Non-Conforming Load Pattern Analysis

Shiuli Subhra Ghosh, Bhawuk Luthra, Jaime De La Ree Jr., and Kevin D. Jones
Engineering Analytics & Modeling
Dominion Energy, Glen Allen, VA, USA
shiuli.s.ghosh@dominionenergy.com

***Abstract*—Scientific proliferation in artificial intelligence and digitization, has substantially increased power demand from data centers, creating new operational challenges for power utilities. Unlike residential and industrial loads, data center consumption exhibits highly non-conforming and irregular patterns. Utilization of data centers can be classified based on hyperscaler, colocation, and enterprise ownership models. Transmission system operators often lack visibility into the utilization of the load for behind-the-meter customers connected via electric distribution companies (EDCs). While there has been ample research about industrial and residential load patterns, research on data center specific loads is limited. This paper presents a study on clustering and identification of load patterns for data centers and residential customers at transformer level. Using hourly SCADA time-series data, we apply K-Means and Agglomerative clustering algorithms to differentiate load behaviors. Our findings emphasize that Agglomerative clustering effectively separates residential (conforming),and data center (non-conforming) profiles, while K-Means further distinguishes between two distinct data center patterns corresponding to different facility types prevalent in Northern Virginia.**

***Index Terms*—Data Centers, Load Patterns, Non-conforming Load, Power Transmission System, Clustering**

I. Introduction

The expansion of the digital economy across the globe has led to a substantial increase in computational demand, driven largely by the proliferation of artificial intelligence (AI) applications such as large language models (LLMs) and cloud-based services. Over the past decade, global server deployments have increased by 647%, storage capacity has expanded by 2,500% and network traffic has surged by 1,000% [1]. Northern Virginia, often referred to as *Data Center Alley*, hosts the world's largest concentration of data centers and serves as a critical nexus for global internet traffic [2]. Its capacity is projected to surpass the combined total of the next three largest U.S. markets [3].

Dominion Energy, the region's primary transmission and distribution utility, supports this need, which now constitutes a significant portion of its system load. The unprecedented growth of data center load has introduced new challenges for power system operators like understanding non-conforming load behaviors and operating the grid under heavily constrained circumstances. Unlike residential or commercial loads, data center consumption patterns are characterized by high load factors, minimal diurnal variation, and abrupt step changes during commissioning or expansion phases. Moreover, transmission operators do not have visibility into behind-the-meter operations, connected via distribution networks.

This paper presents a data-driven clustering framework to identify and characterize load patterns at the transmission level, with a focus on distinguishing conforming (residential and commercial) and non-conforming (data center) behaviors. However, Non-conforming data center loads can have different load profiles depending on their ownership model, such as enterprise, colocation, or hyperscaler.

With the advent of vertical stacking and cloud computing technologies it becomes increasingly important to differentiate data centers by their workloads. For instance, a 100 MW hyperscaler data center providing cloud services may operate significantly different from an equivalently sized colocation or enterprise data center. Using historical time-series data at the transformer level from transmission assets in Northern Virginia, we apply K-Means and Agglomerative clustering algorithms to classify data center load profiles. Our analysis reveals distinct consumption archetypes, including three subcategories of data center loads, with distinct range, variance and load shape characteristics. As Dominion Energy strategically expands to meet the increasing demand, this study of data center load pattern analysis is pivotal as a decision-support resource for load forecasting, short-term planning and operational reliability.

II. Preliminaries and Problem Statement

A. Modeling Substation and Transformer Level Series

Consider a power transmission utility region comprising of S substations. Each substation contains X_p transformers indexed by $p \in \mathcal{K} \triangleq [1, \ldots, S]$. The total number of assets in the hierarchy is defined as, $m = S + m_3$, where, $m_3 = \sum_{p=1}^{S} X_p$.

