

Western University

Scholarship@Western

---

Electrical and Computer Engineering  
Publications

Electrical and Computer Engineering  
Department

---

10-17-2022

## Agglomerative Hierarchical Clustering with Dynamic Time Warping for Household Load Curve Clustering

Fadi AlMahamid

*Western University*, [falmaham@uwo.ca](mailto:falmaham@uwo.ca)

Katarina Grolinger

*Western University*

Follow this and additional works at: <https://ir.lib.uwo.ca/electricalpub>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Other Computer Sciences Commons](#), and the [Power and Energy Commons](#)

---

**Citation of this paper:**

DOI: 10.1109/CCECE49351.2022.9918481



# Agglomerative Hierarchical Clustering with Dynamic Time Warping for Household Load Curve Clustering

Fadi AlMahamid , Katarina Grolinger

*Department of Electrical and Computer Engineering  
Western University*

London, Ontario, Canada

Email: {falmaham, kgroling}@uwo.ca

ORCID:  0000-0002-6907-7626,  0000-0003-0062-8212

**Abstract**—Energy companies often implement various demand response (DR) programs to better match electricity demand and supply by offering the consumers incentives to reduce their demand during critical periods. Classifying clients according to their consumption patterns enables targeting specific groups of consumers for DR. Traditional clustering algorithms use standard distance measurement to find the distance between two points. The results produced by clustering algorithms such as K-means, K-medoids, and Gaussian Mixture Models depend on the clustering parameters or initial clusters. In contrast, our methodology uses a shape-based approach that combines Agglomerative Hierarchical Clustering (AHC) with Dynamic Time Warping (DTW) to classify residential households' daily load curves based on their consumption patterns. While DTW seeks the optimal alignment between two load curves, AHC provides a realistic initial clusters center. In this paper, we compare the results with other clustering algorithms such as K-means, K-medoids, and GMM using different distance measures, and we show that AHC using DTW outperformed other clustering algorithms and needed fewer clusters.

**Index Terms**—Load Curve Clustering, Agglomerative Hierarchical Clustering, Dynamic Time Warping, Shape-Based Clustering, Demand Response, Energy Management

## I. INTRODUCTION

The estimated number of smart meters in the U.S. and China in 2016 was 166 million [1], which is anticipated to increase as International Energy Agency estimates a 90% rise in power demand by 2040 [2]. The European Union (EU) residential sector represents 31% of the total energy consumption [3]. Smart meters enabled energy providers to collect energy usage data more rapidly and accurately, empowering them to analyze and estimate consumer consumption to provide a more cost-effective and reliable Demand Response (DR) [4]. DR balances energy supply and demand by encouraging consumers to use energy when the power supply is ample and affordable. Energy companies apply various marketing strategies to accomplish DR, commonly through incentive programs that target customers based on their consumption patterns to adapt their consumption in response to the change in the prices [5],

such as applying different tariffs during the day (peak vs. off-peak hours) [6]. Moreover, DR actions contribute to reducing CO<sub>2</sub> emissions where energy production is considered one of the most significant contributors to global warming producing two-thirds of human-induced greenhouse gas emissions [2].

Incentive programs aim to segment customers sharing the same consumption, which requires categorizing customers into groups based on their consumption patterns. Customer consumption can be expressed in a load curve representing the consumption pattern in a given period. Office Buildings might share similar load curves because they follow the same daily work routine. In contrast, residential households' load curves vary from one customer to another, and patterns might even differ within the same customer due to changes in device usage [5], which makes it more difficult to segment customers into groups based on their consumption.

Clustering is an unsupervised learning technique that groups similar data instances together [7] that can be used to segment customers based on their energy consumption. Different clustering approaches are used to cluster customers based on their consumption [4], [6], [8]. Some techniques, in addition to the energy consumption data, include customer data such as customer demography and the number of occupants, then apply significance analysis to reduce the number of features (e.g., Principal Component Analysis (PCA)) and then categorize customers into groups [9], [10]. Other approaches focused on clustering the energy consumption using only daily load curves.

K-means and K-medoids are clustering algorithms based on partitioning: they are used to cluster energy data because of their simplicity [11]. However, the final results depend on the quality of the selected initial cluster centers, which means they might be drawn to a local optimum [12], [13]. Other clustering algorithms based on density, such as DBSCAN and OPTICS clustering results, are sensitive to initial parameters and data density distribution [12].

Recent literature suggests using a shape-based approach to cluster energy load curves [4]–[6]. Agglomerative Hierarchical Clustering [14] is a hierarchical-based algorithm suitable for

This research has been supported by NSERC under grant RGPIN-2018-06222.

arbitrary shapes data. All clustering algorithms, including shape-based ones, use a similarity-distance to construct clusters, e.g., Euclidean, Manhattan, Cosine, and Dynamic Time Warping. While metrics such as Euclidean and Manhattan calculate distances between pairs of points (or features), Dynamic Time Warping (DTW) helps compare load curves based on shape and detect shifts in consumer consumption patterns.

