

PRPD – GUI Unificada (README)

Requisitos del sistema

- **Sistema Operativo:** Windows 11 (probado). *Nota:* La aplicación está diseñada principalmente para Windows, aunque podría ejecutarse en otros sistemas con Python instalado.
- **Python:** Version 3.13 o superior ¹.
- **Bibliotecas Python necesarias:** PySide6 (interfaz gráfica), numpy, matplotlib, scikit-learn ². Se recomienda instalar todos los requisitos ejecutando `pip install -r PRPDapp/requirements.txt` en la raíz del proyecto (donde se encuentra el archivo `requirements.txt` con las dependencias).

Hardware: Un PC estándar es suficiente (la carga computacional es baja; por ejemplo, el algoritmo de filtrado tarda ~76 ms en un portátil Core i5 ³). Se recomienda contar con aceleración gráfica básica para renderizar los gráficos de la GUI.

Instalación y ejecución

1. **Obtener el código:** Clone o descargue el repositorio de la aplicación PRPD GUI Unificada en su máquina local, asegurándose de mantener la estructura de archivos (particularmente la carpeta `PRPDapp` y sus submódulos).

2. **Instalar dependencias:** Abra una terminal en la carpeta raíz del repositorio y ejecute:

```
pip install -r PRPDapp/requirements.txt
```

Esto instalará PySide6, numpy, matplotlib, scikit-learn, y cualquier otra dependencia requerida.

3. **Iniciar la aplicación:** Hay dos formas principales de ejecutar la GUI:
4. **Doble clic (Windows):** Use el script de conveniencia `PRPDapp\run_gui.bat` (incluido en el repositorio) haciendo doble clic sobre él en el Explorador de Windows ⁴. Este script lanzará la interfaz gráfica automáticamente.
5. **Línea de comandos:** Navegue hasta la raíz del repositorio (`cd <ruta_del_repo>`) y ejecute:

```
python -m PRPDapp.main
```

Esto iniciará la GUI como un módulo de Python ⁴. *Nota:* Es importante ejecutar la aplicación **como módulo** desde la raíz del proyecto para evitar errores de importación (`ModuleNotFoundError: No module named 'PRPDapp'`) ⁵. En Windows, puede usar el `.bat` mencionado; en otros SO, asegúrese de invocar el módulo con la sintaxis `python -m PRPDapp.main`.

6. **Compatibilidad:** Existen variantes legacy (`PRPDapp.main_utf8`, `PRPDapp.psg_main`), pero **no es necesario usarlas** – todas redirigen a la GUI unificada actual y se mantienen solo por compatibilidad ⁶. Ejecute siempre la aplicación principal como se indicó arriba.

Una vez ejecutada, debería aparecer la ventana principal de la GUI PRPD Unificada. Si tiene problemas (por ejemplo, falta de PySide6), verifique haber instalado los requisitos. En caso de texto codificado extraño en la interfaz, asegúrese de estar ejecutando la versión unificada (usa UTF-8) y **no** un script antiguo separado ⁷.

Descripción general de la interfaz y propósito del software

PRPD GUI Unificada es una herramienta gráfica interactiva para el **análisis de patrones PRPD (Partial Discharge Phase-Resolved)** con capacidades avanzadas de **filtrado de ruido, visualización diagnóstica y exportación de reportes** ¹. Su propósito es facilitar la identificación y clasificación de fuentes de descargas parciales (PD) en equipos eléctricos (p. ej. generadores hidroeléctricos) mediante técnicas modernas de procesamiento de datos y aprendizaje automático.

La aplicación integra en una sola interfaz varias funciones que típicamente requerirían herramientas separadas: ajuste de fase de las descargas, filtrado de ruido basado en *clústeres* y cuantiles, algoritmos de *denoising* de imágenes PRPD, extracción de características (distribuciones de amplitud y fase), visualización de histogramas y nubes de puntos, clasificación automática con **red neuronal (ANN)**, y evaluación de severidad con métricas clave (KPIs). Todo esto permite al usuario analizar un archivo de medición de descargas parciales y obtener en segundos un diagnóstico reproducible con gráficos interpretables y datos exportables.

¿Qué es un patrón PRPD? Es una representación de las descargas parciales registradas en función de su fase respecto al ciclo de alimentación y su amplitud. Cada punto en un diagrama PRPD típico representa una descarga, ubicada por el ángulo de fase (0°–360° en el ciclo de CA) y su magnitud. Los patrones PRPD característicos pueden indicar el tipo de defecto de aislamiento presente (cavidad interna, descarga superficial, corona, etc.), pero suelen estar contaminados de ruido y descargas no dominantes. Esta herramienta implementa un algoritmo novel que **elimina ruido disperso y suprime descargas no dominantes**, aislando esencialmente las “nubes” de descargas principales para un patrón de una sola fuente ⁸. Además, calcula **nuevas características basadas en histogramas** de distribución de descargas en amplitud y fase ⁹, las cuales han demostrado mejorar la clasificación automática de PD en comparación con métodos tradicionales ¹⁰.

En resumen, la GUI PRPD Unificada está diseñada para:

- **Cargar datos PRPD** (archivos CSV o XML de descargas parciales obtenidos de sistemas de monitoreo en línea).
- **Alinear automáticamente la fase** de las descargas (0°, 120° o 240°) para maximizar la concentración, o permitir al usuario fijar manualmente la referencia de fase.
- **Filtrar el ruido** y descargas espurias mediante **gating** inteligente (filtros S1/S2 definidos por cuantiles) antes de la agrupación.
- **Agrupar descargas en “nubes” (clústeres)** usando clustering no supervisado (DBSCAN), combinando y destacando las nubes dominantes correspondientes a la fuente principal de PD.
- **Calcular histogramas ANGPD** (por sus siglas en inglés, *Adjacent to Noise Gap PD* vs *Non-ANGPD*, ver sección correspondiente) que representan la distribución de descargas a lo largo de la fase, tanto normalizada por el total como normalizada por el pico.
- **Generar predicciones de clasificación** de la fuente de PD usando un modelo de Red Neuronal Artificial (ANN) entrenado previamente o mediante heurísticas internas si no se carga un modelo.

- **Mostrar KPIs** (indicadores de desempeño clave), como conteo de descargas, amplitud pico (percentil 95), densidad relativa, concentración de fase, etc., e incluso comparar con un **baseline** (línea base) previo para evaluar la evolución en el tiempo.
- **Exportar resultados** de forma reproducible: gráficos, archivos CSV de datos procesados e incluso reportes resumidos, para documentación y análisis fuera de la herramienta ¹¹.

Esta interfaz ofrece a ingenieros y especialistas en aislamiento una **guía visual y cuantitativa** para diagnosticar descargas parciales, apoyada en investigaciones recientes ⁸. A continuación, se detalla cada componente de la GUI y cómo usarlo para obtener el máximo beneficio.

Descripción detallada de la GUI (botones, secciones y vistas)

Al iniciar la aplicación, se presenta la ventana principal dividida en secciones: en la parte superior, una **barra de herramientas** con botones y controles desplegable; en el centro, un área de **gráficos** dividida en cuadrantes; y en la parte inferior, una barra informativa (que puede mostrar un banner o mensaje de firma si está configurado).

A continuación se enumeran **todos los controles, botones y secciones** de la GUI, con su función:

Barra superior de controles

- **Botón "Abrir PRPD..."** – Abre un diálogo de selección de archivo para cargar datos de descargas parciales. Se aceptan archivos en formato CSV o XML (extensiones `.csv` o `.xml`) ¹². Al seleccionar un archivo y confirmarlo, la aplicación lo carga y muestra inmediatamente la dispersión PRPD **cruda** en el gráfico (parte superior izquierda) ¹³. (*Nota:* El formato CSV esperado es típicamente dos columnas: fase (o tiempo) y amplitud; el formato XML debe contener elementos con la información de descargas, generalmente utilizados por equipos de monitoreo en línea. Si el archivo XML incluye etiquetas `<times>` con tiempos absolutos, estos se utilizan para cálculos de *gap time*, ver más adelante).
- **Botón "Procesar"** – Ejecuta el procesamiento completo sobre el archivo cargado, aplicando los parámetros seleccionados (fase y filtro) paso a paso: alineación de fase, filtrado S1/S2, clustering, cálculo de histogramas ANGPD y clasificación. Al finalizar, actualiza todos los gráficos de la interfaz con los resultados. Además, la acción de "Procesar" **exporta automáticamente** los resultados clave en archivos dentro de la carpeta `out\reports\` (creada si no existe) ¹⁴, usando el nombre base del archivo de datos. Por ejemplo, si cargó `muestra1.xml` y presiona Procesar, obtendrá:
 - `muestra1_angpd.csv` y `muestra1_angpd.png` – datos numéricos y gráfico de las curvas ANGPD/N-ANGPD ¹⁵.
 - `muestra1_clouds_raw.csv`, `muestra1_clouds_combined.csv`, `muestra1_clouds_selected.csv` – detalles de clústeres S3, S4 y S5 respectivamente (centroides de fase y amplitud, conteos, etc.) ¹⁶, junto con `muestra1_clouds.png` (imagen de las nubes de puntos) ¹⁷.
 - `muestra1_metrics.csv` – métricas calculadas por el algoritmo para ese patrón (ver sección de KPIs), incluyendo la clasificación prevista (`predicted`), puntaje de severidad (`severity`), amplitud P95, densidad, concentración de fase, offset de fase aplicado y nivel de filtrado usado ¹⁸.
 - `muestra1_baseline.json` – si presionó "Reset baseline" en algún momento para este archivo, se guarda/actualiza un JSON con valores de referencia (ver sección Baseline) ¹⁹.

Sugerencia: Siempre ejecute **Procesar** después de cargar o cambiar parámetros para refrescar los resultados. Si no se ha cargado ningún archivo, este botón no tendrá efecto. Si intenta exportar PDF o ver resultados sin procesar, la GUI le recordará que primero debe ejecutar el procesamiento.

