CHAPTER

2

Electrical consumption and renewable profile clusterization based on k-medoids method

Paul Arévalo[1,2], *Marcos Tostado-Véliz*[1], *Jimmy Ayala*[1] *and Francisco Jurado*[1]

[1]Department of Electrical Engineering, University of Jaén, Linares, Jaén, Spain [2]Department of Electrical, Electronics and Telecommunications Engineering (DEET), University of Cuenca, Cuenca, Azuay, Ecuador

2.1 Introduction

The current hydrocarbon-based energy system is becoming less sustainable; the negative impact on the environment, the scarcity of oil, and the high prices of fuels have caused several countries to force the investigation of renewable alternatives to maintain a sustainable energy economy [1]. When considering the general terms toward a green transition, renewable energies are the main alternative such as photovoltaic, wind, hydraulic, etc. [2], especially in developing countries [3]. Within this context, renewable systems with various energy sources can supply off-grid and grid-connected consumers [4]. To provide flexibility for renewable systems, energy costs must be affordable, and the technology must be technically reliable. To achieve this goal, scientists recommend optimizing renewable systems, and the use of size optimization algorithms is becoming more frequent along with short-term economic planning and dispatch [5]. The mathematical models of dynamic systems such as wind turbines and hydraulic turbines are more complex; adding storage systems and economic optimization causes problems that require a significant computational effort [6]. In some cases, the optimization problem is complex that the computational cost potentially rises, causing renewable energies to lose credibility on the part of the user. To solve this problem, it is necessary to reduce the data and variables using cluster techniques to select representative days. In this way, the optimization algorithms will not lose their reference point with respect to the real profiles with optimization costs lower than the initial one.

Sustainable Energy Planning in Smart Grids
DOI: https://doi.org/10.1016/B978-0-443-14154-6.00016-8

2.2 Methods to select representative days

This chapter presents a representative day selection study using the k-medoids method for time series. First, the most used methods in the literature are reviewed in an introductory way. Then, an explanation of the k-medoids method is proposed to present the background as a case study. Finally, the results of the method based on optimization indices are analyzed in detail.

2.2.1 State of the art

The energy planning and sizing optimization of renewable systems are challenging, and the high variability of their sources is complemented by the demand profile and energy rates [7]. Generally, renewable production and electricity demand are not simultaneous, and the integration of energy storage allows this problem to increase the complexity of the system. The precision of the energy model will depend on the measurement interval of the time series. Usually, the resolution time intervals are hourly values for a year; in some cases, values are analyzed in intervals of minutes as in this chapter; and while more time series temporarily are considered, the complexity increases drastically. An alternative to reduce the complexity of the optimization problem is to reduce the number of sampling periods using representative days; this allows replicating the behavior of the system for a year in a few days, significantly reducing the computational effort. In the existing literature, there are several methods to select representative days, such as aggregation methods [8], graphical methods [9], OPT methods [10], and statistical methods [11].

The most common methods used in energy system optimization that have been considered are aggregation methods such as averaging, k-medoids, and hierarchical clustering [8]. For example, the k-medoids method has been used to optimally size photovoltaics-battery systems in smart homes taking into account grid outages and demand response [12]. In this sense, Fig. 2.1 schematically shows the most common methods in the optimization of energy systems.

To briefly explain the methods in Fig. 2.1, the time period will be 1 year with t (1−8760 hours) or 1 day with d (1−365 days) and 1 hour with h (1−24 hours) which can be represented as $\chi'_{c,t}$ or $\chi'_{c,d,h}$ [13].

The average method is based on the averaging of hourly data each month, the average value is calculated with the equation shown in Fig. 2.1, the result has a certain order; however, the aggregation is not based on the similarity of the days which can cause deviations with respect to the original series [14]. The aim of the k-medoids method is to group the original data into a few similar series. For the selection of representative days, the k-medoids method uses the global optimal location problem, calculating the Euclidean distance of each element and using a number of predefined clusters; the objective function is shown in Fig. 2.1 [14]. On the other hand, the OPT method consists of fitting the duration curve of the data obtained from representative periods. This parameter is calculated by means of an error for each interval, the model minimizes the error for all the input series by approximating the reductions to the data originals, and the drawback of this technique is to easily reach the global optimum in a reasonable time [10]. Finally, to solve the problems of