This paper facilitates shape-based clustering while remedying the challenge of poor initial clusters using Agglomerative Hierarchical Clustering with Dynamic Time Warping (AHC-DTW) for residential load clustering. Agglomerative Hierarchical Clustering is used as a bottom-up approach for cluster formation, whereas the DTW similarity metrics allow AHC to consider shapes as opposed to individual points. The results show that using the AHC-DTW algorithm outperformed different clustering algorithms such as K-means, K-medoids, and Gaussian Mixture Models (GMM), resulting in fewer clusters. Furthermore, this paper shows that using DTW as a similarity distance with AHC improves clustering results compared to other distance measures such as Euclidean, Manhattan, and Cosine.

This paper is organized as the following: Section II provides an introduction. Section III discuss the related work. Where Section IV explains how AHC-DTW used to cluster load profiles. Section V discuss the results and Section VI concludes the paper and discuss future work.

## II. BACKGROUND

### A. Dynamic Time Warping

Dynamic time warping (DTW) was used initially to compare speech patterns in speech recognition [15] and was then extended to identify the optimal alignment between two time-series sequences subject to specific constraints [16], as illustrated in Figure 1. The DTW cost matrix (distance matrix) must be computed to calculate the DTW value between two time-series, quantifying the alignment of the two sequences under specific constraints. The closer the DTW value is to zero, the more similar the shapes of the curves are.

Consider two time-series  $X$  and  $Y$ , where  $X = \{x_1, x_2, x_3, \dots, x_n\}$  and  $Y = \{y_1, y_2, y_3, \dots, y_m\}$ , and  $C_\ell$  cost value between two points  $x_i$  and  $y_j$ , where  $x_i \in X$ , and  $y_j \in Y$ . Then,  $W_i$  defines a warping path between  $X$  and  $Y$  as follows:

$$W_i = \sum_{\ell=1}^L C_\ell \quad \forall \quad C_\ell \in W_i \quad (1)$$

Therefore, for all warping paths  $W_{\mathbb{P}} = \{W_1, W_2, \dots, W_p\}$ , DTW is defined as the follows:

$$DTW(X, Y) = \min\{W_{\mathbb{P}}\} \quad (2)$$

Equation 2 defines DTW as the minimum of all warping paths between  $X$  and  $Y$ , where a warping path represents the sum of all distances in a distance-matrix path. The solid red line in Figure 2 shows the minimum warping path between

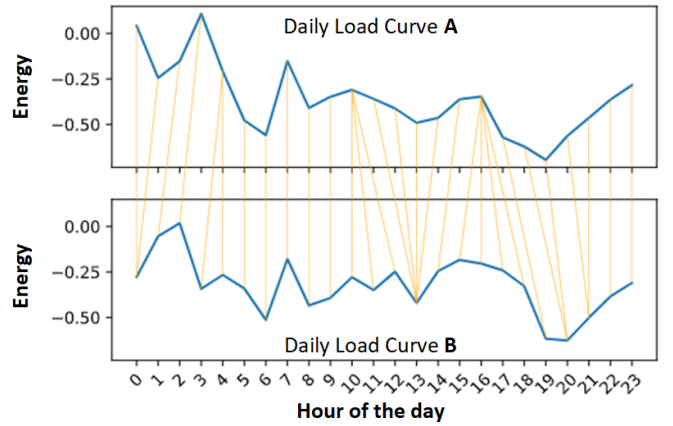


Fig. 1. DTW optimal alignment between two daily load curves showing how DTW aligns the points based on similarity

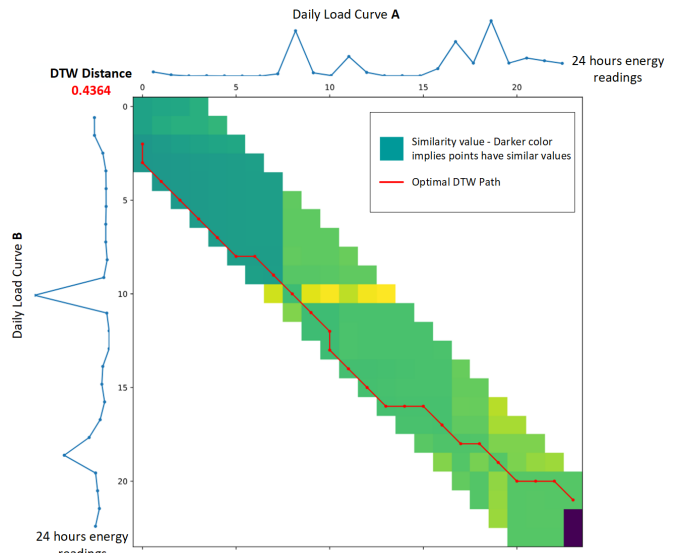


Fig. 2. DTW between two load curves showing the optimal warping path in red line using a window size of four

the two load curves, and the sum of all points in that path represents the DTW value.

