

Advanced Monitoring of Photovoltaic Plants without Irradiance Sensors Using Unsupervised Learning^{*}

Diego Rivelino Espinoza Trejo^{*} Luis Ángel López Vélez^{**}
Vladimir Ilich Ruiz Salinas^{***} Isaac Compeán Martínez^{*}
Daniel Ulises Campos Delgado^{****}
Cristian H. De Angelo^{*****}

^{*} Unidad Académica Multidisciplinaria Región Altiplano, Universidad Autónoma de San Luis Potosí, Carr. Cedral km 5+600, Matehuala, 78700, S.L.P., México (e-mail: drespinozat@ieee.org, isaac.compean@uaslp.mx)

^{**} Benebión de PHYTOSAN S.A. de C.V., Carretera Federal 57 km 7+250, Matehuala, 78760, S.L.P., México (e-mail: luis.lopez@benebion.com)

^{***} Fronius Mexico S.A. DE C.V., Carretera Monterrey Saltillo 3279, Privadas de Santa Catarina, CP.66367, Sta. Catarina, N.L., México (e-mail: ruiz.vladimir@fronius.com)

^{****} Facultad de Ciencias, Universidad Autónoma de San Luis Potosí, Av. Chapultepec 1570, San Luis Potosí, 78295, S.L.P., México (e-mail: ducd@fciencias.uaslp.mx)

^{*****} Universidad Nacional de Río Cuarto, Ruta Nacional 36, km. 601, Río Cuarto, 78295, Córdoba, Argentina (e-mail: cdeangelo@ing.unrc.edu.ar)

Resumen.- Small- and medium-scale photovoltaic (PV) plants commonly rely on monitoring systems limited to electrical parameters, such as generated power. Incorporating sensors to measure plane-of-array irradiance and module operating temperature is often not cost-effective in these installations, which restricts the use of traditional strategies for anomaly detection and performance evaluation. This paper proposes an advanced monitoring scheme based on unsupervised learning, specifically using hierarchical clustering, to identify atypical behaviors and assess the relative performance of neighboring PV inverters. The methodology is grounded in fault diagnosis principles, such as physical redundancy and parity relations, leveraging the expected similarity among power generation profiles. The proposed approach is validated using real data from a 240 kW PV plant consisting of 16 inverters rated at 15 kW, 304 modules of 450 Wp, and 238 modules of 545 Wp.

Keywords: Anomaly detection, Performance evaluation, Photovoltaic plants, Unsupervised learning, Fault diagnosis.

1. INTRODUCCIÓN

El desarrollo de estrategias automáticas para la evaluación del desempeño y la detección de anomalías en plantas fotovoltaicas (FV) de pequeña y mediana escala

^{*} Los autores expresan su agradecimiento a la empresa Benebión de PHYTOSAN S.A. de C.V. y a Fronius México S.A. de C.V. por facilitar el acceso a sus plataformas de monitoreo, lo cual fue fundamental para el desarrollo de este trabajo. Asimismo, se agradece a la Asociación Universitaria Iberoamericana de Postgrado (AUIP) por la concesión de una beca que permitió la realización de una estancia académica en la Universidad Nacional de Río Cuarto (Argentina), llevada a cabo del 7 de diciembre de 2023 al 2 de enero de 2024.

ha cobrado relevancia en los últimos años (Taghezouit et al. 2024), (IEA PVPS 2022). Esta línea de trabajo se ha vuelto particularmente relevante, debido a que las métricas de evaluación propuestas en la última década requieren sensores de irradiancia y temperatura, los cuales rara vez se encuentran en instalaciones de pequeña y mediana escala debido a su costo relativamente elevado en comparación con la propia tecnología FV. Además, es importante destacar que las mediciones requeridas por dichas métricas corresponden a la irradiancia en el plano del arreglo FV y a la temperatura de operación de los módulos. Estas variables no pueden obtenerse directamente mediante estaciones meteorológicas convencionales, que usualmente miden irradiancia global y difusa en plano

horizontal, y temperatura ambiente. Obtener los valores adecuados implica ya sea adaptar los piranómetros a la inclinación del arreglo FV, o estimar la irradiancia en el plano del arreglo a partir de modelos de transposición (Widén and Munkhammar 2019). Asimismo, la medición de la temperatura de operación de los módulos requiere sensores específicos en la superficie posterior del módulo, lo cual representa un gasto adicional que difícilmente se justifica en proyectos de pequeña y mediana escala. En este contexto, resulta crucial desarrollar estrategias que no dependan de sensores adicionales para detectar condiciones anómalas en la operación de las plantas, haciendo uso de los registros de generación de potencia disponibles mediante los sistemas de monitoreo actuales.

