Fuzzy Partitioning of Photovoltaic Solar Power Patterns

Amr Munshi†
aaamunshi@uqu.edu.sa
† Computer Engineering Department, Umm Al-Qura University, Saudi Arabia

Summary
Photovoltaic systems provide a reliable green energy solution. The sustainability and low-maintenance of Photovoltaic systems motivate the integration of Photovoltaic systems into the electrical grid and further contribute to a greener environment, as the system does not cause any pollution or emissions. Developing methodologies based on machine learning techniques to assist in reducing the burden of studies related to integrating Photovoltaic systems into the electric grid are of interest. This research aims to develop a methodology based on a unsupervised machine learning algorithm that can reduce the burden of extensive studies and simulations related to the integration of Photovoltaic systems into the electrical grid.
Keywords:
Fuzzy partitioning, electric grid, photovoltaic, solar panel

1. Introduction

Among renewable energy resources, solar energy has been continuously an area of interest in research. Photovoltaic (PV) systems are a web that captures solar power to transform it into sustainable energy. These photovoltaics, also known as solar panels, provide a reliable green energy solution. The sustainability and low-maintenance of PV systems motivate the integration of PV systems into the electrical grid and further contribute to a greener environment, as the system does not cause any pollution or emissions.

PV systems use photovoltaic cells that capture solar energy from the sunlight, and converts it into Direct Current (DC) electricity. The reflection of the sunlight creates an electric field across the photovoltaic systems, causing electricity to flow. The DC electricity is then transported to an inverter, which converts the DC power into Alternating Current (AC). This AC power is the type of electricity could feed the electric grid including homes. The power output of PV systems is affected by the levels of solar irradiation and ambient temperature. This results in functionality problems and instability in the power output feeding into the electric grid. Accordingly, integrating PV systems requires extensive studies of lengthy data to promote the stability of the electric grid. However, such studies associated with lengthy data is inefficient and computationally expensive. The utilization of machine learning methods in power related studies has become an area of research interest. Models based on machine learning techniques were used to group electrical load patterns of customers in order to assist tariff formation [1]-[8], short-term forecasting [9], and demand response programs to support management decisions [10]-[12]. Thus, developing methodologies based on machine learning techniques to assist in reducing the burden of studies related to integrating PV systems into the electric grid are of interest. Therefore, this research aims to develop a methodology based on a unsupervised machine learning algorithm that can reduce the burden of extensive studies and simulations related to the integration of PV systems into the electrical grid.

The remainder of the paper is structured as follows. Section 2, discusses the solar irradiance time-resolution and required periods. The background of the unsupervised machine learning technology namely, Fuzzy partitioning is presented in Section 3. The methodology to partition the PV power patterns (PVPPs), while reducing the efforts of the lengthy studies for integrating PV solar power systems into the electrical grid, is presented in Section 4. The experimental results of using Fuzzy partitioning on the PVPP data are shown in Section 4. Section 5 concludes the paper.

2. Irradiance Time-resolution and Periods Required

The time resolution of data should be able to apprehend the fluctuations of the solar irradiance as it affects the power output of PV systems. Moreover, the time resolution has a major role in the accuracy of the simulation studies related to integrating PV systems into the electric grid.

To observe the performance and impacts of PV systems, the time resolution of solar irradiance data must be enough to apprehend the sub-hourly fluctuations of irradiance [13]. Moreover, irradiance data with high sub-hourly time steps will have higher auto-correlation coefficients values than irradiance data with a one-hour time resolution [14]. Fig. 1, shows a comparison between fluctuations in solar irradiance of one hour, 30 minutes, and ten minutes of a day. It can be noticed that the one-hour time resolution does not apprehend fluctuations in solar irradiance during the day. The 30-minute time resolution is able to apprehend fluctuations, but much of the temporal information is lost. On the other hand, the ten-minute time

Manuscript received May 5, 2022
Manuscript revised May 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.5.2

resolution apprehends fluctuation with more temporal information and could potentially provide more accurate results.