DTW is capable of capturing the shift in patterns between two time-series using a defined window size that determines the number of preceding points that need to be considered when aligning a selected point. For example, a DTW with window size of four considers four consecutive points  $y_i, y_{i+1}, y_{i+2}$ , and  $y_{i+3}$  in the curve  $Y$  while aligning a selected point  $x_i$  in the curve  $X$ . On the other hand, a window size of one considers only the corresponding point  $(x_i, y_i)$ , which is equal to computing the Euclidean distance between two time-series as shown in Figure 3.

### B. Hierarchical Clustering

Hierarchical Clustering is a clustering algorithm that iteratively partitions the instances into clusters either using a bottom-up (Agglomerative) or top-down (Divisive) approach

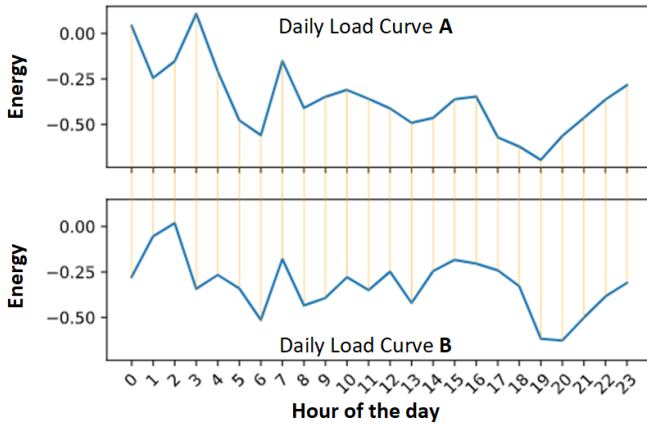


Fig. 3. DTW between two daily load curves using the window size equal to one. This alignment is equivalent to using Euclidean distance

[7]. Agglomerative Hierarchical Clustering (AHC) initially considers each instance as a cluster, and then the clusters are recursively merged until the desired number of clusters is achieved, or all clusters are merged into one cluster. In contrast, Divisive Hierarchical Clustering (DHC) initially merges all instances into a single cluster, and then the clusters are iteratively divided into sub-clusters until the desired number of clusters is achieved.

AHC first computes the distance matrix, which defines the distances between all the instances to be clustered. Then it merges the two instances with the shortest distance, which requires updating the distance between the newly merged instances and all other remaining instances. There are different methods to update the distance matrix, commonly referred to as the linkage criteria, such as:

**Single-linkage clustering** computes a distance matrix between all existing clusters, then merges the two clusters separated by the shortest distance. The newly merged cluster retains the minimum distance between the merged cluster and the remaining clusters.

Let clusters  $A$  and  $B$  have the shortest distance in the distance matrix, and  $K$  represents the remaining clusters. The function  $D(A, B)$  defines the distance between two clusters, and  $C$  represents the new merged cluster. Then distance values are defined as:

$$D(C, K_i) = \text{Min}\{D(A, K_i), D(B, K_i)\} \quad \forall K_i \in K \quad (3)$$

**Complete-linkage clustering**, like single-linkage clustering, computes a distance matrix between all existing clusters, then merges the two clusters separated by the shortest distance. However, in complete-linkage clustering, the new merged cluster retains the maximum distance between the merged clusters and the remaining clusters, as explained in Equation 4.

$$D(C, K_i) = \text{Max}\{D(A, K_i), D(B, K_i)\} \quad \forall K_i \in K \quad (4)$$

**Unweighted average linkage clustering** is also known as Unweighted Pair Group Method with Arithmetic mean (UP-GMA). It is similar to other linkage criteria, except when merging the two clusters separated by the shortest distance, it computes the average distance between the merged clusters and remaining clusters as shown in Equation 5.

$$D(C, K_i) = \text{AVG}\{D(A, K_i), D(B, K_i)\} \quad \forall K_i \in K \quad (5)$$