Las fallas más relevantes en plantas FV incluyen sombreados parciales, acumulación de suciedad, degradación del desempeño, puntos calientes, recortes de potencia y fallas en los strings (IEA PVPS 2021). En este trabajo nos enfocamos únicamente en la detección de anomalías, sin abordar aún la clasificación detallada del tipo de falla ni su cuantificación, lo cual se vislumbra como parte de trabajos futuros que probablemente requieran el uso de sensores adicionales. El inversor, pieza central de la planta FV, concentra funciones críticas como la conversión CD-CA, el MPPT, la protección contra fallas y la comunicación con la red (Doyle et al. 2019), lo que lo convierte también en uno de los componentes más vulnerables (Gunda and Homan 2020). Sin embargo, su naturaleza de caja negra, derivada de la diversidad de topologías y el carácter cerrado de su arquitectura, dificulta el desarrollo de enfoques de diagnóstico basados en modelos determinísticos.

Dado que las plantas FV carecen de componentes rotativos, las técnicas clásicas de diagnóstico basadas en análisis de señales han tenido un papel limitado, restringiéndose en gran medida a la detección de fallas por arcos eléctricos (Zhu et al. 2016). En cambio, los enfoques que han cobrado mayor relevancia en la literatura son los basados en métodos estadísticos y, más recientemente, en aprendizaje automático (IEA PVPS 2021), apoyados en la creciente disponibilidad de datos registrados por los sistemas de monitoreo de los inversores.

De esta manera, nuestro estudio se fundamenta en el uso de los datos disponibles a partir de los sistemas de monitoreo actuales para detectar anomalías y evaluar el desempeño de las plantas FV. La idea central de nuestra propuesta se fundamenta en principios de la teoría del diagnóstico de fallas, a partir de los cuales se plantean alternativas viables para la evaluación del desempeño y la detección automática de anomalías, sin necesidad de sensores adicionales. Un punto clave en esta dirección es que la mayoría de las plantas FV cuentan con redundancia física en los inversores, los cuales se conectan en paralelo para incrementar la capacidad total del sistema. Así, proponemos evaluar la similitud entre los perfiles de generación de inversores vecinos que forman parte de dicha redundancia, utilizando una métrica de disimilitud como base para su agrupamiento. Si todos los inversores operan de manera similar, los perfiles serán consistentes entre

sí y formarán un único grupo. En cambio, si las curvas de generación presentan diferencias significativas en su forma o comportamiento a lo largo del tiempo, el agrupamiento jerárquico las agrupará en distintos cúmulos. Esto permite identificar inversores cuyo perfil de generación se desvía del resto, ya sea por menor producción, mayor variabilidad, o una forma de curva atípica. Consideramos que esta propuesta representa una aportación valiosa, ya que permite detectar bajo desempeño relativo entre los inversores que forman el sistema FV completo, sin requerir sensores de irradiancia y temperatura que normalmente serían necesarios para establecer una referencia externa de generación solar.

El artículo se organiza de la siguiente manera: la Sección 2 describe la metodología basada en un algoritmo de agrupamiento jerárquico; la Sección 3 presenta el caso de estudio y los resultados obtenidos; y la Sección 4 expone las conclusiones y posibles líneas de investigación futura.

2. METODOLOGÍA DEL CLUSTERING JERÁRQUICO

El **clustering jerárquico aglomerativo** (HAC, por sus siglas en inglés) es un método no supervisado de agrupamiento que construye una jerarquía de cúmulos a partir de una matriz de disimilitudes (Hastie et al. 2009). Es ampliamente utilizado para explorar similitudes entre variables o series temporales, y visualizarlas mediante un dendrograma.

2.1 Datos y Representación

Supongamos que tenemos p variables (en nuestro caso, potencias generadas por 16 inversores fotovoltaicos), cada una representada por un vector:

$$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p \in \mathbb{R}^n$$

donde $\mathbf{x}_j = (x_{j1}, x_{j2}, \dots, x_{jn})^\top$ es la serie temporal de observaciones correspondientes a la variable j , con $j \in \{1, 2, \dots, p\}$. Cada \mathbf{x}_j representa una secuencia de n observaciones (por ejemplo, mediciones de potencia registradas cada 5 minutos durante un día) asociadas a la j -ésima variable del conjunto de datos.