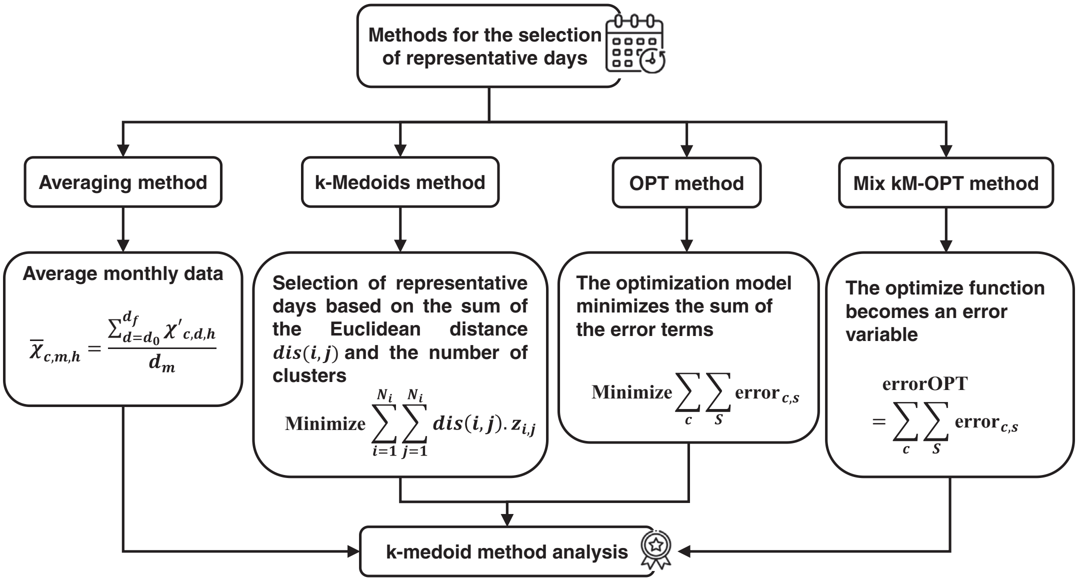

FIGURE 2.1 Optimization of renewable energy systems; methods of determining the most common representative days.

the OPT method of reaching the global optimum, the combination of the k-medoids and OPT method is proposed in the literature, improving the optimization results for systems with multiple renewable sources; in this case, the optimization function becomes an OPT error variable as shown in Fig. 2.1. Correcting this error minimizes the distance of each group component from the medoid and brings it closer to the original series. The more the error is reduced, the more difficult it will be to find the global optimum. Thus a compensation value must be defined without compromising the computational effort [13]. As salient features of the developed approach, within the aggregation methods, k-medoids is considered the most reliable [8,13,14]. In this sense, this chapter analyzes the feasibility of k-medoids method to calculate the representative days for various data series such as solar irradiance and ambient temperature for a year with time intervals of 1 minute. This way, the operating principle of the k-medoids method is explained below.

2.2.2 Representative days using the k-medoids method

To calculate the representative days using the k-medoids method, several steps must be followed. Fig. 2.2 schematically shows the process. First, using forecast information, the scenario space for the uncertain parameters is constructed. Then, these data serve as input variables for the representative days construction algorithm, in which the process is started by selecting the total number of the cluster size (starts $R = 1$), where R is the space corresponding to the representative days. Here, the k-medoids method is applied to reduce the scenario space, and the Bouldin index should be considered until the optimal value is achieved.