### III. RELATED WORK

This section focuses on clustering load curves for residential households using a shape-based approach that has been dominant in recent years. Various clustering algorithms have been used with different similarity distances for comparing the similarity between two time series such as Euclidean, Manhattan, Squared Euclidean, Mahalanobis distance, and Dynamic Time Warping. Zhang *et al.* [17] used the K-means clustering with Euclidean distance to cluster 24-hour load curves. Jin and Bi [6] used Affinity Propagation (AP) clustering using DTW as a distance measure, where data points exchange messages between each other until a high-quality set of exemplars and corresponding clusters emanate [18]. Teeraratkul *et al.* [5] used K-medoids replacing Euclidean distance with DTW and produced better quality clusters compared to K-means and Gaussian-based Expectation-Maximization (EM) algorithms. Yilmaz *et al.* [3] considered two approaches and compared them. The first approach considered the shape of the daily profiles using the K-means with standard Euclidean distance to cluster daily load profiles and tried to optimize the  $K$  value using silhouette score [19], while the second approach used specific features to cluster daily profiles. Zhang *et al.* [20] applied hierarchical clustering using DTW to cluster load profiles of a *ground source heat pump system* and determined the best cluster number using the sum of squares of errors, then analyzed the pattern of each cluster class.

Dynamic Time Warping (DTW) distance is used in some literature as a replacement for the Euclidean distance [4]–[6] in clustering load profiles. DTW finds the optimal alignment between two time-series compared to Euclidean distance, which measures the distance between two straight points. The optimal alignment is obtained by stretching or compressing the series' segments [5].

Using a shape-based approach to cluster load curves was not only limited to DTW; for example, Dasgupta *et al.* [4] uses elastic shape analysis to cluster and analyze load curves according to their shapes. On the other hand, Eskandarnia *et al.* [21] developed a framework that uses auto-encoders to perform dimensionality reduction and provide clustering-friendly representations that maintain the original data characteristics before performing clustering using KL divergence.

Our approach is different since it combines Agglomerative Hierarchical Clustering (AHC) with Dynamic Time Warping. While AHC provides realistic cluster centers since it uses actual load curves as initial class centers (prototypes), DTW assists by comparing the curve shapes and finding the optimal

alignment between two curves rather than the straight distance and, therefore, detecting shifts in consumption patterns within predefined sliding-window.

#### IV. METHODOLOGY

As seen in Figure 4, the proposed clustering process includes several steps to achieve high-quality clusters and evaluate the findings. The following subsections discuss each step in more detail.

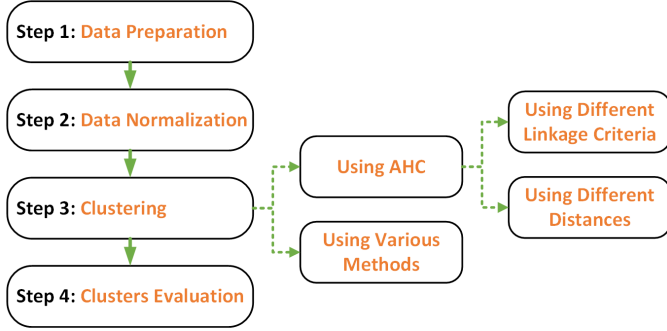


Fig. 4. Clustering Methodology performed with various clustering methods, distances, and linkage criteria

##### A. Dataset Preparation

The dataset contains hourly energy consumption for 19 households captured between 2014 and 2016. The households' data provided by London Hydro (London, Ontario, Canada) has no personal information identifying the customers, only generic values to identify the hourly energy consumption reading per household, e.g., *House#1*.

The data were reformatted so each sample (clustering instance) would represent a daily load curve containing 24 hourly energy consumption readings for a particular house during a specific day of the year. Therefore, the household energy consumption during the day is represented using a *Daily Load Curve* (DLC) as shown in Figure 5.

##### B. Data Normalization

Data normalization is essential for comparing DLCs shapes using DTW for two reasons: 1) DTW is used to compare the time-series shapes, not the actual values, and 2) the differences in the mean or standard deviation (STD) impact any shape similarity. Therefore, Z-normalization in Equation 6 is applied to normalize DLC features.

$$Z_i = \frac{X_i - \bar{X}}{\sigma} \quad (6)$$

Here  $Z_i$  represents the normalized value which is the difference between the actual value  $X_i$  and the mean, divided by the STD. Figure 6 shows the data distribution before and after normalization. Figure 6a shows some outliers and a high STD during most of the day before applying normalization. In contrast, Figure 6b shows that applying normalization reduced the STD and the outliers.

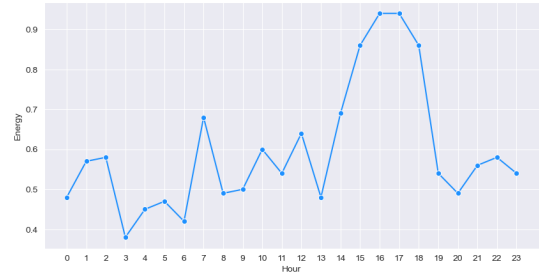


Fig. 5. Daily Load Curve (DLC) using 24 hourly energy readings

##### C. Clustering

DLCs are clustered using different clustering methods AHC-DTW, K-means, K-means++, K-medoids with DTW, and Gaussian Mixture Model with Expectation-Maximization (EM GMM).

