



Identificación de anomalías y tendencias estacionales sobre la demanda eléctrica en instalaciones universitarias mediante algoritmos de clustering

Anomaly and seasonal trend identification of electricity demand in university facilities using clustering algorithms

Patricio G. Donato¹ , Marcos A. Funes¹ , Carlos M. Orallo¹ 

¹ Universidad Nacional de Mar del Plata. Consejo Nacional de Investigaciones Científicas y Tecnológicas. Instituto de Investigaciones Científicas y Tecnológicas en Electrónica. Mar del Plata, Argentina.

DOI: 10.64966/ingeniare.v34.09

Resumen

El aumento constante de la demanda de energía, impulsado por el crecimiento poblacional, la expansión industrial y la digitalización, ejerce una presión creciente sobre los sistemas de distribución eléctrica. En este contexto, en los últimos años se han incorporado algoritmos de aprendizaje automático para el procesamiento de datos eléctricos y la optimización operativa de la red. Evaluar el desempeño de estos algoritmos con datos del mundo real es esencial para comprender su potencial en aplicaciones concretas. Este artículo presenta un estudio de series temporales de demanda eléctrica de dos edificios públicos mediante algoritmos de clustering no supervisados (K-Means, Fuzzy C-Means y Gaussian Mixture Models). Si bien estos algoritmos se han empleado ampliamente en diferentes campos de la ingeniería, las bases de datos de demanda eléctrica analizadas exhiben una combinación de patrones estacionales y eventos anómalos que pueden dificultar la correcta identificación de grupos y, en consecuencia, afectar su agrupamiento. La calidad de los resultados se evalúa mediante los índices Silhouette y Davies-Bouldin. El objetivo de este estudio es explorar la capacidad de las técnicas de clustering para extraer información relevante de series temporales de consumo que abarcan varios años, sentando las bases para futuras investigaciones orientadas a la optimización de la gestión de la energía en edificios públicos y en otros clientes con perfiles de consumo similares.

Palabras clave: Algoritmos de clustering, aprendizaje automático, K-Means, Fuzzy C-Means, modelos de mezclas Gaussianas, demanda de energía, perfiles de consumo.

Abstract

The continuous increase in energy demand, driven by population growth, industrial expansion, and digitalization, places increasing pressure on electrical distribution systems. In this context, machine learning algorithms have been increasingly adopted in recent years to process electrical data and improve the operational performance of power networks. Evaluating their performance using real-world datasets is essential to understanding their potential in practical applications. This article presents a study of electrical demand time series from two public buildings using unsupervised clustering algorithms (K-Means, Fuzzy C-Means, and Gaussian Mixture Models). Although these algorithms have been widely applied in various engineering domains, the analyzed electrical demand datasets exhibit a combination of seasonal patterns and anomalous events that can hinder accurate cluster identification and, consequently, affect clustering quality. The quality of the results is assessed using the Silhouette and Davies-Bouldin indices. This study aims to explore the ability of clustering techniques to extract relevant information from multi-year consumption time series, laying the groundwork for future research focused on optimizing energy management in public buildings and in other customers with similar demand profiles.

Keywords: Clustering algorithms, machine learning, K-Means, Fuzzy C-Means, Gaussian mixture models, power demand, demand profiles.

INTRODUCCIÓN

El incremento constante de la demanda de electricidad, impulsado por el crecimiento demográfico y los hábitos de consumo de la sociedad moderna, crea incertidumbre sobre el escenario energético del futuro. La solución a este problema no sólo pasa por implantar dispositivos de generación de electricidad basados en fuentes renovables, sino también por reducir la demanda a través de un consumo más racional y eficiente. Ambas cuestiones forman parte del concepto más amplio de las Redes Eléctricas Inteligentes (REI), las cuales combinan la infraestructura eléctrica tradicional con avanzadas tecnologías de la información y comunicaciones (TIC), sistemas de generación distribuida (basados en fuentes de energía renovables) y almacenamiento de energía [1], [2]. Las REI recolectan datos en tiempo real mediante medidores inteligentes y otros dispositivos de medición, actuando sobre la red en forma remota a través de dispositivos como interruptores, relés y convertidores de potencia. En la actualidad incluso se plantea un concepto más amplio denominado superredes inteligentes, donde el modelo se escala a un nivel global [3]. La transición de las redes tradicionales a las REI es un proceso complejo que requiere un sólido soporte de comunicaciones para gestionar grandes volúmenes de datos procedentes de múltiples puntos de la red, junto con hardware y software avanzados para procesar y extraer la información pertinente.

Las REI producen grandes cantidades de datos, lo que puede conducir a lo que se denomina como escenarios ricos en datos, pero pobres en información [4]. La complejidad y heterogeneidad inherentes a estos datos, sumadas a la falta de formatos y herramientas analíticas estandarizadas, dificultan su aprovechamiento para la gestión eficiente de la red y la toma de decisiones informada. Como respuesta a este problema, en los últimos años ha crecido el interés en el uso de algoritmos de aprendizaje automático para extraer información valiosa de las bases de datos de parámetros eléctricos. Entre las aplicaciones más destacadas se encuentran la detección y localización de fallos en redes de distribución de baja tensión [5], la detección de casos de hurto de electricidad [6] y el pronóstico de potencia a corto plazo [7], por mencionar algunos. En este contexto, los algoritmos de *clustering*, también conocidos como algoritmos de agrupamiento (a lo largo del texto se adoptará la expresión en inglés de *clustering* por ser la más habitual en el contexto del aprendizaje automático), resultan especialmente relevantes, ya que permiten organizar grandes volúmenes de datos en grupos o *clusters* homogéneos en función de sus características [8]. Estos algoritmos son especialmente eficaces para descubrir patrones ocultos, segmentar datos, detectar anomalías, predecir comportamientos y optimizar las operaciones de red tanto en redes de distribución de baja tensión [9], [10], [11] como en redes de transmisión [12]. En el caso particular del consumo eléctrico en redes de distribución, existen numerosos estudios que abordan la segmentación de clientes en función de sus perfiles de carga diaria o la detección de demandas inusuales [13], [14], [15], [16]. La mayoría de estos trabajos se ha centrado en grandes bases de clientes, mientras que son menos frecuentes aquellos que analizan el comportamiento de un único consumidor a lo largo de períodos prolongados. Especialmente si se consideran demandas dependientes de variables como la población, estación del año y/o tipo de actividades.

Adicionalmente, en el contexto específico de América Latina, el despliegue de medidores inteligentes no ha sido ni armonizado ni masivo, lo que se traduce en una limitada disponibilidad y

accesibilidad de los datos [17]. Esta escasez de datos hace que la aplicación de algoritmos de *clustering* en el sector energético, particularmente en edificios públicos y universitarios, sea relativamente infrecuente. Centrar el análisis en estos conjuntos de datos menos explorados permite identificar algunos comportamientos de consumo particulares e inusuales [18]. Estos hallazgos podrían tener importantes implicaciones para la reducción de costos y la mejora en la eficiencia energética en edificios similares. Además, las características únicas de estas instalaciones requieren una consideración especial de factores como la estacionalidad, la periodicidad y la aparición de eventos anómalos.

En este trabajo se presentan los resultados de un estudio sobre bases de datos de series temporales de consumo eléctrico de dos edificios diferentes de la Universidad Nacional de Mar del Plata (UNMDP), en Argentina. Un aspecto interesante de estas bases de datos es que, si bien tienen un consumo similar al de una pequeña industria o establecimiento comercial, exhiben ciclos de funcionamiento en los cuales la actividad cambia abruptamente. Estos incluyen períodos de vacaciones, feriados y cierres ocasionales, que impactan en el comportamiento de los datos. El objetivo es evaluar la capacidad para identificar comportamientos anómalos y patrones relevantes mediante los algoritmos de *clustering* K-Means, Fuzzy C-Means y el modelo de mezclas Gaussianas. Este estudio incluye, además de la identificación de comportamientos particulares, el análisis de la calidad del *clustering*, mediante el cálculo de índices específicos, en función de la estacionalidad de los datos, los patrones de consumo atípicos causados por cambios en el comportamiento de uso y los patrones anómalos de tiempo de uso. La comparación entre estos algoritmos busca poner de relieve cómo cada uno de ellos enfatiza aspectos diferentes de los datos, lo cual resulta útil en diversos contextos.

METODOS DE CLUSTERING

Los algoritmos de *clustering* sirven para agrupar grandes volúmenes de datos en conjuntos homogéneos (*clusters*) basados en características similares, cada uno de ellos caracterizado por un centroide. Estos centroides son puntos en un espacio multidimensional que representan los centros de cada cluster. Se utilizan para determinar los límites de los *clusters* al servir como puntos de referencia en torno a los cuales se agrupan los datos. Los algoritmos de *clustering* son muy útiles para la identificación de patrones ocultos, la detección de anomalías y la segmentación de datos. Entre los numerosos algoritmos de *clustering* reportados en la bibliografía específica [19], [20], este trabajo selecciona tres que se describen a continuación.

K-Means (KM): Este algoritmo divide un conjunto de datos en K grupos definidos por la distancia euclidiana entre cada dato individual y los centroides de cada grupo [21]. Se inicializa con K centroides elegidos aleatoriamente, y cada dato del conjunto se asigna al centroide más cercano a él (véase el ejemplo de la Figura 1 para un caso con dos variables independientes). A continuación, el algoritmo vuelve a calcular la posición de los centroides como la media de los puntos asignados a cada grupo y repite este proceso hasta que los centroides convergen o hasta que se alcanza un número máximo de iteraciones. El algoritmo K-Means está diseñado para minimizar la varianza intragrupo, es decir, intenta garantizar que los puntos de un grupo sean lo más parecidos posible entre sí y que cada grupo esté representado por su centroide.

Fuzzy C-Means (FCM): Este algoritmo puede considerarse como una variante de K-Means que incorpora el concepto de lógica difusa, permitiendo que los datos pertenezcan a varios *clusters* simultáneamente, asignando un valor de pertenencia difuso a cada uno en lugar de una asignación binaria [22]. Se basa en la minimización de una función objetivo que considera la distancia euclidiana ponderada intracluster entre cada punto y los centroides de los *clusters*, utilizando una función de pertenencia, m , que indica el grado de pertenencia de cada punto a cada clúster (Figura 1). En cada iteración del algoritmo, se recalculan los centroides y la matriz de pertenencia hasta alcanzar la convergencia. Fuzzy C-Means es especialmente útil cuando los datos tienen una pertenencia ambigua a varios conglomerados. Cuando la aplicación lo requiere, se pueden convertir las asignaciones difusas en binarias, asociando cada dato exclusivamente a un único conglomerado. Este proceso se realiza transformando la matriz de pertenencia difusa en una matriz binaria, en la que cada punto se asigna exclusivamente al clúster con el valor de pertenencia más alto.

Modelo de Mezclas Gaussianas (GMM, Gaussian Mixture Models): El modelo es un algoritmo probabilístico de *clustering* basado en el supuesto de que los datos se generan a partir de una mezcla de varias distribuciones gaussianas [23]. Su objetivo es ajustar estas distribuciones de forma que se maximice la probabilidad de los datos dados. Durante el proceso de entrenamiento, el algoritmo estima la probabilidad de que cada dato pertenezca a cada uno de los componentes gaussianos mediante el algoritmo de maximización de expectativas. Los modelos de mezclas gaussianas se caracterizan por su flexibilidad a la hora de modelar conglomerados de formas y tamaños variables, y por su capacidad para asignar una pertenencia probabilística a cada punto de datos de cada conglomerado (Figura 1). El GMM suele funcionar bien en conjuntos de datos en los que los conglomerados presentan formas elipsoidales y no están necesariamente bien separados. Sin embargo, es muy sensible a la inicialización, lo que puede dar lugar a soluciones subóptimas. Además, su complejidad computacional crece de manera lineal con el número de muestras y componentes, y también con la dimensionalidad de los datos, lo que puede volverlo impracticable en bases de datos con millones de instancias o con un gran número de variables.

