IEEE Open Access Journal of
Power and Energy

Received 5 May 2021; revised 7 July 2021; accepted 10 July 2021.
Date of publication 15 July 2021; date of current version 9 November 2021.

Digital Object Identifier 10.1109/OAJPE.2021.3097366

A Shape-Based Clustering Framework for Time Aggregation in the Presence of Variable Generation and Energy Storage

NIMA SARAJPOOR [1], LOGAN RAKAI [1] (Member, IEEE), JUAN ARTEAGA [2] (Member, IEEE), AND HAMIDREZA ZAREIPOUR [1] (Fellow, IEEE)

[1] Department of Electrical and Computer Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada
[2] Facultad de Ingenieria, Universidad Panamericana, Zapopan, Jalisco 45010, Mexico
CORRESPONDING AUTHOR: H. ZAREIPOUR (h.zareipour@ucalgary.ca)

This work was supported by the Natural Sciences and Engineering Research Council of Canada.

ABSTRACT A common solution to mitigate the complexity of power system studies is time aggregation. This is to replace the actual data set for all time intervals with representative time periods. Previous research confirms that when energy storage systems are involved in the study, preserving the overall shape of the original data is crucial. This paper proposes a new time aggregation framework to incorporate a shape-based distance to jointly extract representative periods of wind and demand data. The duration curve of the net demand is used as a data-based validation index to compare the performance of the proposed method against other techniques. Also, a 3-bus case study that includes a wind resource, an energy storage system, and two conventional generators is designed. Four model-based validation indices are defined and applied for performance measurement, including the annual operation cost of the system, the annual wind curtailment in the system, the energy throughput of the storage facility, and the daily average of the state of the charge of the energy storage for each 365 days of the year.

INDEX TERMS Power system planning, aggregation, clustering, dynamic time warping, storage.

NOMENCLATURE

$\|C\|$	number of clusters in the cluster set C obtained at the end of a clustering process,
$\bar{C}_k$	centroid of the k^{th} cluster,
$\| C_k \|$	size of the k^{th} cluster in the cluster set C,
$\|.\|$	norm two operator,
π	a warping path between two time series data in dynamic time warping,
C	a set of clusters resulted from a clustering process,
C_k	k^{th} cluster of the set of clusters C,
D	a distance matrix,
$d(.)$	a distance function that measure the dissimilarity of two sequences using a user-defined distance measure,
H_k	total number of timestamps of all the sequences that grouped into k^{th} cluster,
H_S	total number of timestamps in a data set S considering all sequences,
$I_{k,a}$	$I_{k,a}$ is the mean value of index I corresponds to the k-th cluster obtained by using the whole data
$I_{k,c}$	$I_{k,c}$ is the mean value of index I corresponds to the k-th cluster obtained by using the representative data
I_{yr}	I_{yr} is the annual value of index I
n	number of data sets,
N	number of sequences in any data set,
S	any data set,
S_H	a horizontally-configured data set of several data sets,
S_m	m^{th} data set in the joined configuration of the data sets involved in the clustering process,
S_V	a vertically-configured data set of several data sets,
$t_{h,a}^{C_k}$	h^{th} element of duration curve created from the sequences of an original data set grouped into the k^{th} cluster,
$t_{h,c}^{C_k}$	h^{th} element of duration curve created from the sequences of a data set replicated by the representative periods of an original data set, grouped into the k^{th} cluster,
$T_{i,j}^S$	j^{th} timestamp of the sequence T_i of the data set S,

T_i^S	i^{th} sequence of the data set S,
$t_a^{(max)}$	largest value in the duration curve constructed by the original data set,
$t_a^{(min)}$	smallest value in the duration curve constructed by the original data set,
$U; V$	two set of label assignments for the same objects,
U_i	set of objects with label $= i$,
w	warping window size in dynamic time warping,
w_k	number of days grouped into the k-th cluster.

I. INTRODUCTION

THE complexities of practical power system planning problems are often rooted in either a large network size, the presence of multiple uncertainty sources in the operation layer, a long planning horizon, or a combination of these three. Without proper mitigation measures, long-term planning problems are often intractable. To reduce the network size, planners may choose to only include the very high-voltage backbone systems instead of the entire transmission network [1]. To address the curse of dimensionality imposed by multiple uncertain variables, a small number of scenarios are used to represent uncertainty [2]. Model simplification on the time dimension, sometimes referred to as time aggregation, is achieved by using representative time periods in lieu of every single time interval in the planning horizon [3]. Over the past two decades, the use of high performance computing resources have also been experimented to solve large power system planning at a reasonable time [4]. The focus of our paper is on time aggregation in planning problems where both wind energy and electricity demand time series play a key role. The goal is to select a smaller set of data that effectively represents the complete data set.

It is worth noting that scenario reduction is a general term that can be used for any kind of scenarios. In power system studies, for instance, a scenario can be temporal or spatial. Whether to install a transmission line at some specific parts of the network can be considered as a spatial scenario. However, what typical wind power patterns can be observed from a wind farm throughout one (or more) year is a scenario reduction problem associated with time. Time aggregation can be considered as a subset of temporal scenario reduction. If the actual historical data are considered as the scenarios, time aggregation via clustering can be analogized to scenario reduction on the time horizon. However, if some scenario generation is used to create multiple scenarios for each day or time window, then time aggregation may not be the proper wording.

With the growing trends in energy storage integration, more and more power system operation and planning studies contain storage systems. The shape of fluctuations of wind (and similarly solar) energy resources, the sequences of "on" and "off" periods, and what happens in between has a direct impact on storage operation and planning decisions [5]. Similarly, system loading and availability of transmission corridors play a role in effectively utilizing hybrid energy systems (e.g., wind plus storage) [6]. Thus, when aggregating power system data for studies that include wind and energy storage, the representative data needs to reflect the overall shape of both wind energy and system demand variability.

In the past, representative patterns were selected heuristically by selecting a number of days as the representative periods. To capture the seasonality, four different daily profiles are often considered [7]. This approach is prone to inconsistency depending on the preferences of the decision maker, and there is no guarantee that those selected periods are the best representation of the entire data. In recent years, however, various data mining tools are applied for time aggregation in the power system community [5], [8]–[10]. Baringo and Conejo employed K-Means clustering to select representative hours out of a large data set composed of wind energy and demand time series [8]. In order to maintain intra-hour information, Qiu *et al.* used the same clustering tool as in [8] to discover daily patterns in wind and load data [9]. The authors used the Euclidean Distance (ED) measure with K-Means clustering. However, in contrast to Dynamic Time Warping, Euclidean Distance might not be able to capture the shape of patterns [10]. To mitigate this, Liu *et al.* [10] proposed a hybrid clustering process incorporating Dynamic Time Warping (DTW) distance. DTW is a shape-based distance measure that is commonly used for measuring the similarity between two time series sequences [11]. By concatenating the time series data of several data sets, including electricity demand, wind power, and solar generation, for each individual day, the authors constructed a long time series for each day and fed it into the clustering process. As discussed in the next section of this paper, horizontally concatenating the data sets can result in the mismappping of timestamps and it may mistreat the joint behavior between the wind power and electricity demand data. Teichgraeber and Brandt [5], on the other hand, proposed to use k-shape instead of DTW. K-shape [12] defines similarity of two time series as the highest cross-correlation that can be obtained by allowing one of the sequences being shifted against the other. Focusing on electricity market price data, they conclude that k-shape performed better compared to DTW distance when the shape of representative periods matters. However, it was observed that electricity market price time series, unlike wind energy, show a great deal of seasonality and cyclic behavior [13]. In this paper, we analyze both k-shape and DTW distances and discuss their challenges when applied to wind energy data. As shown in this paper, the proposed k-shape method, however, cannot preserve the shape of patterns when a volatile data such as wind power is involved.