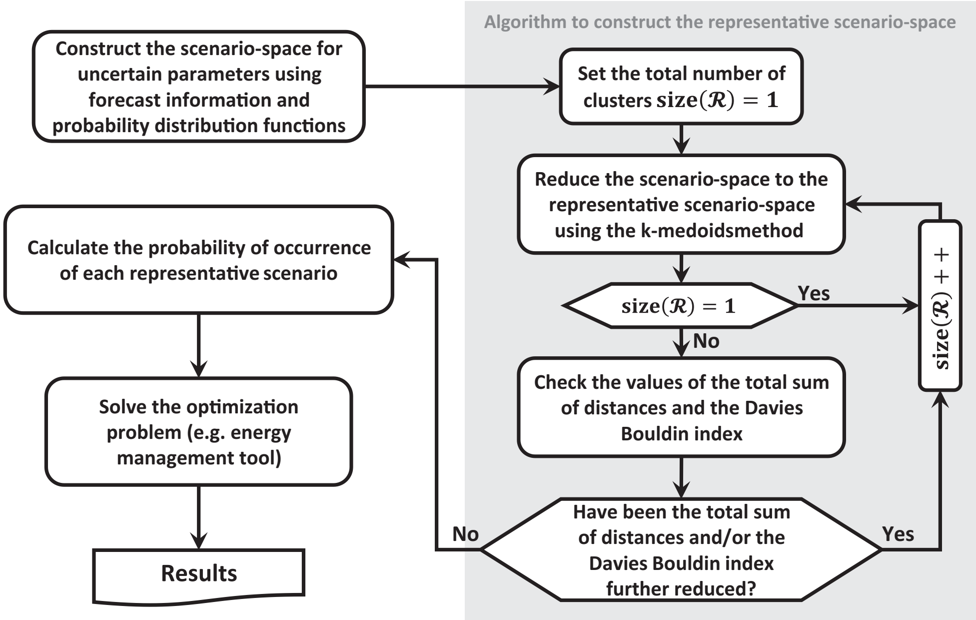

FIGURE 2.2 Schematic representation of the process for the construction of representative days.

To reduce computational time in optimization problems, the k-medoids method is applied to reduce the scenarios from 1 year to a few days. This method has been applied in [12,13,15,16]. Typical values have been selected from solar irradiance and ambient temperature distributions. Cluster configurations are evaluated according to the Davies–Bouldin index and the total sum of the distances for different numbers of clusters. The best number of clusters between both indices is obtained, where the partition being better the lower both indices.

First, if the scale of the time series is different, it is necessary to normalize [15].

$$\chi_{c,t} = \frac{\chi'_{c,t} - \min\chi'_c}{\max\chi'_c - \min\chi'_c}, \chi_{c,t} \in [0, 1] \tag{2.1}$$

Therefore the matrix Ψ containing all the time series (attributes) is defined, where N_h is the number of columns, and N_c is the number of rows corresponding to the number of periods N_i (in this case 365 days). The third step is to define the elements of dissimilarity matrix D, which is composed of the distance (dissimilarity) that is calculated with Eq. (2.2).

$$\mathrm{dis}(i,j) = \left(\sum_{h=1}^{N_h} \left|\Psi_{i,j} - \Psi_{j,h}\right|^2\right)^{1/2} \tag{2.2}$$

The objective function of the k-medoids method is defined by Eq. (2.3).

$$\text{Minimize} \sum_{i=1}^{N_i} \sum_{j=1}^{N_i} \mathrm{dis}\left(i,j\right).z_{i,j} \tag{2.3}$$

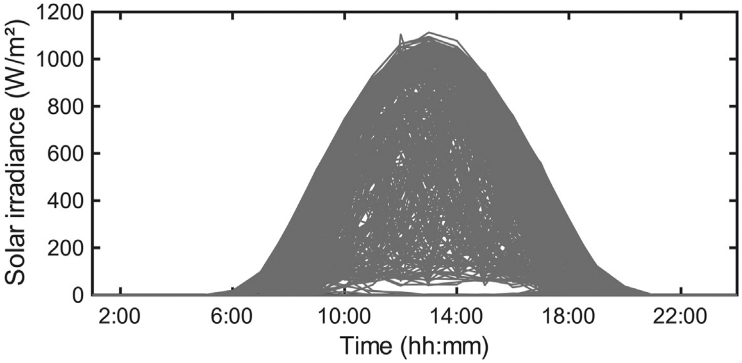

FIGURE 2.3 Solar irradiance in Madrid, Spain, during 2016, resolution (minute).