Given the hierarchical structure of the power transmission network, load in power transmission assets can be effectively modeled as time series. Let $\mathcal{J} \triangleq \{(p, j) \,|\, p \in [S],\, j \in [X_p]\}$ denote the index set of all transformers, with $|\mathcal{J}| = m_3$. For each transformer $(p, j) \in \mathcal{J}$, the observed time series load measurements within a time window of T steps is defined as,

$$\mathbf{y}_{p,j} \;=\; \left\langle y_{p,j}[1],\, y_{p,j}[2],\, \ldots,\, y_{p,j}[T] \right\rangle^{\top} \in \mathbb{R}^{T} \;. \qquad (1)$$

B. Problem Statement

We aim to cluster all m_3 transformers into n clusters, based on their load behavior. Optionally, we map each time series to a d-dimensional feature vector $\mathbf{x}_{p,j} \in \mathbb{R}^{d}$ via a feature map

979-8-3315-8138-1

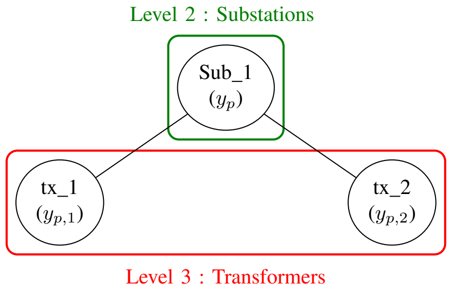

Fig. 1. Dominion Transmission Hierarchy for Data Centers

$\phi : \mathbb{R}^T \to \mathbb{R}^d$ (e.g., hourly statistics, seasonal profiles, spectral features):

$$\mathbf{x}_{p,j} \;=\; \phi\big(\mathbf{y}_{p,j}\big)\;. \tag{2}$$

We collect the feature set as $\mathcal{X} \triangleq \{\mathbf{x}_{p,j} : (p,j) \in \mathcal{J}\}$. For formalizing the clustering problem, we seek a partition $\{\mathcal{J}_k\}_{k=1}^n$ of $\mathcal{J}$ such that,

$$\bigsqcup_{k=1}^{n} \mathcal{J}_k \;=\; \mathcal{J}, \qquad \mathcal{J}_k \cap \mathcal{J}_\ell = \emptyset \;\text{ for } k \neq \ell\;, \tag{3}$$

that minimizes a within-cluster dispersion objective $\mathsf{disp}(\cdot)$ under a chosen dissimilarity $D(\cdot,\cdot)$ (e.g., Euclidean distance on $\mathbf{x}_{p,j}$, or DTW on raw $\mathbf{y}_{p,j}$):

$$\min_{\{\mathcal{J}_k\}_{k=1}^n} \;\sum_{k=1}^{n} \sum_{(p,j)\in\mathcal{J}_k} \mathsf{disp}\left((p,j),\mathcal{J}_k\right), \tag{4}$$

where a common choice is the variance-like dispersion

$$\mathsf{disp}\left((p,j),\mathcal{J}_k\right) \;=\; D\big(\mathbf{x}_{p,j},\, \mathsf{rep}(\mathcal{J}_k)\big)^2,$$

and $\mathsf{rep}(\mathcal{J}_k)$ is a cluster representative (e.g., centroid or medoid). Finally, we denote the cluster set as $\mathcal{C}_n \triangleq \{1,\ldots,n\}$.

III. Load Pattern Clustering Algorithms

Relay measurement data collected by SCADA systems in the transformer-level usually have high temporal resolution. Therefore, feature identification for clustering plays a pivotal role. In this section, we introduce feature identification and weighting mechanism in the context of load pattern identification followed by statistical formulation of clustering algorithms.

A. Feature Identification and Feature Weighing

We hypothesize that data center load patterns can be differentiated by it's regularity in shape, range and variance in a daily time window. Therefore, we define the *shape feature* $\mathbf{x}^s_{p,j} = \phi_s(\mathbf{y}_{p,j})$ where, $\phi_s : \mathbb{R}^T \to \mathbb{R}^H$ aggregates $\mathbf{y}_{p,j}$ to H hourly means, where H is the number of hours in the observation window. Let $\mathcal{I}_h \subseteq \{1,\ldots,T\}$ denote the index set of time steps corresponding to hour h. The h-th component of $\phi_s(\mathbf{y}_{p,j})$ is

$$\big[\phi_s(\mathbf{y}_{p,j})\big]_h = \frac{1}{|\mathcal{I}_h|}\sum_{t\in\mathcal{I}_h} y_{p,j}[t], \qquad h = 1,\ldots,H\;. \tag{5}$$