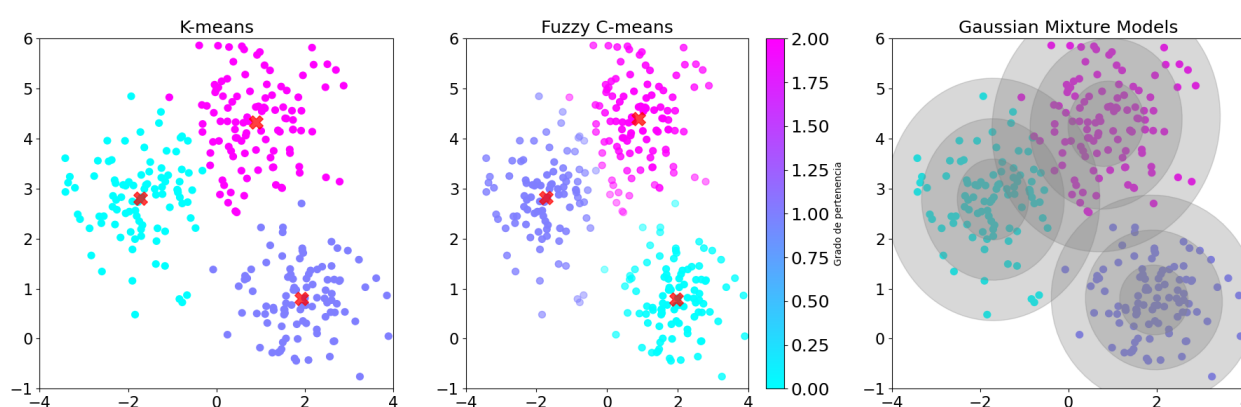


Figura 1. Ejemplo cualitativo de aplicación de algoritmos de clustering a un conjunto de datos aleatorios, donde los colores señalan el clúster al que pertenecen, las cruces de color representan los centroides y las elipses muestran las distribuciones Gaussianas de GMM.

Índices de calidad y ubicación de los centroides

Una de las herramientas habituales usadas para evaluar un determinado agrupamiento es el cálculo de índices de calidad. Se han definido diferentes índices para evaluar distintos aspectos de los *clusters*, pero los más conocidos y utilizados (que proporcionan una evaluación más amplia de un *clustering* determinado) son los índices de Silhouette y Davies-Bouldin [24]. El índice de Silhouette evalúa la cohesión y separación de los *clusters*, determinando en qué medida los datos se agrupan de forma natural [25]. Asigna un valor a cada dato en función de su similitud con su propio clúster en comparación con otros *clusters*. Un valor cercano a 1 indica que el *clustering* está bien definido y que cada dato está correctamente asignado a su clúster. Un valor cercano a 0 indica que los datos se encuentran en el límite entre dos *clusters*, por lo que su asignación es ambigua. Un valor cercano a -1 indica una asignación incorrecta, lo que significa que el punto está más cerca de un conglomerado distinto del suyo.

En cambio, el índice de Davies-Bouldin mide la similitud media entre cada clúster y su clúster más similar, al tiempo que tiene en cuenta la cohesión interna dentro de cada clúster [26]. Este índice parte de la premisa de que un buen *clustering* tiene *clusters* compactos y bien separados. Se calcula utilizando dos componentes: la dispersión *intracluster* (la distancia media entre puntos dentro del mismo clúster) y la dispersión *intercluster* (la distancia entre los centroides de diferentes *clusters*). Este índice puede tomar cualquier valor ≥ 0 , donde un valor cercano a cero indica que los *clusters* están bien definidos y son distintos. El valor ideal es cero, que representa una separación y cohesión perfectas.

Además de evaluar estos índices, resulta relevante examinar la ubicación de los centroides. La evolución temporal de los centroides proporciona información valiosa sobre la magnitud de las diferencias entre *clusters* y ofrece información práctica sobre las variables en cuestión, tal como se verá más adelante.

CASO DE APLICACIÓN

Descripción general de las bases de datos de demanda de energía

Las bases de datos utilizadas para este análisis corresponden a las series temporales de datos de demanda de energía eléctrica de los edificios de la Facultad de Ingeniería (FI) y el Instituto de Investigaciones en Ciencia y Tecnología de Materiales (INTEMA), ambos pertenecientes a la Universidad Nacional de Mar del Plata (UNMDP) pero ubicados en dos zonas muy diferentes de la ciudad, a casi 8 km uno del otro (Figura 2). El edificio FI está emplazado en la zona urbana, protegido del viento y tiene una población mixta de estudiantes, profesores e investigadores. Se conecta a la red pública de distribución de electricidad de baja tensión, pero también cuenta con una instalación fotovoltaica en el tejado con una capacidad de generación de aproximadamente 4 kW. Por otro lado, el edificio INTEMA está situado en una zona suburbana de la ciudad, por lo que está más expuesto al viento, y su población está compuesta mayoritariamente por investigadores. Además, la red de distribución de baja tensión donde está conectado INTEMA es relativamente débil, lo que se refleja en una mala calidad del servicio eléctrico, situación que fue detectada en un estudio previo [27]. Las bases de datos incluyen medidas de potencia activa y reactiva, abarcando el periodo comprendido entre enero de 2021 y junio de 2025, con muestras adquiridas

cada 15 minutos. La adquisición de las muestras se realizó a través de medidores inteligentes DIMET3-G-CT empleados por la empresa distribuidora local y en ambos casos corresponde a la acometida principal del edificio, por lo que comprende el consumo total. En la base de datos del edificio FI se computaron 262 muestras faltantes (0,17% del total de muestras), y en la base de datos de INTEMA se computaron 749 datos faltantes (0,48% del total). Las muestras faltantes corresponden a dos días de 2025 (domingo 2 de febrero y jueves 11 de junio) y conjuntos de muestras dispersas en diferentes meses y años. En todos los casos se trata de faltantes producto de problemas de comunicación entre los medidores y la base de datos. Considerando que el objeto del trabajo es agrupar mediciones reales, y que las muestras faltantes son una fracción menor al 0,5%, no se consideró necesario rellenar las series temporales. En la [Tabla 1](#) se presentan las principales características de ambos conjuntos de datos, y en la [Figura 3](#) y [Figura 4](#) se muestran las curvas de demanda de potencia en función del tiempo para ambas bases de datos.



Figura 2. Ubicación de los edificios analizados en este trabajo (FI e INTEMA).

Tabla 1. Características principales de las bases de datos empleadas en este estudio.

	FI	INTEMA
Número de muestras	157370	156833
Potencia Activa Máxima (kW)	104.98	315.93
Potencia Activa Media (kW)	21.79	61.53
Desviación estándar de la Potencia Activa (kW)	13.76	36.71
Potencia Reactiva Máxima (kVAR)	21.99	62.01
Potencia Reactiva Media (kVAR)	8.68	8.56

Tabla 1. Características principales de las bases de datos empleadas en este estudio. (continuación)

	FI	INTEMA
Desviación estándar de la Potencia Reactiva (kVAR)	3.49	3.39

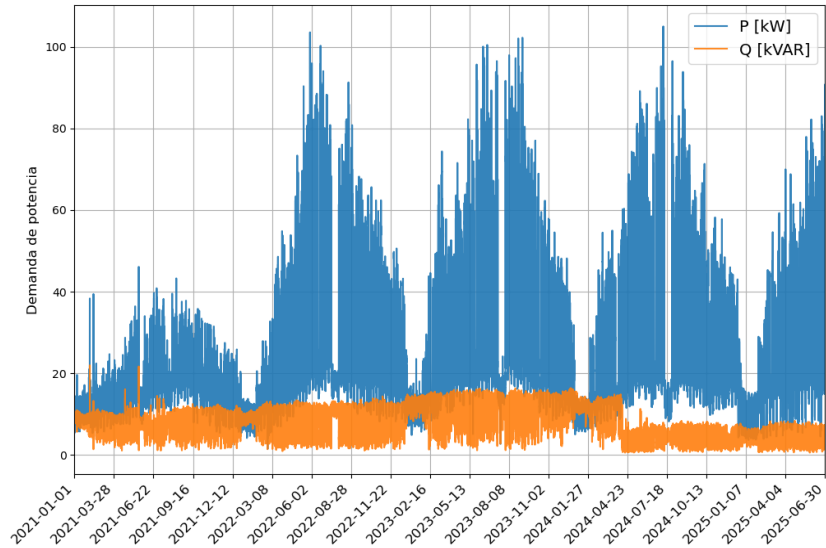


Figura 3. Demanda de potencia activa y reactiva del edificio FI (enero 2021 - junio 2025).

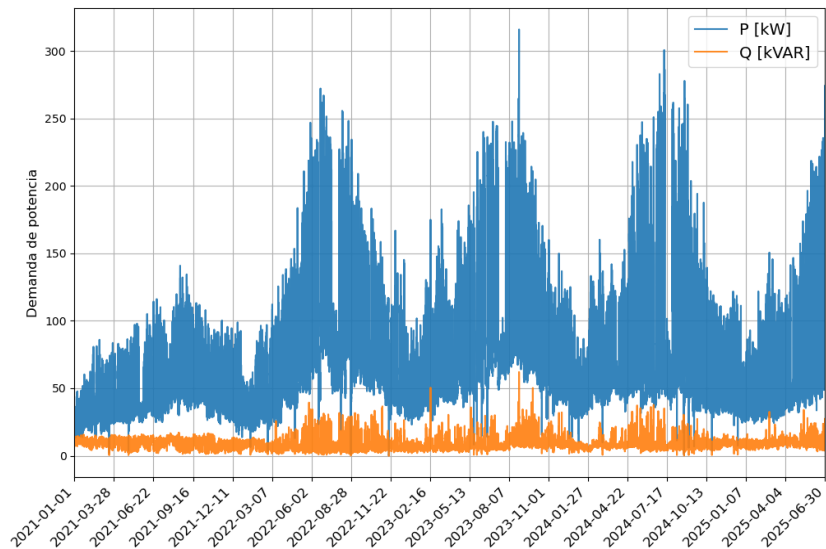


Figura 4. Demanda de potencia activa y reactiva del edificio INTEMA (enero 2021 - junio 2025).

Procesamiento del conjunto de datos

Los datos se estandarizaron antes de aplicar cada algoritmo de *clustering*. En este contexto, estandarizar significa transformar las variables de entrada para que todas tengan una media igual a cero y una desviación estándar unitaria, utilizando el método StandardScaler de la librería Scikit-learn de Python [28]. De esta forma, las magnitudes de potencia activa y reactiva, que original-

mente tienen escalas diferentes, se representan en una misma escala estadística. Esta operación es fundamental porque garantiza que todas las características contribuyan de manera uniforme al proceso de agrupamiento, evitando que aquellas con valores absolutos más grandes (como la potencia activa) dominen sobre las demás y distorsionen los resultados.

El intervalo temporal de ambas bases de datos incluye años con comportamientos diferentes, como el año 2021, en el que muchas de las directivas de restricción de movilidad y confinamiento seguían en vigor debido a la pandemia de Covid-19. En cambio, de 2022 a 2025 hubo un funcionamiento normal, con plena asistencia en aulas, despachos y ámbitos de investigación. La estandarización facilitó la comparación entre estos períodos, reduciendo la influencia de diferencias de escala y dispersión numérica en la asignación de *clusters*.