2.2 Cálculo de la Distancia entre Vectores

Para comparar la similitud entre las series de potencia de distintos inversores, se emplea una métrica de distancia que combina dos componentes: la **forma** del perfil (capturada por la correlación de Pearson) y la **magnitud** (medida mediante la distancia euclidiana).

De esta manera, la distancia entre dos vectores (dos series temporales) \mathbf{x}_i y \mathbf{x}_j se define como:

$$d_{i,j} \equiv d(\mathbf{x}_i, \mathbf{x}_j) = \lambda \cdot d_c(\mathbf{x}_i, \mathbf{x}_j) + (1 - \lambda) \cdot d_m(\mathbf{x}_i, \mathbf{x}_j), \quad (1)$$

donde:

- $d_c(\mathbf{x}_i, \mathbf{x}_j) = 1 - \rho(\mathbf{x}_i, \mathbf{x}_j)$ es la disimilitud basada en la correlación de Pearson.
- $d_m(\mathbf{x}_i, \mathbf{x}_j) = \frac{\|\mathbf{x}_i - \mathbf{x}_j\|_2}{\max_{l, n \in \{1, \dots, p\}} \|\mathbf{x}_l - \mathbf{x}_n\|_2}$, donde el denominador corresponde a la mayor distancia euclidiana

encontrada entre cualquier par de series del conjunto. Esta métrica representa la distancia euclidiana normalizada entre series.

- $\lambda \in [0, 1]$ es un parámetro que pondera la importancia relativa de la forma frente a la magnitud.

El coeficiente de correlación entre \mathbf{x}_i y \mathbf{x}_j se define como:

$$\rho(\mathbf{x}_i, \mathbf{x}_j) = \frac{\sum_{t=1}^n (x_{it} - \bar{x}_i)(x_{jt} - \bar{x}_j)}{\sqrt{\sum_{t=1}^n (x_{it} - \bar{x}_i)^2} \sqrt{\sum_{t=1}^n (x_{jt} - \bar{x}_j)^2}},$$

donde \bar{x}_i y \bar{x}_j representan las medias de las respectivas series.

La distancia combinada $d(\cdot, \cdot)$ permite identificar tanto diferencias en el *perfil temporal* (por ejemplo, forma de la curva) como en el *nivel de producción* (magnitud). Una distancia cercana a cero indica que dos inversores presentan un comportamiento muy similar, tanto en forma como en magnitud, mientras que valores elevados reflejan diferencias sustanciales. Además, observar que por la definición de $d(\cdot, \cdot)$, la métrica está acotada

$$0 \leq d(\mathbf{x}_i, \mathbf{x}_j) \leq 1 \quad \forall i, j.$$

2.3 Cálculo de la Matriz de Distancias

La disimilitud entre cada par de vectores $(\mathbf{x}_i, \mathbf{x}_j)$ se representa mediante $d_{i,j}$, que combina diferencias de forma y magnitud conforme a (1). Observar que la distancia euclidiana se normaliza, la cual se realiza una única vez, antes del agrupamiento, y permite que ambas componentes (forma y magnitud) sean comparables en escala. A partir de esta métrica, se construye una matriz de distancias simétrica $\mathbf{D} \in \mathbb{R}^{p \times p}$, con ceros en la diagonal:

$$\mathbf{D} = \begin{bmatrix} 0 & d_{1,2} & \cdots & d_{1,p} \\ d_{2,1} & 0 & \cdots & d_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ d_{p,1} & d_{p,2} & \cdots & 0 \end{bmatrix}.$$

Esta matriz es la base del algoritmo de agrupamiento jerárquico, ya que determina qué pares de series son más similares y deben fusionarse en cada iteración del proceso.

2.4 Inicialización

Al comienzo, cada variable \mathbf{x}_i forma su propio cúmulo:

$$\mathcal{C}_i = \{\mathbf{x}_i\}, \quad i \in \{1, \dots, p\}.$$

En consecuencia, hay un total de p cúmulos en esta etapa inicial.

