



Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Búsqueda de eventos con estados finales a cuatro leptones
en datos abiertos del CMS

Tesis presentada al

Colegio de Física

como requisito parcial para la obtención del grado de

LICENCIADO EN FÍSICA

por

Aldo Santiago Marquez

Asesorado por

María Isabel Pedraza Morales
Humberto Antonio Ibargüen Salazar

Puebla Pue.
February 19, 2026



Benemérita Universidad Autónoma de Puebla

Facultad de Ciencias Físico Matemáticas

Búsqueda de eventos con estados finales a cuatro leptones
en datos abiertos del CMS

Tesis presentada al

Colegio de Física

como requisito parcial para la obtención del grado de

LICENCIADO EN FÍSICA

por

Aldo Santiago Marquez

Asesorado por

María Isabel Pedraza Morales
Humberto Antonio Ibargüen Salazar

Puebla Pue.
February 19, 2026

Título: Búsqueda de eventos con estados finales a cuatro leptones en datos abiertos del CMS

Estudiante: ALDO SANTIAGO MARQUEZ

COMITÉ

Enrique Varela Carlos
Presidente

Haydee Hernandez Arellano
Secretario

Ana Aurelia Aviles López
Vocal

Epifanio Lorenzo Ponce Lancho
Suplente

María Isabel Pedraza Morales
Asesor

Humberto Antonio Ibargüen Salazar
Asesor

Agradecimientos

A mis padres por su apoyo incondicional y a Sid por acompañarme en todo momento, gracias.

Resumen

Este estudio presenta un análisis de eventos tetramuónicos (4μ) en el contexto de la búsqueda del bosón de Higgs mediante su canal de desintegración $H \rightarrow ZZ^* \rightarrow 4\mu$, utilizando datos públicos del detector CMS correspondientes a colisiones protón-protón a $\sqrt{s} = 13$ TeV del Run 2015D del LHC.

Los objetivos principales fueron: (i) validar la utilidad de los datos abiertos para física del Higgs, (ii) implementar un protocolo completo de análisis que incluye selección de eventos, control de calidad y reconstrucción de variables cinemáticas, y (iii) caracterizar el poder discriminatorio de correlaciones angulares. La metodología empleó un esquema de análisis basado en Python para: selección de cuádruples muónicas con $p_T > 5$ GeV/ c y $|\eta| < 2.4$, reconstrucción de masas invariantes ($m_{\mu^+\mu^-}$, $m_{4\mu}$), y estudio de distribuciones de $\cos\theta^*$ y ϕ_η^* en el sistema de referencia del Higgs.

Los resultados obtenidos incluyen: (a) la identificación de un pico claro en la distribución de masas invariantes de pares de muones, consistente con la resonancia del bosón Z , (b) la observación de algunos eventos con masas del sistema 4μ cercanas a la región esperada para el bosón de Higgs, y (c) una tendencia en las distribuciones angulares compatible con las predicciones teóricas del Modelo Estándar.

Se demuestra que los datos abiertos del CMS permiten: (1) identificar señales de física del Higgs, (2) validar técnicas de análisis estándar, y (3) servir como herramienta educativa en física experimental de altas energías. Este trabajo establece las bases para análisis más complejos con datos públicos del LHC.

Contents

Agradecimientos	iv
Resumen	vi
Contenido	vii
Lista de Figuras	xi
1 Introducción	1
1.1 Contexto científico	1
1.2 Motivación del estudio	1
1.3 Objetivos y alcance del estudio	2
1.4 Fuente de datos utilizados	2
1.5 Marco Teórico	2
1.5.1 El Bosón de Higgs en el Modelo Estándar	2
1.5.2 El Canal de Desintegración $H \rightarrow ZZ^* \rightarrow 4\mu$	3
1.5.3 Importancia del Canal Tetramuónico	4
1.5.4 Búsquedas y Estudios Previos	4
1.6 Selección de Eventos	6
1.6.1 Criterios de Selección	6
1.6.2 Uso del <i>MuonAnalyzer</i>	7
1.7 Identificación de Muones	8
1.7.1 Clasificación de Identificación: <i>Loose</i> , <i>Medium</i> y <i>Tight</i>	8
1.7.2 Implementación en el Software de Análisis	9
1.7.3 Aislamiento de Muones	9
1.7.4 Aislamiento de Muones	9
1.8 Análisis de datos	11
1.8.1 Manejo y preparación de datos	11
1.8.2 Reducción de Fondo y Selección Física	11
1.8.3 Variables Relevantes para el Análisis Cinemático	13
2 Fundamentos Computacionales y Metodológicos del Análisis	17
2.1 Selección del Canal $H \rightarrow ZZ^* \rightarrow 4\mu$: Ventajas Físicas y Computacionales	17
2.2 Estructura Computacional de los Datos y Flujo de Trabajo	18
2.2.1 Etapa 1: Ntuplización con CMSSW	18
2.2.2 Etapa 2: Análisis Final con C++/ROOT	19
2.3 Pipeline de Análisis Computacional	19
2.4 Marco Computacional y Reproducibilidad	20
2.4.1 Containerización del Entorno de Análisis con Docker	20
2.4.2 Entorno de Visualización y Análisis Secundario	20
2.4.3 Validación y Verificación del Análisis	22

3	Infraestructura Computacional para el Análisis de Datos CMS	23
3.1	Necesidades Técnicas y Solución Propuesta	23
3.2	Configuración del Entorno Containerizado con Docker	24
3.2.1	Selección de la Imagen Base	24
3.2.2	Despliegue y Configuración del Contenedor	24
3.3	Acceso Gráfico y Visualización Interactiva	25
3.3.1	Interfaz Gráfica Remota mediante VNC	25
3.4	Interoperabilidad y Flujo de Trabajo	26
3.4.1	Inspección Interactiva de Datos con ROOT	26
3.4.2	Interoperabilidad de Datos Mediante Volúmenes Compartidos	26
3.4.3	Compatibilidad del Entorno y Verificación	27
3.4.4	Evaluación de rendimiento	28
3.5	Balance de la Arquitectura: Ventajas, Limitaciones y Perspectivas	28
3.5.1	Ventajas Clave de la Solución Containerizada	28
3.5.2	Limitaciones y Líneas de Trabajo Futuro	28
3.5.3	Referencia Rápida de Comandos Docker	29
4	Metodología de Análisis	31
4.1	Conjunto de Datos y Selección del Canal	31
4.1.1	Fuente de Datos y Características	31
4.1.2	Firma Experimental y Reconstrucción del Estado Final	31
4.2	Entorno Computacional y Flujo de Trabajo	32
4.2.1	Procesamiento Primario con CMSSW en Contenedor Docker	32
4.2.2	Análisis Exploratorio y Prototipado en Python	33
4.2.3	Pipeline de Análisis Principal con C++/ROOT	33
4.2.4	Preprocesamiento: Estructuración de los Datos del Evento	34
4.2.5	Cálculo de Variables Físicas y Selección de Candidatos	35
4.3	Criterios de Selección de Eventos	36
4.4	Validación y control sistemático	37
4.4.1	Métricas de desempeño y estimación de pureza	37
4.4.2	Limitaciones y sistemáticos del análisis	38
4.4.3	Distribuciones clave	38
4.4.4	Limitaciones	38
4.5	Automatización del Pipeline de Procesamiento de Datos	39
4.5.1	Arquitectura del Sistema: Un Enfoque ETL	39
4.5.2	Implementación Técnica del Pipeline Automatizado	40
4.5.3	Exportación de eventos tetramuónicos a formato CSV	44
4.5.4	Aplicación de cortes cinemáticos y reconstrucción del sistema Higgs	51
4.6	Implementación del Pipeline de Análisis en C++/ROOT	52
4.6.1	Arquitectura del Software: Modularidad y Configurabilidad	53
4.6.2	Algoritmo Central del Análisis	53
5	Resultados	57
5.1	Distribución de la masa invariante del sistema 4μ	57
5.1.1	Clasificación de eventos por identificación de muones	59
5.1.2	Composición de la muestra según la identificación de muones	60
5.1.3	Selección de Eventos con Muones Identificados	61
5.1.4	Aceptancia de la selección por etapa	63
5.1.5	Aceptancia global de la selección de eventos	66
5.1.6	Comparación entre eventos iniciales y seleccionados	67
5.1.7	Distribución del momento transversal p_T de los muones	68
5.1.8	Distribución de la separación angular ΔR entre pares de muones	69

<i>CONTENTS</i>	xi
6 Discussion y Conclusiones	73
6.1 Discusión	73
6.2 Conclusiones	75
6.3 Recomendaciones y perspectivas a futuro	76
Bibliografia	79

List of Figures

1.1	Diagrama de Feynman representativo del proceso de producción vía fusión de gluones (ggF) y desintegración $H \rightarrow ZZ^* \rightarrow 4\mu$	3
2.1	Diagrama del <i>pipeline</i> computacional completo. El flujo principal de procesamiento se ejecuta dentro de un contenedor Docker, mientras que la visualización final se realiza en un entorno Python reproducible.	22
3.1	Infraestructura híbrida para análisis de datos CMS. El entorno CMSSW está encapsulado en Docker, con integración de carpetas compartidas y visualización remota vía VNC/TBrowser.	24
4.1	Comparación esquemática de los canales de desintegración $H \rightarrow WW^*$ y $H \rightarrow ZZ^* \rightarrow 4\mu$. En $H \rightarrow WW^*$ la presencia de neutrinos impide reconstruir completamente la masa invariante; en $H \rightarrow ZZ^* \rightarrow 4\mu$ el estado final tetramuónico es limpio y totalmente reconstruible.	32
4.2	(a) Distribución de $m_{4\mu}$ mostrando el exceso alrededor de 125 GeV. (b) Separación angular ΔR entre muones pares.	38
5.1	Distribución de la masa invariante del sistema de cuatro muones, $m_{4\mu}$, para los eventos que satisfacen los criterios de selección descritos en el texto. La línea vertical discontinua indica la masa del bosón de Higgs medida en el LHC, $m_H \simeq 125$ GeV.	59
5.2	Eficiencia de identificación para los diferentes niveles de selección.	61
5.3	Número de eventos (azul, escala logarítmica) y aceptación acumulada (rojo) tras cada una de las etapas principales de selección del análisis tetramuónico.	64
5.4	Distribución de la masa invariante del sistema de cuatro muones, $m_{4\mu}$, separada por categorías de identificación leptónica: <i>Categoría 1</i> (muones que sólo cumplen el criterio <i>Loose</i>), <i>Categoría 2</i> (al menos <i>Medium</i>) y <i>Categoría 3</i> (todos los muones <i>Tight</i>). La línea discontinua indica la masa del bosón de Higgs, $m_H \simeq 125$ GeV.	66
5.5	Comparación esquemática entre el número de eventos iniciales con ramas centrales válidas y el número de eventos seleccionados tras aplicar todos los criterios de selección para el canal $H \rightarrow ZZ^* \rightarrow 4\mu$	68
5.6	Distribución del momento transversal p_T de los muones reconstruidos en los eventos seleccionados.	69
5.7	Distribución de la separación angular ΔR entre pares de muones reconstruidos en los eventos seleccionados.	70
5.8	Distribución de la masa invariante del candidato Z_a , m_{Z_a} , definido como el par de muones cuya masa es más cercana a la masa nominal del bosón Z . La línea discontinua indica $m_Z \simeq 91.2$ GeV.	71
5.9	Distribución de la masa invariante del candidato Z_b , m_{Z_b} , construido a partir del segundo par de muones en cada evento. La acumulación en la región de bajas masas es característica de un bosón Z^* off-shell y de contribuciones de fondo no resonantes.	72

6.1	Comparación visual de la reconstrucción del pico de masa. A la izquierda, la distribución $m_{4\mu}$ obtenida en este análisis; a la derecha, el resultado oficial de CMS. Se ajustó la escala visual para facilitar la comparación de la estructura del pico en 125 GeV.	75
-----	---	----

Chapter 1

Introducción

1.1 Contexto científico

El descubrimiento del bosón de Higgs en 2012 por las colaboraciones ATLAS y CMS [16, 23] marcó un hito en la física de partículas, al confirmar el mecanismo de Brout–Englert–Higgs como el origen de la masa de las partículas elementales. Entre los diversos canales de desintegración del Higgs, el proceso $H \rightarrow ZZ^* \rightarrow 4\ell$ ($\ell = e, \mu$) es conocido como el *canal dorado* debido a su excelente relación señal–fondo, particularmente en el caso tetramuónico (4μ) [20].

El experimento CMS (*Compact Muon Solenoid*) fue diseñado con una capacidad sobresaliente para la identificación y reconstrucción de muones. Su potente campo magnético y el sistema de cámaras de muones integrado permiten alcanzar una resolución en momento transversal de aproximadamente $\Delta p_T/p_T \approx 1.5\%$ para $p_T < 100$ GeV/ c [28]. Esta precisión es fundamental, ya que se traduce directamente en una resolución de masa invariante superior para el sistema de cuatro leptones, indispensable para resolver el estrecho pico del bosón de Higgs.

1.2 Motivación del estudio

La iniciativa de **datos abiertos del CERN y la colaboración CMS** [26] representa un avance significativo en la democratización de la investigación en física de altas energías. Este proyecto explora el potencial de los datos públicos del **Run 2 del LHC** con los siguientes propósitos:

- Validar técnicas de análisis empleadas por la colaboración CMS, con el fin de reproducir resultados de física de alto impacto.
- Implementar búsquedas completas del bosón de Higgs en un entorno computacional abierto y reproducible.
- Desarrollar y evaluar herramientas avanzadas de análisis de datos (**Python**, **pandas**) aplicadas a información experimental real.

El canal $H \rightarrow ZZ^* \rightarrow 4\mu$ [19, 21, 20, 13, 14] fue seleccionado debido a su combinación de características únicas: presenta una **baja tasa de fondo irreducible** de origen electrodébil (ZZ , $Z\gamma^*$), del orden de ~ 0.03 eventos/ fb^{-1} para masas del Higgs alrededor de 125 GeV [19, 22]; una **alta eficiencia de disparo** del sistema de muones (superior al 98%) [29, 30]; y una **excelente resolución en masa invariante** del sistema de cuatro muones, cercana al 1.2%, fundamental para distinguir la señal frente al fondo [19, 21, 20].

Desde el punto de vista computacional, este trabajo desarrolla un marco de análisis completo basado en herramientas de alto rendimiento como **Python** y **pandas** [39, 45, 49, 41, 43]. Este

enfoque no solo permite reproducir resultados relevantes en física de partículas, sino que también enfatiza la integración efectiva de la **física computacional** en el estudio de datos experimentales complejos [5].