Los datos de entrada de cada algoritmo consistieron en vectores compuestos por medidas de potencia activa (P) y potencia reactiva (Q). Se realizaron algunas pruebas añadiendo un tercer vector que contenía datos de potencia aparente (S), pero los resultados no difirieron significativamente del caso en el que sólo se utilizaron P y Q, en el que la complejidad es menor. Esto puede explicarse por la relación directa entre la potencia aparente y las potencias activa y reactiva, donde $S^2 = P^2 + Q^2$. Por el contrario, los comportamientos de P y Q no son redundantes entre sí, tal como se puede apreciar en la evolución temporal de ambas variables (Figura 3 y Figura 4). Inicialmente, el análisis se planificó por días, aplicando el *clustering* a todos los lunes del año, a todos los martes, y así sucesivamente. Sin embargo, este planteamiento se ve influido por la estacionalidad, ya que las características del consumo eléctrico varían mucho entre un día laborable de enero y uno de junio o septiembre. A modo de ejemplo se puede mencionar que un lunes de enero puede tener demandas máximas de hasta 15,5 kW (6/1/2025), mientras que uno de junio puede llegar a 100 kW (12/6/2023) y uno de septiembre hasta 93,7 kW (4/9/2023). En consecuencia, los agrupamientos resultantes no siempre son intuitivos, lo que dificulta la interpretación física de los agrupamientos obtenidos. Por este motivo el proceso de *clustering* se realizó de manera independiente para cada mes del período 2021-2025, mitigando el problema de la estacionalidad, ya que los perfiles de consumo dentro de un mismo mes presentan una mayor homogeneidad temporal y estacional. La estandarización también se realizó en forma mensual, para evitar distorsionar los agrupamientos combinando conjuntos de datos con distribuciones estadísticas y niveles de demanda significativamente diferentes. En consecuencia, los índices de calidad Silhouette y Davies-Bouldin reportados a lo largo del artículo corresponden a evaluaciones mensuales del desempeño del *clustering*.

RESULTADOS

En este estudio, todos los algoritmos de *clustering* se configuraron con $K = 2$, dado que al incrementar el número de *clusters* se observó una notable disminución de la calidad de los agrupamientos. La Tabla 2 presenta los resultados de un ensayo realizado sobre ambas bases de datos, en el cual se calcularon los valores mínimo, promedio y máximo de los índices de Silhouette y Davies-Bouldin utilizando el algoritmo KM. Los valores se computaron a partir de los índices de calidad determinados para cada mes entre enero de 2021 y junio de 2025. Se observa que el índice de Silhouette tiene su mejor desempeño en todos los casos cuando el número de *clusters* es $K = 2$. En cuanto al índice de Davies-Bouldin, se observa que tanto el valor promedio como el

mínimo alcanzan su mejor desempeño con $K = 2$, y solamente se observa una leve mejoría para el caso del valor máximo cuando $K = 3$ (FI) o $K = 4$ (INTEMA). Considerando que, además, los patrones de consumo eléctrico típicamente muestran comportamientos dicotómicos (día/noche, día hábil/no hábil, etc.), se decidió emplear $K = 2$ para todo el estudio.

El algoritmo KM se configuró con un máximo de 300 iteraciones y una tolerancia $1e-4$, sin semilla fija para reproducibilidad. En lo que respecta al algoritmo FCM, se seleccionó el parámetro $m = 10$ como coeficiente de pertenencia, que determina el grado de solapamiento de los *clusters*, controlando hasta qué punto un dato puede pertenecer a más de un clúster simultáneamente. Al evaluar FCM con valores más bajos ($m \rightarrow 1$), se obtuvieron resultados casi idénticos a los de KM, como era esperable, lo que indica un solapamiento mínimo. Por el contrario, aumentar el grado de pertenencia m por encima de 10 no tuvo un impacto significativo en los resultados del *clustering*, lo que sugiere un punto de saturación en la sensibilidad del algoritmo a este parámetro. En la [Tabla 3](#) se muestran los resultados de un ensayo realizado sobre ambas bases de datos, donde se evidencia que valores de $m \rightarrow 1$ dan resultados similares a KM, mientras que valores superiores a 10 casi no muestran variaciones. Para la conversión de pertenencias difusas a etiquetas categóricas, se empleó el criterio de máxima pertenencia, asignando cada dato al clúster con la mayor pertenencia.

Tabla 2. Índices de Silhouette y Davies-Bouldin para el algoritmo KM empleando diferentes números de clusters (K). Se resaltan los casos donde el agrupamiento tiene mayor calidad (Silhouette $\rightarrow 1$, Davies-Bouldin $\rightarrow 0$).

Índice	Valor	FI			INTEMA		
		K = 2	K = 3	K = 4	K = 2	K = 3	K = 4
Silhouette	Mínimo	0,3377	0,3536	0,3318	0,3376	0,3373	0,3248
	Promedio	0,6632	0,5425	0,451	0,6374	0,552	0,4782
	Máximo	0,824	0,7496	0,6724	0,8101	0,761	0,7196
Davies-Bouldin	Mínimo	0,3259	0,4896	0,6295	0,4518	0,488	0,5583
	Promedio	0,5591	0,7575	0,8411	0,6958	0,7282	0,7719
	Máximo	1,1639	0,9826	0,9999	1,1675	1,0024	0,976

Tabla 3. Índices Silhouette y Davies-Bouldin para diferentes configuraciones del algoritmo FCM.

Índice	FI					INTEMA				
	<i>K-Means</i>	<i>m</i> = 2	<i>m</i> = 5	<i>m</i> = 10	<i>m</i> = 20	<i>K-Means</i>	<i>m</i> = 2	<i>m</i> = 5	<i>m</i> = 10	<i>m</i> = 20
Silhouette	0,66 ± 0,1	0,66 ± 0,1	0,64 ± 0,11	0,64 ± 0,12	0,63 ± 0,12	0,63 ± 0,11	0,63 ± 0,11	0,56 ± 0,13	0,54 ± 0,13	0,54 ± 0,13
Davies-Bouldin	0,56 ± 0,18	0,56 ± 0,18	0,59 ± 0,22	0,6 ± 0,22	0,61 ± 0,22	0,69 ± 0,16	0,76 ± 0,18	0,83 ± 0,2	0,87 ± 0,2	0,88 ± 0,19

Localización de los centroides

Los algoritmos de *clustering* producen dos resultados principales: la asignación de cada dato a un clúster específico y la determinación de las posiciones de los centroides. Los centroides representan los puntos del espacio de datos en torno a los cuales se forman los *clusters*. Sus ubicaciones proporcionan información sobre los patrones de comportamiento observados. En la [Figura 5](#) se

muestra un ejemplo de la ubicación de los centroides para un mes, un año y un edificio concretos. Se pueden observar los dos *clusters*, agrupados según el criterio K-Means, y la ubicación de los correspondientes centroides (P_1 , Q_1) y (P_2 , Q_2).

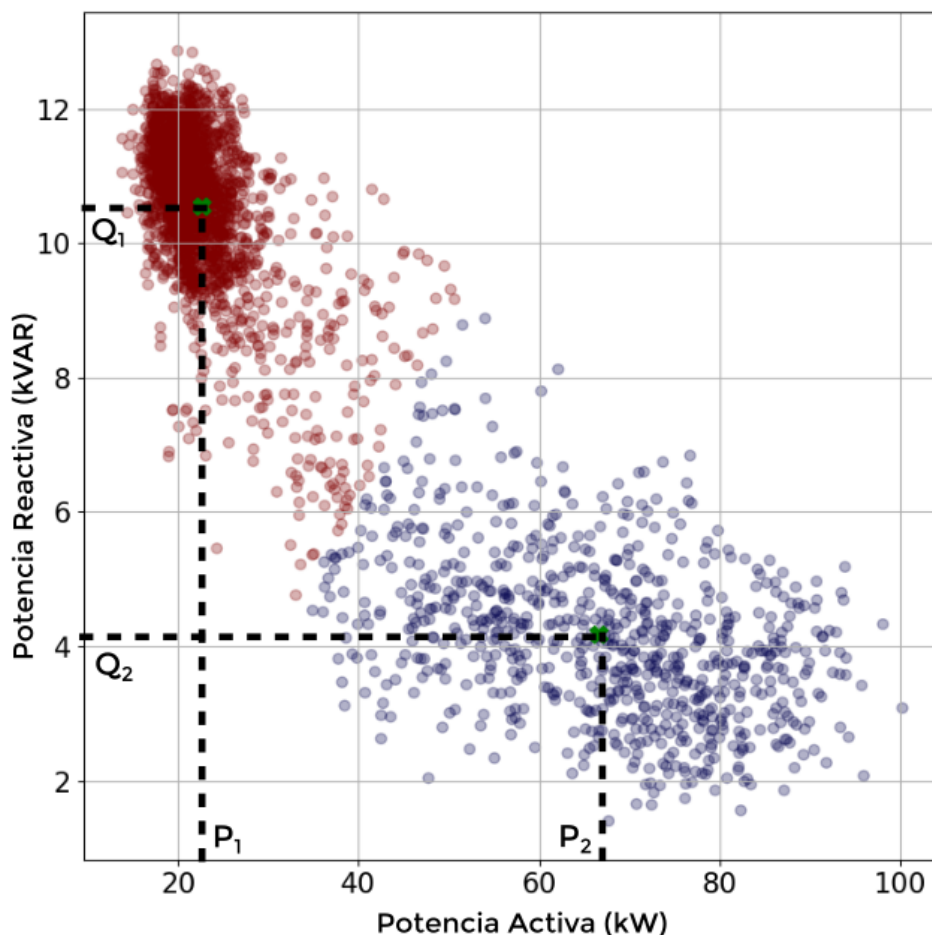


Figura 5. Gráfico de dispersión correspondiente al edificio FI para el mes de junio de 2022.

La [Figura 6](#) ilustra la evolución de las posiciones de los centroides de cada mes correspondientes a los edificios FI e INTEMA, considerando tanto la potencia activa como la reactiva durante el periodo comprendido entre enero de 2021 y junio de 2025. Del análisis de esta figura se pueden extraer las siguientes observaciones:

- En enero de cada año, los centroides de potencia activa de todos los algoritmos son casi idénticos en ambos edificios.
- Durante 2021 los centroides de potencia activa de ambos edificios muestran pocas diferencias. Este comportamiento se explica porque la demanda de energía ese año fue bastante plana y menor de lo habitual debido a la menor presencia de personal como consecuencia de las restricciones de movilidad durante la pandemia de COVID.
- Durante los meses de mayor consumo, de febrero a diciembre, los centroides muestran la mayor diferencia. Esto refleja el importante contraste entre los periodos de actividad e inactividad en estos meses.

- Salvo diferencias relativamente pequeñas, la ubicación de los centroides de potencia activa en los tres algoritmos es notablemente similar en el caso del edificio FI. Sin embargo, en el caso del edificio INTEMA, se observa una discrepancia significativa en julio de 2022, donde FCM sitúa ambos centroides de potencia activa casi en la misma posición, mientras que GMM y KM arrojan dos centroides claramente diferentes. En menor medida, esta diferencia también es apreciable en julio de 2023 y 2024. Este fenómeno puede explicarse por el receso invernal en Argentina durante ese mes, que impacta significativamente en el comportamiento de la demanda eléctrica, como se verá con más detalle en las secciones siguientes.
- El comportamiento de los centroides de potencia reactiva es diferente en ambos casos. En el edificio FI, presentan una forma de evolución estacional similar a la observada en la potencia activa. El cambio significativo en la ubicación de los centroides a partir de abril de 2024 se debe a la desconexión de una batería de condensadores obsoleta que venía afectando al factor de potencia de la instalación eléctrica. Por el contrario, los centroides de potencia reactiva en el edificio INTEMA muestran un comportamiento errático, sin que se aprecie a priori un patrón claro o consistente.