In addition to hourly averages, we define a *daily range feature* for each transformer. Let $\phi_r : \mathbb{R}^T \to \mathbb{R}$ denote the mapping that returns the mean of daily ranges over D days. For transformer (p,j) with time series $\mathbf{y}_{p,j}$, partition the index set $\{1,\ldots,T\}$ into D disjoint subsets $\mathcal{I}_d$ corresponding to day d. Consequently, the mean daily range feature is defined as,

$$\mathbf{x}^r_{p,j} = \phi_r(\mathbf{y}_{p,j}) = \frac{1}{D}\sum_{d=1}^{D}\Big(\max_{t\in\mathcal{I}_d} y_{p,j}[t] - \min_{t\in\mathcal{I}_d} y_{p,j}[t]\Big)\;. \tag{6}$$

Finally, we define a function $\phi_v : \mathbb{R}^T \to \mathbb{R}$ that returns a mean of *daily variance feature*,

$$\mathbf{x}^v_{p,j} = \phi_v(\mathbf{y}_{p,j}) \;=\; \frac{1}{D|\mathcal{I}_d|}\sum_{d=1}^{D}\sum_{t\in\mathcal{I}_d}\Big(y_{p,j}[t] - \mu_{p,j}[d]\Big)\;. \tag{7}$$

where, $\mu_{p,j}[d] = \frac{1}{|\mathcal{I}_d|}\sum_{t\in\mathcal{I}_d} y_{p,j}[t]$. For scaling the feature parameters, we associate α_s, α_r, and α_v corresponding to shape, range, and variance features respectively and concatenate to get a feature vector $\mathbf{x}_{p,j} = \big[\alpha_s \cdot \mathbf{x}^s_{p,j}, \alpha_r \cdot \mathbf{x}^r_{p,j}, \alpha_v \cdot \mathbf{x}^v_{p,j}\big]$ The final feature set $\mathcal{X}$ is obtained as mentioned in section II-B.

B. K-Means formulation (centroid-based clustering).

K-means clustering algorithm solves the minimization problem [4] defined in equation (4), where $D(\cdot,\cdot)$ is defined as the euclidean distance and $\mathsf{rep}(\mathcal{J}_k) = \boldsymbol{\mu}_k$ are the centroid of each cluster k. The algorithm is initialized by selecting the random $\boldsymbol{\mu}_k$ and iteratively solves the objective function in (4), until convergence, by updating the centroids

$$\boldsymbol{\mu}_k \;=\; \frac{\sum_{(p,j)\in\mathcal{J}} z_{(p,j),k}\, \mathbf{x}_{p,j}}{\sum_{(p,j)\in\mathcal{J}} z_{(p,j),k}}\,, \qquad \forall\, k \in \mathcal{C}\;, \tag{8}$$

where, $z_{(p,j),k} \in \{0,1\}$ indicating whether transformer (p,j) is assigned to cluster k.

C. Agglomerative clustering (hierarchical partitioning).

Agglomerative clustering is a bottom-up hierarchical algorithm that builds a tree-like structure of nested clusters [5]. Unlike K-Means, it does not require pre-specifying centroids or solving an optimization problem. Instead, it begins with each transformer (p,j) as its own singleton cluster and iteratively merges the pair of clusters with the smallest inter-cluster dissimilarity, based on a chosen linkage criterion.

The dissimilarity between two clusters A and B is computed using a linkage function $\mathsf{Link}(A,B)$, which may be defined as:

- **Single linkage:** $\min_{a\in A,\, b\in B} D(a,b)$
- **Complete linkage:** $\max_{a\in A,\, b\in B} D(a,b)$
- **Average linkage:** $\frac{1}{|A||B|}\sum_{a\in A,\, b\in B} D(a,b)$

Here, $D(\cdot,\cdot)$ is the pairwise Euclidean distance between feature vectors on raw time series. The algorithm proceeds until exactly n clusters remain, forming a partition $\{\mathcal{J}_k\}_{k=1}^n$ consistent with the objective in equation (4).

Agglomerative clustering is particularly effective in identifying nested or non-convex structures in the data, and in our case, it successfully distinguishes residential, data center, and mixed load patterns without requiring initialization or assumptions about cluster shapes.

979-8-3315-8138-1

IV. Case Study

In this section, we apply the clustering methodologies outlined in Section III to analyze transmission-level load measurements from Dominion Energy. The primary objective is to identify and characterize conforming and non-conforming load patterns associated with behind-the-meter customers, particularly data centers. For this study, we focus on a transmission zone in Northern Virginia, a region with a high concentration of data center facilities. Dominion Energy has observed a significant increase in data center electricity consumption since 2020. The recent contracted capacity projects a load of 40.2GW, an 88% increase from just 21.4GW in July 2024.

