

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingeniería

**Análisis de la radiación secundaria en
controles de calidad de protonterapia
utilizando el detector Timepix3**

Alex Paul Álvarez Cayamcela

Física

Trabajo de titulación presentado como requisito
para la obtención del título de

Licenciado en Física

22 de septiembre de 2025

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingeniería

**HOJA DE CALIFICACIÓN DE TRABAJO DE FIN DE
CARRERA**

Alex Paul Álvarez Cayamcela

Nombre del profesor, Título académico:

Carlos Granja, PhD

Pedro Ducos, PhD

22 de septiembre de 2025

© Derechos de Autor

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: Alex Paul Álvarez Cayamcela

Código: 00208970

Cédula de Identidad: 0350120515

Lugar y fecha: 22 de septiembre de 2025

ACLARACIÓN PARA LA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETheses>

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETheses>

Agradecimientos

Expreso mi profunda gratitud a la Universidad San Francisco de Quito , por brindarme una formación de excelencia. Agradezco de manera especial a mi director de tesis, Dr. Carlos Granja, por su valiosa orientación y constante apoyo durante el desarrollo de este trabajo. Su guía no solo enriqueció este proyecto, sino que también fortaleció mi formación como físico. Extiendo asimismo mi gratitud a los profesores de la carrera de Física, quienes a lo largo de estos años compartieron sus conocimientos con dedicación y entusiasmo.

En el ambito personal, quiero agradecer profundamente a mis padres, Genaro y Janeth, y a mis abuelos, Martha y Luis, por su apoyo incondicional, paciencia y confianza en mí, así como por su motivación constante y por recordarme la importancia de mantener el equilibrio en los momentos más difíciles. Finalmente, dedico este logro a todas las personas que creyeron en mí y me acompañaron en este camino. Sin su apoyo, esta meta no habría sido posible.

Resumen

El análisis de la radiación secundaria producida durante las pruebas diarias de control de calidad en protonterapia utilizando el detector Minipix Timepix3 proporcionó información relevante sobre las propiedades del haz de manera no invasiva. El objetivo de este estudio fue evaluar la capacidad del detector para identificar diferencias entre las pruebas de control de calidad. Se analizaron las distribuciones temporales de la tasa de dosis correspondientes a cada prueba, aplicando posteriormente un análisis de clustering y evaluaciones estadísticas. Los resultados revelaron la existencia de dos distribuciones distintas en los campos B y C, con variaciones tanto en la dosis total como en los tipos de partículas que más contribuyeron a la misma. Estos hallazgos indican que el detector tiene el potencial de ser utilizado como una herramienta no invasiva para verificar las propiedades del haz en los procedimientos de control de calidad.

Palabras clave: *Protonterapia, Control de calidad, Radiación secundaria, Minipix Timepix3, Distribuciones de tasa de dosis, Análisis de clustering.*

Abstract

The analysis of secondary radiation produced during daily quality assurance test in proton therapy using the Minipix Timepix3 detector provided relevant information on beam properties in a non-invasive manner. The aim of this study was to evaluate the detector's ability to identify differences across QA tests by processing raw measurement data collected at the Proton Therapy Center in Prague. Temporal dose rate distributions from each test were analyzed, followed by clustering and statistical evaluation. The results revealed two distinct distributions in field B and C, with variations in both total dose and the particle types contributing most to it. These findings indicate that the detector has the potential to be used as a non-invasive tool for verifying beam properties in QA procedures.

Keywords: *Proton therapy, Quality assurance, Secondary radiation, Minipix Timepix3, Dose rate distributions, Clustering analysis.*

Índice general

1. Introducción	15
1.1. Estado del arte	17
1.2. Objetivos	19
1.2.1. Objetivo general	19
1.2.2. Objetivos secundarios	20
2. Métodos	21
2.1. Técnica basada en detectores no invasivos y miniaturizados de dis- positivo único	21
2.2. Instrumentación basada en detectores px/TPX y cámaras para de- tección de radiación	22
2.3. Configuración geométrica del montaje experimental	24

2.3.1.	Caracterización composicional y espectral de alta resolución y amplio rango de campos terapéuticos de radiación comple- ja/mixta	26
2.4.	Procesamiento de datos	27
2.4.1.	Selección de datos de los controles de calidad	27
2.4.2.	Flujo de Partículas, Tasa de Dosis, Distribuciones Tempora- les y Espectros de Energía Depositada	30
2.5.	Procesamiento: análisis dosimétrico	31
2.5.1.	Distribución temporal de tasas de dosis	31
2.5.2.	Campos de QA en el PTC: macrocampos A, B, C, D (pro- tones, electrones) y rayos x (diagnóstico)	31
2.5.3.	Microcampos de cada macrocampos: Agrupamiento (Clus- tering)	32
3.	Resultados	40
3.1.	Preprocesamiento de Datos (Física Nuclear/Detectores)	40
3.1.1.	Procesamiento de Datos Crudos del Detector	40
3.1.2.	Flujo de partículas, tasa de dosis, distribuciones temporales y espectros de energía depositada	41

	9
3.2. Procesamiento: Análisis Dosimétrico/Médico/QA	41
3.2.1. Distribución temporal de flujos de partículas y tasas de dosis	41
3.2.2. Campos de QA en el PTC: Macrocampos A, B, C, D (pro- tones, electrones) y rayos x (diagnóstico)	43
3.2.3. Microcampos de Cada Macrocampos: Agrupamiento (Clus- tering)	43
4. Conclusiones	58
Bibliografía	61
A. Tablas adicionales	70
B. Figuras	79
B.1. Distribuciones diarias de tasas de dosis por campo y clusters corres- pondientes	79
B.1.1. Campo A	79
B.1.2. Campo B	81
B.1.3. Campo C	83

	10
C. Código en python	87
C.1. Función de clustering con DTW	87

Índice de cuadros

2.1. Clasificación de partículas detectadas según su transferencia de energía. Adaptado de [1].	28
A.2. Asignación de días a clusters para el Campo A	70
A.3. Asignación de días a clusters para el Campo B	72
A.1. Ejemplo de registros generados por DPE/TraX en formato <code>.csv</code> . Se muestran variables de tiempo, conteo de clústeres (<code>clu</code>), flujo de partículas (<code>flux</code>), tasa de dosis (<code>DR</code>), dosis acumulada (<code>D</code>) y tiempo UNIX.	75
A.4. Asignación de días a clusters para el Campo C	76

Índice de figuras

2.1. Visualización de la radiación dispersa producida por un haz de protones de 100 MeV, medida con el detector Timepix3 con sensor de silicio de 300 μm . Datos integrados en un intervalo de irradiación de 1.03 s. Se muestran todos los píxeles del detector. Reimpresa con permiso de C. Granja.	24
2.2. Sala de tratamiento con haz fijo en el Proton Therapy Center, Praga. El campo de radiación dispersa se midió de manera no invasiva con una cámara MiniPIX-Timepix, ubicada en el techo de la sala a 5.7 m del isocentro. La imagen muestra la perspectiva general de la sala, con un recuadro que presenta la vista desde la parte posterior. Tomada de [1].	25

2.3. Esquema de la disposición experimental en la sala de tratamiento con haz fijo del Proton Therapy Center, Praga. Se muestra la posición del detector MiniPIX-Timepix en el techo, a 5.7 m del isocentro, para la medición no invasiva del campo de radiación dispersa. Tomada de [1].	26
2.4. Resultado de la etapa de procesamiento de los datos brutos mediante el motor de procesamiento de datos DPE/TraX. Reimpresión con permiso de C. Granja.	29
2.5. Ilustración de los campos de control de calidad (A, B, C, D). El eje Y corresponde a la intensidad del haz medida (flujo/tasa de dosis) y el eje X al tiempo transcurrido durante la adquisición.	33
3.1. Distribución temporal de la tasa de dosis de radiación correspondiente al control de calidad del día 16 de febrero, 2024 (ver etiquetas dentro del gráfico).	42
3.2. Centroides obtenidos a partir del proceso de clustering en el campo A. Se agruparon 15 días en el cluster 0, 6 en el cluster 1 y 22 en el cluster 2.	44
3.3. Centroides del campo B	46
3.4. Centroides obtenidos a partir del proceso de clustering en el campo C. Se agruparon 10 días en el cluster 0, 22 días en el cluster 1 y 11 días en el cluster 2.	53

B.1. Tasa de dosis, Cluster 0, Campo A	80
B.2. Tasa de dosis, Cluster 1, Campo A	80
B.3. Tasa de dosis, Cluster 2, Campo A	81
B.4. Tasa de dosis, Cluster 0, Campo B	81
B.5. Tasa de dosis, Cluster 1, Campo B	82
B.6. Tasa de dosis, Cluster 2, Campo B	83
B.7. Tasa de dosis, Cluster 3, Campo B	83
B.8. Tasa de dosis, Cluster 0, Campo C	84
B.9. Tasa de dosis, Cluster 1, Campo C	85
B.10. Tasa de dosis, Cluster 2, Campo C	86

Capítulo 1

Introducción

El desarrollo de la física moderna durante el siglo 20 han sido implementados ampliamente para mejorar la calidad de vida en las personas. Una contribución importante ha sido en el tratamiento contra el cáncer con el surgimiento de la radioterapia. Este tratamiento se utiliza especialmente en etapas avanzadas de la enfermedad debido a su costo-beneficio. Estudios que consideraron cánceres de notificación sanitaria obligatoria como tumores del sistema nervioso central concluyeron que el 52.2% de todos los pacientes con cáncer deben usar radioterapia [2]. Desafortunadamente en países con ingresos bajos y medios la radioterapia se encuentra disponible en aquellos que cuenta con grandes poblaciones [3], lo que representa una brecha considerable que afecta la calidad de vida de las personas diagnosticadas con cáncer. Conforme a los datos del Directorio de Centros de Radioterapia de la IAEA (DIRAC), Ecuador tiene 2.22 unidades de radioterapia por millón de habitantes: 12 centros de radioterapia, 35 equipos de radioterapia de

megavoltios (MV), y 5 centros de braquiterapia, y una población de 18.06 millones [4]. De modo que el país no cumple con el las 4 unidades de radioterapia recomendadas, aunque 2.22 está por encima de el número mínimo recomendado, que es 2.5[5]. Este es un ejemplo de un país que debe invertir en centros de radioterapia nuevos siguiendo las recomendaciones internacionales, que contempla el mantenimiento de equipos viejos y la opción de técnicas de radioterapia nuevas y modernas.

En la actualidad la radioterapia con protones ha empezado a tener mayor protagonismo debido a la distribución de dosis en profundidad que presenta en el pico de Bragg, posición donde se deposita la mayor cantidad de energía; es decir, obteniendo una buena definición de profundidad [6]. Esta característica que presentan los protones permite tener mayor control en los órganos sanos alrededor del tumor. Por lo que este tratamiento es útil para tratar tumores que se encuentran cerca de estructuras delicadas, donde es crucial minimizar el daño en los órganos de riesgo. En los últimos años, la protonterapia se empezó a utilizar para tratamientos de tumores pediátricos debido a que los niños tienden a presentar mayor sensibilidad a la radiación, que puede llegar a producir enfermedades secundarias considerables. Las estructuras críticas donde se utiliza esta terapia son el cerebro, cráneo, cuello, ojos, o la espinal. La radioterapia con fotones falla en estas estructuras a causa del desbalance entre la toxicidad y el control tumoral que se puede llegar a obtener, lo que se soluciona mayoritariamente con los protones, que reducen las complicaciones a largo plazo [7].

El sistema de planificación de tratamiento es crucial al momento de usar las ca-

racterísticas físicas de los haces de protones en protocolos clínicos. Estos sistemas se encargan de calcular la distribución espacial de dosis en el paciente, considerando la heterogeneidad de los tejidos del cuerpo, la geometría del tumor, y los órganos críticos alrededor del tumor [8]. Entonces los sistemas de planificación permiten que equipos de oncólogos y físicos optimicen la configuración de los haces y las energías que conformarán la región de alta dosis para el tumor minimizando la exposición de los órganos sanos a la radiación. Desafortunadamente, la planificación de tratamientos en protonterapia es un reto debido a la incertidumbre en la predicción del alcance, en la colocación del paciente, y los efectos biológicos asociados con las variaciones del coeficiente de transferencia lineal de energía (LET, linear energy transfer). Por estas razones es necesario un sistema preciso de control de calidad y la investigación en detectores avanzados capaces de mejorar la verificación dosimétrica [9].

1.1. Estado del arte

Los controles de calidad son esenciales en la radioterapia. Estos procedimientos permiten verificar el funcionamiento, ya sea de los aceleradores lineales en el caso de fotónterapia, o de los ciclotrones en el caso de la protónterapia. Los controles de calidad se los realiza en pruebas estandarizadas diarias, mensuales y anuales para chequear distintos parámetros como dosis de campo de radiación, salida en función del ángulo del gantry, la coincidencia de los campos de protones o rayos x, entre otros. También es necesario realizar controles de calidad específicos para cada paciente para garantizar que las dosis planificadas por el sistema de planifi-

cación del tratamiento coincidan con las dosis realmente entregadas, para ello se consideran medidas de verificación de las unidades monitor (MU), mediciones de dosis puntuales en fantomas, mediciones de dosis en profundidad en fantomas sólidos, las cuales tienen distintas tolerancias [10]. Para ello se utilizan instrumentos de medición de dosis como cámaras de ionización, arreglos de estas, o dosimetría basada en película, [11, 12, 13, 9]. Desafortunadamente, con estos instrumentos no es posible calcular directamente el espectro de transferencia lineal de energía (LET, linear energy transfer) debido a que esto requiere de detección de partículas individuales y modelos complejos [14]. La transferencia lineal de energía es importante debido a su relación directa con la densidad de ionización y al daño inducido en el ADN por la acción de las partículas [15].

Sin embargo, los detectores de píxeles, especialmente el Timepix, han abierto nuevas posibilidades dentro del área de física médica, al ser instrumentos no invasivos y de tamaño considerablemente menor en comparación con los detectores tradicionales de radiación [16]. Novak, et al., evaluaron la capacidad de espectrometría de partículas cargadas, especialmente protones, del MiniPIX Timepix3 demostrando una respuesta homogénea del sensor, buena resolución energética y capacidad para detectar protones incluso en condiciones de alta energía y ángulos extremos, validando su potencial para aplicaciones en radioterapia y detección de neutrones [17]. El detector también se utilizó para medir experimentalmente el espectro de transferencia lineal de energía de protones individuales en terapia de haces de protones demostrando que los valores medidos de LET y LET promedio por dosis (LET_d) coinciden con simulaciones de Monte Carlo, con una diferencia del 17 %, validando su potencial para mejorar la planificación clínica en protonterapia

[15]. En otro estudio, el detector se usó par caracterizar la radiación dispersada de un haz de protones tipo lápiz en rangos de dosis convensionales y ultra-altos (UHDR), mostrando su utilidad para dosimetría en radioterapia Flash [18]. Por otro lado, Stasika-Dudek et al., [19] realizaron una calibración del Timepix3 para usarlo en la detección de partículas con un alto LET considerando futuras aplicaciones en radioterapia donde normalmente se presentan campos de radiación mixtos. Granja et al. [1], utilizaron el detector Timepix3 junto con la cámara de radiación Minipix para caracterizar la radiación secundaria producida durante los controles de calidad en el Centro de terapia con protones (PTC, proton therapy center) de Praga, República Checa, obteniendo información de interés como distribuciones de tasa de dosis y flujo para las distintas partículas presentes en la radiación secundaria, con una resolución temporal de milisegundos.

Este estudio se centra en el análisis de los datos que se obtienen del detector Timepix3 de la radiación secundaria producida en los controles de calidad del PTC, siguiendo el estudio de Granja et al. [1], como se presenta en los siguientes objetivos.

1.2. Objetivos

1.2.1. Objetivo general

Evaluar la radiación secundaria producida durante los controles de calidad diarios en el PTC mediante el detector Minipix Timepix3, para evaluar la capacidad

del detector de identificar posibles diferencias entre los haces de QA diarios.

1.2.2. Objetivos secundarios

1. Depurar los datos crudos provenientes del detector para los días analizados, asegurando su correspondencia con los controles de calidad realizados.
2. Procesar los datos mediante el motor de procesamiento de datos DPE/TraX para obtener las distribuciones temporales de tasa de dosis y flujo de partículas.
3. Identificar y segmentar las distintas etapas del control de calidad en cada archivo diario procesado.
4. Aplicar un método de agrupamiento sobre las distribuciones temporales de tasa de dosis , empleando clustering con la función de distancia Dynamic Time Warping, para clasificar posibles patrones de distribución.
5. Evaluar estadísticamente las diferencias entre los días dentro de cada grupo y entre los grupos identificados.