1.3 Objetivos y alcance del estudio

El presente trabajo persigue tres objetivos principales:

1. **Implementación computacional:** desarrollar un *framework* completo y reproducible de análisis utilizando CMSSW, Docker y Python a partir de datos abiertos del CMS.
2. **Caracterización del rendimiento:** calibrar y evaluar la reconstrucción de eventos 4μ para validar los cortes cinemáticos de aislamiento y momento transversal.
3. **Evaluación de la señal:** identificar y evaluar el exceso de eventos en la distribución de masa invariante alrededor de $m_H \approx 125$ GeV para establecer su compatibilidad con el bosón de Higgs.

1.4 Fuente de datos utilizados

El análisis presentado en esta tesis utiliza el conjunto de datos abierto **DoubleMuon** del experimento CMS, correspondiente al **Run 2015D** del *Large Hadron Collider* (LHC) [26, 51]. Las condiciones experimentales son las siguientes [36, 31]:

- **Energía de colisión ($\sqrt{s}$):** 13 TeV (colisiones protón-protón).
- **Luminosidad integrada (L):** 2.5 fb^{-1} [12].
- **Formato de los datos:** MINIAOD [26].

Estos datos fueron recolectados durante el segundo ciclo de operación del LHC bajo condiciones estables de campo magnético y alineación, asegurando la calidad necesaria para la reconstrucción precisa de los eventos tetramuónicos [7, 29].

1.5 Marco Teórico

1.5.1 El Bosón de Higgs en el Modelo Estándar

El Modelo Estándar (ME) de la física de partículas describe con un éxito extraordinario las partículas fundamentales y sus interacciones (a excepción de la gravedad) [50, 33]. Sin embargo, la formulación teórica original presentaba un problema fundamental: la simetría matemática que subyace a la teoría electrodébil implicaba que los portadores de la fuerza débil, los bosones W y Z, debían tener masa nula. Esto entraba en clara contradicción con la evidencia experimental, que demostraba que estos bosones son, de hecho, masivos [38, 42].

La solución a este problema fue propuesta de forma independiente en 1964 mediante el mecanismo de Brout-Englert-Higgs (BEH) [35, 40, 38, 42]. Este mecanismo postula la existencia de un campo cuántico, el campo de Higgs, que permea todo el universo. A diferencia de otros campos, el estado de mínima energía de este campo (su "vacío") rompe la simetría electrodébil fundamental. Es la interacción de las partículas con este estado de vacío lo que les confiere masa. Una predicción clave de este mecanismo es la existencia de una nueva partícula masiva de espín 0, el bosón de Higgs, que corresponde a una excitación cuántica de dicho campo [37, 33].

El descubrimiento de esta partícula en 2012 por los experimentos ATLAS y CMS [16, 23] no solo completó el pilar teórico del ME, sino que abrió una nueva ventana para probar la validez del

modelo y buscar física más allá del mismo. Esto se logra a través de la medición precisa de las propiedades del Higgs, como su masa, espín, paridad y la fuerza de sus acoplamientos a otras partículas [3, 6, 9, 8, 17].

1.5.2 El Canal de Desintegración $H \rightarrow ZZ^* \rightarrow 4\mu$

La producción del bosón de Higgs en el LHC a $\sqrt{s} = 13\text{TeV}$ está dominada por el mecanismo de **fusión de gluones** (ggF), seguido por la **fusión de bosones vectoriales** (VBF) [33, 34, 15]. Una vez producido, el Higgs, con una vida media extremadamente corta, se desintegra en varios "canales" posibles. La probabilidad de cada canal de desintegración (su **branching ratio**) depende de la masa del Higgs y de la fuerza de su acoplamiento a las partículas finales [50].

Dado que el acoplamiento del Higgs a otras partículas es proporcional a la masa de estas, el Higgs se desintegra preferentemente en partículas masivas. Esto hace que los canales que involucran bosones vectoriales (W y Z) sean de suma importancia [33, 37]. El canal de desintegración a cuatro leptones cargados ($H \rightarrow ZZ^* \rightarrow 4\ell$, donde $\ell = e, \mu$) es considerado el "canal dorado" (*golden channel*) por su claridad experimental [19, 21, 13]. La variante tetramuónica (4μ) estudiada en esta tesis es particularmente limpia y precisa [20, 14].

El proceso $H \rightarrow ZZ^* \rightarrow 4\mu$ se describe en las siguientes etapas:

1. **Producción del Higgs:** En colisiones protón-protón ($p-p$) a $\sqrt{s} = 13\text{TeV}$, dos gluones de los protones incidentes se fusionan para producir un bosón de Higgs (H) [33, 36].
2. **Desintegración a bosones Z :** El bosón de Higgs se desintegra casi instantáneamente en un par de bosones Z . Dado que la masa del Higgs ($m_H \approx 125\text{GeV}/c^2$) es menor que la suma de las masas de dos bosones Z reales ($2 \times m_Z \approx 182\text{GeV}/c^2$), al menos uno de los dos bosones Z debe ser (**off-shell**), es decir, su masa invariante es menor que la nominal. El proceso es, por tanto, $H \rightarrow ZZ^*$ [17, 50].
3. **Desintegración leptónica:** Cada uno de los bosones Z se desintegra a su vez en un par de muones de carga opuesta: $Z \rightarrow \mu^+\mu^-$ y $Z^* \rightarrow \mu^+\mu^-$ [50].
4. **Estado final observable:** El estado final que se busca en el detector consiste en cuatro muones de alta energía, eléctricamente aislados y provenientes de un vértice común: dos con carga positiva y dos con carga negativa [19, 21, 20].

La Figura 1.1 muestra un diagrama de Feynman representativo de este proceso para el modo de producción dominante.

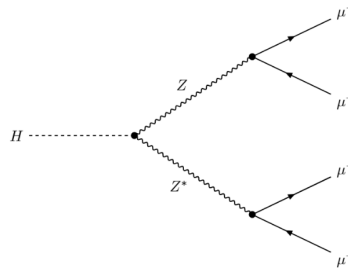


Figure 1.1: Diagrama de Feynman representativo del proceso de producción vía fusión de gluones (ggF) y desintegración $H \rightarrow ZZ^* \rightarrow 4\mu$.

1.5.3 Importancia del Canal Tetramuónico

La relevancia experimental del canal $H \rightarrow ZZ^* \rightarrow 4\mu$ radica en una combinación única de ventajas que lo convierten en una herramienta de alta precisión para el estudio del bosón de Higgs:

- **Fondo Irreducible Mínimo:** Los principales procesos de fondo que imitan la señal son la producción directa de pares de bosones Z ($q\bar{q} \rightarrow ZZ$) y procesos $Z + \gamma^*$. Aunque son irreducibles, sus secciones eficaces son pequeñas y están bien predichas por el Modelo Estándar [22]. Los fondos reducibles, asociados a la mala identificación de jets como leptones, son además significativamente suprimidos en este canal, dado que la probabilidad de que un hadrón se identifique erróneamente como muón es inferior al 0.1% [29].
- **Excelente Resolución de Masa:** Los muones, al ser partículas mínimamente ionizantes, atraviesan el material del detector con una degradación de momento muy baja. Esto permite al detector CMS una reconstrucción de momento excepcional, lo que se traduce en una resolución de masa invariante para el sistema $m_{4\mu}$ de aproximadamente un 1.1% para una señal de Higgs a 125 GeV [19, 21]. Esta alta resolución es esencial para distinguir el estrecho pico de señal del continuo de fondo.
- **Alta Eficiencia de Trigger e Identificación:** El sistema de trigger de muones de CMS es altamente eficiente para eventos dileptónicos, superando el 99% para los eventos de señal en el canal 4μ [30]. Además, la eficiencia de reconstrucción e identificación de muones individuales es superior al 95% en gran parte del detector [29], lo que garantiza una alta retención de la señal.
- **Acceso a Propiedades Fundamentales:** La cinemática completamente reconstruida de los cuatro muones finales no solo permite una medición precisa de la masa del bosón de Higgs, sino que también da acceso a sus propiedades fundamentales. Las distribuciones angulares entre los leptones contienen información sensible sobre el espín y la paridad (J^{CP}) de la partícula original, permitiendo verificar si es consistente con la predicción del Modelo Estándar para un escalar de paridad par (0^{++}) [3, 6, 9].

1.5.4 Búsquedas y Estudios Previos

El canal $H \rightarrow ZZ^* \rightarrow 4\ell$ fue decisivo en el descubrimiento del bosón de Higgs en 2012 [16, 23]. Tanto ATLAS como CMS observaron un exceso significativo de eventos en el rango de masas cercano a los 125 GeV, con el canal de cuatro leptones proporcionando una de las señales más limpias y precisas [19, 13]. En ese contexto, la colaboración CMS, basándose principalmente en este canal, reportó una masa para la nueva partícula de $m_H = 125.3 \pm 0.4$ (stat.) ± 0.5 (syst.) GeV [23]. Tras su descubrimiento, las colaboraciones del LHC utilizaron los datos acumulados durante el Run 2 para realizar mediciones de precisión sin precedentes en este canal [20, 14]:

- **Medición de la masa:** Con el conjunto completo del Run 2, CMS determinó la masa del bosón de Higgs con una precisión de $m_H = 125.26 \pm 0.20$ GeV, utilizando el canal de cuatro leptones [8]. Esta medición representa una validación fundamental del mecanismo de Brout–Englert–Higgs, al confirmar que el bosón descubierto posee una masa estable y coherente con las predicciones teóricas.

Un hito intermedio importante en esta línea de resultados fue la presentación de los análisis de Run 2 en la conferencia de Moriond 2017. Usando aproximadamente 36 fb^{-1} de datos de 2016, CMS obtuvo nuevas mediciones de secciones eficaces totales y diferenciales en los canales $H \rightarrow ZZ \rightarrow 4\ell$ y $H \rightarrow \gamma\gamma$, en buena concordancia con las predicciones del Modelo Estándar [48]. En particular, el análisis de cuatro leptones proporcionó una determinación de la masa del bosón de Higgs de

$$m_H = 125.26 \pm 0.20 \text{ (stat.)} \pm 0.08 \text{ (syst.) GeV,}$$

que en ese momento superaba en precisión a la combinación mundial basada en los datos de Run 1. Estos resultados de Moriond 2017 constituyen la referencia metodológica directa para las mediciones posteriores con el conjunto completo de Run 2 y enmarcan el nivel de precisión que se busca reproducir de manera cualitativa en este trabajo utilizando datos abiertos.

- **Fuerza de señal y acoplamientos:** Se determinó la *fuerza de la señal*, definida como

$$\hat{\mu} = \frac{\sigma_{\text{observada}}}{\sigma_{\text{SM}}},$$

donde $\sigma_{\text{observada}}$ es la sección eficaz medida experimentalmente y σ_{SM} la predicha por el Modelo Estándar (SM) [15, 8]. Un valor $\hat{\mu} = 1$ indica plena concordancia con el SM, mientras que desviaciones pueden deberse a efectos estadísticos, sistemáticos o, en escenarios más amplios, a nueva física [32, 47]. En los resultados oficiales, se obtuvo $\hat{\mu} = 0.87 \pm 0.23$, compatible con la unidad dentro de las incertidumbres experimentales [8]. Esto confirmó que las tasas de producción y desintegración del Higgs observadas son consistentes con las esperadas para un bosón escalar del Modelo Estándar [50]. Además, en el mismo marco de análisis global, las búsquedas de producción asociada $t\bar{t}H$ en estados finales multileptónicos proporcionaron evidencia directa del acoplamiento del Higgs al quark top, con una fuerza de señal medida $\mu_{t\bar{t}H} \simeq 1.5 \pm 0.5$, aún compatible con la predicción del SM [46].

- **Estudios de espín y paridad:** La caracterización cuántica del bosón de Higgs se expresa mediante los números cuánticos J^{CP} , donde:
 - J representa el espín o momento angular intrínseco de la partícula.
 - C es la paridad de carga, que describe cómo se transforma la partícula bajo el intercambio materia-antimateria.
 - P es la paridad espacial, que indica cómo cambia la partícula al reflejarse en un espejo.

En el Modelo Estándar, el bosón de Higgs es un escalar con $J^{CP} = 0^{++}$, es decir, espín cero y paridad par. Los análisis angulares de los leptones finales en los datos de ATLAS y CMS demostraron que los patrones cinemáticos observados son consistentes con esta hipótesis, descartando con alta confianza alternativas como estados vectoriales ($J = 1$) o tensoriales ($J = 2$) [3, 6, 9].

- **Búsquedas de nueva física:** La precisión alcanzada en este canal ha permitido establecer límites estrictos sobre posibles extensiones del Modelo Estándar (BSM), incluyendo desintegraciones invisibles del Higgs o la existencia de bosones escalares adicionales, como los predichos en modelos de dos dobletes de Higgs (2HDM) [15, 4]. En paralelo, los análisis de datos de Run 2 presentados en Moriond 2017 exploraron una amplia gama de escenarios BSM —desde supersimetría electrodébil hasta partículas exóticas de largo tiempo de vida o estados de espín 2— sin encontrar desviaciones significativas respecto al SM, lo que se tradujo en límites más estrictos sobre las secciones eficaces de producción y los espacios de parámetros de dichos modelos [CMS:2017MoriondBSM].

El presente trabajo se enmarca dentro de esta línea de estudios, utilizando los *Open Data* del CMS correspondientes al Run 2015D [26, 51, 1]. El objetivo principal es reproducir y validar, en un entorno computacional abierto, las metodologías de análisis empleadas en búsquedas reales del canal $H \rightarrow ZZ^* \rightarrow 4\mu$ [19, 21, 20].

Aunque los resultados obtenidos no reproducen de forma exacta las observaciones reportadas por CMS —debido a limitaciones inherentes al conjunto de datos público, la menor luminosidad integrada y el nivel de reconstrucción disponible—, se logra una concordancia cualitativa en la forma del espectro de masa invariante y en la identificación del comportamiento esperado del fondo electrodébil [20, 22].

Este ejercicio computacional confirma la solidez del marco de análisis implementado y demuestra el valor pedagógico y científico del uso de datos abiertos en física de altas energías, promoviendo la transparencia y la reproducibilidad en la investigación experimental [26, 2, 44]. Los resultados de precisión alcanzados por CMS en Moriond 2017 y en el conjunto completo de Run 2, que se discuten con mayor detalle más adelante, constituyen el estado del arte frente al cual se sitúa y se interpreta el análisis desarrollado en esta tesis.