Recently, some clustering methods on spaces other than the input (i.e., raw data) has been proposed [14]–[18]. In these studies, the impact of each series on a particular output is considered as a feature. Then, the feature was used to cluster the data. Since the clustering is being conducted directly on the output space, there is no doubt that the aggregated output is closer to the actual value of that particular output. However, such approach cannot guarantee to provide a meaningful representation on the behavior of different parts of a network and different outcomes such as wind curtailment or the behavior of the energy storage where the shape of patterns matters [5]. In addition, the wind power and electricity demand data have intrinsic structure of clusters. Therefore, it is reasonable to cluster them into a certain number of groups. On the other hand, to the best of our knowledge, the papers that proposed a clustering method based on a feature on the output space did not provide a proof to show that the transformed data still have a clustered structure. Furthermore, the clustering on the output space can lead to an over-fitting problem. In other words, any small deviation in the data (e.g. a measurement error at some timestamps) can considerably impact the final result. Clustering the raw data, however, is more robust and avoids the over-fitting problem.

Some previous studies have proposed methods for enhancing the representative days selected by a clustering method, either by including some extreme cases to the set of representative days [19] or using methods to reduce the loss of information after choosing the representative days [20]. While the focus of these works is to modify an existing clustering method, ours is the clustering framework itself that refines the whole set of selected representative days. Therefore, the aforementioned studies can be mounted on top of this framework to further improve the outcomes of time aggregation.

Even though several clustering methods have been reported in the literature for power system time aggregation, no conclusive study exists that compares the performance of alternative techniques for aggregating wind power data for power system operation models that includes energy storage systems. This paper is an attempt to fill this gap.

This paper proposes a novel shape-based clustering framework for time aggregation that aims to not only preserve the shape of both wind power and system demand data, but also preserves their joint behavior. In contrast to the previous works, the present paper proposes to construct the time series in a multi-dimensional format [21] for the purpose of finding the representative days in the power system studies. The proposed approach tries to group similar patterns of the wind power and electricity demand of the days that are changing in the same direction. Therefore, the *joint* behavior of the wind power and electricity demand are preserved when DTW is in use. The dependent DTW can be applied to the vertically structured data to measure the distance between two time series. To the best of our knowledge, this the first time that the joint behavior of sequences are considered in selecting the representative days.

Furthermore, the proposed framework also enhances the quality of clusters by mitigating the chance of misclustering. It is important to note that, regardless of the structure of the data, DTW distance might result in some extreme matches between sequences. In other words, it might match two sequences that have considerably different values. This is a misclustering case that should be avoided. This challenge has not been addressed nor tackled properly in the literature. This paper proposes to use the consensus clustering [22] together with the mutual information to resolve the aforementioned issue. The consensus clustering is first used to discover a set of structure-free clusters, being referred to as neighborhood partitions. Next, the mutual information score between the neighborhood partitions and the set of clusters discovered by the dependent DTW can be used to measure the misclustering. A higher mutual information score corresponds to a set of clusters with lower misclustering and higher quality.

To summarize, the main contribution of this paper is to propose a new time aggregation framework that (i) uses *dependent* DTW distance to group days with similar *joint* behavior; and, (ii) incorporates consensus clustering and mutual information to mitigates the chance of misclustering; and therefore, enhances the quality of clusters. Alternative methods are examined for simultaneously clustering the wind power and electricity demand data. To demonstrate the merits of the proposed method, both data-driven and model-based validation indices are used in this paper.

The rest of this paper is organized as follows. Section II provides a quick review on the clustering process. In Section III the proposed clustering process is explained in detail. In Section IV different indices are explained for measuring the performance of the clustering methods, followed by the numerical results. Section V concludes the paper and future work is discussed.

II. BACKGROUND

In this section, we review the steps of the data clustering process. We can look at the clustering problem from three points of view: (1) clustering subsequences in the streaming data, (2) updating clusters prior to performing a close-to-real-time task such as forecasting, or (3) clustering a given data in off-line mode (i.e., the whole data set is already provided and the goal is to extract salient patterns). In regard to clustering the subsequences in the streaming data, E. Keogh *et al.* showed that clustering of a streaming time series often does not result in meaningful clusters [23]. To forecast a particular day, one can cluster similar days and then build a prediction model for each cluster. The cluster might be updated as time goes on. In the third case which is the one considered in this paper, we assume that the complete data set is already available, and the goal is to choose a proper set of representative days to properly replicate the behavior of data to make the planning study tractable. In other words, the purpose of time aggregation is to represent the original time series by a smaller set of representative sequences [24].

We also discuss the challenges of shape-based clustering when applied to aggregating data sets including wind power data in power system studies. In particular, we investigate the challenges of two shape-based clustering methods that incorporate DTW distance and K-Shape.

A. STEPS OF THE CLUSTERING PROCESS

A clustering process generally consists of the following steps: (1) scaling the data set, (2) measuring the distance, (3) grouping sequences, and (4), choosing the representative sequences. These steps are reviewed next.

1) SCALING DATA SETS

Scaling is a pre-processing step where features of data are normalized. The purpose of scaling is to prevent the domination of some features over the others [25]. Min-Max Scaling and Standardization are among the common scaling methods that have been used in the field of machine learning [26]. The standardization technique is more effective in reflecting the extreme values, which is usually of significance in power system studies for obtaining a feasible yet realistic solution. Therefore, in this paper, we only consider the standardization technique.

There are three approaches for applying the standardization to a data set S [5]. First, the data-wise standardization model can be built by considering all elements of a data set at once. In other words, the goal is to bring the mean and standard deviation of all elements included in the data into 0 and 1, respectively. Second, the time series standardization can be applied onto each individual time series T included in S. The goal is to bring all sequences to the same mean of 0 and standard deviation of 1. Finally, the element-wise scaling process, where the t_i element is treated as the i^{th} feature of T and the standardization is applied onto each i^{th} feature across all the sequences included in the data set.

2) DISTANCE MEASURES

The two common approaches in measuring the similarity of time series sequences are Euclidean and DTW distance measures. The Euclidean distance treats the two time series as two points in the Euclidean Space and considers the length of the line segment connecting the two points as their distance. In contrast to ED, DTW distance provides an elastic approach that allows one element of a sequence to be mapped onto one or more elements of another sequence. This particular feature can match two sequences if one of them is just a stretched or compressed version of the other. The DTW distance between two sequences T_i and T_j of a data set S, can be calculated as follows:

$$DTW(T_i, T_j) = \min_{\pi} \sqrt{\sum_{(r,s)\in\pi} (T_{i,r} - T_{j,s})^2} \tag{1}$$

The distance measure (1) allows expansion and compression of sequences, with no restriction, which may not be desirable in power system studies [5]. A warping window, w, is usually used to limit the amount of expansion and/or compression that is allowed for the alignment [27], i.e., $|r - s| \leq w$.

3) GROUPING THE SEQUENCES

After defining the dissimilarity, the distance between sequences can be calculated. The goal is to group similar sequences together. Common methods for grouping the sequences are briefly explained next.

a: K-MEANS

A partitioning algorithm that tries to create a set of clusters where the square root of the total sum of distances of observations to their corresponding centroids is minimized [28], using the formula below:

$$\min \sum_{k=1}^{|C|} \sum_{T\in C_k} d(T, \bar{C_k})^2 \tag{2}$$

where, the function d(.) is the distance function being utilized in the K-Means clustering.