Capítulo 2

Métodos

2.1. Técnica basada en detectores no invasivos y miniaturizados de dispositivo único

Los detectores no invasivos no interfieren con el haz terapéutico de radiación. De este modo el haz terapéutico no presenta interacciones indeseadas que pueden llegar a modificar la posición del pico de Bragg, afectando negativamente a los órganos de riesgo (órganos sanos alrededor del tumor), o que puedan modificar la dosis calculada [18]. En comparación con los detectores como cámaras de ionización o una matriz de estos, estos detectores son de dimensiones relativamente pequeños, permitiendo que con un solo dispositivo se pueda obtener información relevante a partir de la radiación secundaria, lo que resulta ventajoso al no requerir montajes complejos.

Los detectores de pixeles modernos necesitan tiempos de exposición corto para registrar tres grados de libertad de las partículas (especies, energías y direcciones) al momento de detectar y discriminar entre partículas, con una cobertura angular de hasta 4π [20], obteniendo una eficiencia cercana al 100 % para las partículas pesadas cargadas como los protones, e información detallada de los campos de rayos x [21]. Estas características hacen que estos detectores sean de gran utilidad para los controles de calidad en radioterapia con partículas.

Estos dispositivos utilizan distintos materiales para la capa del sensor en función de su aplicación, entre ellos silicio (Si), arseniuro de galio (GaAs), o telururo de cadmio (CdTe) [22]. La radiación ionizante incidente interactúa con la capa del sensor generando pares electrón-hueco, que inducen una señal en el amplificador de cada pixel [23], los cuales funcionan de manera independiente, en conjunto con su propio contador y discriminador. La información obtenida de este proceso incluye la energía depositada, la posición espacial, el tiempo de llegada y el conteo de partículas [24].

2.2. Instrumentación basada en detectores px/TPX y cámaras para detección de radiación

En este estudio se utilizó el detector Timepix3 en conjunto con la cámara de radiación Minipix, equipada con un sensor de Silicio (Si) de $300\text{ }\mu\text{m}$ de espesor [1]. El chip ASIC del timepix3 cuenta con una matriz de 256×256 píxeles, cada uno con la capacidad de registrar tanto el tiempo de llegada (ToA, time of arrival)

coo el tiempo sobre umbral (ToT, time over threshold) de las partículas incidentes. Dado que una sola partícula puede activar múltiples píxeles por encima del umbral, es necesario calibrar el conversor analógico-digital pixel por pixel para aplicaciones específicas en protonterapia [19].

Para la detección de partículas de bajo LET, se configuró el umbral de energía en 3 keV, el valor más bajo posible. Con estas calibraciones se alcanza una eficiencia cercana al 100 % en la detección de partículas cargadas y una eficiencia considerable en la detección de fotones [1]. En el modo data-driven, el detector alcanza un rendimiento de aproximadamente 40 millones de conteos por segundo por centímetro cuadrado, con la capacidad adicional de detectar partículas incidentes desde casi cualquier dirección, siempre que logren incidir en la superficie sensible del sensor [25].

Los datos brutos obtenidos consisten en los valores de ToA y ToT de cada píxel activado. Posteriormente, los píxeles vecinos que se activan dentro de una ventana temporal definida se agrupan para reconstruir la trayectoria de cada partícula. Con esta información es posible obtener la distribución espacial bidimensional de las partículas detectadas (Figura 2.1), que constituyen la base de los análisis posteriores [26].

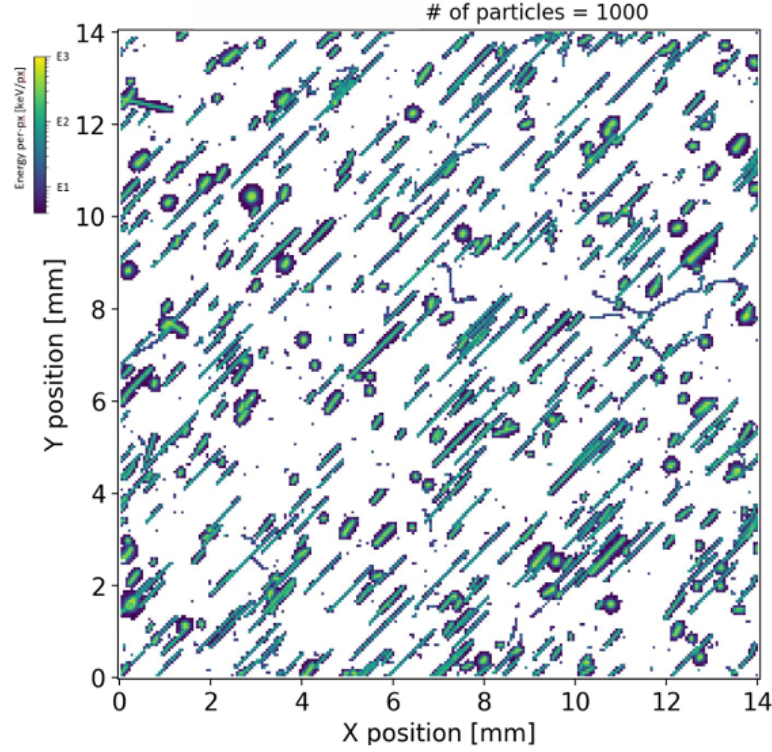


Figura 2.1: Visualización de la radiación dispersa producida por un haz de protones de 100 MeV, medida con el detector Timepix3 con sensor de silicio de 300 μm . Datos integrados en un intervalo de irradiación de 1.03 s. Se muestran todos los píxeles del detector. Reimpresa con permiso de C. Granja.

2.3. Configuración geométrica del montaje experimental

Los datos analizados en este estudio fueron obtenidos de los controles de calidad diarios realizados en el PTC de Praga, República Checa. Estos controles consisten en tres campos de haz de protones, producidos por el ciclotrón IBA Medical. La radiación secundaria generada por estos haces es registrada por el detector Minipix Timepix3. El detector se colocó paralelo al suelo, a una distancia de 572.0

cm del isocentro, y a un ángulo de 12.8° respecto a la horizontal en el plano de referencia. Además, el detector se rotó 45° con respecto al vector de trayectoria de las partículas dispersadas. Por su parte, el isocentro se encuentra a 37.5 cm del gantry (Figura 2.2, Figura 2.3) [1]. Esta configuración permite capturar la distribución espacial de las partículas dispersadas y analizar la información energética y direccional obtenida por el detector en condiciones representativas de los controles de calidad diarios.

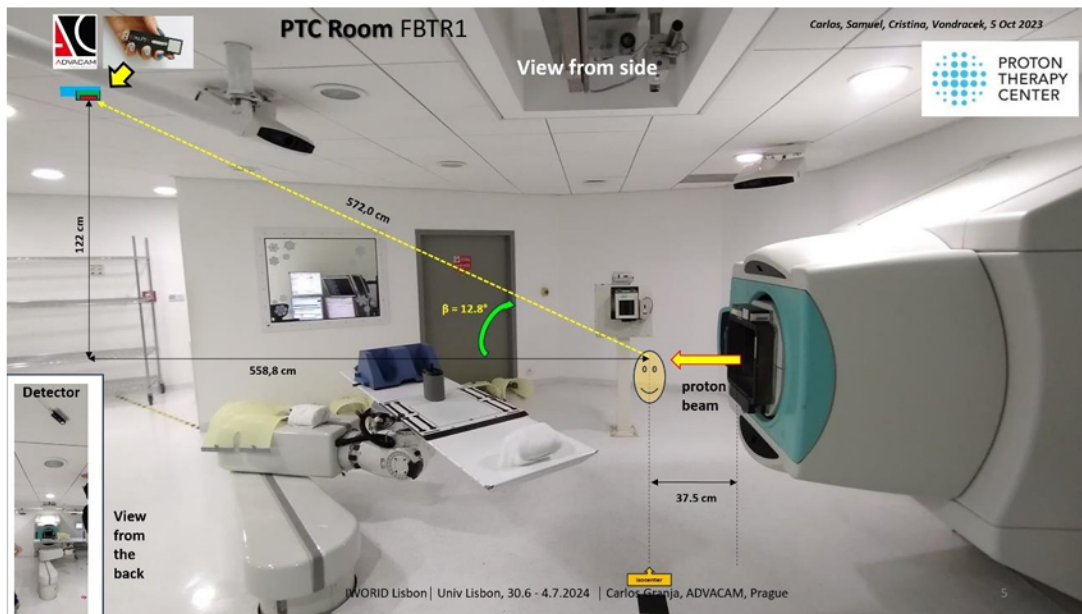


Figura 2.2: Sala de tratamiento con haz fijo en el Proton Therapy Center, Praga. El campo de radiación dispersa se midió de manera no invasiva con una cámara MiniPIX-Timepix, ubicada en el techo de la sala a 5.7 m del isocentro. La imagen muestra la perspectiva general de la sala, con un recuadro que presenta la vista desde la parte posterior. Tomada de [1].

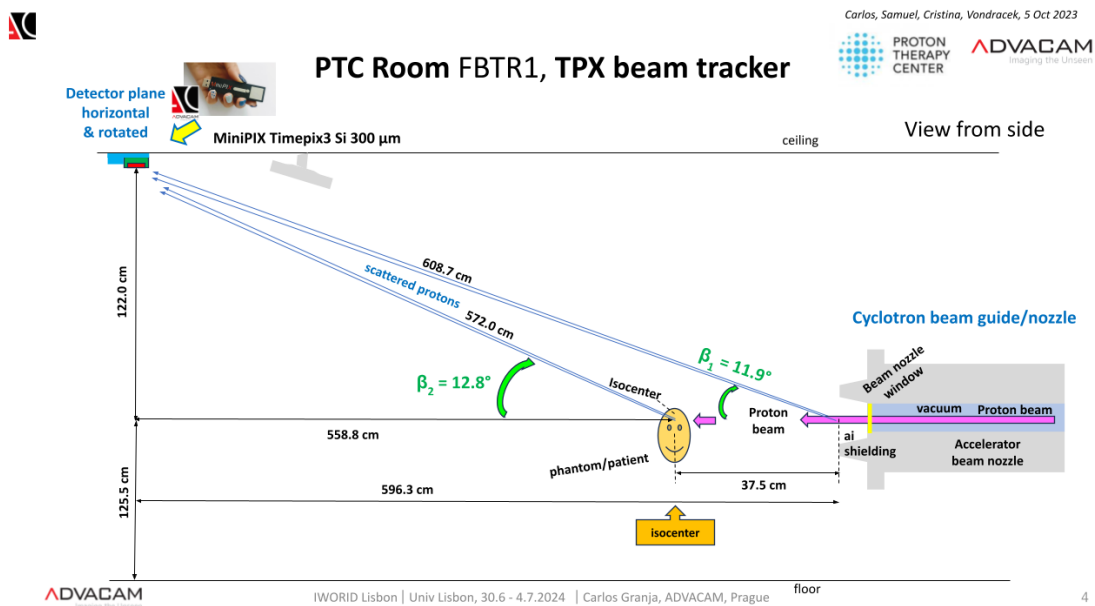


Figura 2.3: Esquema de la disposición experimental en la sala de tratamiento con haz fijo del Proton Therapy Center, Praga. Se muestra la posición del detector MiniPIX-Timepix en el techo, a 5.7 m del isocentro, para la medición no invasiva del campo de radiación dispersa. Tomada de [1].

2.3.1. Caracterización composicional y espectral de alta resolución y amplio rango de campos terapéuticos de radiación compleja/mixta

El estudio busca determinar la distribución de la radiación secundaria de los distintos haces utilizados en el proceso de control de calidad diario. Es importante mantener un control de estas características para determinar si el haz está correctamente configurado respecto a lo calculado. Para ello se utiliza el detector Minipix Timepix3, el cual registra el ToA y el ToT (Sección 2.2); posteriormente se utiliza el software del motor DPE/TraX. El timestamp para registrar las partículas tiene una resolución temporal de aproximadamente 50 nanosegundos. En este caso, las

partículas se dividieron en dos clases (ver Tabla 2.1) [1].

El motor de procesamiento de datos DPE/TraX del Advacam [27] consiste en una serie de scripts de Python que realizan un preprocesamiento, procesamiento y posprocesamiento de los datos crudos del detector. La etapa de preprocesamiento realiza el reconocimiento y discriminación de los distintos tipos de partículas, produciendo histogramas que se utilizan para el reconocimiento de los campos de radiación. En la etapa de procesamiento se realiza el reconocimiento de los campos de radiación y la clasificación de las partículas, cuyo algoritmo depende de los objetivos del usuario (Figura 2.4). Finalmente, en la etapa de posprocesamiento, la información obtenida en las etapas anteriores se convierte en información física, incluyendo el conteo de partículas, el flujo y la tasa de dosis, y se guarda en un archivo con formato .csv [28].

2.4. Procesamiento de datos

2.4.1. Selección de datos de los controles de calidad

El primer paso del procesamiento consistió en identificar y seleccionar los archivos correspondientes a los controles de calidad diarios dentro del conjunto de datos de cada jornada. Este procedimiento se realizó de manera manual. La correcta selección de los archivos es fundamental, ya que una clasificación inadecuada podría incluir, por error, los datos correspondientes a los tratamientos de pacientes. Por esta razón, fue necesario verificar cuidadosamente la hora de creación de

Clase	Subgrupo	Rango de energía	Direccionalidad / Observaciones
HETP	Protones de baja energía (LEp)	1–70 MeV	Omnidireccionales
HETP	Neutrones térmicos rápidos	-	Omnidireccionales
HETP	Protones de alta energía (HEp)	>70 MeV	No perpendiculares
LETP	Electrones de alta energía (HEe)	>300 keV	Omnidireccionales
LETP	Rayos gamma	>30 keV	Omnidireccionales, eficiencia limitada a <100 keV
LETP	Electrones de baja energía (LEe)	10–300 keV	Perpendiculares u omnidireccionales
LETP	Rayos X	5–30 keV	Omnidireccionales
LETP	Protones	>150 MeV	Perpendiculares
LETP	Neutrones rápidos	Energía intermedia	-

Cuadro 2.1: Clasificación de partículas detectadas según su transferencia de energía. Adaptado de [1].

cada archivo registrada en la computadora frente a la hora real en la que fueron generados (hora del país de origen).

Por ejemplo, en la computadora utilizada para este trabajo, los controles de calidad se ejecutaban entre las 6:30 am y las 7:00 am. Sin embargo, al analizar los mismos datos en otro equipo, debido a la configuración horaria, este intervalo correspondía a entre las 5:00 am y las 5:30 am. De ahí la importancia de identificar correctamente el rango horario asociado al inicio de los controles de calidad.