Periods when no solar irradiance is available can be removed, as no PVPPs can be obtained during those periods. Fig. 2 shows the daily solar irradiance values for a whole year; it can be noticed that periods from 8:00 PM to 4:00 AM do not contain any solar irradiance throughout the year and may be removed to reduce the dimensionality of the data, accordingly, this will reduce the burden of studies related to integrating PV systems into the electric grid.

3. Fuzzy C-means (FCM) Clustering Algorithm

The Fuzzy C-means (FCM) algorithm is one of the oldest and most ubiquitous fuzzy clustering algorithms developed by Dunn in 1973 and improved by Bezdek in 1981. It allows each data point to belong to more than one cluster with different membership degrees. This method behaves in a similar fashion to the K-means algorithm but each data point has a membership degree with respect to each cluster. Various fuzzy clustering algorithms have been developed based on the optimization and modification of the FCM algorithm [15]. The main objective of the FCM is based on the minimization of the following objective function:

$$J_m(U,C)=\sum_{i=1}^{N}\sum_{j=1}^{C}(\mu_{ij})^m d^2(x_i\ C_j) \qquad (1)$$

where $U = [\mu_{ij}]_{n\times k}$ is a matrix with degrees of memberships of each data point in each cluster and $\mu_{ij} \in [0, 1]$. This implies that the sum of the membership values for each data point on the K clusters must be equal to "1". The function $d(x_i, C_j)$ is the distance between x_i (the ith data point) and C_j (the centroid of the jth cluster), and $m \in [1,\infty)$ is the fuzziness parameter. The selection of these parameters is not an easy task and must be made by experience or by trial-and-error. Fig. 3 shows the basic FCM flowchart.

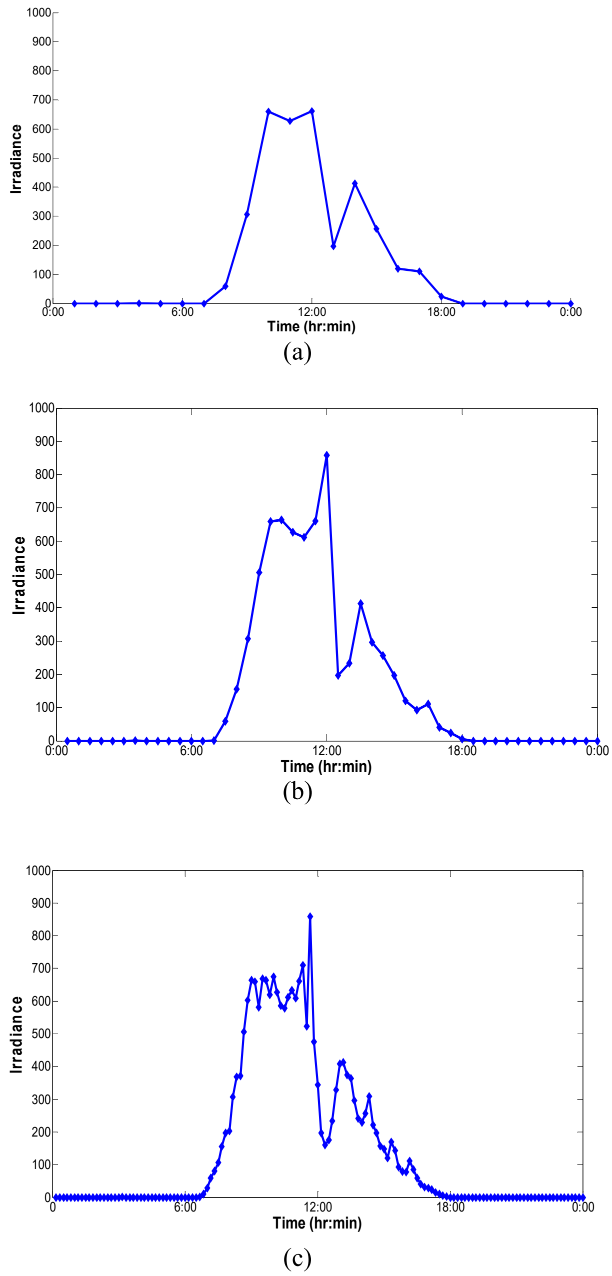

Figure 1: Aprehended fluctuations in solar irradiance for various time resolutiond: a) one-hour b) 30-minutes c) ten-minutes.