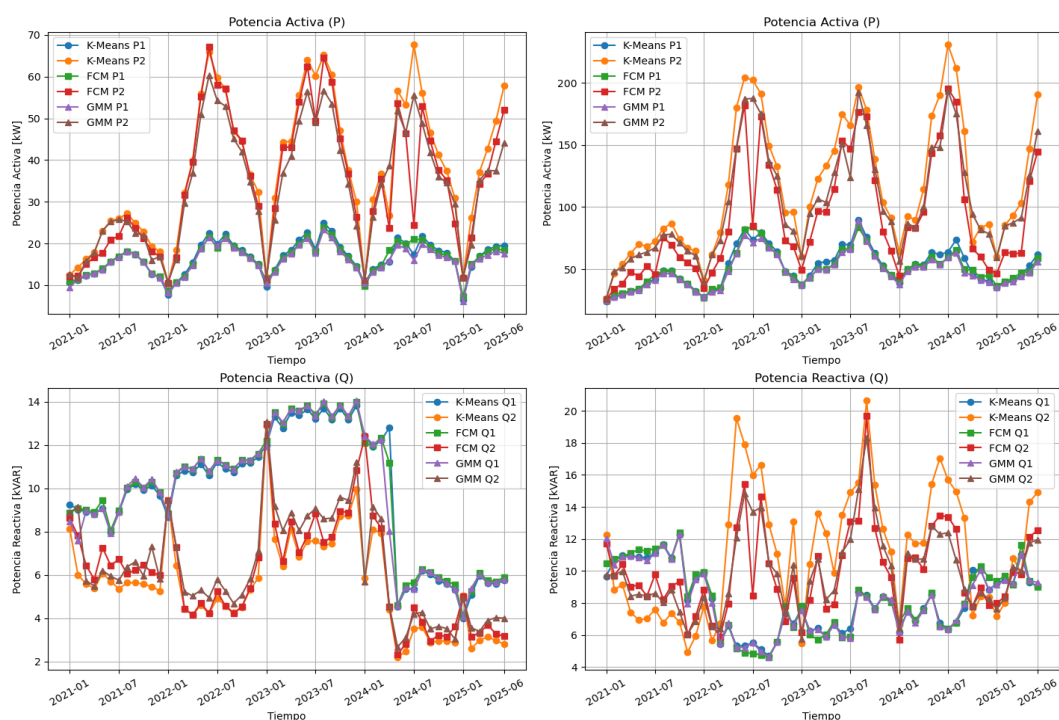


Figura 6. Localización de los centroides de potencia activa (arriba) y reactiva (abajo) durante el periodo 2021-2025, para los edificios FI (izquierda) e INTEMA (derecha).

Índices de calidad de los clusters

Como se mencionó en la sección anterior, se realizó una evaluación numérica utilizando los índices Silhouette y Davies-Bouldin. En términos generales, un valor de Silhouette más alto y uno de Davies-Bouldin más bajo reflejan una mejor calidad de agrupamiento. Si bien los umbrales específicos dependen del conjunto de datos y del algoritmo utilizado, se suele considerar que un Silhouette cercano a 1 y un Davies-Bouldin próximo a 0 son indicativos de un buen desempeño. No obstante, la literatura específica no presenta criterios unívocos para definir el umbral que

separa un agrupamiento muy bueno de uno regular o malo. En este trabajo se adoptaron como referencia los siguientes rangos para valorar la calidad del agrupamiento [29], [30]:

- Muy buena calidad: Silhouette $> 0,5$ y Davies-Bouldin $< 0,5$
- Calidad regular: $0,3 < \text{Silhouette} < 0,5$ y $0,5 < \text{Davies-Bouldin} < 1$
- Baja calidad o ausencia de estructura: Silhouette $< 0,3$ y Davies-Bouldin > 1

En la [Figura 7](#) y [Figura 8](#) se presenta la evolución en el tiempo de ambos índices, abarcando el periodo comprendido entre enero de 2021 y junio de 2025. A primera vista, ambos índices presentan rangos de variación similares en los dos edificios, oscilando el índice Silhouette entre 0,3 y 0,85 aproximadamente, y el índice Davies-Bouldin entre 0,3 y 1,7. En la [Tabla 4](#) y [Tabla 5](#) se resumen los valores medios y la desviación típica de estos índices en el periodo 2021-2025 para cada edificio. A modo de referencia, en las figuras se han resaltado los valores que indican una buena calidad del agrupamiento (fondo gris), una calidad regular (fondo verde) y una mala calidad (fondo blanco).

En el caso del edificio FI ([Figura 7](#)), cabe destacar que ambos índices experimentan un descenso significativo en el mes de enero. El índice de Silhouette cae bruscamente por debajo de 0,45, alcanzando 0,32 para el algoritmo FCM en enero de 2021. Mientras tanto, durante el mismo periodo, el índice Davies-Bouldin aumenta significativamente en comparación con los meses anteriores y posteriores. En ambos casos, esto sugiere que la calidad del *clustering* es deficiente. Un análisis más detallado revela que en enero, cuando se observaron los peores índices, el consumo de potencia reactiva fue, proporcionalmente, mayor que en otros meses ([Figura 6](#)).

Otra posible explicación de este fenómeno es que en enero el consumo total de energía disminuye significativamente en comparación con el consumo habitual durante los meses de actividad normal (de febrero a diciembre). Dado que el consumo es menor y no hay variaciones significativas según franjas horarias (como ocurre en otros meses, en los que se observa una diferencia notable entre el consumo en horas laborables y no laborables), los datos son muy similares entre sí, lo que dificulta su separación. Esto es coherente con las ubicaciones de los centroides comentadas anteriormente ([Figura 6](#)), donde los centroides aparecen más próximos entre sí, lo que dificulta la distinción entre *clusters*.

Por otro lado, se observa que en los meses de mayor consumo de potencia activa, asociados a los periodos de tiempo en los que hay actividad académica, asistencia de estudiantes y mayor número de personal universitario en el edificio, los índices muestran los mejores resultados. Esto está relacionado con lo visto en la [Figura 6](#), donde la localización de los centroides es más diferenciada en esos meses, permitiendo una mejor separación de los datos que se agrupan alrededor de cada centroide.

En general, el análisis subraya el impacto de los patrones de consumo de energía en el rendimiento del *clustering*. Los periodos de baja actividad, como enero, dan lugar a una menor diferenciación de los *clusters* debido a la falta de variación en el consumo de energía, mientras que los periodos de alta actividad potencian la separación de los *clusters*, mejorando la calidad general del *clustering*.

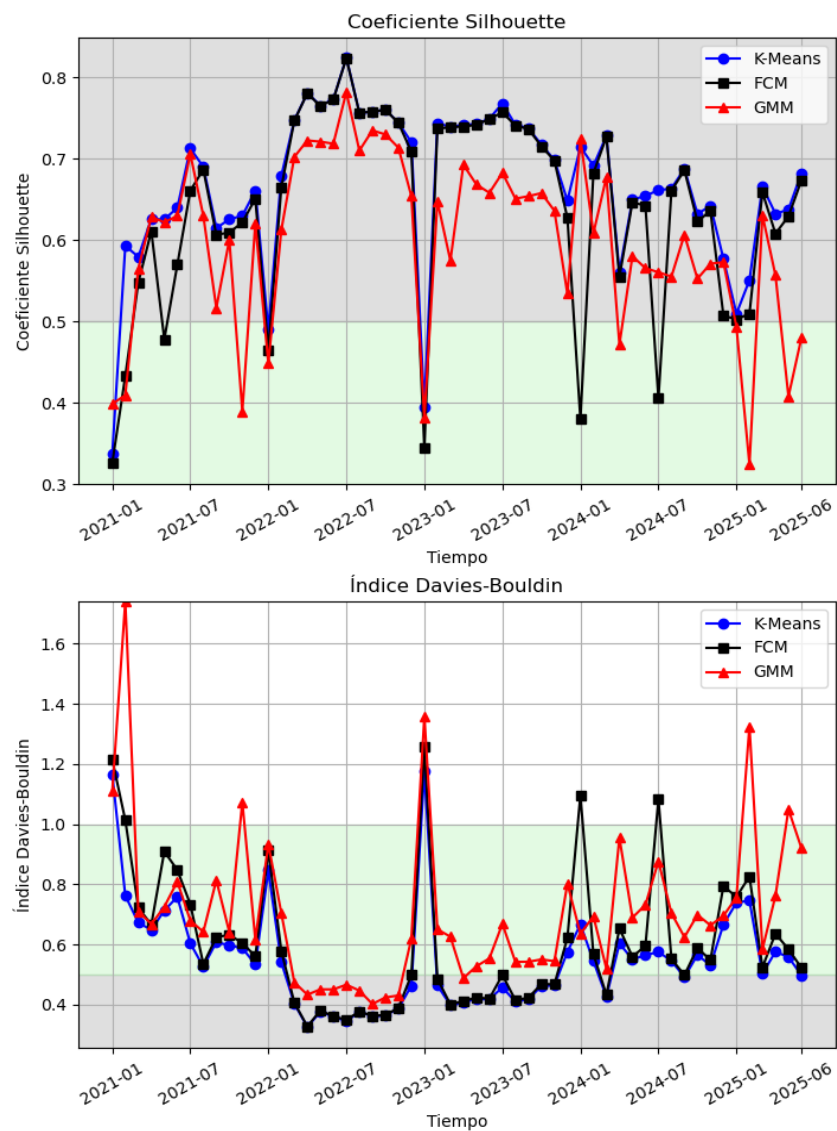


Figura 7. Métricas Silhouette y Davies-Bouldin en el periodo 2021-2025, edificio FI.

Tabla 4. Promedio y desviación estándar de los índices de calidad del edificio FI.

		Promedio	Desviación estándar
Silhouette	K-Means	0,66	0,1
	Fuzzy C-Means	0,64	0,12
	Mezclas Gaussianas	0,6	0,11
Davies-Bouldin	K-Means	0,56	0,18
	Fuzzy C-Means	0,6	0,23
	Mezclas Gaussianas	0,71	0,26

En el caso del edificio INTEMA, la evolución de los índices de calidad es menos estacional que la observada para el edificio FI (Figura 8). Mientras que Silhouette muestra un menor rendimiento en enero y un buen desempeño durante los meses de otoño e invierno, el comportamiento de los

tres métodos de *clustering* difiere significativamente en 2022 y 2023. Según el índice Silhouette, KM mantiene un rendimiento relativamente constante a lo largo de la línea temporal, con picos notables a mediados de 2022 y mediados de 2023. FCM, por su parte, muestra fluctuaciones más pronunciadas, con fuertes caídas durante el primer trimestre de cada año, lo que sugiere una mayor sensibilidad a las variaciones estacionales o a los cambios de datos. GMM muestra un comportamiento intermedio, con sus picos y valles alineados a veces tanto con KM como con FCM, lo que indica que capta algunos cambios estructurales compartidos, pero también reacciona de forma única a las variaciones de los datos.

Se observa una dinámica similar en la evolución del índice de Davies-Bouldin. En general, KM alcanza valores de Davies-Bouldin menores, lo que indica *clusters* más compactos y bien separados, especialmente a mediados de 2022 y de 2024. El algoritmo FCM presenta una mayor volatilidad, con picos en diferentes meses de 2022 y a comienzos de 2023, y un deterioro entre finales de 2024 y comienzos de 2025, lo que sugiere que las asignaciones difusas de datos conducen a configuraciones de *clusters* más dispersas durante esos periodos. Por su parte, el algoritmo GMM presenta un desempeño peor que KM, mientras que respecto de FCM muestra agrupamientos de mejor y peor calidad según el intervalo de tiempo considerado.

