

DISCRIMINACIÓN DEL RUIDO DE FONDO EN MUOGRAFÍA USANDO
TÉCNICAS DE APRENDIZAJE AUTOMATIZADO

ALEJANDRO RAMIREZ MUÑOZ
DAVID VILLABONA ARDILA

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA

2021

DISCRIMINACIÓN DEL RUIDO DE FONDO EN MUOGRAFÍA USANDO
TÉCNICAS DE MACHINE LEARNING

ALEJANDRO RAMIREZ MUÑOZ
DAVID VILLABONA ARDILA

Trabajo de Grado para optar al título de
Ingeniero de Sistemas.

Director
JESÚS PEÑA RODRÍGUEZ

Codirector
LUIS ALBERTO NÚÑEZ DE VILLAVICENCIO MARTÍNEZ
FABIO MARTÍNEZ CARILLO

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICOMECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA

2021

Después de un potente período de largos meses, hoy es el día en el que escribo este apartado de agradecimientos para concluir mi proyecto de fin de grado. Ha sido una época de aprendizaje intenso no solo en el campo académico sino también a nivel personal. Escribir este trabajo ha tenido un gran impacto en mi persona y por ello me gustaría dar las gracias a todas aquellas personas que me han ayudado y apoyado durante este proceso. En primer lugar, me gustaría agradecer a mis compañeros del grupo HALLEY por su colaboración. Me han brindado un gran apoyo en la parte operativa y siempre han estado ahí para ayudarme cuando lo he necesitado. Particularmente me gustaría nombrar a mi director del proyecto Jesús Peña Rodríguez. Me gustaría agradecer tu cooperación y todas las oportunidades que me has dado durante la investigación sobre mi actual

proyecto.

Además, me gustaría dar las gracias a mi compañero David Villabona Ardila por su valiosa ayuda. Definitivamente me ofreció todas las herramientas necesarias para completar nuestro proyecto de fin de grado de forma satisfactoria.

También me gustaría agradecer a mi padre y a mi madre por sus sabios consejos y su comprensión. Siempre han estado ahí cuando los he necesitado. Finalmente, mis amigos. No solo han estado a mi lado para apoyarnos entre nosotros en los momentos más complicados (que hay por montones), sino que también hemos tenido conversaciones sobre otras cosas no relacionadas con universidades y artículos académicos que me llevaron a ser lo hoy soy y motivarme de la mejor manera para seguir adelante.

¡Muchas gracias a todos!

Alejandro Ramirez Muños.

Este ciclo de culminación repercute para todo estudiante sentimientos de nostalgia, que a la vez mezclados con alegría permiten entender como la universalidad es importante en el crecimiento personal. Porque es esa universidad la que permite, en primer lugar, crear mi opinión y respetar la del otro. Cuando miro hacia atrás debo estar agradecido con esta oportunidad de tener otro estilo de vida y tener otras oportunidades a nivel personal y social. Importante el acompañamiento de mis profesores, compañeros, amigos y todas esas personas que estuvieron cerca del proceso. El agradecimiento primero a mi familia, mis padres y mi hermana siempre están apoyándome en todo, gracias por su cariño y comprensión infinita. El grupo Halley por abrirnos las puertas y su constante apoyo. También nuestro director de proyecto Jesús, por su alta disposición y cooperación. A mi compañero Alejandro ramirez Muñoz, quien fue importante en todo este ciclo, del cual aprendí mucho y siempre estuvo presente. En general mucho aprendizaje y agradecimientos, de todas esas experiencias donde muchas personas dejaron huella, y me hicieron crecer en muchos ámbitos desconocidos.

David Villabona Ardila

CONTENIDO

	pág.
INTRODUCCIÓN	15
1. MUOGRAFÍA	17
1.1. MUONES ATMOSFÉRICOS	18
1.2. FUENTES DE RUIDO EN MUOGRAFÍA	18
1.2.1. Dispersión de muones de baja energía	19
1.2.2. Partículas de trayectoria inversa	20
1.2.3. Componente electromagnética	22
1.2.4. Incidencia múltiple de partículas	24
1.3. ELIMINACIÓN DE RUIDO EN MUOGRAFÍA	25
1.3.1. Métodos activos	26
1.3.2. Métodos pasivos	29
2. EL DETECTOR MUTE	30
2.1. FLUJO DE DATOS	31
3. CLASIFICACIÓN SUPERVISADA DE RUIDO EN MUOGRAFÍA POR PÉRDIDA DE ENERGÍA.	32
3.1. EXTRACCIÓN DE CARACTERÍSTICAS.	32
3.2. DISCRIMINACIÓN ENTRE MUONES, COMPONENTE ELECTROMAGNÉTICA Y MULTI-PARTÍCULA.	32
3.3. MINERÍA DE DATOS	33
3.4. MODELO DE MEZCLA DE GAUSSIANAS Y PARAMETRIZACIÓN	34
3.5. CONSTRUCCIÓN DE GAUSSIANAS Y ETIQUETADO	36

3.6. CLASIFICACIÓN SUPERVISADA, DIVISIÓN Y ENTRENAMIENTO	37
4. CLASIFICACIÓN DEL RUIDO COMBINACIONAL POR MEDIO DEL TIEMPO DE VUELO USANDO ALGORITMOS DE APRENDIZAJE NO SUPERVISADO	41
4.1. EVENTOS DE PARTÍCULA SIMPLE, CORRELACIONADOS Y NO-CORRELACIONADOS TENIENDO EN CUENTA EL TOF	42
4.1.1. Tratamiento y análisis de datos.	42
4.2. CLASIFICACIÓN Y SEGMENTACIÓN.	43
5. CLASIFICACIÓN NO SUPERVISADA DE LOS MUONES DE BAJO MOMENTUM.	46
5.1. MANEJO DE VARIABLES	46
5.2. ESTIMACIÓN DE MOMENTUM	46
5.3. CLASIFICACIÓN	47
5.4. ANÁLISIS DE RESULTADOS.	48
6. CONCLUSIONES	51
BIBLIOGRAFÍA	53

LISTA DE FIGURAS

	pág.
Figura 1. Componentes de la lluvia aérea extendida(EAS)	19
Figura 2. Dispersión de los muones incidentes de baja energía sobre la superficie	20
Figura 3. Evento falso debido a un muón que incide por la parte posterior del detector	21
Figura 4. Resultado de la tomografía de alta definición en el Monte La Soufriere, sin corrección de flujo ascendente (Arriba) y con corrección de flujo ascendente (Abajo). Densidad de roca en $g.cm^{-3}$	22
Figura 5. Detección de un evento falso debido a la incidencia de un e^- generado en una EAS entre el objeto escaneado y el detector	23
Figura 6. Electrones generados por una EAS, causando un evento de múltiple partícula que imita un evento causado por una partícula única.	24
Figura 7. (Arriba a la izquierda) Muón falso. Causado por un hadrón de EAS (Arriba a la derecha) Ruido combinatorio. (Abajo a la izquierda) Muón de bajo momentum. (Abajo a la derecha) Muón de de trayectoria inversa	26
Figura 8. El sistema de observación muográfica basado en MWPC (mMOS). La vista esquemática de mMOS que consta de siete cámaras proporcionales multi-hilo y cinco placas de blindaje de plomo con un grosor de 2 cm cada una	27

- Figura 9. Distribución del ToF en función del ángulo cenital para los datos registrados en el volcán La Soufriere. El horizonte está representado por la línea discontinua. Las elipses sólidas azul y roja muestran respectivamente los eventos hacia atrás ($\alpha_B < 0$ y $\Delta t < 0$) y hacia adelante ($\alpha_F < 0$ y $\Delta t > 0$) correspondientes a los flujos descendentes. Las elipses discontinuas muestran eventos correspondientes a muones ascendentes desde adelante (elipse roja, $\alpha_B < 0$ y $\Delta t > 0$) y hacia atrás (elipse azul, $\alpha_F < 0$ y $\Delta t < 0$) 28
- Figura 10. Vista lateral del detector. El WCD contiene $1.7m^3$ de agua y se ubica sobre el soporte de elevación. El hodoscopio compuesto por dos paneles centelladores de $120\text{ cm} \times 120\text{ cm}$ ubicados dentro de cajas metálicas que los protegen de la contaminación lumínica y de la humedad. 30
- Figura 11. Histograma de carga para un registro de datos de una hora del WCD de MuTe. (arriba). Datos del WCD de eventos que han atravesado el hodoscopio (abajo). 34
- Figura 12. Datos de Carga crudos. ARRIBA: Datos completos de carga. ABAJO: Datos acotados por el umbral ($\text{Carga} < 800$) 35
- Figura 13. Histograma de frecuencia con los datos y la función de densidad de probabilidad ajustada por GMM. Con las medias calculadas, para apreciar el ajuste a los datos. 36
- Figura 14. Las tres componentes: la muónica (verde), la electromagnética (azul) y multi-partícula (roja), representadas por su respectiva función de densidad de probabilidad. 37
- Figura 15. Separación por *k-fold*, se tomaron 20 *splits* y en cada uno de ellos se dividió unos para prueba y otros para entrenamientos, en cada uno de los 20 grupos se cambió la distribución para pruebas y entrenamientos . 38

Figura 16. Resultados del etiquetado de los datos por el clasificados y ajuste de las distribuciones.	39
Figura 17. Partículas incidentes que cruzan los paneles. Estos se clasifican en correlacionados (verde oscuro), no correlacionados (verde claro), y eventos simples (rojo).	41
Figura 18. Tiempo de vuelo de cada evento. Un umbral en 300ns (línea negra) permite dividir los eventos correlacionados (<300ns) de los no correlacionados (>300ns).	42
Figura 19. Histograma de frecuencia de los datos del tiempo de vuelo.	43
Figura 20. Tiempo de vuelo en orden logarítmico. Se distinguen los eventos de partícula simple y correlacionadas con un ToF menor a 300ns, mientras que los eventos no correlacionados tienen ToF mayores a 300 ns	44
Figura 21. $ToF < 300ns$. Eventos correlacionados. Los datos entregados por el detector y el modelo ajustado, muestran que los eventos se concentran en un ToF menor a 40ns.	45
Figura 22. ToF vs momentum. El 54% de las partículas tiene momentum mayor a 1GeV/c (línea púrpura).	47
Figura 23. ToF vs momento entregados por los clasificadores <i>AgglomerativeClustering</i> y <i>Kmeans</i> , se muestra la comparativa de 2 a 4 clusters, con el umbral en 1GeV/c. (línea roja)	48
Figura 24. ToF vs momento entregados por los clasificadores <i>Agglomerative Clustering</i> y <i>Kmeans</i> , se muestra la comparativa de 2 a 4 clusters. Muestra el porcentaje de cada grupo en cada cluster.	49

LISTA DE TABLAS

	pág.
Tabla 1. Matriz de confusión del <i>naive gaussian</i> . Donde las columnas 1, 2 y 3 son valores reales que representan las componentes: electromagnética, muónica y multi-partícula. Las filas 1, 2 y 3 son los valores predichos que definen las componentes electromagnética, muónica y multi-partícula.	40
Tabla 2. Métricas de precisión definidas para cada componente, calculadas mediante la matriz de confusión, mostrada en la Tabla 1	40

GLOSARIO

Se se presentan algunos términos utilizados en el trabajo desarrollado:

CR: Radiación cósmica, son partículas subatómicas de origen extraterrestre que impactan nuestro planeta. Este flujo está compuesto por aproximadamente 90 % de protones y 9 % de partículas alfa, el resto son electrones e iones pesados¹.

EAS: Por sus siglas en inglés Extensive Air Shower, se le llama a la cascada de millones de partículas generadas cuando un rayo cósmico interactúa con los átomos terrestres.

Muón: El muón es una partícula elemental, cuya masa es ~ 200 veces la del electrón: $105,6 \text{ MeV}/c^2$.

ToF: Tiempo de vuelo.

MuTe: Telescopio de Muones, proyecto el cual lleva a cabo un estudio muográfico de volcanes en Colombia ².

WCD: Detectores Cherenkov de agua. Es un detector de partículas cargadas conformado por un volumen de agua ultra pura y un elemento sensible (foto-multiplicador)

GMM: Modelos probabilísticos que representan subpoblaciones normalmente distribuidas dentro de una población general.

PDF: Función de densidad de probabilidad.

¹ S Procureur. "Muon imaging: Principles, technologies and applications". En: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 878 (2018), págs. 169-179.