AHC was performed utilizing various linking criteria, including complete, average, and single linkage. Furthermore, to determine the influence of DTW on AHC and how it enhances the clustering quality, AHC was performed using several similarity-distances such as DTW, Euclidean, Cosine, and Manhattan.

Despite occasional claims that DTW could add a shape distortion (pinching effect [22]), [4]; DTW is used here to detect similar energy consumption within four hours period, which would not cause DLC shape distortion while finding the optimal alignment as shown in the experiments. Therefore, DTW is strictly applied with  $Window = 4$  and was not optimized for better clustering results. Note that the approach still clusters the daily load curves and this  $Window = 4$  only characterizes the alignment in DTW. The window size of four is an appropriate setting to spot shifts in daily consumption patterns while still not allowing overly high distortions.

##### D. Clusters Evaluation

Several clustering evaluation metrics may be used to assess the quality of clustering, including the Silhouette Coefficient [19] and Dunn's Index [23]. The quality of clusters is determined in this study using the ratio of within-cluster distances to between-cluster distances (WCBCR) [5], [24], as denoted in Equation 7 because it takes into account both the similarity of the instances within each cluster (individual cluster quality) and the dissimilarity between clusters (all clusters quality). WCBCR is calculated as follows:

$$WCBCR = \frac{\sum_{k=1}^K \sum_{X \in C_k} d(X, \mu_k)}{\sum_{i \neq j}^K d(\mu_i, \mu_j)} \quad (7)$$

WCBCR represents the ratio of the sum of the distances between all DLCs ( $X \in C_k$ ) in a cluster ( $C_k$ ) and their corresponding cluster's prototype ( $\mu_k$ ) to the sum between all clusters' prototypes  $d(\mu_i, \mu_j) \forall i \neq j \in K$ . In other words, it measures the ratio between the similarity within clusters to the



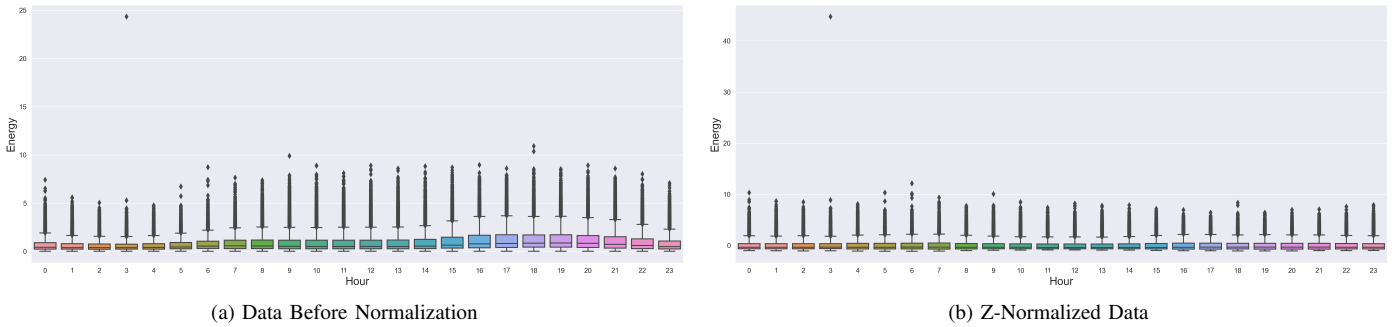


Fig. 6. Data visualization before and after applying normalization shows that applying normalization reduced the STD and the outliers

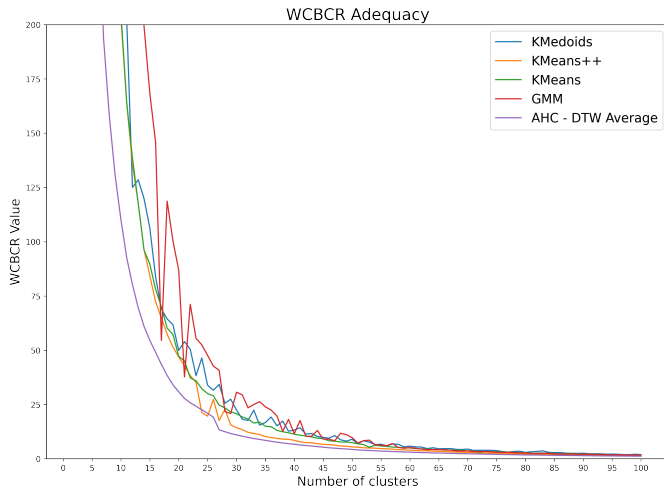


Fig. 7. WBCBR results using different clustering methods. The figure shows that using AHC with DTW and average linkage criteria outperformed other clustering methods

diversity between clusters, where a smaller value of WBCBR indicates a better clustering quality.

