



Generation of energy demand and PV production profiles based on Markov chains for the design and operation of microgrids

Hugo Radet, Xavier Roboam, Bruno Sareni

► To cite this version:

Hugo Radet, Xavier Roboam, Bruno Sareni. Generation of energy demand and PV production profiles based on Markov chains for the design and operation of microgrids. Electrimacs, May 2022, Nancy, France. hal-03796717

HAL Id: hal-03796717

<https://hal.science/hal-03796717>

Submitted on 4 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Generation of energy demand and PV production profiles based on Markov chains for the design and operation of microgrids

Hugo Radet · Bruno Sareni · Xavier Roboam

Abstract This work provides a simple and straightforward method based on Markov chains to generate a large number of probabilistic energy production and demand profiles when historical measurements are available. The method is intended for energy modelers seeking a simple generation approach to test different design and operation methods for microgrids without spending too much time in this generation phase. Results show that the proposed method can capture the main statistical features and the temporal variability of real data at both long and short time scales, despite the method's simplicity.

1 Introduction

The design and operation of microgrids are challenging, especially because many parameters (e.g., future energy demands, renewable production, electricity tariffs) are inherently uncertain, so their future values cannot be predicted with perfect accuracy when making decisions. On one hand, the design of microgrids under uncertainty might be based on stochastic programming optimization techniques [1] where a large number of scenarios are required. On the other hand, once the size of the assets has been fixed, short-term probabilistic forecasts might be needed by real-time operation strategies to optimize the power flows between the equipment under uncertainty. For instance, look-ahead control strategies solve, at each time step, a multi-stage optimization problem, based on several probabilistic forecasts, each of them associated with a given probability. In both

H. Radet · B. Sareni · X. Roboam
LAPLACE, UT3, CNRS, INPT, Toulouse, France
e-mail: name@laplace.univ-tlse.fr

cases, a large number of data over multi-time scales are essential to accurately solve the problems.

Having said that, decision-makers and modelers often lack appropriate data to run the models, especially in a stochastic context. In many real case studies, no historical data are available or the dataset is of poor quality, over short periods. Therefore, decision-makers might come up with inappropriate design decisions while modelers do not have enough data to assess the design and control approaches they are implementing. To overcome these difficulties, *scenario generation* methods have been widely implemented in the literature [2, 3]. This work mainly focuses on the generation of solar production and energy demands (i.e. electricity and heat) profiles at an hourly time step.

1.1 Literature review

While short-term forecasting is a relatively new topic (late 20th century) driven by efficient real-time operation needs, long-term forecasting for energy systems has been studied for a long time [2]. Indeed, the latter has been used for decades, to anticipate the energy demand growth in order to plan future energy production and transmission infrastructures. However, the recent and strong development of variable renewable energy (VRE) has led to new long-term forecast requirements where short temporal granularity (i.e. at an hourly time step) is needed to cope with the short-time scale variability of the production [4]. Also, as noticed by Hong et al in [2] "*another important step in the recent history, is the transition from a deterministic to a probabilistic point of view*": instead of single values, the output of probabilistic forecasts are probability distributions of the uncertain parameters.

Recently, Mavromatidis et al [3] draw a great review of uncertainty characterization for the design of distributed energy systems, which is of first interest for this work. A large number of methods are documented for both the generation of solar production and energy demand profiles. The readers could refer to this article for an in-depth discussion about the different approaches. The objective of this part is to summarize the main conclusions and provide a clear insight into the direction of this paper. Therefore, the first observation from their review is that the generation method depends on whether or not historical data are available. These approaches can be classified into *top-down* (i.e., historical data are available) and *bottom-up* categories, respectively. While obtaining solar production data is relatively straightforward [5], the availability of energy demand measurements is generally rarer.

In the *top-down* case, the most frequent and easiest generation method is the use of probability distribution functions (PDFs), derived from historical profiles for each hour. Then, a scenario is built by sampling from the PDFs. The drawback of such a method is that the uncertain parameters are treated as independent random variables between consecutive time steps, which might lead to unrealistic behavior where the autocorrelation and periodicity of the initial dataset are lost. To overcome this issue, more sophisticated and hybrid methods have been developed such as autoregressive models [6], Markov approaches [7], and machine learning based methods [8] to name just a few. The latter is probably the most popular approach for both the production and energy demands when large datasets are available [9].