² <http://halley.uis.edu.co/fuego/>

RESUMEN

TÍTULO: DISCRIMINACIÓN DEL RUIDO DE FONDO EN MUOGRAFÍA USANDO TÉCNICAS DE APRENDIZAJE AUTOMATIZADO. *

AUTOR: ALEJANDRO RAMIREZ MUÑOZ, DAVID VILLABONA ARDILA **

PALABRAS CLAVE: Rayos cósmicos, Muografía, Aprendizaje automatizado, ruido de fondo.

DESCRIPCIÓN:

La muografía es una técnica no-invasiva que se utiliza para escanear grandes estructuras antrópicas o naturales. Su principio de funcionamiento consiste en la medición del flujo de muones que cruzan la estructura en diferentes direcciones. Esta técnica tiene aplicaciones en campos tales como: mediciones subterráneas, arqueología, detección de materiales ocultos en contenedores, reactores y residuos nucleares.

Esta técnica se ve afectada por una subestimación de la densidad del objeto, producto del ruido de fondo (falsos-positivos) que se pueden clasificar en: partículas cargadas procedentes de lluvias aéreas extensas (EAS), las partículas que inciden desde la parte trasera del detector, los muones de baja energía que son dispersados por la superficie del volcán y eventos de múltiple partícula.

Para la eliminación del ruido se han desarrollado técnicas pasivas como la instalación de paneles absorbentes, para filtrar las partículas de baja energía y el aumento de la cantidad de paneles sensibles, para disminuir la probabilidad de detectar eventos combinacionales. En la actualidad se plantea la eliminación del ruido de fondo con sistemas ToF e identificación de partículas¹.

En este trabajo se desarrolla un clasificador de aprendizaje automatizado que disminuya las princi-

* Trabajo de grado

** Facultad de Ingeniería Físico-Mecánica. Escuela de Ingeniería de Sistemas e Informática, Director: Jesús Peña Rodríguez

¹ **muones que impactan en el detector por la parte de posterior.**

pales fuentes de ruido que pueden afectar la muografía, basados en los datos del detector MuTe. El proyecto se divide en 2 partes:

- En la primera instancia se desarrolla un clasificador de aprendizaje supervisado para separar la componente electromagnética, muónica y de múltiple partícula.
- En la segunda parte se desarrolla un clasificador de aprendizaje no-supervisado el cual discrimina los muones de bajo momentum ($< 1 \text{ GeV}/c$).

ABSTRACT

TITLE: DISCRIMINATION OF BACKGROUND NOISE IN MUOGRAPHY USING MACHINE LEARNING TECHNIQUES *

AUTHOR: ALEJANDRO RAMIREZ MUÑOZ, DAVID VILLABONA ARDILA **

KEYWORDS: Cosmic rays, Muography, machine learning, background noise, MuTe.

DESCRIPTION:

Muography is a non-invasive technique used to scan large anthropic or natural structures. Its operating principle consists of measuring the flux of muons that cross the structure in different directions. This technique has applications in fields such as: underground measurements, archeology, detection of hidden materials in containers, reactors and nuclear waste.

This technique is affected by an underestimation of the object density, as a consequence of background noise (false-positives) that can be classified into: charged particles from extensive air showers (EAS), the incident particles from the rear of the detector, low energy muons that are scattered by the surface of the volcano and multiple particle events.

Techniques have been developed for the elimination of noise, such as the installation of absorbent panels, to filter low-energy particles and increasing the number of sensitive panels, to reduce the probability of detecting combinational events. At present, the elimination of background noise with ToF systems and particle identification is being considered¹.

In this work, an automated learning classifier is developed that reduces the main sources of noise that can affect muography, based on data from the MuTe detector.

* Bachelor Thesis

** Faculty of Physical-Mechanical Engineering. School of Systems Engineering and Informatics. Director: Jesús Peña Rodríguez

¹ **muons that impact the detector from the rear.**

INTRODUCCIÓN

La muografía se ha enfocado principalmente en el estudio de volcanes y los fenómenos relacionados. Proyectos como el telescopio DIAPHANE², ubicado en el volcán La Soufrière, analiza las variaciones del contenido interno de líquido/vapor relacionados con su dinámica hidrotermal³. Esto es posible teniendo en cuenta que el flujo de muones que atraviesa la estructura volcánica varía dependiendo de la densidad del material: a mayor densidad, el flujo será menor y viceversa.

Debido a que el flujo de muones a ángulos de observación típicos de la muografía es bajo, los niveles de ruido pueden generar una sobre-estimación del flujo penetrante y como consecuencia una sub-estimación en la densidad del objeto estudiado⁴. Los principales actores del ruido de fondo son los componentes electromagnéticos (electrones, positrones y rayos gama) de las EAS, muones de bajo momentum (< 1 GeV/c) que desvían su trayectoria inicial por interacción con objetos externos, muones que ingresan desde la parte posterior del detector y eventos combinacionales de múltiple-partícula⁵.

² Jacques Marteau y col. "DIAPHANE: muon tomography applied to volcanoes, civil engineering, archaeology". En: *Journal of Instrumentation* 12.02 (2017), pág. C02008.

³ H Gómez y col. "Forward scattering effects on muon imaging". En: *Journal of Instrumentation* 12.12 (2017), P12018.

⁴ Taro Kusagaya e Hiroyuki KM Tanaka. "Muographic imaging with a multi-layered telescope and its application to the study of the subsurface structure of a volcano". En: *Proceedings of the Japan Academy, Series B* 91.9 (2015), págs. 501-510.

⁵ R Nishiyama, S Miyamoto y N Naganawa. "Experimental study of source of background noise in muon radiography using emulsion film detectors". En: *Geoscientific Instrumentation, Methods and Data Systems* 3.1 (2014), págs. 29-39.

Durante el desarrollo de este proyecto, se propone una solución para la disminución del ruido de fondo en muografía usando técnicas de aprendizaje automatizado de forma sistemática, usando los datos del proyecto MuTe ⁶. Por medio de un algoritmo de clasificación se abordará el problema desde el análisis de los datos, haciendo en primera instancia un clasificador de aprendizaje supervisado para separar la componente electromagnética de la muónica, guiados por las distribuciones obtenidas por el detector Cherenkov de agua de MuTe ⁶. En la segunda parte se desarrolla un clasificador de aprendizaje no-supervisado, para separar los muones de bajo momentum ($< 1 \text{ GeV}/c$) contra los muones que tienen baja probabilidad de desviación.

⁶ Adriana y Sanabria Gómez José D y Núñez Luis A y Sierra-Porta David y Asorey Hernán Peña-Rodríguez Jesús y Vásquez-Ramírez. "Calibración y primeras mediciones de MuTe: un telescopio muón híbrido para estructuras geológicas". En: *arXiv preprint arXiv: 1909.09732* ().

1. MUOGRAFÍA

El estudio de grandes estructuras mediante muografía se basa en el principio de la radiografía, es decir, la radiación a la que se expone un objeto es absorbida parcialmente dependiendo de su densidad. La muografía utiliza como fuente de radiación los muones atmosféricos creados por la interacción de los rayos cósmicos con los átomos que componen la atmósfera terrestre. Los muones interactúan con los átomos que conforman el objeto escaneado, creando procesos de pérdida de energía y dispersión múltiple ⁷.

Todas las aplicaciones de la muografía se basan en la atenuación del flujo de muones al atravesar un objetivo, aprovechando el flujo natural de muones producido por las interacciones de los rayos cósmicos en la atmósfera⁸.

La primera aparición práctica de muografía se remonta a la década de 1950, cuando George estudió la viabilidad de emplear un telescopio Geiger para inferir el espesor del hielo sobre un túnel en una mina australiana ⁹.

Hoy en día, la vulcanología es el área que más implementa la muografía, ya que esta proporciona información del comportamiento dinámico de dichas formaciones geológicas de manera no invasiva y con una resolución espacial de decenas de metros¹⁰.

⁷ S Procureur. "Muon imaging: Principles, technologies and applications". En: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 878 (2018), págs. 169-179.

⁸ Lorenzo Bonechi, Raffaello D'Alessandro y Andrea Giammanco. "Atmospheric muons as an imaging tool". En: *Reviews in Physics* (2020), pág. 100038.

⁹ EP George. "Cosmic rays measure overburden of tunnel". En: *Commonwealth Engineer* 455 (1955).

¹⁰ K. Nagamine y col. "Probing the inner structure of blast furnaces by cosmic-ray muon radio-

1.1. MUONES ATMOSFÉRICOS

Cuando la CR llega a la Tierra colisiona con los núcleos atómicos en la atmósfera. Produciendo nuevas partículas que posteriormente chocan con otros núcleos, creando una nueva generación de partículas que continúan el proceso. La cascada de partículas resultante es llamada lluvia aérea extensa (EAS por sus siglas en inglés). Las EAS pueden llegar al nivel del suelo extendiéndose sobre grandes áreas ($\approx \text{km}^2$).

Las EAS están conformadas por tres componentes: la **electromagnética**, la **hadronica** y la componente **muónica** como se muestra en la Fig. 1. La interacción de la CR en la atmósfera produce cascadas de partículas más ligeras como los piones (π^+ , π^0 , π^-) y kaones (K^+ , K^0 , K^-) los cuales se descomponen principalmente en muones ¹¹. Los muones, en comparación con otras partículas inestables, pueden desplazarse grandes distancias en la atmósfera sin decaer. Los muones conforman más de la mitad de la radiación cósmica a nivel del mar, siendo el resto principalmente electrones, positrones y fotones provenientes de las EAS.

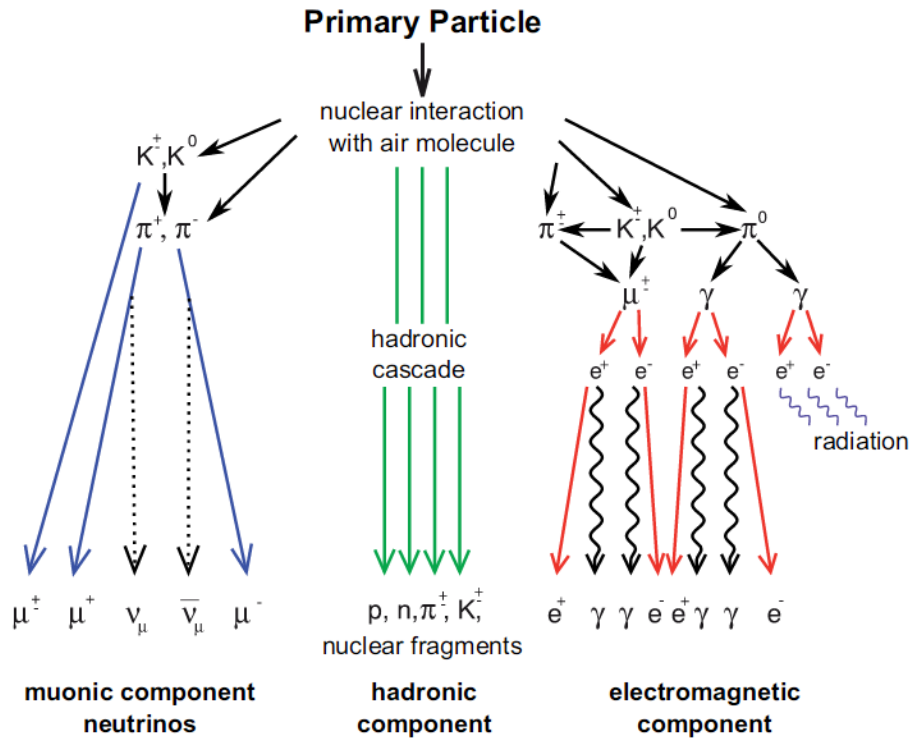
1.2. FUENTES DE RUIDO EN MUOGRAFÍA

La muografía tiene principalmente tres fuentes de ruido: los muones de baja energía dispersados por la superficie de la estructura escaneada, las partículas que ingresan por la parte posterior del detector y la componente electromagnética de las EAS. Si el ruido de fondo es dominante en la observación, la densidad estimada será menor que la densidad real.

graphy". En: *Proceedings of the Japan Academy, Series B* 81.7 (2005), págs. 257-260. DOI: 10.2183/pjab.81.257.

¹¹ Masaharu Tanabashi y col. "Review of particle physics". En: *Physical Review D* 98.3 (2018), pág. 030001.