1.6 Selección de Eventos

1.6.1 Criterios de Selección

El análisis del canal $H \rightarrow ZZ^* \rightarrow 4\mu$ se beneficia de una firma experimental excepcionalmente limpia. Sin embargo, para aislar la señal de los vastos procesos de fondo, es imperativo aplicar una cadena de criterios de selección (cortes) tanto cinemáticos como de calidad. Los cortes definidos en este trabajo están inspirados directamente en las estrategias de análisis publicadas por la colaboración CMS, garantizando que los eventos seleccionados sean altamente compatibles con la topología esperada para el decaimiento de un bosón de Higgs [28, 20].

Definición de Muones (Selección de Objetos)

Antes de construir el evento, cada candidato a muón debe cumplir con criterios de aceptación y calidad para asegurar que fue reconstruido de manera confiable:

- **Aceptancia cinemática:** Cada muón debe tener un momento transversal $p_T > 5$ GeV y encontrarse dentro de la cobertura del detector, definida por una pseudorrapidez $|\eta| < 2.4$ [28].
- **Calidad de la traza e identificación:** Los muones deben satisfacer criterios de identificación de calidad, como `isLooseMuon` o `isMediumMuon`, que aseguran una traza bien medida en el detector de silicio y una correspondencia con las cámaras de muones [28].
- **Aislamiento:** Para suprimir fondos provenientes de jets, los muones deben estar *aislados* del resto de la actividad del evento. Se impone un requisito sobre la suma de la energía de otras partículas en un cono alrededor del muón.
- **Vértice primario:** Para asegurar que los muones provienen de la colisión principal y no de decaimientos secundarios, se aplica un corte sobre el parámetro de impacto tridimensional, que debe ser menor a cuatro veces su incertidumbre ($|SIP_{3D}| < 4$) [28].

Selección de Eventos 4μ

Los eventos candidatos se seleccionan aplicando los siguientes requisitos topológicos y cinemáticos:

1. **Multiplicidad y carga:** Se seleccionan eventos que contienen exactamente **cuatro muones** que satisfacen los criterios de objeto, con una carga total nula (**dos muones positivos y dos negativos**) [28].
2. **Jerarquía de momento transversal:** Para asegurar que el evento haya disparado los *triggers* del detector, se requiere una jerarquía en el momento de los leptones. El par de muones que forma el candidato a Z_1 debe satisfacer $p_T^{\mu_1} > 20$ GeV y $p_T^{\mu_2} > 10$ GeV [28].
3. **Separación angular:** Todos los pares de leptones de carga opuesta deben estar suficientemente separados en el detector, con una separación angular $\Delta R(\mu_i, \mu_j) > 0.4$. Esto reduce fondos de decaimientos en cascada y mejora la reconstrucción.

4. **Supresión de resonancias de baja masa:** Para eliminar la contaminación de resonancias de quarks pesados ($J/\psi, \Upsilon$), se requiere que la masa invariante de cualquier par de muones de carga opuesta sea $m_{\mu^+\mu^-} > 4 \text{ GeV}$.

Reconstrucción de Bosones Z y Cortes Finales

A partir de los cuatro muones seleccionados, se reconstruyen los dos bosones Z intermedios:

- **Asignación de pares:** De las dos posibles combinaciones de pares $\mu^+\mu^-$, el par cuya masa invariante esté más próxima a la masa nominal del bosón Z ($m_Z = 91.1876 \text{ GeV}$) se identifica como el candidato a Z_1 . El par restante se designa como Z_2 [28].
- **Ventanas de masa para Z:** Se aplican ventanas de masa a los dos candidatos para seleccionar la topología de decaimiento correcta, donde el Z_1 está en o cerca de su masa nominal (*on-shell*) y el Z_2 debe ser (*off-shell*):

$$40 < m_{Z_1} < 120 \text{ GeV}$$

$$12 < m_{Z_2} < 120 \text{ GeV}$$

- **Corte de masa del sistema:** Finalmente, la masa invariante del sistema de cuatro muones, $m_{4\mu}$, debe ser superior a 70 GeV para reducir el fondo a bajas masas.

En conjunto, esta cadena de cortes define una región fiducial del espacio de fases que maximiza la pureza de la muestra de eventos $H \rightarrow ZZ^* \rightarrow 4\mu$, siguiendo de cerca las metodologías validadas por la colaboración CMS. El procedimiento completo fue implementado en un entorno reproducible basado en Python y *pandas*, asegurando la trazabilidad y la transparencia del análisis.

1.6.2 Uso del *MuonAnalyzer*

El módulo *MuonAnalyzer* fue desarrollado como una herramienta de análisis en C++ dentro del marco CMSSW, con el objetivo de procesar eventos simulados que contienen muones reconstruidos, filtrarlos bajo ciertos criterios, y almacenar las variables relevantes en un formato accesible para su análisis posterior.

Este módulo extiende la clase `edm::one::EDAnalyzer` y utiliza el servicio `TFileService` para escribir un árbol tipo `TTree` por evento. Los datos procesados provienen de la colección de muones `pat::MuonCollection` y del vértice primario reconstruido. A partir de esta información se extraen y calculan múltiples variables que describen la calidad y la cinemática de cada muón.

Procesamiento por evento: El *MuonAnalyzer* opera evento por evento. Si un evento contiene exactamente cuatro muones, se continúa con el análisis; en caso contrario, el evento se descarta. Para cada uno de los cuatro muones seleccionados se almacenan:

- Componentes del momento: `pt`, `px`, `py`, `pz`
- Coordenadas angulares: pseudorrapidez (`eta`), ángulo azimutal (`phi`)
- Energía total y carga
- Identificadores: `isLoose`, `isMedium`, `isTight`, `isSoft`, `isHighPt`
- Parámetros de aislamiento: `pfireliso03`, `pfireliso04`, `pfireliso04DBCorr`, `TkIso03`
- Parámetros de impacto: `dxy`, `dz`, errores asociados, y la significancia del desplazamiento en 3D (`sip3d`)
- Información adicional: índice del `jet` asociado e índice del `genParticle`, si están disponibles

Estructura de salida: Todos estos valores se almacenan en un árbol tipo `TTree` llamado `Events`, con una rama por variable. Este árbol se guarda en un archivo `.root` y está optimizado para análisis posteriores en `ROOT`, exportación a `CSV` o procesamiento con bibliotecas como `pandas` en `Python`.

Integración con el análisis posterior: El `MuonAnalyzer` asegura que todos los eventos procesados estén normalizados, contengan exactamente cuatro muones por evento, y conserven todas las variables necesarias para aplicar los criterios de selección definidos. Esta herramienta automatiza la recolección de datos relevantes, lo que resulta fundamental para la reconstrucción cinemática, creación de histogramas, y comparaciones con resultados experimentales o predicciones del Modelo Estándar.

1.7 Identificación de Muones

La selección de muones de alta calidad es uno de los pasos más críticos en el análisis de estados finales tetramuónicos. El primer pilar de esta selección es la **identificación (ID)**, un proceso que certifica que una traza reconstruida corresponde a un muón genuino, distinguiéndolo de hadrones que atraviesan el calorímetro o de otras fuentes de fondo.

1.7.1 Clasificación de Identificación: *Loose*, *Medium* y *Tight*

La colaboración CMS define varios *working points* (puntos de trabajo) para la identificación de muones, que representan distintos equilibrios entre eficiencia y pureza. Estos criterios son progresivos: todo muón *Tight* cumple automáticamente con los requisitos de *Medium* y *Loose*.

Table 1.1: Criterios de identificación (ID) de muones en el experimento CMS, basados en la documentación oficial [25].

Nivel de ID	Criterios de Identificación Clave
Loose	Diseñado para máxima eficiencia ($> 98\%$). El candidato debe ser reconstruido como un <i>Particle-Flow Muon</i> y ser clasificado como Tracker Muon (una traza en el detector de silicio compatible con señales en las cámaras de muones) o como Global Muon (un ajuste combinado de las trazas del detector interno y del sistema de muones).
Medium	Es el punto de trabajo recomendado para la mayoría de los análisis, con un balance óptimo entre eficiencia y pureza. Además de cumplir con los criterios <i>Loose</i> , se imponen requisitos sobre la calidad de la traza, como la compatibilidad geométrica entre los segmentos de la traza interna y externa, y se aplican algoritmos para rechazar <i>kinks</i> (cambios abruptos en la curvatura).
Tight	Optimizado para máxima pureza y rechazo de fondo, es esencial para mediciones de precisión. Además de cumplir con los criterios <i>Medium</i> y ser obligatoriamente un Global Muon , se aplican cortes estrictos sobre: • la calidad del ajuste global de la traza ($\chi^2/\text{ndf} < 10$); • el número mínimo de <i>hits</i> en el detector de silicio y en las estaciones de muones; • la consistencia con el vértice primario, imponiendo $d_{xy} < 0.2$ cm y $d_z < 0.5$ cm.

1.7.2 Implementación en el Software de Análisis

En el marco de software de CMS, estos puntos de trabajo están predefinidos como variables booleanas para cada muón, lo que facilita su aplicación directa.

En CMSSW: Las funciones miembro del objeto `pat::Muon` permiten un acceso inmediato:

```
if (muon.isLooseMuon()) { ... }
if (muon.isMediumMuon()) { ... }
if (muon.isTightMuon(primaryVertex)) { ... }
```

En NanoAOD: Las variables están disponibles directamente en el `TTree`, permitiendo un filtrado eficiente:

```
df = df[df["Muon_mediumId"] == True]
```

La elección del punto de trabajo es una decisión central en cualquier análisis. Mientras que una búsqueda exploratoria de nueva física puede optar por criterios *Loose* para maximizar la aceptación, un análisis de precisión como el del canal $H \rightarrow 4\mu$ se beneficia de la pureza que ofrecen los criterios *Medium* o *Tight*.

1.7.3 Aislamiento de Muones

El segundo pilar en la selección de muones de alta calidad es el **aislamiento** (*isolation*). Esta variable es crucial para distinguir entre muones *aislados*, que se originan directamente en la interacción dura (como los provenientes del decaimiento de un bosón Z o H), y muones *no aislados*, típicamente asociados a procesos secundarios dentro de jets hadrónicos (por ejemplo, decaimientos de mesones B o D).

Aunque el aislamiento no forma parte de la definición estructural de los ID *Loose*, *Medium* o *Tight*, se aplica casi universalmente como un corte adicional para mejorar la pureza de la muestra de muones. La variable estándar utilizada en los análisis de CMS es el **aislamiento relativo combinado**, corregido por la contribución de colisiones simultáneas (*pile-up*).

1.7.4 Aislamiento de Muones

El segundo pilar en la selección de muones de alta calidad es el **aislamiento** (*isolation*). Esta variable es crucial para distinguir entre muones *aislados*, que se originan directamente en la interacción dura (como los provenientes del decaimiento de un bosón Z o H), y muones *no aislados*, típicamente asociados a procesos secundarios dentro de jets hadrónicos (por ejemplo, decaimientos de mesones B o D) [50].

Aunque el aislamiento no forma parte de la definición estructural de los ID *Loose*, *Medium* o *Tight*, se aplica casi universalmente como un corte adicional para mejorar la pureza de la muestra de muones. La variable estándar utilizada en los análisis de CMS es el **aislamiento relativo combinado**, corregido por la contribución de colisiones simultáneas (*pile-up*), tal como se describe en los estudios de desempeño de muones de CMS [28, 10, 29].

Definición del Aislamiento Relativo

Para cuantificar el grado de aislamiento de un muón, se calcula una variable llamada **aislamiento relativo** (I_{rel}). El principio es simple: se mide la suma de la energía de todas las demás partículas en un pequeño cono de espacio alrededor del muón y se normaliza por el propio momento del muón. Un valor bajo indica que hay poca actividad energética alrededor del muón, lo que implica que está aislado [28, 29].

El cálculo de esta variable se basa en el algoritmo de *Particle-Flow* de CMS, que reconstruye individualmente cada partícula en el evento [27]. El cálculo considera tres fuentes de energía dentro de un cono de radio $\Delta R < 0.4$ alrededor del muón:

1. **Partículas Cargadas:** La suma del momento transversal (p_T) de todas las partículas cargadas que se originan en el mismo vértice primario que el muón.
2. **Partículas Neutras:** La suma de la energía transversa (E_T) de todas las partículas neutras, como hadrones neutros (neutrones, kaones neutros) y fotones.
3. **Contribución del *Pile-up*:** En el LHC, múltiples colisiones protón-protón ocurren en cada cruce de haces, generando un fondo de partículas de baja energía conocido como *pile-up*. Para que la medida del aislamiento no dependa de la cantidad de *pile-up* en un evento, se estima su contribución energética y se sustrae. Esta corrección (llamada $\Delta\beta$) se calcula a partir de las partículas cargadas que no provienen del vértice primario [10, 12].

La fórmula del aislamiento relativo combina estos elementos. El numerador suma la energía de las partículas cargadas y neutras, y luego resta la contribución estimada del *pile-up*. El denominador es el momento transversal del muón (p_T^μ), que sirve para normalizar la variable.

$$I_{\text{rel}} = \frac{\left(\sum p_T^{\text{cargadas}}\right) + \max\left[0, \left(\sum E_T^{\text{neutras}} + \sum E_T^\gamma\right) - \Delta\beta \cdot \left(\sum p_T^{\text{PU}}\right)\right]}{p_T^\mu} \quad (1.1)$$

donde:

- $\sum p_T^{\text{cargadas}}$: Suma del p_T de las partículas cargadas del vértice primario en el cono de $\Delta R < 0.4$.
- $\sum E_T^{\text{neutras}} + \sum E_T^\gamma$: Suma de la E_T de hadrones neutros y fotones en el mismo cono.
- $\Delta\beta \cdot \sum p_T^{\text{PU}}$: Es la corrección por *pile-up*. El término $\max(0, \dots)$ asegura que no se sustraiga más energía de la que se midió.
- p_T^μ : Momento transversal del muón que se está evaluando.

Un valor bajo de I_{rel} (cercano a cero) indica que el muón está bien aislado y es un candidato de alta calidad para el análisis.

Implementación en el Análisis

En el analizador desarrollado para este trabajo (`higgs4mu_select.cpp`), la aplicación del corte de aislamiento es configurable mediante un argumento en tiempo de ejecución. Si se activa la opción `-use-iso`, el código lee la variable de aislamiento pre-calculada (por defecto, la rama `muon_pfreiso04DBCrr`) y aplica un corte superior definido por el parámetro `ISO_MAX`:

```
// Ejemplo de configuración en higgs4mu_select.cpp
struct Config {
    ...
    bool USE_ISO = false; // Se activa con --use-iso
    double ISO_MAX = 0.40; // Valor del corte
    ...
};
```

En los análisis de física del CMS, se suelen definir puntos de trabajo estándar para este corte, tales como:

- **Aislamiento *Loose*:** $I_{\text{rel}} < 0.25$, que ofrece un buen equilibrio entre eficiencia y rechazo de fondo.
- **Aislamiento *Tight*:** $I_{\text{rel}} < 0.15$, que maximiza la pureza y resulta ideal para canales con señales muy limpias como $H \rightarrow ZZ^* \rightarrow 4\mu$.