On the other hand, when the case study lacks adequate energy demand measurements (e.g., newly built buildings), model-based methods are usually implemented to generate profiles. The most common approaches are probably the use of ready-made Building Performance Simulation (BPS) tools (e.g., energyPlus [10]), but other model-based techniques are also implemented (e.g., resistance-capacitance (RC) models [11], a stochastic model where the input parameters are characterized based on interview information [12]). More elaborate methods are derived for large-scale districts where the previous approaches might not be appropriate (creating a model for each building of a district is quite laborious...) [13]. In the *bottom-up* case, uncertainty is added to the input parameters of the simulation. The drawback of these methods is that a non-negligible amount of time is usually required to get familiar with BPS tools and collect all the numerous input parameters. Thus, energy modelers who are only seeking a fast generation

method to test their design and operation algorithms might be discouraged by these approaches.

1.2 Main contributions

The main objective of this work is to provide a unique and straightforward method to generate a large number of probabilistic energy production and demand profiles when historical measurements are available. The energy modeler point of view is deliberately adopted in this work. The focus is more on creating a dataset to test different microgrid design and operation algorithms rather than the scenario generation accuracy. Nevertheless, the last section will show that the proposed method can capture the main statistical features and variations of real data despite the method's simplicity. Also, another important aspect is that the generation approach can be used simultaneously to generate long-term scenarios for design and short-term forecasts for operation purposes. Hence, the method is intended for modelers seeking a simple generation approach without spending too much time in this phase.

Therefore, the method implemented in this work is based on Markov chains over representative periods. The model only requires historical measurements of the uncertain time series in order to provide a wide range of contingencies. The methodology is introduced in section 2. Next, the performance of the approach is demonstrated on a residential case study from the Ausgrid dataset in section 3. Finally, the discussion and conclusion are drawn in section 4.

2 Methodology

The uncertain parameters (i.e. energy demands and solar production) are modeled as discrete random variables over a probability space (π, Ω) . The following work aims at providing a method to build the discrete sample space Ω where a scenario is a sequence of all the random variable realizations over a given horizon H , associated with a probability π_s .

2.1 Building the Markov chains

Therefore, as said previously in the introduction, the generation method is based on Markov chains over representative periods. A Markov chain is a stochastic model where the main ingredients are the *states* and the *transition matrix*:

- *States* are observable realizations of the underlying random variables. The finite set of observed states is

called the *state space*. In our case study, they are derived from the energy demand and production measurements.

- The *transition matrix* is a probability matrix where each cell is associated with the probability of going from one state to another.

The first step of the methodology (step 1 in figure 2) is to identify representative periods from the initial annual dataset to account for the different time scales variability. The Markov chains will be later computed over these periods. Therefore, each month of the year is gathered to avoid seasonality issues. Then, for each month, week and weekend days are divided into two classes as the energy demand pattern usually depends on the working activity. Finally, each day is segmented into 23 hourly periods to account for intraday variability. Thus, $23 \times 2 \times 12$ Markov chains will be computed from the historical dataset. The classification of the representative periods is depicted in figure 1. Note that the representative periods are arbitrarily set based on both statistical explorations of the initial dataset and intuitions of the authors. Any other classes could be adopted depending on the nature of the random processes. This latter issue is definitely a weakness of this approach, which is further discussed in the last section.

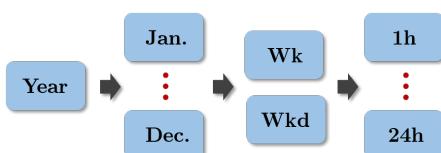


Fig. 1 Representative periods classification to account for the different time scales variability.

Once the representative periods have been identified, a Markov chain is built for each hour where:

- *States* are aggregated (to keep the synchronicity) and normalized vectors of energy demands (i.e., electrical and thermal consumption) and production. To limit the number of state values for each hour, a given number k of relevant states is selected using the k-medoids clustering algorithm [14]. Therefore, each hour of each representative day is represented by $\{1, \dots, k\}$ state values, associated with k vectors of aggregated and normalized power levels (step 2 in figure 2).
 - *Probabilities* are computed based on the transition from one state to another between two consecutive hours. Thus, a 24 hours day is associated with 23 $k \times k$ transition matrices (one for each hour) with k state values for each hour (step 3 in figure 2).