- **Desplegable "Fase:"** – Permite seleccionar el **alineamiento de fase** de los datos:
- **Auto (0/120/240)** – (*Por defecto*) La herramienta determinará automáticamente un **desplazamiento de fase óptimo** entre 0°, 120° o 240° para alinear las descargas ²⁰. Este algoritmo elige el offset (0, 120 o 240) que maximiza la concentración de las descargas en el ciclo, es decir, que hace más patente la separación entre polos de descarga. Esto es útil en sistemas trifásicos donde la referencia de fase de la medición puede ser arbitraria; la auto-alineación situará las “nubes” principales en posiciones estándar para facilitar la comparación.
- **0°, 120°, 240°** – Seleccionan manualmente un desplazamiento fijo. Use estas opciones si desea **forzar una referencia de fase** concreta. Por ejemplo, si sabe que el defecto corresponde a la fase A de un sistema trifásico, podría alinear a 0° esa fase. En general, se recomienda dejar *Auto* para la mayoría de análisis iniciales (la herramienta aplicará 0°, 120° o 240° automáticamente según los datos). Al cambiar esta opción, vuelva a presionar **Procesar** para aplicar el nuevo alineamiento.
- **Desplegable "Filtro:"** – Selecciona el nivel de **filtrado de ruido (gating)** antes del clustering. Hay tres niveles disponibles ²¹:
 - **S1 Weak:** Filtrado débil. Aplica el criterio S1 para eliminar ruido mínimo. Concretamente, **elimina el quintil más frecuente de la métrica** `quantity` de las descargas ²². Aquí `quantity` se refiere al número de descargas repetidas en la misma posición de fase (si el XML lo provee; ver nota abajo). En términos simples, S1 descarta el grupo de eventos más repetitivo (por ejemplo, ruido periódico) pero conserva la mayoría de las descargas. Use S1 cuando quiera observar prácticamente todas las descargas excepto quizás un ruido muy obvio y repetitivo.
 - **S2 Strong:** Filtrado fuerte. Aplica S1 **y adicionalmente elimina el quintil inferior de amplitud** ²³. Esto remueve también las descargas de amplitud más baja (que suelen ser ruido de fondo). Use S2 para condiciones de ruido moderado, donde hay muchos eventos pequeños irrelevantes que conviene descartar junto con la componente más repetitiva.
 - **S2 Stronger:** Filtrado muy fuerte. **Elimina los dos quintiles más frecuentes de** `quantity` **y el quintil inferior de amplitud** ²⁴. Es el filtrado más agresivo: quita tanto las descargas más repetitivas (posibles interferencias periódicas) como las de amplitud muy baja. Esto deja idealmente solo las descargas más significativas del patrón principal, pero también podría descartar descargas reales de bajo nivel, así que úselo cuando el ruido sea abundante o para enfocar únicamente en los pulsos más dominantes.

Notas sobre filtrado: Los filtros S1/S2 se aplican **antes del clustering (etapa S3)**, afectando qué eventos se consideran en los análisis posteriores ²⁵ ²⁶. Si el archivo PRPD **no contiene el campo** `quantity`, el componente de filtrado por cantidad se omite automáticamente (por ejemplo, S1 no tendrá efecto en ausencia de `quantity`) ²⁷, y solo se aplicará la parte de amplitud en S2/S2 stronger. En cambio, si el XML sí provee `quantity` (conteo de descargas por fase o por pixel), el algoritmo **nunca descartará completamente esa información:** la utiliza para decidir qué eventos filtrar pero preserva el campo `quantity` de los eventos “kept” (conservados) para cálculos posteriores ²⁸. Tras el gating, los eventos restantes pasan a la etapa de clustering y cálculo de características.

- **Checkbox "S1+S2 PNG":** (Opcional) Si se activa, la aplicación generará **imágenes combinadas** comparativas de filtros S1 y S2. En la versión actual, esta opción está prevista para automatizar la

exportación de gráficos comparando la vista con filtro débil vs. fuerte. Por ejemplo, podría guardar un PNG adicional mostrando lado a lado la nube con S1 vs S2. *Actualmente*, no obstante, esta funcionalidad está **en desarrollo** y su comportamiento puede ser limitado. Se incluye para futuras versiones donde al exportar resultados se generarán visualizaciones complementarias con ambos filtros para análisis más completo.

- **Checkbox "Densidad (hist2D)":** Activado por defecto. Controla la **visualización de los gráficos PRPD como mapa de densidad** en lugar de dispersión de puntos individuales ²⁹ ³⁰. Cuando está marcado, la GUI dibuja los gráficos de PRPD crudo y filtrado usando un **histograma 2D (Fase vs Amplitud)** con color representando la densidad de puntos ²⁹. Esto es útil para archivos con muchísimos puntos donde una nube dispersa sería difícil de ver; en su lugar, verá una imagen estilo calor (heatmap) donde áreas más "calientes" (colores más intensos) indican mayor concentración de descargas. La escala de color es relativa (se agrega un pequeño valor constante para evitar divisiones por cero, por lo que zonas sin descargas aparecen en un color base muy tenue).
- **Overlay de ruido:** Incluso con hist2D activo, la GUI superpone **puntos grises semitransparentes** para indicar eventos considerados ruido/eliminados ³¹ ³². Así, podrá apreciar dónde estaban los puntos filtrados. Estos aparecen como puntos gris claro (alpha ~0.1–0.15) tanto sobre el gráfico crudo como en el filtrado, ayudando a ver qué fue descartado por el gating.
- Si desactiva "Densidad", entonces los gráficos PRPD se dibujarán con **puntos individuales** (scatter plot): círculos semitransparentes (por defecto de tamaño 3–4 px) representando cada descarga ³³. Esto puede ser preferible con menos datos o si desea ver la distribución exacta de puntos. En cualquier caso, los ejes muestran Fase (°) y Amplitud (típicamente en mV o unidades normalizadas 0–100).
- **Botón "Cargar ANN":** Permite cargar un **modelo de Red Neuronal Artificial** entrenado para la clasificación de patrones PD. Al hacer clic, abre un diálogo para seleccionar un archivo de modelo (*.pkl o *.joblib) ³⁴ ³⁵. Debe ser un modelo compatible con `scikit-learn` u otro formato serializado con joblib. Una vez seleccionado, la aplicación intentará cargarlo:
 - Si el modelo fue guardado usando el *loader* personalizado (`models/ann_loader.py` incluido), se usará esa rutina para cargar y extraer también los nombres de clases ³⁶.
 - Si no, se intentará usar un objeto interno `PRPDANN` que soporta cargar modelos `MLPClassifier` de scikit-learn por defecto ³⁵.
 - Si la carga tiene éxito, mostrará un mensaje "Modelo cargado" ³⁷ ³⁸. A partir de entonces, cada vez que procese un archivo, las **probabilidades y predicciones** mostradas corresponderán a este modelo. (Si ya había resultados procesados sin modelo, la GUI recalculará resultados con el nuevo modelo cargado automáticamente para actualizar la vista ³⁹).
- **Feature Order:** Si en la carpeta `models/` existe un archivo `feature_order.json`, la aplicación lo utilizará para **alinear el vector de entrada de características** del patrón con el orden esperado por el modelo ⁴⁰. Esto es importante porque la red neuronal espera los inputs (por ejemplo, los 64 valores de histogramas más otros KPIs) en cierto orden. El JSON debe listar el orden de características usado en el entrenamiento; la GUI reorganizará sus propias features calculadas para coincidir con ese orden antes de predecir. *Consejo:* Asegúrese de que el modelo y la definición de features correspondan; de lo contrario, la predicción puede ser errónea.

Heurística por defecto: Si no se carga ningún modelo ANN, la herramienta puede proporcionar una clasificación aproximada basada en reglas o un modelo interno por defecto. En la implementación actual, se calcula un conjunto de indicadores (p. ej., patrón

de fase, simetría, densidad de ruido) y se asigna una categoría tentativa. Esto se refleja en la vista de **Probabilidades** como barras; aún si no hay ANN, verá probabilidades estimadas (por ejemplo, si detecta mucha dispersión quizás “ruido” tenga alta probabilidad). Para mejores resultados, se recomienda entrenar un modelo con datos conocidos y cargarlo.

- **Botón "3D"**: Abre una **visualización 3D** de los datos PRPD filtrados y clusterizados. Este botón estará habilitado después de procesar un archivo (es decir, cuando existan datos de puntos y etiquetas de clúster disponibles). Al hacer clic, si hay datos:
 - Se lanza una ventana o gráfico 3D (utilizando la función `plot_prpd_3d`) mostrando los puntos de descarga en un espacio tridimensional ⁴¹. Típicamente, los ejes son Fase, Amplitud y una tercera dimensión que puede ser la densidad o simplemente un índice; en nuestro caso, el 3D está configurado para ilustrar mejor la separación de clústeres. Los puntos aparecerán coloreados según su pertenencia a clúster (igual que en 2D) y posiblemente con distinta forma para resaltar los centroides.
 - Esta visualización es útil para inspeccionar la distribución de descargas cuando la dimensión de *quantity* o la separación entre clústeres necesita otro ángulo. Por ejemplo, a veces dos clústeres pueden solaparse en 2D pero distinguirse al “girar” la vista.
- *Nota*: Requiere soporte de Matplotlib 3D o bibliotecas similares. Si ocurre algún error al generar el 3D, se notificará en un mensaje (p. ej., falta de backend 3D) ⁴². No es crítica esta función para el análisis principal, pero es un complemento visual poderoso.
- **Botón "Procesar carpeta"**: Ejecuta un **análisis por lotes (batch)** en todos los archivos XML de una carpeta elegida. Al pulsarlo, se abre un diálogo para seleccionar un directorio (carpeta raíz que contenga subcarpetas o archivos XML) ⁴³. Tras seleccionar la carpeta, la aplicación:
 - Busca recursivamente todos los archivos `*.xml` en esa ruta ⁴⁴. Si no encuentra XML, lanza una advertencia ⁴⁵.
 - Crea una carpeta de salida bajo `out/batch/` con el nombre de la carpeta y un *timestamp* ⁴⁵ (por ejemplo, si seleccionó `DatosEntrenamiento`, la salida podría ser `out/batch/DatosEntrenamiento_20251114_2055` con la fecha y hora actual).
 - Por cada archivo XML encontrado, ejecuta el procesamiento completo **tres veces (S1, S2, S2 Stronger)** ⁴⁶ ⁴⁷, aplicando el offset de fase seleccionado globalmente (Auto o fijo) en cada caso ⁴⁸. Para cada combinación archivo+filtro, guarda todos los resultados igual que en el modo individual (CSV de ANGPD, imágenes de nubes, histogramas, etc.) ⁴⁹ ⁵⁰.
 - Genera un **resumen** en pantalla y en archivos: muestra una lista con cada archivo procesado y su resultado principal (clase predicha, severidad, número de clústeres, si se detectó ruido) ⁵¹ ⁵² para el filtro S1 Weak (como referencia). Este mismo resumen se guarda en `batch_summary.txt` y en formato JSON (`batch_summary.json`) dentro de la carpeta de batch ⁵³.
 - Muestra un mensaje emergente con el texto resumen al finalizar ⁵⁴ ⁵⁵, y otro indicando la ruta de la carpeta de salida donde hallar los archivos generados ⁵⁵. Esto permite procesar grandes conjuntos de muestras de forma automática. *Ejemplo de uso*: puede correr “Procesar carpeta” sobre datos de entrenamiento para generar características e indicadores de muchos archivos de PD, y luego usar los CSV para entrenar su propio modelo ANN.
- **Botón "Exportar PDF"**: Genera un **informe PDF** del resultado actual. Una vez que haya procesado un archivo (de lo contrario le avisará “Ejecuta primero el procesamiento” si no hay resultados ⁵⁶), al pulsar Exportar PDF se compilará un documento PDF con los gráficos y datos

principales. Guardará el PDF en la carpeta de salida del archivo actual (típicamente junto a los CSV/PNG en `out\reports\`) y notificará la ubicación exacta ⁵⁷.