2.5 Matriz de Distancias entre Cúmulos

La distancia entre dos cúmulos \mathcal{C}_i y \mathcal{C}_j se define a través de una función de enlace (*linkage*), la cual determina cómo calcular la disimilitud entre ellos $d(\mathcal{C}_i, \mathcal{C}_j)$. Existen distintas funciones de enlace que permiten definir esta medida de distancia entre cúmulos. Una vez calculada la matriz de distancias entre cada par de vectores, utilizando la métrica combinada de forma y magnitud previamente definida, es posible aplicar diferentes criterios de enlace para guiar el proceso de fusión.

2.6 Funciones de Enlace Comunes

A continuación, se describen tres de las funciones de enlace más utilizadas en el análisis de agrupamiento jerárquico:

- **Single linkage (mínima distancia):** define la distancia entre dos clústers \mathcal{C}_i y \mathcal{C}_j como la menor distancia entre cualquier par de elementos $\mathbf{x} \in \mathcal{C}_i$ y $\mathbf{y} \in \mathcal{C}_j$:

$$d(\mathcal{C}_i, \mathcal{C}_j) = \min_{\mathbf{x} \in \mathcal{C}_i, \mathbf{y} \in \mathcal{C}_j} d(\mathbf{x}, \mathbf{y}).$$

- **Complete linkage (máxima distancia):** define la distancia entre dos clústers \mathcal{C}_i y \mathcal{C}_j como la mayor distancia entre cualquier par de elementos $\mathbf{x} \in \mathcal{C}_i$ y $\mathbf{y} \in \mathcal{C}_j$:

$$d(\mathcal{C}_i, \mathcal{C}_j) = \max_{\mathbf{x} \in \mathcal{C}_i, \mathbf{y} \in \mathcal{C}_j} d(\mathbf{x}, \mathbf{y}).$$

- **Average linkage (distancia promedio):** define la distancia entre dos clústers \mathcal{C}_i y \mathcal{C}_j como el promedio de todas las distancias entre pares de elementos $\mathbf{x} \in \mathcal{C}_i$ y $\mathbf{y} \in \mathcal{C}_j$:

$$d(\mathcal{C}_i, \mathcal{C}_j) = \frac{1}{|\mathcal{C}_i||\mathcal{C}_j|} \sum_{\mathbf{x} \in \mathcal{C}_i} \sum_{\mathbf{y} \in \mathcal{C}_j} d(\mathbf{x}, \mathbf{y}).$$

donde $|\cdot|$ representa la cardinalidad del cúmulo. Este criterio busca un equilibrio entre los enfoques de mínima y máxima distancia, y tiende a formar clústers de forma y tamaño más homogéneos.

Cabe señalar que en estas funciones de enlace, cada clúster puede contener múltiples vectores, especialmente en etapas posteriores del algoritmo jerárquico.

2.7 Algoritmo Jerárquico Aglomerativo

El procedimiento de agrupamiento jerárquico aglomerativo parte de los vectores individuales como cúmulos iniciales y, en cada iteración, combina los dos cúmulos más cercanos hasta formar un único grupo. El proceso puede describirse de la siguiente manera:

1. **Inicialización:** cada vector \mathbf{x}_i se considera como un cúmulo individual, es decir, $\mathcal{C}_i = \{\mathbf{x}_i\}$.
2. **Cálculo inicial de distancias:** se construye la matriz de distancias \mathbf{D} entre todos los cúmulos. Dado que al inicio cada cúmulo contiene un solo vector, la distancia entre cúmulos en esta fase inicial coincide con la distancia base entre vectores definida en (1).
3. **Selección de cúmulos a fusionar:** se identifica el par de cúmulos más próximos de acuerdo con la función de enlace elegida:

$$(\mathcal{C}_a, \mathcal{C}_b) = \arg \min_{i \neq j} d(\mathcal{C}_i, \mathcal{C}_j).$$

Aquí, $d(\mathcal{C}_i, \mathcal{C}_j)$ representa la distancia entre cúmulos, calculada a partir de las distancias entre los vectores de ambos cúmulos mediante la regla de enlace (single, complete, average, etc.).

4. **Fusión:** se combinan los dos cúmulos seleccionados en uno nuevo:

$$\mathcal{C}_{ab} = \mathcal{C}_a \cup \mathcal{C}_b,$$

donde la operación corresponde a la unión de conjuntos, es decir, \mathcal{C}_{ab} contiene todos los vectores de \mathcal{C}_a y de \mathcal{C}_b . No se trata de una concatenación, sino de una fusión en un único cúmulo.