2.2 Scenario generation

Based on the Markov model previously introduced, energy demand and production scenarios are generated by giving an initial state, a timestamp and the length of the horizon. In practice, the power values (which correspond to the Markov chain states) are sampled from the transition matrices where categorical distributions are built for each hour. The probability of each scenario π_s is given by the product of all the hourly probabilities. The generation process corresponds to the last step in figure 2.

3 Evaluation on a case study

The generation method is evaluated using the Ausgrid dataset [15] where the 39th customer is arbitrarily chosen. Figure 3 shows the 3-year time series at an hourly time step for the electrical and thermal demands, in addition to the normalized solar production (in gray). Note that the first hour corresponds to the 1st of July as the season cycle is opposite to Europe.

While well-established metrics (e.g. root-mean-square error (RMSE), mean absolute error (MAE), etc.) are usually derived to assess the performance of short-term forecasting methods, the evaluation of long-term scenarios is less obvious at first glance. Therefore, following [7], [8] and [12] the evaluation for long-term scenarios will be based on a combination of both statistical and visual examination in comparison with the measured data.

3.1 Statistical assessment over the representative periods

To run the evaluation, Markov chains are built from the 3-year dataset of measured data. Then, 1000 scenarios of one year at an hourly time step are generated for the study. A single scenario is plotted in figure 3 with colored lines for comparison. A first general observation is that the shape of the profiles seems consistent with the measured data depicted in gray in the figure. This conclusion is also verified at a lower time scale as depicted in figure 4 and 5. Indeed, the latter show the comparison between the real data and the Markov model for both the week and weekend days of each month. As observed in the figures, it seems that the Markov model correctly reproduces the main statistical features of the initial dataset for each of the representative days (e.g., the model mean values match those of the historical dataset). Furthermore, the seasonal issues are accurately addressed by the model as

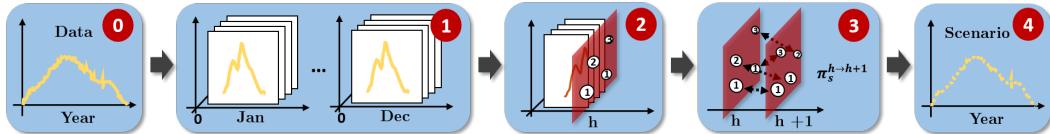


Fig. 2 Description of the scenario generation method based on Markov chains: from historical data (0), days are classified into representatives week and week-end days for each month (1), for each hour, a given number of states is selected using the k-medoids algorithm (2), then the transition matrices based on the probabilities of going from one state to another between two consecutive hours are computed (3) and finally, synthetic scenarios are generated by giving an initial state, a timestamp and the length of the horizon (4).

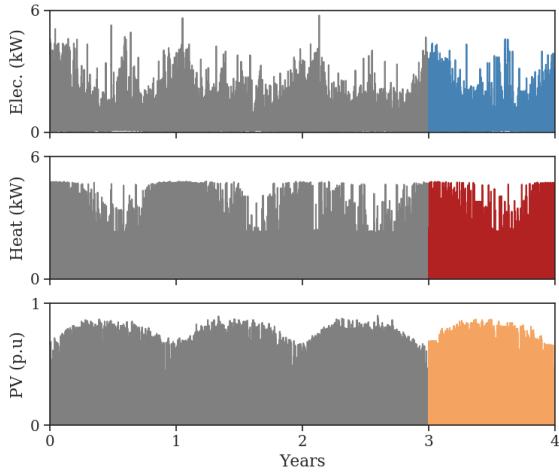


Fig. 3 Overview of the 3-year time series from the 39th Ausgrid customer (in gray) and a one-year scenario generated with the Markov model (in color) for example.

it follows the monthly variations of the real data. This latter observation is reinforced by comparing the power level amplitudes, in addition to the sunrise and sunset times of the different months. Note that for this case study, there are no major differences between the week and weekend days energy demand patterns. This latter observation might not be true with other residential customers.