- Este reporte PDF incluirá normalmente: el gráfico PRPD con y sin filtrado, las curvas ANGPD, quizás los histogramas e información de métricas y clasificación. Es una manera rápida de obtener un documento para compartir o archivar los resultados de cierto archivo.
- *Nota:* Esta función está en una fase inicial de implementación; el formato del PDF podría ser básico. En futuras versiones, se planea un PDF más completo con todos los gráficos y explicaciones. Actualmente, considere el PDF como un *resumen visual rápido*.
- **Checkbox "Gap-time XML":** (Seguimiento) Al marcar esta casilla, le indica a la aplicación que utilizará información de **Gap-Time** para el análisis de criticidad. Gap-time se refiere al intervalo temporal entre descargas, un indicador de qué tan frecuente ocurre la actividad PD. Muchos equipos no proveen esto directamente, pero si tiene un XML con los tiempos absolutos de cada evento, puede aprovecharlo.
- Para usarlo, active la casilla y luego haga clic en el botón "..." a la derecha (botón "..." descrito a continuación) para seleccionar un archivo XML que contenga datos de tiempos. Ese archivo puede ser el mismo que cargó para análisis (si contiene `<times>`), u otro archivo histórico para comparar.
- La casilla por sí sola solo habilita la consideración de gap-time; el análisis concreto sucede al comparar con baseline (ver "Comparar vs base"). Si está marcada, la aplicación calculará métricas de gap-time cuando corresponda (p50, p5, explicado más adelante). Si está desmarcada, ignora cualquier cálculo de tiempos incluso si existen.
- **Botón "..." (Gap-time XML selector):** Este pequeño botón al lado de "Gap-time XML" abre un diálogo para elegir un archivo XML específico que contenga datos de tiempos (etiquetas `<times>` con los instantes de cada descarga) ¹². Al seleccionar el archivo, la ruta se guarda internamente y la GUI muestra un mensaje confirmando "Usando XML para gap-time: [ruta]" ⁵⁸. **Importante:** Esto **no** carga el PRPD para análisis visual, solo guarda el archivo para extraer datos de tiempos cuando se haga una comparación de criticidad.
- *Ejemplo:* Puede cargar y procesar `muestra_actual.xml` (quizá un CSV sin tiempos), pero seleccionar como gap-time un archivo completo `muestra_actual_detallado.xml` que sí tiene los tiempos de cada evento. Así, la comparación vs baseline podrá evaluar cómo han cambiado los intervalos entre descargas usando ese archivo detallado.
- **Botón "Comparar vs base":** Realiza una **comparación de la situación actual contra una línea base (baseline)** registrada previamente, arrojando indicadores de tendencia y criticidad. Esta función está orientada a análisis de mantenimiento: permite ver si un patrón ha empeorado o mejorado en el tiempo con respecto a una referencia.
- **Uso:** Primero debe existir un baseline para el archivo actual. El baseline se genera usando el botón *Reset baseline* (descrito abajo). Supongamos que procesó el archivo `equipox.csv` hace 6 meses y guardó baseline. Ahora procesa nuevamente `equipox.csv` (datos recientes). Al pulsar "Comparar vs base", la aplicación cargará los datos baseline guardados (de

`equipox_baseline.json`) y los comparará con las métricas actuales ⁵⁹ ⁶⁰. Si la comparación es posible (el baseline corresponde al mismo archivo/activo) ⁶¹, calculará:

- **Diferencia en conteo de descargas:** cuánto ha cambiado el número total de eventos (en % respecto al baseline) ⁶² ⁶³.
- **Cambio en amplitud (TEV) de pico:** la diferencia en dB del percentil 95 de amplitud actual vs baseline ⁶⁴ ⁶⁵. Este valor es similar a comparar un nivel de actividad PD – un aumento significativo en dB indica que las descargas más fuertes ahora son mucho mayores.
- **Cambio en anchura de fase:** compara el **ancho de distribución de fase** (calculado como desviación circular o rango) actual vs baseline, indicando si las descargas están más dispersas en fase que antes ⁶⁶ ⁶⁷. Un mayor ancho puede significar que el patrón se extendió, a veces señal de degradación creciente.
- **Cambio en centro de fase:** la variación en grados del centro de fase de la actividad principal ⁶⁶ ⁶⁸ (por ejemplo, si antes descargaba a 90° y ahora a 110°, hay un desplazamiento). Cambios grandes pueden sugerir un cambio de fuente o condiciones.
- **Gap time (si se proporcionó):** si la casilla Gap-time estaba activa y se seleccionó un XML, calculará **p50** (mediana) y **p5** (percentil 5%) de los intervalos entre descargas actuales, y si el baseline tenía valores comparativos, evaluará cómo han cambiado ⁶⁹ ⁷⁰. Un *p50* significativamente menor que antes, o muy por debajo de cierto umbral, indica que las descargas ocurren mucho más frecuentemente (lo cual es grave). Especialmente se revisa si $p50_{actual} < \sim 7$ ms o $p5_{actual} < \sim 3$ ms, marcando severidad roja ⁶⁹ (estos umbrales corresponden a actividad muy continua, casi arco eléctrico).
- Tras calcular todo lo anterior, el sistema asigna **banderas de alerta** por criterio: `flag_count`, `flag_tev`, `flag_ancho`, `flag_phase`, `flag_gap` con valores **verde**, **naranja** o **rojo** según la magnitud del cambio ⁶³ ⁶⁵ ⁷¹ ⁶⁹. Por ejemplo, si el conteo de eventos aumentó $\geq 50\%$, `flag_count` será rojo; si el pico de amplitud subió más de 6 dB, `flag_tev` rojo; aumento moderado 3-6 dB sería naranja, etc. Estas reglas incorporan también consideraciones combinadas (p.ej., si el pico subió pero es simétrico, aún rojo) ⁶⁵.
- Finalmente, determina una **críticidad global**: rojo si alguna bandera es roja, naranja si no hay rojas pero hay al menos una naranja, o verde si todas verdes ⁷². En base a eso sugiere una **"decisión recomendada"**: *Inspección prioritaria* (rojo), *Monitorear y re-evaluar* (naranja) u *OK* (verde) ⁷² ⁷³.
- Todos estos resultados se guardan en un archivo JSON `*_kpi_tracking.json` dentro de `out\reports` para registro histórico ⁷⁴, y se añade (append) una línea en un archivo CSV resumen (`*_summary.csv`) si existe, de modo que se puede construir un histórico de evolución ⁷⁵ ⁷⁶.
- La GUI presenta un cuadro de mensaje con un resumen de los deltas y flags ⁷⁷. Por ejemplo:

```
?Conteo: +0.5000 | ?TEV(dB): 7.2 | Asim: 0.0
Ancho act/base/?%: 45.000 / 30.000 / +0.5000 | Fase ?°: 15.000
Gap p50/p5 ms: 6.000 / 2.000
Flags: count=red, tev=red, ancho=red, phase=orange, gap=red
Críticidad: red | Decisión: Inspección prioritaria | summary.csv:
OK
```

(Esto indicaría, por ejemplo: 50% más eventos, TEV 7.2 dB arriba, ancho de fase 50% mayor, fase 15° corrida, gap times más cortos; varias alertas rojas => inspección urgente).

- Si no hay baseline previo disponible, el sistema informará que no puede comparar (o asumirá baseline vacío = sin cambio).

- Esta función es **experimental** y muy poderosa: permite implementar un programa de mantenimiento predictivo, donde cada vez que analice un activo pueda comparar con su estado base. Sin embargo, requiere que el **mismo archivo (mismo nombre)** se use para baseline y para el seguimiento, ya que la correspondencia se verifica por nombre de archivo fuente ⁶¹. (Consejo: Para seguimiento, conserve el mismo nombre de archivo de medición en sucesivas pruebas de un equipo).

- **Botón "Reset baseline"**: Sirve para **establecer o actualizar la línea base (baseline)** de un archivo de descargas. Al pulsarlo, toma los resultados actuales procesados y los guarda como referencia base para futuras comparaciones ⁷⁸ ⁷⁹. Sus efectos son:

- Genera o actualiza el archivo `<nombre>_baseline.json` en `out\reports` con las métricas actuales ⁷⁸ ⁸⁰. Este JSON incluye los KPIs principales (p95_amp, dens, R_phase, std_circ_deg, severity) y también un bloque extendido `__meta__` con metadata (nombre de archivo original, fecha de creación, filtro usado, offset de fase) ⁸¹, y un sub-bloque `kpi_ext` con métricas adicionales calculadas (total_count, tev_db que es p95 en unidades dB, ang_width_deg, phase_center_deg, y campos para gap time p50_ms/p5_ms inicialmente como `null`) ⁸² ⁸³.
- En la interfaz, internamente, "resetear baseline" también recalcula las **bandas de referencia de amplitud $\pm 20\%$** en la visualización ¹⁹ ⁸⁴. Es decir, tras establecer baseline, la GUI podría mostrar en los gráficos una banda o línea indicando el 100% del valor base y límites del 120% y 80%. Esto le ayuda visualmente a ver si las descargas actuales exceden significativamente la amplitud típica base. *(Actualmente, esta representación es muy sutil; en futuras versiones se resaltarán más claramente. La funcionalidad principal es el guardado en JSON.)*
- Use este botón **después** de procesar un archivo que considere representativo de condiciones normales o aceptables, de modo que sirva como baseline. Por ejemplo, tras instalar sensores y obtener una primera medición de PD en buen estado, procese ese archivo y haga "Reset baseline". Meses después, cuando mida de nuevo y procese, podrá usar "Comparar vs base" para ver cambios.