To run this algorithm, an initial guess of centroids needs to be provided. A drawback of this technique is that different initializations can lead to different outcomes. To mitigate this issue, a K-Means++ algorithm is used to make a better initial guess [29]. Furthermore, if there are clusters with unbalanced sizes, the performance of K-Means will be decreased [30].

b: K-MEDOIDS

In contrast to K-Means, K-Medoids considers one of the members of each cluster as its centroid and it is more robust to the outliers [31]. The objective is to minimize the sum of distances from each centroid to their corresponding cluster's members, as follows:

$$\min \sum_{k=1}^{|C|} \sum_{T\in C_k} d(T, \bar{C_k}) \tag{3}$$

Since the clusters can be achieved by only distance matrix, it can be associated with any distance measure.

c: HIERARCHICAL CLUSTERING

A partitioning algorithm in which each individual observation is considered as one cluster in the beginning of the grouping process. Then, at each step of the clustering process, the two closest groups will be merged to form a bigger group. The algorithm ends when all sequences have been merged into a single group, or the number of clusters reaches a user-defined value. To gauge the similarity between groups, a measure known as linkage, is required. The four common linkages are ward, single, complete, and pair-wise linkage. Except for the first one, i.e., ward linkage, the other linkages can lead to clusters with no clear centroids. In addition to the aforementioned linkages, the min-max linkage has recently been used [10].

In this paper, all three grouping methods have been considered. For the hierarchical clustering, the only linkages that are examined in this paper are ward, associated with Euclidean distance, and min-max, associated with DTW distance, due to the fact that a clear centroid can be considered for their clusters. The first method is referred to HC-Ward and the second method is referred to HC-minmax in this paper. These two linkages are calculated as follows:

$$L_w(C_i, C_j) = \sum_{T \in C_{ij}} d(T, \bar{C}_{ij})^2 - \sum_{T_i \in C_i} d(T_i, \bar{C}_i)^2 - \sum_{T_j \in C_j} d(T_j, \bar{C}_j)^2 \quad (4)$$

$$L_m(C_i, C_j) = \min_{T_i \in C_{ij}} (\max_{T_j \in C_{ij}} d(T_i, T_j)) \quad (5)$$

where, $L_w(C_i, C_j)$ and $L_m(C_i, C_j)$ are, respectively, the ward and min-max linkages between the i^{th} and j^{th} cluster, and C_{ij} is a group that includes the members of both C_i and C_j.

4) CHOOSING REPRESENTATIVE SEQUENCES

This last step of clustering is optional as one may have no interest in finding the centroids. However, for the purpose of this paper, choosing the proper set of representative periods is necessary. A simple average of sequences grouped as one cluster is not desirable when Euclidean distance is in use. This is because the averaging will smooth the sequences and cannot properly reflect the volatility of the patterns [32]. To avoid this issue, one can choose medoid or close-to-average member. Both of these approaches have been considered in this paper.

B. CHALLENGES TO DTW AND K-SHAPE

DTW distance is the most commonly used shape-based distance measure in analyzing time series data [11]. K-shape is a parameter-free clustering method that shows a comparable performance to that DTW [12]. Both of these methods have previously been applied in power system studies [5], [7]. DTW and K-shape have been used in the power system community for the purpose of clustering [5], [10]. However, there are some caveats when applying these two approaches, which are discussed here.

1) DTW

DTW distance can consider expansion and/or compression [5]. However, it is recommended to normalize the sequences in a time series-wise manner [33], otherwise, the true similarity of the sequences may not be revealed as DTW is sensitive to noises. To show the difference, three toy time series sequences are created as follows: $x = \{0, 0, 1, 1\}$, $y = \{0.05, -0.1, 1.1, 0.95\}$, and $z = \{0, 1, 0.9, 1.1\}$. Figure. 1 shows these sequences. By merely taking a look at the sequences, it can be inferred that there is a similarity between x and y. They are both different than z. However, calculating the DTW distance will show that $DTW(x, y)$ is 0.16 while $DTW(x, z)$ is 0.14. Hence, according to the DTW distance, the sequences x and z are more similar than x and y. Although x and z share a similar shape, it may not be meaningful to let x and z be grouped together if they are far from each other according to their Euclidean distance. Normalizing the sequences to the mean of zero and a standard deviation of one can solve this issue; however, this cannot be applied to volatile time series where the magnitude matters.

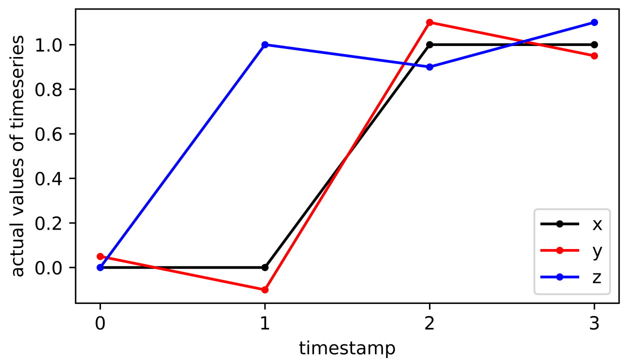

FIGURE 1. Three toy time series sequences.

Another problem that might appear in the use of DTW distance is when more than one data set is involved (e.g., wind data and load data). In such case, the clustering process should group days that exhibit similar patterns on both data sets, and the clustering process should be applied to a joint configuration of the data sets. In previous power systems studies, such as [10] and [34], this configuration is performed by joining the data sets on the time axis to create a new sequence. For instance, a wind power data set and electricity demand data set, each with size (365, 24) and (365, 24), can be concatenated to create a new data set with length (365, 48), such that in each sequence, the first 24 elements are from the first data set and the last 24 elements from the second data set. Generalization of this configuration for s number of data sets can be presented as follows:

$$T_i^{S_H} = T_i^{S_1} \cup_H \cdots \cup_H T_i^{S_n} = (T_i^{S_1}, \ldots, T_i^{S_n}) \quad (6)$$

Although the configuration in (6) can work well under Euclidean distance, it cannot properly handle the DTW distance. In fact, it might lead to the mapping of an element $T_{i,24}^{S_m}$ onto $T_{j,0}^{S_{m+1}}$, where S_m and S_{m+1} are two adjacent data sets considered in the horizontal configuration. For instance, in the case of wind power data and electricity demand data, this means allowing an element of wind power data to be mapped onto an element of electricity demand data, which is not reasonable. One solution is to use a very large value between $T_{i,24}^{S_m}$ and $T_{j,0}^{S_{m+1}}$. This sentical value can isolate the mapping of sequences of one data set from another data set. In other words, DTW distance tries to find the best match between the sequences of each data set included in the joint configuration without being affected by the mapping of other sequences from other data sets. The reason is that the large sentical value enforces $T_{i,24}^{S_m}$ to be mapped onto $T_{j,24}^{S_m}$, and the mapping of sequences of the next adjacent data set will start by mapping $T_{i,0}^{S_{m+1}}$ and $T_{j,0}^{S_{m+1}}$ onto each other.

2) K-SHAPE

K-Shape was initially introduced by Paparrizos *et al.* [12] to find correlation-based similarity between sequences. For this, each individual sequence should be normalized to have

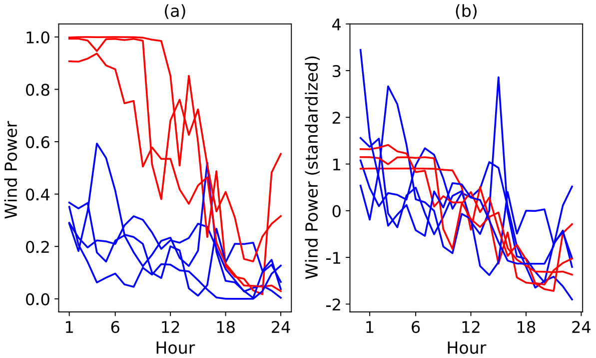

FIGURE 2. **One of the clusters extracted by K-Shape clustering with k = 40.**

a mean of zero and standard deviation of one before being fed to the K-Shape algorithm [5], [12]. Authors in [5] used a normalized-denormalized technique to first normalize the data and then denormalize it and bring it back to its actual values. The authors used K-Shape for clustering the electricity price data. However, the proposed method cannot be generalized to other data set such as wind power data where there is a higher level of volatility. Suppose two time series sequences T_1 and T_2 of wind power data (normalized by the installed capacity of its corresponding wind farm), where T_1 has a mean of 0.1 and standard deviation σ_1, and T_2 has a mean of 0.9 with the standard deviation σ_2. Normalizing these two sequences and bringing their K-Shape average back to the actual space results in a sequence with mean of 0.5, which means this approach will group two extreme cases into one group which is not desirable in the power system studies. Fig. 2 shows one of the clusters of wind power data obtained via the K-Shape clustering. As depicted in Fig. 2-(a), there is a distinctive difference between the days shown in red and the ones in blue. However, these days exhibit a similar pattern after being standardized, as illustrated in 2-(b). The reason, for this is that the K-Shape clustering is based on the correlation, which neglects the actual magnitude of a time series. Therefore, in highly volatile data, using K-Shape can result in grouping sequences that are different regarding their actual values.