3.2 Short time scale variability

Despite those statistical similarities, the Markov model still introduces short time scale variability from one scenario to another as shown in figure 6, where the energy demands and production are depicted over one week for 10 scenarios randomly chosen in July. Indeed, power values are not simultaneously the same between scenarios, which leads to a wide range of contingencies. This latter aspect is of first importance when dealing with the design and operation under uncertainties of microgrids. Also, remember that each scenario is associated with a given probability which is computed thanks to the transition matrices (see section 2). Thus, the generation procedure is also suitable for short-term proba-

bilistic forecasts, which can be later used by look-ahead control strategies to operate microgrids.

3.3 Autocorrelation and duration curves

Autocorrelation refers to the correlation of a time series with a lagged copy of itself. The goal is to determine if the signal shows similarities between observations at different time lags. The result is given as a function of the delay (also called *lags* in figure 7). Despite the Markovian property attached to the generation method (i.e. the future state of the stochastic process only depends on the current state, without any memory of the past), the autocorrelation of the three variables is also recovered by the model as shown in figure 7. This might be explained as Markov chains are computed for each hour of representative days, leading to realistic power level sequences.

Finally, figure 7 also shows the *duration curves* of the three variables. With this representation, the values are sorted in descending order, which makes easier the comparison between the real data and the synthetic scenarios at a yearly time scale. The area under the curve corresponds to the total energy consumed (or produced) over the horizon. As shown in the figure, while model peak values are consistent with real data, the Markovian approach tends to generate scenarios with annual energy demands close to the average. Indeed, the model's blue curves are delimited by the real data. This latter observation is not verified for the production profiles. This is probably because the energy demand profiles have redundant patterns, leading to more peaked state probability distributions than for production. Indeed, despite the deterministic meteorological characteristics (i.e. the sunrise and sunset only depend on the position of the earth), the PV production is more likely to have random variations during the day, leading to more diversity in the generation process.

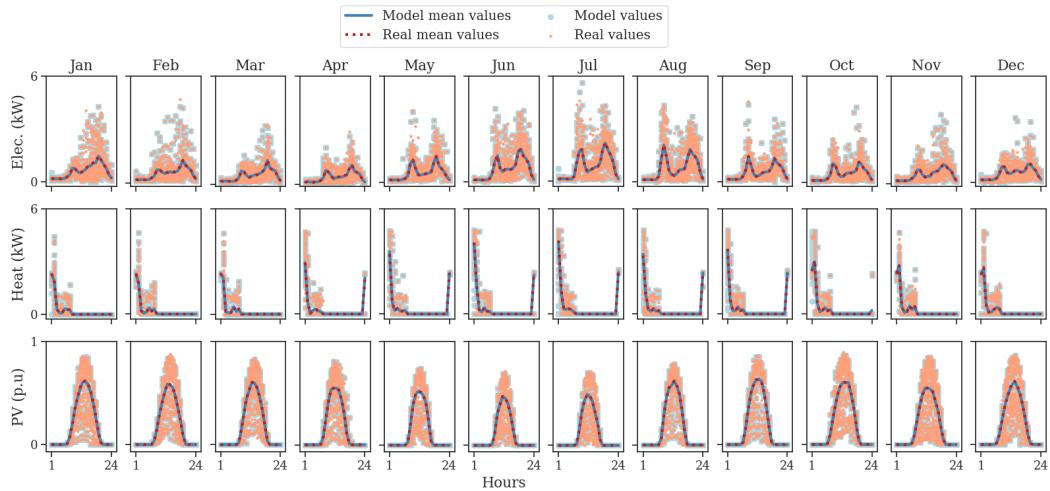


Fig. 4 Comparison between the Markov model (in blue) and the real data (in red) for each **week** day of each month. Mean values are depicted with a solid and dash line for the model and the real data, respectively. All the values are given in the background of each figure for both cases.