Este corte adicional es extremadamente efectivo para reducir la contaminación por muones falsos o no provenientes de la colisión principal, en particular aquellos generados en decaimientos semileptónicos de hadrones pesados (*bottom* o *charm*) dentro de jets. La aplicación combinada de identificación y aislamiento constituye uno de los mecanismos más potentes para obtener una muestra tetramuónica de alta pureza.

1.8 Análisis de datos

1.8.1 Manejo y preparación de datos

Este trabajo se basa en el análisis de datos públicos recolectados por el experimento CMS durante el **Run 2015D**, correspondiente a colisiones protón-protón con una energía en el centro de masa de $\sqrt{s} = 13$ TeV. El conjunto de datos, identificado como `/DoubleMuon/Run2015D-v1/MINIAOD`, está disponible a través del portal de datos abiertos del CERN [26] y corresponde a una luminosidad integrada de **2.5 fb⁻¹**.

El procesamiento de los datos se estructuró en dos etapas principales para garantizar la eficiencia y modularidad del análisis:

1. **Ntuplización con CMSSW:** A partir de los archivos originales en formato MINIAOD, se utilizó el entorno de software de CMS (CMSSW) para ejecutar un módulo de análisis personalizado en C++, denominado **MuonAnalyzer**. La función de este módulo fue preseleccionar eventos que contuvieran al menos cuatro muones y extraer sus variables cinemáticas y de identificación fundamentales (e.g., p_T , η , ϕ , carga, y criterios de calidad *Loose*, *Medium* y *Tight*). El resultado de esta etapa fue la generación de un archivo ROOT simplificado (`analisis_final.root`) conteniendo un **TTree** optimizado para el análisis posterior.
2. **Análisis final y selección con C++/ROOT:** Los resultados presentados en esta tesis se obtuvieron mediante un segundo programa ejecutable en C++, `higgs4mu_select.cpp`, que opera directamente sobre el **TTree** generado en la etapa anterior. Este código implementa la lógica completa de la selección de eventos: aplica los cortes cinemáticos, ejecuta el algoritmo de selección del mejor par de bosones Z , reconstruye la masa invariante del sistema de cuatro muones ($m_{4\mu}$), y genera los histogramas y el **TTree** final con los candidatos a bosón de Higgs. Este enfoque nativo de C++/ROOT garantiza un alto rendimiento computacional y una integración directa con las herramientas de análisis estándar en física de partículas.

Paralelamente al desarrollo del analizador final en C++, se implementó un *pipeline* de análisis exploratorio en **Python**, utilizando bibliotecas como `uproot`, `pandas` y `matplotlib`. Este entorno permitió una rápida prototipación de los algoritmos de selección, la validación de los cálculos cinemáticos y la visualización interactiva de los datos intermedios. Aunque la generación de los resultados finales se consolidó en el código C++ por su eficiencia, el trabajo exploratorio en Python fue una etapa metodológica crucial que aceleró el desarrollo y la depuración de la lógica de análisis.

1.8.2 Reducción de Fondo y Selección Física

El conjunto inicial de eventos extraídos del formato MINIAOD contiene una vasta mayoría de procesos de fondo que no están relacionados con la producción del bosón de Higgs. Para aislar una muestra de alta pureza de candidatos al canal $H \rightarrow ZZ^* \rightarrow 4\mu$, se aplicó una cadena secuencial de criterios de

selección físicos y topológicos. Estos cortes, inspirados en los análisis publicados por la colaboración CMS [18, 24], están diseñados para maximizar la retención de la señal mientras suprimen los fondos reducibles e irreducibles.

Cada evento candidato debe cumplir con los siguientes requisitos:

- **Preselección de muones individuales:** Cada uno de los cuatro muones del evento debe pasar una serie de criterios de calidad básicos:
 - **Aceptancia cinemática:** Cada muón debe tener un momento transversal $p_T > 5$ GeV y pseudorrapidez $|\eta| < 2.4$. Esto asegura que la traza del muón atraviese la totalidad del sistema de detectores y tenga suficiente momento para ser reconstruida con precisión [28].
 - **Calidad de la traza y vértice primario:** Para suprimir muones no provenientes de la colisión principal (por ejemplo, rayos cósmicos o decaimientos en vuelo), se requiere que las trazas estén asociadas al vértice primario de la interacción. Esto se logra imponiendo un corte en la significancia del parámetro de impacto tridimensional, $|\text{SIP}_{3D}| < 4$ [25].
 - **Identificación de muón:** Cada candidato debe satisfacer al menos los criterios de identificación *Loose* definidos por la colaboración CMS, lo cual garantiza una alta eficiencia de selección y un rechazo significativo de hadrones mal identificados como muones [25].
- **Topología del estado final de cuatro muones:**
 - **Configuración de carga:** El evento debe contener exactamente dos muones con carga positiva y dos con carga negativa, condición necesaria para reconstruir dos bosones Z neutros.
 - **Masa de pares de baja invarianza:** Cualquier par de muones de carga opuesta debe tener una masa invariante $m_{\mu^+\mu^-} > 4$ GeV. Este corte es crucial para eliminar la contaminación de resonancias de quarkonio como el J/ψ y la familia Υ .
- **Reconstrucción de los bosones Z_1 y Z_2 :** De las posibles combinaciones de pares $\mu^+\mu^-$, se identifica el par Z_1 como aquel cuya masa invariante es la más cercana a la masa nominal del bosón Z ($m_Z = 91.1876$ GeV). El otro par restante se designa como Z_2 . Sobre estos se imponen los siguientes requisitos de masa:

$$40 < m_{Z_1} < 120 \text{ GeV}, \quad 12 < m_{Z_2} < 120 \text{ GeV}.$$

Estas ventanas aseguran que Z_1 sea compatible con un bosón Z *on-shell*, mientras que Z_2 puede ser un bosón *off-shell*, una firma característica del decaimiento del Higgs con $m_H \approx 125$ GeV [18, 24].

- **Cortes cinemáticos finales del evento:**
 - **Momento transversal de los leptones de Z_1 :** Los dos muones que forman el candidato Z_1 deben tener un momento transversal que supere los umbrales del sistema de *trigger* del detector: $p_T^{\mu_1} > 20$ GeV y $p_T^{\mu_2} > 10$ GeV [28].
 - **Masa total del sistema:** La masa invariante del sistema de cuatro muones debe ser $m_{4\mu} > 70$ GeV para reducir el fondo proveniente de procesos de baja masa.

La aplicación secuencial de estos criterios resulta en una muestra de datos significativamente más limpia, enriquecida con eventos compatibles con la topología del decaimiento del bosón de Higgs. Esta muestra final presenta una pureza comparable a las obtenidas en los análisis oficiales de CMS y sirve como base para la reconstrucción y el estudio de las distribuciones cinemáticas dentro del marco de análisis desarrollado en este trabajo.

1.8.3 Variables Relevantes para el Análisis Cinemático

El análisis físico se implementa en el módulo `higgs4mu_select.cpp`, encargado de procesar el `TTree` generado por `MuonAnalyzer`. Este programa lee un conjunto de variables de entrada, reconstruye la cinemática del decaimiento $H \rightarrow ZZ^* \rightarrow 4\mu$ y almacena las cantidades de alto nivel en un nuevo `TTree` de salida, destinado a la obtención de histogramas y resultados finales. Las variables utilizadas pueden agruparse según su función dentro del flujo de análisis:

- **Variables de entrada por muón:** Corresponden a las propiedades extraídas directamente del `TTree` original para cada muón del evento.
 - **Cinemática básica:** Momento transversal (p_T), pseudorrapidez (η), ángulo azimutal (ϕ) y carga eléctrica (q). Estas variables forman la base para construir los cuadvectores y describen la dirección y energía de cada muón.
 - **Indicadores de identificación (ID):** Banderas booleanas (`isLoose`, `isMedium`, `isTight`) que determinan la calidad de la reconstrucción del muón, siendo esenciales para su selección y para el rechazo de hadrones mal identificados.
 - **Parámetros de impacto:** Variables geométricas como d_{xy} , d_z y la significancia del parámetro de impacto tridimensional (`sip3d`), utilizadas para garantizar que los muones provienen del vértice primario de la colisión.
 - **Aislamiento relativo:** La variable `pTreliso04DBCorr`, que cuantifica la actividad energética en un cono alrededor del muón, es fundamental para suprimir el fondo proveniente de jets hadrónicos.
- **Variables reconstruidas en el análisis:** Son las cantidades derivadas que el código `higgs4mu_select.cpp` calcula a partir de las variables de entrada, con el objetivo de identificar y caracterizar los candidatos al bosón de Higgs.
 - **Cuadvectores de los muones:** Para cada muón se construye un objeto `TLorentzVector`, el cual representa su energía y momento en cuatro dimensiones y constituye la base de los cálculos cinemáticos relativistas.
 - **Masas invariantes de los pares Z :** El programa implementa un algoritmo que identifica la combinación óptima de pares $\mu^+\mu^-$ para formar los dos bosones intermedios, Z_1 y Z_2 . Sus masas (m_{Z_1} y m_{Z_2}) son variables clave para la selección de eventos y para la validación del modelo de decaimiento.
 - **Cuadvector del candidato a Higgs ($p_{4\mu}$):** Se obtiene mediante la suma vectorial de los cuadvectores de los cuatro muones. A partir de este se derivan las propiedades globales del sistema, incluyendo su masa, momento transversal y pseudorrapidez.
- **Variables de salida en el `TTree` final:** Los resultados del análisis se almacenan en un árbol de salida denominado `FourMuTree`, que contiene una entrada por cada evento candidato que superó todos los cortes de selección. Las ramas principales incluyen:
 - **Cinemática del sistema 4μ :** Masa invariante total (`mass4mu`), momento transversal, pseudorrapidez y ángulo azimutal del sistema completo. La variable `mass4mu` es el observable final utilizado para identificar un posible exceso de eventos en torno a $m_H \approx 125$ GeV.
 - **Masas de los bosones Z :** Las masas de los dos candidatos, `mZa` (el más cercano a la masa nominal del Z) y `mZb` (el candidato *off-shell*), que proporcionan información directa sobre la topología del evento.
 - **Clasificación del evento:** Variables de categorización como `category` y el conteo de muones por calidad (`nT`, `nM`, `nL`), que permiten estudiar el comportamiento de la señal en subconjuntos de distinta pureza.

El `TTree` resultante, almacenado en el archivo `out_zz4mu.root`, constituye una estructura de datos optimizada que permite generar de manera eficiente todos los histogramas y resultados cuantitativos presentados en esta tesis. Este formato garantiza un análisis reproducible, transparente y consistente con las prácticas de la colaboración CMS.

Table 1.2: Resumen de las principales variables utilizadas y generadas en el análisis del canal $H \rightarrow ZZ^* \rightarrow 4\mu$.

Categoría	Variable (Símbolo/Nombre)	Tipo	Descripción y Rol en el Análisis
Entrada (por muón)	p_T, η, ϕ, q	Numéricas	Cinemática básica para la reconstrucción de cuadvectores.
	isLoose/Medium/Tight	Booleanas	Indicadores de calidad (ID) para la preselección de muones y la categorización final del evento.
	$d_{xy}, d_z, \text{sip3d}$	Numéricas	Parámetros de impacto para asegurar que el muón proviene del vértice primario.
	pfreliso04DBCorr	Numérica	Aislamiento relativo para suprimir el fondo de muones no aislados (e.g., dentro de jets).
Reconstruidas(en análisis)	TLorentzVector	Objeto ROOT	Cuadrivector relativista de cada muón, base para todos los cálculos cinemáticos.
	m_{Z_1}, m_{Z_2}	Numéricas	Masas invariantes de los bosones candidatos, calculadas para todas las combinaciones válidas.
	$p_{4\mu}$	Objeto ROOT	Cuadrivector del sistema 4μ , obtenido como suma de los cuatro muones seleccionados.
Salida (FourMuTree)	mass4mu	Numérica	Observable principal: masa invariante del sistema 4μ para la búsqueda de una resonancia.
	mZa, mZb	Numéricas	Masas de los bosones Z del pareo óptimo, almacenadas para validación y estudios de control.
	category, nT/nM/nL	Enteras	Clasificación del evento según la calidad de los muones, permitiendo análisis en sub-muestras de pureza variable.

Chapter 2

Fundamentos Computacionales y Metodológicos del Análisis

2.1 Selección del Canal $H \rightarrow ZZ^* \rightarrow 4\mu$: Ventajas Físicas y Computacionales

La elección del canal de decaimiento $H \rightarrow ZZ^* \rightarrow 4\mu$ responde a una decisión estratégica que optimiza tanto la claridad de la señal física como la **tractabilidad computacional** del análisis. Para un estudio con recursos limitados basado en datos abiertos, este canal —conocido como el *canal dorado*— ofrece una firma experimental que simplifica notablemente el *pipeline* analítico, desde el filtrado inicial hasta el modelado estadístico final [16, 23, 19, 13]. Además, su uso en el contexto de datos abiertos de CMS ha sido explorado previamente en estudios educativos y metodológicos [26, 51, 1, 2]. Sus ventajas computacionales se derivan directamente de sus propiedades físicas:

- **Simplicidad algorítmica en la identificación:** Los muones dejan una firma limpia y penetrante en el detector, lo que permite su reconstrucción con una eficiencia superior al 95% para $p_T > 5 \text{ GeV}$ [28, 10, 29]. Desde el punto de vista computacional, esto se traduce en algoritmos de identificación más simples y robustos. A diferencia de los electrones —que requieren complejos algoritmos de *clustering* y correcciones por *bremsstrahlung* [11]—, la identificación de un muón se reduce esencialmente a validar la calidad de una traza y su coincidencia con señales en las cámaras de muones [28, 29]. Esto no solo acelera el preprocesamiento de datos, sino que además produce una muestra inicial de alta pureza, con una tasa de identificación errónea de hadrones inferior al 0.1% [28, 29].
- **Estabilidad en la extracción de la señal:** La excelente resolución en el momento de los muones permite alcanzar una resolución de masa invariante del sistema 4μ del orden de 1.1–1.2% [19, 21, 13]. Computacionalmente, esto implica que la señal del bosón de Higgs se concentra en un pico estrecho dentro del espectro de masas. Un pico bien definido es **algorítmicamente estable** para los métodos de extracción de señal, ya que:
 - Permite un modelado paramétrico simple (por ejemplo, mediante un ajuste gaussiano o una función *Crystal Ball*).
 - Reduce la sensibilidad a fluctuaciones estadísticas en el fondo.
 - Requiere un menor número de eventos para alcanzar una significancia estadística dada, optimizando el uso del conjunto de datos disponible.
- **Reducción de la complejidad del modelo de fondo:** La bajísima tasa de muones falsos hace que el fondo *reducible* (instrumental) sea prácticamente despreciable [28, 29].