Importante: El baseline es **específico por archivo/activo**. Si procesa otro archivo distinto y pulsa comparar, la herramienta notará que el baseline guardado pertenece a otra fuente y lo ignorará ⁶¹ para evitar confusiones. Debe establecer baseline para cada objeto de análisis individualmente y usar el mismo nombre de archivo (o renombrar su nueva medida con el mismo nombre) para que la comparación lo detecte.

- **Botón "Ayuda/README"**: Abre este mismo documento de ayuda (README) de la aplicación. Al hacer clic, la GUI localizará el archivo `README.md` en el directorio raíz y lo abrirá con la aplicación asociada en su sistema ⁸⁵. En Windows, normalmente se abrirá en su navegador o editor de texto predeterminado; en otros sistemas se usará el navegador web por defecto para mostrarlo. De esta manera, siempre tiene a mano la guía de uso mientras opera la herramienta. Si actualiza el archivo README (por ejemplo, añadiendo instrucciones o notas propias), el botón siempre mostrará la versión más reciente, ya que simplemente lee el archivo del disco. *Consejo:* Si prefiere ver la ayuda dentro de la ventana de la aplicación, puede instalar un visor Markdown o asociar `.md` con su editor favorito.

Área central de gráficos

La ventana principal muestra **cuatro subgráficos** organizados en una matriz de 2x2. Estos paneles presentan distintas vistas de los datos y resultados, y su contenido varía según la opción elegida en el combo "Vista:" descrito más adelante.

Por defecto, tras cargar un archivo y pulsar Procesar, los gráficos se interpretan así (en el modo de vista por defecto "Probabilidades"):

- **Gráfico superior izquierdo – PRPD crudo:** Muestra la distribución bruta de descargas tal cual se leyeron del archivo, **antes** de cualquier alineación de fase o filtrado. El eje X es la Fase (0° a 360°) y el eje Y es la Amplitud (en porcentaje normalizado 0–100, o unidades relativas a la escala máxima) ⁸⁶. Aquí usted ve todos los puntos registrados. Si "Densidad" está activo, verá un mapa de calor de densidad de descargas ²⁹; si no, verá cada descarga como un punto. Este gráfico ayuda a identificar a simple vista el patrón original y la presencia de ruido: por ejemplo, *descargas verdaderas* suelen agruparse en ciertas fases (p. ej. alrededor de 90° y 270° para descargas internas), mientras que *ruido aleatorio* tiende a llenar muchos ángulos con amplitudes bajas.
- **Título:** "PRPD crudo".
- **Overlay de ruido:** En caso de tener activa la densidad y haber completado el procesamiento, es posible que se muestren puntos grises indicando cuáles eventos fueron considerados ruido y removidos tras el filtrado ³¹. En el gráfico crudo, estos serían esencialmente los mismos puntos que en filtrado se descartaron; aquí se marcan para referencia.
- **Gráfico superior derecho – PRPD alineado/filtrado:** Muestra las descargas **después de aplicar la alineación de fase seleccionada y el filtro S1/S2** elegido ⁸⁷ ⁸⁸. Esto representa el *resultado de los datos preprocesados* sobre el cual se hacen el clustering y demás cálculos. Eje X nuevamente es fase (ya desplazada según offset escogido) y eje Y amplitud (0–100). Aquí debería notar algunas diferencias respecto al crudo:
 - Las **nubes principales** de PD probablemente estén centradas en fases estándar (si auto-fase estaba activo, las verá cerca de 0°/360° o 120° o 240°, dependiendo del mejor alineamiento encontrado). El título del gráfico indica el offset aplicado, ej. "Alineado/filtrado (offset=120°)" ⁸⁸.
 - Las **descargas de ruido filtradas** ya no aparecen como puntos de color; si la opción densidad está activa, solo verá la densidad de los puntos conservados. Sin embargo, de nuevo, se superponen puntos grises para indicar dónde estaban los descartados ⁸⁹. Notará que en este gráfico normalmente las zonas de baja amplitud muy pobladas (ruido de fondo) desaparecen si aplicó S2, y que la concentración de descargas es mayor en torno a ciertas fases.
 - **Escala:** La escala vertical se fija de 0 a 100% de amplitud para permitir comparar con el crudo fácilmente ⁹⁰ (de hecho, ambos gráficos usan 0–100 en Y).
- Este gráfico es clave para entender qué está entrando al algoritmo de clustering: idealmente, tras buen filtrado, ver aquí 1 o 2 nubes bien definidas. Si aún ve muchas "nubes" pequeñas o dispersas en este panel, quizás necesite un filtrado más fuerte, o indica múltiples fuentes.
- **Gráfico inferior izquierdo – panel multifunción (Probabilidades / Histogramas / Nubes):** Por defecto (vista *Probabilidades*) en este panel se muestra un **gráfico de barras con las probabilidades de clasificación** para cada tipo de descarga parcial ⁹¹. Las clases típicas son: *cavidad interna*, *descarga superficial*, *corona*, *electrodo flotante* (y puede haber "ruido" como categoría según el modelo). Cada barra se extiende de 0 a 1 (100%) en el eje vertical. Por

ejemplo, si después de procesar el patrón el modelo (o la heurística) determina 90% de probabilidad de que sea una descarga de cavidad y 10% superficial, verá una barra de 0.9 en "cavidad" y 0.1 en "superficial", y prácticamente 0 en las demás. El título del gráfico es "Probabilidades".

- Si no se cargó ANN y la heurística interna no asigna valores, es posible que todas las barras salgan 0 o valores bajos arbitrarios. Pero generalmente, debería reflejar la mejor conjetura del sistema. Un resultado muy "concentrado" (una barra casi 1.0) indica alta confianza en esa clase; una distribución más plana (todas barras ~0.2–0.3) indica incertidumbre o mezcla de patrones.
- **Interpretación práctica:** Las clases corresponden a defectos típicos en máquinas eléctricas: *cavidad* (descargas internas en huecos del aislamiento, suelen dar dos grupos simétricos por polaridad), *superficial* (tracking en superficies, a menudo más distribuido en fase), *corona* (descargas en aire, generalmente unipolares, amplitud menor), *flotante* (electrodos flotantes, patrón similar a corona pero con ciertas características), y *ruido* (no PD, interferencia). Use estas probabilidades como apoyo, no como verdad absoluta – siempre contraste con los gráficos. Por ejemplo, si "ruido" está alto, verifique si su patrón se ve atípico (muy disperso, sin forma); a veces un modelo puede confundir un patrón muy enmascarado con ruido.
- **Gráfico inferior derecho – indicadores de histograma (vista por defecto):** En el modo por defecto, este panel no muestra contenido específico de probabilidades ni nubes (podría quedar en blanco o con texto informativo). Sin embargo, es utilizado en otros modos de vista: en la vista **Histogramas**, aquí aparecerá el **histograma de fase**, y en vista de **Nubes** se usa para mensajes si no hay nubes. Lo detallaremos en la siguiente sección ("Modos de vista"). En esencia, en el modo Probabilidades inicial, puede ignorar este panel (en versiones futuras podría contener un texto resumen de métricas, pero actualmente esas métricas se consultan vía CSV o "Comparar vs base").

Modos de vista (selector "Vista:")

Debajo de los gráficos hay un selector desplegable etiquetado "**Vista:**" ⁹², que permite alternar qué información se muestra en los paneles inferiores (e incluso modifica el contenido de los superiores en ciertos casos). Las opciones son:

- **Probabilidades:** (*Vista predeterminada*). Muestra el resultado de clasificación (barras de probabilidad) en el panel inferior izquierdo ⁹¹, como ya se describió. Es la vista enfocada en diagnosticar **qué tipo de PD** podría ser, usando la ANN o heurística. Use esta vista inmediatamente después de procesar para ver la clasificación sugerida del patrón. Mientras esté seleccionada, los gráficos superiores muestran PRPD crudo y filtrado, y el inferior derecho queda sin uso específico (podría considerarse para texto en el futuro).
- **Histogramas:** Esta vista está orientada al **análisis de distribuciones de amplitud y fase** detalladas, complementando la información del PRPD. Al seleccionar "Histogramas", la GUI reorganiza los gráficos de la siguiente manera ⁹³ ⁹⁴ :
- El **gráfico superior izquierdo** en esta vista muestra las curvas **ANGPD y N-ANGPD** calculadas a partir de los datos filtrados. Es decir, en vez de la dispersión cruda, verá dos curvas: una curva

ANGPD (normalmente de color azul) y una **N-ANGPD** (naranja) ⁹⁵. El eje X sigue siendo Fase (0–360°) y el eje Y se etiqueta como "Densidad". Aquí:

- **ANGPD (Área=1):** Es la **distribución angular de las descargas normalizada por el área** ⁹⁶. Se calcula acumulando la amplitud (peso = |amplitud|) de todas las descargas en bins de fase (por defecto 72 bins para cubrir 0–360°) ⁹⁷. Luego se normaliza para que la suma sobre todos los bins sea 1 (por eso área unitaria) ⁹⁶. Representa esencialmente qué fracción de la "energía" de descarga ocurre en cada porción de fase ⁹⁸. La curva azul sube en las fases donde hubo más actividad (ponderada por amplitud) y baja donde casi no hubo descargas de importancia.
 - **N-ANGPD (Pico=1):** Es la **distribución angular normalizada al valor pico** ⁹⁹. Se parte del mismo histograma de amplitud vs fase, pero en lugar de normalizar por el total, se divide cada valor por el máximo valor obtenido, de modo que el bin más alto queda en 1.0 (100%) ¹⁰⁰. Esto realza la forma relativa de la distribución sin importar cuánta energía total hubo; muestra el perfil de fase de las descargas (curva naranja). Por ejemplo, si hay dos picos de igual altura en azul, en naranja ambos llegarán a ~1; si uno era ligeramente menor, en naranja igualmente llegará a 1 si es el mayor, y el otro quedará algo por debajo.
 - En el gráfico, ambas curvas permiten ver **dónde** en fase se concentran las descargas y con qué forma. La escala "Densidad" es adimensional (fracción o porcentaje).
Interpretación: Si la curva azul ANGPD tiene dos picos claramente alrededor de, digamos, 90° y 270°, significa que ~ esos son los ángulos favoritos con, sumados, todo el peso (típico de defectos internos). La naranja N-ANGPD ayudará a ver si hay más de dos picos significativos (picos secundarios) porque todos sus picos se igualan en altura máxima. Si la curva azul es muy plana (sin picos, más o menos nivelada), indica que las descargas están distribuidas en muchas fases (no concentradas) ¹⁰¹.
 - **Leyenda:** El gráfico incluye una leyenda "ANGPD" (azul) y "N-ANGPD" (naranja) ⁹⁵ para distinguirlas.
 - Si no hubiera datos (por ejemplo, filtró todo o no hay eventos), se mostrará "Sin datos" en lugar de las curvas ¹⁰².
- El **gráfico superior derecho** cambia para mostrar el **histograma de amplitud por semiciclos** ¹⁰³ ¹⁰⁴. En lugar del PRPD filtrado, ahora este panel (titulado "Histograma de Amplitud (N=16)") presenta dos series de datos: **H_amp+** (en color azul) y **H_amp-** (en rojo) ¹⁰⁵. Estas corresponden al histograma de amplitud de las descargas durante el **semiciclo positivo** y el **semiciclo negativo**, respectivamente. ¿Cómo se construyen?
- Se toma el conjunto de amplitudes de las descargas alineadas/filtradas. Se separan en dos grupos: aquellas cuya fase está entre 0°–180° (semiciclo positivo, H_amp+) y 180°–360° (semiciclo negativo, H_amp-) ¹⁰⁶.
 - Para cada grupo, se crea un histograma de amplitudes con **N=16 intervalos** iguales de 0 a amplitud máxima (0–100%) ¹⁰⁶ ¹⁰⁷. Se cuenta cuántas descargas caen en cada intervalo de amplitud.
 - Antes de graficar, se aplica un escalado logarítmico: se usa $\log_{10}(1 + \text{conteo})$ en cada bin para reducir el efecto de diferencias muy grandes ¹⁰⁷. Luego, para facilitar la comparación, tanto H_amp+ como H_amp- se **normalizan dividiendo por el máximo de ambas** (es decir, se escala para que el valor más alto entre los dos histogramas sea 1) ¹⁰⁸. Esto permite ver ambas curvas superpuestas en la misma escala (0 a 1).
 - En el gráfico, el eje X muestra el **índice de ventana 1..16** (cada ventana es un rango de amplitud fijo, por ejemplo ventana 1 corresponde a amplitudes muy bajas, ventana 16 a amplitudes cercanas al máximo). El eje Y muestra H_amp (norm), es decir el valor log-

normalizado ¹⁰⁹. La curva azul representa la distribución de amplitudes en semiciclo positivo, la roja en semiciclo negativo.

- **Interpretación:** Comparar H_{amp+} vs H_{amp-} permite cuantificar la **simetría de amplitud** de las descargas positivas vs negativas ¹¹⁰. Idealmente, en ciertos defectos internos, ambas curvas se parecen (similar cantidad de descargas de cada amplitud en + y -). Si ve, por ejemplo, que la curva roja (negativa) está consistentemente por debajo de la azul, significa que en el semiciclo negativo hubo menos descargas en todos los niveles de amplitud, sugiriendo asimetría (podría ser corona unipolar, o un defecto que solo descarga en un pico). Un histograma muy concentrado hacia las ventanas altas indica la mayoría de descargas con amplitud alta (posible PD severo), mientras que concentrado hacia ventanas bajas indica muchas descargas pequeñas (posible ruido o PD incipiente). Como se usa $\log(1+N)$, incluso diferencias pequeñas se notan. Esta técnica proviene de la referencia ¹¹⁰: en PD, la simetría de amplitud +/- es un rasgo diagnóstico reconocido en normas IEC.
- El **gráfico inferior derecho** mostrará el **histograma de fase por semiciclos** ¹¹¹ ¹¹². Titulado "Histograma de Fase (N=16)", este gráfico presenta **H_{ph+}** (azul) y **H_{ph-}** (rojo). La construcción es similar a la de amplitud pero a lo largo del eje de fase:
 - Se toma el conjunto de fases de las descargas filtradas. Para las del semiciclo positivo (0–180°), se mide su distribución dentro de ese rango en 16 ventanas iguales (cada ventana ~11.25°) ¹¹¹. Para las del semiciclo negativo (180–360°), se restan 180° a las fases para traerlas al rango 0–180° y también se cuentan en 16 ventanas ¹¹¹. Así obtenemos H_{ph+} (frecuencia de descargas en cada segmento de fase del primer semiciclo) y H_{ph-} (lo mismo para el segundo semiciclo).
 - De nuevo, se aplica $\log_{10}(1+N)$ a los conteos ¹¹³ y se normaliza tomando el mayor valor entre ambas distribuciones como referencia 1 ¹¹⁴.
 - El eje X va de 1 a 16 indicando la **ventana de fase** (1 corresponde a inicio de semiciclo, 16 al final de semiciclo, i.e., 0° y 180° respectivamente). El eje Y es H_{ph} (norm) log-normalizado.
 - **Interpretación:** Estas curvas indican cómo están **repartidas las descargas dentro de cada medio ciclo**. Por ejemplo, una típica descarga interna en cavidad ocurrirá predominantemente alrededor del pico de tensión en cada polaridad (aprox. 90° en el semiciclo pos. y 270° en el neg., que aquí se mapearían a ventana ~8 para H_{ph+} y ventana ~8 para H_{ph-} porque 90° es mitad del semiciclo). Así, ambas H_{ph+} y H_{ph-} podrían tener su máximo cerca de la ventana 8. En cambio, descargas de tracking superficial suelen iniciar unas decenas de grados después del cruce por cero y extenderse hasta cerca del pico, dando una distribución más ancha. Corona puede aparecer solo en un semiciclo (por ejemplo solo H_{ph+} tiene valores, H_{ph-} casi cero) porque quizás solo con polaridad positiva hay emisión. Al comparar H_{ph+} vs H_{ph-} , se puede ver si la actividad en fase es similar en ambos sentidos o no ¹¹⁰.
 - Un histograma de fase más *angosto* (curva con pico afilado en 2-3 ventanas y casi cero en otras) significa descargas agrupadas en fase (indicador de patrón definido, e.g. interna). Uno más *ancho* (valores repartidos en muchas ventanas) implica descargas a lo largo de gran parte del semiciclo (e.g. tracking superficial, que a menudo abarca una amplia región en fase) ¹¹⁰. Esta medida complementa las curvas ANGPD de arriba pero sin ponderar por amplitud (cada descarga cuenta igual aquí, tras binarización se considera solo presencia/ausencia ¹¹⁵).
 - Un punto importante: antes de calcular estos histogramas, **todas las descargas cuentan como una unidad independientemente de su multiplicidad** (se aplica *binarización*) ¹¹⁵. Esto fue diseñado así para que ruido de muchos pulsos pequeños no domine la forma del

histograma ¹¹⁶ . Por ello, los histogramas H_ph y H_amp son sensibles a diferencias sutiles de patrón más que a volumen bruto ¹¹⁷ .

En conjunto, la vista **Histogramas** proporciona una mirada cuantitativa detallada a características del patrón PD: **simetría de amplitud** entre semiciclos, **localización en fase** de las nubes, y las curvas ANGPD que resumen la distribución global. Estas características (64 valores: 16 H_amp+/- y 16 H_ph+/- concatenados) son justamente las **features de entrada para la ANN** en el método propuesto ¹¹⁸ , debido a su poder para resaltar diferencias importantes entre tipos de PD (más sensibles que las proyecciones tradicionales) ¹¹⁷ . Como usuario, puede usar esta vista para validar la clasificación: por ejemplo, cavidades tienden a dar histogramas bastante simétricos y con picos pronunciados en ambas fases, corona podría mostrar histograma de fase con un solo pico en un semiciclo y nada en el otro, etc. ¹¹⁹ .

- **Nubes (S3), Nubes (S4), Nubes (S5):** Estas tres vistas muestran las **agrupaciones (clústeres) de descargas** identificadas en las etapas S3, S4 y S5, respectivamente. Son útiles para entender la estructura interna del patrón y cómo el algoritmo de *denoising* y clustering está segregando las descargas. Al seleccionar cualquiera de ellas, el panel inferior izquierdo se utilizará para graficar las nubes de puntos, reemplazando al gráfico de probabilidades.

Común a todas las vistas de Nubes: El gráfico mostrado es un scatter plot de Fase vs Amplitud de las descargas **alineadas y filtradas**, donde los puntos están **coloreados según su pertenencia a clúster** ¹¹⁹ . Cada clúster (nube) tiene un color distinto; los puntos considerados ruido (outliers del clustering) aparecen típicamente en gris o sin color especial. Además, los **centroides** de los clústeres pueden indicarse con un marcador destacado (por ejemplo, un círculo más grande o con borde), a veces etiquetados con un ID de clúster si hay pocos clústeres. El título del gráfico indicará qué nivel Sx se está mostrando.

- **Nubes crudas (S3):** Muestra los clústeres **originales** encontrados por el algoritmo de clustering (DBSCAN) directamente sobre los puntos filtrados ¹²⁰ . El título será "Nubes crudas (S3)". Aquí verá posiblemente varios colores distintos dispersos en el gráfico, cada conjunto representando un clúster de descargas que el algoritmo encontró densamente agrupadas. Por ejemplo, un patrón con una fuente principal puede tener 2 clústeres grandes (positivo y negativo) y quizás algunos clústeres pequeños de puntos esparcidos (que pueden ser ruido residual que DBSCAN agrupó en su propio clúster o pequeñas segundas fuentes).

- **Interpretación:** El número de clústeres S3 (`n_clusters`) da una idea de cuántas agrupaciones distinguió el algoritmo en los datos. Un valor bajo (1-3) es lo esperado si hay una sola fuente dominante; valores más altos (5, 6, 7...) indican que o bien hay múltiples patrones superpuestos (varias fuentes de PD o ruido) o que el algoritmo fragmentó una nube grande en subgrupos. En la mayoría de los casos, un `n_clusters` elevado acompañado de `has_noise = Sí` (ruido detectado) sugiere que había bastante dispersión que formó clústeres pequeños.
- En la GUI, al poner esta vista, puede identificar visualmente dónde están las nubes: ¿agrupadas cerca del pico de voltaje positivo/negativo? ¿Existe una nube muy pequeña aislada? ¿Hay nubes "en fila" a lo largo de amplitud baja (indicando ruido de pulso de baja amplitud en muchas fases)? Esta inspección le permite validar qué está considerando el algoritmo como grupos.