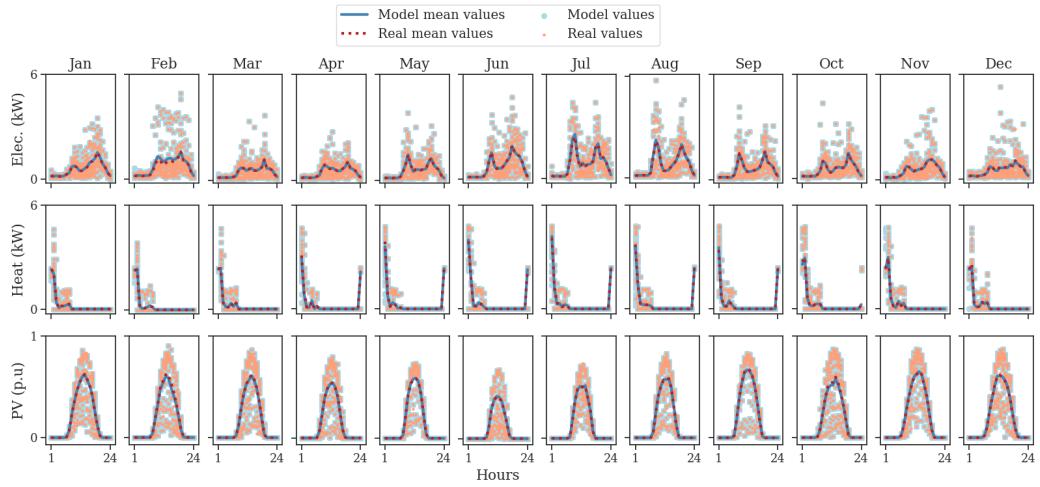


Fig. 5 Comparison between the Markov model (in blue) and the real data (in red) for each **weekend** day of each month. Mean values are depicted with a solid and dash line for the model and the real data, respectively. All the values are given in the background of each figure for both cases.

4 Discussion and conclusion

In order to generate scenarios for both long and short-term applications, a simple stochastic model based on Markov chains was presented in this paper. First, the methodology was introduced where the Markov chains are computed over representative periods to account for the different time scales variability. Then, the method was applied to a residential case study where the objective was to build several energy demands and production scenarios. The results have shown that the main statistical features of the initial dataset have been recovered with this simple Markov model while introducing realistic temporal variability to the annual time series. Finally, the last section has demonstrated that the

Markovian approach is also suitable to generate short-term probabilistic forecasts, later used to control microgrids.

The first limitation of this work comes from the classification procedure manually operated to identify the representative periods. Indeed, the performance of the Markov method is directly related to the expert knowledge concerning the structure and patterns of the initial dataset. Other approaches (mostly based on machine learning as in [8] for instance) do not require this first step and might be more relevant if little information is available about the stochastic processes. Moreover, although the Markov model introduces temporal variability into the scenario set, epistemic uncertainties are not addressed within this approach. Indeed, the model

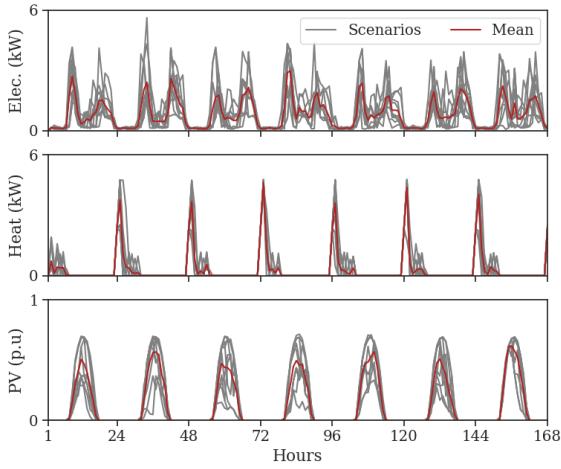


Fig. 6 Short time scale variability over one week for 10 randomly chosen scenarios in July. The mean value is depicted in red.

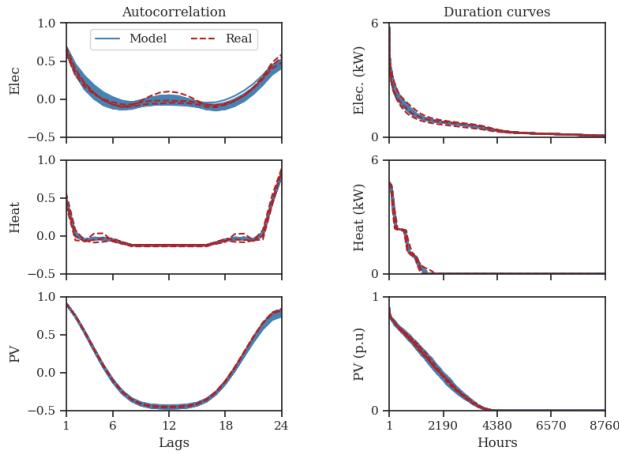


Fig. 7 (Left) Autocorrelation of the three variables, and (Right) Load and production duration curves for both the synthetic scenarios (in blue) and the 3-year historical dataset (in red).