Given that load patterns during the initial ramp-up phase of data center operations may not reflect steady-state behavior, we restrict our analysis to facilities that have reached operational stability. Specifically, we examine load data from $S = 12$ substations comprising $m_3 = 46$ transformers (see Figure 1 for the network topology). Load measurements are collected at the transformer level using field-deployed relays, retrieved via Dominion's SCADA ADMS infrastructure, and archived in the SCADA Historian database.

A. Data Collection and Feature Set Formation

Raw historical load data retrieved from the SCADA historian are sampled at approximately 4-second intervals. However, asynchronous data collection across transformers results in non-uniform timestamps, making direct shape comparison inconsistent. To address this, we aggregate the raw data into time weighted hourly averages as defined in equation (5).

Operational experience at Dominion Energy indicates that load shapes vary significantly by hour, day, and season. Seasonal differences (e.g., summer vs. winter) produce distinct daily profiles for conforming and non-conforming loads. For clustering purposes, daily load curves ($\mathbf{x}^s_{p,j}$), along with derived features such as mean daily range ($\mathbf{x}^r_{p,j}$) and variance ($\mathbf{x}^v_{p,j}$), provide more discriminative power than sparse statistics like weekly or monthly peaks. While coarse resampling is useful for long-term forecasting, accurate load pattern identification requires preserving daily shape characteristics.

Therefore, to ensure the integrity of the clustering analysis, we carefully selected two representative days ($D = 2$) during the summer of 2024 when no transformer switching events were recorded. This selection minimizes the risk of data contamination due to topology changes in the distribution network. We collected historical load profiles from 46 transformers across 12 substations during this period. Finally, we obtain $\mathcal{X}$ by computing the engineered feature vector $\mathbf{x}_{p,j}$.

B. Agglomerated Clustering of Transformers

The electrical transmission network exhibits an inherently hierarchical structure. Since a single data center facility can be connected to multiple upstream transformers, certain similarities in their load profiles are expected. To determine the optimal number of clusters, we apply the agglomerative clustering algorithm to the feature set $\mathcal{X}$, using weights $\alpha_s = 0.3$, $\alpha_r = 8$, and $\alpha_v = 8$. A lower weight is assigned to the shape parameter because it resides in a high-dimensional space ($\mathbb{R}^{24}$), whereas the daily range and variance parameters are scalar features in $\mathbb{R}$. This weighting strategy ensures that the clustering process is not dominated by the shape feature, thereby preserving the influence of variance and range characteristics in the final cluster formation.

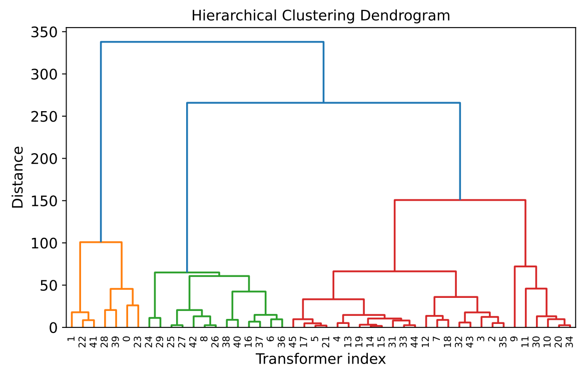

Fig. 2. Dendogram of Agglomerated Clustering on $m_3 = 46$

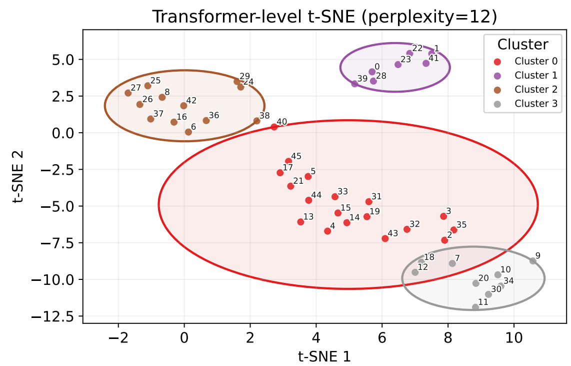

Fig. 3. Visualization of Clusters in K-Means Clustering

Figure 2 illustrates the hierarchical structure of the three dominant load pattern types identified through agglomerative clustering. The corresponding representative load profiles for each type are shown in Figure 4. To preserve data confidentiality and enhance interpretability, all load profiles are mean-centered and expressed in megawatts (MW) in the figure.