Clustering evaluation is a challenging task since the evaluation metrics depend on the similarity distance used for evaluation, and using different distance similarity measures makes it even harder to compare. Therefore, we decided to use the same similarity distance and metric for clustering evaluation to overcome this challenge, i.e., Euclidean distance.

## V. RESULTS AND DISCUSSION

WBCBR is computed using various clusters' numbers for AHC-DTW with average linkage criteria and compared to other clustering methods as shown in Figure 7. The WBCBR value drops as the number of clusters increases for all clustering methods, with AHC-DTW outperforming all other clustering methods by achieving a smaller WBCBR value for the same number of clusters.

Using the Elbow method that utilizes the curve's elbow to decide on the best number of clusters, we can infer that AHC-DTW outperformed other clustering methods, including K-medoids using DTW. AHC-DTW requires fewer clusters (ap-

proximately 27 clusters) while providing a smaller WBCBR value as indicated by the highlighted green cell in Table I.

TABLE I  
WBCBR VALUE USING VARIOUS CLUSTERING METHODS. THE GREEN HIGHLIGHTED CELL PERTAINS THE BEST WBCBR 13.33 FROM AHC-DWT USING AVERAGE LINKAGE CRITERIA WITH THE OPTIMAL NUMBER OF CLUSTERS 27 DETERMINED USING THE ELBOW METHOD.

Clusters #	K-medoids	K-means	K-means++	GMM	AHC-DTW Avg
20	49.89	47.16	46.89	86.93	30.85
21	53.99	45.09	42.96	37.71	27.78
22	50.42	37.25	38.35	71.15	25.81
23	38.28	35.92	35.18	55.47	24.21
24	46.44	32.29	21.31	52.47	22.55
25	33.91	30.05	19.64	47.61	20.98
26	31.59	29.01	27.35	42.70	19.17
<b>27</b>	34.23	24.70	17.65	40.75	<b>13.33</b>
28	25.57	23.45	22.82	21.77	12.47
29	27.43	21.57	15.61	20.82	11.63
30	22.46	20.80	14.31	30.66	11.01

While Table I and Figure 7 compare AHC-DTW with average linkage, Table II and Figure 8 are strictly using AHC with different similarity distances and linkage criteria. The linkage criteria in Table II are labeled as single-linkage **S**, complete-linkage **C**, and unweighted-average labeled **A**. We can conclude that AHC using Complete linkage criteria achieved better WBCBR results when the number of clusters used was less than 23. However, when clusters were more than 26, AHC using the Average Linkage criteria provided better WBCBR values. Furthermore, using the Elbow method, we infer that AHC using Average Linkage requires fewer clusters (27 clusters) with the best WBCBR value of 13.33.

As seen from Table I, for 27 clusters, AHC with Cosine distance achieved the worst performance irrelevant of which linkage criteria were used. Results obtained with Manhattan and Euclidean distances are close to results of DTW for the same linkage criteria. However, using complete linkage, Euclidean distance achieved a slightly better WBCBR value of 15.59 than DTW's value of 15.72. Nevertheless, the best clustering was achieved with DTW with average linkage.

## VI. CONCLUSION AND FUTURE WORK

Energy companies aim to provide an efficient and reliable Demand Response (DR). Therefore, it is imperative to under-