- **Nubes combinadas (S4):** Muestra los clústeres después de la **combinación/agrupación de S3** ¹²¹ . El algoritmo de combinación toma los centroides de S3 y une aquellos que están muy próximos o corresponden a la misma nube física. Por ejemplo, si en S3 hubo 2 clústeres muy cercanos en fase y amplitud (quizás porque la nube tenía una forma extraña y DBSCAN los

separó), en S4 se consolidarán en uno solo. El título "Nubes combinadas (S4)" sugiere que aquí cada color corresponde a un *grupo consolidado*.

- Visualmente, notará que en S4 puede haber menos colores que en S3 (porque varios clústeres originales se fusionaron). Los **centroides consolidados** también se recalculan y se muestran.
 - *Interpretación:* Esta vista refleja mejor la **cantidad de nubes físicas distintas** presentes. Idealmente, un patrón de una sola fuente PD debería tener 1 nube principal por semiciclo (pos/neg), así que quizá 2 clústeres consolidados. Si aún en S4 ve 4 o 5 nubes consolidadas, podría significar que hay más de una fuente significativa de PD en la señal, o algunas descargas extras (por ejemplo corona + superficial juntas, etc.).
 - En S4, la densidad de puntos todavía se muestra con color (cuando `color_points=True` en el código ¹²²), así que se aprecia dónde hay más densidad dentro de cada nube consolidada. Esto ayuda a ver si la nube combinada abarca un rango grande o sigue concentrada.
- **Nubes dominantes (S5):** Presenta únicamente las **nubes dominantes seleccionadas** por el algoritmo ¹²³. En esta etapa, el algoritmo elige las nubes que representan la fuente principal de PD y descarta las demás (no-dominantes) ¹²⁴ ¹²⁵. Normalmente, para un patrón de PD de una sola fuente, habrá **un par de nubes dominantes: una en semiciclo positivo y otra en negativo** (o solo una si la descarga es unipolar). El título "Nubes dominantes (S5)" encabeza este gráfico.
- Aquí probablemente verá solo 1 o 2 colores de puntos (más algún punto gris si quedó ruido suelto). Esas son las descargas consideradas relevantes. Las nubes más pequeñas habrán desaparecido.
 - *Interpretación:* S5 es el **resultado final del algoritmo de denoising** propuesto en la investigación ¹²⁴ ¹²⁵. Lo que queda deberían ser las descargas pertenecientes al patrón principal. Si S5 muestra efectivamente dos nubes limpias (una en cada semiciclo), es señal de que la clasificación puede ser más confiable y que los histogramas calculados correspondan a esa fuente principal. Si S5 por ejemplo muestra 3 nubes, puede que la herramienta haya considerado que hay dos fuentes en un semiciclo (p.ej. dos tipos de PD superpuestos en positivo) – caso complejo. En general, *usted querrá que S5 contenga la esencia del patrón*. Si nota que en S5 se perdió algo que usted cree importante (p. ej., tal vez eliminó una nube relevante de menor densidad), puede optar por revisar el filtrado S1/S2 o los parámetros de clustering. No obstante, la configuración predeterminada está calibrada para la mayoría de casos típicos (según el paper, remover nubes no dominantes mejora la clasificación global ¹²⁶ ¹²⁷).

En la GUI, al cambiar entre S3, S4, S5, se recalculan los clústeres en ese momento para asegurarse de reflejar los datos filtrados actuales ¹²⁸ ¹¹⁹. (Esto significa que si ajusta filtro y reprocesa, las nubes se actualizan; si solo cambia de vista S3->S4->S5, no necesita re-procesar, es instantáneo). Además, se fija la escala vertical 0-100 en la vista de nubes para unificar comparación ¹²⁹.

Ejemplo de uso de vistas de nubes: Supongamos que tiene un patrón ruidoso. En S3 ve 7 clústeres (muchos colores), S4 combina a 3 clústeres (quizá 2 principales y 1 pequeño), y S5 deja 2 (los 2 principales). Esto le dice que había una fuente principal (2 nubes, pos/neg) y otros grupitos de ruido que fueron eliminados. Ahora suponga otro caso: S3 da 4 clústeres, S4 combina a 4 (no pudo reducir ninguno, lo que sugiere que ya eran bastante separados), y S5 escoge 2 de ellos. Puede ser que esas otras 2 nubes eran descargas de superficie de muy baja densidad que se eliminaron, quedando por ejemplo solo las de cavidad. Estos detalles pueden ayudar a justificar por qué la ANN clasifica de tal

forma (p.ej., si se eliminaron nubes de superficie, la ANN verá un patrón más puro de cavidad y dará ese resultado).

KPIs y métricas calculadas

Además de las visualizaciones gráficas, el software calcula una serie de **métricas cuantitativas** con cada procesamiento. Muchas de estas aparecen en los archivos CSV exportados (`*_metrics.csv`) y en el JSON de baseline. A continuación describimos las principales KPIs (Indicadores Clave de Desempeño) y cómo interpretarlos:

- **predicted (clase predicha):** Es la **clase de PD estimada** por la herramienta para el patrón analizado (texto, por ejemplo `"cavidad"` o `"superficial"`). Corresponde a la categoría con mayor probabilidad en la vista de Probabilidades. En `metrics.csv` aparece como `predicted`. Úselo como identificación principal del tipo de defecto, sabiendo que proviene del modelo entrenado o heurística.
- **severity (score de severidad):** Un valor numérico (generalmente de 0 a 100, o 0 a 10) que indica la **severidad relativa del patrón**. Este puntaje se calcula en base a la magnitud de las descargas y su concentración. Por ejemplo, una posible fórmula es combinar la amplitud pico normalizada, la densidad de descargas y la extensión en fase: patrones con descargas muy intensas, muy frecuentes y extendidas en fase tienden a mayor severidad. En la implementación actual, `severity_score` se deriva de componentes internos (puede ser similar a la "criticidad" pero calibrado en continuo). Úselo para rankear qué tan preocupante es el patrón: valores altos significan PD más fuertes/dispersos (más preocupantes). Un incremento en severidad con el tiempo puede disparar atención.
- **p95_amp (amplitud P95):** El **percentil 95 de amplitud** de las descargas registradas (tras filtrado). Es aproximadamente la amplitud pico observada, ignorando el 5% de los valores más altos para evitar outliers extremos. Expresado en las unidades normalizadas (0-100, que podrían correlacionarse con mV absolutos según calibración). Este valor es importante porque representa un nivel por encima del cual muy pocas descargas alcanzan; en PD a menudo se asocia con severidad (descargas >100 mV, etc.). Aparece en el baseline y métricas. Convertido a dB ($20\log_{10}$ relación con un ref), se le llama a veces *TEV (Transiente en dB)*. De hecho, en baseline JSON se guarda `tev_db` que es $20\log_{10}(P95_actual / P95_base)$ para comparación ⁶⁴. Un p95_amp que crece en el tiempo sugiere que los pulsos están alcanzando mayor magnitud (aislamiento empeorando).
- **dens (densidad):** Métrica de **densidad de descargas**. Representa cuán poblado está el patrón. Puede definirse como número de descargas por ciclo o un proxy similar. En nuestras exportaciones, `dens` es un valor normalizado (0 a 1) calculado en la etapa de severidad breakdown. Por ejemplo, podría ser el ratio de eventos conservados vs totales, o alguna medida de densidad angular (`R_phase`, etc.). Un valor mayor indica muchos eventos (alta actividad PD), uno bajo indica pocos eventos dispersos. En baseline se guarda dens y su variación se usa para severidad global. Úselo junto con el conteo total.
- **total_count (conteo total):** Cantidad total de descargas registradas (post-filtrado). No aparece directamente en metrics.csv pero sí en baseline `kpi_ext.total_count`. Sirve para seguimiento: un aumento significativo en la cantidad de eventos PD puede indicar un empeoramiento (más puntos débiles disparando). Por ejemplo, de 1000 eventos a 2000 eventos

es un salto notable (50% más). La comparación vs base utiliza este valor (lo convierte en porcentaje delta).

- **R_phase (factor de concentración de fase):** Es la **longitud resultante de fase** calculada como $|\text{promedio vectorial de los ángulos}|$. Oscila entre 0 y 1: si las descargas están perfectamente concentradas en un ángulo, $R_{\text{phase}} \approx 1$; si están uniformemente distribuidas en 0–360°, $R_{\text{phase}} \approx 0$. En metrics.csv se entrega R_phase, y su complementario suele relacionarse con la desviación estándar circular. En baseline guardamos `std_circ_deg` (desviación circular en grados) ¹³⁰. Ambas métricas hablan de lo **compactas o dispersas en fase** que están las descargas. Un std_circ_deg bajo (y R_phase alto) suele ser bueno (descargas bien agrupadas típicas de un único origen). Si con el tiempo std_circ_deg aumenta significativamente (y R_phase cae), podría implicar que la actividad se está extendiendo a más zonas del ciclo (posible señal de empeoramiento o aparición de otra fuente).
- **phase_center_deg (centro de fase):** Ángulo medio (en grados) de las descargas principales ⁸². Indica alrededor de qué fase están ocurriendo mayoritariamente. Por ejemplo, 85° sugiere cerca del pico positivo, 170° cerca del cruce, 250–270° pico negativo, etc. En un patrón simétrico, podríamos tener un centro ~0° porque uno en 90° y otro en 270° median la vuelta completa. Pero en baseline guardamos phase_center_deg (posiblemente para un cluster dominante). Cambios en este valor (reportados en "Fase Δ" en comparar) indican desplazamiento del patrón en fase, a veces asociado a cambio de condición (ej: si una cavidad empieza a disparar más tarde en el ciclo, puede indicar mayor humedad, etc.).
- **has_noise (indicador de ruido):** Es un booleano (Sí/No) que indica si el algoritmo detectó **ruido disperso significativo**. Es determinado típicamente por el clustering: si muchos puntos quedaron sin clúster (labels = -1 en DBSCAN) o en clústeres atípicos, se marca `has_noise=True`. En el resumen de batch verá "ruido=Sí/No". Úselo para saber si debe confiar plenamente en la clasificación – si hay ruido, los resultados pueden ser menos seguros o requerir filtrado más fuerte.
- **filter_level (nivel de filtro):** En metrics, se anota qué filtro S1/S2 se aplicó (`weak`, `strong`, `stronger`). Esto es para trazar qué resultados corresponden a qué limpieza. Si compara análisis, tenga en cuenta que con `stronger` quizás se omitieron más datos (lo cual puede subir R_phase artificialmente, etc.). Generalmente use siempre el mismo filtro para comparar severidad en el tiempo.
- **phase_offset (offset de fase usado):** 0, 120 o 240 (o otro entero si manual). Para registro simplemente.