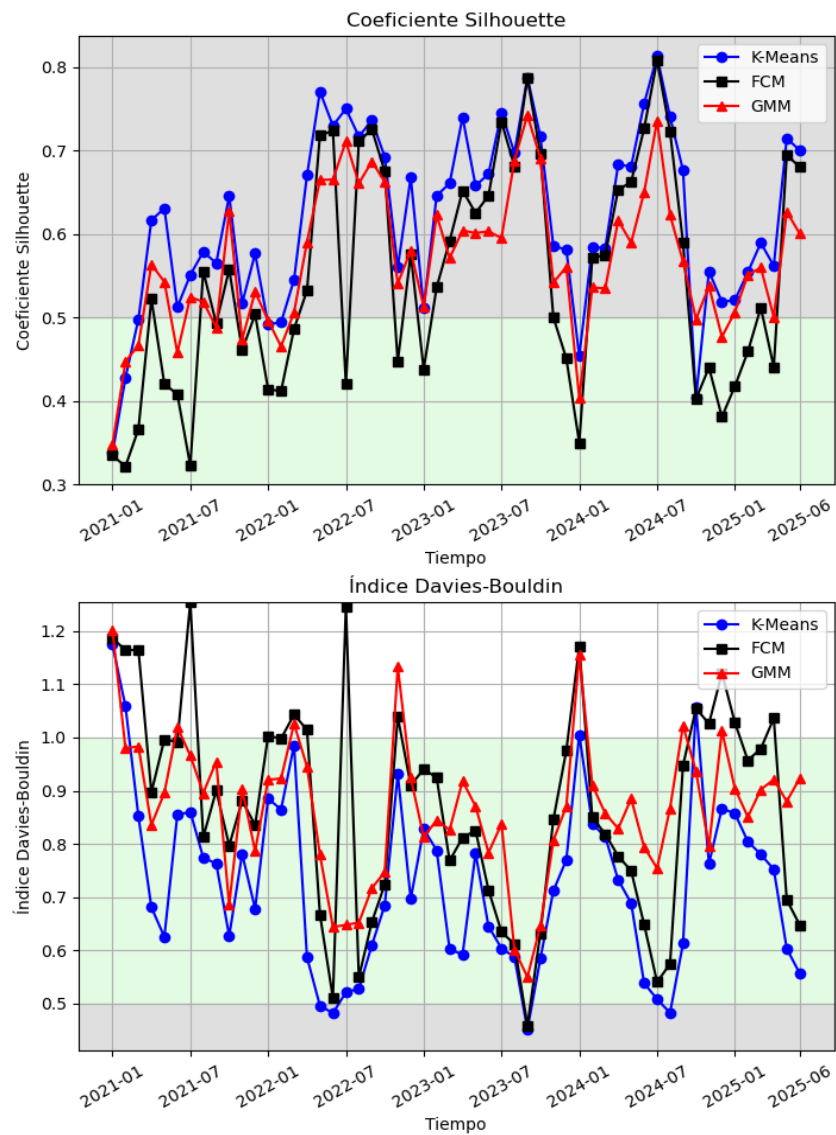


Figura 8. Métricas Silhouette y Davies-Bouldin en el periodo 2021-2025, edificio INTEMA.

Tabla 5. Promedio y desviación estándar de los índices de calidad del edificio INTEMA.

		Promedio	Desviación estándar
Silhouette	K-Means	0,63	0,1
	Fuzzy C-Means	0,57	0,12
	Mezclas Gaussianas	0,58	0,08
Davies-Bouldin	K-Means	0,69	0,16
	Fuzzy C-Means	0,82	0,2
	Mezclas Gaussianas	0,84	0,18

Eventos particulares y anomalías en los clusters

En esta sección se presentan resultados específicos de *clustering* que ponen de manifiesto las diferencias de criterio entre los algoritmos, así como su sensibilidad o inmunidad frente a los comportamientos anómalos. Desde una perspectiva temporal, se observan ciertas variaciones entre las agrupaciones obtenidas según el algoritmo empleado y el periodo analizado (mes y año). En general, las agrupaciones son similares, identificando los intervalos de tiempo en los que hay presencia/ausencia de personas en el edificio y diferenciando los días laborables de los días correspondientes a fines de semana y festivos. En la mayoría de los casos, los periodos analizados corresponden a los intervalos que presentan buenos índices de calidad. Sin embargo, se han detectado pequeñas diferencias entre los métodos, principalmente entre los *clusters* obtenidos con KM y FCM, respecto a GMM. En las figuras que se muestran en esta sección, los agrupamientos se representan mediante las etiquetas “Clúster 1” y “Clúster 2”, asignadas según el par de valores de potencia activa (P) y reactiva (Q). El Clúster 1 corresponde al grupo en torno al centroide con menor potencia activa, que a su vez corresponde con las situaciones donde la actividad es nula o reducida, mientras que el Clúster 2 se asocia al centroide con mayor potencia activa, que suele corresponder con los intervalos de tiempo de mayor actividad y presencia de alumnos y empleados de la institución (Figura 6). Este mismo criterio se aplica de manera uniforme a ambos edificios analizados.

Un primer ejemplo corresponde al edificio FI en mayo de 2022, donde todos los algoritmos coincidieron en los *clusters*, excepto el miércoles 18 de mayo (Figura 9). Ese día se realizó el Censo Nacional 2022 en Argentina, por lo que no hubo actividad universitaria. Esta situación es evidente en la curva de potencia activa, donde el consumo diurno fue el mismo que el nocturno, y es apenas perceptible en la potencia reactiva. En este caso los tres algoritmos muestran comportamientos diferentes, donde KM y FCM identifican alguna anomalía en el comportamiento de la demanda, mientras que GMM identifica la demanda de ese día como la de un día laborable. A pesar de esta diferencia, los índices de calidad de ese mes son similares, y razonablemente buenos, para los tres algoritmos. Se puede deducir que en este caso GMM agrupa los datos tomando como referencia la ligera variación de potencia reactiva.

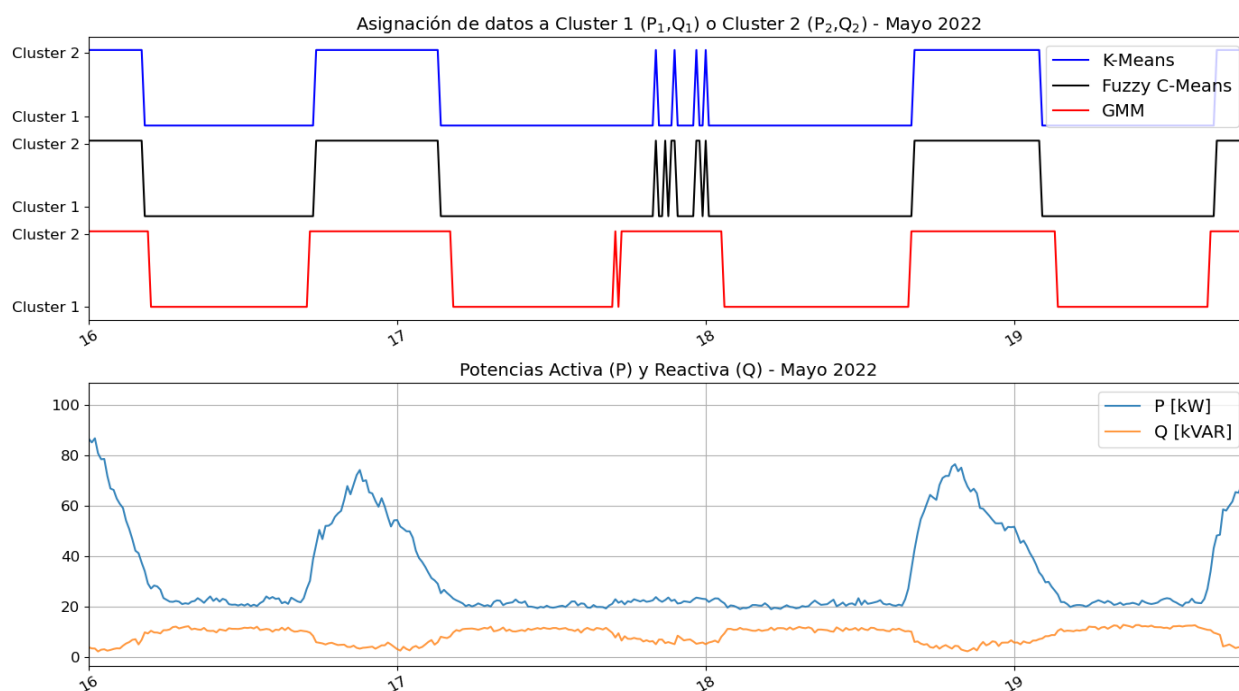


Figura 9. Superior: Asignación de los datos temporales a cada clúster. Inferior: Potencias activa y reactiva del edificio FI para la tercera semana de mayo de 2022.

En julio de 2022, se observa una anomalía en los *clusters* al procesar los datos de INTEMA (Figura 10). En Argentina, no hay actividad académica durante la segunda quincena de julio debido al receso invernal. Sin embargo, dado que INTEMA es un instituto de investigación, algunas actividades continúan incluso durante este período. Esto se refleja en el patrón de consumo de potencia activa, que muestra una caída significativa a partir del 15 de julio. Antes de esa fecha, los picos de consumo de energía alcanzaban casi 250 kW durante los días laborables, mientras que, durante el receso, los picos disminuyen a aproximadamente 100 kW. Este período de actividad reducida es interpretado de manera diferente por los algoritmos KM y GMM, ambos clasifican toda la segunda quincena del mes en el clúster de correspondiente al centroide de menor potencia activa, con la excepción de algunos datos de potencia agrupados por GMM el día 19 de julio. Por el contrario, FCM tiene dificultades para agrupar estos datos de manera consistente, asignando los datos de la segunda quincena a los *clusters* sin un patrón claro, lo que resulta en un *clustering* fragmentado y carente de coherencia. Esto puede verse en forma de un gráfico de dispersión en la Figura 11, y sugiere que FCM puede ser menos efectivo para manejar transiciones graduales o niveles de actividad mixtos, a diferencia de KM y GMM, que demuestran una mayor robustez para identificar cambios sostenidos en la actividad durante el período de receso. Se observa que la frontera que separa los *clusters* con los algoritmos KM y GMM parece estar definida principalmente por la potencia activa, mientras que en el caso de FCM el criterio de separación se basa en mayor medida en la potencia reactiva. Además, el comportamiento de FCM es coherente con la posición de los centroides correspondiente a ese mes (Figura 6), aunque es importante destacar que los índices de calidad se deterioraron, ya que el valor de Silhouette es 0,42 y Davies-Bouldin es 1,24 (Figura 8). Por el contrario, KM y GMM tienen valores mayores a 0,7 en Silhouette y menores a 0,67 en Davies-Bouldin.

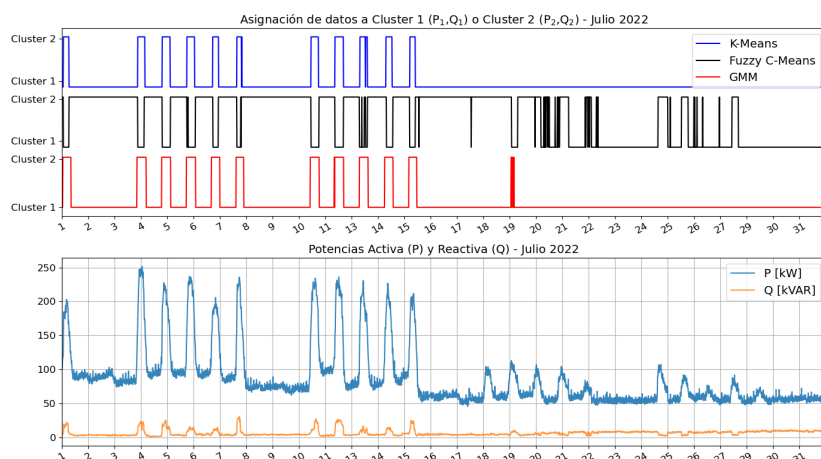


Figura 10. Superior: Asignación de los datos temporales a cada clúster. Inferior: Potencias activa y reactiva de julio de 2022 para el edificio de INTEMA.