Constraints

$$\sum_{i=1}^{N_i} z_{i,j} = 1, \forall j \in 1, 2, \ldots N_i \tag{2.4}$$

$$z_{i,j} \le u_i, \forall i \in 1, 2, \ldots N_i \tag{2.5}$$

$$\sum_{i=0}^{N_i} u_i = N_k \tag{2.6}$$

$$u_i, z_{i,j} \in \{0, 1\} \tag{2.7}$$

2.2.3 Background

In order to determine the feasibility of the proposed cluster method, real data of solar irradiance and ambient temperature have been used for the year 2016 in the city of Madrid, Spain [17]; in each case, the amount of data is 525,600 values. Fig. 2.3 shows solar irradiance, and Fig. 2.4 shows ambient temperature. The simulations have been run over a time horizon of 1 year with a time step of 1 minute.

2.3 Results

With the input data collected, in this case solar irradiance and ambient temperature, the resulting profiles are reduced using the k-medoid technique to consider only the most representative. The total number of clusters has been selected based on the Davies–Bouldin index and the total sum of distances. Thus a compromise solution is taken between these two indicators [18]. Fig. 2.5 shows the result of Davies–Bouldin index. In the first place, only the solar irradiance has been considered, and the solution shows the relationship between the sum of the distances and the Davies–Bouldin index. It can be clearly verified that for values greater than 19, the result does not improve. Then, it can be seen that with 19 clusters the lowest Davies–Bouldin index is obtained; therefore 19 clusters are selected.

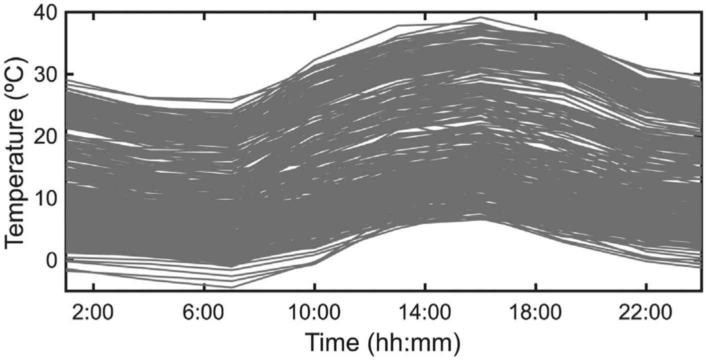

FIGURE 2.4 Ambient temperature in Madrid, Spain, during 2016, resolution (minute).

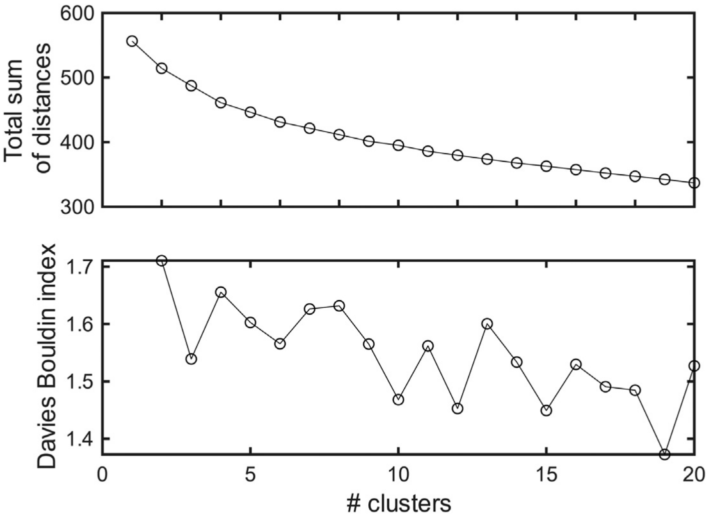

FIGURE 2.5 Total sum of distances with respect to the Davies—Bouldin index.

Website Figure 6 shows the solar irradiance and ambient temperature for the calculated representative days. Using the k-medoids method, the number of days has been reduced from 365 to 19; this means a radical computational optimization.

On the other hand, when considering a series other than solar irradiance, in this case ambient temperature, the pattern of this series is not as unpredictable as the previous one. Therefore, the Davies—Bouldin index is lower; in Website Figure 7, it is evident that for a value of 9 clusters the index is the lowest. Therefore for the data series (ambient temperature) 9 clusters have been selected. The result is shown in Website Figure 8. The difference is remarkable, and 365 values have been reduced to 9.

Applying this methodology, it is possible to analyze more than one profile; in this case, the profile of solar radiation and ambient temperature has been analyzed simultaneously. The results based on the value of the Davies—Bouldin index and the total sum of the distances for different numbers of clusters using k-medoids are shown in Website Figure 9. In this case, if two simultaneous profiles are used, the result of the number of clusters is 19.