Therefore, the K-Shape method will not be investigated further in this paper. However, DTW distance, in spite of having its own caveats, can be used in power system studies provided certain modifications are applied in the clustering process; the next section proposes a clustering framework that mitigates the caveats of DTW distance when applied to wind power.

III. THE PROPOSED CLUSTERING FRAMEWORK

In this section, we introduce a shape-based clustering framework (SBCF) that can alleviate the challenges of DTW, described in Section II-B. In particular, the proposed framework mitigates the chance of grouping sequences that are not close to each other according to their Euclidean distance by identifying the local neighbors through a consensus clustering process. Then, the mutual information score can be used to choose a set of clusters, obtained by DTW distance, that shows a high share with the local neighbors. The proposed framework properly employs elastic distance to preserve the between-data set relative behaviour, and finally uses a two-step clustering to provide the final representative days.

To scale the data, a data-wise normalization is applied to each data set. The reason behind this choice is that element-wise normalization cannot preserve the shape of each sequence and therefore the DTW distance will not be applicable.

After normalizing the data sets, we propose to construct a vertical configuration [21]. The vertical configuration adds features, including information from different data sets, into each timestamp of newly-joined sequences. The following expression shows this configuration in its mathematical form:

$$T_i^{S_V} = T_i^{S_1} \cup_V \cdots \cup_V T_i^{S_n} = \{t_q : t_q = (T_{i,q}^{S_1}, \ldots T_{i,q}^{S_n})\} \quad (7)$$

Accordingly, mapping a timestamp to another timestamp means that, in general, the value of each feature in a time stamp is close to its corresponding value in the other timestamp. In other words, the mapped timestamps share similar-in-values features, where each feature corresponds to one data set. Therefore, the vertical configuration can better preserve the joint behavior between data sets. Consequently, the representative days can better represent cases where the wind power is high, and electricity demand is low, or vice versa, which is of significance in impacting the operation of an energy storage system. The DTW distance in the vertical configuration can be calculated as follows:

$$DTW(T_i^{S_V}, T_j^{S_V}) = \min_{\pi} \sqrt{\sum_{(r,s)\in\pi} \|T_{i,r}^{S_V} - T_{j,s}^{S_V}\|^2} \quad (8)$$

To visualize the difference between the horizontal configuration and the advantage of vertical configuration, we separately use (1) and (8) to calculate the distance matrix on the wind power and electricity demand data. Then, we use a K-Medoids to obtain the clusters. Figure 3 shows two days that are grouped together with the use of the (independent) DTW distance (1), calculated through a horizontal configuration of the two data sets (6). The dashed lines show the optimal path discovered by the DTW distance. The circles show the part of the patterns where the joint behavior of the two data sets are not preserved. The wind power sequence on day 352 is mapped onto some previous hours of day 47. However, the electricity demand sequences of those same days show a different behavior. In fact, in the electricity demand, the day 47 is ahead of the day 352.

Fig. 4 shows two days as co-members of a cluster achieved by the use of dependent DTW through a vertical configuration (7). As opposed to the (independent) DTW, DTW distance associated with the vertical configuration can preserve the joint behavior of the sequences. For instance, the circles

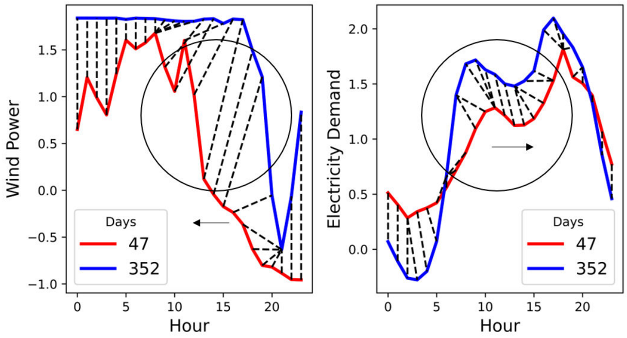

FIGURE 3. **Two days as co-member of one of the clusters achieved by DTW (1).**

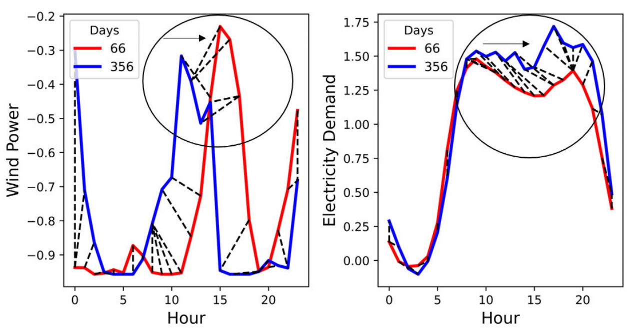

FIGURE 4. **Two days as co-member of one of the clusters achieved by (dependent) DTW (8).**

show the day 66 is a little ahead of day 356 regarding both the wind power and the electricity demand.

Owing to its vertical configuration, the time complexity of calculating the DTW distance decreases from $O(n^2H^2)$ to $O(H^2)$, where n is the number of data sets. Hence, in contrast to [10], there is no need to break down data into smaller partitions when a clustering method can be performed with only utilizing the distances between data points.

After configuring the data set, a DTW-based clustering algorithm can group the sequences into a set of clusters. However, as depicted in Fig. 1, grouping two sequences that are far from each other in the Euclidean distance should be avoided. Therefore, we are interested in finding clusters whose members are neighbors in the Euclidean distance as well. However, different distance measures result in different shapes for cluster [35]. Therefore, it is not reasonable to use a particular grouping method with Euclidean distance as it forces the neighbors to form a particular shape. We prefer to find the neighborhoods in the Euclidean spaces while avoiding a particular structure.

In this paper, we propose to use consensus clustering [22] to find the neighborhoods of data points in the Euclidean space. The goal is to generate a single consensus cluster using different set of clusters obtained from multiple clustering outcomes. To this end, we use Cluster-based Similarity Partitioning Algorithm (CSPA) [22]. In this algorithm, the input is a set of labels, each obtained by a clustering process; and the output is the clusters of the objects such that the average normal mutual information between themselves and the set of clusters provided as the input is maximized [22].

Therefore, we first perform several clustering processes associated with Euclidean distance and K-Medoids on the data set using different initialization. The reason for choosing K-Medoids is that it considers the centroids as one of the members of the clusters throughout the clustering process and therefore it will be less affected by the smoothing problem that might appear in the K-Means algorithm. Then, we use the consensus clustering algorithm to achieve the final groups, hereafter in this paper, referred to as neighborhoods partition.