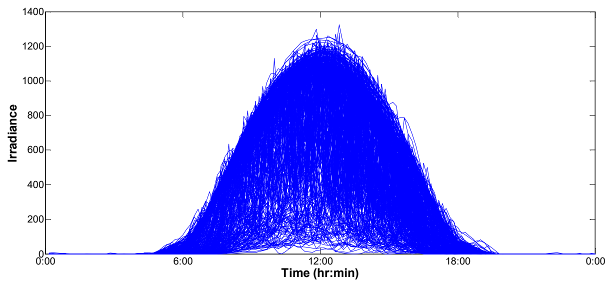

Figure 2: Daily solar irradiance patterns for one whole year.

The steps to perform FCM algorithm are [16]:

1- Select values for the number of clusters K, fuzziness parameter m, a small positive threshold value ε, and a random set of centroids $\boldsymbol{C}$.
2- If $t = 0$, calculate, or if $t > 0$, update the membership matrix U:

$$U_{ij}^{(t+1)} = \frac{1}{\sum_{k=1}^{K}\left(\frac{d(x_i, C_j)}{d(x_i, C_k)}\right)^{\left(\frac{2}{m-1}\right)}} \tag{2}$$

where $i = 1, \ldots, N$ and $j = 1, \ldots, K$

3- Update the fuzzy centers ($\boldsymbol{C}$) by:

$$C_j^{(t+1)} = \frac{\sum_{i=1}^{N} \mu_{ij}^{m} . x_i}{\sum_{i=1}^{N} \mu_{ij}^{m}}, \text{ for } j = 1, \ldots, K \tag{3}$$

4- Repeat steps 2 and 3 until the objective function J_m converges to a local minimum. This means that $\|U^{t+1} - U^{t}\| < \varepsilon$.

Various FCM clustering algorithms have appeared as a result of the utilization of different distance metrics and fuzziness control [17]. It should be noted that the fuzziness parameter m is a positive value greater than “1” and as it increases the fuzziness increases.

Inspired by the effectiveness of neural network self-organization maps, the partitioning of PV solar power systems is obtained by applying such technique on historical data of solar irradiance and ambient temperature time-series observations at a site of interest for previous years. The time-step observations must be able to preserve the short-term instabilities in solar irradiance and ambient temperature. The data is then transformed to daily PV solar power observations. The self-organization maps partition the daily time-series daily PV solar power observations that have comparable observations. A representative from each partition group is achieved. The representative PV solar power is then used in simulation studies from the integration of PV solar power systems into the grid.

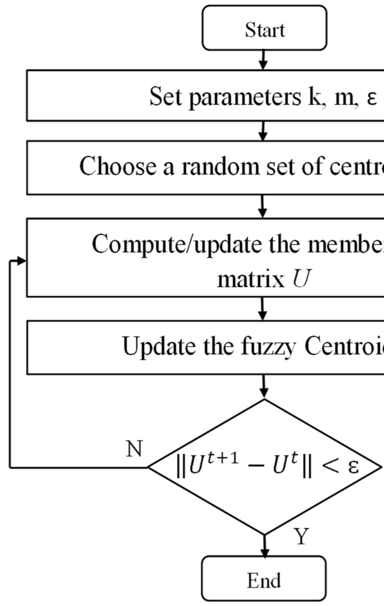

Figure 3: Flowchart of FCM.

3. 1. Preprocessing

The obtained data from weather stations are likely to include outliers due to errors in measurements. The days where irradiance values greater than 1360W/m^2 [18] and night-time periods where irradiance is observed are removed. Also, days where ambient temperature values are unavailable or abnormal are removed.