The result is greater than the previous case due to the consideration of two data series (solar irradiance and ambient temperature), and the selection of representative days optimizes the data treatment of 730 profiles to only 19.

As outstanding features of the developed approach, solar irradiance and ambient temperature profiles have been obtained simultaneously using the k-medoids method for an optimal number of clusters of 19. The resulting profiles are shown in Website Figure 10. These data could serve as entry to analyze size optimization and energy planning of hybrid renewable systems analyzed for a year. When considering temporary resolutions of 1 minute, during a year 525,600 data each series must be analyzed for each profile, if using the reduction of representative days, this speed is reduced to 12,960 values per minute.

In optimization problems, the amount of input data is essential to obtain accurate results, in this case, by having a large amount of input data due to temporal resolution (525,600 data in each sample) and by considering various combinations and optimization of system size, electricity rates, energy storage, and electricity demand profiles. The solution would have a large computational cost, assuming that the computer does not freeze. By reducing the data with the selection of representative days using the k-medoids method, the computational effort is radically reduced. This is how the authors demonstrate it in [18]. When considering a time window of 365 days, optimization of a system with three electricity tariffs has been analyzed, the resulting size is unaffordable, and the computer (Intel Core i5-9400F 2.90 GHz 8.00 GB RAM under MATLAB® R2019a environment) has thrown an error. Using representative day selection using k-medoids method as in this chapter, the solution has converged to 155.28 seconds using 15 clusters, highlighting the importance of reducing the available data.

2.4 Conclusions

This chapter presents a data reduction technique based on the k-medoids method for a PV system. Average data have been considered every minute for a year, and the time series studied have been solar irradiance and ambient temperature. The results are summarized below.

If the solar irradiance is analyzed, the method reduces the scenarios from 365 to 19 representative days by considering the variability of the original time series, i.e., the presence of clouds, change of seasons, etc.

When analyzing the ambient temperature data for a year, the proposed technique shows a reduction from 365 to 9 representative days. In this case, the difference is greater than the solar irradiance because the ambient temperature does not undergo sudden changes such as solar irradiance, and it only shifts in temperature value during the seasons.

During optimization of real systems, it will be necessary to analyze several time series simultaneously. In this case, solar irradiance and ambient temperature have been analyzed, both variables have 730 values in total, and the k-medoids method has reduced them to 19 representative days.

When considering annual data, the problem could become computationally unsolvable, it has been shown to stop an optimization process of a residential renewable system connected to the grid, considering three electricity tariffs through a computer (Intel® Core i5-9400F 2.90 GHz 8.00 GB RAM under MATLAB R2019a environment), and the problem is unsolvable and produces an error. Using the results using the k-medoids method with 15 clusters, the result converges in 155.28 seconds.

This representative day selection technique is useful for planning and optimization of systems with multiple variables. This chapter could promote the development of complex energy planning systems, with multiple variables and mathematical modeling for dynamic systems.

Acknowledgment

The author (Paul Arévalo) thanks the Call for Grants for the Requalification of the Spanish University System for 2021−23, Margarita Salas Grants for the training of young doctors awarded by the Ministry of Universities and financed by the European Union—Next Generation EU.

The icons used in this document were developed by Freepik, monkik, Smashicons, and Pixel Perfect, from http://www.flaticon.com.

References

[1] K.H. Solangi, M.R. Islam, R. Saidur, N.A. Rahim, H. Fayaz, A review on global solar energy policy, Renewable and Sustainable Energy Reviews 15 (2011) 2149−2163. Available from: https://doi.org/10.1016/J.RSER.2011.01.007.

[2] X. Yuan, C.W. Su, M. Umar, X. Shao, O.R. LOBONŢ, The race to zero emissions: can renewable energy be the path to carbon neutrality? Journal of Environmental Management 308 (2022). Available from: https://doi.org/10.1016/J.JENVMAN.2022.114648.

[3] K. Alford-Jones, How injustice can lead to energy policy failure: a case study from Guatemala, Energy Policy 164 (2022). Available from: https://doi.org/10.1016/J.ENPOL.2022.112849.