Todas estas métricas juntas brindan un cuadro cuantitativo del patrón. Por ejemplo, un **patrón grave** podría tener: severity ~ 8/10, p95_amp muy alto, dens alta, R_phase bajo (descargas por todas partes), has_noise=False (porque es genuino, no ruido), etc. Un **patrón leve**: severity ~ 2/10, p95_amp bajo, dens baja, R_phase alto (descargas concentradas), etc.

Puede consultar `out/reports/<file>_metrics.csv` después de procesar para ver los valores exactos de su caso. Y usando la función de comparar vs base, las diferencias relevantes se le resumirán como vimos.

Ejemplos de uso práctico

A continuación presentamos algunos escenarios prácticos y consejos de cómo usar las distintas funciones y cómo interpretar los resultados visuales, para que pueda aplicar esta herramienta de la manera más efectiva en el diagnóstico real:

• Ejemplo 1: Descarga interna en aislante sólido (Cavidad interna)

Situación: Usted tiene un patrón PRPD típico de descargas internas en una bobina: suelen aparecer dos grupos densos de puntos, aproximadamente simétricos en magnitud, uno durante el semiciclo positivo y otro en el negativo, alrededor del pico de tensión de cada ciclo (cerca de 90° y 270°).

Uso: Cargue el archivo y use **Fase: Auto** para alinear (así, aunque el sensor estuviera desfasado, el programa alineará esas nubes a ~90/270). Aplique **Filtro S1 (Weak)** inicialmente, ya que en este caso el patrón es relativamente claro y solo hay leve ruido de fondo. Procese.

Interpretación: En la vista *Probabilidades*, debería ver una clara predicción de "**cavidad**" con probabilidad alta (por ejemplo >0.8), ya que las características se corresponden con descargas internas. Confirme esto viendo la vista *Histogramas*: las curvas ANGPD mostrarán seguramente **dos picos** grandes, indicando distribución bimodal en fase (propio de cavidades) ¹³¹. El histograma de amplitud H_amp+ vs H_amp- probablemente muestre curvas similares (ambos semiciclos descargan con amplitudes comparables), reflejando **simetría** ¹¹⁰. El histograma de fase H_ph tendrá picos agudos cerca de la mitad del semiciclo (ventanas 8±1), mostrando que las descargas se concentran cerca del pico de voltaje en ambos semiciclos, lo cual concuerda con descargas internas típicas ¹¹⁰. En vista *Nubes S3*, es posible que vea unos 3 clústeres: dos clústeres grandes (uno en ~90° y otro ~270°) y quizá algún clúster pequeño de ruido. S4 puede combinarlos o no (si estaban ya bien separados no habrá cambio), y S5 deberá mostrar sólo esos **2 nubes dominantes** (una positiva y una negativa). Esto confirma que la fuente principal es una y simétrica. Con estos outputs, usted concluiría que el patrón corresponde a PD interno (cavidad) en el aislamiento, de severidad acorde a la amplitud (vea p95_amp: si es alto, severo; moderado, menos crítico). *Acción:* Podría establecer este resultado como baseline si es la primera medición; o si ya tenía baseline, usar comparar vs base para ver si p95_amp o count aumentaron.

• Ejemplo 2: Descarga superficial (Tracking)

Situación: Un aislador sucio produce descargas superficiales. Estas a menudo empiezan tras el arranque de la semionda (después de 0°) y pueden extenderse durante una porción significativa del semiciclo, dependiendo de la humedad. El patrón PRPD se ve como puntos que cubren quizás 30° a 150° en el semiciclo positivo, y 210° a 330° en el negativo, con amplitudes crecientes a medida que avanza la semionda (forma "lengua" o difuso). Es menos concentrado que una cavidad.

Uso: Cargar archivo, usar **Fase: Auto** (el auto-align aún pondrá el grosso en posiciones canónicas, aunque la distribución es más amplia). Para el filtrado, si observa mucho ruido aleatorio entre las dos bandas principales, podría probar **S2 (Strong)** para eliminarlo. Procese.

Interpretación: En vista *Probabilidades*, es probable que la ANN dé "**superficial**" o quizá confunda algo con corona si amplitudes son bajas. De cualquier forma, se espera que cavidad tenga baja prob, corona mediana, superficial alta, etc. Compruebe *Histogramas*: Es característico de tracking que el **histograma de fase** sea **más ancho** que en cavidad – es decir, H_ph+ y H_ph- tendrán valores significativos en muchas ventanas (no un pico único) ¹¹⁰. Puede que H_ph incluso sea casi plano o con una meseta en medio, indicando descargas a lo largo de gran parte del medio ciclo. El **histograma de amplitud** podría mostrar asimetría si, por ejemplo, en semiciclo positivo hubo más descargas que en negativo (dependiendo de polaridad del electrodo sucio). Si hay humedad, a veces las descargas conducen más en un semiciclo que en el otro; esto lo vería como H_amp+ notablemente diferente de H_amp-. ANGPD azul quizás no tenga dos picos

separados, sino una curva amplia con un pico dominante y “colas”, reflejando actividad prolongada en fase. N-ANGPD (naranja) destacará esa forma extendida. En *Nubes S3*, podría ver 2-3 clústeres: quizás uno principal en pos, uno en neg, y alguno pequeño de ruido. S4 probablemente quede en 2 clústeres consolidados (pos/neg), y S5 mostrará esos 2 dominantes. La clave es que estos clústeres abarcan mayor ancho en fase (usted lo nota en S3: los puntos de cada color ocupan una franja vertical más ancha que en el caso de cavidad). *Acción:* La clasificación superficial suele indicar contaminación o daño superficial; combine con los KPIs: si *std_circ_deg* (ancho) es muy alto comparado con baseline, y amplitudes subieron, puede ameritar limpieza/inspección.

• Ejemplo 3: Corona unipolar con ruido ambiental

Situación: Está midiendo en campo abierto y capta **corona** en un conductor (descargas corona típicamente ocurren solo en el semiciclo donde el conductor es positivo w.r.t. aire, por lo que las descargas salen solo en una mitad del ciclo). Además, hay bastante ruido de radiofrecuencia y quizá algún pulso externo esporádico. El PRPD se ve como un grupo de puntos de baja amplitud concentrados, digamos, entre 70°-110° solo en semiciclo positivo; el resto del patrón son puntos aislados dispersos (ruido).

Uso: Cargue datos. Aquí es útil aplicar **Filtro S2 Stronger** de entrada porque sabemos que hay mucho ruido (quintiles más frecuentes y amplitudes bajas son probablemente ruido). Esto quitará gran parte de los puntos dispersos. Procese.

Interpretación: En *Probabilidades*, el modelo podría dar algo como "corona" con probabilidad moderada, pero ojo: si el entrenamiento incluía corona como bipolo (muchos modelos no incluyen unipolar), quizá clasifique como "flotante" o confunda con ruido. No confíe solo en la etiqueta; mire *Histogramas*: El **histograma de fase H_{ph}** mostrará una distribución en un semiciclo y prácticamente nada en el otro. Con nuestro método de graficación, H_{ph} probablemente esté cercano a cero en todas las ventanas, y H_{ph}+ tenga un pico en ciertas ventanas. Eso es una **señal clara de actividad unipolar**. El **histograma de amplitud H_{amp}** puede mostrar más pulsos en positivo que en negativo (rojo casi cero en todos bins, azul con forma decente). ANGPD azul tendrá un pico en la región de 90° y luego cercano a cero en 270°; la curva naranja N-ANGPD igualmente marcará pico=1 en ~90° y estará en cero en la mitad opuesta. Básicamente, ANGPD/N-ANGPD *reflejarán que la mitad del ciclo no contribuye nada* (lo cual es propio de corona fase única). En *Nubes S3*, seguramente verá solo **1 clúster principal** (en semiciclo positivo) y quizás algunos puntitos dispersos no clusterizados (ruido) que DBSCAN dejó fuera. S4 y S5 probablemente resulten en solo **1 nube dominante**. Si S5 muestra 1 nube, la ANN puede tener problemas porque generalmente espera 2 nubes para una fuente – pero usted sabrá que es corona por la distribución. *Acción:* Este patrón, si es estable, quizás no sea gravísimo a nivel de aislamiento (corona externa), pero en subestaciones corona puede indicar puntos filosos que conviene suavizar. Los KPIs: severity podría ser medio (no por amplitud – suele ser baja – sino por densidad continua en un semiciclo), R_{phase} bastante alto (porque está muy concentrado en fase, solo que en un semiciclo), etc. Si baseline era sin corona y ahora aparece, flags se pondrán rojos en phase (cambio de centro) y quizá gap (si corona es muy continua, p5_{ms} bajo).

• Ejemplo 4: Múltiples fuentes (mezcla PD)

Situación: Un caso complejo: suponga que en un generador tiene **descarga interna** y también algo de **tracking superficial** ocurriendo simultáneamente. El patrón PRPD sería la superposición de ambos: se verán los dos lobulos típicos de cavidad, pero además un “velo” más amplio de puntos de menor densidad extendidos en fase. Esto confunde la clasificación automática.

Uso: Cargar datos y probar con **Filtro S1 Weak** primero, para no eliminar demasiado y ver toda la imagen. Procesar.

Interpretación: La ANN quizás dé una salida incierta (por ejemplo, prob. cavidad 0.5, superficial

0.4). En *Histogramas*, verá tal vez una mezcla: H_ph puede mostrar un pico fuerte (por la cavidad) pero también valores en otras ventanas elevadas (por el tracking). ANGPD azul puede no ser tan limpio de dos picos, sino un pico grande y una meseta. *Nubes S3* aquí es revelador: posiblemente verá **4 clústeres** en S3 (dos densos = cavidad pos/neg, y dos difusos = tracking pos/neg). En S4, el algoritmo podría combinarlos si se solapan, pero si son distintos en fase, quedarán 4. S5 luego tendrá que elegir los dominantes: típicamente seleccionaría las nubes de mayor densidad (probablemente las cavidad) y descartaría las de tracking (menos densas) ¹²⁶. Así S5 mostraría 2 nubes (cavidad). Esto significa que la herramienta enfocó la fuente principal (cavidad) y “quitó” la superficial para fines de caracterización principal. Esto se reflejará en la clasificación final inclinándose a cavidad.

Si usted sospecha esta múltiple fuente, puede mirar S5 vs S3: S3 mostraba 4, S5 dejó 2 – las otras 2 eran superficiales. Para confirmarlo manualmente, podría filtrar de otra manera: por ejemplo, procese nuevamente con **Filtro S2 Stronger** – eso probablemente eliminará las descargas más débiles (posiblemente las superficiales) y dejar solo las cavidad. Entonces la ANN sin duda daría cavidad. Alternativamente, puede procesar en otra pasada con filtro más suave S1 y examinar los clústeres.