Una vez establecido el rango temporal característico, se procedió a localizar el primer archivo de tamaño considerable (generalmente > 500 kB). A partir de este archivo, se identificaron tres conjuntos adicionales cuyos tamaños también supe-

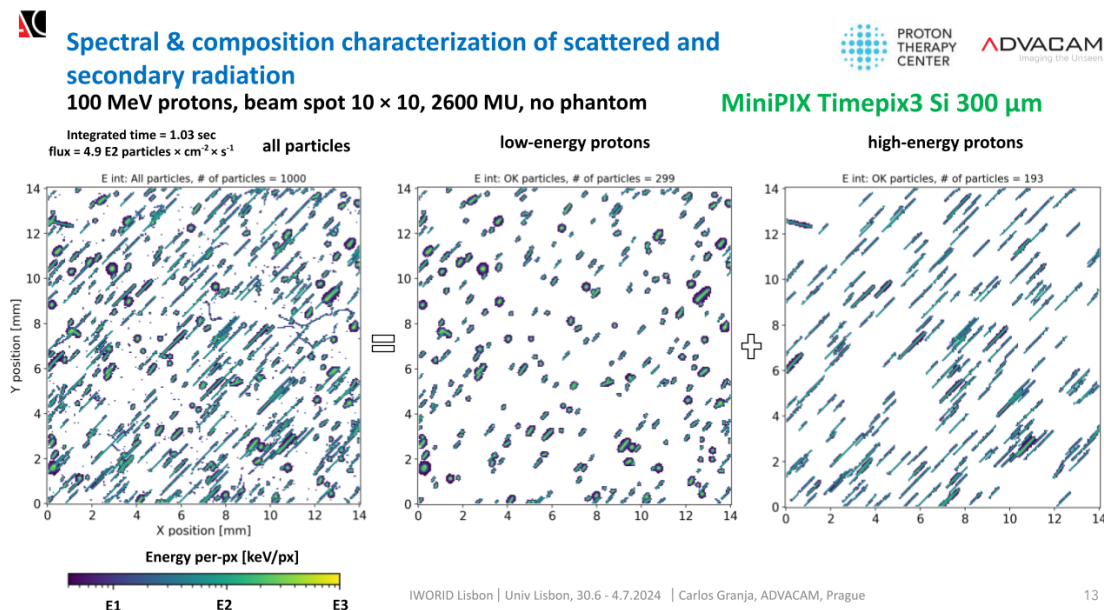


Figura 2.4: Resultado de la etapa de procesamiento de los datos brutos mediante el motor de procesamiento de datos DPE/TraX. Reimpresa con permiso de C. Granja.

rabán los 500 kB. Posteriormente, se seleccionaron los datos comprendidos entre cinco pares de archivos previos al primer archivo grande y cinco pares posteriores al último archivo grande del último conjunto identificado. Cabe recalcar que los datos estaban organizados cronológicamente por fecha de creación. En la mayoría de los casos, los controles de calidad tuvieron una duración aproximada de 40 minutos; no obstante, en situaciones particulares fue necesario ampliar la ventana de análisis hasta cubrir una hora completa para asegurar la inclusión de toda la información relevante.

Finalmente, para facilitar el análisis posterior, los datos seleccionados fueron copiados a una carpeta principal. Dentro de esta carpeta se organizaron subcarpetas correspondientes a los distintos meses analizados (febrero, marzo, abril). A

su vez, cada mes se subdividió en carpetas por día, cuyo nombre contenía la fecha en formato **año-mes-día**, seguida de las tres letras iniciales del día de la semana. Por ejemplo, los datos de una sesión de control de calidad realizada el martes 13 de febrero de 2024 se almacenaron en la carpeta denominada **20240213tue**.

2.4.2. Flujo de Partículas, Tasa de Dosis, Distribuciones Temporales y Espectros de Energía Depositada

El motor de procesamiento de datos DPE/TraX genera como salida un archivo en formato **.csv** denominado **sampling_list** (Tabla A.1), en el cual se recopila la información temporal y las magnitudes físicas derivadas del análisis de los datos crudos del detector. Este archivo contiene las columnas de fecha y tiempo (**year**, **month**, **day**, **hour**, **minute**, **second**), así como las variables de interés: **DR_all**, **DR_1**, **DR_2**, **DR_3**, **DR_4**, **flux_all**, **flux_1**, **flux_2**, **flux_3**, **flux_4**, y las análogas para el conteo de partículas. Aquí, **DR** corresponde a la tasa de dosis, definida como la dosis depositada por unidad de tiempo ($\mu\text{Gy h}^{-1}$), mientras que **flux** representa el flujo de partículas, definido como el número de partículas detectadas por unidad de tiempo y de área ($\text{ctn cm}^{-2} \text{s}^{-1}$).

Las variables con subíndice **all** incluyen todas las partículas detectadas, mientras que las clasificadas como **1-4** corresponden a las categorías de partículas definidas en la Tabla 2.1:

- 1: protones de baja energía,

- 2: protones de alta energía,
- 3: electrones de alta energía y rayos gamma,
- 4: electrones de baja energía y rayos X.

Para obtener las distribuciones temporales, las variables de tiempo (`hour`, `minute`) fueron transformadas en segundos, de modo que se pudieron sumar las filas correspondientes. Así, se construyeron las distribuciones temporales del flujo y de la tasa de dosis de radiación, tanto para el conjunto total de partículas como para cada una de las categorías individuales.

2.5. Procesamiento: análisis dosimétrico

2.5.1. Distribución temporal de tasas de dosis

En este estudio se consideraron las distribuciones temporales de tasas de dosis (Figura 2.5) correspondientes a los 43 días identificados como sesiones de control de calidad válidas.

2.5.2. Campos de QA en el PTC: macrocampos A, B, C, D (protones, electrones) y rayos x (diagnóstico)

Una vez obtenidos los datos correspondientes a los días válidos, se identificaron manualmente las distintas etapas del control de calidad en cada registro

(Figura 2.5):

1. **Campo A:** primera prueba del control.
2. **Campo B:** segunda prueba del control.
3. **Campo C:** tercera prueba del control.
4. **Campo D:** intervalo de tiempo comprendido entre los campos A y B.

Los datos se agruparon dentro de un dataframe específico para cada campo para el posterior análisis.

2.5.3. Microcampos de cada macrocampos: Agrupamiento (Clustering)

Debido al limitado conocimiento sobre la naturaleza de los datos procesados, más allá de que corresponden a controles de calidad diarios que deben ser pruebas estandarizadas, es necesario realizar primero una clasificación de los mismos. Para ello se emplean algoritmos de data clustering, los cuales permiten identificar y agrupar conjuntos de datos homogéneos [29]. La selección del algoritmo adecuado resulta crucial, ya que cada enfoque se centra en propiedades específicas de los datos.

Diversos algoritmos de data clustering han sido desarrollados con aplicaciones en campos específicos. Por ejemplo, los algoritmos basados en la dinámica de Langevin modelan los datos como partículas que siguen las ecuaciones de Langevin,

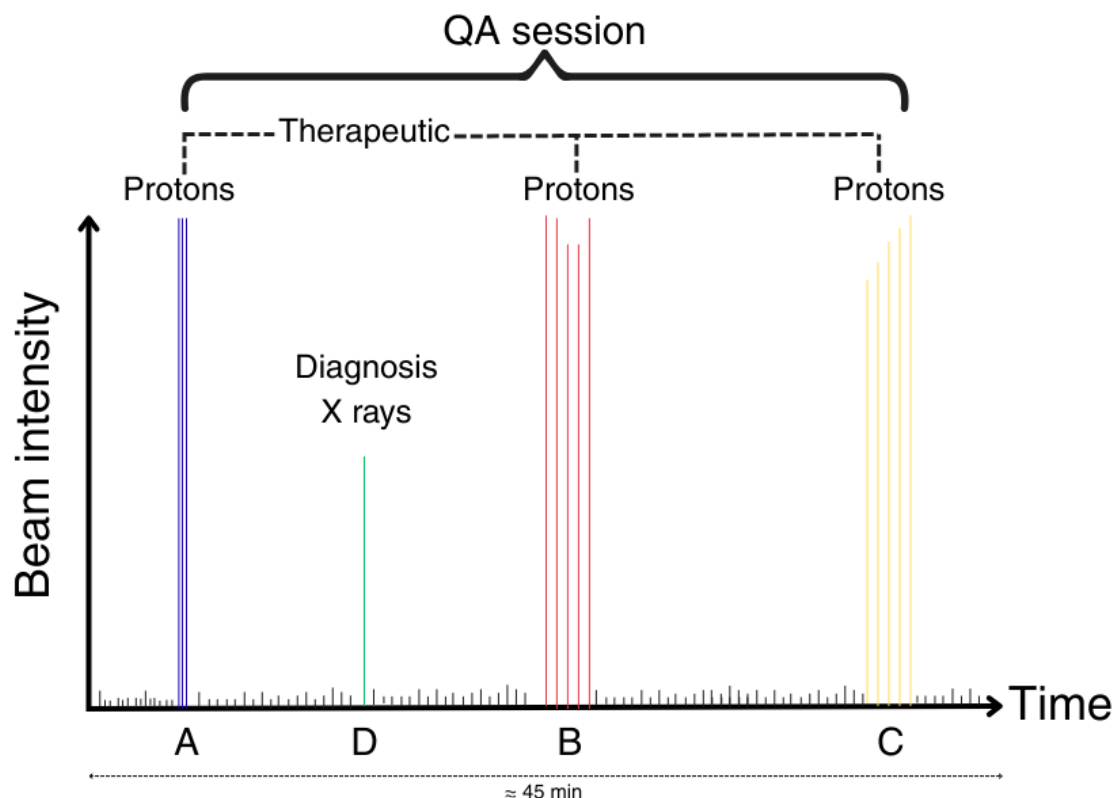


Figura 2.5: Ilustración de los campos de control de calidad (A, B, C, D). El eje Y corresponde a la intensidad del haz medida (flujo/tasa de dosis) y el eje X al tiempo transcurrido durante la adquisición.

lo que puede resultar útil para analizar datos complejos como los que aparecen en Física Médica [30]. Otro caso es el algoritmo de agrupamiento super-paramagnético, que presenta la ventaja de no requerir la especificación previa del número de clústeres y ha demostrado resultados favorables en el análisis de genes humanos que comparten una historia evolutiva común [31].

El algoritmo de k-means es, sin embargo, el más ampliamente utilizado. Un ejemplo relevante se encuentra en su aplicación para identificar subpoblaciones tumorales a partir de representaciones de t-SNE, que resumen las relaciones entre

las firmas moleculares de diferentes subpoblaciones. Este enfoque ha permitido diferenciar grupos y demostrar la existencia de diferencias significativas en las tasas de supervivencia entre ellos [32].

El algoritmo es el siguiente [33]:

Algoritmo 1: Algoritmo K-means

Input: Conjunto de datos $X = \{x_1, x_2, \dots, x_n\}$, número de clusters k

Output: Partición de X en k clusters y sus centroides correspondientes

1 Objetivo: minimizar la suma de errores cuadrados intra-cluster (SSE):

$$\text{SSE} = \sum_{j=1}^k \sum_{x_i \in C_j} \|x_i - \mu_j\|^2$$

```

2 Inicializar  $k$  centroides seleccionando  $k$  objetos aleatorios de  $X$ ;
3 while los centroides cambian significativamente y no se alcanza el número
    máximo de iteraciones do
4   for cada objeto  $x_i \in X$  do
5     Asignar  $x_i$  al cluster  $C_j$  cuyo centroide  $\mu_j$  esté más cercano, según
        una función de distancia;
6   for cada cluster  $C_j, j = 1, \dots, k$  do
7     Recalcular el centroide  $\mu_j$  como el promedio de los objetos
        asignados a  $C_j$ ;
    // Actualizar la SSE con los nuevos centroides
8 return clusters finales  $C_1, C_2, \dots, C_k$  y centroides  $\mu_1, \mu_2, \dots, \mu_k$ ;

```

La función de distancia utilizada en el clustering de datos temporales es la

llamada Dynamic Time Warping (DTW), que ha demostrado ser eficiente en la obtención de clústeres. DTW es una submétrica, ya que no satisface la desigualdad triangular, por lo que no es posible utilizar directamente la optimización de SSE para obtener los clusters [33].

En este caso se emplea el método DTW Barycenter Averaging (DBA). El algoritmo sigue la misma lógica de k-means, pero primero se calcula la distancia DTW entre cada serie temporal individual y la secuencia promedio temporal provisional, con el fin de establecer correspondencias entre las coordenadas de la secuencia promedio y las de cada serie del conjunto. Luego, cada coordenada de la secuencia promedio se actualiza calculando su baricentro, considerando las coordenadas de las series individuales asociadas a ella en el primer paso [34].

La distancia de Dynamic Time Warping (DTW) entre dos series temporales unidimensionales T_1 y T_2 , de longitudes m y n respectivamente, se puede calcular mediante programación dinámica de la siguiente manera [33]:

$$\text{DTW}(T_1, T_2) = \begin{cases} 0, & \text{si } m = n = 0, \\ \infty, & \text{si una de las series está vacía y la otra no,} \end{cases} \quad (2.1)$$

y para los demás casos

$$\text{DTW}(T_1, T_2) = \text{dist}(T_{1,1}, T_{2,1}) + \min \begin{cases} \text{DTW}(\text{Rest}(T_1), \text{Rest}(T_2)), \\ \text{DTW}(\text{Rest}(T_1), T_2), \\ \text{DTW}(T_1, \text{Rest}(T_2)), \end{cases} \quad (2.2)$$

Aquí, $\text{Rest}(T_i)$ denota la serie T_i sin su primer elemento, y $\text{dist}(T_{1,1}, T_{2,1})$ corresponde a la distancia entre los primeros elementos de cada serie, usualmente la norma L_2 (distancia Euclídea).

En este estudio, por lo tanto, se realizó el clustering utilizando la función de medida DTW. Este algoritmo de clustering se encuentra implementado en la librería de python `tslearn`, que utiliza DBA para el cálculo de los clústeres [35]. El código utilizado se presenta en la Sección C.1.

Una vez obtenido los clusteres para los distintos campos del control de calidad se realiza un análisis adicional dentro de cada cluster.

Tests estadísticos

Test de Kruskal-Wallis

Para determinar de manera estadística la variación en las distribuciones de radiación dentro de cada cluster, se aplicó la prueba de Kruskal–Wallis. En los casos en que esta prueba indicó la existencia de diferencias significativas, se realizó un análisis post hoc mediante la prueba de Dunn, con el fin de identificar los pares de días que difieren entre sí.

El test de Kruskal–Wallis es un método no paramétrico análogo al análisis de varianza (ANOVA) paramétrico, que permite identificar la existencia de diferencias estadísticamente significativas entre tres o más grupos [36]. Este test asume que las distribuciones a comparar son independientes entre sí. La hipótesis nula

establece que todas las muestras provienen de la misma población, mientras que la hipótesis alternativa plantea que al menos una de las muestras procede de una población distinta. Una limitación del test es que, en caso de detectar diferencias, no permite identificar directamente entre qué grupos se encuentran; por ello, resulta necesario realizar un análisis post-hoc, en este caso mediante la prueba de Dunn, para determinar qué pares de días difieren entre sí [37].

Test de Dunn

La prueba post-hoc de Dunn es un método no paramétrico cuyo objetivo es identificar diferencias entre pares de grupos de datos [38]. Dado que la realización de múltiples comparaciones incrementa la probabilidad de cometer errores de tipo I, esta prueba se combina habitualmente con el ajuste de Bonferroni, el cual corrige el nivel de significancia y reduce dicho efecto, aumentando así la robustez del método [39].

Test de Kendall Tau

Este test estadístico no paramétrico permite analizar la tendencia general de series temporales en función de sus rangos. La hipótesis nula establece que no existe relación monotónica entre las series comparadas, lo que se representa por un coeficiente de Kendall Tau $T = 0$. Por su parte, la hipótesis alternativa plantea que sí existe una relación monotónica, en cuyo caso el coeficiente varía entre -1 para asociación negativa perfecta y 1 para asociación positiva perfecta [40].

Test de Mahalanobis para la detección de outliers

La distancia de Mahalanobis se utiliza para detectar valores atípicos en datos multidimensionales, identificando observaciones que se encuentran lejos de la media teniendo en cuenta la covarianza de los conjuntos de datos. Para la identificación de los valores atípicos se compara la distancia de Mahalanobis con el umbral derivado de la distribución chi cuadrada con grados de libertad igual al número de variables [41].

Test de Mann–Whitney U

El test no paramétrico de Mann–Whitney U permite comparar dos grupos independientes sin requerir que tengan el mismo tamaño de muestra, lo que lo hace especialmente útil cuando los datos se presentan en forma de rangos, series temporales o distribuciones que no cumplen con los supuestos de normalidad. La hipótesis nula establece que ambos grupos provienen de la misma población; es decir, que no existen diferencias significativas en sus distribuciones, mientras que la hipótesis alternativa plantea que los grupos pertenecen a poblaciones diferentes [42, 43]. Cabe señalar que este test únicamente permite determinar la existencia de una diferencia estadísticamente significativa, pero no proporciona información sobre la magnitud de dicha diferencia.

Implementación

Para todas las pruebas se consideró un nivel de significancia de $\alpha = 0,05$. El test de Kruskal-Wallis se utilizó para identificar si, luego del proceso de clustering, los datos dentro de cada cluster tienen consistencia entre ellos; es decir, si siguen la misma distribución. En los casos donde este test mostró evidencia estadísticamente

significante de diferencias entre las distribuciones, se procedió con el test pos-hoc de Dunn para determinar que par de días difieren entre sí. Adicional a el test de Kruskal-Wallis, se utilizó el test de Kendall tau para comparar cada distribución de tasa de dosis frente a la media histórica del cluster al que pertenece, y la distancia de Mahalanobis para detectar outliers considerando la dosis total, y el pico de tasa de dosis de cada distribución. Finalmente, Se utiliza el test de Mann–Whitney U para comparar entre dos clusters de las distribuciones derivadas de dosis total (calculada considerando un umbral de $0.5 \mu\text{Gy h}^{-1}$) y pico máximo de tasa de dosis de cada distribución diaria.