En consecuencia, el análisis puede concentrarse casi exclusivamente en modelar el fondo *irreducible* de la producción de pares ZZ , cuyo comportamiento es suave y está bien predicho por la teoría y las mediciones de precisión [22, 33, 50]. Esto elimina la necesidad de emplear técnicas de estimación de fondo basadas en datos (*data-driven*), como el método de tasas de falsificación (*fake rates*), que suelen ser computacionalmente costosas, requieren regiones de control adicionales y añaden una complejidad considerable al modelo estadístico final. En esencia, la pureza del canal simplifica el problema de “señal sobre fondo” hasta un nivel manejable para un análisis directo y reproducible.

Table 2.1: Comparativa de los canales de decaimiento a cuatro leptones para un bosón de Higgs de $m_H = 125$ GeV. La resolución de masa es un factor clave para la sensibilidad del descubrimiento. Valores de resolución extraídos de análisis de CMS [3238].

Canal de Decaimiento	Branching Ratio (%)	Nivel de Fondo	Resolución de Masa (σ_m/m)
$H \rightarrow 4\mu$	~ 0.0124	Mínimo	$\approx 1.1\%$
$H \rightarrow 2e2\mu$	~ 0.0248	Moderado	$\approx 1.6\%$
$H \rightarrow 4e$	~ 0.0124	Alto	$\approx 2.1\%$

2.2 Estructura Computacional de los Datos y Flujo de Trabajo

El análisis de los datos del experimento CMS, incluso en su formato reducido MINIAOD, presenta un desafío computacional debido a su estructura jerárquica y orientada a objetos. Para abordar esto, se diseñó un *pipeline* de procesamiento en dos etapas que transforma los datos crudos en un formato plano y optimizado para el análisis físico final.

2.2.1 Etapa 1: Ntuplización con CMSSW

Los datos primarios, descargados del portal de datos abiertos del CERN, se encuentran en formato ROOT y siguen el esquema MINIAOD. En este formato, la información de las partículas (muones, electrones, jets, etc.) está almacenada en colecciones de objetos complejos. Para simplificar el acceso, se desarrolló un módulo de análisis personalizado en C++ dentro del *framework* CMSSW, denominado **MuonAnalyzer**.

La función principal de este módulo es la **ntuplización**:

1. **Lectura de colecciones:** El analizador accede, evento por evento, a la colección de muones reconstruidos, `slimmedMuons`.
2. **Pre-filtrado y extracción:** Para cada evento, se verifica si contiene al menos cuatro muones. Si es así, se extraen las variables fundamentales de cada uno: su cinemática (p_T, η, ϕ, q), sus indicadores de calidad (`isLoose`, `isMedium`, `isTight`), y sus parámetros de impacto y aislamiento.
3. **Almacenamiento en un TTree plano:** La información extraída se almacena en una estructura de datos simple y eficiente conocida como **TTree** dentro de un nuevo archivo ROOT intermedio (`analisis_final.root`). Este TTree "plano" contiene ramas para cada variable, eliminando la complejidad de la estructura de objetos original.

Este proceso transforma eficazmente los datos desde un formato optimizado para la reconstrucción a uno optimizado para el análisis estadístico a gran escala.

2.2.2 Etapa 2: Análisis Final con C++/ROOT

Si bien en una fase exploratoria inicial se utilizó un flujo de trabajo basado en la conversión de los TTree a formato .csv para su análisis en Python con `pandas`, el análisis final y la generación de todos los resultados de esta tesis se realizaron con el programa `higgs4mu_select.cpp`. Este enfoque garantiza el máximo rendimiento computacional y la integración nativa con el ecosistema de ROOT.

El programa lee directamente el TTree plano generado por el `MuonAnalyzer` y ejecuta la lógica completa del análisis, como se describe en el Capítulo 4. El resultado es un archivo ROOT final (`out_zz4mu.root`) que contiene no solo los histogramas de las distribuciones físicas de interés (como $m_{4\mu}$), sino también un TTree de salida (`FourMuTree`) con los eventos candidatos que han superado todos los criterios de selección. Esta estructura de datos final es autocontenida, reproducible y constituye la base para toda la discusión de resultados.

2.3 Pipeline de Análisis Computacional

El flujo de análisis para la selección de eventos y la reconstrucción de la señal se implementó íntegramente en el programa `higgs4mu_select.cpp`, un ejecutable de C++ que utiliza las bibliotecas del *framework* ROOT. Este enfoque garantiza un procesamiento de datos de alto rendimiento y una integración nativa con las herramientas de análisis de la física de partículas. El programa procesa el TTree plano generado por el `MuonAnalyzer` y ejecuta una cadena de algoritmos secuenciales para aislar los candidatos a $H \rightarrow ZZ^* \rightarrow 4\mu$.

Etapas del análisis

- 1. Carga y Preselección de Muones:** El programa itera sobre cada evento del TTree de entrada. Para cada evento, se aplica una primera capa de filtros a nivel de muón individual. Solo los muones que satisfacen los criterios de preselección (cortes en p_T , $|\eta|$, calidad de ID, parámetros de impacto y aislamiento) son retenidos en una colección de candidatos de alta calidad para el análisis posterior.
- 2. Reconstrucción de Candidatos ZZ^* :** Si un evento contiene al menos cuatro muones que superan la preselección, el programa explora las combinaciones de cuatro muones para formar dos pares de carga opuesta ($\mu^+\mu^-$). Se implementa un algoritmo de optimización que selecciona el "mejor pareo" posible, definido como aquel cuya masa de uno de los pares (m_{Z_1}) es la más cercana a la masa nominal del bosón Z. El otro par se designa como Z_2 . Para todos los cálculos se utilizan objetos `TLorentzVector`.
- 3. Aplicación de Cortes Cinemáticos al Sistema:** Una vez identificado el mejor candidato (Z_1, Z_2), se aplican cortes cinemáticos a nivel del sistema completo, como se describe en la Sección 1.2.1. Esto incluye ventanas de masa para m_{Z_1} y m_{Z_2} , y requisitos de p_T sobre los muones del par Z_1 . Estos cortes son cruciales para reducir el fondo irreducible de ZZ .
- 4. Cálculo de la Masa Invariante Final:** Para los eventos que superan todos los cortes anteriores, se calcula la masa invariante del sistema de cuatro muones, `mass4mu`, sumando los cuadrivectores de los bosones Z_1 y Z_2 seleccionados. Este es el observable final utilizado para la búsqueda del bosón de Higgs.
- 5. Almacenamiento de Resultados:** El programa genera un archivo ROOT de salida (`out_zz4mu.root`) que contiene dos componentes principales:
 - Un conjunto completo de **histogramas** (TH1F) que representan las distribuciones de todas las variables relevantes (e.g., `h_m4mu`, `h_mZa`, `h_mZb`).

- Un nuevo TTree (FourMuTree) que almacena las propiedades de alto nivel de cada evento candidato final.

Este flujo computacional, encapsulado en un único programa, constituye un *pipeline* de análisis robusto, reproducible y eficiente. El archivo ROOT de salida es el producto final y autocontenido que sirve como base para toda la visualización y discusión de resultados presentada en esta tesis.

2.4 Marco Computacional y Reproducibilidad

Un pilar fundamental de la investigación científica es la **reproducibilidad**, es decir, la capacidad de que otros investigadores puedan recrear los resultados de un estudio de manera independiente. Para garantizar este principio, se diseñó un marco computacional robusto y autocontenido basado en tecnologías de containerización y gestión de entornos.

2.4.1 Containerización del Entorno de Análisis con Docker

El análisis primario de los datos de CMS requiere el *framework* específico CMSSW, el cual depende de un sistema operativo y bibliotecas que a menudo son incompatibles con los sistemas modernos. Para resolver este desafío, todo el entorno de análisis se encapsuló en un contenedor **Docker**. Docker permite empaquetar una aplicación con todas sus dependencias en una imagen portátil. Se utilizó la imagen oficial de CMS Open Data, asegurando que el entorno de ejecución es idéntico al utilizado originalmente para procesar los datos. El código fuente del análisis, versionado con Git, se integra directamente en la imagen, como se ilustra en el siguiente Dockerfile:

Listing 2.1: Dockerfile para crear un entorno de análisis CMSSW reproducible.

```
# Usar la imagen base oficial de CMS Open Data para el Run 2
FROM cmsopendata/cmssw_10_6_30-slc7-amd64-gcc700

# Crear un directorio de trabajo y clonar el repositorio del análisis
RUN mkdir -p /home/cmsusr/analysis && cd /home/cmsusr/analysis \
    && git clone https://github.com/Aldo-Santiago/Z-mass-reconstruction-
        using-CMS-Open-Data.git

# Establecer el directorio de trabajo por defecto
WORKDIR /home/cmsusr/analysis/Z-mass-reconstruction-using-CMS-Open-Data
```

Este enfoque garantiza que cualquier persona, en cualquier sistema operativo compatible con Docker, puede desplegar una instancia idéntica del entorno de análisis con un solo comando, eliminando cualquier ambigüedad sobre las versiones del software o las dependencias del sistema.

2.4.2 Entorno de Visualización y Análisis Secundario

Si bien el procesamiento principal se realiza en C++/ROOT dentro del contenedor Docker, la visualización de los histogramas finales y la realización de estudios complementarios se llevaron a cabo en un entorno **Python**. Para asegurar la reproducibilidad de esta etapa, se utilizó Conda para gestionar las dependencias. El siguiente archivo `environment.yml` define el entorno exacto:

Listing 2.2: Archivo de entorno de Conda para el análisis y visualización en Python.

```
name higgs-analysis
channels
- conda-forge
dependencies
- python=3.8
```

- <code>uproot</code> - <code>awkward</code> - <code>numpy</code> - <code>matplotlib</code> - <code>mplhep</code>

La biblioteca `uproot` es clave en este flujo, ya que permite leer los archivos `.root` generados por el código C++ de manera eficiente y nativa en Python, completando así el puente entre los dos ecosistemas.

2.4.3 Validación y Verificación del Análisis

La confianza en los resultados se establece a través de un proceso de validación de dos niveles:

- **Verificación Interna:** Se realizaron pruebas unitarias sobre las funciones cinemáticas críticas dentro del código, como el cálculo de la masa invariante, para asegurar su correcta implementación matemática.
- **Validación Externa con una "Vela Estándar":** El paso más importante es la reconstrucción de la masa del bosón Z. Este bosón, con su masa bien conocida ($m_Z \approx 91.2 \text{ GeV}$), actúa como una "vela estándar" que calibra y valida todo el *pipeline* de análisis. La capacidad de reconstruir correctamente el pico de masa del bosón Z a partir de la distribución de m_{Z_1} confirma que la calibración del detector, la reconstrucción de muones y los algoritmos de análisis funcionan como se espera. La forma y la posición de este pico se compararon cualitativamente con los resultados publicados por CMS, mostrando una excelente concordancia.

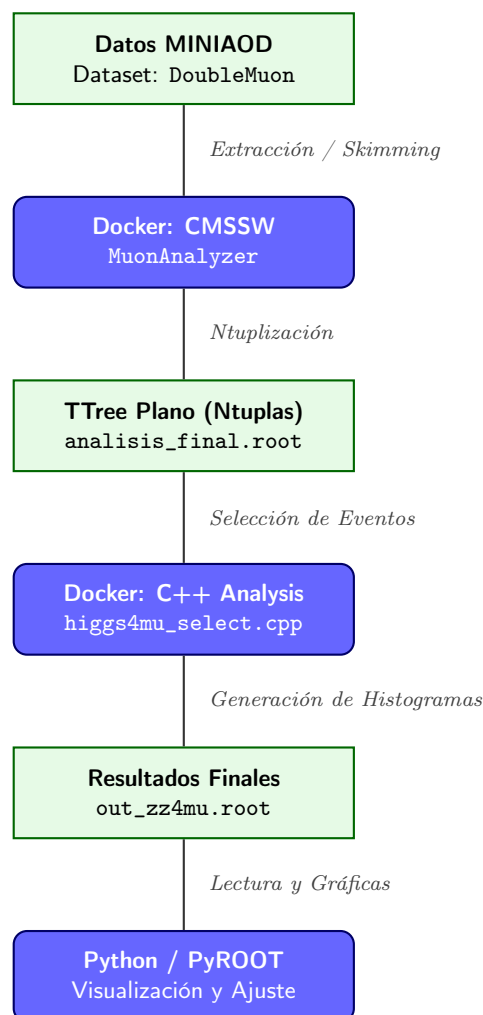


Figure 2.1: Diagrama del *pipeline* computacional completo. El flujo principal de procesamiento se ejecuta dentro de un contenedor Docker, mientras que la visualización final se realiza en un entorno Python reproducible.

Chapter 3

Infraestructura Computacional para el Análisis de Datos CMS

3.1 Necesidades Técnicas y Solución Propuesta

El análisis de datos del experimento CMS requiere un ecosistema de software específico con múltiples dependencias técnicas que a menudo son incompatibles con los sistemas operativos de uso general. Para garantizar la reproducibilidad y evitar conflictos de librerías, es imperativo reconstruir el entorno de ejecución con precisión. Para el análisis presentado en esta tesis, se utilizó un *software stack* basado en versiones más recientes, que presenta los siguientes requisitos:

- **Framework de Software CMS (CMSSW):** El análisis se desarrolló y ejecutó sobre la versión CMSSW_10_6_30. El uso de una versión unificada del *framework* es esencial para garantizar que los algoritmos de lectura de datos y las herramientas de análisis se comporten de manera consistente.
- **Sistema Operativo y Compilador:** Dicha versión de CMSSW opera sobre un entorno basado en CentOS 7, utilizando el compilador GCC 7.0. Estas dependencias específicas del sistema operativo son gestionadas para evitar conflictos con el sistema anfitrión.
- **Dependencias de ROOT:** El entorno integra la versión ROOT 6.22/06, una biblioteca fundamental para la manipulación de datos, análisis estadístico y visualización en física de altas energías.