Acción: En un caso así, vale la pena hacer análisis separado: trate de identificar cada patrón por separado (puede ser necesario exportar los CSV y filtrar por clúster ID fuera de la herramienta). La GUI en su estado actual prioriza la fuente dominante, pero su ojo experto, apoyado por las gráficas S3/S4, puede detectar la secundaria. Esto le indicaría que potencialmente tiene dos defectos distintos co-ocurriendo. Para mantenimiento, se tendría que planificar abordar ambos (por ejemplo, hay corona o tracking además de una cavidad principal).

• Ejemplo 5: Seguimiento en el tiempo

Situación: Usted midió un transformador al ponerlo en servicio (baseline) y luego 6 meses después. En el baseline las métricas eran: p95_amp 20, total_count 500, R_phase 0.8 (std_circ ~40°), etc. Se establece baseline. Ahora, 6 meses después, procesa de nuevo con las mismas condiciones. Obtiene p95_amp 28, total_count 800, R_phase 0.6 (std_circ ~60°).

Uso: Tras procesar, presione **Comparar vs base** (habiendo marcado Gap-time si también tenía esa info, por ejemplo).

Interpretación: El mensaje de salida podría indicar: Conteo +60% (flag count naranja/rojo), TEV +3 dB (flag tev naranja), ancho fase +50% (flag ancho roja), shift fase 5° (flag phase verde quizás), gap time sin grandes cambios (flag gap verde). Criticidad global saldría roja (por el ancho). Decisión: “Inspección prioritaria”. Esto le alertaría de un empeoramiento significativo: más descargas y más dispersas en fase (posible propagación del defecto). Aunque el incremento de amplitud fue moderado (3 dB), la combinación es preocupante. Este resultado sugiere que debe programarse una parada para inspección interna del activo antes de que falle.

Acción: Gracias a la comparación, usted puede justificar intervenciones. Si los flags hubiesen sido todos verdes (cambios menores), podría decidir que el PD está estable y re-medirse en otros 6 meses. Este enfoque de *trend* es uno de los valores añadidos de la herramienta: mantener un registro numérico de la evolución PD, más allá de la inspección visual de patrones.

En general, siempre **corrobore la información de múltiples vistas**: por ejemplo, no tome la clasificación ANN ciegamente, sino verifique que las características (histogramas, ANGPD) la respalden. La experiencia humana más la consistencia de varios indicadores da la confianza final en el diagnóstico.

Créditos y autoría

Esta herramienta GUI ha sido desarrollada por **Hugo Emilio Tenorio Caballero** en colaboración con **Franco Cruz Martínez**. La metodología implementada está inspirada en el trabajo de investigación de Araújo *et al.* (2021) ⁸, quienes propusieron un algoritmo de denoising de patrones PRPD y extracción

de features por histogramas que mejoró la clasificación de descargas parciales en generadores hidroeléctricos ¹⁰ . Hemos adaptado y expandido sus ideas en esta aplicación práctica.

Agradecimientos especiales a los grupos de investigación y laboratorios que generaron los datos de PD para entrenamiento y validación, y a los estándares IEC 60270 / IEEE 1434 que sirvieron de guía para interpretar los patrones.

Derechos: Este software se distribuye con fines académicos y de mejora del mantenimiento en la industria eléctrica. Puede utilizarlo y modificarlo libremente, mencionando la procedencia. Para cualquier duda o sugerencia, contacte a los autores.

Notas finales sobre funcionalidades en desarrollo

La PRPD GUI Unificada se encuentra en evolución constante. Algunas funciones presentes en la interfaz están **en desarrollo o experimentales**, por ejemplo:

- **Gap-time tracking:** La integración del análisis de *gap time* (tiempos entre descargas) está en proceso. En la versión actual, puede cargar un XML con tiempos y se calcularán los percentiles p50 y p5 para comparativas, pero la herramienta aún no muestra un gráfico de distribución de tiempos ni utiliza gap time durante el procesamiento principal (solo en la comparación baseline). En futuras versiones, planeamos incorporar una vista de histograma de tiempos entre descargas y quizá utilizar esa información para ajustar la clasificación de severidad (p. ej., reconocer actividad tipo “arco” si los tiempos son muy cortos). Por ahora, **interprete manualmente** los valores p50/p5 que se den en la comparación: p50 bajo (por ej <10 ms) significa descargas muy frecuentes, lo cual es crítico.
- **Baseline comparativo multi-archivo:** Actualmente, la comparación vs baseline funciona mejor cuando el nuevo archivo tiene el mismo nombre base que el baseline (indicando mismo activo). Si sus archivos tienen nombres diferentes para la misma unidad en distintas fechas, es posible engañar al sistema renombrando o editando el JSON meta, pero eso no es amigable. Se está estudiando permitir cargar un baseline específico manualmente para comparar contra el archivo actual (independiente del nombre). Por ahora, siga la convención de nombres constantes para facilitar el tracking.
- **Reporte PDF mejorado:** La generación de PDF actualmente exporta la figura principal y algunos datos básicos. En versiones futuras se pretende incluir todos los gráficos (ANGPD, histogramas, etc.) y quizás interpretación automática en el PDF. Si necesita un informe detallado ahora, puede complementar el PDF con los gráficos individuales exportados (PNG) y los comentarios de este README.
- **Interfaz gráfica:** Podrían añadirse **gráficos adicionales** como la vista de gap time mencionada, o un panel de métricas numéricas en la GUI (por ejemplo, que en el cuarto cuadrante se muestren los valores de KPIs directamente). Estamos evaluando qué información adicional en pantalla beneficia al usuario sin sobrecargar la vista.
- **Optimización y ML:** El rendimiento con archivos muy grandes es bueno, pero siempre se puede optimizar. Asimismo, la ANN por defecto es simple; se planea permitir integraciones con otros modelos (ej. SVM, Random Forest) de manera plug-and-play en el futuro. Igualmente, la identificación automática de múltiples fuentes PD (clusters dominantes múltiples) es un campo

abierto; se podría incluir un aviso si S5 suprime nubes que aún tenían tamaño considerable, advirtiendo posible mezcla de fuentes.

Agradecemos cualquier retroalimentación de los usuarios para seguir puliendo estas funciones en desarrollo.

Actualización del README dentro de la GUI

Mantener la documentación actualizada es vital. Esta aplicación facilita eso: el botón **Ayuda/README** carga directamente el archivo `README.md` desde la carpeta raíz ⁸⁵. Por lo tanto, **para actualizar la ayuda que ven los usuarios**, solo tiene que **editar este archivo README.md**. Por ejemplo, si agrega nuevas características en el código, documente su uso en este README y guárdelo. La próxima vez que usted u otro usuario haga clic en Ayuda en la GUI, verá inmediatamente los cambios reflejados (no requiere recompilar la aplicación).

Recomendamos conservar una copia en texto plano (`README.txt`) o exportar a PDF si lo desea) junto a este Markdown, en caso que el sistema donde se abra no renderice formato Markdown. En nuestro caso, la ventana lo abrirá probablemente en un navegador con soporte MD, pero nunca está de más prever.

En resumen, **la guía de usuario está ligada al propio archivo README.md del proyecto**. Por favor, asegúrese de actualizarla con cada cambio de funcionalidad. Un buen README es tan importante como un buen código: ayuda a que la herramienta sea utilizada correctamente y aprovechada al máximo.

Fin del README. Si ha leído hasta aquí, ahora conoce en detalle cómo usar la PRPD GUI Unificada para analizar descargas parciales. ¡Que le resulte útil en sus diagnósticos y contribuya a un mantenimiento más inteligente y proactivo! ¹

¹ ² ⁴ ⁵ ⁶ ⁷ ¹¹ ¹⁴ ¹⁵ ¹⁶ ¹⁷ ¹⁸ ¹⁹ ⁴⁰ ⁸⁴ READMEp.txt

file://file_00000000839471f7ad86fb9fb01c2df3

³ ⁸ ⁹ ¹⁰ ¹¹⁰ ¹¹⁵ ¹¹⁶ ¹¹⁷ ¹¹⁸ ¹²⁴ ¹²⁵ ¹²⁶ ¹²⁷ energies-14-03267[1].pdf

file://file_00000000ce7c71f7a009c56c67f2b3b9

¹² ¹³ ²¹ ²⁹ ³⁰ ³¹ ³² ³³ ³⁴ ³⁵ ³⁶ ³⁷ ³⁸ ³⁹ ⁴¹ ⁴² ⁴³ ⁴⁴ ⁴⁵ ⁴⁶ ⁴⁷ ⁴⁸ ⁴⁹ ⁵⁰ ⁵¹ ⁵² ⁵³ ⁵⁴ ⁵⁵
⁵⁶ ⁵⁷ ⁵⁸ ⁵⁹ ⁶⁰ ⁶¹ ⁶² ⁶³ ⁶⁴ ⁶⁵ ⁶⁶ ⁶⁷ ⁶⁸ ⁶⁹ ⁷⁰ ⁷¹ ⁷² ⁷³ ⁷⁴ ⁷⁵ ⁷⁶ ⁷⁷ ⁷⁸ ⁷⁹ ⁸⁰ ⁸¹ ⁸² ⁸³ ⁸⁵
⁸⁶ ⁸⁷ ⁸⁸ ⁸⁹ ⁹⁰ ⁹¹ ⁹² ⁹³ ⁹⁴ ⁹⁵ ¹⁰² ¹⁰³ ¹⁰⁴ ¹⁰⁵ ¹⁰⁶ ¹⁰⁷ ¹⁰⁸ ¹⁰⁹ ¹¹¹ ¹¹² ¹¹³ ¹¹⁴ ¹¹⁹ ¹²² ¹²⁸ ¹²⁹ ¹³⁰

main (6).txt

file://file_00000000922471f582824d350f8ff38c

²⁰ ²⁵ ²⁶ ²⁸ ¹²⁰ ¹²¹ ¹²³ PIPELINE_FLOW.txt

file://file_00000000e1fc71f79692080c16ed6581

²² ²³ ²⁴ ²⁷ FILTERS_SPEC.txt

file://file_000000002be471f591b207325a2693b0

⁹⁶ ⁹⁷ ⁹⁸ ⁹⁹ ¹⁰⁰ ¹⁰¹ ¹³¹ ANGPD_NOTES.txt

file://file_0000000072b471f5a159d2b6a1f46c60