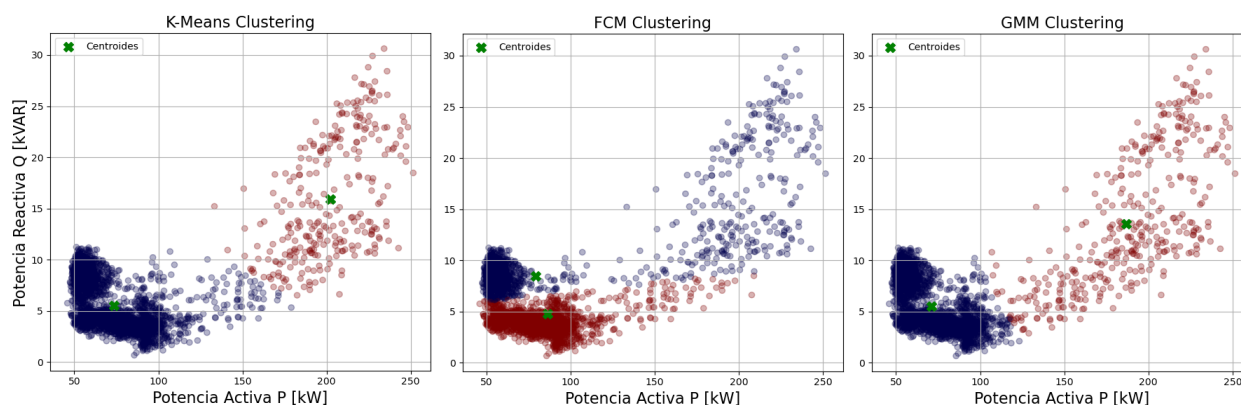


Figura 11. Gráficos de dispersión de potencia activa y reactiva correspondientes al edificio INTEMA en julio de 2022.

En febrero de 2023, se observa una clara diferencia en el comportamiento de los *clusters* al procesar los datos del edificio FI. Durante este mes, hubo varios días no laborables debido a feriados locales y nacionales (Figura 12). Todos los algoritmos de *clustering* clasificaron los datos de manera consistente en los días de actividad regular. Sin embargo, GMM agrupó parte de los datos de los días no laborables (específicamente el 4, 5, 18, 20 y 21 de febrero) en el Clúster 2, correspondiente a los consumos asociados con la mayor demanda de potencia activa. Esta clasificación podría relacionarse con la reducción de la demanda de potencia activa durante los días no hábiles, debido a la inyección de energía proveniente de los paneles solares instalados en la azotea del edificio. Durante esta época del año, la irradiación solar es particularmente alta, mientras que el consumo interno de energía se mantiene bajo. Como resultado, la demanda de potencia activa cae por debajo de los niveles mínimos habituales, lo que podría explicar la clasificación única realizada por el algoritmo GMM. Sin embargo, esto no explica porque algunos días, como es el caso del 10 y el 19 de febrero, se agruparon en el Clúster 1. Cabe resaltar que en este mes los indicadores de calidad son buenos para todos los algoritmos. No obstante, la calificación de KM y FCM es superior a la de GMM, indicando una mayor consistencia en los agrupamientos obtenidos con esos algoritmos.

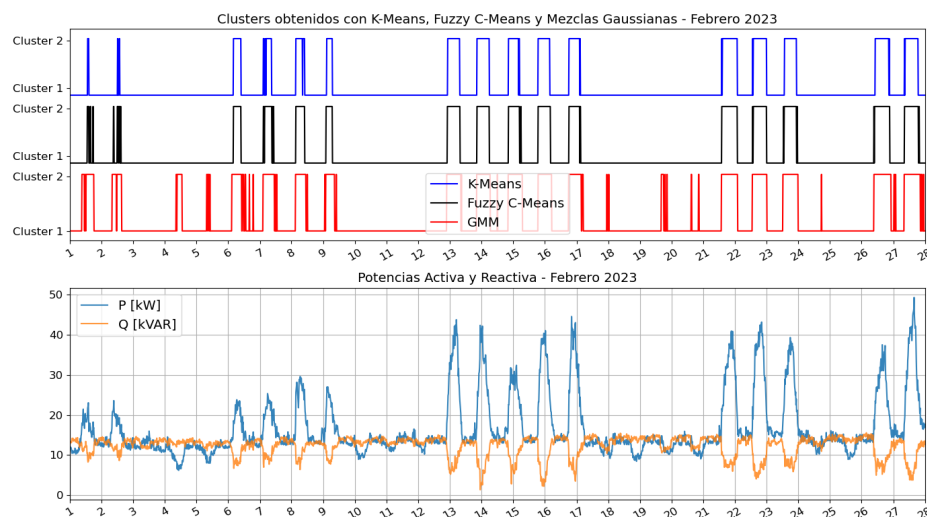


Figura 12. Superior: Asignación de los datos temporales a cada clúster. Inferior: Potencias activa y reactiva para febrero de 2023 en el edificio FI.

En la [Figura 13](#) se observa una anomalía importante en el edificio FI en abril de 2024, donde los resultados de *clustering* obtenidos mediante GMM difieren notablemente de los producidos por los otros dos métodos. Mientras KM y FCM agruparon los datos de manera similar, GMM identificó grupos distintos, destacando un patrón de segmentación alternativo. Un factor clave que influyó en el comportamiento de *clustering* de KM y FCM parece ser la potencia reactiva, que experimentó una fuerte caída a partir del 11 de abril. Esta disminución puede atribuirse a la desconexión de un banco de capacitores obsoleto que había estado compensando la potencia reactiva en la Facultad de Ingeniería. Este evento tuvo un impacto directo en el factor de potencia, lo que llevó a una reducción en la demanda de potencia reactiva. El algoritmo GMM reaccionó a este evento de manera diferente, probablemente incorporando matices adicionales en la distribución de datos que KM y FCM no pudieron captar. Esto sugiere que GMM, con su enfoque probabilístico, podría ser más sensible a variaciones sutiles tanto en la potencia activa como en la reactiva, lo que lo hace más adecuado para detectar cambios complejos en los patrones de consumo de energía. Sin embargo, el índice de Silhouette para este caso indica que los algoritmos KM y FCM tienen mejor desempeño que GMM, mientras que el índice de Davies-Bouldin muestra en todos los casos un desempeño regular, aunque peor para GMM.

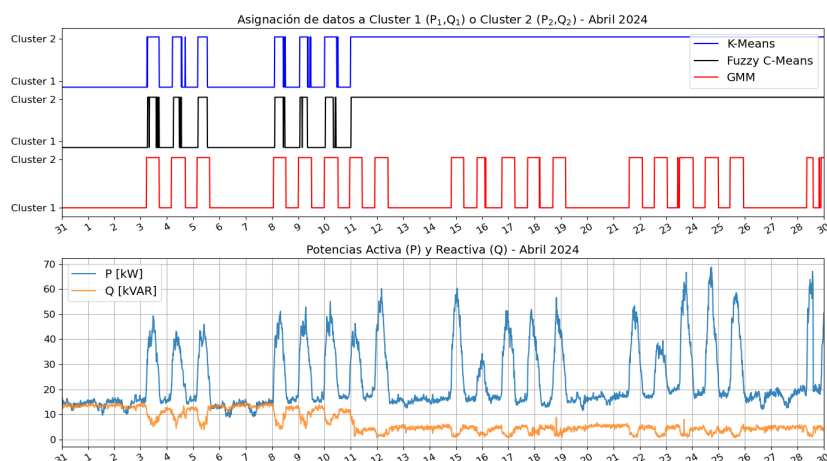


Figura 13. Superior: Asignación de los datos temporales a cada clúster. Inferior: Potencias activa y reactiva para abril de 2024 en el edificio FI.

Siguiendo con el caso de abril de 2024 en el edificio FI, se puede analizar el comportamiento de los algoritmos a partir de los gráficos de dispersión de potencia activa y reactiva. En la [Figura 14](#) se puede ver cómo se modifican los *clusters* entre los meses de marzo y abril de 2024. En la parte superior de la figura se puede ver que los tres algoritmos de *clustering* agrupan los datos de manera muy similar, con un clúster para los datos de baja potencia activa y potencia reactiva alta, y otro para los datos con potencia reactiva media o baja y potencia activa media o alta. El gráfico de dispersión de abril es muy diferente, porque el cambio de valor medio de la potencia reactiva distribuye los datos de otra manera. Aquí puede apreciarse que los algoritmos KM y FCM agrupan los datos básicamente en función de la potencia reactiva, mientras que GMM agrupa tomando como referencia la potencia activa. Como dato adicional, si se analizan los valores de las métricas de calidad ([Figura 7](#)), se ve que en los tres algoritmos se da la misma situación, en la que valores de Silhouette y Davies-Bouldin son peores en abril de 2024 que en marzo del mismo año. Esto podría ser comprensible en el caso de KM y FCM, de acuerdo a lo que se ve en la [Figura 14](#), pero es más difícil de entender en el caso de GMM, donde la calidad del agrupamiento parece mucho más intuitiva.

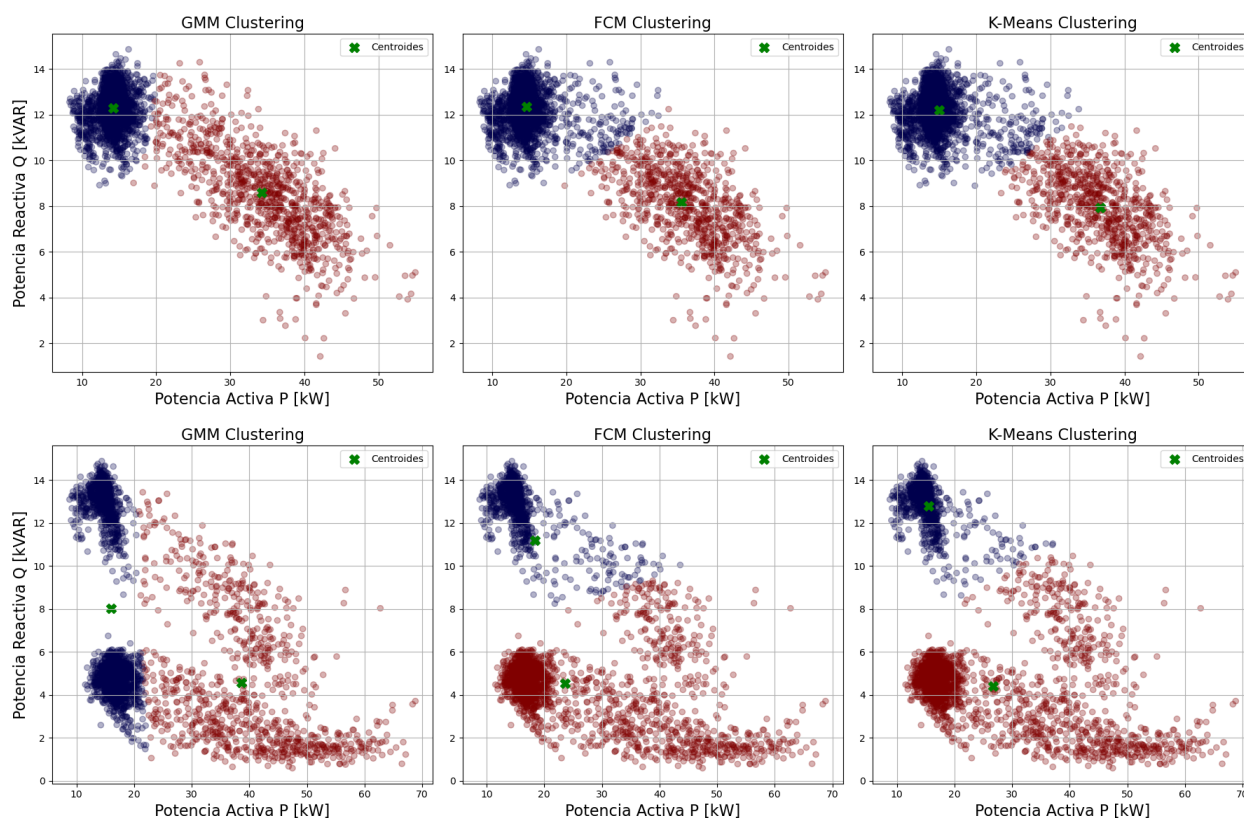


Figura 14. Gráficos de dispersión de potencia activa y reactiva del edificio FI. En la parte superior se puede ver el caso de marzo de 2024, mientras que en la parte inferior se ve el correspondiente a abril de 2024.