Type 1 clusters exhibit the highest mean daily range and display regular, periodic load shapes. A strong correspondence is observed in Figure 4 between peak load and peak ambient temperature when compared with hourly temperature data from the Northern Virginia region. This suggests that transformers classified under Type 1 predominantly serve residential and commercial customers.

Type 2 clusters demonstrate the lowest average daily range and exhibit relatively flat and irregular load profiles compared to Type 0 transformers. These patterns are indicative of large hyperscaler data centers, typically drawing power in excess of 100 MW.

Type 0 clusters show highly irregular load profiles, with daily average range and variance values similar to those of Type 2. These patterns often peak between 12:00 p.m. and 6:00 p.m., making the timing of peak load highly unpredictable.The irregularity of the load and peak load characteristics suggest

979-8-3315-8138-1

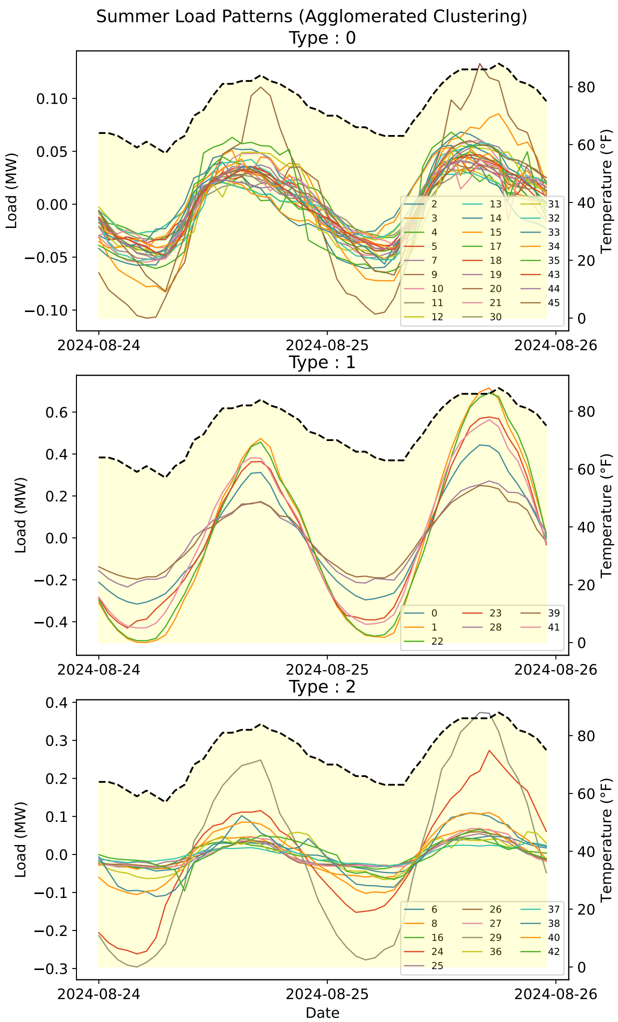

Fig. 4. Shapes of load types obtained using Agglomerated clustering

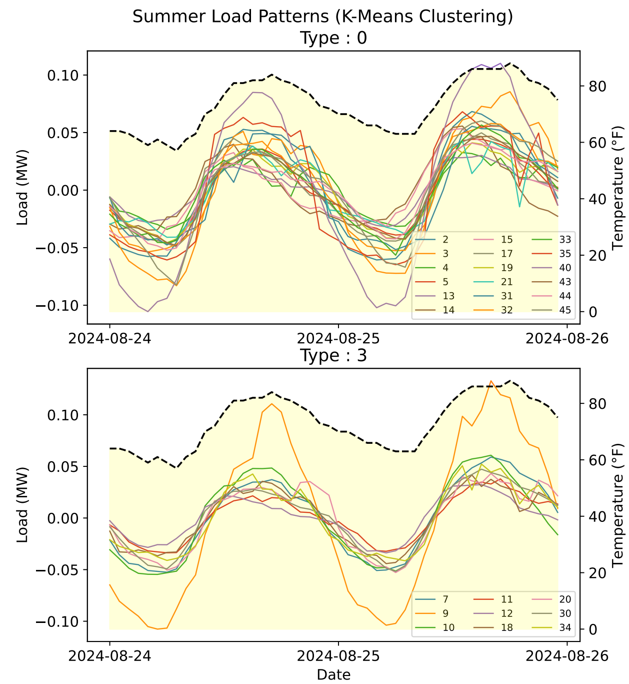

Fig. 5. Shapes of load types obtained using K-Means clustering (results for Type 0 and Type 3)

