain more realistic scenarios necessary to begin MG planning for small-scale projects, ensuring feasibility even in worst-case scenarios.

The proposed methodology presented in this work can be improved in some ways. The measurement of data for longer periods may allow the determination of seasonal modifications in consumption or habits, in which case it is possible to adjust new MCs or to use other strategies that account for these changes in demand. In the same direction, the dwelling clusters from another sociodemographically similar community can be associated with the cluster data obtained in this work to validate or adjust the load profiles generated from this study with respect to actual consumption measurements from the other similar community. Regarding the generation of scenarios, a stochastic approach can be compared against the robust approach of this work for microgrid sizing, simulation and selection to evaluate the performance and obtained final costs from the evaluation given a certain risk level.

CRediT authorship contribution statement

Raúl Morales: Conceptualization, Methodology, Software, Writing – original draft. **Luis G. Marín:** Methodology, Software, Writing – review & editing. **Tomislav Roje:** Methodology, Software, Writing – review & editing. **Víctor Caquilpan:** Methodology, Software, Writing – review & editing. **Doris Sáez:** Methodology, Visualization, Writing – review & editing. **Alfredo Nuñez:** Visualization, Resources, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was funded by Instituto Sistemas Compejos de Ingeniería (ISCI), Chile under grant ANID PIA/PUENTE AFB220003, the Solar Energy Research Center SERC-Chile under grant ANID/FONDAP/1522A0006, and the ANID/FONDECYT 1220507 grant.

Appendix A. Determination of the Markov chain number of states

One important task is to prove that the transition probabilities are constant within each set S_q, that is, that $\hat{p}_{ij} = \hat{p}_{ij}(t) \quad \forall t$. Thus, the χ^2 test is performed for the MC transition matrix (Anderson & Goodman, 1957), where the value of χ^2 is calculated as:

$$\chi_q^2 = \sum_{t \in S_q} \sum_{i=1}^{m} \sum_{j=1}^{m} n_i(t) \cdot \frac{\left(\hat{p}_{ij}(t) - \hat{p}_{ij}\right)^2}{\hat{p}_{ij}} \quad \forall q = 1 \dots s \tag{A.1}$$

where t belongs to set S_q.

The null (H_0) and alternative (H_a) hypotheses are proposed, for which the p-value obtained from the χ_q^2 test is compared against the threshold defined by the α-value of 0.01:

$$H_0 : \quad p_{ij}(t) = p_{ij}, \quad \forall t \quad t = 1 \dots T \tag{A.2a}$$

$$H_a : \quad p_{ij}(t) \neq p_{ij}, \quad \exists t \quad t = 1 \dots T \tag{A.2b}$$

The degrees of freedom of the χ^2 distribution to obtain the p-value are $DOF = m(m-1)(T-1)$. If p-value $\geq \alpha$-value, there is evidence against the null hypothesis H_0; otherwise, H_0 is accepted as valid.

Based on the method of Navarrete (2014), the determination of the MC number of states is performed by using the Markov inequality, which states that:

$$\mathbb{P}\{X \geq t\} \leq \mathbb{P}\{\phi(X) \geq \phi(t)\} \leq \frac{\mathbb{E}\{\phi(X)\}}{\phi(t)}, \tag{A.3}$$

with ϕ a monotonically increasing and nonnegative function. Let $\phi(t) = t$; we have:

$$\mathbb{P}\{X \geq t\} \leq \frac{\mathbb{E}\{X\}}{t}. \tag{A.4}$$

Chebyshev's inequality, the product of taking $\phi(t) = t^2$ and $X = |Z - \mathbb{E}(Z)|$, expresses that Boucheron et al. (2003):

$$\mathbb{P}\{|X - \mathbb{E}\{X\}| \geq t\} \leq \frac{Var\{X\}}{t^2} \tag{A.5}$$