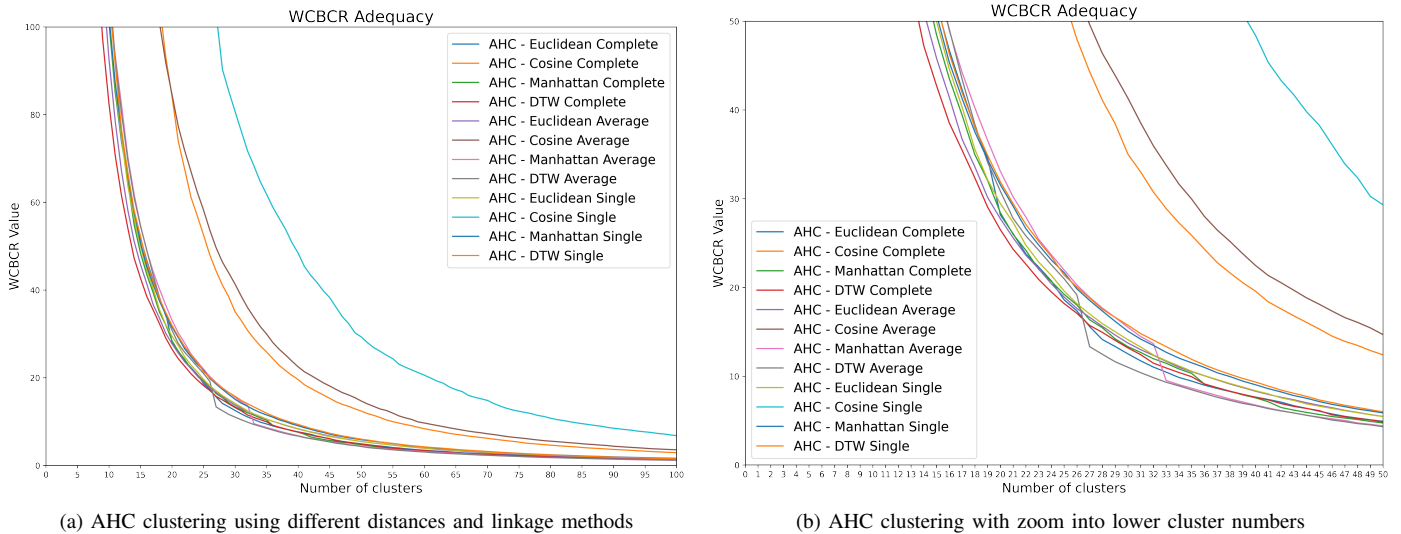


Fig. 8. WBCBR results using AHC with different similarity distances. The figure shows that using AHC with DTW and average linkage criteria outperformed other clustering methods with the optimal number of clusters 27 determined using the elbow method

TABLE II

WBCBR USING AHC WITH DIFFERENT DISTANCES AND LINKAGE CRITERIA: SINGLE-LINKAGE (S), COMPLETE-LINKAGE (C), AND UNWEIGHTED-AVERAGE (A). THE GREEN HIGHLIGHTED CELL PERTAINS TO THE BEST WBCBR 13.33 FROM AHC-DTW USING AVERAGE LINKAGE CRITERIA WITH THE OPTIMAL NUMBER OF CLUSTERS 27 DETERMINED USING THE ELBOW METHOD.

Clustering #	DTW S.	DTW C.	DTW A.	Manhattan S.	Manhattan C.	Manhattan A.	Euclidean S.	Euclidean C.	Euclidean A.	Cosine S.	Cosine C.	Cosine A.
20	31.91	26.51	30.85	31.57	28.43	33.10	29.45	28.18	27.75	221.52	84.15	84.58
21	29.45	24.31	27.78	29.15	25.99	30.20	27.33	26.06	25.60	202.14	73.86	77.27
22	27.10	22.63	25.81	26.62	24.19	27.94	24.77	23.78	23.64	165.57	67.61	71.86
23	25.21	20.91	24.21	24.80	22.05	25.43	22.87	22.31	22.10	139.30	61.18	67.06
24	23.48	19.49	22.55	23.09	20.68	23.67	21.36	20.74	20.41	130.07	56.95	62.36
25	21.39	18.19	20.98	21.67	19.23	21.97	19.60	18.61	18.93	121.26	52.55	58.22
26	20.05	17.09	19.17	19.84	18.02	20.28	18.17	17.36	17.63	113.46	47.97	53.63
27	18.79	15.72	13.33	18.54	16.36	18.90	17.01	15.59	16.64	103.08	44.31	49.62
28	17.55	14.94	12.47	17.34	15.46	17.68	15.92	14.13	15.62	90.27	41.08	46.36
29	16.63	14.04	11.63	16.18	14.21	16.62	15.02	13.35	14.64	85.33	38.45	43.89
30	15.75	13.19	11.01	15.07	13.33	15.52	14.11	12.52	13.74	80.71	35.02	41.33

stand and categorize customers' consumption patterns to help energy companies reduce consumption during peak hours by creating incentive programs to target customers based on their consumption patterns. Targeting the right customer group to reduce energy consumption requires placing the customers into categories based on their similarity of consumption patterns. Grouping residential-households load curves are challenging since load curves vary between customers and within the same customer.

This paper proposed AHC with DTW to categorize DLCs based on shape similarity. AHC provides a better starting point for clusters since each cluster prototype is initiated by an actual load curve, compared to K-means and K-medoids, where initial cluster centers are randomly selected. On the other hand, DTW tries to find the optimal alignment between two curves by detecting shifts in the patterns, making DTW a better measurement distance than other distance measures.