After obtaining the neighborhoods, we perform the clustering process using DTW (8) to find the pairwise distance between any two sequences and construct the distance matrix D. For the grouping stage, we employ K-Medoids. K-Medoids can find the clusters faster than K-Means when DTW is in use as it only requires the distance matrix to obtain the clusters. In the DTW-based K-Means, however, the average, known as DTW Barycenter Average (DBA), is required during the K-Means algorithm and it should be optimized for each cluster with the time complexity of $O(IMn_T^2)$ [36]; I is the number of iteration required for achieving a convergence in the DBA, M is the number of sequences in the cluster, and n_T is the length of each sequence, which is 24, as 24 hours per day. Therefore, DTW-based K-Means is a time consuming process.

We repeat the DTW-based K-Medoids clustering process for different initializations. Then, we choose the one that has the highest share of grouped sequences with the neighborhoods partition. For this purpose, we use Mutual Information (MI) [37] to find the similarity score between the neighborhoods partition and each set of clusters obtained by the DTW-based K-Medoids. The definition of MI is as follows:

$$MI(U,V) = \sum_{i=1}^{|U|}\sum_{j=1}^{|V|} \frac{|U_i \cap V_j|}{N} \times \log\left(\frac{N|U_i \cap V_j|}{|U_i||V_j|}\right) \quad (9)$$

where, U and V are two label assignments of the same N objects in the data set, and U_i and V_j are, respectively, the i^{th} and j^{th} cluster in set U and V.

After choosing the cluster set with highest MI score, the centroid of a cluster C is its medoid, which is one of its members that has the smallest sum of distances to the other members of the same cluster, which can be obtained using the following formula:

$$\bar{C}_i = \arg\min_{T \in C_i} \left(\sum_{T' \in C_i} d(T, T')\right) \quad (10)$$

The proposed clustering framework is summarized into the following six steps:

Step I: Scale each data set through data-wise standardization, referred to Section II-A.1,

Step II: Make a joint data sets by using vertical configuration, using (7),

TABLE 1. Different clustering methods considered for comparison ($w = 0$ corresponds to Euclidean distance).

Name of Method	Scaling / Configuration	warping size / Grouping	Centroid
E-Kmeans(I)	Data-wise/ Horizontal	0 / K-Means	Closest-to-Average
E-Kmeans(II) [5]	Data-wise/ Horizontal	0 / K-Means	Medoid
HC-Ward [5]	Data-wise/ Horizontal	0/ HC(ward linkage)	Closest-to-Average
K-Medoids [5]	Data-wise/ Horizontal	0/ K-Medoids	Medoid
HC-2	Data-wise/ Horizontal	2/ HC (minmax linkage)	min-max
HC-24 [10]	Data-wise/ Horizontal	24/ HC (minmax linkage)	min-max
SBCF: The proposed clustering framework	Data-wise/ Vertical	2/ ED-based Consensus clustering, followed up by DTW-based K-Medoids	Medoid

Step III: Use consensus clustering to find the neighborhoods partition,
Step IV: Calculate the distance matrix with DTW distance measure, using (8) and feed it into the K-Medoids algorithm,
Step V: Repeat k-medoids algorithm and choose the set of clusters that has highest MI (9) with the neighborhoods partition,
Step VI: Obtain the medoid of each cluster using (10), and consider it as the representative period of the corresponding cluster.

IV. NUMERICAL RESULTS AND DISCUSSION

In this section, first the data sets and the clustering methods considered for comparison are described. Then, a number of performance validation measures are discussed and the performance of the clustering methods are analyzed accordingly.

We use the hourly Alberta Internal Load data along with the energy production of a single wind farm in the center of Alberta for the year 2019 to build the vertically configured data set. Although we use one year in this case study, the approach can be applied to more than one year. Thus, the data set consist of wind and demand data for 365 individual days. The objective of the aggregation process is to select a reduced number of days (e.g. 30 days [10]) that effectively represent the whole year. We apply the methods considering different numbers of representative days, i.e., $|C| \in \{20, 30, 40, 50\}$.

We compare the performance of the proposed data aggregation method to those of other alternative clustering methods listed in Table 1. The same number of initialization is considered for all methods. All methods are applied to the data scaled in the data-wise manner. In previous works [10], HC-minmax was applied with warping size of $w = 24$. In the present paper, however, we slightly modify this method by inserting a large sentical value between the wind power and electricity demand sequences in the horizontal configuration to resolve the challenge described in Section II-B.1. We implement this method with both $w = 24$ and $w = 2$, and refer to them by HC-24 and HC-2 in the following sections. In addition, E-Kmeans (II), HC-Ward, and K-Medoids. E-Kmeans, with average of the centroid as the representative periods, were used in [5]. E-Kmeans was also considered in [5] where the average of the centroid was used as the representative days. However, as discussed in [32], the average of a centroid obtained by the Euclidean distance is usually smoothed and may not be a proper representative period for sequences grouped together. Therefore, in this paper, we consider the representative period as one of the members of the cluster that is closest to the average. For the proposed framework of this paper we use $w = 2$. To be consistent in the consensus clustering stage in the proposed clustering framework, the neighborhood partition is extracted using $\frac{|C|}{5}$ as the number of local neighbors.

A. AGGREGATION PERFORMANCE MEASURES

To compare alternative clustering methods, a number of criteria, which are independent of the distance or grouping method chosen for the clustering process, are defined. This is achieved by two different approaches: (1) Data-based Validation, and (2) Model-based Validation [3]. In Data-based Validation, those criteria are determined directly from the data. This is used to understand the quality of clusters with respect to a particular aspect of the data. In model-based validation, the performance of the clustering methods is illustrated for a case study. In the following sections, A-DATA refers to the original data set whereas C-DATA refers to the time aggregated data set that is constructed based on the representative periods. To apply the performance measures, C-DATA is built to have the same size of A-DATA by replacing each day of A-DATA with its corresponding representative period.

For the data-based validations, we use Duration Curve Error Index for the net demand, i.e., demand minus wind production. Duration Curve of a data set ignores the temporal information and sorts the data according to the magnitude of instances. The Duration Curve Error (DCE) measures how well the C-DATA's duration curve represents A-DATA's. Some researchers compare the duration curve of the entire data set [10]. However, such approach disregards the relationship between each cluster and its corresponding centroid. Another approach is to compare the duration curve of each sequence with its corresponding representative sequence [34]. However, this approach ignores the fact that a centroid is impacted by the members of its corresponding cluster. Hence, it is not meaningful to individually compare each sequence with its representative period without considering the other members of a cluster.

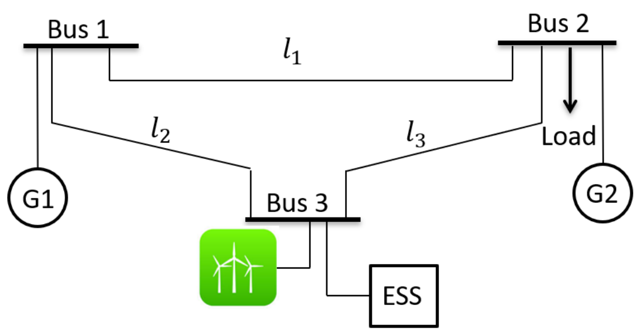

FIGURE 5. A three-bus network with one wind farm and one load.

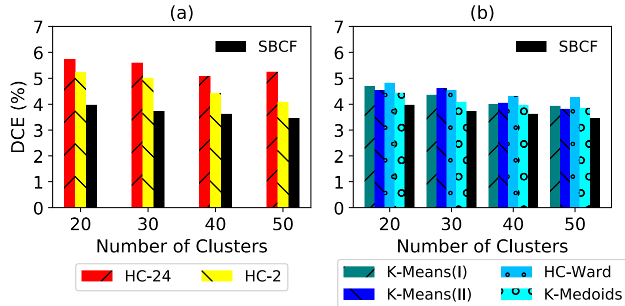

FIGURE 6. Net demand's DCE of the proposed framework compared with methods using: (a) DTW distance, and (b) Euclidean distance.