Let $S_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, where $X_1, \ldots, X_n$ are independent random variables corresponding to a Bernoulli experiment that has a value of 1 with probability p_{ij} if the transition is from state i to j, and 0 with probability $(1 - p_{ij})$ otherwise. Then:

$$\mathbb{P}\left\{|S_n - \mathbb{E}\{S_n\}| \geq t\right\} = \mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mathbb{E}\{X_i\}\right)\right| \geq t\right\} \tag{A.6a}$$

$$\leq \frac{Var\{S_n\}}{t^2} = \frac{\sum_{i=1}^{n} Var\{X_i\}}{n^2 t^2} = \frac{n\sigma^2}{n^2 t^2} = \frac{\sigma^2}{nt^2} \tag{A.6b}$$

Furthermore, since $\mathbb{E}\{X_i\} = p_{ij}$:

$$\mathbb{E}\{S_n\} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\{X_i\} = \frac{1}{n} n \cdot \hat{p}_{ij} = \hat{p}_{ij} \tag{A.7}$$

By substituting this result into inequality (A.4), we obtain:

$$\mathbb{P}\left\{|p_{ij} - \hat{p}_{ij}| \geq t\right\} \leq \frac{\sigma^2}{nt^2} = \frac{p_{ij}(1 - p_{ij})}{nt^2} \tag{A.8}$$

The upper bound is obtained with $p_{ij} \cdot (1 - p_{ij}) = 0.25$, which occurs when $p_{ij} = 0.5$. Thus, we finally obtain:

$$\mathbb{P}\left\{|p_{ij} - \hat{p}_{ij}| \geq t\right\} \leq \frac{1}{4nt^2} \tag{A.9}$$

On the other hand, Hoeffding's inequality (Boucheron et al., 2003) for X_i independent variables, in which X_i belongs to the interval $[a_i, b_i]$ with probability 1, establishes that:

$$\mathbb{P}\left\{|S_n - \mathbb{E}\{S_n\}| \geq t\right\} \leq 2\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}\left(\frac{(b_i - a_i)}{n}\right)^2}\right) \tag{A.10}$$

Given that $X_i \in [0, 1]$ with probability 1, then $a_i = 0$ y $b_i = 1$, and replacing (A.7), we finally obtain:

$$\mathbb{P}\left\{|p_{ij} - \hat{p}_{ij}| \geq t\right\} \leq 2e^{-2nt^2} \tag{A.11}$$

By using (A.9) and (A.11) as the upper bounds for the probability $p^* = \mathbb{P}\left\{|p_{ij} - \hat{p}_{ij}| \geq t\right\}$ together with $p^* \leq 1$, we verify that $c(n_i) \leq p^*$, where n_i is the amount of data that belongs to state I and $c(n_i)$ is defined as:

$$c(n_i) = \min\left\{1, 2e^{-2n_i t^2}, \frac{1}{4n_i t^2}\right\} \tag{A.12}$$

References

Alcántara, A., Galván, I. M., & Aler, R. (2022). Direct estimation of prediction intervals for solar and wind regional energy forecasting with deep neural networks. *Engineering Applications of Artificial Intelligence*, *114*, Article 105128. http://dx.doi.org/10.1016/j.engappai.2022.105128, URL https://www.sciencedirect.com/science/article/pii/S0952197622002573.

Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2020). A systematic review on supervised and unsupervised machine learning algorithms for data science. In M. W. Berry, A. Mohamed, & B. W. Yap (Eds.), *Supervised and unsupervised learning for data science* (pp. 3–21). Cham: Springer International Publishing, http://dx.doi.org/10.1007/978-3-030-22475-2_1.

Anderson, T. W., & Goodman, L. A. (1957). Statistical inference about Markov chains. *The Annals of Mathematical Statistics*, *28*(1), 89–110. http://dx.doi.org/10.1214/aoms/1177707039.

Anuar, N., & Zakaria, Z. (2012). Electricity load profile determination by using fuzzy cmeans and probability neural network. *Energy Procedia*, *14*, 1861–1869. http://dx.doi.org/10.1016/j.egypro.2011.12.1180, URL https://www.sciencedirect.com/science/article/pii/S1876610211046005. 2011 2nd International Conference on Advances in Energy Engineering (ICAEE).

Asan, U., & Ercan, S. (2012). An introduction to self-organizing maps. In C. Kahraman (Ed.), *Computational intelligence systems in industrial engineering: With recent theory and applications* (pp. 295–315). Paris: Atlantis Press, http://dx.doi.org/10.2991/978-94-91216-77-0_14.