5. **Actualización de la matriz de distancias:** se recalculan las distancias entre el nuevo cúmulo \mathcal{C}_{ab} y todos los cúmulos restantes, aplicando nuevamente la función de enlace.
6. **Iteración:** se repite el proceso desde el paso 3 hasta que todos los vectores estén agrupados en un único cúmulo. Este procedimiento genera una estructura jerárquica que puede representarse gráficamente mediante un dendrograma.

2.8 Dendrograma

El resultado final se representa en un **dendrograma**, un árbol binario que muestra visualmente cómo se fusionaron los cúmulos. La altura de cada fusión refleja la disimilitud entre los cúmulos.

Este gráfico permite identificar:

- Cúmulos con comportamientos similares (por ejemplo, inversores con perfiles de potencia parecidos).
- Agrupaciones naturales dentro de los datos.
- Niveles de corte adecuados para particionar los datos en un número deseado de cúmulos.

3. RESULTADOS EXPERIMENTALES

La planta fotovoltaica (FV) considerada en este estudio está conectada a la red de distribución de media tensión (34.5 kV) mediante una subestación con capacidad de 300 kVA. Está compuesta por 16 generadores fotovoltaicos (GFV) trifásicos, cada uno con una capacidad nominal de 15 kW, como se muestra en la Fig. 1. Los inversores están conectados en paralelo en el lado de carga del sistema, operando a un nivel de tensión de 220 V.

En la Fig. 2 se presentan los detalles de configuración de cada uno de los GFV. Aunque existen distintos arreglos fotovoltaicos entre inversores, la potencia máxima de salida de todos ellos está limitada a 15 kW. La distribución de los arreglos FV es la siguiente:

- Grupo1 (18 kWp):** cuatro GFVs (9, 11, 13, 15) están compuestos por cuatro cadenas en paralelo cada una con 10 módulos de 450 Wp;
- Grupo2 (16.2 kWp):** cuatro GFVs (10, 12, 14, 16) tienen cuatro cadenas en paralelo cada una con 9 módulos de 450 Wp;
- Grupo3 (16.35 kWp):** siete GFVs (2-8) están formados por dos cadenas en paralelo con 15 módulos de 545 Wp cada una;
- Grupo4 (15.26 kWp):** un GFV (1) está formado por dos cadenas en paralelo cada una con 14 módulos de 545 Wp.

A través del sistema de monitoreo, se tiene acceso a parámetros eléctricos tanto en el lado de CD como en CA de cada inversor. Específicamente, se dispone

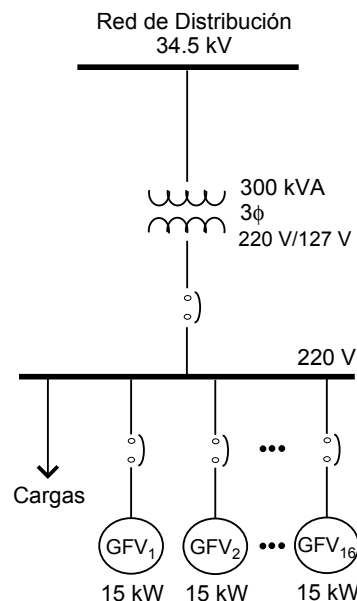


Figura 1. Diagrama unifilar simplificado de la planta fotovoltaica de 240 kW (Generación Neta).

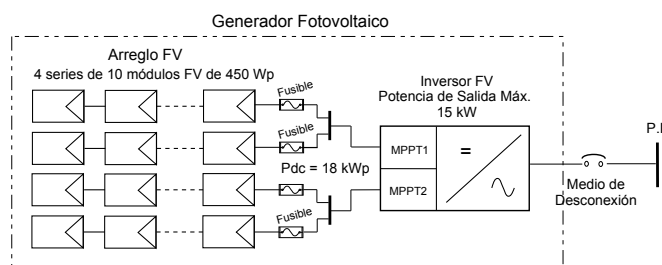


Figura 2. Diagrama esquemático de un generador fotovoltaico (GFV) que contiene 4 cadenas en paralelo de 10 módulos FV cada una.