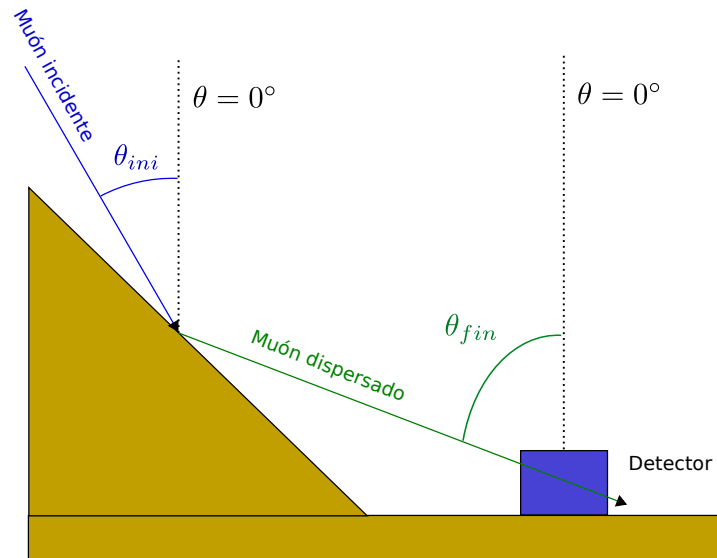
Figura 1. Componentes de la lluvia aérea extendida (EAS): hadrónica (verde), electromagnética (roja) y muónica (azul).



Fuente. A. Haungs y col. "KCDC — The KASCADE Cosmic-ray Data Centre". En: *Journal of Physics: Conference Series* 632.1 (2015), pág. 012011.

1.2.1. Dispersión de muones de baja energía El flujo de muones a grandes ángulos cenitales es bajo y los muones dispersados en la estructura escaneada pueden volverse fácilmente dominantes. En este caso, la dirección del muón incidente varía debido a la dispersión múltiple de Coulomb generada por su interacción con la estructura escaneada ⁸. La dispersión angular causa un desenfoque de la imagen final de densidad, afectando su resolución espacial con la consiguiente pérdida de detalles. Ver Fig. 2.

Figura 2. Dispersión de los muones incidentes de baja energía sobre la superficie. El ángulo de incidencia del muón θ_{ini} varía debido a su interacción con el material que compone la estructura resultando en un ángulo dispersado θ_{fin} .

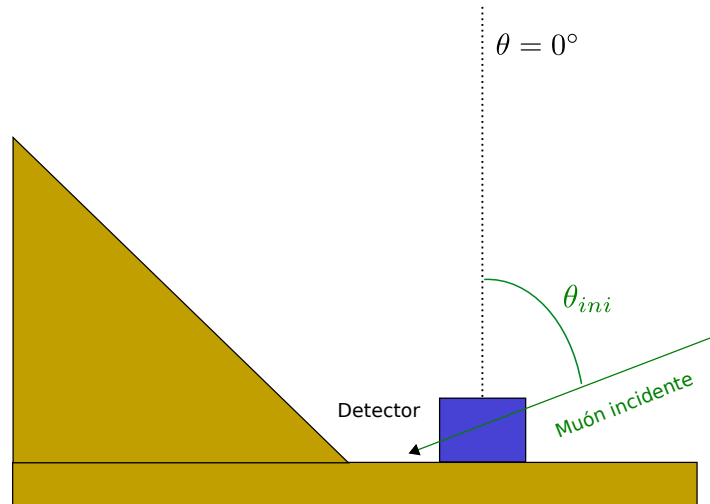


Fuente. Jesús Peña Rodríguez. “Diseño y calibración de un telescopio de muones híbrido para estudios vulcanológicos.” Tesis doct. Facultad de Ciencias. Escuela de Física. Universidad Industrial de Santander, 2020.

1.2.2. Partículas de trayectoria inversa Otra fuente de contaminación en la muografía son las partículas que impactan en el detector desde la parte posterior¹², creando trayectorias similares a los muones provenientes desde la estructura escaneada. Ver Fig. 3.

¹² R. Nishiyama, S. Miyamoto y N. Naganawa. “Experimental study of source of background noise in muon radiography using emulsion film detectors”. En: *Geoscientific Instrumentation, Methods and Data Systems* 3.1 (2014), págs. 29-39. DOI: 10.5194/gi-3-29-2014.

Figura 3. Evento falso debido a un muón que incide por la parte posterior del detector.

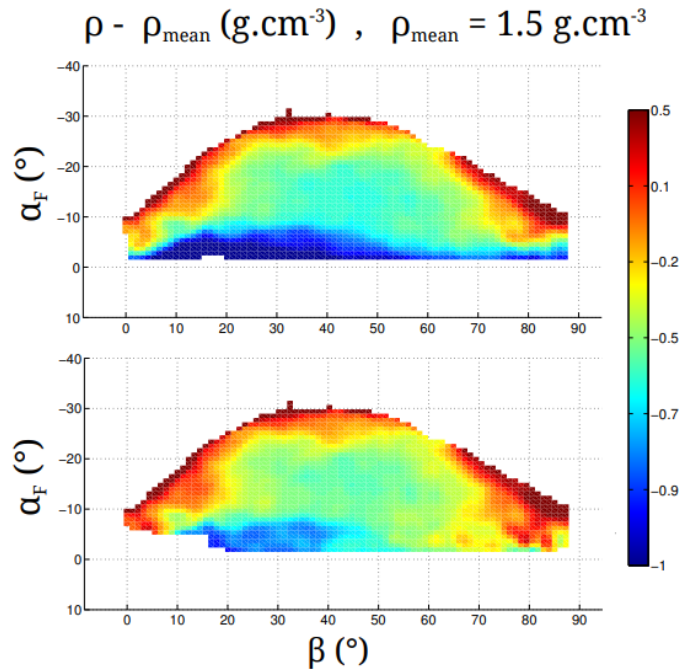


Fuente. Jesús Peña Rodríguez. “Diseño y calibración de un telescopio de muones híbrido para estudios vulcanológicos.” Tesis doct. Facultad de Ciencias. Escuela de Física. Universidad Industrial de Santander, 2020.

Los flujos traseros se detectan cuando la parte posterior del telescopio está expuesta a un amplio volumen de atmósfera ubicado por debajo del nivel de medición, como se observa en la Fig. 4. El análisis de datos discutido por Jourde et. al. ¹³, demuestra la existencia de un flujo de mounes traseros cuyas trayectorias podrían confundirse con la de los mounes descendentes que cruzan el volcán.

¹³ Kevin Jourde y col. “Experimental detection of upward going cosmic particles and consequences for correction of density radiography of volcanoes”. En: *Geophysical Research Letters* 40.24 (2013), págs. 6334-6339.

Figura 4. Resultado de la tomografía de alta definición en el Monte La Soufriere, sin corrección de flujo ascendente (Arriba) y con corrección de flujo ascendente (Abajo). Densidad de roca en $g.cm^{-3}$

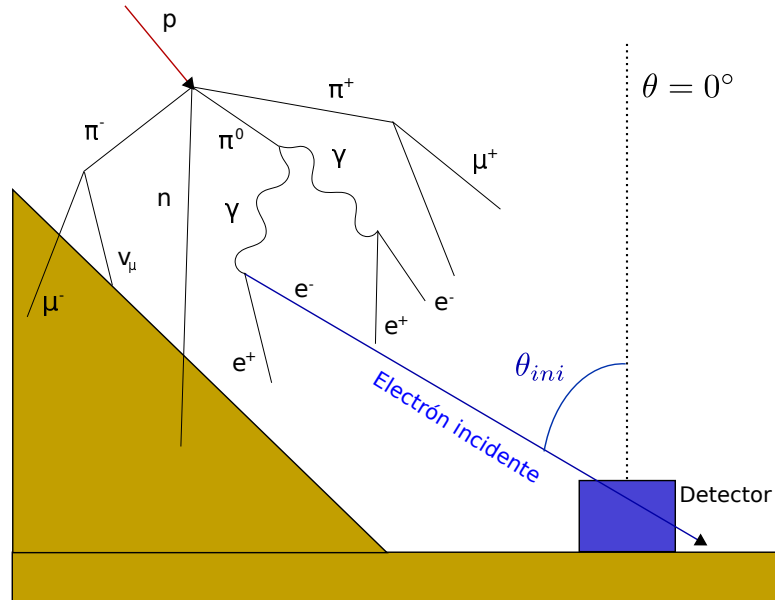


Fuente. Kevin Jourde y col. "Effects of upward-going cosmic muons on density radiography of volcanoes". En: *arXiv preprint arXiv:1307.6758* (2013).

Se reporta que los muones posteriores superan a los frontales dos a uno, respectivamente ¹³. El ruido inverso disminuye al aumentar el ángulo de elevación del telescopio.

1.2.3. Componente electromagnética Otra fuente de contaminación en la muografía ocurre a las partículas secundarias generadas por EAS entre la estructura y el detector. El flujo de PS a nivel del suelo está conformado principalmente por μ^{\pm}, e^{\pm} y γ . Ver Fig. 5.

Figura 5. Detección de un evento falso debido a la incidencia de un e^- generado en una EAS entre el objeto escaneado y el detector.



Fuente. Jesús Peña Rodríguez. “Diseño y calibración de un telescopio de muones híbrido para estudios vulcanológicos.” Tesis doct. Facultad de Ciencias. Escuela de Física. Universidad Industrial de Santander, 2020.

Las partículas secundarias generan falsa información de dos maneras:

- Mediante la coincidencia accidental de dos o más partículas incidentes en el detector lo cual imita una trayectoria generada por un muón ¹⁴.
- Un electrón/positrón con energía suficiente para atravesar todo el detector.

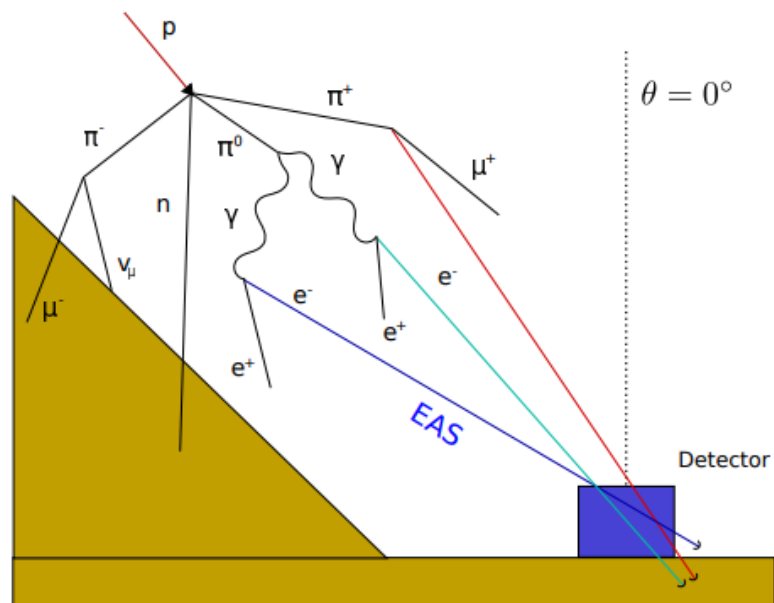
Teniendo en cuenta que la muografía calcula la distribución densidad interna de la estructura dependiendo del flujo diferencial de los mounes que la atraviesan, un

¹⁴ T. Kusagaya y H. K. Tanaka. “Muographic imaging with a multi-layered telescope and its application to the study of the subsurface structure of a volcano”. En: *Proceedings of the Japan Academy, Series B* 91.9 (2015), págs. 501-510. DOI: 10.2183/pjab.91.501.

aumento en el flujo registrado debido a los eventos falsos repercute en la medición de la densidad de la estructura. ¹².

1.2.4. Incidencia múltiple de partículas El ruido combinacional es causado por varias partículas individuales que golpean los paneles sensibles, confundiendo al detector como si fuera un evento generado por una única partícula (Ver Fig. 6). Estos eventos tienen la singularidad que las partículas están correlacionadas temporalmente, ya que son producidas por la misma EAS ¹⁵.

Figura 6. Electrones generados por una EAS, causando un evento de múltiple partícula que imita un evento causado por una partícula única.



Fuente. Jesús Peña Rodríguez. “Diseño y calibración de un telescopio de muones híbrido para estudios vulcanológicos.” Tesis doct. Facultad de Ciencias. Escuela de Física. Universidad Industrial de Santander, 2020.