3. 2. Data Conversion

The irradiance and corresponding ambient temperature data are used in a PV model to calculate the maximum DC power output of the PV system. In order to calculate the AC power generated from the PV system, it should be noted that the DC power generated from a PV array is affected by several factors, such as the power loss due to dust, the power loss due to module parameter mismatch, and the power loss due to the DC current ripple caused by the converter [19].

The AC power output time series of the PV system can be estimated from the irradiance and ambient temperature time series data by the following two steps [19]:

1. The calculation of the DC power (P_{dc}) output generated from the PV system using a suitable PV model and using the PV module data sheet is as follows:

$$T_c = T_{amb} + \left(\frac{NOCT - 20}{0.8}\right).\, S \tag{3}$$

$$P_{dc} = P_{max}\left(S/1000\right)\left[1 - 0.005\left(T_c - 25\right)\right] \tag{4}$$

where T_c is the cell temperature, T_{amb} is the ambient temperature, S is the irradiance, $NOCT$ is the operating cell temperature at Standard Test Conditions (STC: S=1000W/m^2, T_{amb}=25ºC), and P_{max} is the maximum rated power.

The calculation of the AC power (P_{ac}) output generated from the PV system using the manufacturer's efficiency curve is as follows:

$$P_{ac} = P_{dc} \times \eta_{mismatch} \times \eta_{dirt} \times \eta_{inverter} \tag{5}$$

where $\eta_{mismatch}$, $\eta_{dirt,}$ and $\eta_{inverter}$ are array to mismatched modules, dirt loss, and inverter efficiency, respectively.

The AC power output of the PV system for each day at each observation constructs the daily PVPP data.

3. 3. Data Partitioning

For each segment of PVPP time series data, The FCM partitioning methods is used to assign the PVPP into groups, such that PVPPs in the same group are more similar to each other than those in other groups. From each partition, a representative PVPP (centroid) can be obtained. Thus, the set of centroids can be used to represent the whole data set.

3. 4. Validation of Partitioning

The Mean Square Error or Error Function (J) [20] expresses the distance of each data point from its cluster centroid with the same weight values:

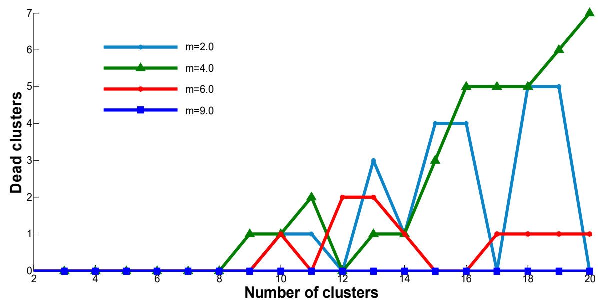

Figure 4: Dead clusters for the FCM method on fall data with m = {2, 4, 6, 9} for two to 20 clusters.

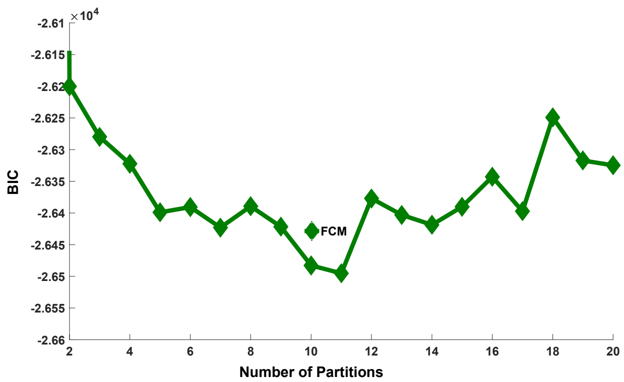

Figure 5: The best results for each grouping for the Fall data set of PVPP for 2 to 20 partitions on the J validy index.

$$J = \frac{1}{N}\sum_{i=1}^{N} d\left(x_i, C_j\right) \quad \text{for } j = 1, \ldots, K \tag{6}$$

The J function is a decreasing function regarding the number of clusters. When it reaches a knee point, the optimum number of clusters can be found.