de registros de voltaje, corriente en ambos lados del inversor, y potencia del lado de CA. En este estudio sólo estamos analizando las potencias de los arreglos FV denominadas $P_{dc1}, P_{dc2}, \dots, P_{dc16}$, las cuales se calculan a partir de los voltajes y corrientes de CD. Proponemos realizar la evaluación del desempeño diariamente. En consecuencia, se obtienen 16 series de tiempo por día, es decir, P_{dcj} con $j = \{1, 2, \dots, 16\}$, cada una con 288 observaciones, correspondientes a registros en intervalos de 5 minutos. Estas series temporales constituyen la base para evaluar la similitud en el comportamiento de los inversores vecinos, con el objetivo de identificar posibles desviaciones o patrones anómalos en su desempeño. En la Fig. 3 se presentan las curvas de generación de potencia en corriente directa (CD) correspondientes al 19 de mayo de 2025 para cada uno de los inversores que conforman el sistema fotovoltaico (FV). A partir de estas curvas, resulta difícil identificar patrones evidentes, lo que sugiere que una inspección visual por parte de un operador podría llevar a la conclusión errónea de

que el comportamiento de los inversores es uniforme o aparentemente normal.

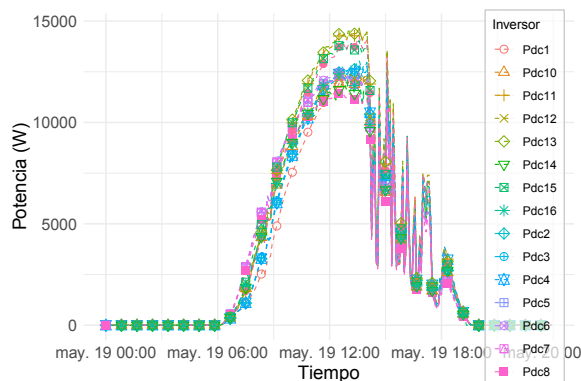


Figura 3. Curvas de generación del 19 de mayo de 2025.

La Fig. 4 presenta el resultado del análisis de agrupamiento jerárquico, representado mediante un dendrograma construido a partir de los datos correspondientes al 19 de mayo de 2025. El análisis de agrupamiento jerárquico se implementó en RStudio (Posit Team 2025), utilizando las librerías `dplyr` para la preparación de los datos, `stats` para el cálculo de distancias y agrupamiento, y `dendextend` para la visualización del dendrograma. Se empleó una métrica combinada basada en una ponderación entre disimilitud por correlación inversa y distancia euclidiana normalizada, controlada mediante un parámetro de mezcla $\lambda = 0.2$. En el gráfico se identifican claramente cuatro grupos principales.

- **Grupo A1:** ($P_{dc9}, P_{dc11}, P_{dc13}, P_{dc15}$)
- **Grupo B1:** ($P_{dc1}, P_{dc2}, P_{dc3}, P_{dc4}$)
- **Grupo C1:** ($P_{dc10}, P_{dc12}, P_{dc14}, P_{dc16}$)
- **Grupo D1:** ($P_{dc5}, P_{dc6}, P_{dc7}, P_{dc8}$)

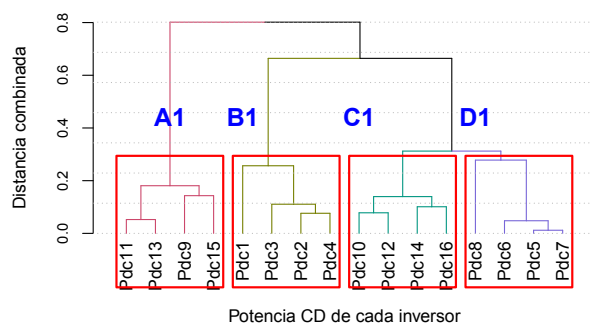


Figura 4. Dendrograma del 19 de mayo de 2025.

Dentro del grupo D1, se observa en Fig. 4 que el inversor P_{dc8} presenta un comportamiento distinto, separándose ligeramente de los demás de su grupo. Esta diferencia puede deberse a uno o varios de los siguientes factores:

- Valores de generación significativamente más bajos,
- Valores anómalamente altos,

- Mayor variabilidad en el perfil de generación,
- Forma de la curva de producción distinta al resto.

La Fig. 5 muestra las curvas de generación de los cuatro inversores del grupo D1. En esta comparación descriptiva se observa que P_{dc8} presenta una producción consistentemente inferior respecto a los demás, lo cual constituye un indicio de posible suciedad o sombreado parcial, aunque no se ha aplicado en esta etapa un contraste estadístico formal.