Las pruebas estadísticas se implementaron en **Python** utilizando la librería **scipy**, específicamente el módulo **stats** con la función **kruskal** para realizar el análisis de Kruskal–Wallis, y **kendalltau** para el análisis de Kendall Tau. Para el análisis post-hoc se empleó la librería **scikit_posthocs**, mediante la función **posthoc_dunn**, la cual incluye la corrección de Bonferroni.

Capítulo 3

Resultados

3.1. Preprocesamiento de Datos (Física Nuclear/Detectores)

3.1.1. Procesamiento de Datos Crudos del Detector

Durante el proceso de selección de los datos pertenecientes a los controles de calidad se identificaron ciertas particularidades en los registros disponibles. En algunos días no se realizaron los controles de calidad, de manera que los primeros archivos del día correspondían a tratamientos clínicos. Además, los controles de calidad se ejecutaban únicamente de lunes a viernes, por lo que los fines de semana carecieron de datos relevantes. Finalmente se observaron días en los que el centro permaneció cerrado, reflejados en archivos con tamaños menores a 5 kB, in-

suficientes para evidenciar actividad de radiación. Por lo tanto se registraron en total 43 días con controles de calidad válidos entre los días 19 de febrero al 28 de marzo del 2024. Entre ellos, los días 23 de febrero, 04, 06 y 18 de marzo, y los días 03 y 10 de abril no presentaron la primera prueba del control de calidad correspondiente al campo A.

3.1.2. Flujo de partículas, tasa de dosis, distribuciones temporales y espectros de energía depositada

Los archivos resultantes del procesamiento con el software DPE/TraX contienen la información como se presenta en la Tabla A.1.

3.2. Procesamiento: Análisis Dosimétrico/Médico/QA

3.2.1. Distribución temporal de flujos de partículas y tasas de dosis

En la Figura 3.1 se muestra un ejemplo de la distribución temporal de la tasa de dosis de radiación. En la Figura, la primera fila corresponde a la tasa de dosis correspondiente a todas las partículas, la segunda fila corresponde a los protones de baja energía, la tercera a los protones de alta energía, la cuarta a los electrones

de alta energía y rayos gamma; y la última fila a los electrones de baja energía y rayos x. El eje y se encuentra en escala logarítmica.

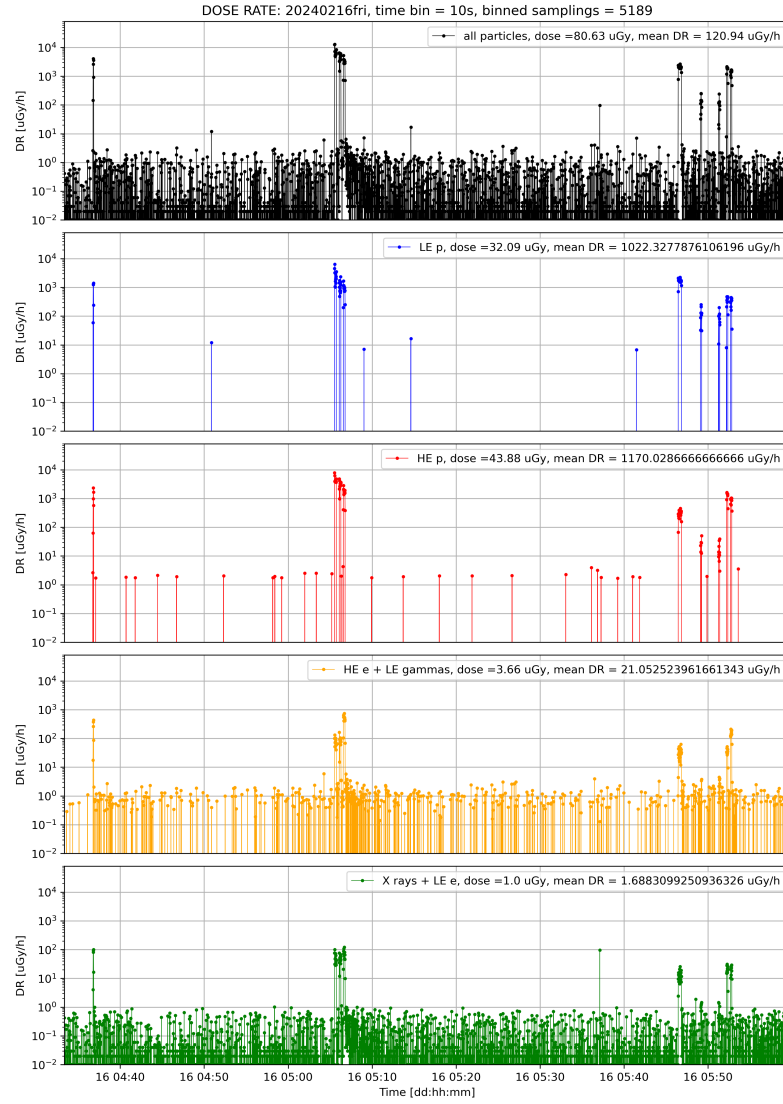


Figura 3.1: Distribución temporal de la tasa de dosis de radiación correspondiente al control de calidad del día 16 de febrero, 2024 (ver etiquetas dentro del gráfico).

3.2.2. Campos de QA en el PTC: Macrocampos A, B, C, D (protones, electrones) y rayos x (diagnóstico)

3.2.3. Microcampos de Cada Macrocampos: Agrupamiento (Clustering)

Campo A

Del proceso de clustering se obtuvieron 3 grupos de datos, como se muestra en la Fig. 3.2, donde se presentan los centroides obtenidos, los cuales representan los patrones característicos de cada grupo y que sirvieron como base para la clasificación de los datos en los distintos clusters. En la Tabla A.2 se presenta la asignación exacta de cada cluster.

Cluster 0

Las 15 distribuciones agrupadas en este cluster (Figura B.1) presentan alta inconsistencia en cuanto al número de picos que se presentan. Los días 20 y 27 de febrero presenta 4 picos, los días 19, 22, 29 de febrero, 07, 13, 28, de marzo, 11, y 17 de abril presentan dos picos; mientras que los días 01, 12, 14, y 25 de marzo presentan 3 picos.

Cluster 1

En este cluster se agruparon correctamente los días donde no se realizó la primera prueba del control de calidad: 23 de febrero, 04, 06, y 18 de marzo, y el

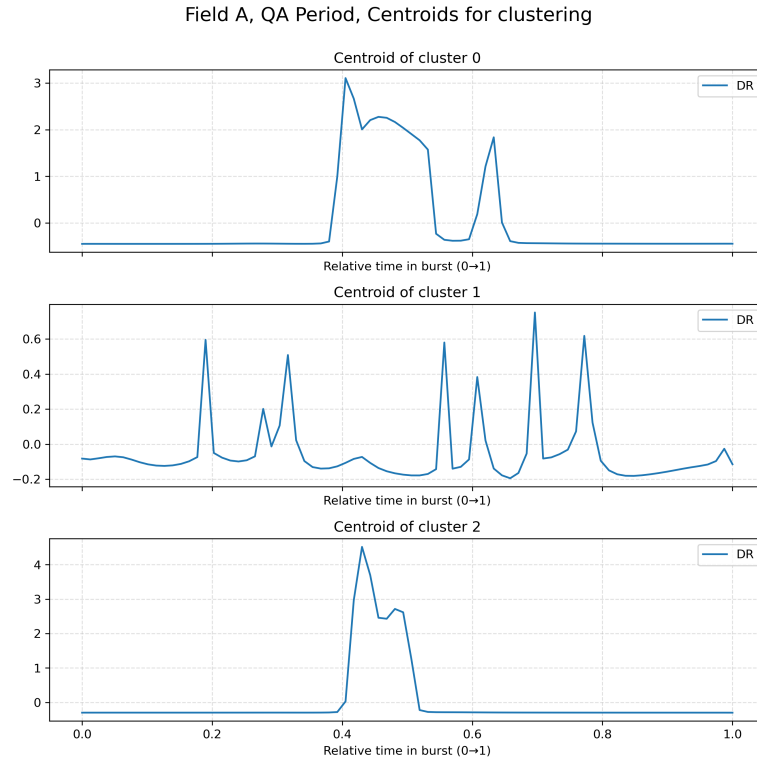


Figura 3.2: Centroides obtenidos a partir del proceso de clustering en el campo A. Se agruparon 15 días en el cluster 0, 6 en el cluster 1 y 22 en el cluster 2.

03 y 10 de abril (Figura B.2).

Cluster 2

En este cluster se agruparon 22 días, los cuales presentaron únicamente un pico de tasa de dosis. Sin embargo, como se observa en la Figura B.3, la duración del pico es diferente en los distintos días. Sin embargo, la mediana de la dosis total, considerando todas las partículas sin distinción de su tipo, fue de $10.20 \mu\text{Gy}$ [$8.51 - 12.99 \mu\text{Gy}$], el pico de tasa de dosis fue de $3333.07 \mu\text{Gy h}^{-1}$ [$2587.80 - 3641.73 \mu\text{Gy h}^{-1}$], y el estadístico de Kendall tau fue de 0.60 [$0.53 - 0.68$], lo que indica alta concordancia. Los protones de alta energía fueron los que más contruyeron a la

dosis total con $4.95 \mu\text{Gy}$ [$3.13 - 6.03 \mu\text{Gy}$], y un pico en la tasa de dosis de $1708.11 \mu\text{Gy h}^{-1}$ [$1108.21 - 1879.02 \mu\text{Gy h}^{-1}$]; mientras que los protones de baja energía presentaron una dosis total de $4.00 \mu\text{Gy}$ [$2.80 - 4.72 \mu\text{Gy}$], y un pico máximo de tasa de dosis de $1285.38 \mu\text{Gy h}^{-1}$ [$1097.80 - 1351.98 \mu\text{Gy h}^{-1}$]. EL estadístico de kendall tau para las distribuciones de tasa de dosis correspondiente a los protones de baja y alta energía fue mayor a 0.60 en cada caso, demostrando que tienen considerable concordancia; mientras que el test de Kruskal-Wallis demostró en ambos casos que existe concordancia entre las distribuciones de tasa de dosis para estas partículas.

Los electrones y los fotones fueron las partículas que menos contribuyeron a la dosis total. En el caso de los electrones de alta energía y rayos gamma de baja energía su contrubución fue de $1.32 \mu\text{Gy}$ [$0.98 - 1.60 \mu\text{Gy}$], mientras que para los electrones de baja energía y rayos x su contribución fue de $0.33 \mu\text{Gy}$ [$0.23 - 0.39 \mu\text{Gy}$], bastante menor a las demás. Aunque ambos canales de partículas de mostraron tener considerable concordancia respecto a la media histórica de tasa de dosis, con un estadístico de Kendall tau de 0.53 [$0.37 - 0.63$] y 0.62 [$0.44 - 0.70$], respectivamente. De igual manera el test de Kruskal-Wallis reafirma esta consistencia dentro los días del cluster con un p-valor de 0.36 y 0.05 respectivamente; aunque las distribuciones correspondientes a electrones de baja energía y rayos x se encuentra en el límite. La distancia Mahalanobis tampoco detectó outliers dentro de este cluster. De modo que las distribuciones dentro de este cluster son consistentes entre ellas.

Campo B

Para el campo B se consideraron 4 clusters, cuyos centroides calculados se muestran en la Fig. 3.3. En la Tabla A.3 se presenta la asignación exacta de cada cluster.

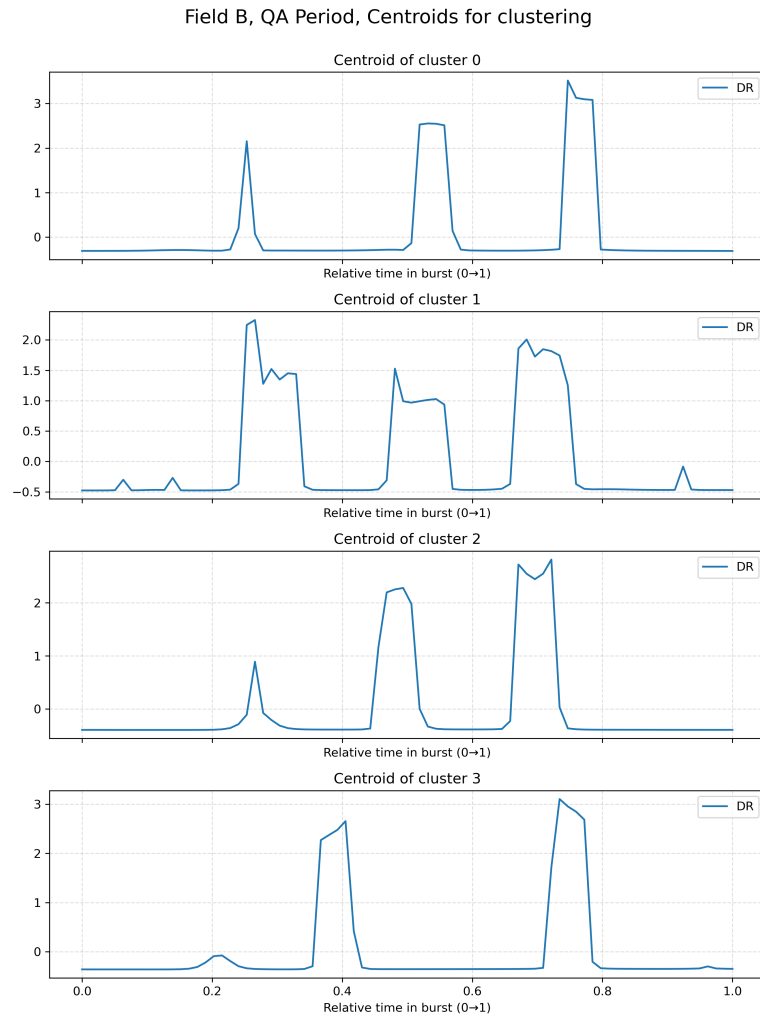


Figura 3.3: Centroides obtenidos a partir del proceso de clustering en el campo B. En este caso se agruparon 5 días en el cluster 0, 20 en el cluster 1, 9 en el cluster 2 y 9 en el cluster 3.

Cluster 0

En este cluster se agruparon los días que contienen 1 o 2 picos extras antes de los picos principales correspondientes al control de calidad (Figura B.4). Las distribuciones diarias de tasa de dosis, para DR_all (todas las partículas) presentan una concordancia moderada, según el estadístico de Kendall Tau (0.69 [0.68 - 0.72]), y el error de predicción, evaluado con RMSE, fue de $855.71 \mu\text{Gy h}^{-1}$ [528.32 - 961.49 $\mu\text{Gy h}^{-1}$]. El pico de tasa de dosis alcanzó $2575.61 \mu\text{Gy h}^{-1}$ [2451.33 - 2776.45 $\mu\text{Gy h}^{-1}$]; mientras que la dosis total presentó un valor mediano de $96.39 \mu\text{Gy}$ [65.86 - 139.35 μGy]. Adicionalmente, la prueba de Kruskal-Wallis presentó un estadístico de 3.77 con un p-valor de 0.44, reafirmando la estabilidad interdiaria.

Los protones de baja energía contribuyeron principalmente a la dosis total con $58.42 \mu\text{Gy}$ [38.28 - 80.70 μGy]. El pico de tasa de dosis de estas partículas fue de $2241.35 \mu\text{Gy h}^{-1}$ [2061.58 - 2427.85 $\mu\text{Gy h}^{-1}$], y la concordancia entre las distribuciones de tasa de dosis de estas partículas fue alta, según Kendall Tau (0.77 [0.72 - 0.79]), y el estadístico de Kruskal-Wallis fue de 7.29 ($p = 0.12$). Los protones de alta energía presentaron una dosis total de $44.07 \mu\text{Gy}$ [33.75 - 48.63 μGy] y la tasa de dosis tuvo un pico de $1601.29 \mu\text{Gy h}^{-1}$ [1554.88 - 1674.79 $\mu\text{Gy h}^{-1}$], Kendall tau de 0.63 [0.63 - 0.69] y Kruskal-Wallis con $p = 0.77$.