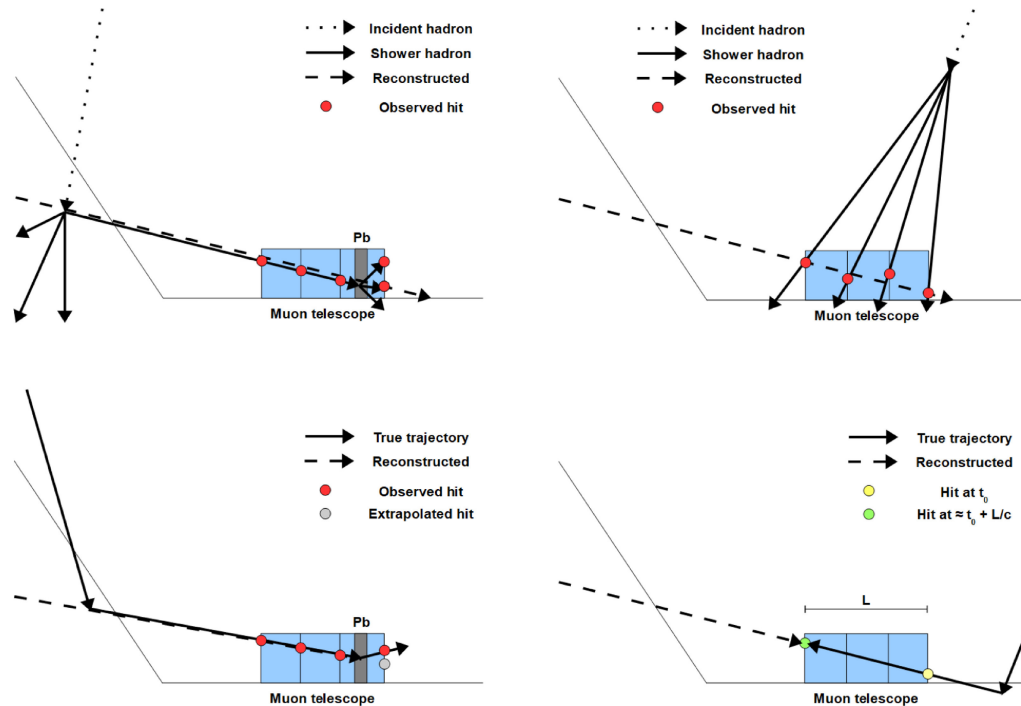
¹⁵ Jesús Peña Rodríguez. “Diseño y calibración de un telescopio de muones híbrido para estudios vulcanológicos.” Tesis doct. Facultad de Ciencias. Escuela de Física. Universidad Industrial de Santander, 2020.

El retardo en promedio de los muones para una distancia al centro de la EAS menor a 1km es < 100 ns, se puede jugar con esos datos para combatir con el ruido combinacional. También se puede reducir aumentando el número de planos sensibles.

1.3. ELIMINACIÓN DE RUIDO EN MUOGRAFÍA

Debido a los altos niveles de ruido en la muografía generados por las diferentes fuentes antes mencionadas, ver Fig. 7, la eliminación de ruido en muografía se convierte en un campo para indagar. Se han creado distintos métodos para la disminución del ruido en la muografía, algunos que sobresalen son la adicción de capas del material absorbente, el aumento de capas sensibles y la medición del ToF. Estos se pueden clasificar en métodos activos y pasivos.

Figura 7. (Arriba a la izquierda) Muón falso. Causado por un hadrón de EAS (Arriba a la derecha) Ruido combinatorio. (Abajo a la izquierda) Muón de bajo momentum. (Abajo a la derecha) Muón de de trayectoria inversa



Fuente. Lorenzo Bonechi, Raffaello D'Alessandro y Andrea Giammanco. "Atmospheric muons as an imaging tool". En: *Reviews in Physics* (2020), pág. 100038.

1.3.1. Métodos activos Estos métodos buscan la eliminación de fuentes de ruido mediante el análisis de los datos entregados, para encontrar un patrón que muestre el comportamiento de los eventos falsos.

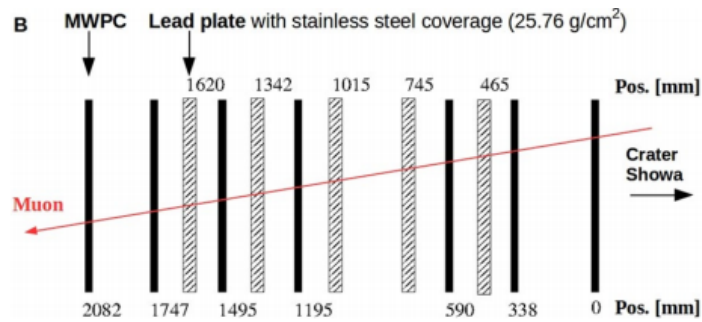
L. Oláh et. al. ¹⁶ implementaron siete detectores tipo cámara proporcional de múltiple hilo (MWPC), con un tamaño de $(80 \times 80 \text{ cm}^2)$, con cinco láminas de protección contra la radiación construidas de plomo, cada placa está oculta dentro de otra caja de

¹⁶ László Oláh y col. "High-definition and low-noise muography of the Sakurajima volcano with gaseous tracking detectors". En: *Scientific reports* 8.1 (2018), págs. 1-13.

acero inoxidable para protegerla mecánicamente y evitar la exposición a la toxicidad del plomo. La Fig. 8 muestra el detector desde una vista lateral.

Debido a la buena resolución espacial, la dispersión de partículas de baja energía en las placas de plomo se puede medir a lo largo de la trayectoria y, por lo tanto, estas partículas de fondo $< 2 \text{ GeVc}^{-1}$ se pueden eliminar de los datos registrados.

Figura 8. El sistema de observación muográfica basado en MWPC (mMOS). La vista esquemática de mMOS que consta de siete cámaras proporcionales multi-hilo y cinco placas de blindaje de plomo con un grosor de 2 cm cada una



Fuente. L. Oláh y col. “High-definition and low-noise muography of the Sakurajima volcano with gaseous tracking detectors”. En: *Scientific Reports* 8.1 (2018). DOI: 10.1038/s41598-018-21423-9.

Dentro de los métodos activos se encuentra la medición del tiempo de vuelo (ToF). EL ToF es el tiempo que tarda una partícula en recorrer una distancia determinada. Con esta estimación se pueden calcular otras variables como velocidad, momento, dirección e identidad. El momento p se relaciona con el ToF mediante:

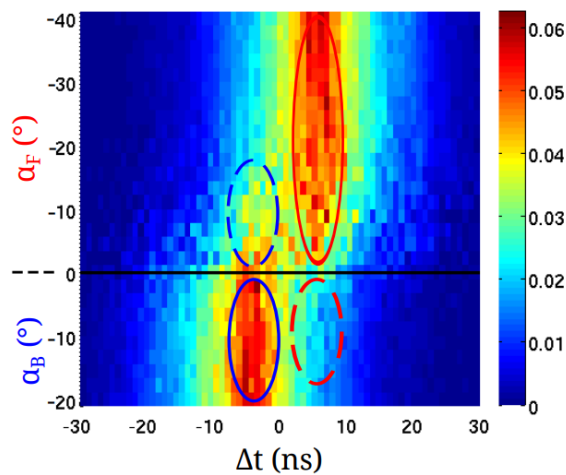
$$p = \frac{m_0 c d}{\sqrt{c^2 t^2 - d^2}} \quad (1)$$

Donde m_0 es la masa de la partícula cargada en reposo, c es la velocidad de la luz, d es la distancia recorrida y t es el tiempo de vuelo.

La Fig. 9 muestra el ToF en función del ángulo cenital, las elipses sólidas identifican

a los muones que se propagan hacia abajo (componente principal del flujo), las elipses discontinuas muestran la existencia de muones con propagación hacia arriba identificadas por el detector. Los resultados para la medición presentada en la figura muestran que la contaminación asciende al 70 % del flujo total, y disminuye al 30 % en ángulos de -10° y en cero por encima de -20° . Estos resultado permiten corregir las imágenes obtenidas por el detector ¹⁷.

Figura 9. Distribución del ToF en función del ángulo cenital para los datos registrados en el volcán La Soufriere. El horizonte está representado por la línea discontinua. Las elipses sólidas azul y roja muestran respectivamente los eventos hacia atrás ($\alpha_B < 0$ y $\Delta t < 0$) y hacia adelante ($\alpha_F < 0$ y $\Delta t > 0$) correspondientes a los flujos descendentes. Las elipses discontinuas muestran eventos correspondientes a muones ascendentes desde adelante (elipse roja, $\alpha_B < 0$ y $\Delta t > 0$) y hacia atrás (elipse azul, $\alpha_F < 0$ y $\Delta t < 0$)



Fuente. Kevin Jourde y col. “Effects of upward-going cosmic muons on density radiography of volcanoes”. En: *arXiv preprint arXiv:1307.6758* (2013).

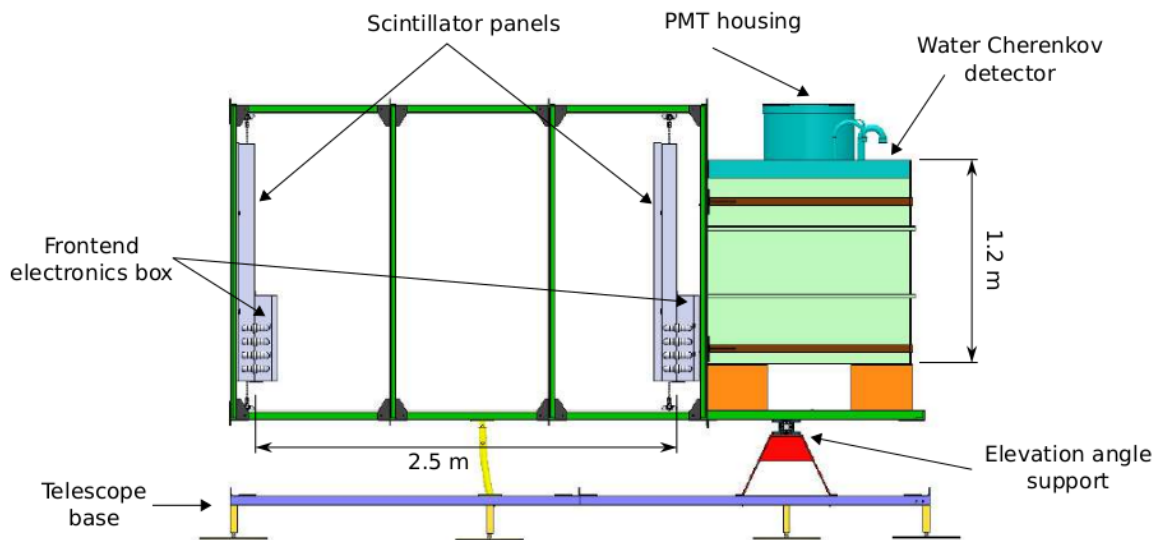
¹⁷ J. Marteau y col. “Implementation of sub-nanosecond time-to-digital convertor in field-programmable gate array: applications to time-of-flight analysis in muon radiography”. En: *Measurement Science and Technology* 25.3 (2014), pág. 035101. DOI: 10.1088/0957-0233/25/3/035101.

1.3.2. Métodos pasivos Este método consiste en adicionar capas de material absorbente entre los paneles del hodoscopio. Lo que se busca es que la partícula que atraviesa el detector pierda toda su energía en el material absorbente. La primera implementación de esta técnica fue hecha por Nagamine et. al. ¹², donde introducen dos placas absorbentes de hierro (40 g/cm^2) dentro del sistema de detección para filtra los muones de baja energía ($< 1\text{GeV}c^{-1}$).

2. EL DETECTOR MUTE

El telescopio de muones (MuTe) es un detector creado para realizar muografía de volcanes en Colombia, ver Fig. 10. Tiene como característica principal separar los muones que ingresan al detector de las fuentes de ruido que afectan la muografía. Aplica dos técnicas para el reconocimiento de partículas: la medición del ToF de partículas incidentes y la pérdida de energía. MuTe esta compuesto por dos detectores: un hodoscopio y un detector Cherenkov de agua (WCD).

Figura 10. Vista lateral del detector. El WCD contiene $1.7m^3$ de agua y se ubica sobre el soporte de elevación. El hodoscopio compuesto por dos paneles centelladores de $120\text{ cm} \times 120\text{ cm}$ ubicados dentro de cajas metálicas que los protegen de la contaminación lumínica y de la humedad.



Fuente. Jesús Peña Rodríguez. “Diseño y calibración de un telescopio de muones híbrido para estudios vulcanológicos.” Tesis doct. Facultad de Ciencias. Escuela de Física. Universidad Industrial de Santander, 2020.

El hodoscopio está conformado por dos paneles sensibles, cada panel contiene 30 barras centelladoras verticales y 30 horizontales, creando una matriz de 900 píxeles. El hodoscopio de MuTe es un dispositivo de conteo de sucesos y su principal función es la estimación del flujo de muones.