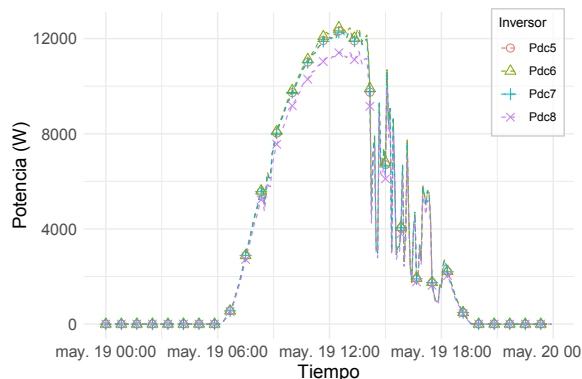


Figura 5. Curvas P_{dc5} , P_{dc6} , P_{dc7} y P_{dc8} del 19 de mayo de 2025.

De manera similar, dentro del grupo B1, el inversor P_{dc1} también presenta un comportamiento atípico, separándose sutilmente del resto de los inversores en su grupo. La Fig. 6 muestra las curvas de generación de los cuatro inversores de dicho grupo, donde se constata que P_{dc1} presenta una producción inferior al promedio de sus compañeros de grupo.

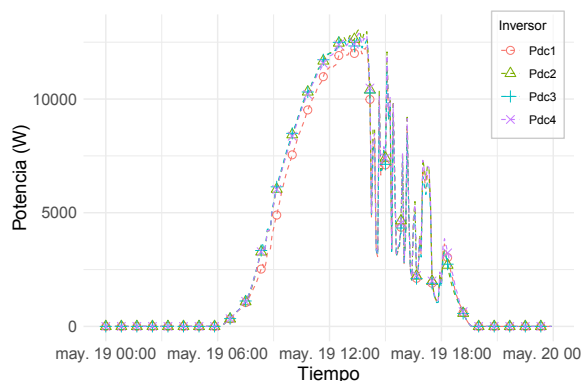


Figura 6. Curvas P_{dc1} , P_{dc2} , P_{dc3} y P_{dc4} del 19 de mayo de 2025.

A continuación se presentan los resultados correspondientes al 21 de mayo de 2025, con el objetivo de fortalecer la validación de las ideas propuestas en este trabajo. En la Fig. 7 se muestra el dendrograma obtenido a partir de los datos de generación del sistema FV en dicho día. En el gráfico se identifican nuevamente cuatro grupos principales.

- **Grupo A2:** ($P_{dc9}, P_{dc11}, P_{dc13}, P_{dc15}$)
- **Grupo B2:** ($P_{dc5}, P_{dc6}, P_{dc7}, P_{dc8}$)
- **Grupo C2:** (P_{dc14}, P_{dc16})
- **Grupo D2:** ($P_{dc1}, P_{dc2}, P_{dc3}, P_{dc4}, P_{dc10}, P_{dc12}$)

A través de este análisis, se observa que los grupos A2 y B2 se mantienen sin cambios respecto al 19 de mayo. Nuevamente P_{dc8} se destaca en el cúmulo B2 por su comportamiento diferente, como en el día anterior, lo que ratifica el análisis previo. Por otro lado, los grupos C2 y D2 han sufrido modificaciones. En particular, se identifica un nuevo subgrupo conformado únicamente por los inversores P_{dc14} y P_{dc16} , los cuales se agrupan de forma aislada respecto al resto.

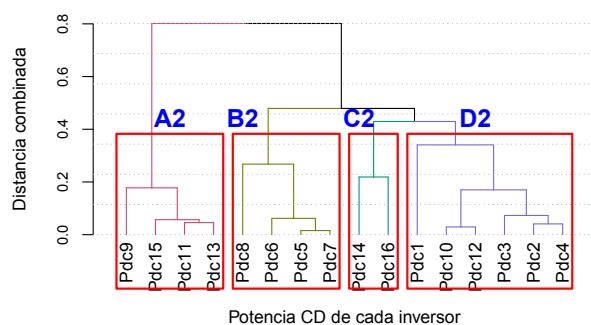


Figura 7. Dendrograma del 21 de mayo de 2025.