En mayo de 2024, se observó una situación similar a la del Censo de 2022 en los datos del edificio FI, aunque en este caso debido a otro evento no programado, que fue la huelga nacional del jueves 9 de mayo (Figura 15). En este caso se observa una marcada disminución de la potencia activa, del orden del 40%, en comparación con el promedio de los otros días jueves del mes de mayo. Esta situación hizo que los algoritmos KM y FCM agrupen parte de las mediciones de todo el día como correspondientes a un período de inactividad. Sin embargo, GMM identifica estos consumos como correspondientes a un día de actividad. Aquí se puede destacar que los tres algoritmos detectan actividad en el día de la huelga, pero lo hacen en diferentes bandas horarias, evidenciando ciertas diferencias de sensibilidad. Cabe aclarar que los tres algoritmos presentan indicadores de calidad aproximadamente iguales, con índices de Silhouette buenos e índices de Davies-Bouldin moderadamente buenos.

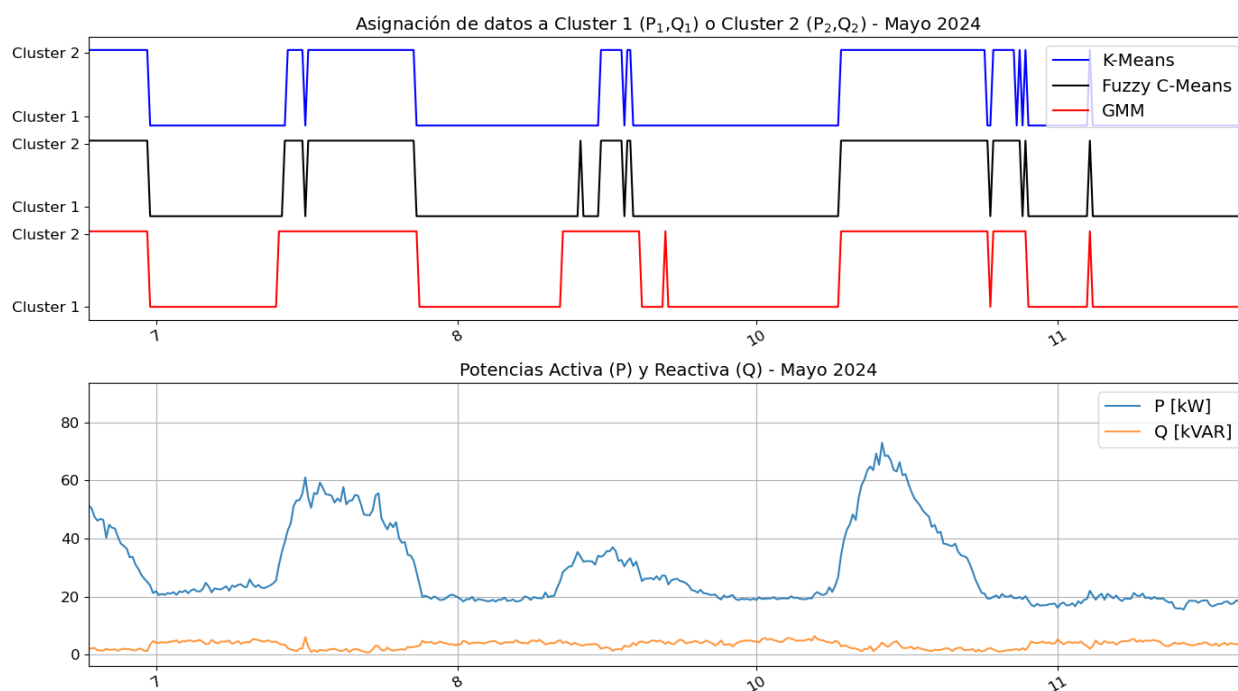


Figura 15. Superior: Asignación de los datos temporales a cada clúster. Inferior: Potencias activa y reactiva para la segunda semana de mayo de 2024 en el edificio FI.

Finalmente, en octubre de 2024, se observó otro patrón de *clustering* inusual en la base de datos de INTEMA. Como se muestra en la [Figura 16](#), los algoritmos KM y GMM agruparon los datos de manera similar, basándose principalmente en los niveles de potencia activa. Un clúster incluye datos correspondientes a las horas de actividad, mientras que el otro abarca datos de las horas de inactividad. Sin embargo, el algoritmo FCM exhibe un comportamiento errático a partir del 13 de octubre, desviándose del patrón esperado. No se identificó un patrón claramente interpretable o una estructura discernible respecto de cómo FCM organiza los datos durante este período, lo que sugiere una posible sensibilidad a fluctuaciones o ruido en el conjunto de datos. Esta divergencia resalta las diferencias en la robustez y adaptabilidad de los algoritmos de *clustering* cuando se aplican a datos reales de consumo de energía. Además, es importante señalar que los índices de calidad empeoraron durante este mes ([Figura 8](#)), lo que es consistente con el comportamiento errático de FCM pero no con el desempeño de los otros dos algoritmos.

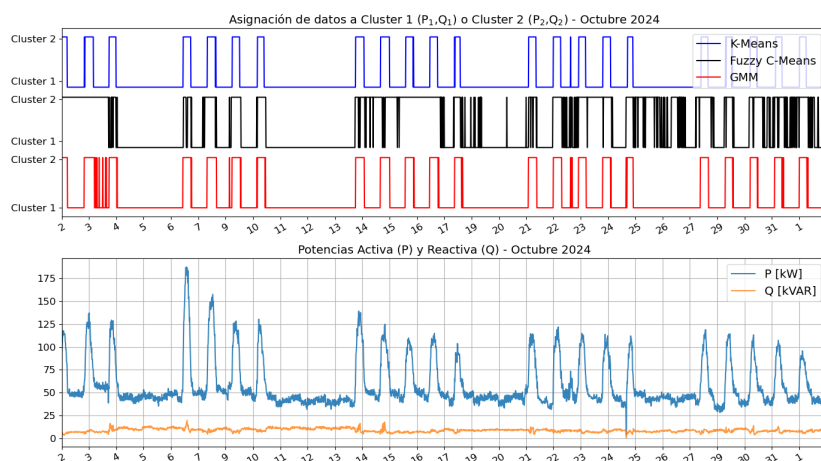


Figura 16. Superior: Asignación de los datos temporales a cada clúster. Inferior: Potencias activa y reactiva para octubre de 2024 en el edificio INTEMA.

DISCUSIÓN

Los resultados obtenidos evidencian que los algoritmos de *clustering* aplicados a series temporales de demanda eléctrica permiten identificar patrones de consumo y eventos anómalos en instalaciones universitarias. El desempeño de cada método estuvo condicionado tanto por la estacionalidad de los datos como por la naturaleza de los eventos analizados. Por este motivo, el proceso de *clustering* se realizó de manera independiente para cada mes del período 2021-2025, mitigando el problema de la estacionalidad, ya que los perfiles de consumo dentro de un mismo mes presentan una mayor homogeneidad temporal y estacional.

En el caso del algoritmo K-Means se observó un desempeño más consistente a lo largo del período de estudio, alcanzando los valores promedio más altos del índice de Silhouette ($\rightarrow 1$) y los más bajos del índice de Davies-Bouldin ($\rightarrow 0$). Estos resultados indican una mejor separación y mayor cohesión interna de los *clusters* obtenidos. Su capacidad para separar estados de actividad e inactividad sugiere que este enfoque resulta apropiado cuando los perfiles de consumo presentan fronteras bien definidas. Por su parte, Fuzzy C-Means mostró una mayor dispersión en los resultados, posiblemente debido a la función de pertenencia difusa. La asignación binaria aplicada al definir los *clusters* podría no representar adecuadamente algunos estados intermedios de demanda. Por último, el algoritmo de Mezclas Gaussianas demostró un desempeño más flexible frente a cambios combinados de potencia activa y reactiva, permitiendo detectar algunas variaciones sutiles o comportamientos inusuales.

Las discrepancias observadas entre los métodos durante algunos meses y eventos específicos (censo nacional, recesos académicos y modificaciones al circuito interno del edificio) destacan el potencial del *clustering* como herramienta para la detección temprana de anomalías y la caracterización de patrones energéticos en edificios con ocupación variable. No se observó que la mala calidad del suministro eléctrico en el edificio de INTEMA [27] haya afectado a la calidad de los agrupamientos. Los índices de calidad mostraron valores compatibles con agrupamientos de calidad buena o intermedia durante la mayor parte del período analizado, con excepción de

algunos meses de demanda baja, y no se observaron más anomalías que las presentadas en la Figura 10 y la Figura 16. En cuanto al edificio FI, se observó que la inyección de potencia de los paneles solares puede afectar a algunos de los algoritmos de agrupamiento, particularmente en aquellos casos donde el consumo es bajo (por ejemplo, los fines de semana, como se vio en la Figura 12).

Los resultados pusieron de manifiesto que la calidad del agrupamiento está fuertemente condicionada por la estacionalidad y por la magnitud de las fluctuaciones en el consumo eléctrico. La calidad del *clustering* se evaluó principalmente mediante los índices Silhouette y Davies-Bouldin, elegidos por su solidez en la mayoría de los escenarios. En términos generales, valores de Silhouette superiores a 0,5 y valores de Davies-Bouldin inferiores a 0,5 indicaron agrupamientos compactos y bien separados, mientras que valores de Silhouette inferiores a 0,4 o Davies-Bouldin superiores a 1 reflejaron una menor diferenciación entre *clusters*. En los meses de baja actividad o con demanda uniforme, como por ejemplo enero, la diferenciación entre *clusters* tendió a disminuir, lo que se refleja en el índice de Silhouette, que cae por debajo de 0,5, y en el índice de Davies-Bouldin, que en algunos casos supera el valor 1. Esta situación confirma que la uniformidad en la demanda eléctrica durante los períodos de receso dificulta la separación de *clusters*, reduciendo la confiabilidad de la estructura de *clustering* identificada en dichos períodos.

El estudio realizado tiene algunas limitaciones que se han mencionado a lo largo del texto. Sin embargo, los resultados alientan a profundizar en algunos aspectos que permitan sortear dichas limitaciones. A modo de resumen se pueden mencionar las siguientes:

- El estudio se concentra en dos edificios con comportamientos diferenciados, lo que podría limitar la extrapolación de resultados. Sería conveniente incluir nuevas bases de datos para ampliar el alcance del estudio.
- La escasa cantidad de eventos anómalos (censo, huelga y períodos vacacionales) limita la robustez del análisis.
- La sensibilidad de los algoritmos a la inicialización puede generar diferencias en algunos *clusters* entre ensayos sucesivos. En trabajos futuros se debería evaluar la sensibilidad de los *clusters* mediante repeticiones iterativas del proceso de *clustering*.

De acuerdo a los resultados presentados por Yang et al. [14], otra posible extensión de este trabajo sería la integración de algunos de los algoritmos de *clustering* analizados con un sistema de detección automática de anomalías. Este enfoque permitiría no solo identificar patrones y eventos de consumo atípicos, sino también ayudaría a caracterizar mejor las posibles causas en entornos reales.

Por último, es importante subrayar que el proceso de *clustering* es sensible a la estandarización de los datos, por lo que es esencial definir cuidadosamente los criterios de aplicación. Por ejemplo, se observó que la estandarización de los datos realizada en una base diaria no arrojó buenos resultados, mientras que la estandarización realizada sobre una base mensual demostró mejor calidad del agrupamiento de datos. Esta cuestión sugiere la necesidad de realizar un análisis más amplio y una comparación a gran escala utilizando diferentes ventanas temporales para la estandarización de los datos, con el fin de comprender mejor su impacto en la forma y distribución de los *clusters*.