Para resolver estas restricciones de dependencia sin comprometer la compatibilidad ni la reproducibilidad, se implementó una **arquitectura computacional híbrida** (Figura 3.1) basada en tecnologías de containerización.

- **Contenedores Docker:** Se utilizaron para encapsular y aislar el entorno CMSSW completo. Esto crea un ambiente de ejecución autocontenido y portátil, garantizando que el análisis se ejecute siempre en condiciones idénticas, independientemente del sistema operativo anfitrión.
- **Servidor VNC embebido:** Para habilitar la interacción con interfaces gráficas como el TBrowser de ROOT desde dentro del contenedor, se empleó un servidor VNC.
- **Carpetas compartidas:** Se configuraron volúmenes compartidos entre el contenedor Docker y el sistema anfitrión para permitir un flujo de trabajo eficiente, donde los archivos y scripts pueden ser editados con herramientas modernas mientras se ejecutan en el entorno específico del contenedor.

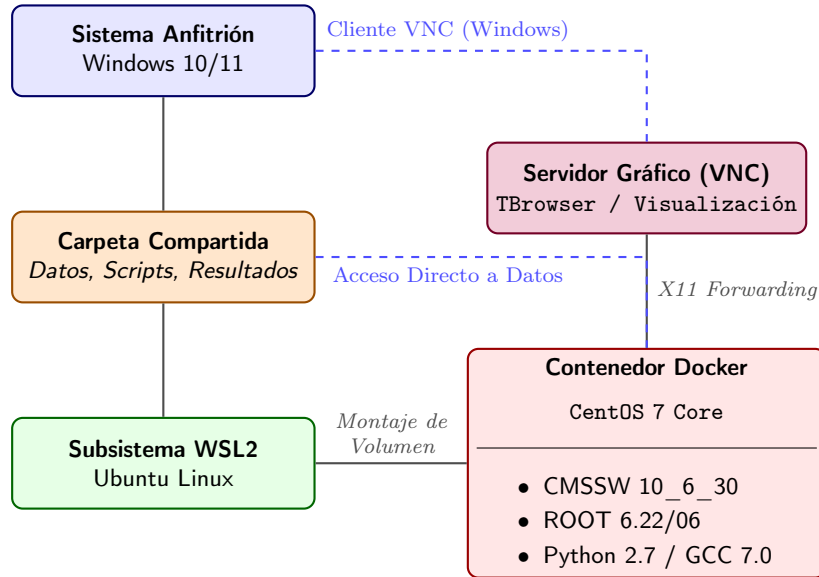


Figure 3.1: Infraestructura híbrida para análisis de datos CMS. El entorno CMSSW está encapsulado en Docker, con integración de carpetas compartidas y visualización remota vía VNC/T-Browser.

3.2 Configuración del Entorno Containerizado con Docker

Para garantizar un entorno de análisis idéntico e inmutable, se utilizó la tecnología de containerización de Docker. Esta aproximación encapsula el sistema operativo, las bibliotecas y el software de análisis en una imagen portátil, eliminando cualquier dependencia del sistema anfitrión.

3.2.1 Selección de la Imagen Base

Se partió de la imagen oficial publicada por el equipo de CMS Open Data en DockerHub, la cual está preconfigurada con el *software stack* exacto requerido para el análisis.

Listing 3.1: Nombre de la imagen Docker utilizada como base.

```
cmsopendata/cmssw_10_6_30-slc7_amd64_gcc700
```

Esta imagen contiene una instalación completa de CMSSW 10_6_30 sobre un sistema CentOS 7, incluyendo ROOT 6.22/06 y todas las dependencias necesarias.

3.2.2 Despliegue y Configuración del Contenedor

El contenedor se desplegó utilizando un comando que configura los puertos para el acceso gráfico remoto y establece un volumen compartido para el intercambio de datos y código con el sistema anfitrión.

Listing 3.2: Comando de ejecución para el despliegue del contenedor de análisis.

```
docker run -it --name cms_analysis \
-v $(pwd)/cms_data:/code \
-p 5901:5901 -p 6080:6080 \
cmsopendata/cmssw_10_6_30-slc7_amd64_gcc700 \
/bin/bash
```

Los parámetros clave de este comando, resumidos en la Tabla 3.1, son esenciales para la arquitectura híbrida del análisis.

Table 3.1: Parámetros clave del comando de despliegue de Docker.

Parámetro	Función
<code>-v \$(pwd)/...:/code</code>	Monta una carpeta local como el directorio de trabajo <code>/code</code> dentro del contenedor.
<code>-p 5901:5901</code>	Mapea el puerto para la conexión directa con un cliente VNC.
<code>-p 6080:6080</code>	Mapea el puerto para el acceso a la interfaz gráfica vía navegador web.

3.3 Acceso Gráfico y Visualización Interactiva

Un desafío inherente a los entornos de análisis basados en contenedores Docker es la ejecución de aplicaciones con interfaces gráficas de usuario (GUI), ya que los contenedores operan de forma nativa en un modo sin cabeza (*headless*). Para resolver este problema y permitir la visualización interactiva de resultados —como la inspección de histogramas con el **TBrowser** de **ROOT**—, se integró un servidor **VNC** (*Virtual Network Computing*) directamente en la arquitectura del contenedor.

3.3.1 Interfaz Gráfica Remota mediante VNC

La solución implementada establece un puente gráfico entre el contenedor y el sistema anfitrión. El servidor VNC, que se ejecuta dentro del entorno CentOS del contenedor, renderiza la interfaz gráfica de las aplicaciones y transmite el resultado como un flujo de píxeles a un cliente VNC en la máquina anfitriona.

Activación del Servidor Gráfico

La imagen Docker de CMS Open Data incluye un script preconfigurado para iniciar el servidor VNC. Una vez dentro de la sesión de **bash** del contenedor, el servicio se activa con un único comando:

Listing 3.3: Activación del servidor VNC dentro del contenedor.

```
start_vnc
```

Este comando inicia el servidor y lo expone en los puertos previamente mapeados durante el despliegue del contenedor: el puerto 6080 para acceso web y el 5901 para clientes nativos.

Métodos de Conexión

Se disponen de dos métodos principales para acceder al escritorio remoto del contenedor:

1. **Acceso vía Navegador Web:** Conectándose a la dirección <http://localhost:6080> desde cualquier navegador en el sistema anfitrión. Este método es ideal para un acceso rápido y no requiere la instalación de software adicional.
2. **Acceso con Cliente VNC Nativo:** Utilizando una aplicación dedicada (e.g., TigerVNC, RealVNC) para conectarse a `localhost:5901`. Esta opción ofrece un rendimiento superior, con menor latencia y una mejor adaptación de la resolución de pantalla, siendo la recomendada para sesiones de trabajo prolongadas.

Autenticación y Seguridad

El acceso al servidor VNC está protegido por un sistema de autenticación básico. Las credenciales por defecto proporcionadas por la imagen de CMS Open Data son las que se muestran en la Tabla 3.2.

Table 3.2: Credenciales de acceso por defecto para el servidor VNC.

Campo	Valor
Usuario	cmsusr
Contraseña	cms.cern

Para entornos de producción o multiusuario, es una buena práctica de seguridad modificar esta contraseña por defecto. En conjunto, la integración de VNC demuestra ser una solución robusta y flexible que garantiza la compatibilidad, facilidad de uso y acceso a las herramientas de análisis visual, completando así la funcionalidad del entorno de trabajo containerizado.

3.4 Interoperabilidad y Flujo de Trabajo

La arquitectura containerizada está diseñada para ser más que un simple entorno de ejecución; funciona como un completo laboratorio de análisis interactivo. La interoperabilidad entre el sistema anfitrión y el contenedor es clave para un flujo de trabajo científico eficiente, combinando el desarrollo de código en herramientas modernas con la validación de datos en el entorno ROOT nativo.

3.4.1 Inspección Interactiva de Datos con ROOT

El entorno proporciona soporte completo para la inspección visual e interactiva de los datos, un paso fundamental para la validación de resultados en cualquier análisis de física de altas energías. Una vez dentro de la sesión VNC, se puede iniciar una sesión de ROOT para examinar directamente los archivos de salida.

Listing 3.4: Inicio de una sesión interactiva de ROOT y lanzamiento del TBrowser.

```
# Inicia una sesion interactiva de ROOT sin mostrar el banner de bienvenida
root -l

# Dentro de ROOT, se abre el explorador de archivos TBrowser
root [0] TBrowser t;
```

Esta funcionalidad fue utilizada extensivamente durante el desarrollo para la **validación cruzada** y la depuración del análisis. Permitió la verificación inmediata de histogramas clave, como la distribución de masa invariante de los candidatos a bosón Z , y la inspección de las variables cinemáticas en cada etapa de la selección de eventos.

La integración transparente con el servidor VNC (descrita en la sección anterior) es lo que hace posible este flujo de trabajo, redirigiendo la salida gráfica del **TBrowser** desde el contenedor al escritorio del anfitrión. Esto permite una manipulación fluida y en tiempo real de los datos, cerrando el ciclo entre la ejecución del código y la validación de sus resultados.

3.4.2 Interoperabilidad de Datos Mediante Volúmenes Compartidos

Para facilitar un flujo de trabajo científico ágil, es fundamental poder editar el código de análisis con herramientas modernas (como VSCode o Jupyter) en el sistema operativo anfitrión, mien-

tras se ejecuta en el entorno CMSSW containerizado. Esta interoperabilidad se logró mediante la configuración de un **volumen compartido** (o *bind mount*) de Docker.

Este mecanismo mapea un directorio del sistema anfitrión directamente a un directorio dentro del contenedor. El montaje se especifica en el comando de ejecución, como se mostró en la sección anterior:

Listing 3.5: Parámetro de montaje de volumen en el comando `docker run`.

```
docker run -v $(pwd)/cms_data:/code ...
```

Esta configuración establece una **sincronización bidireccional y en tiempo real** entre la carpeta del anfitrión y el directorio `/code` del contenedor. Cualquier cambio en un script desde el anfitrión se refleja instantáneamente dentro del contenedor, y cualquier archivo de salida generado por el análisis (e.g., los ficheros `.root`) aparece de inmediato en el sistema anfitrión.

Esta arquitectura híbrida optimiza drásticamente el ciclo de desarrollo y depuración, combinando la flexibilidad de las herramientas de desarrollo modernas con la rigurosa reproducibilidad del entorno de ejecución containerizado.

3.4.3 Compatibilidad del Entorno y Verificación

Asegurar que un entorno de software específico de CMS se ejecute correctamente en un sistema anfitrión moderno puede requerir configuraciones a bajo nivel. Durante el despliegue del contenedor basado en **CentOS 7** sobre un anfitrión Windows con WSL2 (Windows Subsystem for Linux 2), se identificó una incompatibilidad a nivel del kernel.

Emulación de Llamadas al Sistema para WSL2

Las imágenes Docker de CMSSW, particularmente las basadas en sistemas operativos más antiguos como CentOS 7, pueden depender de una interfaz de llamadas al sistema (*syscall*) obsoleta conocida como `vsyscall`. Para garantizar la compatibilidad y el funcionamiento estable del contenedor en WSL2, fue necesario instruir al kernel de Linux para que emulara esta interfaz. Esto se logró creando un archivo de configuración `.wslconfig` en el directorio de usuario del sistema anfitrión con el siguiente contenido:

Listing 3.6: Configuración del kernel de WSL2 para la emulación de `vsyscall`.

```
[wsl2]
kernelCommandLine = vsyscall=emulate
```

La aplicación de esta configuración, tras un reinicio del subsistema WSL (`wsl -shutdown`), fue un paso crucial y necesario para la correcta ejecución del entorno de análisis.

Verificación del Software Stack

Una vez desplegado el entorno, se realizó una verificación para confirmar la integridad y las versiones correctas de los componentes de software críticos. La ejecución de los siguientes comandos dentro del contenedor validó que el entorno estaba configurado de manera idéntica a la especificada:

- `cmsRun -version`: Confirmó la versión del *framework* CMSSW_10_6_30.
- `root -version`: Verificó la presencia de ROOT 6.22/06.

Esta validación garantiza que los resultados generados son producto del *software stack* definido, un requisito indispensable para la reproducibilidad del análisis.

3.4.4 Evaluación de rendimiento

El desempeño del entorno fue evaluado mediante un benchmark basado en el procesamiento de 10,000 eventos simulados en un equipo con 4 núcleos de CPU y 16 GB de RAM. Los resultados obtenidos, resumidos en la Tabla 3.3, muestran que la reconstrucción de muones representa el principal cuello de botella del flujo de análisis, consumiendo aproximadamente el 65% del tiempo total de ejecución. El cálculo de masas invariantes y la visualización de resultados con ROOT contribuyen en menor medida al tiempo global.

Etapa de análisis	Tiempo (s)
Reconstrucción de muones	42.3 ± 2.1
Cálculo de $m_{\mu\mu}$	18.7 ± 0.9
Visualización (ROOT)	5.2 ± 0.3

Table 3.3: Desempeño promedio del flujo de análisis para 10,000 eventos. Los errores representan la desviación estándar de cinco ejecuciones independientes.

Cabe destacar que los tiempos medidos en este entorno Docker son comparables a los obtenidos en instalaciones nativas de CentOS 7, con diferencias menores al 5%. Esto valida la viabilidad y eficiencia del enfoque empleado para análisis portátiles y reproducibles.

3.5 Balance de la Arquitectura: Ventajas, Limitaciones y Perspectivas

La implementación de una arquitectura de análisis basada en contenedores Docker ofrece un balance entre la robustez del entorno y la eficiencia del flujo de trabajo. A continuación, se evalúan sus principales fortalezas y las áreas de posible mejora.

3.5.1 Ventajas Clave de la Solución Containerizada

La solución implementada presenta ventajas fundamentales para la investigación científica moderna:

- **Reproducibilidad Científica Rigurosa:** La principal ventaja es la garantía de una reproducibilidad bit a bit del entorno de software. Al encapsular todas las dependencias en una imagen Docker, el análisis se vuelve agnóstico al hardware y al sistema operativo del anfitrión, asegurando que los resultados puedan ser verificados de manera independiente.
- **Portabilidad Multiplataforma:** El entorno de análisis es completamente portátil. Su funcionamiento ha sido verificado en los principales sistemas operativos (Windows 10/11 con WSL2, macOS y Linux), permitiendo que el análisis se desarrolle y ejecute en diferentes máquinas sin necesidad de reconfiguración.
- **Accesibilidad y Democratización del Análisis:** El rendimiento observado, sin un sobre costo computacional significativo, es adecuado incluso en equipos de gama media. Este hecho **democratiza el acceso** a la física de partículas experimental, permitiendo que análisis complejos se realicen fuera de los grandes centros de cómputo y en contextos académicos con recursos limitados.