El WCD es un tanque de acero que contiene $1,7 \text{ m}^3$ agua y un tubo fotomultiplicador como elemento sensible. Este detector mide la energía recibida por las partículas cargadas y permite diferenciar los eventos registrados en: muones, electrones/positrones y múltiple partícula.

El Sistema ToF de MuTe permite estimar el tiempo que tardan las partículas en atravesar el hodoscopio en una distancia dada y la dirección de estas. Este permite filtrar las partículas que ingresan por la parte posterior. Con el ToF y la energía depositada en el WCD MuTe estima el momentum de las partículas incidentes.

2.1. FLUJO DE DATOS

Los datos adquiridos por MuTe son sincronizados temporalmente por un GPS y guardados en dos discos duros, uno para los datos del hodoscopio y otro para los datos del WCD. Cada hora se crean nuevos archivos que almacenan 1 hora de datos. Los datos almacenados tiene meta-datos de presión atmosférica, temperatura y el consumo energético para hacer correcciones de flujo y evaluar la autonomía del detector. Cabe resaltar que estimación del momentum se hace offline.¹⁵.

3. CLASIFICACIÓN SUPERVISADA DE RUIDO EN MUOGRAFÍA POR PÉRDIDA DE ENERGÍA.

En esta sección se muestra la clasificación de los eventos detectados por MuTe mediante algoritmos de aprendizaje supervisado. En este apartado se hace un recuento de los métodos y propuestas que se emplearon para concluir la primera etapa, con los resultados y los problemas que surgieron durante su desarrollo.

3.1. EXTRACCIÓN DE CARACTERÍSTICAS.

La extracción de características se basa en la parametrización y análisis de los datos suministrados por el WCD por medio de un histograma de carga. A continuación se presentan los pasos empleados:

Algoritmo 1 Extracción de características

- 1: Extraer los datos crudos del detector y posteriormente ordenarlos.
 - 2: Parametrizar los datos para ajustarlo a una distribución probabilística.
 - 3: Implementar un histograma de carga, para identificar y separar la componente muónica de la electromagnética y multipartícula.
 - 4: Encontrar características según los resultados de las variables de entrada de cada componente.
-

3.2. DISCRIMINACIÓN ENTRE MUONES, COMPONENTE ELECTROMAGNÉTICA Y MULTI-PARTÍCULA.

Las características (la carga de las componentes, la variación y la media) se ajustan a diferentes modelos de clasificación como: *naive gaussian*, *support vector machine*, *random forest* y *gaussian mixture model*. Se etiquetan como componente: muónica, electromagnética y multi-partícula. Para ello se emplearon los siguientes pasos:

Algoritmo 2 Discriminación entre muones, electromagnética y multi-partícula.

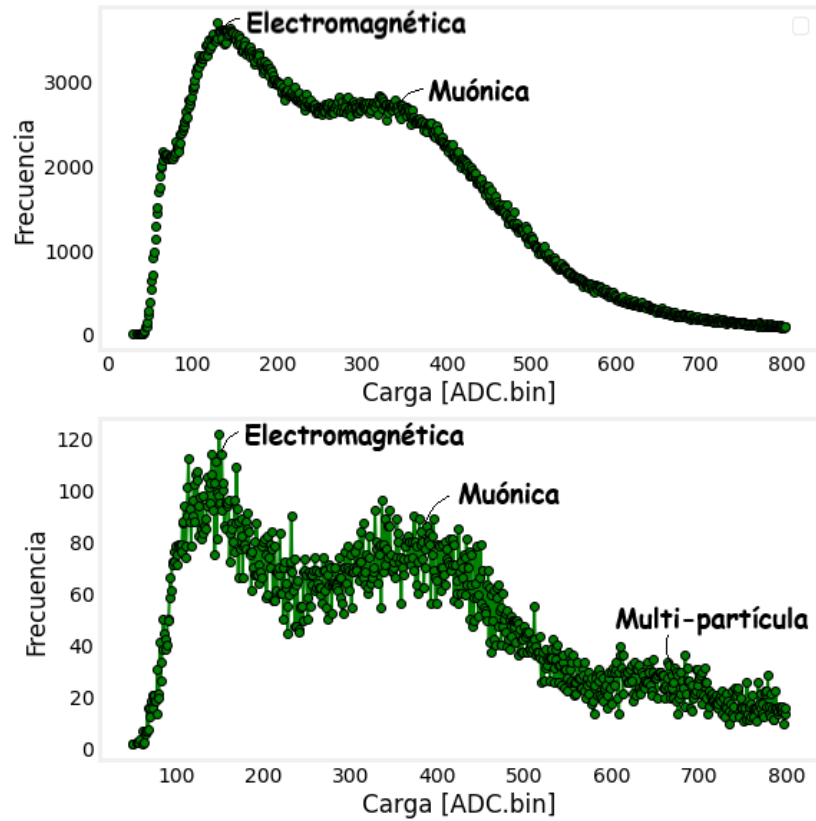
- 1: Definir una función de probabilidad optimizando los valores que ajusten a la curva de la distribución seleccionada.
 - 2: A partir de los valores ajustados con la función de densidad de probabilidad (PDF), analizar independientemente cada componente por medio de la distribución extraída del histograma de carga.
 - 3: Se etiqueta cada tupla de datos que se representan como: 1 para muones, 0 para electromagnético y 2 para multi-partícula.
 - 4: Para el entrenamiento de los datos se emplean los siguientes modelos de clasificación: *naive gaussian*, *support vector machine*, *random forest* y *gaussian mixture model*. Donde se utilizan las características (la carga de las componentes, la varianza y la media) y etiquetas, de tal manera que se pueda evaluar el comportamiento del clasificador a usar¹⁸.
-

3.3. MINERÍA DE DATOS

Inicialmente la base del problema se encuentra en el reconocimiento de los datos entregados por el WCD. Los datos están dados por los valores de carga (energía depositada). Los datos se modelaron con histogramas de frecuencia de carga, de dos dimensiones (Carga vs Frecuencia), como se muestra en la Fig. 11. El histograma evidencia tres componentes, que guiados por un conocimiento a priori, se clasifican como (muónica, electromagnética y multi-partícula).

Al recibir los datos crudos totales, se hizo el procesamiento y modelamiento de los datos para implementarlos en GMM ya ajustado con los datos de carga del histograma. El modelo presentó complicaciones para evidenciar las componentes, esto se presentó debido a un grupo de datos innecesarios (outliers). Separando y visualizando los datos crudos, se estableció un límite donde sólo se tuvo en cuenta valores de carga < 800 ADC.bin, ya que por encima de esto los datos no aportaban información relevante (ver Fig. 12).

Figura 11. Histograma de carga para un registro de datos de una hora del WCD de MuTe. (arriba). Datos del WCD de eventos que han atravesado el hodoscopio (abajo).

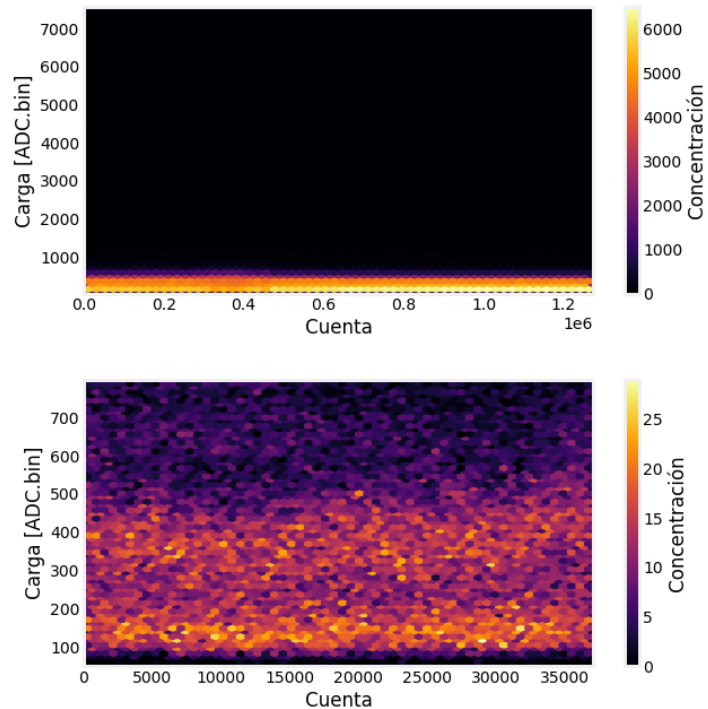


3.4. MODELO DE MEZCLA DE GAUSSIANAS Y PARAMETRIZACIÓN

Los modelos de mezcla gaussiana (GMM), representan subpoblaciones normalmente distribuidas dentro de una población general. Este modelo parametriza un subconjunto que pertenece a un grupo de datos. GMM puede tener más de 2 componentes y estima los parámetros individuales para modelar por medio de distribuciones normales.

GMM necesita parámetros iniciales para obtener un mejor ajuste. Los parámetros usados en este caso fueron: el número de componentes (3), el peso de cada una de

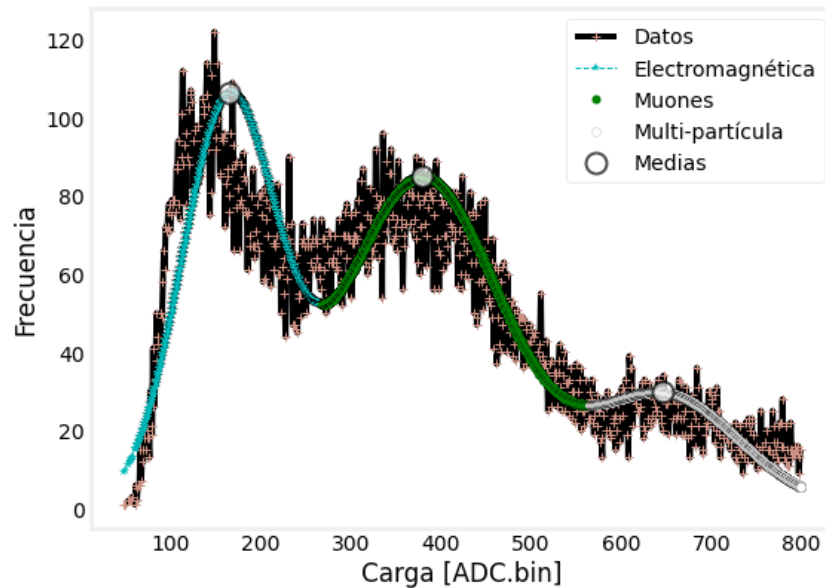
Figura 12. Datos de Carga crudos. ARRIBA: Datos completos de carga. ABAJO: Datos acotados por el umbral (Carga < 800)



las componentes (0,6 para la electromagnética , 0,3 para la muónica y 0,1 para la multi-partícula), el número de inicializaciones (5) y el número de iteraciones (1000). Al converger GMM devuelve una serie de parámetros, de los cuales sólo se tuvieron en cuenta valores como las medias, las varianzas y el peso de cada una de las componentes. Contando ya con estos parámetros se optó por normalizar los datos por medio de una PDF para cada componente (muónica, electromagnética y multi-partícula) permitiendo comparar los resultados con los datos más fácilmente, como se puede ver en la Fig.13.

En la primera separación de las componentes evidenció que el modelo con pocos datos (1.323) no responde acertadamente. Pero al contar con la cantidad total de los datos (1'273.125), el modelo mostró un mejor ajuste.

Figura 13. Histograma de frecuencia con los datos y la función de densidad de probabilidad ajustada por GMM. Con las medias calculadas, para apreciar el ajuste a los datos.