La Fig. 8 permite visualizar con mayor claridad la distorsión en la forma de las curvas de generación de los inversores P_{dc14} y P_{dc16} , en comparación con los demás integrantes del grupo original. Esta alteración en la forma de las curvas podría deberse a la presencia de una sombra temporal que afectó parcialmente la producción de ambos inversores.

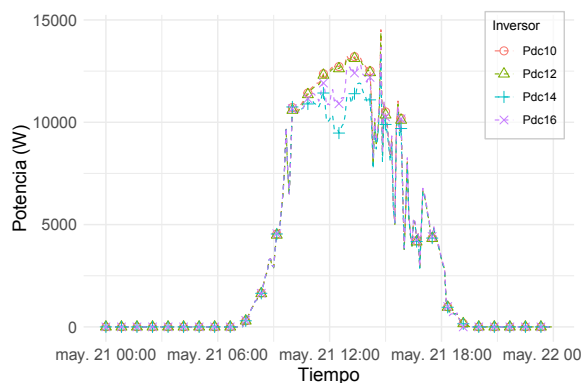


Figura 8. Curvas P_{dc10} , P_{dc12} , P_{dc14} y P_{dc16} del 21 de mayo de 2025.

4. COMENTARIOS FINALES

El análisis de agrupamiento jerárquico aplicado a las curvas de generación de los inversores FV permite transformar un conjunto complejo de datos en una estructura

más interpretable, facilitando la identificación de patrones, similitudes y posibles desviaciones relevantes. Si una serie no aparece en el dendrograma, puede deberse a la falta de variabilidad en sus datos, como valores todos nulos (NA), lo que podría indicar un sensor fuera de servicio. Este enfoque ofrece varias ventajas: permite reducir el número de curvas a revisar enfocándose solo en casos anómalos; entrega una base objetiva y reproducible para el análisis; y posibilita la automatización del monitoreo, generando alertas solo cuando un inversor cambia de grupo o se comporta de forma aislada. Además, puede utilizarse para evaluar la estabilidad de los agrupamientos a lo largo del tiempo, lo cual es útil como herramienta complementaria para la supervisión operativa de sistemas FV.

REFERENCIAS

- Bilal Taghezouit, Fouzi Harrou, Ying Sun, and Walid Merrouche. Model-based fault detection in photovoltaic systems: A comprehensive review and avenues for enhancement. *Results in Engineering*, Volume 21, 2024, Article 101835. ISSN: 2590-1230.
- International Energy Agency (IEA). *Photovoltaic Power Systems Programme – Performance, Operation and Reliability of Photovoltaic Systems*. IEA PVPS Task 13, Report IEA-PVPS T13-25:2022, October 2022. ISBN 978-3-907281-13-0.
- Joakim Widén and Joakim Munkhammar. *Solar Radiation Theory*. Uppsala University, 2019. ISBN 978-91-506-2760-2. DOI: 10.33063/diva-381852.
- International Energy Agency (IEA). *Photovoltaic Power Systems Programme – The Use of Advanced Algorithms in PV Failure Monitoring*. IEA PVPS Task 13, Report IEA-PVPS T13-19:2021, September 2021. ISBN 978-3-907281-07-9.
- Secretaría de Energía (SENER). *NORMA Oficial Mexicana NOM-001-SEDE-2012, Instalaciones Eléctricas (utilización)*. Diario Oficial de la Federación, México, 2012. Artículo 690 [690-11, 690-35 c)].
- Tara Doyle, Ryan Desharnais, and Michael Mills-Price. *2019 PV Inverter Scorecard*. PV Evolution Labs (PVEL), May 2019. <https://www.pvel.com/inverter-scorecard/>.
- Thushara Gunda and Rachel Homan. *Evaluation of Component Reliability in Photovoltaic Systems Using Field Failure Statistics*. Sandia National Laboratories, Albuquerque, NM, Report SAND2020-9231, 2020. DOI: 10.2172/1660804.
- H. Zhu, Z. Wang, and R. S. Balog. Real time arc fault detection in PV systems using wavelet decomposition. In *2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC)*, Portland, OR, USA, pages 1761–1766, 2016. DOI: 10.1109/PVSC.2016.7749926.
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second Edition, Springer, 2009. ISBN 978-0-387-84857-0.
- Posit team. *RStudio: Integrated Development Environment for R*. Posit Software, PBC, Boston, MA, 2025. URL: <http://www.posit.co/>.