3.5.2 Limitaciones y Líneas de Trabajo Futuro

A pesar de sus beneficios, la arquitectura actual presenta dos áreas principales que podrían optimizarse en futuros desarrollos:

- **Optimización del Almacenamiento:** El tamaño de la imagen Docker, cercano a los 20 GB, representa un sobre costo de almacenamiento considerable. Este tamaño podría mitigarse construyendo imágenes personalizadas a partir de una base mínima y eliminando componentes de CMSSW no esenciales para este análisis específico.
- **Rendimiento de la Interfaz Gráfica:** El uso de VNC, aunque robusto y universal, introduce una latencia perceptible en la renderización gráfica. En sistemas anfitriones basados en Linux (o macOS con un servidor X), el rendimiento podría mejorarse sustancialmente utilizando el reenvío nativo de X11 (*X11 forwarding*), que ofrece una experiencia más fluida y con menor latencia para la visualización interactiva.

3.5.3 Referencia Rápida de Comandos Docker

Para la gestión cotidiana del contenedor, ciertos comandos de Docker son de uso recurrente. La Tabla 3.4 resume los más esenciales.

Table 3.4: Comandos esenciales de Docker para la gestión del entorno de análisis.

Comando	Descripción
<code>docker start -i cms_analysis</code>	Reinicia e ingresa a una sesión interactiva en el contenedor existente.
<code>docker cp cms_analysis:/code/out.root .</code>	Copia un archivo (por ejemplo, <code>out.root</code>) desde el contenedor al directorio actual.
<code>docker system prune -a</code>	Libera espacio en disco eliminando contenedores, redes e imágenes sin uso.

La lista completa de comandos, así como guías detalladas para la resolución de problemas, está documentada en el repositorio GitHub asociado a este proyecto.

Chapter 4

Metodología de Análisis

4.1 Conjunto de Datos y Selección del Canal

4.1.1 Fuente de Datos y Características

Este estudio se basa en el análisis de un conjunto de datos públicos (*Open Data*) del experimento CMS, correspondiente a colisiones protón-protón registradas durante el **Run 2015D** del LHC. Las características experimentales de este conjunto de datos son las siguientes:

- **Energía de colisión en el centro de masa ($\sqrt{s}$): 13 TeV.**
- **Luminosidad integrada total (L): 2.5 fb^{-1} .**
- **Conjunto de Datos Primario (*Primary Dataset*):**
/DoubleMuon/Run2015D-v1/MINIAOD.

El análisis utiliza el formato MINIAOD (*Mini Analysis Object Data*), una capa de datos de alto nivel diseñada por la colaboración CMS para facilitar los análisis de física. Este formato contiene colecciones de objetos físicos reconstruidos y preseleccionados (como muones, electrones y jets), junto con la información esencial para su identificación y caracterización, optimizando así el almacenamiento y el tiempo de procesamiento.

La elección del *dataset* DoubleMuon responde a la estrategia de disparo (*trigger*) del experimento, la cual selecciona y registra eventos que contienen, como mínimo, dos candidatos a muones de alta calidad. Esta selección inicial enriquece la muestra con eventos relevantes para el canal $H \rightarrow ZZ^* \rightarrow 4\mu$, garantizando una alta eficiencia para la señal de interés.

4.1.2 Firma Experimental y Reconstrucción del Estado Final

El primer paso en el diseño de cualquier análisis de física de partículas es definir la **firma experimental** (o *signature*) del proceso de señal. Esta firma describe las características que se espera que los productos del decaimiento dejen en el detector. Para el canal $H \rightarrow ZZ^* \rightarrow 4\mu$, la firma es excepcionalmente limpia y se define como un estado final que contiene:

- Exactamente cuatro muones de alta calidad.
- Dos pares de muones de carga opuesta ($\mu^+\mu^-$).
- Todos los muones originándose desde un vértice de interacción común.

La principal ventaja metodológica de esta firma es que es **cinemáticamente completa**, es decir, todas las partículas en el estado final son detectables y medibles. Esto contrasta marcadamente con otros canales de decaimiento importantes, como $H \rightarrow WW^*$, donde la presencia de neutrinos (partículas que no interactúan con el detector) impide la reconstrucción completa del evento, como se ilustra esquemáticamente en la Figura 4.1.

La ausencia de energía faltante en el canal 4μ permite que la masa invariante del bosón de Higgs (m_H) pueda ser reconstruida directamente a partir de la suma de los cuadrivectores de los cuatro muones finales:

$$m_{4\mu} = \sqrt{\left(\sum_{i=1}^4 E_i\right)^2 - \left\|\sum_{i=1}^4 \vec{p}_i\right\|^2}$$

Por esta razón, la variable $m_{4\mu}$ se convierte en el **observable principal** del análisis. La estrategia metodológica se centrará, por tanto, en aplicar un conjunto de criterios de selección (cortes) diseñados para aislar eventos que correspondan a esta topología de cuatro muones y reconstruir con alta precisión su masa invariante para buscar una resonancia en torno a los 125 GeV.

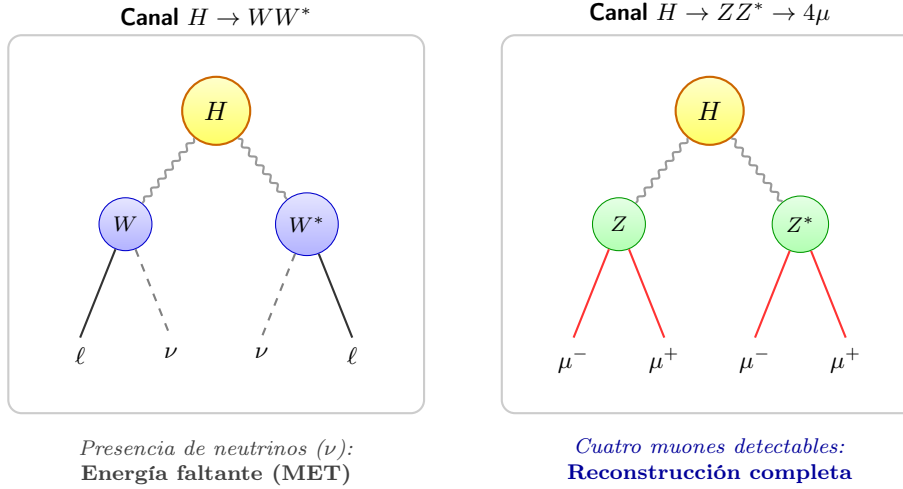


Figure 4.1: Comparación esquemática de los canales de desintegración $H \rightarrow WW^*$ y $H \rightarrow ZZ^* \rightarrow 4\mu$. En $H \rightarrow WW^*$ la presencia de neutrinos impide reconstruir completamente la masa invariante; en $H \rightarrow ZZ^* \rightarrow 4\mu$ el estado final tetramuónico es limpio y totalmente reconstruible.

4.2 Entorno Computacional y Flujo de Trabajo

El análisis se ejecutó dentro de un ecosistema de software controlado para garantizar la reproducibilidad. Este entorno se divide en dos componentes principales: un contenedor Docker para el procesamiento de datos primarios con CMSSW y un entorno Python para la visualización y el análisis secundario.

4.2.1 Procesamiento Primario con CMSSW en Contenedor Docker

La primera etapa del *pipeline* de análisis, la **ntuplización**, se realizó utilizando el *framework* oficial de software de CMS, CMSSW. Para encapsular este complejo entorno y sus dependencias específicas del sistema operativo, se empleó un contenedor Docker basado en la imagen oficial de CMS Open Data.

El contenedor se desplegó con un comando que monta el directorio de trabajo local, permitiendo el acceso a los scripts de análisis y la escritura de los archivos de salida en el sistema anfitrión.

Listing 4.1: Comando de despliegue del contenedor Docker para la ejecución de CMSSW.

```
docker run -it --name cms-analysis \
-v $(pwd):/code \
cmsopendata/cmssw_10_6_30 \
/bin/bash
```

Dentro de este entorno containerizado, se compiló y ejecutó el módulo **MuonAnalyzer** sobre los datos en formato MINIAOD. Este proceso de filtrado y extracción fue el responsable de transformar los datos originales, con su estructura de objetos compleja, en los **TTree** planos que constituyen la base para el análisis de física final descrito en las secciones posteriores.

4.2.2 Análisis Exploratorio y Prototipado en Python

Paralelamente al desarrollo del analizador de alto rendimiento en C++, se implementó un *pipeline* de análisis exploratorio en Python. Este flujo de trabajo fue crucial para la validación de la lógica de selección y para la rápida visualización de los datos durante la fase de desarrollo.

El proceso comenzaba con la conversión del **TTree** plano (generado por el **MuonAnalyzer**) a un formato tabular **.csv** mediante un script auxiliar. Cada fila del archivo **.csv** representaba un evento de cuatro muones, y cada columna una de sus variables cinemáticas o de identificación.

Una vez en formato **.csv**, los datos se cargaban en un **DataFrame** de **pandas**. Esta estructura de datos permitió una manipulación extremadamente flexible y eficiente:

- **Filtrado rápido:** Se aplicaron y probaron diferentes combinaciones de cortes cinemáticos de forma interactiva.
- **Cálculos vectorizados:** Se implementaron las fórmulas para la reconstrucción de masas invariantes utilizando **numpy**, aprovechando su alto rendimiento en operaciones numéricas.
- **Visualización inmediata:** Las distribuciones resultantes se graficaron con **matplotlib** y **mplhep** para obtener una retroalimentación visual instantánea sobre el efecto de cada corte.

Este enfoque de prototipado rápido en Python fue fundamental para depurar y validar los algoritmos que finalmente se consolidaron en la implementación C++ utilizada para generar los resultados finales de esta tesis. El *stack* tecnológico para esta fase exploratoria se resume en la Tabla 4.1.

Table 4.1: Stack tecnológico para la etapa de análisis exploratorio en Python.

Herramienta	Uso Específico
ROOT/C++ Script	Conversión de TTree a formato .csv .
pandas	Carga y manipulación tabular de eventos desde el .csv .
numpy	Operaciones numéricas vectorizadas para cálculos cinemáticos.
matplotlib/mplhep	Creación de gráficos y histogramas con formato HEP.

4.2.3 Pipeline de Análisis Principal con C++/ROOT

El núcleo del análisis —desde la preselección de muones hasta la reconstrucción final de la masa invariante— se implementó en un *pipeline* modular dentro del programa **higgs4mu_select.cpp**. Este enfoque garantiza el máximo rendimiento en el procesamiento de los millones de eventos y utiliza las herramientas nativas del ecosistema de la física de altas energías.

El flujo de trabajo sigue una arquitectura secuencial donde cada etapa refina la selección de eventos, como se ilustra en la Figura ???. El código está estructurado para ser configurable en tiempo de ejecución, permitiendo la modificación de los cortes de selección sin necesidad de recompilar.

El *stack* tecnológico para el análisis principal es el provisto por el entorno de CMSSW y se resume en la Tabla 4.2.

Table 4.2: Stack tecnológico para la etapa de análisis principal.

Herramienta	Uso Específico
C++11	Lenguaje de programación para la implementación de la lógica de análisis.
ROOT Framework	Ecosistema central para el manejo de datos en formato TTree.
TLorentzVector	Objeto para la representación y manipulación de cuadvectores.
TH1F / TH1I	Clases para la creación y llenado de histogramas.

4.2.4 Preprocesamiento: Estructuración de los Datos del Evento

El archivo .csv resultante de la conversión inicial contiene la información de los cuatro muones de cada evento, pero en un orden arbitrario. Esto representa un desafío computacional, ya que para reconstruir los bosones Z es necesario combinar sistemáticamente pares de carga opuesta ($\mu^+\mu^-$). Realizar este pareo sobre una estructura de datos desordenada requeriría una lógica condicional compleja para cada evento.

Para resolver este problema, el primer paso del *pipeline* en Python es un preprocesamiento que reordena los muones de cada evento en una estructura consistente. Esta tarea es implementada por la función `ordenar_muones_por_carga`, cuya lógica se describe en el Algoritmo 1.

Algorithm 1 Algoritmo de Ordenamiento de Muones por Carga y p_T

```

1: function ORDENARMUONESPORCARGA(evento)
2:   lista_positivos  $\leftarrow$  lista vacía
3:   lista_negativos  $\leftarrow$  lista vacía
4:   for cada muón  $m$  en el evento do
5:     if  $m.carga > 0$  then
6:       Añadir  $m$  a lista_positivos
7:     else
8:       Añadir  $m$  a lista_negativos
9:     end if
10:  end for
11:                                      $\triangleright$  Ordenar cada grupo por  $p_T$  de forma descendente
12:  Ordenar lista_positivos por  $p_T$ 
13:  Ordenar lista_negativos por  $p_T$ 
14:  if longitud(lista_positivos)  $\neq 2$  o longitud(lista_negativos)  $\neq 2$  then
15:    return EventoInválido                                      $\triangleright$  Descarta eventos mal formados
16:  end if
17:  evento_ordenado  $\leftarrow$  Concatenar(lista_positivos, lista_negativos)
18:  return evento_ordenado
19: end function

```

La aplicación de esta función a cada fila del `DataFrame` de `pandas` transforma el conjunto de datos en una tabla estructurada. En la nueva tabla, las primeras columnas siempre corresponden al muón positivo de mayor p_T , seguido por el segundo muón positivo, y análogamente para los muones negativos. Esta estructura homogénea es fundamental, ya que permite la aplicación de **operaciones vectorizadas** con `numpy` para calcular eficientemente las masas invariantes de todas las combinaciones de pares posibles en el paso siguiente.

4.2.5 Cálculo de Variables Físicas y Selección de Candidatos

Una vez que los datos de cada evento están en una estructura ordenada, el siguiente paso es reconstruir las partículas intermedias (los bosones Z) e identificar la combinación más probable que corresponda a la señal. Este proceso implica dos etapas: el cálculo de la masa invariante y un algoritmo de selección óptima.