3.5. CONSTRUCCIÓN DE GAUSSIANAS Y ETIQUETADO

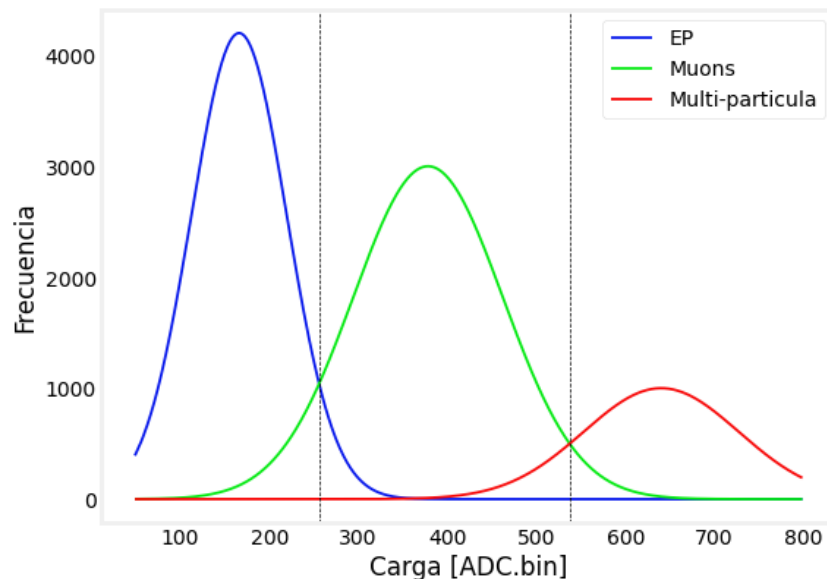
Con los parámetros de cada una de las gaussianas obtenidos previamente, se creó una función gaussiana definida por la Ecuación (2) de densidad de probabilidad.

$$f(x) = A * e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \quad (2)$$

donde A es la escala, μ la media, σ la varianza y x los datos. Se definieron tres funciones para las tres componentes (EP, muón y multi-partícula) como se muestra en la Fig. 14.

Definidas las tres funciones se analizan independientemente etiquetando: 0 para muon, 1 para electromagnético y 2 para multi-partícula. Se crea un vector con las etiquetas de la misma longitud del vector carga.

Figura 14. Las tres componentes: la muónica (verde), la electromagnética (azul) y multi-partícula (roja), representadas por su respectiva función de densidad de probabilidad.

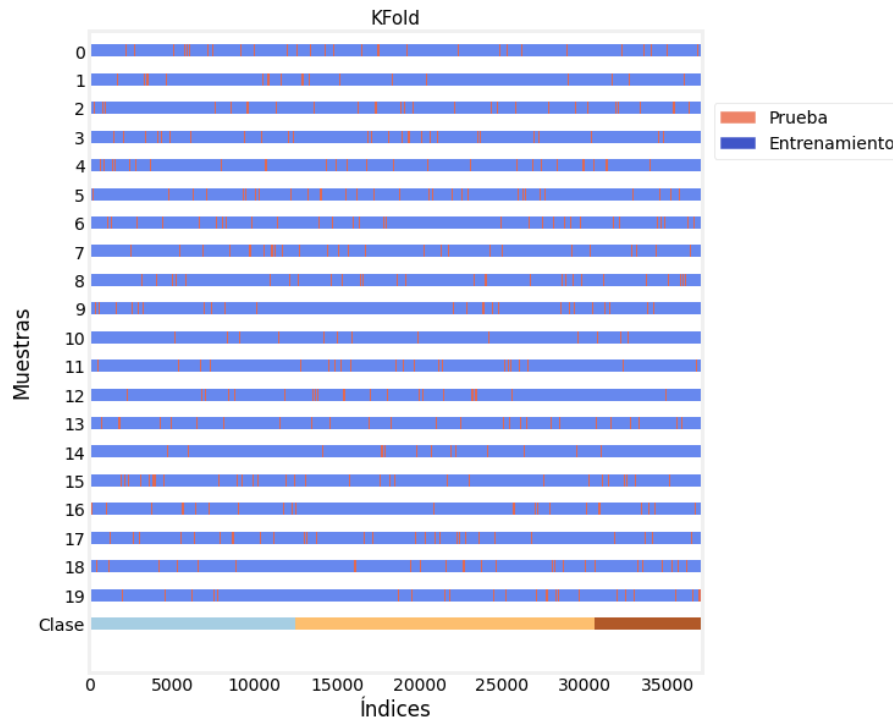


3.6. CLASIFICACIÓN SUPERVISADA, DIVISIÓN Y ENTRENAMIENTO

Inicialmente se hizo una validación cruzada para comparar los resultados obtenidos. Se partieron en 20 secciones los datos y se hizo pruebas con cada clasificador con el fin de fortalecer los resultados al tomar el 100 % de ellos. Ver Fig. 15. El resultado promedio fue de 0,89 en precisión con una desviación estándar de 0,14.

Con los datos etiquetados se debe hacer un *Split* para entrenar el clasificador. En este proyecto se usó *Train Test Split* implementado en la librería de *SKlearn*, ya que además de cumplir con el objetivo principal es una herramienta fácil y rápida de usar. Los datos fueron separados en dos conjuntos, Test (70 % de datos) y train (30 % de datos) para evaluar los modelos de clasificación.

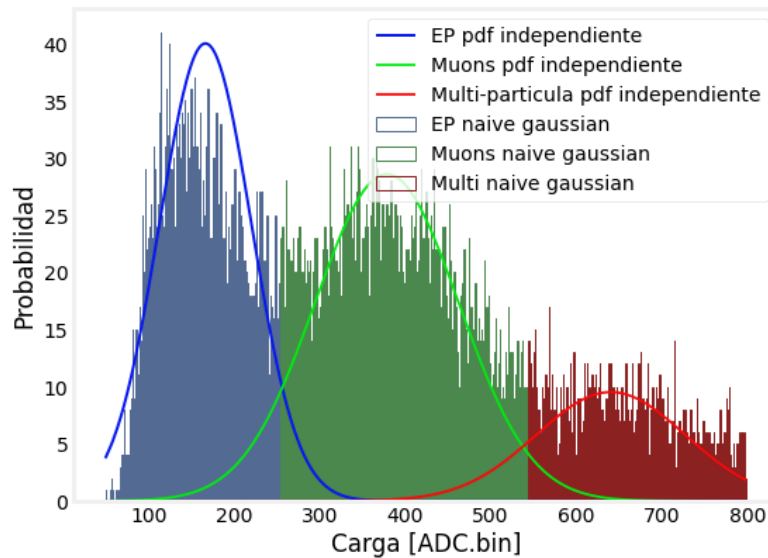
Figura 15. Separación por *k-fold*, se tomaron 20 *splits* y en cada uno de ellos se dividió unos para prueba y otros para entrenamientos, en cada uno de los 20 grupos se cambió la distribución para pruebas y entrenamientos .



En este caso se usaron tres modelos de clasificación supervisada: *Naive Gaussian*, *RandomForest* y *Support Vector Machines* implementados en la librería *SKlearn*. *Naive Gaussian* se escogió ya que la probabilidad de las características con la que opera es gaussiana, dado que los datos se modelan con variables que se ajustan a distribuciones normales, se optó por analizar los resultados con este modelo. Se trabajó con *RandomForest* debido a su versatilidad con los datos, ya que no se hizo preparación para la su calibración y facilitan su uso. Se calibró con 100 arboles. *Support Vector Machines* se usó para analizar la separación por grupos de los datos ajustando el kernel para definir su separación. Este modelo se ajustó con un kernel de función de base radial *RBF* y gamma *auto*. Los clasificadores se evaluaron me-

diante métricas de la matriz de confusión y eficiencia, además se hizo un ajuste de las predicciones respecto a los datos, como se ve en la Fig. 16.

Figura 16. Resultados del etiquetado de los datos por el clasificados y ajuste de las distribuciones.



En la matriz de confusión de la Tabla (1) se muestra el resultado obtenido por el clasificador. Los valores erróneos se dan entre las componentes estrictamente contiguas, como lo es la electromagnético con muónica y la muónica con la multi-partícula. No hubo fallos en la predicción entre muón y multi-partícula.

El valor de precisión en la Tabla (2) se refiere a cuán cerca del valor real se encuentra el valor medido. En la componente electromagnética y multi-partícula se obtuvo una precisión de 100 %. La muónica se redujo a 98 %: predijo que 49 datos de carga pertenecían a la componente muónica, cuando realmente pertenecían a la componente electromagnética y en 51 casos el clasificador los etiquetó como muónica cuando pertenecía verdaderamente a la componente multi-partícula.

Tabla 1. Matriz de confusión del *naive gaussian*. Donde las columnas 1, 2 y 3 son valores reales que representan las componentes: electromagnética, muónica y multi-partícula. Las filas 1, 2 y 3 son los valores predichos que definen las componentes electromagnética, muónica y multi-partícula.

	Electromagnética	Muón	Multi-partícula
Electromagnética	4026	49	0
Muón	0	5297	0
Multi-partícula	0	51	1702

Tabla 2. Métricas de precisión definidas para cada componente, calculadas mediante la matriz de confusión, mostrada en la Tabla 1

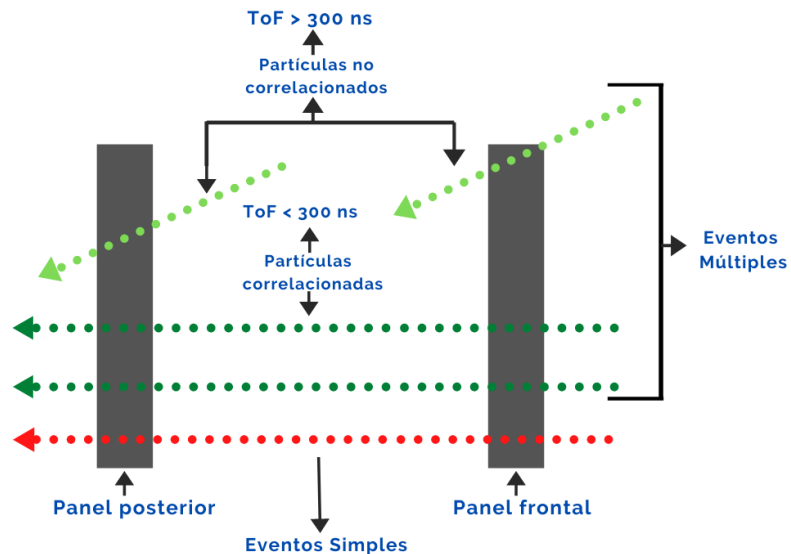
	Precisión	Veracidad	score	Datos
Electromagnético	1.00	0.99	0.99	4075
Muón	0.98	1.00	0.99	5297
Multi-partícula	1.00	0.97	0.99	1753

La veracidad o sensibilidad del modelo determina la proporción de los casos positivos clasificados por el modelo con respecto al total de los datos positivos. En esta métrica el clasificador asignó correctamente el 100 % de los datos pertenecientes a la componente muónica. En la componente electromagnética (99 %) obtuvo 4026 verdaderos positivos y 49 falsos positivos (muónica). En la multi-partícula (97 %) se obtuvieron 1702 verdaderos positivos y 51 falsos positivos (muónica).

4. CLASIFICACIÓN DEL RUIDO COMBINACIONAL POR MEDIO DEL TIEMPO DE VUELO USANDO ALGORITMOS DE APRENDIZAJE NO SUPERVISADO

En este apartado se analizaron los eventos detectados por MuTe mediante las mediciones del ToF y GMM. Los eventos se clasifican como correlacionados y no correlacionados por medio del tiempo de vuelo (ToF). Las partículas correlacionadas e individuales su tiempo de vuelo oscila entre 2 a $30ns$, mientras que las partículas no correlacionadas, la diferencia temporal entre las 2 partículas es $> 300ns$, revisar Fig. 17. La probabilidad de que ocurra un evento por partículas no correlacionadas con un diferencia temporal $< 200ns$ es de $0,05\%$ ¹⁵.

Figura 17. Partículas incidentes que cruzan los paneles. Estos se clasifican en correlacionados (verde oscuro), no correlacionados (verde claro), y eventos simples (rojo).



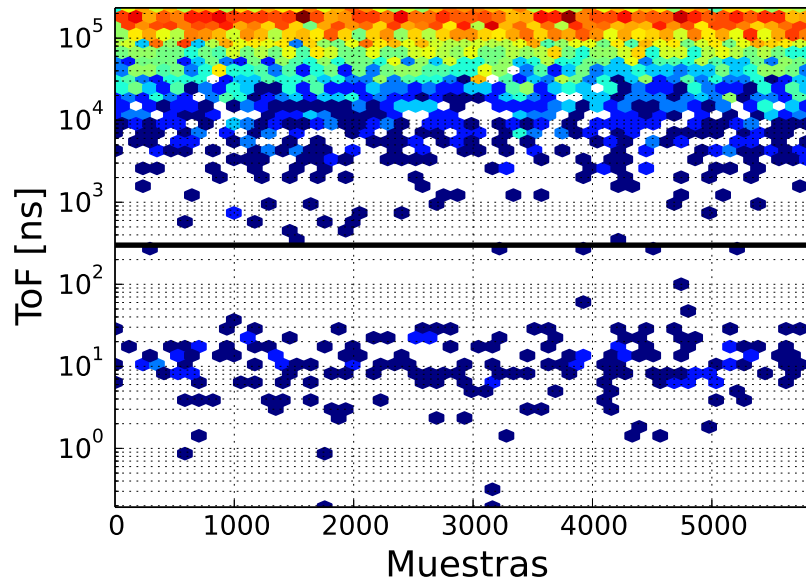
4.1. EVENTOS DE PARTÍCULA SIMPLE, CORRELACIONADOS Y NO-CORRELACIONADOS TENIENDO EN CUENTA EL TOF

4.1.1. Tratamiento y análisis de datos.

Inicialmente, se hizo una visualización de los datos para analizar la relación que tienen los eventos con las etiquetas teniendo en cuenta la información a-priori. La primera gráfica muestra el tiempo de vuelo de cada uno de los eventos. En la Fig. 18 se observa una aglomeración de datos correspondientes a los eventos correlacionados ($\text{ToF} < 300\text{ns}$). Los no correlacionados se dispersan uniformemente sobre los valores $> 300\text{ns}$.