Los electrones de alta energía y rayos gamma de baja energía presentaron una dosis total de $3.33 \mu\text{Gy}$ [3.07 - 4.73 μGy], con un pico de tasa de dosis de $224.67 \mu\text{Gy h}^{-1}$ [219.74 - 244.82 $\mu\text{Gy h}^{-1}$]. Las distribuciones diarias tienen concordancia alta (Kendall Tau = 0.70 [0.70 - 0.74]), con un estadístico de Kruskal-Wallis de 6.42 ($p = 0.17$). Los electrones de baja energía y rayos x presentaron la menor

contribución, con una dosis total de $0.89 \mu\text{Gy}$ [$0.81 - 1.28 \mu\text{Gy}$], con un pico de tasa de dosis de $32.65 \mu\text{Gy h}^{-1}$ [$31.28 - 37.54 \mu\text{Gy h}^{-1}$]. El Kendall tau de las distribuciones fue de 0.66 [0.66 - 0.75] y Kruskal-Wallis de 3.88 con un p-valor de 0.42.

Finalmente, la distancia de Mahalanobis obtenida en todos los casos fue de aproximadamente 1.79, con valores por debajo del umbral de significancia establecido para $\alpha = 0.05$, de modo que ninguna de las jornadas se consideró como outlier. Este resultado combinado con los obtenidos con Kruskal-Wallis muestra que, tanto las distribuciones de tasa de dosis, como las distintas métricas asociadas son consistentes.

Cluster 1

En este clúster se agruparon las distribuciones de 20 días cuyos tres picos principales presentan una altura similar (Figura B.5). Dentro de este grupo, el 19 de febrero muestra seis picos en lugar de tres, lo que podría deberse a que la prueba de control de calidad se repitió dos veces en esa jornada. Por otro lado, el 14 de marzo se observan dos picos adicionales posteriores a la prueba de control, sin información que permita explicar esta anomalía. Por esta razón, ambos días fueron omitidos del análisis.

Las distribuciones diarias de tasa de dosis, que consideran todas las partículas obtuvo un estadístico de Kendall tau de 0.55 [0.49 – 0.61] y el de Kruskal-Wallis fue de 12.83, con un p-valor de 0.85, por lo que las distribuciones tienen concordancia moderada; el RMSE evaluado fue de $1884.79 \mu\text{Gy h}^{-1}$ [1649.55 – 2219.15 $\mu\text{Gy h}^{-1}$].

h^{-1}]. El pico de tasa de dosis fue de $12749.49 \mu\text{Gy h}^{-1}$ [$12316.37 - 13600.52 \mu\text{Gy h}^{-1}$], y la mediana de la dosis total fue de $91.63 \mu\text{Gy}$ [$82.24 - 100.37 \mu\text{Gy}$].

En este caso fueron los protones de alta energía los que contribuyeron mayoritariamente a la dosis total ($58.40 \mu\text{Gy}$ [$52.29 - 63.49 \mu\text{Gy}$]), con un pico de tasa de dosis de $8029.62 \mu\text{Gy h}^{-1}$ [$6934.72 - 8450.99 \mu\text{Gy h}^{-1}$]. Kendall Tau mostro que la concordancia entre las distribuciones fue moderada (0.61 [$0.51 - 0.65$]), y kruskall-Wallis (p-valor = 0.86) también demostro consistencia. Para el caso de los protones de baja energía, la dosis total fue de $28.13 \mu\text{Gy}$ [$25.22 - 30.59 \mu\text{Gy}$], el pico de tasa de dosis fue de $5068.61 \mu\text{Gy h}^{-1}$ [$4690.70 - 5714.88 \mu\text{Gy h}^{-1}$], y, de acuerdo a los estadísticos de Kendall tau (0.60 [$0.52 - 0.63$]) y Kruskall-Wallis (p-valor = 0.85), las distribuciones de tasa de dosis son consistentes.

Otra vez los electrones de alta energía y rayos gamma de baja energía presentaron una contribución menor a la dosis total ($4.00 \mu\text{Gy}$ [$3.67 - 4.48 \mu\text{Gy}$]) con un pico de tasa de dosis de $734.49 \mu\text{Gy h}^{-1}$ [$679.30 - 767.32 \mu\text{Gy h}^{-1}$]. Las distribuciones tienen concordancia moderada (Kendall tau = 0.40 [$0.35 - 0.45$]), y también son consistentes (Kruskall-Wallis con p-valor = 0.91). Los electrones de baja energía y rayos x presentan una dosis total mucho menor ($1.08 \mu\text{Gy}$ [$0.97 - 1.15 \mu\text{Gy}$]), y otra vez el estadístico de Kendall tau (0.47 [$0.40 - 0.54$]) y de Kruskall-Wallis (p-valor = 0.73) encuentran evidencia estadística de que las distribuciones de tasa de dosis son consistentes.

La distancia de Mahalanobis estuvo por debajo del umbral de significancia para $\alpha = 0.05$ en todos los casos, de modo que no se detectaron outliers. Por lo tanto, las distribuciones de tasa de dosis en este cluster son consistentes entre sí.

Cluster 2

En este cluster se agruparon 9 días cuyas distribuciones de tasa de dosis presentan el primer pico un poco reducido en comparación a los otros dos (Figura B.6), a excepción de el día 03 de abril, cuyos picos parecen tener la misma altura, de modo que se excluyó este día del análisis.

Las distribuciones diarias de tasa de dosis que considera todas las partículas presentan una concordancia moderada, según el estadístico de Kendall Tau ($0.52 [0.49 - 0.56]$), con un error de predicción (RMSE) de $586.46 \mu\text{Gy h}^{-1} [542.88 - 645.41 \mu\text{Gy h}^{-1}]$. El pico de tasa de dosis alcanzó $2115.33 \mu\text{Gy h}^{-1} [1994.69 - 2155.06 \mu\text{Gy h}^{-1}]$; y la dosis total presentó un valor de $25.92 \mu\text{Gy} [22.61 - 26.96 \mu\text{Gy}]$. La prueba de Kruskal-Wallis, con un p-valor de 0.81, reafirmó la estabilidad de las distribuciones.

En este caso, fueron los protones de baja energía los que contribuyeron principalmente a la dosis total con $17.73 \mu\text{Gy} [15.96 - 18.80 \mu\text{Gy}]$. El pico de tasa de dosis de estas partículas fue de $1611.82 \mu\text{Gy h}^{-1} [1509.12 - 1631.49 \mu\text{Gy h}^{-1}]$, y la concordancia entre las distribuciones de tasa de dosis de estas partículas fue moderada, según Kendall Tau ($0.55 [0.54 - 0.57]$); el estadístico de Kruskal-Wallis fue de 3.04 ($p = 0.93$). Los protones de alta energía presentaron una dosis total de $6.07 \mu\text{Gy} [5.09 - 6.46 \mu\text{Gy}]$ y la tasa de dosis tuvo un pico de $516.39 \mu\text{Gy h}^{-1} [490.70 - 586.63 \mu\text{Gy h}^{-1}]$, Kendall tau de $0.45 [0.38 - 0.50]$ y Kruskal-Wallis con $p = 0.60$; de modo que son consistentes y con una consistencia moderada.

Los electrones de alta energía y rayos gamma de baja energía presentaron una

dosis total de $1.59 \mu\text{Gy}$ [$1.40 - 1.73 \mu\text{Gy}$], con un pico de tasa de dosis de $232.66 \mu\text{Gy h}^{-1}$ [$217.92 - 263.29 \mu\text{Gy h}^{-1}$]. Las distribuciones diarias tienen concordancia moderada-baja (Kendall Tau = 0.36 [$0.30 - 0.40$]), con un estadístico de Kruskal-Wallis de 4.17 (p-valor = 0.84). Los electrones de baja energía y rayos x presentaron nuevamente la menor contribución, con una dosis total de $0.32 \mu\text{Gy}$ [$0.29 - 0.36 \mu\text{Gy}$], con un pico de tasa de dosis de $36.91 \mu\text{Gy h}^{-1}$ [$31.57 - 39.72 \mu\text{Gy h}^{-1}$]. El Kendall tau de las distribuciones fue de 0.39 [$0.32 - 0.48$] y Kruskal-Wallis de 9.68 con un p-valor de 0.28 . Finalmente, la distancia de Mahalanobis no detectó outliers, de modo que las distribuciones presentes en este cluster, nuevamente, son consistentes.

Cluster 3

En este cluster se agruparon 9 días, cuyas distribuciones de tasa de dosis se asemejan a las del cluster 2 (Figura B.7). En la comparación entre los clusters 2 y 3 mediante la prueba de Mann-Whitney U no se encontraron diferencias estadísticamente significativas ($p < 0.05$) en ninguna de las métricas (dosis total y pico de la tasa de dosis) evaluadas. De modo que, esto demuestra que el cluster 2 y 3 son consistentes entre sí.

Comparación entre el cluster 1 y 2

La comparación entre los clusters 1 y 2 muestra diferencias altamente significativas en varias métricas de dosis. Para el caso de todas las partículas, la dosis total acumulada y el pico de tasa de dosis presentan p-valores extremadamente bajos ($\approx 6.43 \times 10^{-7}$), lo que indica que los días agrupados en el cluster 1 difieren clara-

mente de los del cluster 2 en términos de magnitud de la dosis y de los picos. Los protones de baja energía y los protones de alta energía también muestran diferencias significativas en la dosis total y el pico de tasa de dosis ($\approx 6,43 \times 10^{-7}$), al igual que los electrones de alta energía y rayos gamma de baja energía ($\approx 6,43 \times 10^{-7}$). Por otra parte, los electrones de baja energía y rayos x no presentaron diferencias significativas en la dosis total ni en la tasa de dosis ($\approx 6,43 \times 10^{-7}$ para dosis total y $\approx 0,218$ para pico de tasa de dosis), indicando que la contribución de estas partículas a la separación de clusters es menor. Estos resultados confirman que los clusters 1 y 2 se diferencian de manera clara tanto en la magnitud de la dosis como en las distribuciones de tasa de dosis, evidenciando características distintivas entre los días agrupados en cada cluster.

Campo C

En este caso se consideraron 3 centroides distintos, los cuales se presentan en la Fig. 3.4, y en la Tabla A.4 se presenta la asignación exacta de cada cluster.

Cluster 0

En este cluster se agruparon 10 días. La característica principal de estas distribuciones es que los tres picos presentan la misma altura (Figura B.8).

Las distribuciones diarias de tasa de dosis muestran que los protones de alta energía fueron los principales contribuyentes a la dosis total, con $51.04 \mu\text{Gy}$ [$49.82 - 54.41 \mu\text{Gy}$] y un pico de tasa de dosis de $6796.43 \mu\text{Gy h}^{-1}$ [$6511.74 - 7188.23 \mu\text{Gy h}^{-1}$]. La concordancia entre las distribuciones de estos protones fue moderada,

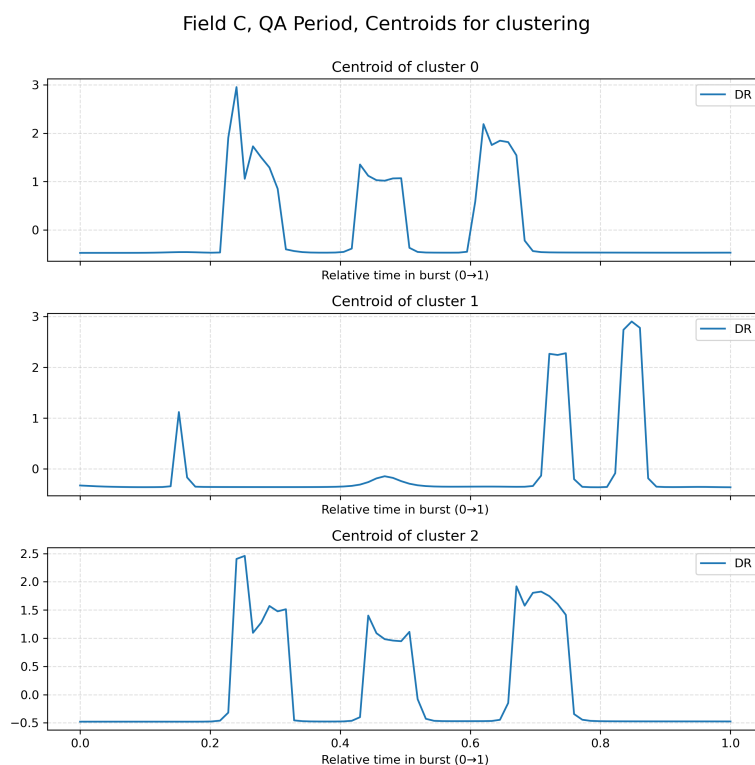


Figura 3.4: Centroides obtenidos a partir del proceso de clustering en el campo C. Se agruparon 10 días en el cluster 0, 22 días en el cluster 1 y 11 días en el cluster 2.

como lo indica el estadístico de Kendall tau ($0.64 [0.59 - 0.67]$). Los protones de baja energía también aportaron significativamente a la dosis total ($24.69 \mu\text{Gy} [23.80 - 26.97 \mu\text{Gy}]$), con un pico de tasa de dosis de $5484.48 \mu\text{Gy h}^{-1} [4402.34 - 5980.02 \mu\text{Gy h}^{-1}]$, cuyas distribuciones tuvieron concordancia moderada (Kendall tau = $0.64 [0.56 - 0.66]$).

Los electrones de alta energía y rayos gamma de baja energía presentaron una contribución menor ($3.54 \mu\text{Gy} [3.35 - 3.80 \mu\text{Gy}]$), con un pico de $697.27 \mu\text{Gy h}^{-1} [634.40 - 770.88 \mu\text{Gy h}^{-1}]$ y concordancia moderada a baja (Kendall tau = $0.41 [0.35 - 0.48]$), mientras que los electrones de baja energía y rayos X mostraron la

menor contribución ($0.97 \mu\text{Gy}$ [$0.91 - 0.98 \mu\text{Gy}$]), con un pico de $111.94 \mu\text{Gy h}^{-1}$ [$108.64 - 117.27 \mu\text{Gy h}^{-1}$] y concordancia moderada (Kendall tau = 0.49 [$0.44 - 0.59$]).

La consistencia de estas distribuciones a lo largo de los días se confirmó mediante Kruskal-Wallis, la cual no detectó diferencias significativas entre los días del cluster para ninguno de las distribuciones de tasa de dosis correspondiente a los distintos tipos de partículas (p-values > 0.05 en todos los casos, desde 0.98 hasta 0.99), lo que indica que las distribuciones diarias son estadísticamente equivalentes. La prueba de Mahalanobis tampoco detectaron outliers en las distintas métricas analizadas.

Cluster 1

En este cluster se agruparon 22 días. La característica principal que se observó en las distribuciones de tasa de dosis es que el primer pico es de menor altura respecto a los otros dos. Sin embargo, los días 19, 20, y 28 de febrero, el 04, 08, 12, 14, 25 y 28 de marzo, y 09 de abril presentan ciertos picos extras, algunos antes y otros después de los tres picos correspondientes al control de calidad, que probablemente correspondan a disparos fuera de la prueba de control de calidad, aunque sin los registros de las actividades realizadas esos días no se puede comprobar, de modo que se omitieron estos días del análisis, restando 12 días válidos.

Los resultados para el conjunto de todas las partículas muestran una dosis total de $22.43 \mu\text{Gy}$ [$20.05 - 36.67 \mu\text{Gy}$], con un pico de tasa de dosis de $2185.30 \mu\text{Gy h}^{-1}$ [$2025.85 - 2354.48 \mu\text{Gy h}^{-1}$]. La correlación medida con Kendall tau, fue

moderada (0.39 [0.34 – 0.45]). El análisis de Kruskal-Wallis no mostró diferencias estadísticamente significativas entre los días de medida (p-valor = 0.455). Al considerar únicamente los protones, los de baja energía presentaron una dosis integrada de $5.32 \mu\text{Gy}$ [4.69 - 8.84 μGy] y un pico de tasa de dosis de $556.61 \mu\text{Gy h}^{-1}$ [502.54 - 640.32 $\mu\text{Gy h}^{-1}$]. Por otro lado, los protones de alta energía alcanzaron una dosis total de $15.59 \mu\text{Gy}$ [13.80 - 19.06 μGy], con un pico de $1636.53 \mu\text{Gy h}^{-1}$ [1583.38 - 1705.75]. EN ambos casos se obtuvieron correlaciones Kendall tau moderadas (0.31 [0.26 - 0.38] y 0.42 [0.35 - 0.46], respectivamente). Las pruebas de Kruskal-Wallis confirmaron la ausencia de diferencias significativas (p-valor = 0.652 para protones de baja energía y p-valor = 0.494 para protones de alta energía).