Borghei, M., & Ghassemi, M. (2021). Optimal planning of microgrids for resilient distribution networks. *International Journal of Electrical Power & Energy Systems*, *128*, Article 106682. http://dx.doi.org/10.1016/j.ijepes.2020.106682, URL https://www.sciencedirect.com/science/article/pii/S0142061520342277.

Boucheron, S., Lugosi, G., & Bousquet, O. (2003). Concentration inequalities. In O. Bousquet, U. von Luxburg, & G. Rätsch (Eds.), *Advanced lectures on machine learning: ML Summer Schools 2003, Canberra, Australia, February 2 - 14, 2003, TÜBingen, Germany, August 4 - 16, 2003, revised lectures* (pp. 208–240). Berlin, Heidelberg: Springer Berlin Heidelberg, http://dx.doi.org/10.1007/978-3-540-28650-9_9.

Cartagena, O., Parra, S., Muñoz-Carpintero, D., Marín, L. G., & Sáez, D. (2021). Review on fuzzy and neural prediction interval modelling for nonlinear dynamical systems. *IEEE Access*, *9*, 23357–23384. http://dx.doi.org/10.1109/ACCESS.2021.3056003.

Chen, X., Dong, W., & Yang, Q. (2022). Robust optimal capacity planning of grid-connected microgrid considering energy management under multi-dimensional uncertainties. *Applied Energy*, *323*, Article 119642. http://dx.doi.org/10.1016/j.apenergy.2022.119642, URL https://www.sciencedirect.com/science/article/pii/S0306261922009436.

Del Carpio Huayllas, T. E., Ramos, D. S., & Vasquez-Arnez, R. L. (2010). Microgrid systems: Current status and challenges. In *2010 IEEE/PES transmission and distribution conference and exposition: Latin America* (pp. 7–12).

Dominguez, C., Orehounig, K., & Carmeliet, J. (2021). Estimating hourly lighting load profiles of rural households in east africa applying a data-driven characterization of occupant behavior and lighting devices ownership. *Development Engineering*, *6*, Article 100073. http://dx.doi.org/10.1016/j.deveng.2021.100073, URL https://www.sciencedirect.com/science/article/pii/S2352728521000154.

Ferrer-Martí, L., Garwood, A., Chiroque, J., Ramirez, B., Marcelo, O., Garfí, M., & Velo, E. (2012). Evaluating and comparing three community small-scale wind electrification projects. *Renewable and Sustainable Energy Reviews*, *16*(7), 5379–5390. http://dx.doi.org/10.1016/j.rser.2012.04.015, URL http://www.sciencedirect.com/science/article/pii/S1364032112002730.

Greene, D., Cunningham, P., & Mayer, R. (2008). *Lecture notes in applied and computational mechanics*, *Unsupervised Learning and Clustering* (pp. 51–90). http://dx.doi.org/10.1007/978-3-540-75171-7-3.

Guo, L., Wang, N., Lu, H., Li, X., & Wang, C. (2016). Multi-objective optimal planning of the stand-alone microgrid system based on different benefit subjects. *Energy*, *116*, 353–363. http://dx.doi.org/10.1016/j.energy.2016.09.123, URL http://www.sciencedirect.com/science/article/pii/S0360544216313950.

Hafez, O., & Bhattacharya, K. (2012). Optimal planning and design of a renewable energy based supply system for microgrids. *Renewable Energy*, *45*, 7–15. http://dx.doi.org/10.1016/j.renene.2012.01.087, URL http://www.sciencedirect.com/science/article/pii/S0960148112000985.

Jovanović, S., & Hikawa, H. (2022). A survey of hardware self-organizing maps. *IEEE Transactions on Neural Networks and Learning Systems*, 1–20. http://dx.doi.org/10.1109/TNNLS.2022.3152690.

Kangas, J., & Kohonen, T. (1996). Developments and applications of the self-organizing map and related algorithms. *Mathematics and Computers in Simulation*, *41*(1), 3–12. http://dx.doi.org/10.1016/0378-4754(96)88223-1, URL http://www.sciencedirect.com/science/article/pii/0378475496882231.