To this end, we first compare the duration curves that are calculated using A-DATA and C-DATA for each cluster. Then, similar to [10], we normalize the root square mean of errors across the clusters as follows:

$$\mathrm{DCE}^S = \frac{\sqrt{\frac{1}{H_S}\sum_{k=1}^{|C|}\sum_{h=1}^{H_k}(t_{h,a}^{C_k} - t_{h,c}^{C_k})^2}}{|\ t_a^{(max)} - t_a^{(min)}\ |} \times 100 \tag{11}$$

where, DCE^S is the DCE corresponding to the data set S. k is the index of the cluster. H_S represents the total hours in data set S, while H_k represents the total number of hours in the k^{th} cluster. $t_{h,a}$ is the h^{th} element of the duration curve constructed by the actual data set, while $t_{h,a}^{C_k}$ is the h^{th} element of the duration curve constructed by the sequences of actual data set grouped into the cluster C_k. $t_{h,c}^{C_k}$ is the h^{th} element of the duration curve constructed by using the sequences of C-DATA corresponding to the cluster C_k. Lower values of DCE means the selected representative periods can better replicate the magnitudes of the original data sets.

The DCE index does not measure the performance of the clustering methods in detecting temporal patterns as they disregard the time dependency of the elements. Thus, to fill the gap, additional model-based measures are often used [10] [5] where a power system study is set up to further validate the performance of the clustering methods. To develop model-based validation measures, we employ a three-bus network where a wind farm is co-located with an energy storage system at one of the buses-see Fig. 5.

The objective of operating this network is to minimize the cost of daily unit commitment operation, together with the economic dispatch over a full year [38]. The operation cost of the network is optimized for each day and their summation over one year gives the annual cost of operation [5]. The reason for this is to make sure that the inter-day information does not affect the result. The energy storage is co-located with the wind farm at bus 3 and is being operated by the operator of the network. Four performance indices are defined for this system, i.e., the annual operation cost of the system, the total annual wind curtailment, the total annual storage energy throughput defined as the summation of discharged energy, and the daily average of the state of the charge of the energy storage for each 365 days of the year. Thus, we have four model-based indices.

The installed wind capacity is 250 MW and the peak load is 400 MW. The penetration level is 25%. The complete input data and parameters for this case study are available in [39].

To compare the performance of each of the data aggregation methods according to each of the indices defined above, we calculate each index when using the actual complete data set. After performing the clustering process by each method, a set of representative days are extracted. A weight is associated with each representative day that reflects the impact of that representative day. The weight is the number of days in its corresponding cluster. The weighted mean absolute error for each index I is then calculated as follows:

$$\mathrm{WMAE}_I = \frac{\sum_{k=1}^{|C|} w_k \left|I_{k,c} - I_{k,a}\right|}{I_{yr}} \times 100 \tag{12}$$

where, I_{yr} is the annual value of index I, the weight w_k is the number of days grouped into the k^{th} cluster, $I_{k,a}$ is the mean value of I corresponds to the k^{th} cluster obtained by A-DATA, and $I_{k,c}$ is the value of index represented by the representative period of the same cluster.

We compare the proposed method, i.e., SBCF, versus the methods listed in Table 1 based on two data-based and three model-based indices. Given that the proposed method is shape-based, we compare it against HC-2 and HC-24 that incorporate DTW distance in a separate graph, and then against other Euclidean distance-based methods in another graph.

The DCE indices applied to net demand for the proposed SBCF and alternative methods are presented in Fig. 6; comparison with two shape-based methods are shown ion Fig. 6-(a) whereas the comparative results for four Euclidean-based methods are demonstrated in Fig. 6-(b). As illustrated in Fig. 6-(a), the proposed method outperforms the other two shape-based methods. This is because, in contrast to HC-24 with warping size 24, SBCF has a restriction on the warping size, and hence, it avoids extreme expansion or compression of sequences. Therefore, the distortion of the

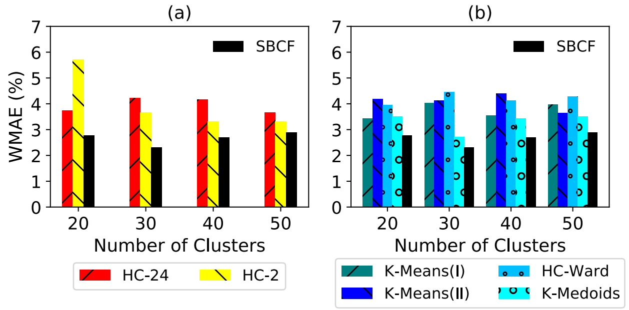

FIGURE 7. **WMAE$_{\text{Cost of Operation}}$ of the proposed framework compared with methods using: (a) DTW distance, and (b) Euclidean distance.**

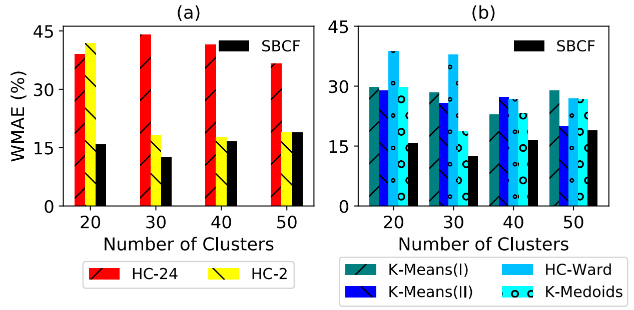

FIGURE 8. **WMAE$_{\text{Wind Curtailment}}$ of the proposed framework compared with methods using: (a) DTW distance, and (b) Euclidean distance.**

duration curve is negligible in the SBCF. Although using the same warping size of 2 can mitigate such distortion in HC- method, it can be seen that HC-2 still shows larger error compared to SBCF. The reason has its root in the fact that the mapping of sequences of wind power and electricity demand are dependent on each other in the SBCF. Therefore, as opposed to HC-2, SBCF has further restrictions in mapping the sequences, and therefore, it shows better performance in replicating the duration curve.

Figure. 6-(b) compares the DCE index of SBCF against ED-based methods. K-Means(I), (II), and HC-ward suffer from the smoothing challenge caused by the averaging process [32]. Although K-Medoids uses the medoid as the representative periods, it is still unable to detect the similarity of two sequences if there is a small lag in one of the sequences. Owed to the use of dependent DTW distance, the proposed clustering framework benefits from matching sequences in an elastic manner while preserving the relative shape in the duration curve of the net demand. This is why the proposed method slightly outperforms the ED-based techniques.

Figure 7 shows the performance of the representative days of alternative clustering methods in approximating the operation cost of the network. Observe from Fig. 7-(a) that the proposed framework shows better performance than HC-2 and HC-24 in reflecting the operation cost of the network. It is worthwhile to note that HC-24 in spite of exhibiting high error in replicating the duration curve, shows low errors in reflecting the cost of operation. The reason has its root in the fact that the unit commitment is mostly driven by demand. Since the compression/expansion usually affect the wind profiles rather the load profile, the high error of HC-24 in replicating the duration curve may not be reflected in the operation cost.

Comparing the SBCF against the ED-based clustering methods in Fig. 7-(b) shows that the performance of the proposed SBCF outperforms other methods in approximating the operation cost. As discussed earlier, the operation of a unit commitment is usually driven by the electricity demand. Therefore, the performance of the methods in approximating the operation cost are close to each other.