CONCLUSIONES

El uso de algoritmos de aprendizaje automático se ha extendido a muchos ámbitos, incluido el sector de la energía y la distribución eléctrica. Los algoritmos de *clustering* no supervisado tienen un gran potencial para el tratamiento de datos en el sector eléctrico, especialmente cuando se trata de identificar patrones de consumo y detectar comportamientos anómalos. En este estudio se han presentado los resultados obtenidos mediante el uso de algoritmos de *clustering* no supervisados aplicados a series temporales de demanda de potencia activa y reactiva de dos edificios universitarios durante el periodo 2021-2025, caracterizadas por variaciones asociadas a la población, estación del año y/o tipo de actividades.

Se evaluaron tres algoritmos de *clustering* diferentes (K-Means, Fuzzy C-Means y mezclas Gaussianas) utilizando diferentes experimentos. Se evaluaron no solo los índices de calidad (Silhouette y Davies-Bouldin) y la posición de los centroides, sino también los agrupamientos en función de los días y horarios, para tratar de vincular los resultados con sus posibles causas. El análisis reveló que, si bien los tres algoritmos evaluados presentan en general un rendimiento similar, surgen diferencias al considerar eventos de consumo anómalos, como cambios en la demanda de potencia activa o reactiva (por modificaciones en la topología de la red o por alteraciones en la rutina laboral producto de recesos), o por las variaciones estacionales propias de los datos.

Los resultados obtenidos aportan evidencia sobre la utilidad del aprendizaje automático como herramienta de apoyo en la gestión energética de instalaciones. Las limitaciones de este estudio, que fueron mencionadas a lo largo del texto, impulsan a la extensión de la investigación mediante la incorporación de bases de datos correspondientes a otros edificios, y la incorporación de variables climáticas y/o de ocupación, que permitan obtener mejores agrupamientos.

Trabajos futuros deberían explorar métricas de evaluación alternativas para ofrecer una perspectiva más amplia, incluyendo métodos de *clustering* basados en la densidad, como OPTICS y DBSCAN, y/o empleando estrategias de estandarización/normalización y segmentación temporal adaptativa. Este estudio sirve de referencia para futuras investigaciones, en estas u otras instituciones, especialmente para analizar series temporales con perfiles muy diversos. Incluso se podría integrar con otras líneas de investigación de la institución, como por ejemplo el pronóstico de demanda de energía eléctrica.

AGRADECIMIENTOS

Los autores agradecen al Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo (CYTED) por el apoyo brindado a través de las Redes Temáticas RIPCEL (726RT0203) y RIIIAinSG (726RT0204).

FINANCIAMIENTO

Este trabajo fue financiado por la Universidad Nacional de Mar del Plata (UNMDP), a través del proyecto N° 80020240500011MP, y el Consejo Nacional de Investigaciones Científicas y Tecnológicas.

gicas (CONICET), a través del proyecto PIP0007.

REFERENCIAS

- [1] J. Powell, A. McCafferty-Leroux, W. Hilal, and S. Andrew Gadsden, "Smart grids: A comprehensive survey of challenges, industry applications, and future trends," *Energy Reports*, vol. 11, 2024, pp. 5760-5785, <https://doi.org/10.1016/j.egyr.2024.05.051> [↑]
- [2] M. Mahmood, P. Chowdhury, R. Yeassin, M. Hasan, T. Ahmad, and N.U.R. Chowdhury, "Impacts of digitalization on smart grids, renewable energy, and demand response: An updated review of current applications," *Energy Conversion and Management: X*, vol. 24, pp. 1-31, 2024, <https://doi.org/10.1016/j.ecmx.2024.100790> [↑]
- [3] N. Ahmad, Y. G. Ghadi, M. Adnan, and M. Ali, "From smart grids to super smart grids: A roadmap for strategic demand management for next generation SAARC and European power infrastructure," *IEEE Access*, vol. 11, pp. 12303-12341, 2023, <https://doi.org/10.1109/ACCESS.2023.3241686> [↑]
- [4] F. de Caro, A. Andreotti, R. Araneo, M. Panella, A. Rosato, A. Vaccaro, and D. Villacci, "A review of the enabling methodologies for knowledge discovery from smart grids data," *Energies*, vol. 13, no. 24, Art. no. 6579, pp. 1-25, 2020, <https://doi.org/10.3390/en13246579> [↑]
- [5] N. Sapountzoglou, J. Lago, B. De Schutter, and B. Raison, "A generalizable and sensor-independent deep learning method for fault detection and location in low-voltage distribution grids," *Applied Energy*, vol. 276, Art. no. 115299, pp. 1-22, 2020, <https://doi.org/10.1016/j.apenergy.2020.115299> [↑]
- [6] Y. Bai, H. Sun, L. Zhang, and H. Wu, "Hybrid CNN-transformer network for electricity theft detection in smart grids," *Sensors*, vol. 23, no. 20, Art. no. 8405, pp. 1-21, 2023, <https://doi.org/10.3390/s23208405> [↑]
- [7] W. Wu and M. Peng, "A data mining approach combining K-means clustering with bagging neural network for short-term wind power forecasting," *IEEE Internet of Things Journal*, vol. 4, no. 4, pp. 979-986, 2017, <https://doi.org/10.1109/JIOT.2017.2677578> [↑]
- [8] I.H. Witten, E. Frank, and M.A. Hall, *Data Mining: Practical machine learning tools and techniques*, 3rd ed. New York, USA: Morgan Kaufmann, 2011, <https://doi.org/10.1016/C2009-0-19715-5> [↑]
- [9] L. Sun, Y. Chen, Q. Du, R. Ding, Z. Liu, and Q. Cheng, "Topology identification of low-voltage power lines based on IEC 61850 and the clustering method," *Energies*, vol. 16, no. 3, Art. no. 1126, pp. 1-20, 2023, <https://doi.org/10.3390/en16031126> [↑]
- [10] L. Sun, Y. Chen, Q. Du, H. Xu, and W. Wang, "Identification of low-voltage phase lines using IEC 61850 and K-means clustering," *Electric Power Systems Research*, vol. 234, Art. no. 110597, 2024, <https://doi.org/10.1016/j.epsr.2024.110597> [↑]
- [11] P.D. Patil, R. Patil, P. Ahire, R. Bharati, and Y. Dongre, "An adaptive methodology based on predictive deep learning and context aware clustering for electricity power usage mining and optimization at different granularity levels," *e-Prime, Advances in Electrical Engineering, Electronics and Energy*, vol. 8, Art. no. 100628, 2024, <https://doi.org/10.1016/j.prime.2024.100628> [↑]

- [12] E. Priya and J. Preetha Roselyn, "A multi-layer constrained spectral k-embedded clustering methodology approach for intelligent partitioning of power grid to enhance resiliency in transmission networks," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 10, Art. no. 100813, pp. 1-16, 2024, <https://doi.org/10.1016/j.prime.2024.100813> [↑]
- [13] M. Charwand, M. Gitizadeh, P. Siano, G. Chicco, and Z. Moshavash, "Clustering of electrical load patterns and time periods using uncertainty-based multi-level amplitude thresholding," *International Journal of Electrical Power & Energy Systems*, vol. 117, 105624, Art. no. 2020, <https://doi.org/10.1016/j.ijepes.2019.105624> [↑]
- [14] J. Yang, H. Que, W. Liu, and J. Xiao, "A monitoring model for abnormal electricity consumption based on K-Means++ clustering and improved K-Nearest neighbor algorithm," *Smart Grids and Energy*, vol. 9, Art. no. 29, pp. 1-16, 2024, <https://doi.org/10.1007/s40866-024-00213-5> [↑]
- [15] G.E. Okereke, M.C. Bali, C.N. Okwueze, E.C. Ukekwe, S.C. Echezona, and C.I. Ugwu, "*K-means clustering of electricity consumers using time-domain features from smart meter data*," *Journal of Electrical Systems and Information Technology*, vol. 10, Art. no. 2, 2023, <https://doi.org/10.1186/s43067-023-00068-3> [↑]
- [16] C. Zhang and J. Mo, "A clustering analysis of electricity consumption behavior," *5th International Conference on Energy, Electrical and Power Engineering (CEEPE)*, Chongqing, China, 2022, pp. 962-969, <https://doi.org/10.1109/CEEPE55110.2022.9783311> [↑]
- [17] P.G. Donato, M.N. Hadad, I. Carugati, M.A. Funes, J.L. Strack, and C.M. Orallo, "Smart grids in developing countries: current status, challenges and upcoming trends," *Advances in Energy Research*, vol. 33, M. J. Acosta, Ed., New York, USA: Nova Science Publishers, 2020, pp. 85-122. [↑]
- [18] L. Marrero, D. Carrizo, L. García-Santander y F. Ulloa-Vásquez, "Using K-means algorithm to classify customer profiles with data from smart energy consumption meters: A case study," *Ingeniare*, vol. 29, no. 4, pp. 778-787, 2021, <https://doi.org/10.4067/S0718-33052021000400778> [↑]
- [19] A.E. Ezugwu, A.M. Ikotun, O.O. Oyelade, L. Abualigah, J.O. Agushaka, C.I. Eke, and A.A. Akinyelu, "A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects," *Engineering Applications of Artificial Intelligence*, vol. 110, Art. no. 104743, pp. 1-43, 2022, <https://doi.org/10.1016/j.engappai.2022.104743> [↑]
- [20] Scikit-learn, "Clustering," scikit-learn.org. Accessed: August 16, 2024. [Online]. Available: scikit-learn.org/stable/modules/clustering.html [↑]
- [21] A.K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651-666, 2010, <https://doi.org/10.1016/j.patrec.2009.09.011> [↑]
- [22] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, New York, USA: Springer, 1981, <https://doi.org/10.1007/978-1-4757-0450-1> [↑]
- [23] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, 1977, pp. 1-22, <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x> [↑]

- [24] O. Arbelaiz, I. Gurrutxaga, J. Muguerza, J.M. Pérez, and I. Perona, “An extensive comparative study of cluster validity indices,” *Pattern Recognition*, vol. 46, Issue 1, pp. 243-256, 2013, <https://doi.org/10.1016/j.patcog.2012.07.021> [↑]
- [25] Scikit-learn “Silhouette_Score,” [scikit-learn.org](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html). Accessed: February 28, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.silhouette_score.html [↑]
- [26] Scikit-learn, “Davies_bouldin_score,” [scikit-learn.org](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.davies_bouldin_score.html). Accessed: February 28, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.davies_bouldin_score.html [↑]
- [27] I. Carugati, M. A. Funes, P. G. Donato, J. L. Strack, C. M. Orallo, and M. Hadad, “Calidad de energía en ciudades inteligentes: primeros resultados en la ciudad de Mar del Plata,” *2020 IEEE Congreso Bienal de Argentina (ARGENCON)*, Resistencia, Argentina, diciembre 2020, pp. 1-6, <https://doi.org/10.1109/ARGENCON49523.2020.9505326> [↑]
- [28] Scikit-learn “StandardScaler,” [scikit-learn.org](https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html#standardscaler). Accessed: October 15, 2025. [Online]. Available: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html#standardscaler> [↑]
- [29] L. Kaufman and P.J. Rousseeuw, *Finding Groups In Data. An Introduction To Cluster Analysis*, New Jersey, USA: John Wiley & Sons, 1990, <https://doi.org/10.1002/9780470316801> [↑]
- [30] A.A. Mekonnen, T. Sipos, and N. Krizsik, “Identifying Hazardous crash locations using empirical bayes and spatial autocorrelation,” *International Journal of Geo-Information*, vol. 12, no. 3, Art. no. 85, 2023, pp. 1-17, <https://doi.org/10.3390/ijgi12030085> [↑]