Karnik, N. N., & Mendel, J. M. (2001). Operations on type-2 fuzzy sets. *Fuzzy Sets and Systems*, *122*(2), 327–348. http://dx.doi.org/10.1016/S0165-0114(00)00079-8.

Kharrich, M., Sayouti, Y., & Akherraz, M. (2018). Microgrid sizing with environmental and economic optimization. In *2018 Renewable energies, power systems & green inclusive economy* REPS-GIE, (pp. 1–6). http://dx.doi.org/10.1109/REPSGIE.2018.8488864.

Khayatian, A., Barati, M., & Lim, G. J. (2018). Integrated microgrid expansion planning in electricity market with uncertainty. *IEEE Transactions on Power Systems*, *33*(4), 3634–3643. http://dx.doi.org/10.1109/TPWRS.2017.2768302.

Khodaei, A., Bahramirad, S., & Shahidehpour, M. (2015). Microgrid planning under uncertainty. *IEEE Transactions on Power Systems*, *30*(5), 2417–2425.

Khosravi, A., Nahavandi, S., & Creighton, D. (2010). Load forecasting and neural networks: A prediction interval-based perspective. In B. K. Panigrahi, A. Abraham, & S. Das (Eds.), *Computational intelligence in power engineering* (pp. 131–150). Berlin, Heidelberg: Springer Berlin Heidelberg, http://dx.doi.org/10.1007/978-3-642-14013-6_5.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*(9), 1464–1480.

Kohonen, T., Kaski, S., Lagus, K., Salojarvi, J., Honkela, J., Paatero, V., & Saarela, A. (2000). Self organization of a massive document collection. *IEEE Transactions on Neural Networks*, *11*(3), 574–585.

Lasseter, R. H. (2002). MicroGrids. In *2002 IEEE power engineering society winter meeting. conference proceedings (Cat. No.02CH37309), vol. 1* (pp. 305–308).

Leary, J., While, A., & Howell, R. (2012). Locally manufactured wind power technology for sustainable rural electrification. *Energy Policy*, *43*, 173–183. http://dx.doi.org/10.1016/j.enpol.2011.12.053, URL http://www.sciencedirect.com/science/article/pii/S0301421511010597.

Lee, K. H. (2005). *First course on fuzzy theory and applications* (1st edition). (p. 335). Berlin: Springer-Verlag, http://dx.doi.org/10.1007/3-540-32366-X_5, (Chapter 5).

Levin, D., & Peres, Y. (2017). *MBK, Markov chains and mixing times*. American Mathematical Society.

Levorato, M., Figueiredo, R., & Frota, Y. (2022). Robust microgrid energy trading and scheduling under budgeted uncertainty. *Expert Systems with Applications*, *203*, Article 117471. http://dx.doi.org/10.1016/j.eswa.2022.117471, URL https://www.sciencedirect.com/science/article/pii/S0957417422008028.

Li, J., Liu, P., & Li, Z. (2020). Optimal design and techno-economic analysis of a solar-wind-biomass off-grid hybrid power system for remote rural electrification: A case study of west China. *Energy*, *208*, Article 118387. http://dx.doi.org/10.1016/j.energy.2020.118387, URL https://www.sciencedirect.com/science/article/pii/S0360544220314948.

Llanos, J., Morales, R., Núñez, A., Sáez, D., Lacalle, M., Marín, L. G., Hernández, R., & Lanas, F. (2017). Load estimation for microgrid planning based on a self-organizing map methodology. *Applied Soft Computing*, *53*, 323–335. http://dx.doi.org/10.1016/j.asoc.2016.12.054, URL http://www.sciencedirect.com/science/article/pii/S1568946617300030.

Llanos, J., Sáez, D., Palma-Behnke, R., Núñez, A., & Jiménez-Estévez, G. (2012). Load profile generator and load forecasting for a renewable based microgrid using self organizing maps and neural networks. In *The 2012 international joint conference on neural networks* IJCNN, (pp. 1–8).