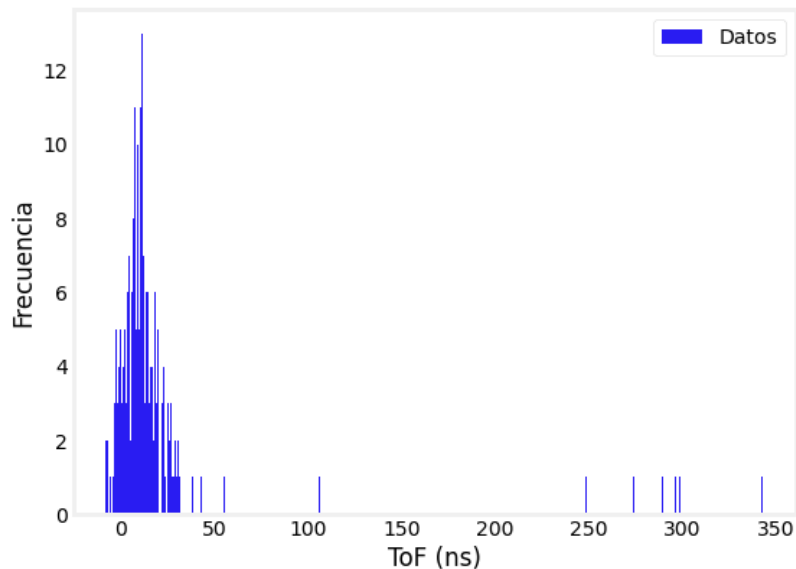
El histograma de frecuencia, evidencia la distribución de los eventos con $\text{ToF} < 300\text{ns}$, como se muestra en la Fig. 19.

Figura 18. Tiempo de vuelo de cada evento. Un umbral en 300ns (línea negra) permite dividir los eventos correlacionados ($< 300\text{ns}$) de los no-correlacionados ($> 300\text{ns}$).



También se visualizaron los datos en orden logarítmico para corroborar la conclusión obtenida de las dos gráficas anteriores. En la Fig. 20, se evidencia una concentración de muestras alrededor de 10 ns corroborando un mayor número de eventos correlacionados. Ubicadas estas zonas de interés, se parametrizan los datos por medio de distribuciones.

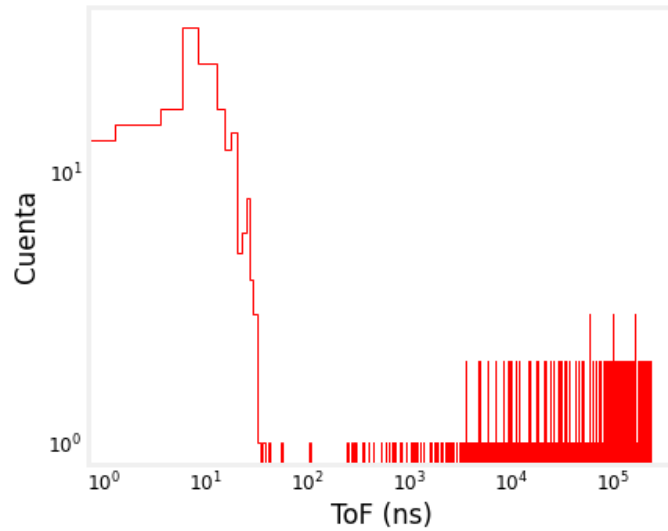
Figura 19. Histograma de frecuencia de los datos del tiempo de vuelo.



4.2. CLASIFICACIÓN Y SEGMENTACIÓN.

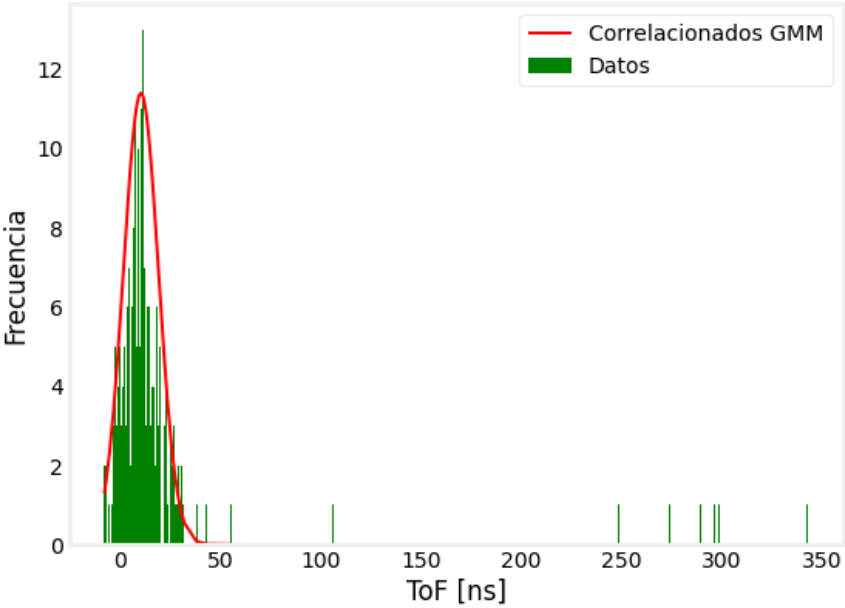
Bajo las métricas del ToF se ingresaron los datos a GMM para parametrizar cada componente. Se trabajó con 5883 datos definiendo parámetros iniciales como: el número de componentes (2), el número de iteraciones (1000), los pesos de cada componente (0.9 para correlacionados y 0.1 para no-correlacionados) y el número de iniciaciones que tuvo el modelo (5). Todo escogido para que se ajustara mejor a los eventos de partícula simple y correlacionada.

Figura 20. Tiempo de vuelo en orden logarítmico. Se distinguen los eventos de partícula simple y correlacionadas con un ToF menor a 300ns, mientras que los eventos no correlacionados tienen ToF mayores a 300 ns



Al parametrizar las componentes, se validó el modelo con los datos reales. De los parámetros de salida se tomó la varianza (76.80), la media (10.22) y los pesos de las componentes (0.96 y 0.037) para empezar con la separación de los eventos correlacionados y la eliminación de los no-correlacionados. Como se ve en la Fig. 21.

Figura 21. ToF < 300ns. Eventos correlacionados. Los datos entregados por el detector y el modelo ajustado, muestran que los eventos se concentran en un ToF menor a 40ns.



5. CLASIFICACIÓN NO SUPERVISADA DE LOS MUONES DE BAJO MOMENTUM.

Para la discriminación de los muones de bajo momentum ($<1 \text{ GeV}/c$), se trabajó con la relación de dos variables: el momento y el ToF. Se graficó en escala logarítmica para ver la relación entre las 2 variables e identificar su umbral. Los muones con momento menor a $1\text{GeV}/c$ son filtrados ya que sufren una dispersión angular que afecta la resolución espacial del MuTe. El proceso de clasificación se llevó a cabo gracias a algoritmos de clasificación no supervisada.

5.1. MANEJO DE VARIABLES

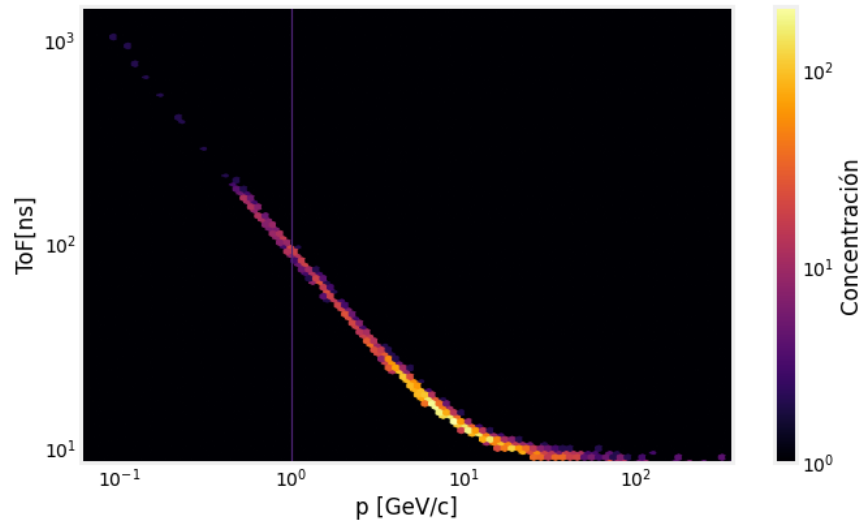
Inicialmente se eliminaron datos NaN de los datos crudos que dificultaban el procesamiento. Los datos resultantes se redujeron \sim en un 40%. En esta parte del proyecto ya se filtró la mayoría de las fuentes de ruido. Tras relacionar el ToF y el momentum, ver Fig. 22, se clasificó los muones de interés que presentan un momentum mayor a $1\text{GeV}/c$. A continuación se explica la relación entre el ToF y el momento.

5.2. ESTIMACIÓN DE MOMENTUM

El sistema de ToF de MuTe mide el tiempo de vuelo que tardan las partículas cargadas de energía en atravesar una determinada trayectoria en el hodoscopio. Utilizando el ToF y la energía depositada en el WCD, MuTe estima el momentum de todas las partículas que ingresan al detector, estableciendo un umbral para discriminar las partículas de baja energía ($<1\text{Gev}/c$)¹⁵.

La estimación del momentum se lleva a cabo con el fin de encontrar los eventos que sufren de dispersión múltiple. El cálculo del momentum depende de 2 variables, el

Figura 22. ToF vs momentum. El 54 % de las partículas tiene momentum mayor a 1 GeV/c (línea púrpura).



ToF y la distancia recorrida entre las placas, que se define en la Ecuación (1).

5.3. CLASIFICACIÓN

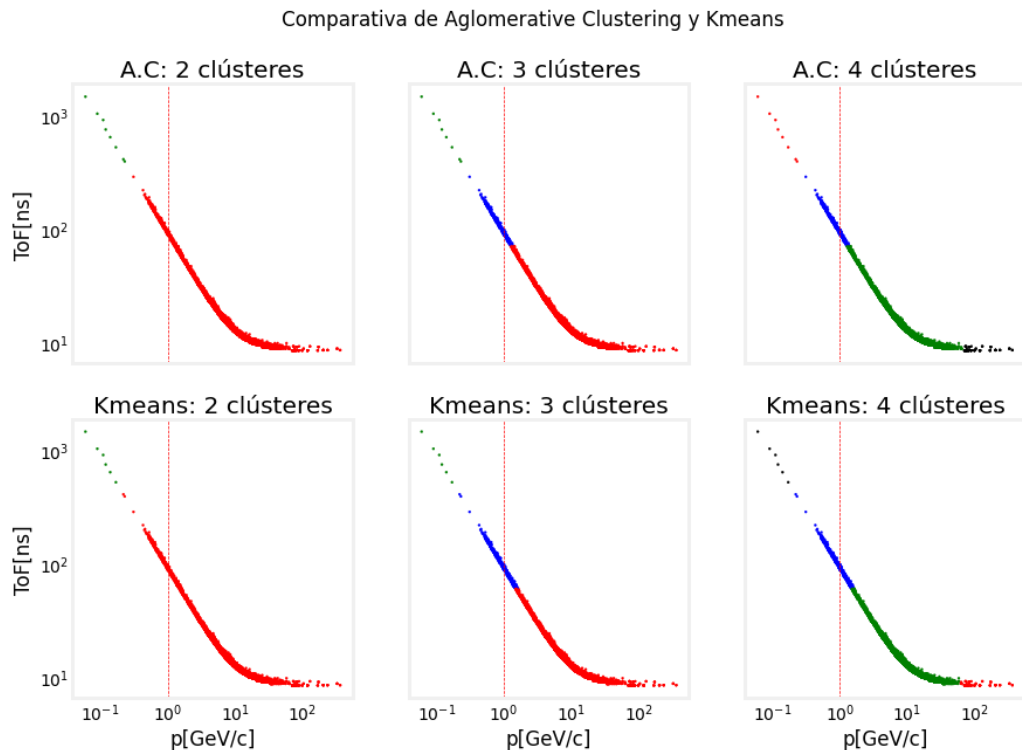
Bajo el criterio de la naturaleza de los datos se buscó encontrar algoritmos capaces de localizar grupos en los datos. En este caso puntualmente se conoce que a partir de 1 GeV/c los eventos coinciden cada vez más, así que la búsqueda de los algoritmos para clasificar usando aprendizaje no supervisado se limita a encontrar opciones que ejecuten bien la tarea de localizar agrupaciones de altas y bajas densidades en los datos. En este proceso se encontró y probó los métodos mencionados a continuación: *Agglomerative Clustering*, *Kmeans* y *Spectral Clustering*. También se trabajó con *PCA* para ver alguna relación entre los datos o si se podía obviar alguna componente, pero la resultante arrojó un comportamiento similar a los datos originales.