only recovers power levels and daily patterns that were already present in the historical dataset. When applied to a real case study, a strong assumption made by using this method is that the values of the uncertain parameters will remain the same in the future, regardless of their temporal variability. But what happens if the future energy demands increase or if the shape of the daily consumption changes? These latter issues are not properly addressed by only using the Markov model.

This work aimed at developing a simple method to generate a large number of scenarios that will be later used to assess the different design and operation approaches. Decision-makers seeking quantitative and realistic results must spend a significant amount of time towards this generation phase.

Acknowledgements This work has been supported by the ADEME (french national agency on environment, energy and sustainable development) in the framework of the HYMAZONIE project.

References

1. A. J. King, *Modeling with stochastic programming*. Springer series in operations research and financial engineering, New York: Springer, 2012.
2. T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, “Energy Forecasting: A Review and Outlook,” *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 376–388, 2020.
3. G. Mavromatidis, K. Orehoungi, and J. Carmeliet, “A review of uncertainty characterisation approaches for the optimal design of distributed energy systems,” *Renewable and Sustainable Energy Reviews*, vol. 88, pp. 258–277, May 2018.
4. N. E. Koltsakis and A. S. Dagoumas, “State-of-the-art generation expansion planning: A review,” *Applied Energy*, vol. 230, pp. 563–589, Nov. 2018.
5. S. Pfenniger and I. Staffell, “Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data,” *Energy*, vol. 114, pp. 1251–1265, Nov. 2016.
6. K. B. Debnath and M. Mourshed, “Forecasting methods in energy planning models,” *Renewable and Sustainable Energy Reviews*, vol. 88, pp. 297–325, May 2018.
7. S. Patidar, D. P. Jenkins, and S. A. Simpson, “Stochastic modelling techniques for generating synthetic energy demand profiles,” *International Journal of Energy and Statistics*, vol. 04, p. 1650014, Sept. 2016.
8. Y. Chen, Y. Wang, D. Kirschen, and B. Zhang, “Model-Free Renewable Scenari Generation Using Generative Adversarial Networks,” *arXiv:1707.09676 [cs, math]*, Feb. 2018. arXiv: 1707.09676.
9. I. Ghalekhondabi, E. Ardjmand, G. R. Weckman, and W. A. Young, “An overview of energy demand forecasting methods published in 2005–2015,” *Energy Systems*, vol. 8, pp. 411–447, May 2017.
10. D. B. Crawley, C. O. Pedersen, L. K. Lawrie, and F. C. Winkelmann, “EnergyPlus: Energy Simulation Program,” *ASHRAE Journal*, vol. 42, pp. 49–56, 2000.
11. T. Berthou, P. Stabat, R. Salvazet, and D. Marchio, “Development and validation of a gray box model to predict thermal behavior of occupied office buildings,” *Energy and Buildings*, vol. 74, pp. 91–100, May 2014.
12. F. Lombardi, S. Balderrama, S. Quoilin, and E. Colombo, “Generating high-resolution multi-energy load profiles for remote areas with an open-source stochastic model,” *Energy*, vol. 177, pp. 433–444, June 2019.
13. J. A. Fonseca and A. Schlueter, “Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts,” *Applied Energy*, vol. 142, pp. 247–265, Mar. 2015.
14. E. Schubert and P. J. Rousseeuw, “Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms,” in *Similarity Search and Applications* (G. Amato, C. Gennaro, V. Oria, and M. Radovanović, eds.), Lecture Notes in Computer Science, (Cham), pp. 171–187, Springer International Publishing, 2019.
15. E. L. Ratnam, S. R. Weller, C. M. Kellett, and A. T. Murray, “Residential load and rooftop PV generation: an Australian distribution network dataset,” *International Journal of Sustainable Energy*, vol. 36, pp. 787–806, Sept. 2017.