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: Statistics* (pp. 281–297). Berkeley, Calif.: University of California Press, URL https://projecteuclid.org/euclid.bsmsp/1200512992.

Marín, L. G., Cruz, N., Sáez, D., Sumner, M., & Núñez, A. (2019). Prediction interval methodology based on fuzzy numbers and its extension to fuzzy systems and neural networks. *Expert Systems with Applications*, *119*, 128–141. http://dx.doi.org/10.1016/j.eswa.2018.10.043.

McLoughlin, F., Duffy, A., & Conlon, M. (2015). A clustering approach to domestic electricity load profile characterisation using smart metering data. *Applied Energy*, *141*, 190–199. http://dx.doi.org/10.1016/j.apenergy.2014.12.039, URL http://www.sciencedirect.com/science/article/pii/S0306261914012963.

Mendel, J. (2017). *Uncertain rule-based fuzzy logic systems: Introduction and new directions* (2nd ed.). (p. 684). Springer International Publishing, http://dx.doi.org/10.1007/978-3-319-51370-6.

Ministerio de Desarrollo Social (2013). *CASEN 2013 Pueblos Indígenas*: *Technical report*, Ministerio de Desarrollo Social.

Ministerio de Desarrollo Social y Familia (2022). URL http://sni.gob.cl/sector/6.

Mohammadi, M., Hosseinian, S., & Gharehpetian, G. (2012). GA-based optimal sizing of microgrid and DG units under pool and hybrid electricity markets. *International Journal of Electrical Power & Energy Systems*, *35*(1), 83–92. http://dx.doi.org/10.1016/j.ijepes.2011.09.015, URL https://www.sciencedirect.com/science/article/pii/S0142061511002213.

Mumtaz, F., & Bayram, I. S. (2017). Planning, operation, and protection of microgrids: An overview. *Energy Procedia*, *107*, 94–100. http://dx.doi.org/10.1016/j.egypro.2016.12.137, URL https://www.sciencedirect.com/science/article/pii/S187661021631726X. 3rd International Conference on Energy and Environment Research, ICEER 2016, 7-11 September 2016, Barcelona, Spain.

National Renewable Energy Laboratory NREL (2005). *Getting started guide for HOMER version 2.1*. U.S. Department of Energy.

Navarrete, H. (2014). *Caracterización estadística del perfil de uso de baterías para el pronóstico del Estado-de-carga* (B.S. Thesis), Santiago, Chile: Universidad de Chile - Facultad de Ciencias Físicas y Matemáticas, URL http://repositorio.uchile.cl/handle/2250/117108.

Nayak, J., Naik, B., & Behera, H. S. (2015). Fuzzy C-means (FCM) clustering algorithm: A decade review from 2000 to 2014. In L. C. Jain, H. S. Behera, J. K. Mandal, & D. P. Mohapatra (Eds.), *Computational intelligence in data mining - Volume 2* (pp. 133–149). New Delhi: Springer India.

Niez, A. (2010). Comparative study on rural electrification policies in emerging economies: Keys to successful policies. http://dx.doi.org/10.1787/5kmh3nj5rzs4-en, OECD, International Energy Agency, IEA Energy Papers.

Pereira, M. G., Freitas, M. A. V., & da Silva, N. F. (2010). Rural electrification and energy poverty: Empirical evidences from Brazil. *Renewable and Sustainable Energy Reviews*, *14*(4), 1229–1240. http://dx.doi.org/10.1016/j.rser.2009.12.013, URL http://www.sciencedirect.com/science/article/pii/S1364032109003025.

Prahastono, I., King, D. J., Ozveren, C., & Bradley, D. (2008). Electricity load profile classification using fuzzy C-means method. In *2008 43rd International universities power engineering conference* (pp. 1–5). http://dx.doi.org/10.1109/UPEC.2008.4651527.

Richardson, I., Thomson, M., Infield, D., & Clifford, C. (2010). Domestic electricity use: A high-resolution energy demand model. *Energy and Buildings*, *42*(10), 1878–1887. http://dx.doi.org/10.1016/j.enbuild.2010.05.023, URL http://www.sciencedirect.com/science/article/pii/S0378778810001854.