Con *Agglomerative Clustering* y *Kmeans* hubo un mejor ajuste de los datos, como se

ve en la Fig. (23). Luego estos clasificadores se probaron con diferente número de clúster, para compararlos minuciosamente mediante la medición de la precisión.

5.4. ANÁLISIS DE RESULTADOS.

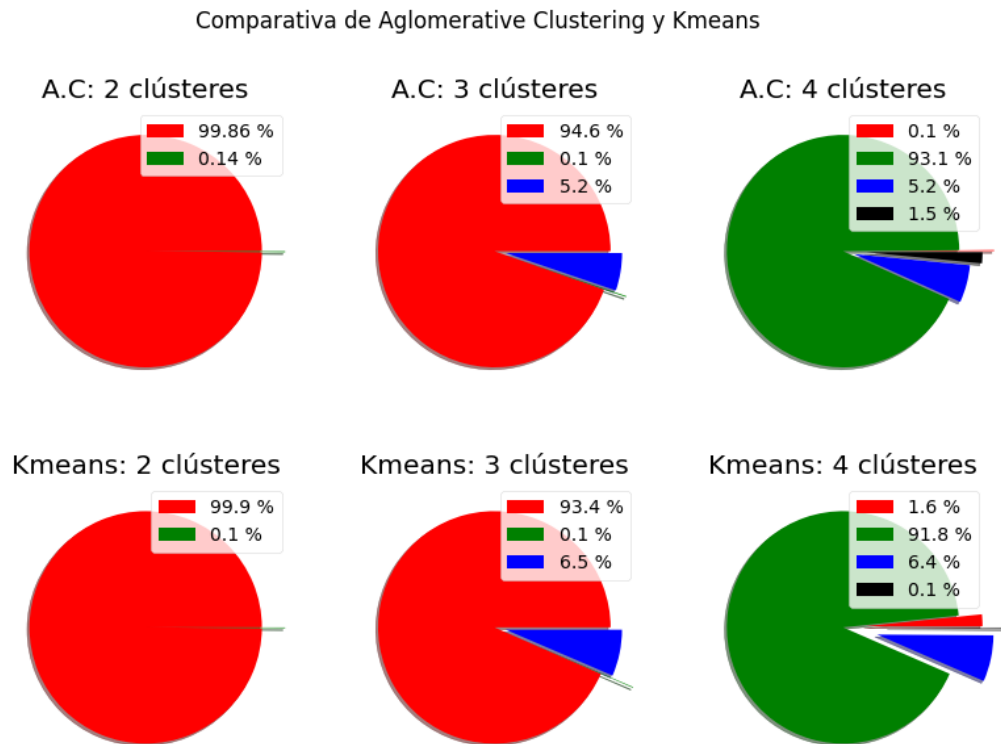
Figura 23. ToF vs momento entregados por los clasificadores *AgglomerativeClustering* y *Kmeans*, se muestra la comparativa de 2 a 4 clusters, con el umbral en 1GeV/c. (línea roja)



Para la decisión final se tuvo en cuenta 2 factores: la precisión de la clasificación y la complejidad computacional. En este orden de ideas, según los resultados obtenidos se escogieron tres clústeres con el algoritmo de *Agglomerative Clustering*. Al aumentar el volumen de datos y el número de clústeres, se afecta el funcionamiento

fluído del algoritmo, así que se validó con el resultado más óptimo.

Figura 24. ToF vs momento entregados por los clasificadores *Aglomerative Clustering* y *Kmeans*, se muestra la comparativa de 2 a 4 clusters. Muestra el porcentaje de cada grupo en cada cluster.



En términos generales \sim el 90 % de los eventos estaban en la zona de aceptación (>1 GeV/c), esta condición se presentó en todos los clusters. Otra cosa que cabe mencionar, es que en todos los clusters el grupo de mayor cantidad de eventos estuvo estrictamente después del umbral de 1 GeV/c como se muestra en la Fig. 23. No hubo diferencias significativas en los resultados entre los 2 algoritmos al evaluarse con el mismo número de clusters, diferían por $\sim 3\%$ (200 datos). La diferencia a validar fue la posición del grupo en la muestra de datos. Por ejemplo en 2 clusters

con *Kmeans*, el grupo *rojo* con 5817 datos, fue el menos preciso con respecto al umbral de 1GeV/c. Para 3 clusters se tuvo una mejor precisión, con un 94.6% para *AgglomerativeClustering* y 93.4% para *Kmeans* como se puede ver en la Fig. 24.

6. CONCLUSIONES

Los resultados obtenidos con los modelos de mezcla gaussiana (GMM) permitieron parametrizar las componentes que conforman los eventos registrados por el detector de MuTe. El clasificador tuvo una precisión de 100 % etiquetando los eventos pertenecientes a la componente electromagnética y multi-partícula; en la muónica fue del 98 %.

Los resultados con *support vector machine* y *random forest* fueron superiores en precisión al 95 % . Esto gracias a la naturaleza de los datos, ya que son altamente normalizables y de 1 dimensión, minimizando el impacto computacional en los clasificadores. *Random forest* y *support vector machine* fueron idóneos para el desarrollo del proyecto ya que necesitó una preparación de datos mínima, pero si la complejidad en los datos aumenta puede causar un sobre ajuste.

Con los datos del ToF se filtraron las partículas correlacionadas con un ToF < 300 ns. El modelo ajustó bien a los datos ya que se presentaba una aglomeración marcada entre 0 y 50 ns. Los eventos multi-partícula no correlacionados presentaron un ToF por encima de 300 ns.

De las mediciones de momentum se buscó encontrar y separar los eventos que son afectados por dispersión múltiple. Se estableció un umbral para discriminar las partículas de baja energía con un momentum menor a 1 GeV/c. En el proceso de clasificación se indagaron métodos como: *Agglomerative Clustering*, *Kmeans*, *PCA* y *Spectral Clustering*, variando el número de clusters. Encontramos que los mejores resultados están entre 2 y 4 clusters.

La aglomeración parcial de los datos en zonas específicas fue clave para que los algoritmos identificaran de acuerdo al número de clusters las clases. Con 3 clusters para *Agglomerative Clustering* se tuvo 5511 (rojo), 8 (azul) y 304 (verde), y con *Kmeans* 5441 (rojo), 6 (azul), 376 (verde). Dejando en ambos casos el grupo de mayor número de datos (cluster rojo) por encima del umbral de momentum.

BIBLIOGRAFÍA

Bonechi, Lorenzo, Raffaello D'Alessandro y Andrea Giammanco. "Atmospheric muons as an imaging tool". En: *Reviews in Physics* (2020), pág. 100038 (vid. págs. 17, 19, 26).

George, EP. "Cosmic rays measure overburden of tunnel". En: *Commonwealth Engineer* 455 (1955) (vid. pág. 17).

Gómez, H y col. "Forward scattering effects on muon imaging". En: *Journal of Instrumentation* 12.12 (2017), P12018 (vid. pág. 15).

Haungs, A. y col. "KCDC — The KASCADE Cosmic-ray Data Centre". En: *Journal of Physics: Conference Series* 632.1 (2015), pág. 012011 (vid. pág. 19).

Jourde, Kevin y col. "Effects of upward-going cosmic muons on density radiography of volcanoes". En: *arXiv preprint arXiv:1307.6758* (2013) (vid. págs. 22, 28).

Jourde, Kevin y col. "Experimental detection of upward going cosmic particles and consequences for correction of density radiography of volcanoes". En: *Geophysical Research Letters* 40.24 (2013), págs. 6334-6339 (vid. págs. 21, 22).

Kusagaya, T. y H. K. Tanaka. "Muographic imaging with a multi-layered telescope and its application to the study of the subsurface structure of a volcano". En: *Proceedings of the Japan Academy, Series B* 91.9 (2015), págs. 501-510. DOI: 10.2183/pjab.91.501 (vid. pág. 23).

Kusagaya, Taro e Hiroyuki KM Tanaka. "Muographic imaging with a multi-layered telescope and its application to the study of the subsurface structure of a volcano".

En: *Proceedings of the Japan Academy, Series B* 91.9 (2015), págs. 501-510 (vid. pág. 15).

Marteau, J. y col. "Implementation of sub-nanosecond time-to-digital convertor in field-programmable gate array: applications to time-of-flight analysis in muon radiography". En: *Measurement Science and Technology* 25.3 (2014), pág. 035101. DOI: 10.1088/0957-0233/25/3/035101 (vid. pág. 28).

Marteau, Jacques y col. "DIAPHANE: muon tomography applied to volcanoes, civil engineering, archaeology". En: *Journal of Instrumentation* 12.02 (2017), pág. C02008 (vid. pág. 15).

Nagamine, K. y col. "Probing the inner structure of blast furnaces by cosmic-ray muon radiography". En: *Proceedings of the Japan Academy, Series B* 81.7 (2005), págs. 257-260. DOI: 10.2183/pjab.81.257 (vid. pág. 17).

Nishiyama, R, S Miyamoto y N Naganawa. "Experimental study of source of background noise in muon radiography using emulsion film detectors". En: *Geoscientific Instrumentation, Methods and Data Systems* 3.1 (2014), págs. 29-39 (vid. pág. 15).

Nishiyama, R., S. Miyamoto y N. Naganawa. "Experimental study of source of background noise in muon radiography using emulsion film detectors". En: *Geoscientific Instrumentation, Methods and Data Systems* 3.1 (2014), págs. 29-39. DOI: 10.5194/gi-3-29-2014 (vid. págs. 20, 24, 29).

Oláh, L. y col. "High-definition and low-noise muography of the Sakurajima volcano with gaseous tracking detectors". En: *Scientific Reports* 8.1 (2018). DOI: 10.1038/s41598-018-21423-9 (vid. pág. 27).

Oláh, László y col. “High-definition and low-noise muography of the Sakurajima volcano with gaseous tracking detectors”. En: *Scientific reports* 8.1 (2018), págs. 1-13 (vid. pág. 26).

Peña-Rodríguez Jesús y Vásquez-Ramírez, Adriana y Sanabria Gómez José D y Núñez Luis A y Sierra-Porta David y Asorey Hernán. “Calibración y primeras mediciones de MuTe: un telescopio muón híbrido para estructuras geológicas”. En: *arXiv preprint arXiv: 1909.09732* () (vid. pág. 16).

Pérez, Aritz, Pedro Larranaga e Inaki Inza. “Modelos gráficos probabilísticos para la clasificación supervisada empleando la estimación basada en kernels Gaussianos esféricos”. En: *III Taller Nacional de Minería de Datos y Aprendizaje* (2005), págs. 125-134 (vid. pág. 33).

Procureur, S. “Muon imaging: Principles, technologies and applications”. En: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 878 (2018), págs. 169-179 (vid. págs. 11, 17).

Rodríguez, Jesús Peña. “Diseño y calibración de un telescopio de muones híbrido para estudios vulcanológicos.” Tesis doct. Facultad de Ciencias. Escuela de Física. Universidad Industrial de Santander, 2020 (vid. págs. 20, 21, 23, 24, 30, 31, 41, 46).

Tanabashi, Masaharu y col. “Review of particle physics”. En: *Physical Review D* 98.3 (2018), pág. 030001 (vid. pág. 18).