Finalmente, los electrones de alta energía junto con rayos gamma de baja energía presentaron una dosis total de $1.43 \mu\text{Gy}$ [1.31 - 1.79 μGy] y un pico de tasa de dosis de $217.65 \mu\text{Gy h}^{-1}$ [207.78 - 226.39 $\mu\text{Gy h}^{-1}$]. Los electrones de baja energía y rayos X alcanzaron una dosis total mucho menor, de $0.29 \mu\text{Gy}$ [0.27 - 0.43 μGy], con un pico de $33.61 \mu\text{Gy h}^{-1}$ [30.60 - 36.50 $\mu\text{Gy h}^{-1}$]. En ambos casos, la correlación temporal fue baja a moderada (Kendall tau = 0.33 [0.27 - 0.39] y 0.40 [0.29 - 0.46]). El análisis de Kruskal-Wallis no identificó diferencias estadísticamente significativas en estos subgrupos (p-valor = 0.629 y p-valor = 0.833, respectivamente). La prueba adicional con la distancia Mahalanobis encontró outliers ni en las distribuciones ni en la dosis total, ni en el pico de tasa de dosis. De modo que las distribuciones de este cluster son consistentes entre sí, aunque con correlación moderada.

Cluster 2

En este cluster se agruparon 11 días (Figura B.10), cuyas distribuciones de tasa de dosis comparten similitud con el cluster 1, a excepción del día 10 de abril, que presenta 2 picos adicionales, de modo que este día se descartó al realizar el análisis estadístico. Al comparar los clusters 0 y 2 mediante la prueba de Mann–Whitney U no se encontraron diferencias estadísticamente significativas ni en la Dosis total ni en el pico de tasa de dosis, ni en ninguna de las distribuciones correspondientes a un tipo de partícula en específico. Los p-valores fueron elevados en todos los casos (0.12–1.0), lo cual indica que las distribuciones son equivalentes entre clusters. Por esta razón, los clusters 0 y 2 pueden considerarse representaciones de un mismo comportamiento, y su diferenciación se debe más a la segmentación del algoritmo de agrupamiento que a una diferencia física o estadística en las distribuciones de tasa de dosis.

Comparación entre cluster 0 y 1

Para el total de partículas, la comparación entre los clusters 0 y 1 mostró diferencias estadísticamente significativas tanto en la dosis total como en el pico de tasa de dosis (p-valor < 0.001 en ambos casos). En el caso de los protones, tanto de baja como de alta energía, se observaron resultados consistentes en el área bajo la curva y en el pico de tasa de dosis, con diferencias altamente significativas en ambos casos (p-valor < 0.001). Finalmente, al considerar los electrones y rayos x, se encontraron diferencias significativas en la dosis total y en el pico de tasa de dosis (p-valor < 0.001 en ambos casos). Esto demuestra que los dos clusters presentan dos distribuciones de tasa de dosis completamente distintas entre sí.

Campo D

En este campo, que comprende el intervalo de tiempo entre al campo A y C no presentó ninguna característica distintiva. La dosis total para todos los canales de partículas fue menor a $1 \mu\text{Gy}$. el estadístico de Kandall Tau resultó mayor a 0.41, por lo que existe concordancia en las distribuciones. Sin embargo, el test de Kruskall-Wallis demostro que existen diferencias estadísticamente significativas en los canales de protones de baja energía, y el de electrones de baja energía y rayos x. El test pos-hoc de Dunn obtuvo que el día que la distribución que varía significativamente respecto a las demás de la del 27 de marzo, en el caso de los protones de baja energía; mientras que el 07 y 26 de marzo presentan variaciones significativas en el caso de los electrones baja energía. Finalmente, la distancia Mahalanobis no detectó Outliers.

Capítulo 4

Conclusiones

Luego de la depuración de los datos crudos provenientes del detector se obtuvieron 43 días válidos que presentaron datos pertenecientes a los controles de calidad diarios. De estos días, 8 no presentaron el primer campo correspondiente a la primera prueba del control. Luego del procesamiento de los datos mediante el motor DPE/Trax se obtuvieron los archivos con las distribuciones temporales de tasa de dosis, además de las distribuciones de flujo, entre otras. La identificación manual de los distintos campos del control en estos archivos permitió verificar nuevamente que existían días donde no se realizó la primera prueba, como se muestra en la Figura B.2.

En las distribuciones del campo A se pudieron identificar tres clústeres distintos, donde los días sin datos de controles se agruparon exitosamente en un solo clúster (clúster 1). Dentro del clúster 0 se agruparon los días que contenían más de un pico de tasa de dosis. Es probable que estos picos adicionales se deban a

una repetición de la prueba, aunque para poder verificarlo sería necesario acceder a los registros de los controles. En el clúster 2 se agruparon aquellos días donde se obtuvo un solo pico de tasa de dosis. La prueba de Kruskal-Wallis presentó evidencia estadística de que las distribuciones de este clúster son consistentes; sin embargo, la variabilidad relativa de la dosis total respecto a su mediana fue de aproximadamente el 44 %. Como se muestra en la Figura B.3, a pesar de que las distribuciones tienen un solo pico, la duración del mismo varía en ciertos días, lo cual puede ser la razón de la alta variabilidad de la dosis total.

En el caso del campo B fue necesario considerar cuatro clústeres distintos. En el clúster 0 se agruparon los días cuyas distribuciones presentaban más picos de los esperados para esta prueba (Figura B.4). En el clúster 1 se agruparon los días cuyas distribuciones tienen tres picos similares en altura; el test de Kruskal-Wallis presentó evidencia estadística de que las distribuciones son consistentes, y la dispersión relativa de la dosis total respecto a la mediana fue de aproximadamente el 20 %, siendo los protones de alta energía los que mayor contribución tuvieron en la tasa de dosis. En los clústeres 2 y 3 se agruparon los días con distribuciones cuyo primer pico tenía menor altura en comparación con los otros dos. El test de Mann-Whitney U no presentó evidencia estadística de que las distribuciones de estos dos clústeres sean diferentes. Por otro lado, el test de Mann-Whitney U presentó evidencia estadísticamente significativa de que las distribuciones de los clústeres 1 y 2 son diferentes. En sí, existe una gran diferencia en la mediana de la dosis total de cada clúster: en el clúster 1 la dosis total fue de $91.63 \mu\text{Gy}$ [$18.13 \mu\text{Gy}$], siendo los protones de alta energía los que más aportaron; mientras que en el clúster 2 la mediana fue de $25.92 \mu\text{Gy}$ [$4.35 \mu\text{Gy}$], con un mayor aporte de los protones

de baja energía. De modo que, los clústeres tienen diferencias significativas en las magnitudes de dosis total y en las partículas que aportaron mayoritariamente a la misma.

Para el campo C se consideraron tres clústeres distintos. En los clústeres 0 y 2 se agruparon aquellas distribuciones con los tres picos de la misma altura. El test de Mann-Whitney U no encontró evidencia estadísticamente significativa de que las distribuciones de estos dos clústeres fueran distintas. En ambos casos fueron los protones de alta energía los que aportaron mayoritariamente a la dosis total, cuya mediana fue de $51.05 \mu\text{Gy}$ [$4.59 \mu\text{Gy}$]. El test de Kruskal-Wallis obtuvo evidencia estadísticamente significativa de que las distribuciones de dosis de radiación en los clústeres son consistentes. Finalmente, en el clúster 1 se agruparon las distribuciones cuyo primer pico tenía menor altura respecto a los dos restantes. En este clúster fueron los protones de alta energía los que más contribuyeron a la dosis total, con una mediana de $15.59 \mu\text{Gy}$ [$5.26 \mu\text{Gy}$]. El test de Mann-Whitney U entre los clústeres 0 y 1 encontró evidencia estadísticamente significativa de que las distribuciones de estos clústeres son completamente diferentes, como se evidencia también en las diferencias en la magnitud de dosis total obtenidas en cada clúster. Finalmente, en el campo D el test de Kruskal-Wallis arrojó evidencia estadísticamente significativa de que al menos dos distribuciones difieren entre sí. Posteriormente, mediante el test pos-hoc de Dunn se determinó que la distribución de tasa de dosis producida por los protones de baja energía del 27 de marzo es la distribución que varía frente a las demás.

Estos resultados demuestran que el detector Minipix Timepix3 es capaz de

encontrar diferencias en los distintos campos de los controles de calidad, siendo una potencial herramienta como sistema de verificación de los controles de calidad diarios. Sin embargo, es necesario antes analizar si se realizaron cambios en las pruebas de control de calidad, que también podrían explicar estas variaciones detectadas con el Timepix3. Futuros estudios pueden enfocarse en analizar solamente los datos tomados por el detector el día posterior a las calibraciones para determinar experimentalmente cuál es el porcentaje de variación que es posible detectar con el dispositivo, o contrastar los resultados obtenidos en este estudio con simulaciones de Monte Carlo. Además, de ser el caso que no se realizaron cambios intencionados en los haces de los controles de calidad, sería de interés estudiar la razón detrás de las variaciones, de modo que se pueda incluir una base de datos que permita que, en los controles de calidad, mediante métodos de clustering en línea, se obtenga información sobre lo que está ocurriendo en el haz que sale del gantry.

Bibliografía

- [1] C. Granja, S. Kurucz, J. Solc, V. Vondracek, J. Pacik, and C. Oancea. Spectral-tracking characterization and wide-range monitoring of scattered and secondary radiation in proton therapy. *Journal of Instrumentation*, 20(04):C04026, 2025.
- [2] G. Delaney, S. Jacob, C. Featherstone, and M. Barton. The role of radiotherapy in cancer treatment: Estimating optimal utilization from a review of evidence-based clinical guidelines. *Cancer*, 104(6):1129–1137, 2005.
- [3] M. Barton, M. Frommer, and J. Shafiq. Role of radiotherapy in cancer control in low-income and middle-income countries. *The Lancet Oncology*, 7(7):584–595, 2006.
- [4] International Atomic Energy Agency. Number of radiotherapy machines per million people [data set]. Directory of Radiotherapy Centres (DIRAC), n.d. Accessed March 2025.
- [5] P. Maitre, R. Krishnatry, S. Chopra, S. Gondhowiarjo, B. M. Likonda, Q. M. Hussain, E. H. Zubizarreta, and J. P. Agarwal. Modern radiotherapy techno-

- logy: Obstacles and opportunities to access in low- and middle-income countries. *JCO Global Oncology*, 8:e2100376, 2022.
- [6] E. Natelauri, M. Pkhaladze, and M. Atskvereli. Radiobiology of proton therapy and its clinical implications. In T. J. FitzGerald, editor, *Proton Therapy - Scientific Questions and Future Direction*, chapter 3. IntechOpen, London, 2024.
- [7] P. A. Taylor. Clinical trials evaluating proton therapy. In T. J. FitzGerald and M. Bishop-Jodoin, editors, *Proton Therapy - Current Status and Future Directions*, chapter 5. IntechOpen, London, 2021.
- [8] R. Hazem. Interaction of proton beam with human tissues in proton therapy. In T. J. FitzGerald, editor, *Proton Therapy - Scientific Questions and Future Direction*, chapter 5. IntechOpen, London, 2023.
- [9] T. Nishio, M. Nakamura, H. Okamoto, S. Kito, T. Minemura, S. Ozawa, Y. Kumazaki, M. Ishikawa, N. Tohyama, M. Kurooka, T. Nakashima, H. Shimizu, R. Suzuki, S. Ishikura, and Y. Nishimura. An overview of the medical-physics-related verification system for radiotherapy multicenter clinical trials by the medical physics working group in the japan clinical oncology group-radiation therapy study group. *Journal of Radiation Research*, 61(6):999–1008, 2020.
- [10] Editor(s): Hasan Murshed. Proton radiation therapy. In Hasan Murshed, editor, *Fundamentals of Radiation Oncology*, pages 117–132. Academic Press, fourth edition edition, 2024.

- [11] R. Zhu, X. F. Poenisch, F. Song, X. L. Johnson, J. G. Ciangaru, B. Taylor, M. F. Lii, M. C. Martin, B. Arjomandy, K. Lee, A. S. Choi, N. Nguyen, Q. T. Gillin, M. and N. Sahoo. Patient-specific quality assurance for prostate cancer patients receiving spot scanning proton therapy using single-field uniform dose. *International Journal of Radiation Oncology, Biology, Physics*, 81(2):552–559, 2011.
- [12] N. Bizzocchi, F. Fracchiolla, M. Schwarz, and C. Algranati. A fast and reliable method for daily quality assurance in spot scanning proton therapy with a compact and inexpensive phantom. *Medical Dosimetry*, 42(3):238–246, 2017.
- [13] S. A. Park, J. W. Kwak, M. G. Yoon, D. H. Shin, S. B. Lee, K. H. Cho, S. K. Kang, K. J. Kim, H. S. Bae, and S. Y. Park. Dose verification of proton beam therapy using the gafchromic ebt film. *Radiation Measurements*, 46(8):717–721, 2011.
- [14] Y. Gao, C.-W. Chang, S. Pan, J. Peng, C. Ma, P. Patel, J. Roper, J. Zhou, and X. Yang. Deep learning-based synthetic dose-weighted let map generation for intensity modulated proton therapy. *Physics in Medicine & Biology*, 69(2):025004, 2024.
- [15] P. Stasica, H. Nguyen, C. Granja, R. Kopeć, L. Marek, C. Oancea, Ł. Raczynski, A. Rucinski, M. Rydygier, K. Schubert, R. Schulte, and J. Gajewski. Single proton let characterization with the timepix detector and artificial intelligence for advanced proton therapy treatment planning. *Physics in Medicine and Biology*, 68(10):104001, 2023.

- [16] T. K. Rügheimer, U. Gebert, T. Michel, G. Anton, J. Séguinot, and C. Joram. Experimental demonstration of a hybrid photon detector concept based on the timepix detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 595(2):353–358, 2008.
- [17] A. Novak, C. Granja, A. Sagatova, J. Jakubek, B. Zatko, V. Vondráček, M. Andrlík, V. Zach, S. Polanský, and A. Rathi. Silicon carbide Timepix3 detector for quantum-imaging detection and spectral tracking of charged particles in wide range of energy and field-of-view. *Journal of Instrumentation*, 18(11):C11004, nov 2023.
- [18] C. Oancea, C. Granja, L. Marek, J. Jakubek, J. Šolc, E. Bodenstein, S. Gantz, J. Pawelke, and J. Pivec. Out-of-field measurements and simulations of a proton pencil beam in a wide range of dose rates using a timepix3 detector: Dose rate, flux and LET. *Physica Medica*, 106:102529, 2023.
- [19] P. Stasica-Dudek, C. Oancea, C. Granja, K. Guguła, J. Jakubek, R. Kopec, D. Krzempek, M. Matous, A. Rucinski, M. Rydygier, and J. Gajewski. Configuration of Timepix3 read-out parameters for spectral measurements in proton therapy applications. *Physica Medica*, 130:104885, 2025.
- [20] C. Granja, J. Jakubek, M. Martišíková, V. Vondráček, L. Opalka, and J. Šolc. Dynamic range and resolving power of the timepix detector to heavy charged particles. *Journal of Instrumentation*, 13(11):C11003, 2018.
- [21] A. Rosenfeld, S. Alnaghy, M. Petasecca, D. Cutajar, M. Lerch, S. Pospisil, V. Giacometti, R. Schulte, V. Rosso, M. Würfl, C. Granja, M. Martišíková,

- and K. Parodi. Medipix detectors in radiation therapy for advanced quality-assurance. *Radiation Measurements*, 130:106211, 2020.
- [22] A. Novak, C. Granja, A. Sagatova, V. Zach, J. Stursa, and C. Oancea. Spectral tracking of proton beams by the timepix3 detector with gaas, cdte and si sensors. *Journal of Instrumentation*, 18(01):C01022, jan 2023.
- [23] J. H. Jungmann, L. MacAleese, R. Buijs, F. Giskes, A. de Snaijer, J. Visser, J. Visschers, M. J. J. Vrakking, and R. M. A. Heeren. Fast, high resolution mass spectrometry imaging using a medipix pixelated detector. *International Journal of Mass Spectrometry*, 341–342:34–44, 2013.
- [24] J. Fey, S. Böhm, P. Köhler, M. Börner, M. Hörmann, A. Häusler, M. Rösch, M. Schürmann, A. Heering, and M. Hörmann. Investigations on performance and spectroscopic capabilities of a 3 mm cdte sensor connected to a medipix3rx readout chip. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 957:164308, 2020.
- [25] S. Nievaart, B. Janssen, R. Middendorp, J. Pacík, et al. Pairing cdte array detector capabilities with traditional gamma-spectrometry setup for revealing betas in mixed gamma-beta energy spectra for free release measurements. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1078:170643, 2025.
- [26] L. Ghesquière-Déricks, A. Schlechter, R. Félix-Bautista, T. Gehrke, G. Echner, L. Kelleter, and M. Martišíková. Investigation of suitable detection angles

for carbon-ion radiotherapy monitoring in depth by means of secondary-ion tracking. *Frontiers in Oncology*, 11:780221, 2021.