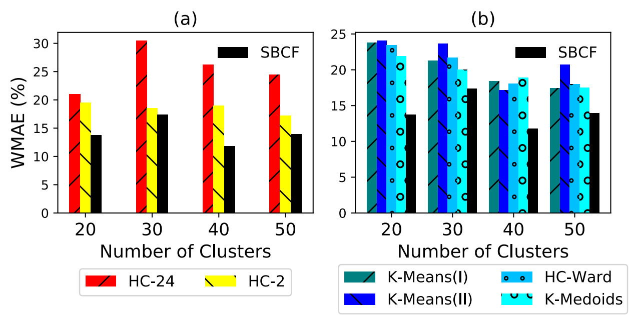

FIGURE 9. **WMAE$_{\text{Energy throughput}}$ of the proposed framework compared with methods using: (a) DTW distance, and (b) Euclidean distance.**

Figure. 8 compares the error of the alternative methods in approximating the annual wind curtailment in the case study. Observe from Fig. 8-(a) that the proposed clustering framework significantly outperforms HC-24. The reason lies in the fact that its warping size is not restricted, which can lead to extreme compression or expansion of sequences to match them onto each other. Therefore, a particular curtailment at some intervals can be reflected as a small or large wind curtailment in the representative periods. Compared to HC-2, the proposed clustering framework has lower or comparable error. In fact, to achieve as small error as SBCF, higher number of clusters should be used for HC-2 to mitigate its problem in preserving the relative behavior between the data sets. Compared to the ED-based clustering methods, as shown in Fig. 8-b, observe that the proposed framework has smaller errors in approximating the curtailed wind energy for all different number of clusters.

The performances of alternative methods in reflecting the true data when it comes to the annual energy throughput of the storage system for in the 3-bus system is illustrated in Fig. 9. Observe from Fig. 9-(a) that the proposed framework outperforms others in replicating the storage energy throughput. The reason is that the operation of the energy storage depends on both the wind power and the load of the network, and SBCF considers both due to the vertical configuration of the data.

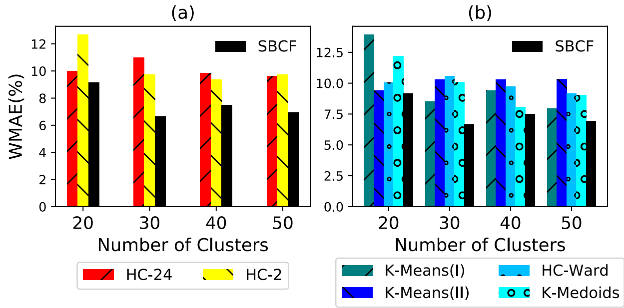

FIGURE 10. **$WMAE_{SOC}$ of the proposed framework compared with methods using: (a) DTW distance, and (b) Euclidean distance.**

As depicted in Fig. 9-(b), the proposed SBCF also shows lower errors compared to the ED-based clustering methods, for different number of representative days.

In addition to the energy throughput of the energy storage, the performances of alternative methods in reflecting the daily average value of the state-of-charge (soc) of the energy storage, calculated by taking the average of the hourly soc throughout a day, is illustrated in Fig. 10.

As depicted in Fig. 10-(a), the proposed SBCF shows a better performance compared to the other DTW-based methods. The reason lies in the fact that the proposed method can preserve the joint behavior of wind power and electricity demand. Therefore, it can better capture the moments of high wind with low demand (energy storage is charging) and low wind with high demand (energy storage is discharging). The proposed SBCF can generally preserve the overall operation of the energy storage and hence, the grouped days have similar daily average soc values.

Compared with ED-based methods in Fig. 10-(b), it is observed that the proposed SBCF outperforms the other methods since it utilizes the elastic distance.

Finally, to show the effectiveness of the proposed method under different levels of wind power penetration, we applied all the methods with $k = 30$ clusters when there are $\{20\%, 25\%, 30\%, 35\%, 40\%\}$ level of wind power penetration. Figure 11 shows the performance of the methods in reflecting the operational cost of the network.

As depicted in Fig. 11, increasing the wind power penetration generally results in higher error. However, the proposed clustering framework, SBCF, shows better performance compared to the other clustering methods. In fact, at a higher level of wind power penetration, the difference in the error of the proposed method becomes bigger which shows the effectiveness of the proposed method in the highly volatile context.

To summarize, and considering all the model-based and data-based indices, the proposed method better reflects the overall behaviour of both wind and demand data when applied to systems that include energy storage. This is consistent with the findings of [5] where shape-based methods are preferred when dealing with electricity market price data involving energy storage. The key difference, however, is that instead of k-shape that is found useful in [5], we demonstrate the good performance of DTW distance when wind power data is involved. Also, compared to the results of HC-24 presented in [10], since we employed DTW distance in a multi-dimensional format, and considered the local neighbors during the clustering process, the performance is improved.

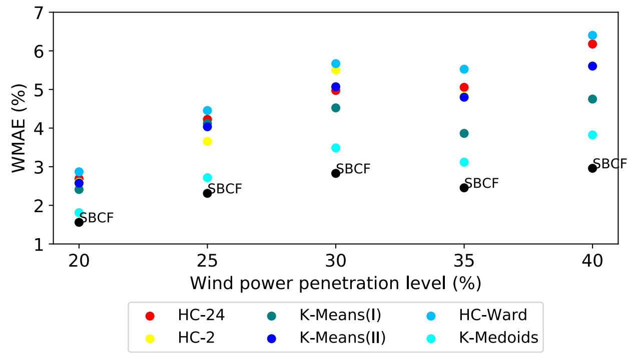

FIGURE 11. **$WMAE_{\text{Cost of Operation}}$ of the proposed framework compared with other methods for 30 clusters under different levels of penetration.**

V. CONCLUSION AND FUTURE WORK

In this paper, we propose a time aggregation framework for choosing representative periods for studies that include both wind and load data. We focus on the shape of the data sets because we are interested in system studies that involve energy storage. The proposed method uses the DTW distance in the multi-dimensional format. Therefore, it takes advantage of an elastic distance without mistreating the joint behavior of the data sets. Furthermore, to avoid grouping sequences that are far from each other according to their Euclidean distance, the proposed framework first utilizes consensus clustering to extract group of sequences as local neighbors. Then, mutual information score is applied to find a set of clusters obtained by the use of DTW distance that show more common partitions with the extracted neighbors. Compared to other clustering methods in the power system community, the proposed framework can jointly preserve the shape and behavior of the included data sets which can play an important role in the operation of the network when the energy storage is present. We use data-based and model-based validation indices to compare the performance of the proposed framework against six other methods. The results show that the proposed method consistently outperforms the alternatives in replicating the duration curve of the net demand- defined as load minus wind-, as well as reflecting the cost of operation, the amount of wind curtailment, and the energy throughput of the energy storage system.

Therefore, as long as the data sets are related to a particular area, they should have joint relationship. For instance, when the wind is low and the load is high during a day, the price is usually high. Such observation means that preserving the joint behavior of data sets is of significance. So, the data sets can be structured vertically and the proposed clustering framework is recommended.

In cases concerned with multiple sources of uncertainty where the sources are far from each other from a geographical point of view, the proposed method should be used with caution as the sources might not necessarily have joint behavior. It is important to note that other current methods have a similar challenge. According to a recent study [40], analyzing multi-variate time series data for similar patterns in high dimensional space does not lead to meaningful results. In other words, the grouped patterns and particularly their centroid may not properly show similar behavior in all dimensions. This task is under the study of the authors of this paper.

REFERENCES

[1] M. Majidi-Qadikolai and R. Baldick, "A generalized decomposition framework for large-scale transmission expansion planning," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1635–1649, Mar. 2018.

[2] E. Shayesteh, B. F. Hobbs, and M. Amelin, "Scenario reduction, network aggregation, and DC linearisation: Which simplifications matter most in operations and planning optimisation?" *IET Gener., Transmiss. Distrib.*, vol. 10, no. 11, pp. 2748–2755, Aug. 2016.

[3] S. Buchholz, M. Gamst, and D. Pisinger, "A comparative study of time aggregation techniques in relation to power capacity expansion modeling," *TOP*, vol. 27, no. 3, pp. 353–405, Oct. 2019.