4. Application of FCM

The methodology was applied on data concerning three consecutive past years (2010-2012) with ten-minute time-steps of irradiation and ambient temperature from the Solar Radiation Research Laboratory [21]. The location of the obtained data has a latitude of 39.74ºN and a longitude of 105.18ºW. The irradiance data with this high time

Table 1: Validity indices of clustering algorithms for eleven partitions of Fall.

Validity index	J
FCM	426.35

Table 2: Compactness, separation, and CPU for eleven partitions on Fall data.

	Compactness	Separation	CPU time(second)		
			Best	Worst	Average
FCM	398.74	800.90	*0.068*	0.383	0.164

resolution (ten minutes) can lead to better accuracy due to the autocorrelation coefficients that will have higher positive values as compared to those obtained for data with lower time resolutions [22]. Thus, the 10-minute time resolution will result in 144 observations per day.

The FCM algorithm was applied with different fuzziness values $m = \{2, 4, 6, 9\}$. The maximum number of epochs is 500 for the four scenarios and the upper limit of weight change between sequential iterations $\varepsilon = 10^{-4}$. The results of all adequacy values improved as the fuzziness parameter increases. For lower fuzziness values $m = \{2, 4, 6\}$, dead clusters were produced, as shown in Fig. 4. Hence, the simulations were applied using FCM with $m = 9$.

Fig. 5 shows the J validity index plot for 2 to 20 partitions of PVPPs. It can be observed from Fig. 5, that eleven clusters can be considered as the optimum number of clusters to represent the Fall PVPPs data. The compactness and separation of the FCM algorithm on the J validity index results showed that the best results for compactness and separation were obtained from FCM at eleven partitions for Fall season (Table 1). It should be noted that in Table 2 the best partition results were chosen based on the J validity index value. From the compactness and separation results, it can be observed that the FCM algorithm with the J validity index presented well partitions of PVPPs and the centroid of each partition can be used as a representative PVPP that can be used in studies for integrating PV systems into the electric grid.

5. Conclusions

This paper attempts at presenting a methodology based on unsupervised machine learning techniques to assist in reducing the burden of studies related to integrating PV systems into the electric grid. A thorough discussion on the solar irradiance time-resolution and periods required to apprehend and obtain the PVPPs was presented. The FCM algorithm was used in partitioning PVPPs. The application pf FCM with the J validity index presented well-partitioning of PVPPs, this improves the results of representatives that can represent the whole PVPP data. The resulted partition representatives from FCM can be utilized in PV system integration studies while leveraging the reliability of PV systems during the integration into the electric grid.

References

[1] G. Chicco, R. Napoli, P. Postolache, M. Scutariu, and C. Toader, "Customer characterization options for improving the tariff offer," *IEEE Trans. Power Syst.*, vol. 18, no. 1, pp. 381–387, Feb. 2003.

[2] V. Figueiredo, F. Rodrigues, Z. Vale, and J. B. Gouveia, "An electric energy consumer characterization framework based on data mining techniques," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 596–602, May 2005.

[3] D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, "Determining the load profiles of consumers based on fuzzy logic and probability neural networks," *Proc. Inst. Elect. Eng., Gener. Transm., Distrib.*, vol. 151, no. 3, pp. 395–400, 2004.

[4] G. Chicco, R. Napoli, F. Piglione, P. Postolache, M. Scutariu, and C. Toader, "Load pattern-based classification of electricity customers," *IEEE Trans. Power Syst.*, vol. 19, no. 2, pp. 1232–1239, May 2004.

[5] G. Chicco, R. Napoli, and F. Piglione, "Comparisons among clustering techniques for electricity customer classification," *IEEE Trans. Power Syst.,* vol. 21, no. 2, pp. 933–940, May 2006.