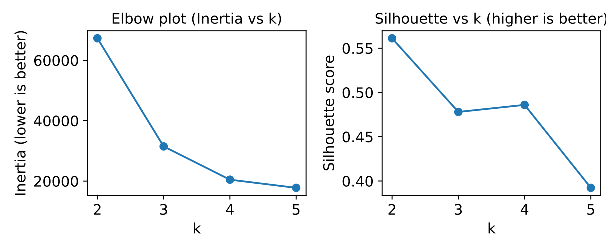

Fig. 6. Elbow silhouette plot for increasing k in K-Means Algorithm

that Type 0 clusters are primarily associated with edge and colocation data center customers.

Table I summarizes the average statistical characteristics of each identified load pattern type. During summer, non-conforming loads exhibit a strong average correlation with ambient temperature due to substantial cooling requirements, while maintaining distinct differences in average range and variance compared to conforming loads. It is important to note that, as a transmission-level utility, Dominion Energy has access only to SCADA data from high-voltage distribution transformers connected to the transmission network. Consequently, precise classification of data center types is not feasible, as downstream distribution networks may host a mix of residential and various data center facilities connected to a single transformer. At this aggregation level, the dominant power-consuming entity tends to shape the observed load pattern. Nonetheless, load pattern analysis at the transmission level provides valuable insights for explainable feature extraction, which is critical for day-ahead and short-term load forecasting. Moreover, this approach offers a pathway for forecasting load behavior at newly commissioned transformers that lack sufficient historical data.

TABLE I
CHARACTERISTIC OF SUMMER LOAD PATTERN TYPES

Algorithm	Type	Range	Variance	Temperature correlation
Agglomerated	0	3.45	1.84	0.91
	1	13.44	7.39	0.96
	2	2.47	1.32	0.90
K-Means	0	3.04	1.64	0.91
	1	13.44	7.39	0.96
	2	2.30	1.22	0.90
	3	4.40	2.31	0.93

TABLE II
CHARACTERISTIC OF WINTER LOAD PATTERN TYPES

Algorithm	Type	Range	Variance	Temperature correlation
Agglomerated	0	1.23	0.63	0.05
	1	6.29	3.01	-0.27
	2	1.31	0.67	-0.007

C. K-Means and Non-Conforming Load Pattern Analysis

To further interpret the clustering results, we visualized the transformer-level feature vectors $\mathbf{x}_{p,j}$ using the t-distributed Stochastic Neighbor Embedding (t-SNE) technique [6], which projects high-dimensional data into a two-dimensional space

979-8-3315-8138-1

while preserving local structure. Figure 3 illustrates the four clusters identified by the K-Means algorithm. Notably, the transformer indices classified as Type 1 and Type 2 closely match those identified by the Agglomerative clustering method shown in Figure 4, with the exception of transformer index 40, which is assigned to Type 0 in the K-Means result.

A closer inspection of Figure 3 reveals that several transformers previously grouped under Type 0 in the Agglomerative clustering are further separated into a distinct **Type 3** cluster by K-Means. This new Type 3 cluster exhibits higher daily range and variance compared to Type 0, along with smoother load profiles. These observations suggest that K-Means provides finer granularity in distinguishing load behaviors, particularly among non-conforming patterns. Despite minor differences in cluster assignments, both algorithms yield consistent results when applied to the same engineered feature set $\mathbf{x}_{p,j}$, reinforcing the robustness of the clustering framework.

D. Winter Load Patterns by the Summer Month Labels

The clustering analysis initially focused on load profiles from summer months, selecting two representative days without any recorded switching events to ensure data integrity. Summer load patterns were more distinguishable due to pronounced diurnal variations. Understanding load behavior during peak winter months is equally important, as residential electricity consumption typically exhibits a negative correlation with ambient temperature. Figure 7 illustrates winter load patterns classified using the labels derived from Agglomerative clustering based on summer data, while Table II summarizes the statistical characteristics of each cluster type during winter.