- [27] Advacam. Trax engine cli: Manual de usuario. https://wiki.advacam.cz/wiki/TraX_Engine_CLI. Accedido en marzo de 2025.
- [28] L. Marek, C. Granja, J. Jakubek, J. Ingerle, D. Turecek, M. Vuolo, and C. Oancea. Data processing engine (dpe): data analysis tool for particle tracking and mixed radiation field characterization with pixel detectors timepix. *Journal of Instrumentation*, 19(04):C04026, apr 2024.
- [29] Z. Li, X. Hu, K. Li, F. Zhou, and F. Shen. Inferring the outcomes of rejected loans: an application of semisupervised clustering. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 183(2):pp. 631–654, 2020.
- [30] K. Lafata, Z. Zhou, J.-G. Liu, and F.-F. Yin. Data clustering based on langevin annealing with a self-consistent potential. *Quarterly of Applied Mathematics*, 77(3):pp. 591–613, 2019.
- [31] A. Murua and N. Wicker. The conditional-potts clustering model. *Journal of Computational and Graphical Statistics*, 23(3):717–739, 2014.
- [32] W. M. Abdelmoula, B. Balluff, S. Englert, J. Dijkstra, M. J. T. Reinders, A. Walch, L. A. McDonnell, and B. P. F. Lelieveldt. Data-driven identification of prognostic tumor subpopulations using spatially mapped t-sne of mass spectrometry imaging data. *Proceedings of the National Academy of Sciences of the United States of America*, 113(43):12244–12249, 2016.

- [33] D. Kotsakos, G. Trajcevski, D. Gunopulos, and C. Aggarwal. Time-series data clustering. In Charu C. Aggarwal and Chandan K. Reddy, editors, *Data Clustering: Algorithms and Applications*, chapter 15, pages 358–376. Taylor & Francis Group, LLC, 2014.
- [34] F. Petitjean, A. Ketterlin, and P. Gançarski. A global averaging method for dynamic time warping, with applications to clustering. *Pattern recognition*, 44(3):678–693, 2011.
- [35] R. Tavenard, J. Faouzi, G. Vandewiele, F. Divo, G. Androz, C. Holtz, M. Payne, R. Yurchak, M. Rußwurm, K. Kolar, and E. Woods. Tsllearn, a machine learning toolkit for time series data. *Journal of Machine Learning Research*, 21(118):1–6, 2020.
- [36] X. Xu, C. Huang, and W. U. H. Shah. Financial resources utilization efficiency in sports infrastructure development, determinant of total factor productivity growth and regional production technology heterogeneity in china. *Heliyon*, 10(5):e26546, 2024.
- [37] J. A. Pérez-Méndez and A. Machado-Cabezas. Relationship between management information systems and corporate performance. *Revista de Contabilidad*, 18(1):32–43, 2015.
- [38] D. Güneş and H. Hasegawa. Optimizing uav sprayer performance using field data and machine learning approaches. *Smart Agricultural Technology*, 11:101013, 2025.

- [39] A. Mendieta-Aragón, R. Arguedas-Sanz, L.M. Ruiz-Gómez, et al. Tackling the challenge of peer learning in hybrid and online universities. *Education and Information Technologies*, 28:4505–4529, 2023.
- [40] S. C. Eggenberger, N. L. James, C. Ho, S. S. Eamegdool, V. Tatarinoff, N. A. Craig, B. S. Gow, S. Wan, C. W. D. Dodds, D. La Hood, A. Gilmour, S. L. Donahoe, M. Krockenberger, K. Tumuluri, M. J. da Cruz, J. R. Grigg, P. McCluskey, N. H. Lovell, M. C. Madigan, A. T. Fung, and G. J. Suaning. Implantation and long-term assessment of the stability and biocompatibility of a novel 98 channel suprachoroidal visual prosthesis in sheep. *Biomaterials*, 279:121191, 2021.
- [41] Peng Shi, Huanchang Yan, Xingjun Fan, and Shuhua Xi. A benchmark dose analysis for urinary cadmium and type 2 diabetes mellitus. *Environmental Pollution*, 273:116519, 2021.
- [42] M. E. Tuftedal, D. J. Delene, and Andrew Detwiler. Precipitation evaluation of the north dakota cloud modification project (ndcmp) using rain gauge observations. *Atmospheric Research*, 269:105996, 2022.
- [43] B. Wellington and S. Rouhani. Environmental statistics. In Kandi Brown, William L. Hall, Marjorie Snook, and Kathleen Garvin, editors, *Sustainable Land Development and Restoration*, pages 311–323. Butterworth-Heinemann, 2010.

Apéndice A

Tablas adicionales

Cuadro A.2: Asignación de días a los distintos clusters determinados por el algoritmo de clasificación aplicado a los datos del *Campo A*. La primera columna indica las fechas analizadas, la segunda columna especifica el cluster al que corresponde cada día, y la tercera columna muestra la duración de los datos (en segundos) que fueron considerados por el algoritmo para asignar el cluster correspondiente. Esta tabla proporciona un resumen formal de la organización de los días dentro de cada cluster.

Día	Cluster asignado	Duración de datos (s)
2024-02-19	0	87.0
2024-02-20	0	113.0
2024-02-21	2	78.0
2024-02-22	0	92.0
2024-02-23	1	36.0
2024-02-26	2	87.0
2024-02-27	0	277.0
2024-02-28	2	88.0

Tabla A.2 continúa de la página anterior

Día	Cluster asignado	Duración de datos (s)
2024-02-29	0	88.0
2024-03-01	0	111.0
2024-03-04	1	53.0
2024-03-05	0	95.0
2024-03-06	1	92.0
2024-03-07	0	112.0
2024-03-08	2	83.0
2024-03-11	2	70.0
2024-03-12	0	109.0
2024-03-13	0	92.0
2024-03-14	0	118.0
2024-03-15	2	75.0
2024-03-18	1	59.0
2024-03-19	2	72.0
2024-03-20	2	76.0
2024-03-21	2	68.0
2024-03-22	2	73.0
2024-03-25	0	106.0
2024-03-26	2	66.0
2024-03-27	2	78.0
2024-03-28	0	88.0
2024-04-02	2	83.0

Tabla A.2 continúa de la página anterior

Día	Cluster asignado	Duración de datos (s)
2024-04-03	1	52.0
2024-04-04	2	84.0
2024-04-05	2	84.0
2024-04-08	2	79.0
2024-04-09	2	82.0
2024-04-10	1	35.0
2024-04-11	0	106.0
2024-04-12	2	97.0
2024-04-15	2	79.0
2024-04-16	2	78.0
2024-04-17	0	90.0
2024-04-18	2	92.0
2024-04-19	2	86.0

Cuadro A.3: Asignación de días a los distintos clusters determinados por el algoritmo de clasificación aplicado a los datos del *Campo B*. La primera columna indica las fechas analizadas, la segunda columna especifica el cluster al que corresponde cada día, y la tercera columna muestra la duración de los datos (en segundos) que fueron considerados por el algoritmo para asignar el cluster correspondiente. Esta tabla proporciona un resumen formal de la organización de los días dentro de cada cluster.

Día	Cluster asignado	Duración de datos (s)
2024-02-19	1	342.0
2024-02-20	1	149.0

Tabla A.3 continúa de la página anterior

Día	Cluster asignado	Duración de datos (s)
2024-02-21	2	134.0
2024-02-22	1	153.0
2024-02-23	2	148.0
2024-02-26	1	160.0
2024-02-27	0	369.0
2024-02-28	1	143.0
2024-02-29	0	275.0
2024-03-01	3	183.0
2024-03-04	1	158.0
2024-03-05	2	143.0
2024-03-06	1	173.0
2024-03-07	3	144.0
2024-03-08	1	175.0
2024-03-11	3	173.0
2024-03-12	1	171.0
2024-03-13	3	157.0
2024-03-14	1	268.0
2024-03-15	2	148.0
2024-03-18	1	145.0
2024-03-19	3	150.0
2024-03-20	2	143.0
2024-03-21	1	153.0

Tabla A.3 continúa de la página anterior

Día	Cluster asignado	Duración de datos (s)
2024-03-22	2	148.0
2024-03-25	1	149.0
2024-03-26	3	154.0
2024-03-27	1	137.0
2024-03-28	1	172.0
2024-04-02	2	157.0
2024-04-03	2	130.0
2024-04-04	0	556.0
2024-04-05	1	171.0
2024-04-08	3	145.0
2024-04-09	1	167.0
2024-04-10	3	178.0
2024-04-11	0	465.0
2024-04-12	2	173.0
2024-04-15	1	164.0
2024-04-16	3	263.0
2024-04-17	1	155.0
2024-04-18	1	160.0
2024-04-19	0	354.0

Cuadro A.1: Ejemplo de registros generados por DPE/TraX en formato .csv. Se muestran variables de tiempo, conteo de clústeres (clu), flujo de partículas (flux), tasa de dosis (DR), dosis acumulada (D) y tiempo UNIX.

year	month	day	hour	minute	second	ch-all	clu-1	clu-2	clu-3	clu-4	flux-all	flux-1	flux-2	flux-3	flux-4	DR-all	DR-1	DR-2	DR-3	DR-4	D-all	D-1	D-2	D-3	D-4	unixtime	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2024	2	16	4	36	43.00	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00	0.00	0	0	0	0	0	0	0	0	0	1708058203.00	
2024	2	16	4	36	44.00	1	0	1	0	0	0.51	0.00	0.51	0.00	0.00	2.62	0	2.62	0	0	0	0	0	0	0	1708058204.00	
2024	2	16	4	36	45.00	0	0	0	0	0	0.00	0.00	0.00	0.00	0.00	0	0	0	0	0	0	0	0	0	0	1708058205.00	
2024	2	16	4	36	46.00	68	7	16	11	34	34.34	3.54	8.08	5.56	17.17	143.33	59.13	62.74	17.40	4.07	398.14	16425	17427	4833	1129	1708058206.00	
2024	2	16	4	36	47.00	1122	140	264	211	507	566.67	70.71	133.33	106.57	256.06	2633.81	1309.72	981.67	251.94	80.49	11289.96	363812	272686	72760	22357	1708058207.00	
2024	2	16	4	36	48.00	1285	111	442	241	491	648.99	56.06	223.23	121.72	247.98	4061.17	1232.34	2361.05	378.53	92.24	11289.96	363812	272686	72760	22357	1708058208.00	
2024	2	16	4	36	49.00	1311	116	501	273	591	662.12	58.59	182.32	137.88	258.33	3610.29	1410.41	1661.94	436.81	101.13	10028.96	36178	441651	121365	28091	1708058209.00	
2024	2	16	4	36	50.00	264	17	107	55	115	148.48	8.59	54.48	27.76	38.33	923.32	242.92	576.91	86.85	16.64	254.76	67477	160253	24126	4621	1708058210.00	
2024	2	16	4	36	51.00	6	0	0	3	3	3.03	0.00	0.00	0.00	1.52	2.27	0	0	2.03	0.24	630	0	0	0	0	1708058211.00	
2024	2	16	4	36	52.00	2	0	0	0	2	1.01	0.00	0.00	0.00	1.01	0.03	0	0	0	0.03	9	0	0	0	9	1708058212.00	

Cuadro A.4: Asignación de días a los distintos clusters determinados por el algoritmo de clasificación aplicado a los datos del *Campo C*. La primera columna indica las fechas analizadas, la segunda columna especifica el cluster al que corresponde cada día, y la tercera columna muestra la duración de los datos (en segundos) que fueron considerados por el algoritmo para asignar el cluster correspondiente. Esta tabla proporciona un resumen formal de la organización de los días dentro de cada cluster.

Día	Cluster asignado	Duración de datos (s)
2024-02-19	1	200.0
2024-02-20	1	385.0
2024-02-21	2	138.0
2024-02-22	1	138.0
2024-02-23	0	173.0
2024-02-26	1	142.0
2024-02-27	0	163.0
2024-02-28	1	123.0
2024-02-29	2	147.0
2024-03-01	0	163.0
2024-03-04	1	367.0
2024-03-05	0	168.0
2024-03-06	1	152.0
2024-03-07	0	154.0
2024-03-08	1	183.0
2024-03-11	1	157.0
2024-03-12	1	380.0
2024-03-13	2	146.0

Tabla A.4 continúa de la página anterior

Día	Cluster asignado	Duración de datos (s)
2024-03-14	1	179.0
2024-03-15	2	144.0
2024-03-18	1	160.0
2024-03-19	2	161.0
2024-03-20	0	150.0
2024-03-21	1	165.0
2024-03-22	0	168.0
2024-03-25	1	240.0
2024-03-26	2	142.0
2024-03-27	1	149.0
2024-03-28	1	240.0
2024-04-02	0	150.0
2024-04-03	0	164.0
2024-04-04	0	162.0
2024-04-05	1	130.0
2024-04-08	2	154.0
2024-04-09	1	300.0
2024-04-10	2	199.0
2024-04-11	1	146.0
2024-04-12	2	203.0
2024-04-15	1	138.0
2024-04-16	2	155.0

Tabla A.4 continúa de la página anterior

Día	Cluster asignado	Duración de datos (s)
2024-04-17	1	136.0
2024-04-18	1	138.0
2024-04-19	2	162.0

Apéndice B

Figuras

B.1. Distribuciones diarias de tasas de dosis por campo y clusters correspondientes

B.1.1. Campo A

Cluster 0



Figura B.1: Distribuciones diarias de la tasa de dosis correspondientes al cluster 0 del campo A.

Cluster 1

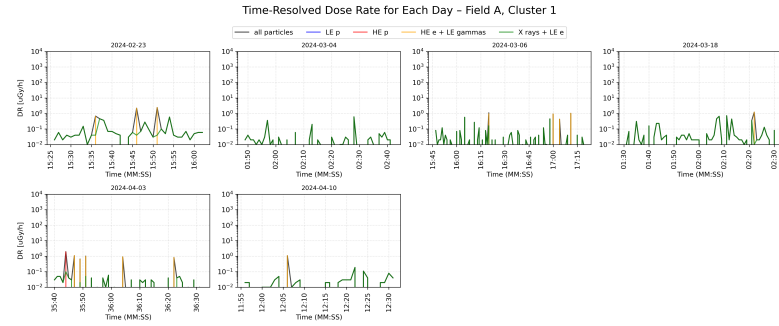


Figura B.2: Distribuciones diarias de la tasa de dosis correspondientes al cluster 1 del campo A.

Cluster 2



Figura B.3: Distribuciones diarias de la tasa de dosis correspondientes a 16 días del cluster 2 del campo A.

B.1.2. Campo B

Cluster 0

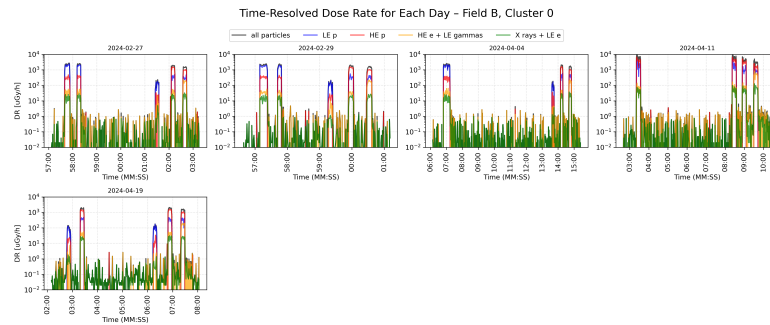


Figura B.4: Distribuciones diarias de la tasa de dosis correspondientes al cluster 0 del campo B.