[6] S. V. Verdu, M. O. Garcia, C. Senabre, A. G. Marin, and F. J. G. Franco, "Classification, filtering, and identification of electrical customer load patterns through the use of Self-Organizing Maps," *IEEE Trans. Power Syst.*, vol. 21, no. 4, pp. 1672–1682, Nov. 2006.

[7] J. Tsekouras, N. D. Hatziargyriou, and E. N. Dialynas, "Two-stage pattern recognition of load curves for classification of electricity customers," *IEEE Trans. Power Syst.,* vol. 22, no. 3, pp. 1120–1128, August. 2007.

[8] G. Chicco, O. M. Ionel, and R. Porumb, "Electrical load pattern grouping based on centroid model with ant colony clustering," *IEEE Power Syst.*, vol. 28, no. 2, pp. 1706-1715, May 2013.

[9] G. Chicco, R. Napoli, and F. Piglione, "Load pattern clustering for short-term load forecasting of anomalous days," *Proc. IEEE PowerTech Conf.*, Porto, 10-13 Sept 2001, vol. 2.

[10] A. Gabaldon, A. Guillamon, M. C. Ruiz, S. Valero, C. Alvarez, M. Ortiz, and C. Senabre, "Development of a methodology for clustering electricity-price series to improve customer response initiatives," *IET Gener., Transm., Distrib.*, vol. 4, no. 6, pp. 706–715, 2010.

[11] G. J. Tsekouras, C. A. Anastasopoulos, F. D. Kanellos, V. T. Kontargyri, I. S. Karanasiou, A. D. Salis, and N. E. Mastorakis, "A demand side management program of vanadium redox energy storage system for an interconnected power system," *Proc. WSEAS EPESE,* Corfu Island, Greece, 2008.

[12] G. J. Tsekouras, F. D. Kanellos, V.T. Kontargyri, I. S. Karanasiou, A. D. Salis, and N. E. Mastorakis, "A new

classification pattern recognition methodology for power system typical load profiles," *WSEAS Trans. Circuits and Systems*, vol. 12, no. 7, pp. 1090-1104, 2008.

[13] G. Vijayakumar, M. Kummert, S. Klein, W. Beckman, "Analysis of short-term solar radiation data," *Solar Energy,* vol.79, pp. 495–504, 2005.

[14] C. Craggs, E. M. Conway and N. M. Pearsall, "Statistical investigation of the optimal averaging time for solar irradiance on horizontal and vertical surfaces in the UK," *Solar Energy*, vol.68, pp. 179–187, 2000.

[15] R. Babuska, *Fuzzy and Neural Control DISC Course Notes*, Delft University of Technology, Netherlands, 2009.

[16] A. V. Gawand, P. Lokhande, S. Daware, and U. Kulkarni, "Image Segmentation for Nature Images using K-Mean and Fuzzy C-Mean," *International Journal of Computer Applications*, 2011.

[17] R. Xu, and D. Wunsch, "Survey of Clustering Algorithms," *IEEE Trans. Neural Networks*, vol. 16, no. 3, pp. 645-678, 2005.

[18] G. Kopp, and J. L. Lean, "A new, lower value of total solar irradiance: evidence and climate significance," *Geophys. Res. Lett*., vol. 38, 2011.

[19] G. M. Masters, *Renewable and Efficient Electric Power Systems*, John Wiley & Son, 2004, pp. 505-531.

[20] G. Chicco, R. Napoli, P. Postolache, M. Scutariu, and C. Toader, "Customer characterization options for improving the tariff offer," *IEEE Trans. Power Syst.*, vol. 18, no. 1, pp. 381–387, Feb. 2003.

[21] Solar Radiation Research Laboratory (BMS) available online at: http://www.nrel.gov/midc/srrl_bms. Retrieved January 2014.

[22] H. Mori, and M. Takahashi, "Application of preconditioned generalized radial basis function network to prediction of photovoltaic power generation," *Proc. IEEE ISGT*, pp.1-6, 14-17 Oct. 2012.