[4] IEEE. IEEE Power and Energy Society General Meeting (2010), in: IEEE PES General Meeting, [proceedings], Institute of Electrical and Electronics Engineers, July 25−29, Minneapolis, Minnesota, 2010.

[5] M.Z. Oskouei, A.A. Şeker, S. Tunçel, E. Demirbaş, T. Gözel, M.H. Hocaoğlu, et al., A critical review on the impacts of energy storage systems and demand-side management strategies in the economic operation of renewable-based distribution network, Sustain 14 (2022). Available from: https://doi.org/10.3390/SU14042110.

[6] M. Mustafa, G. Anandhakumar, A.A. Jacob, N.P. Singh, S. Asha, S.A. Jayadhas, Hybrid renewable power generation for modeling and controlling the battery storage photovoltaic system, International Journal of Photoenergy 2022 (2022). Available from: https://doi.org/10.1155/2022/9491808.

[7] K. Tapia-Ahumada, I.J. Pérez-Arriaga, E.J. Moniz, A methodology for understanding the impacts of large-scale penetration of micro-combined heat and power, Energy Policy 61 (2013) 496−512. Available from: https://doi.org/10.1016/J.ENPOL.2013.06.010.

[8] T. Schütz, M.H. Schraven, M. Fuchs, P. Remmen, D. Müller, Comparison of clustering algorithms for the selection of typical demand days for energy system synthesis, Renewable Energy. 129 (2018) 570−582. Available from: https://doi.org/10.1016/J.RENENE.2018.06.028.

[9] J. Ortiga, J.C. Bruno, A. Coronas, Selection of typical days for the characterisation of energy demand in cogeneration and trigeneration optimisation models for buildings, Energy Conversation Management 52 (2011) 1934−1942. Available from: https://doi.org/10.1016/J.ENCONMAN.2010.11.022.

[10] K. Poncelet, H. Hoschle, E. Delarue, A. Virag, W. Drhaeseleer, Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems, IEEE Transactions on Power Systems 32 (2017) 1936−1948. Available from: https://doi.org/10.1109/TPWRS.2016.2596803.

[11] S.G. Chen, Bayesian approach for optimal PV system sizing under climate change, Omega 41 (2013) 176−185. Available from: https://doi.org/10.1016/J.OMEGA.2011.07.013.

[12] M. Tostado-Véliz, D. Icaza-Alvarez, F. Jurado, A novel methodology for optimal sizing photovoltaic-battery systems in smart homes considering grid outages and demand response, Renewable Energy. 170 (2021) 884−896. Available from: https://doi.org/10.1016/J.RENENE.2021.02.006.

[13] E.S. Pinto, L.M. Serra, A. Lázaro, Evaluation of methods to select representative days for the optimization of polygeneration systems, Renewable Energy. 151 (2020) 488−502. Available from: https://doi.org/10.1016/J.RENENE.2019.11.048.

[14] L. Kotzur, P. Markewitz, M. Robinius, D. Stolten, Impact of different time series aggregation methods on optimal energy system design, Renewable Energy. 117 (2018) 474−487. Available from: https://doi.org/10.1016/J.RENENE.2017.10.017.

[15] F. Domínguez-Muñoz, J.M. Cejudo-López, A. Carrillo-Andrés, M. Gallardo-Salazar, Selection of typical demand days for CHP optimization, Energy Build 43 (2011) 3036−3043. Available from: https://doi.org/10.1016/J.ENBUILD.2011.07.024.

[16] S. Swaminathan, G.S. Pavlak, J. Freihaut, Sizing and dispatch of an islanded microgrid with energy flexible buildings, Applied Energy. 276 (2020) 115355. Available from: https://doi.org/10.1016/J.APENERGY.2020.115355.

[17] JRC Photovoltaic Geographical Information System (PVGIS) – European Commission. <https://re.jrc.ec.europa.eu/pvg_tools/en/tools.html>, n.d. (accessed 16.05.22).

[18] M. Tostado-Véliz, S. Mouassa, F. Jurado, A Milp framework for electricity tariff-choosing decision process in smart homes considering 'Happy Hours' tariffs, International Journal of Electrical Power & Energy Systems 131 (2021) 107139. Available from: https://doi.org/10.1016/J.IJEPES.2021.107139.