Our observations indicate that overall variance in winter load profiles is significantly lower than in summer across all cluster types. Specifically, Type 0 and Type 2 loads exhibit similar ranges and variances, with most transformers maintaining nearly constant energy consumption throughout the day. This stability is attributed to the reduced cooling requirements during winter, as data center loads, primarily driven by cooling, remain largely unaffected by temperature fluctuations. Consequently, as shown in Table II non-conforming loads (Types 0, 2, and 3) show minimal correlation with hourly temperature, whereas seasonal conditions (summer, winter, and off-peak periods) exert a notable influence on overall load patterns (high correlation in Table I). In contrast, conforming loads demonstrate strong positive correlation with temperature during summer and strong negative correlation during winter, reflecting typical residential heating and cooling behavior.

V. Concluding Remarks

This study presents a comprehensive analysis of conforming and non-conforming load patterns within Dominion Energy's transmission network. By applying both Agglomerative and K-Means clustering algorithms to transformer-level SCADA data, we successfully identify multiple non-conforming load behaviors and interpreted them in the context of known data center ownership models. Despite limited visibility into behind-the-meter electrical configurations, the clustering framework enables effective classification of aggregated transformer loads into conforming and non-conforming categories. Furthermore, we examine seasonal variations in load patterns by comparing summer and winter load patterns. The results align with expected heating and cooling behaviors, i.e. residential loads show strong temperature-dependent correlations, while data center loads remain relatively stable due to their operational characteristics and cooling requirements.

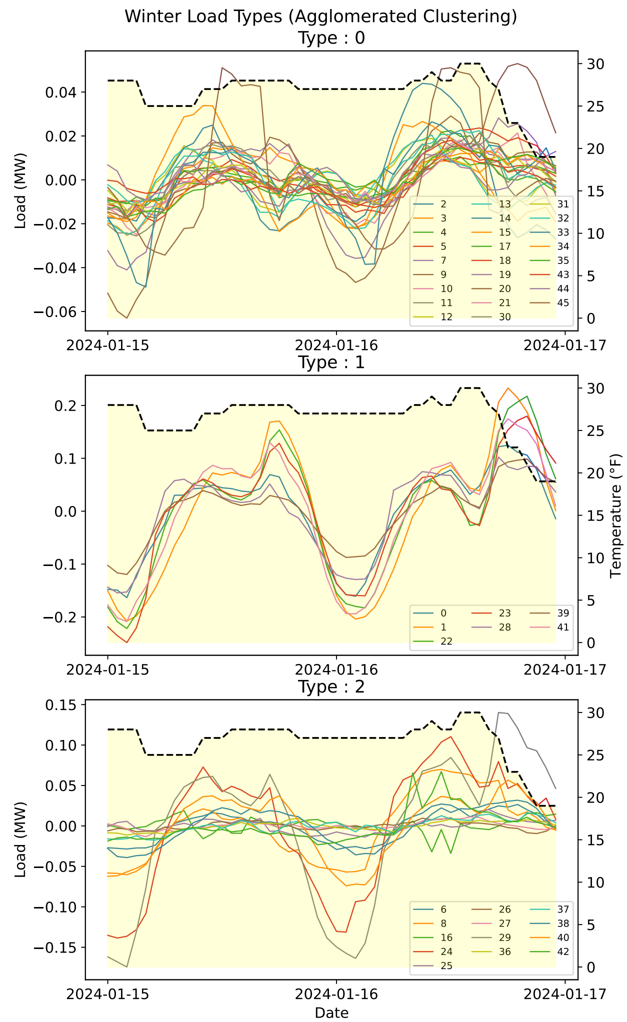

Fig. 7. Winter load shapes

References

[1] D. Mytton and M. Ashtine, "Sources of data center energy estimates: A comprehensive review. joule 6, 2032–2056," 2022.

[2] M. Vincent and D. Chernicoff, "Land and expand: Virginia data centers," *Data Center Frontier*, 2024.

[3] Cushman and Wakefield, "2025 global data center market comparison," *Vantage Point*, 2025.

[4] S. Lloyd, "Least squares quantization in pcm," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[5] D. Müllner, "Modern hierarchical, agglomerative clustering algorithms," 2011.

[6] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008.

979-8-3315-8138-1