Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 12, n. 1, 2026.

Trabalho apresentado no XLIV CNMAC, Fundação Getulio Vargas - Rio de Janeiro - RJ, 2025.

Enhancing Energy Consumption Time Series Clustering with Variational Mode Decomposition

Guillermo Benítez[1], Diego H. Stalder[2], Hans R. E. Mersch Fernandez[3], Carlos Sauer[4]
Facultad de Ingeniería, Universidad Nacional de Asunción, Paraguay

The extensive use of Smart Metering Systems has revolutionized the monitoring and analysis of energy consumption data, enabling the derivation of representative demand profiles [1]. The profiles are essential for optimizing energy distribution, however, clustering energy usage profiles presents a significant challenge, particularly when data inconsistencies arise from unsynchronized timestamps in records.

Clustering algorithms need to separate noise from meaningful components while also managing the dimensionality of the data. This work proposes the use of Variational Mode Decomposition (VMD) [2], which isolates intrinsic mode functions (IMFs), providing a clearer representation of energy consumption patterns. By integrating VMD with K-means clustering, we aim to improve the accuracy of demand profile classification.

In this work we consider a real data sample from 2021 done by the National Electricity Administration of Paraguay (ANDE), encompassing data on energy use from 122 households in the Asunción metropolitan area [3]. First, the Lomb-Scargle Periodogram (LSP) is used to identify dominant periods in the real energy consumption data. Based on these periods, a synthetic energy demand profile is generated to aid in determining the optimal number of modes for VMD. The VMD algorithm is then applied to decompose the signals into IMFs, solving the following minimization problem:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[(\delta(t) + j\pi t) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \tag{1}$$

where u_k are the modes, ω_k are the center frequencies, and $\delta(t)$ is the Dirac delta function.

The mean squared error (MSE) was used to determine the optimal number of decomposition modes K. By comparing the results for 10 modes with an ideal curve based on the periods identified in the periodogram (24, 12, 8, and 6 hours), K=4 achieved the lowest MSE value of 3.3944, confirming it as the optimal number of modes. The signals are reconstructed using the optimal VMD decomposition, resulting in a denoised dataset, which is then aggregated into daily profiles by calculating the average hourly consumption for each user.

Then K-means Clustering (KMC) is applied, with the optimal number of clusters determined using the elbow method. These profiles are subsequently compared with those obtained from other clustering methods using validation indices such as the Silhouette score and the Calinski-Harabasz score to assess clustering quality. The clustering analysis of hourly energy consumption data identified six distinct user groups, as shown in Figure 1. Some clusters exhibit pronounced peaks in energy usage at specific hours, while others maintain a more stable consumption pattern throughout the day. These findings suggest that different user groups follow distinct energy usage routines.

[1] gdbenitezs@fiuna.edu.py
[2] dstalder@ing.una.py
[3] hmersch@fiuna.edu.py
[4] csauer@ing.una.py

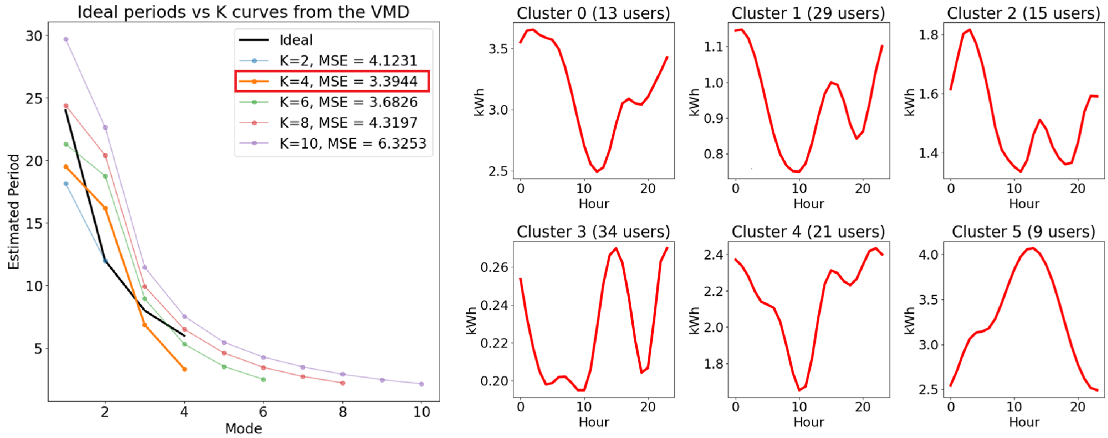

Figure 1: From left to right. Ideal curve obtained with data from Lomb-Scargle Periodogram, compared with different values of decomposition K, the minimum MSE corresponds to K=4. Hourly energy consumption profiles grouped into six clusters. Source: Own elaboration.

The performance of KMC+VMD was compared with other clustering methods obtained in [3], eg Spectral Clustering with Feature Agglomeration (SC+FA), Spectral Clustering with Principal Component Analysis (SC+PCA), with Principal Component Analysis (KMC+PCA), and KMC with Feature Agglomeration (KMC+FA). Table 1, KMC+VMD achieved the highest scores among the tested methods, achieving a Silhouette Score of 0.4334 and a Calinski-Harabasz Score of 161.7985.

Table 1: Silhouette Score and Calinski-Harabasz Score.

Scores	SC+FA	SC+PCA	KMC+PCA	KMC+FA	KMC+VMD
SS	-0.0484	-0.0286	0.1930	0.1888	0.4334
CHS	10.9507	9.6678	22.3376	22.4650	161.7985

The integration of VMD with KMC demonstrates potential for clustering energy consumption time series. These preliminary results suggest that the method can help generate representative consumption profiles, potentially supporting energy management strategies, as indicated by the Silhouette and Calinski-Harabasz scores.

References

[1] F. Morales et al. “Analysis of Electric Energy Consumption Profiles Using a Machine Learning Approach: A Paraguayan Case Study”. In: **Electronics** 11.2 (2022), pp. 267–268. DOI: `10.3390/electronics11020267`.

[2] K. Dragomiretskiy and D. Zosso. “Variational mode decomposition”. In: **IEEE Transactions on Signal Processing** 62.3 (2014), pp. 531–544. DOI: `10.1109/TSP.2013.2288675`.

[3] H. R. E. Mersch Fernandez, D. H. Stalder, C. Sauer, and F. Morales. “Clustering energy consumption time series using singular spectrum analysis and K-means clustering”. In: **Proceeding Series of the Brazilian Society of Computational and Applied Mathematics** 11.1 (Sept. 2024), pp. 1–2. ISSN: 2359-0793. URL: `https://proceedings.sbmac.emnuvens.com.br/sbmac/article/view/4769`.