[4] A. Werner, K. Duwadi, N. Stegmeier, T. M. Hansen, and J.-H. Kimn, "Parallel implementation of AC optimal power flow and time constrained optimal power flow using high performance computing," in *Proc. IEEE 9th Annu. Comput. Commun. Workshop Conf. (CCWC)*, Jan. 2019, pp. 597–602.

[5] H. Teichgraeber and A. R. Brandt, "Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison," *Appl. Energy*, vol. 239, pp. 1283–1293, Apr. 2019.

[6] H. Khani, M. R. Dadash Zadeh, and A. H. Hajimiragha, "Transmission congestion relief using privately owned large-scale energy storage systems in a competitive electricity market," *IEEE Trans. Power Syst.*, vol. 31, no. 2, pp. 1449–1458, Mar. 2016.

[7] A. J. Conejo *et al.*, "Long-term coordination of transmission and storage to integrate wind power," *CSEE J. Power Energy Syst.*, vol. 3, no. 1, pp. 36–43, Mar. 2017.

[8] L. Baringo and A. J. Conejo, "Correlated wind-power production and electric load scenarios for investment decisions," *Appl. Energy*, vol. 101, pp. 475–482, Jan. 2013.

[9] T. Qiu, B. Xu, Y. Wang, Y. Dvorkin, and D. S. Kirschen, "Stochastic multistage coplanning of transmission expansion and energy storage," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 643–651, Jan. 2017.

[10] Y. Liu, R. Sioshansi, and A. J. Conejo, "Hierarchical clustering to find representative operating periods for capacity-expansion modeling," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3029–3039, May 2018.

[11] C. A. Ratanamahatana and E. Keogh, "Making time-series classification more accurate using learned constraints," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2004, pp. 11–22.

[12] J. Paparrizos and L. Gravano, "K-shape: Efficient and accurate clustering of time series," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2015, pp. 1855–1870.

[13] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *Int. J. Forecasting*, vol. 30, no. 4, pp. 1030–1081, Oct. 2014, doi: 10.1016/j.ijforecast.2014.08.008.

[14] D. Z. Fitiwi, F. de Cuadra, L. Olmos, and M. Rivier, "A new approach of clustering operational states for power network expansion planning problems dealing with RES (renewable energy source) generation operational variability and uncertainty," *Energy*, vol. 90, pp. 1360–1376, Oct. 2015.

[15] Q. Ploussard, L. Olmos, and A. Ramos, "An operational state aggregation technique for transmission expansion planning based on line benefits," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 2744–2755, Jul. 2017.

[16] R. Alvarez, A. Moser, and C. A. Rahmann, "Novel methodology for selecting representative operating points for the TNEP," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 2234–2242, May 2017.

[17] M. Sun, F. Teng, I. Konstantelos, and G. Strbac, "An objective-based scenario selection method for transmission network expansion planning with multivariate stochasticity in load and renewable energy sources," *Energy*, vol. 145, pp. 871–885, Feb. 2018.

[18] M. Sun, F. Teng, X. Zhang, G. Strbac, and D. Pudjianto, "Data-driven representative day selection for investment decisions: A cost-oriented approach," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 2925–2936, Jul. 2019.

[19] A. Yeganefar, M. R. Amin-Naseri, and M. K. Sheikh-El-Eslami, "Improvement of representative days selection in power system planning by incorporating the extreme days of the net load to take account of the variability and intermittency of renewable resources," *Appl. Energy*, vol. 272, Aug. 2020, Art. no. 115224.

[20] D. A. Tejada-Arango, M. Domeshek, S. Wogrin, and E. Centeno, "Enhanced representative days and system states modeling for energy storage investment analysis," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6534–6544, Nov. 2018.

[21] M. Shokoohi-Yekta, B. Hu, H. Jin, J. Wang, and E. Keogh, "Generalizing DTW to the multi-dimensional case requires an adaptive approach," *Data Mining Knowl. Discovery*, vol. 31, no. 1, pp. 1–31, Jan. 2017.

[22] A. Strehl and J. Ghosh, "Cluster ensembles—A knowledge reuse framework for combining multiple partitions," *J. Mach. Learn. Res.*, vol. 3, pp. 583–617, Dec. 2002.

[23] E. Keogh and J. Lin, "Clustering of time-series subsequences is meaningless: Implications for previous and future research," *Knowl. Inf. Syst.*, vol. 8, no. 2, pp. 154–177, Aug. 2005.

[24] M. Hoffmann, J. Priesmann, L. Nolting, A. Praktiknjo, L. Kotzur, and D. Stolten, "Typical periods or typical time steps? A multi-model analysis to determine the optimal temporal aggregation for energy system models," 2021, *arXiv:2103.16657*. [Online]. Available: http://arxiv.org/abs/2103.16657

[25] J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*. San Mateo, CA, USA: Morgan Kaufmann, 2011.

[26] S. G. K. Patro and K. K. Sahu, "Normalization: A preprocessing stage," 2015, *arXiv:1503.06462*. [Online]. Available: http://arxiv.org/abs/1503.06462

[27] H. A. Dau *et al.*, "Optimizing dynamic time warping's window width for time series data mining applications," *Data Mining Knowl. Discovery*, vol. 32, no. 4, pp. 1074–1120, Jul. 2018.

[28] J. Wu, *Advances in K-means Clustering: A Data Mining Thinking*. Cham, Switzerland: Springer, 2012.

[29] D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms*. New Orleans, LA, USA: Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035, doi: 10.5555/1283383.1283494.

[30] P. Fränti and S. Sieranoja, "How much can k-means be improved by using better initialization and repeats?" *Pattern Recognit.*, vol. 93, pp. 95–112, Sep. 2019.

[31] H.-S. Park and C.-H. Jun, "A simple and fast algorithm for K-medoids clustering," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3336–3341, Mar. 2009.

[32] M. Hoffmann, L. Kotzur, D. Stolten, and M. Robinius, "A review on time series aggregation methods for energy system models," *Energies*, vol. 13, no. 3, p. 641, Feb. 2020.

[33] E. Keogh and S. Kasetty, "On the need for time series data mining benchmarks: A survey and empirical demonstration," *Data Mining Knowl. Discovery*, vol. 7, no. 4, pp. 349–371, Oct. 2003.

[34] K. Poncelet, H. Hoschle, E. Delarue, A. Virag, and W. Drhaeseleer, "Selecting representative days for capturing the implications of integrating intermittent renewables in generation expansion planning problems," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 1936–1948, May 2017.

[35] T. Velmurugan, "Efficiency of K-means and K-medoids algorithms for clustering arbitrary data points," *Int. J. Comput. Technol. Appl.*, vol. 3, no. 5, pp. 1758–1764, 2012.

[36] F. Petitjean, A. Ketterlin, and P. Gançarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognit.*, vol. 44, no. 3, pp. 678–693, Mar. 2011.

[37] S. Romano, J. Bailey, V. Nguyen, and K. Verspoor, "Standardized mutual information for clustering comparisons: One step further in adjustment for chance," in *Proc. Int. Conf. Mach. Learn.*, PMLR, 2014, pp. 1143–1151.

[38] Y. Dvorkin, R. Fernandez-Blanco, D. S. Kirschen, H. Pandzic, J.-P. Watson, and C. A. Silva-Monroy, "Ensuring profitability of energy storage," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 611–623, Jan. 2017.

[39] *The System Data for the Three-Bus Network*. Accessed: Jul. 15, 2021. [Online]. Available: https://bit.ly/3egp2s7

[40] C.-C.-M. Yeh, N. Kavantzas, and E. Keogh, "Matrix profile VI: Meaningful multidimensional motif discovery," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2017, pp. 565–574.