Cluster 1

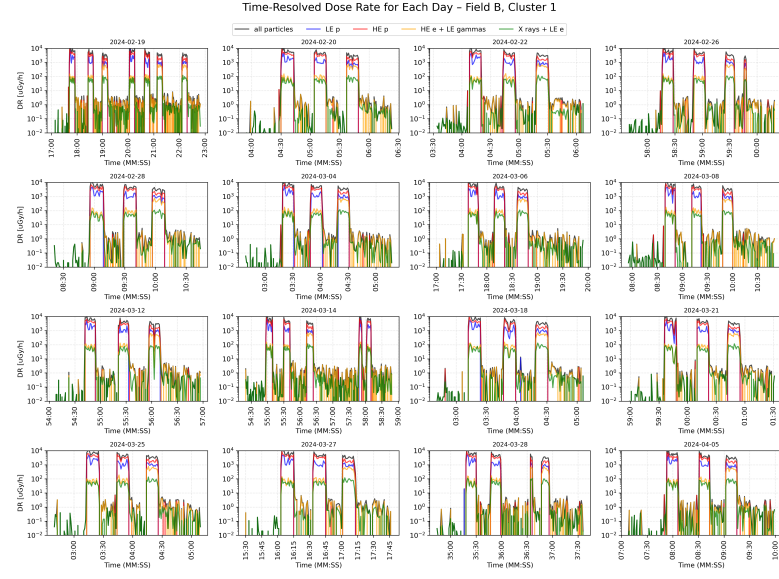


Figura B.5: Distribuciones diarias de la tasa de dosis correspondientes a 16 días del cluster 1 del campo B.

Cluster 2

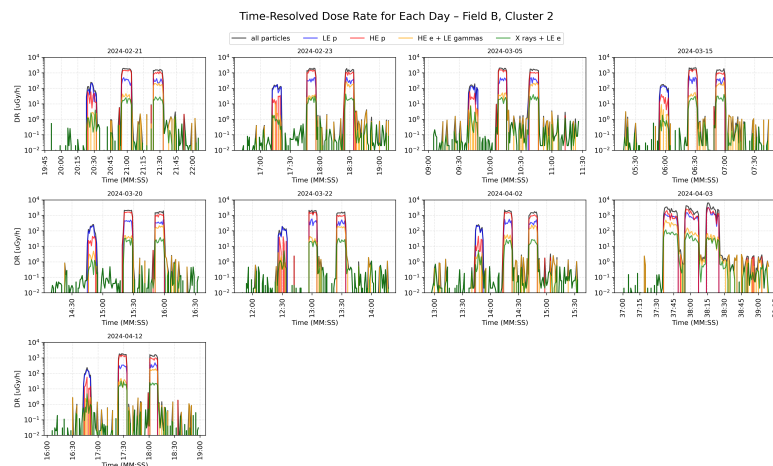


Figura B.6: Distribuciones diarias de la tasa de dosis correspondientes al cluster 2 del campo B.

Cluster 3

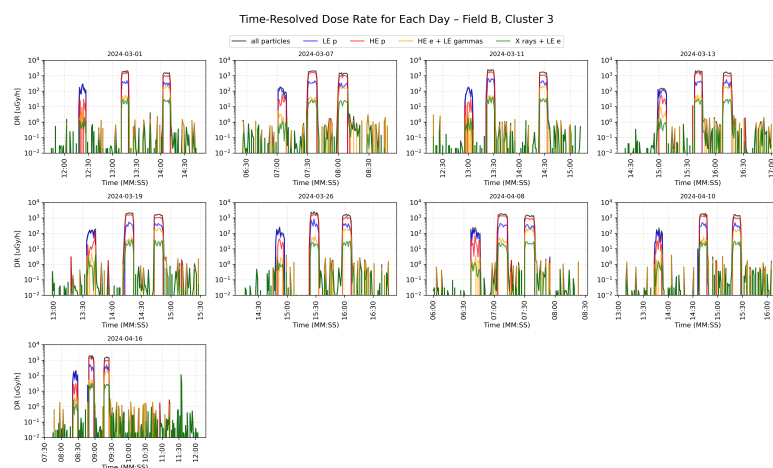


Figura B.7: Distribuciones diarias de la tasa de dosis correspondientes al cluster 3 del campo B.

B.1.3. Campo C

Cluster 0

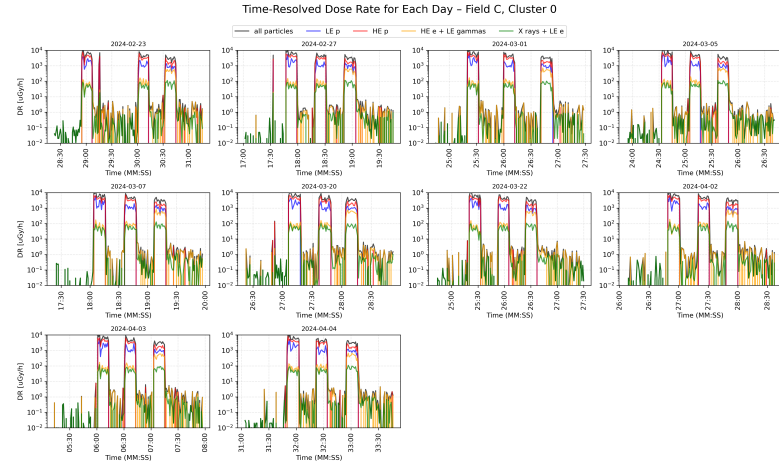


Figura B.8: Distribuciones diarias de la tasa de dosis correspondientes al cluster 0 del campo C.

Cluster 1

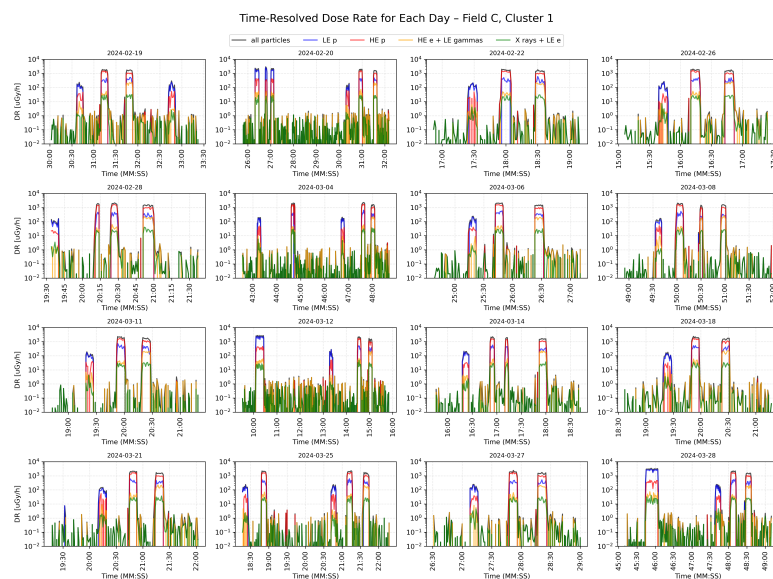


Figura B.9: Distribuciones diarias de la tasa de dosis correspondientes a 16 días del cluster 1 del campo C.

Cluster 2

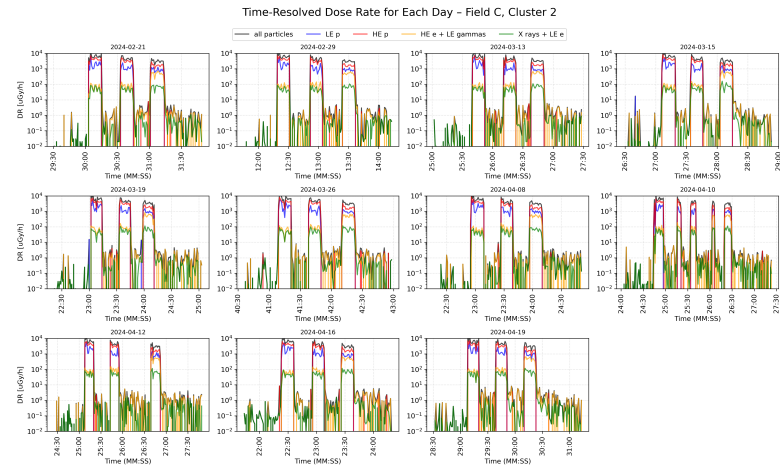


Figura B.10: Distribuciones diarias de la tasa de dosis correspondientes al cluster 2 del campo C.

Apéndice C

Código en python

C.1. Función de clustering con DTW

En esta sección se presenta la función `clustering`, desarrollada en Python, la cual permite extraer ráfagas de los datos, re-muestrearlas, escalarlas y aplicar `TimeSeriesKMeans` con la métrica DTW de la librería `tslearn` [35] para obtener los centroides de los clústeres.

La función tiene como parámetros de entrada lo siguiente:

1. un DataFrame de `pandas` que debe contener las columnas temporales: `year`, `month`, `day`, `hour`, `minute`, `second`, y las variables `flux` y `DR`.
2. el número de clústeres (`num_clusters`) que se generarán en el análisis.
3. una carpeta de salida (`output_folder`) en la que se guardarán los resultados.

La función devuelve un DataFrame con las columnas: `dia`, `cluster` y `duracion_seg`. La columna `dia` registra el día que se analiza, la columna `cluster` indica el clúster asignado a los datos de ese día, y la columna `duracion_seg` corresponde a la duración de la ráfaga considerada para determinar el clúster.

El funcionamiento interno de la función es el siguiente:

1. **Configuración de parámetros iniciales:** se definen las variables de interés (`CATEGORIAS`, `SUFIJOS`), el número de puntos para remuestreo (`RESAMPLED_POINTS`), así como los parámetros para la detección de ráfagas (`MAX_GAP_SEC`, `MIN_BURST_POINTS`).
2. **Funciones auxiliares:**
 - `find_bursts`: identifica ráfagas de datos consecutivos según un intervalo máximo permitido entre puntos y un número mínimo de puntos.
 - `resample_burst`: remuestrea cada ráfaga para que todas tengan el mismo número de puntos, permitiendo comparar series de distinta longitud.
3. **Extracción y remuestreo de ráfagas por día:** se recorren los datos agrupados por día, se selecciona la ráfaga más larga de cada día y se remuestrea según los parámetros definidos.
4. **Escalado de datos:** las series temporales se normalizan mediante `StandardScaler` para que las variables tengan media cero y desviación estándar uno, facilitando la comparación entre ellas.
5. **Clustering con DTW:** se aplica `TimeSeriesKMeans` de `tslearn`, usando la métrica DTW para agrupar las ráfagas en clústeres según su similitud temporal.

6. **Visualización de centroides:** se generan gráficos de los centroides de cada clúster para analizar la forma promedio de las ráfagas agrupadas.

El código se presenta en el Listado C.1.

Listado C.1: Función de clustering con DTW

```
def clustering(data, num_clusters, output_folder):

    # ===== CONFIG =====

    CATEGORIAS = ['flux', 'DR']  # solo flux y DR
    SUFIJOS = ['all', '1', '2', '3', '4']
    RESAMPLED_POINTS = 50
    MAX_GAP_SEC = 60
    MIN_BURST_POINTS = 5
    N_CLUSTERS = num_clusters

    # =====

    df = data.copy()

    for c in ['year', 'month', 'day', 'hour', 'minute']:
        df[c] = pd.to_numeric(df[c], errors='coerce')
    df['second'] = pd.to_numeric(df['second'], errors='coerce').fillna(0).
        round().astype(int)
    dt = pd.to_datetime(df[['year', 'month', 'day', 'hour', 'minute', 'second']],
        errors='coerce')
    df = df[~dt.isna()].copy()
    df['fecha'] = dt[~dt.isna()]
    df = df.sort_values('fecha').reset_index(drop=True)
    expected_cols = [f'{cat}_{s}' for cat in CATEGORIAS for s in SUFIJOS]
```

```

missing = [c for c in expected_cols if c not in df.columns]
if missing:
    print("Faltan columnas:", missing)

# ===== 2) Funciones auxiliares =====

def find_bursts(times, max_gap_sec=60, min_points=3):
    if len(times) == 0:
        return []
    times = pd.to_datetime(times)
    gaps = (times[1:].values - times[:-1].values) / np.timedelta64(1, 's')
    starts = [0]
    for i, g in enumerate(gaps, start=1):
        if g > max_gap_sec:
            starts.append(i)
    bursts = []
    for s, e in zip(starts, starts[1:] + [len(times)]):
        if (e - s) >= min_points:
            bursts.append((s, e))
    return bursts

def resample_burst(time_rel, values, n_points=50):
    if len(time_rel) < 2:
        return np.repeat(values[1:], n_points, axis=0)
    t0, t1 = time_rel.min(), time_rel.max()

```

```

if t1 == t0:

    return np.repeat(values[:1], n_points, axis=0)

x = (time_rel - t0) / (t1 - t0)

xq = np.linspace(0, 1, n_points)

out = np.empty((n_points, values.shape[1]), dtype=float)

for j in range(values.shape[1]):

    out[:, j] = np.interp(xq, x, values[:, j])

return out

# ===== 3) Extraer burst mas largo por dia y resamplear =====
df = df.set_index('fecha')

X_list, dias_list, dur_list = [], [], []

for day, g in df.groupby(pd.Grouper(freq='D')):

    if len(g) == 0:

        continue

    times = g.index.to_series().sort_values()

    bursts = find_bursts(times.values, max_gap_sec=MAX_GAP_SEC,

        min_points=MIN_BURST_POINTS)

    if not bursts:

        continue

    s, e = max(bursts, key=lambda ab: ab[1]-ab[0])

    g_b = g.iloc[s:e].copy()

    t_rel = (g_b.index - g_b.index[0]).total_seconds().values

```

```

V = g_b[expected_cols].values
V_rs = resample_burst(t_rel, V, n_points=RESAMPLED_POINTS)
X_list.append(V_rs)
dias_list.append(day.normalize())
dur_list.append(t_rel.max() if len(t_rel) else 0.0)

if len(X_list) == 0:
    raise ValueError("No se encontro ninguna rafaga valida.")

X = np.stack(X_list, axis=0) # (n_days, RESAMPLED_POINTS, n_feat)
n_days, n_t, n_feat = X.shape
print(f"Rafagas: dias={n_days}, puntos={n_t}, variables={n_feat}")

# ===== 4) Escalado =====
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X.reshape(-1, n_feat)).reshape(n_days, n_t,
    n_feat)

# ===== 5) Clustering con DTW =====
X_tsl = to_time_series_dataset(X_scaled)

km = TimeSeriesKMeans(
    n_clusters=N_CLUSTERS,
    metric="dtw",

```

```

        max_iter=10,
        random_state=0,
        n_init=2
    )

    labels = km.fit_predict(X_tsl)

    clusters_por_dia = pd.DataFrame({
        "dia": pd.to_datetime( dias_list ),
        "cluster": labels ,
        "duracion_seg": dur_list
    }).sort_values("dia").reset_index(drop=True)

    print("\nAsignacion de cluster por dia (primeros 12):")
    print(clusters_por_dia)
    print("\nConteo por cluster:")
    print(clusters_por_dia[ 'cluster' ].value_counts().sort_index())

    # ===== 6) Visualizacion de centroides =====
    time_rel_axis = np.linspace(0, 1, RESAMPLED_POINTS)
    fig, axes = plt.subplots(N_CLUSTERS, 1, figsize=(9, 3 * N_CLUSTERS),
        sharex=True)
    fig.suptitle("Centroids for clustering", fontsize=16)

```

```

if N_CLUSTERS == 1:
    axes = [axes]

for k in range(N_CLUSTERS):
    centroid = km.cluster_centers_[k]
    for j, cat in enumerate(CATEGORIAS):
        idxs = [expected_cols.index(f"{cat}-{s}") for s in SUFIJOS]
        mean_ts = centroid[:, j] if n_feat == 2 else centroid[:, idxs].
            mean(axis=1)
        axes[k].plot( time_rel_axis , mean_ts, label=f"{cat}")
    axes[k].set_title (f"Centroid of cluster {k}")
    axes[k].set_xlabel ("Relative time in burst (0–1)")
    axes[k].grid(True, linestyle='--', alpha=0.4)
    axes[k].legend()

plt.tight_layout ()
plt.tight_layout (rect=[0, 0, 1, 0.98])
plt.show()

return clusters_por_dia

```