Roje, T., Marín, L. G., Sáez, D., Orchard, M., & Jiménez-Estévez, G. (2017). Consumption modeling based on Markov chains and Bayesian networks for a demand side management design of isolated microgrids. *International Journal of Energy Research*, *41*(3), 365–376. http://dx.doi.org/10.1002/er.3607, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/er.3607, arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/er.3607.

Ruspini, E. H., Bezdek, J. C., & Keller, J. M. (2019). Fuzzy clustering: A historical perspective. *IEEE Computational Intelligence Magazine*, *14*(1), 45–55. http://dx.doi.org/10.1109/MCI.2018.2881643.

Sanchez, I. B., Espinos, I. D., Sarrion, L. M., Lopez, A. Q., & Burgos, I. N. (2009). Clients segmentation according to their domestic energy consumption by the use of self-organizing maps. In *2009 6th international conference on the european energy market* (pp. 1–6).

Serrano-Guerrero, X., Briceño-León, M., Clairand, J. M., & Escrivá-Escrivá, G. (2021). A new interval prediction methodology for short-term electric load forecasting based on pattern recognition. *Applied Energy*, *297*, Article 117173. http://dx.doi.org/10.1016/j.apenergy.2021.117173, URL https://www.sciencedirect.com/science/article/pii/S0306261921006048.

Su, W., Yuan, Z., & Chow, M.-Y. (2010). Microgrid planning and operation: Solar energy and wind energy. In *IEEE PES general meeting* (pp. 1–7).

Taylor, H. M., & Karlin, S. (1984). Chapter 3 - Markov chains: Introduction. In H. M. Taylor, & S. Karlin (Eds.), *An introduction to stochastic modeling* (pp. 67–119). Academic Press, http://dx.doi.org/10.1016/B978-0-12-684880-9.50006-X, URL http://www.sciencedirect.com/science/article/pii/B978012684880950006X.

Ubilla, K., Jiménez-Estévez, G. A., Hernádez, R., Reyes-Chamorro, L., Hernández Irigoyen, C., Severino, B., & Palma-Behnke, R. (2014). Smart microgrids as a solution for rural electrification: Ensuring long-term sustainability through cadastre and business models. *IEEE Transactions on Sustainable Energy*, *5*(4), 1310–1318.

Ultsch, A., & Siemon, H. P. (1990). Kohonen's self organizing feature maps for exploratory data analysis. In B. Widrow, & B. Angeniol (Eds.), *Proceedings of the international neural network conference, vol. 1* (pp. 305–308). Dordrecht, Netherlands: Kluwer Academic Press, URL http://www.uni-marburg.de/fb12/datenbionik/pdf/pubs/1990/UltschSiemon90.

Vallvé, X. (2010). Micro-grid power systems based on renewable energy for rural electrification: Benefits, examples and steps to promote these solutions. *International Hearing on Climate Change an Energy Access for the Poor*, 26–28.

Wang, Z., Chen, Y., Mei, S., Huang, S., & Xu, Y. (2017). Optimal expansion planning of isolated microgrid with renewable energy resources and controllable loads. *IET Renewable Power Generation*, *11*(7), 931–940. http://dx.doi.org/10.1049/iet-rpg.2016.0661, URL https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-rpg.2016.0661.

Widén, J., & Wäckelgård, E. (2010). A high-resolution stochastic model of domestic activity patterns and electricity demand. *Applied Energy*, *87*(6), 1880–1892. http://dx.doi.org/10.1016/j.apenergy.2009.11.006, URL http://www.sciencedirect.com/science/article/pii/S0306261909004930.

Yamaguchi, Y., Fujimoto, T., & Shimoda, Y. (2011). Occupant behavior model for households to estimate high-temporal resolution residential electricity demand profile. In *Proceedings of building simulation 2011: 12th conference of international building performance simulation association*.

Yon, K., Alvarez-Herault, M. C., Raison, B., Khon, K., Vai, V., & Bun, L. (2021). Microgrids planning for rural electrification. In *2021 IEEE Madrid PowerTech* (pp. 1–6). http://dx.doi.org/10.1109/PowerTech46648.2021.9494966.