All clustering methods quality was evaluated using WBCBR. The optimal number of clusters was determined

using the Elbow method at 27, with the best WBCBR value of 13.3 obtained from AHC-DTW using average linkage criteria. Future work will examine the optimal number of clusters using Davies–Bouldin index and evaluate results on a large dataset and perform detailed cluster analysis.

#### ACKNOWLEDGMENT

This research has been supported by NSERC under grant RGPIN-2018-06222.

#### REFERENCES

- [1] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3125–3148, 2019.
- [2] L. Sehovac and K. Grolinger, "Deep Learning for Load Forecasting: Sequence to Sequence Recurrent Neural Networks with Attention," *IEEE Access*, vol. 8, pp. 36411–36426, 2020.
- [3] S. Yilmaz, J. Chambers, and M. K. Patel, "Comparison of clustering approaches for domestic electricity load profile characterisation - Implications for demand side management," *Energy*, vol. 180, pp. 665–677, 2019. [Online]. Available: <https://doi.org/10.1016/j.energy.2019.05.124>

- [4] S. Dasgupta, A. Srivastava, J. Cordova, and R. Arghandeh, "Clustering household electrical load profiles using elastic shape analysis," in *IEEE Milan PowerTech*, 2019, pp. 1–6.
- [5] T. Teeraratkul, D. O'Neill, and S. Lall, "Shape-Based Approach to Household Electric Load Curve Clustering and Prediction," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5196–5206, 2018.
- [6] Y. Jin and Z. Bi, "Power Load Curve Clustering Algorithm Using Fast Dynamic Time Warping and Affinity Propagation," *2018 5th International Conference on Systems and Informatics, ICSAI 2018*, no. Icsai, pp. 1132–1137, 2019.
- [7] F. Camastra and A. Vinciarelli, "Clustering methods," *Advanced Information and Knowledge Processing*, no. 9781447167341, pp. 131–167, 2015.
- [8] J. Kwac, J. Flora, and R. Rajagopal, "Household energy consumption segmentation using hourly data," *IEEE Transactions on Smart Grid*, vol. 5, pp. 420–430, 2014.
- [9] Y. Wang, Q. Chen, C. Kang, M. Zhang, K. Wang, and Y. Zhao, "Load profiling and its application to demand response: A review," *Tsinghua Science and Technology*, vol. 20, pp. 117–129, 2015.
- [10] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Transactions on Smart Grid*, vol. 10, pp. 3125–3148, 2018.
- [11] G. Hamerly and C. Elkan, "Learning the k in k-means," *Advances in neural information processing systems*, vol. 16, 2003.
- [12] D. Xu and Y. Tian, "A Comprehensive Survey of Clustering Algorithms," *Annals of Data Science*, vol. 2, no. 2, pp. 165–193, 2015.
- [13] M. E. Celebi, H. A. Kingravi, and P. A. Vela, "A comparative study of efficient initialization methods for the k-means clustering algorithm," *Elsevier Expert systems with applications*, vol. 40, pp. 200–210, 2013.
- [14] F. Murtagh, "A survey of recent advances in hierarchical clustering algorithms," *The computer journal*, vol. 26, no. 4, pp. 354–359, 1983.
- [15] P. Senin, "Dynamic Time Warping Algorithm Review," *Science*, vol. 2007, no. December, pp. 1–23, 2008.
- [16] M. Müller, "Dynamic Time Warping," in *Information Retrieval for Music and Motion*. Springer Berlin Heidelberg, 2007, pp. 69–84.
- [17] Y. Zhang, W. Chen, R. Xu, and J. Black, "A Cluster-Based Method for Calculating Baselines for Residential Loads," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2368–2377, 2016.
- [18] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, 2007.
- [19] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Elsevier Journal of computational and applied mathematics*, vol. 20, pp. 53–65, 1987.
- [20] S. Zhang, L. Zhang, and X. Zhang, "Clustering based on dynamic time warping to extract typical daily patterns from long-term operation data of a ground source heat pump system," *Elsevier Energy*, vol. 249, 2022.
- [21] E. Eskandarnia, H. M. Al-Ammal, and R. Ksantini, "An embedded deep-clustering-based load profiling framework," *Elsevier Sustainable Cities and Society*, vol. 78, 2022.
- [22] A. Srivastava and E. P. Klassen, *Functional and shape data analysis*. Springer, 2016, vol. 1.
- [23] J. C. Dunn, "A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters," *Journal of Cybernetics*, vol. 3, pp. 32–57, 1973.
- [24] G. Tsekouras, F. Kanellos, V. Kontargyri, I. Karanasiou, A. Salis, and N. Mastorakis, "A new classification pattern recognition methodology for power system typical load profiles," *WSEAS Transactions on Circuits & Systems*, vol. 12, no. 7, pp. 1090–1104, 2008.