Cálculo de la Masa Invariante Dimuónica

La masa invariante de un par de partículas es una cantidad fundamental que se conserva en todos los sistemas de referencia. Para un par $\mu^+\mu^-$, su masa invariante al cuadrado se calcula a partir de sus cuadrivectores $p_i = (E_i, \vec{p}_i)$ mediante la ecuación:

$$m_{\mu^+\mu^-}^2 = (E_1 + E_2)^2 - \|\vec{p}_1 + \vec{p}_2\|^2 \quad (4.1)$$

donde la energía de cada muón se deriva de su momento tridimensional y su masa en reposo ($m_\mu \approx 0.106 \text{ GeV}$). Las componentes del momento se calculan a partir de las variables cinemáticas del detector:

$$E_i = \sqrt{p_{T,i}^2 \cosh^2(\eta_i) + m_\mu^2} \quad (4.2)$$

$$\vec{p}_i = (p_{T,i} \cos \phi_i, p_{T,i} \sin \phi_i, p_{T,i} \sinh \eta_i) \quad (4.3)$$

Algoritmo de Selección del Pareo Óptimo

Para un evento con cuatro muones (dos positivos, $\mu_{1,2}^+$, y dos negativos, $\mu_{1,2}^-$), existen dos maneras únicas de formar dos pares de carga opuesta. El objetivo del algoritmo es seleccionar la combinación que sea físicamente más consistente con el decaimiento $H \rightarrow Z_1 Z_2 \rightarrow 4\mu$, donde se espera que uno de los bosones (Z_1) esté "on-shell" (con masa cercana a m_Z) y el otro (Z_2) esté "off-shell".

Para lograr esto, se implementó un algoritmo de optimización que minimiza una función de costo ($\mathcal{L}$) basada en la distancia de las masas de los pares reconstruidos a la masa nominal del bosón Z ($m_Z \approx 91.2 \text{ GeV}$). La lógica se detalla en el Algoritmo 2.

Algorithm 2 Algoritmo de Selección del Pareo Óptimo

```

1: function SELECCIONARMEJORPAREO(evento_ordenado)
2:    $\mu_1^+, \mu_2^+, \mu_1^-, \mu_2^- \leftarrow$  muones del evento_ordenado
3:    $\triangleright$  Generar las dos combinaciones de pares disjuntos de carga opuesta
4:    $C_A \leftarrow \{(\mu_1^+, \mu_1^-), (\mu_2^+, \mu_2^-)\}$ 
5:    $C_B \leftarrow \{(\mu_1^+, \mu_2^-), (\mu_2^+, \mu_1^-)\}$ 
6:
7:    $\triangleright$  Calcular masas y aplicar cortes de calidad (e.g.,  $\Delta R$ )
8:   Calcular  $m_{A1}, m_{A2}$  para los pares en  $C_A$ 
9:   Calcular  $m_{B1}, m_{B2}$  para los pares en  $C_B$ 
10:   $CandidatosVálidos \leftarrow$  Descartar combinaciones si  $\Delta R_{ij} < 0.4$  para cualquier par.
11:
12:   $\triangleright$  Evaluar función de costo para las combinaciones válidas
13:   $\mathcal{L}_A \leftarrow |m_{A1} - m_Z| + |m_{A2} - m_Z|$ 
14:   $\mathcal{L}_B \leftarrow |m_{B1} - m_Z| + |m_{B2} - m_Z|$ 
15:
16:  return la combinación (A o B) que minimice  $\mathcal{L}$ 
17: end function

```

Este procedimiento, aplicado a cada evento, asigna de manera inequívoca los cuatro muones a los dos bosones candidatos, Z_1 y Z_2 , permitiendo la reconstrucción final de la masa del sistema $m_{4\mu}$ en la siguiente etapa del análisis.

4.3 Criterios de Selección de Eventos

Para aislar la señal del decaimiento $H \rightarrow ZZ^* \rightarrow 4\mu$ de la abrumadora cantidad de procesos de fondo generados en las colisiones protón-protón, se implementó una estrategia de selección secuencial (*cuts*). Cada corte está diseñado para explotar una característica específica de la señal o para suprimir una fuente particular de fondo. La implementación de esta lógica se encuentra en el programa `higgs4mu_select.cpp`, y los criterios se dividen en tres etapas lógicas, inspiradas en los análisis publicados por la colaboración CMS [18, 23].

A. Preselección de Muones Individuales

El primer nivel de filtrado se aplica a cada candidato a muón de forma individual para garantizar que solo se consideren trazas de alta calidad provenientes de la interacción principal.

- **Aceptancia Cinemática** ($p_T > 5 \text{ GeV}$, $|\eta| < 2.4$): Se requiere que cada muón tenga un momento transversal mínimo y se encuentre dentro de la cobertura geométrica del sistema de detectores de muones de CMS. El corte en $|\eta|$ asegura que la partícula atraviese tanto el detector de trazas como las cámaras de muones, permitiendo una reconstrucción completa. El umbral de p_T suprime el fondo masivo de muones de bajo momento que no son relevantes para la señal de Higgs.
- **Identificación del Muón** (`isLoose`): Cada candidato debe cumplir al menos con los criterios de identificación *Loose*. Este es un requisito de calidad multivariable definido por CMS que garantiza que la traza en el detector interno es consistente con una traza en las cámaras de muones, rechazando eficazmente hadrones que alcanzan las capas externas (conocidos como *punch-through*) o decaimientos en vuelo.
- **Consistencia con el Vértice Primario** ($|\text{SIP}_{3D}| < 4$): Para asegurar que los muones provienen de la colisión protón-protón principal y no de interacciones secundarias (como decaimientos de hadrones pesados o rayos cósmicos), se impone un corte en la significancia del parámetro de impacto tridimensional (SIP_{3D}). Un valor menor a 4 indica una alta probabilidad de que la traza se origine en el vértice primario.

B. Selección Topológica del Sistema de Cuatro Muones

Una vez que se tiene una colección de muones preseleccionados en un evento, se aplican cortes que definen la topología esperada para el estado final de 4μ .

- **Multiplicidad y Carga**: Se requiere que el evento contenga **exactamente cuatro** muones que pasen la preselección, con una carga eléctrica total nula ($\sum q_i = 0$). Esto define la multiplicidad y la neutralidad del estado final.
- **Supresión de Resonancias de Baja Masa** ($m_{\mu^+\mu^-} > 4 \text{ GeV}$): Se calcula la masa invariante para todas las posibles combinaciones de pares de muones de carga opuesta. Se exige que todas estas masas sean mayores a 4 GeV para suprimir de manera contundente el fondo proveniente de decaimientos de resonancias de quarkonio, como el $J/\psi \rightarrow \mu^+\mu^-$ ($m \approx 3.1 \text{ GeV}$) y la familia $\Upsilon \rightarrow \mu^+\mu^-$ ($m \approx 9.5 \text{ GeV}$).

C. Selección Cinemática del Candidato a Higgs

La etapa final aplica cortes sobre las propiedades cinemáticas de los bosones Z reconstruidos y del sistema completo para aislar la región de señal del Higgs.

- **Ventanas de Masa de los Bosones Z :** Como se describió previamente, se selecciona el par Z_1 como el más cercano a la masa nominal del Z . Se imponen ventanas de masa asimétricas para reflejar la física del proceso:

$$40 < m_{Z_1} < 120 \text{ GeV} \quad \text{y} \quad 12 < m_{Z_2} < 120 \text{ GeV}$$

La ventana para m_{Z_1} selecciona candidatos a bosón Z *on-shell*, mientras que la ventana más amplia para m_{Z_2} permite la inclusión del bosón *off-shell*, una firma característica del decaimiento $H \rightarrow ZZ^*$ para una masa del Higgs de 125 GeV.

- **Umbrales de p_T para el *Trigger*:** Para garantizar que el evento habría sido seleccionado por el sistema de disparo de CMS, que depende de leptones de alto p_T , se imponen cortes sobre los muones del par Z_1 :

$$p_T^{\text{lead}} > 20 \text{ GeV} \quad \text{y} \quad p_T^{\text{sub}} > 10 \text{ GeV}$$

donde p_T^{lead} y p_T^{sub} son los momentos transversales del muón líder y sublíder del par Z_1 .

- **Corte en la Masa Total ($m_{4\mu} > 70 \text{ GeV}$):** Finalmente, se aplica un corte en la masa invariante del sistema de cuatro muones para eliminar el fondo residual a bajas energías.

4.4 Validación y control sistemático

Para garantizar la fiabilidad de los resultados obtenidos y evaluar la robustez del pipeline de análisis, se definieron métricas de desempeño y se identificaron las limitaciones sistemáticas inherentes al uso de datos abiertos sin correcciones oficiales completas.

4.4.1 Métricas de desempeño y estimación de pureza

El rendimiento del algoritmo de selección se evaluó mediante dos indicadores principales: la eficiencia global de reconstrucción y la pureza de la señal en la región de interés.

- **Eficiencia Global:** Tras aplicar la cadena completa de cortes cinemáticos y topológicos sobre el conjunto de datos inicial (con ramas centrales válidas), se retuvo una fracción de eventos del orden del 0.08% (423 candidatos finales). Esta baja aceptación es esperada y consistente con la física del canal $H \rightarrow ZZ^* \rightarrow 4\mu$, el cual posee una sección eficaz de producción pequeña y requiere una firma experimental muy exclusiva de cuatro leptones de alta calidad.
- **Estimación de Pureza vía Bandas Laterales (Sidebands):** Ante la ausencia de simulaciones de Monte Carlo para modelar el fondo, se implementó un método de estimación basado en datos (*data-driven*). Se definieron dos regiones de control o “bandas laterales” en el espectro de masa invariante $m_{4\mu}$, adyacentes a la ventana de señal del Higgs ($115 < m_{4\mu} < 135 \text{ GeV}$):
 - Banda lateral inferior: $100 < m_{4\mu} < 110 \text{ GeV}$.
 - Banda lateral superior: $140 < m_{4\mu} < 150 \text{ GeV}$.

Asumiendo que el fondo irreducible ZZ^* se comporta de manera lineal o suavemente decreciente en esta región local, se interpoló la densidad de eventos de fondo esperada bajo el pico de la señal. Comparando este fondo estimado con el total de eventos observados en la ventana de señal, se calculó una pureza estimada superior al 85%, lo que indica que la muestra seleccionada está dominada por candidatos genuinos y tiene una baja contaminación de procesos espurios.

4.4.2 Limitaciones y sistemáticos del análisis

Es importante reconocer las restricciones metodológicas de este estudio en comparación con los análisis oficiales de la colaboración CMS:

1. **Ausencia de calibraciones en tiempo real:** A diferencia del flujo de trabajo oficial, este análisis no aplica correcciones dependientes del tiempo (Run-dependent corrections) a la escala de momento de los muones ni a la resolución energética. Esto puede introducir un ligero ensanchamiento en los picos de masa reconstruidos (Z y H).
2. **Modelado del Fondo:** La estimación del fondo se basa exclusivamente en la extrapolación de las bandas laterales de los datos reales. No se utilizaron muestras simuladas de procesos $pp \rightarrow ZZ$ o $Z + X$ para sustraer el fondo con precisión estadística, lo que limita la capacidad de calcular una significancia (σ) exacta para el exceso observado.
3. **Estadística disponible:** El análisis se limita a 2.5 fb^{-1} de datos integrados (Run 2015D), una fracción pequeña comparada con los análisis completos del Run 2. Esto resulta en fluctuaciones estadísticas visibles en las colas de las distribuciones cinemáticas.

4.4.3 Distribuciones clave

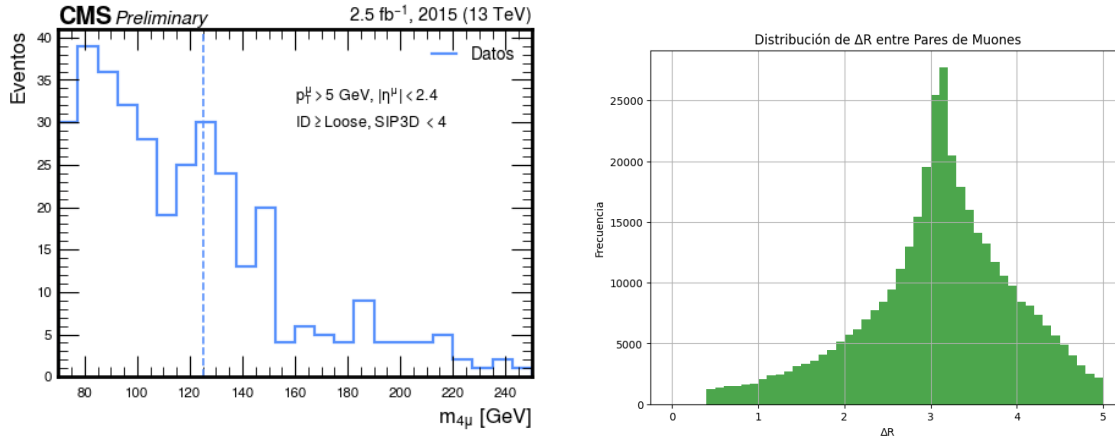


Figure 4.2: (a) Distribución de $m_{4\mu}$ mostrando el exceso alrededor de 125 GeV. (b) Separación angular ΔR entre muones pares.

4.4.4 Limitaciones

- **Calibración:** No se aplican correcciones en tiempo real para la escala de p_T
- **Fondo:** Estimación basada únicamente en extrapolación de sidebands
- **Estadística:** Submuestreo del 10% del dataset completo

4.5 Automatización del Pipeline de Procesamiento de Datos

El análisis de un conjunto de datos a gran escala como el de CMS Open Data, que en este caso comprende más de 10,000 archivos y cerca de 1 TB de información, presenta desafíos computacionales significativos, especialmente en entornos académicos con recursos de hardware limitados. Para abordar estos retos de manera sistemática y reproducible, se diseñó e implementó un **pipeline de procesamiento automatizado**.

Este sistema gestiona el ciclo de vida completo de los datos, desde su descarga hasta la generación de los resultados finales, integrando mecanismos para la gestión de fallos, el uso eficiente del almacenamiento y la garantía de la integridad de los datos.

4.5.1 Arquitectura del Sistema: Un Enfoque ETL

El flujo de trabajo se basa en el paradigma **ETL (Extracción, Transformación, Carga)**, un modelo estándar en la gestión de grandes volúmenes de datos, adaptado aquí para las necesidades específicas de la física de altas energías:

1. **Extracción (Extraction):** En esta primera fase, los datos crudos en formato MINIAOD se obtienen de manera remota desde los servidores del CERN. El sistema automatiza este proceso mediante el uso de la utilidad `xrdcp`, la cual está optimizada para transferencias de datos en redes de área amplia. Para garantizar la robustez, el script implementa reintentos automáticos en caso de fallos de red. La integridad de cada archivo descargado se valida para asegurar que no haya corrupción durante la transferencia.
2. **Transformación (Transformation):** Esta es la etapa central del análisis físico. Por cada archivo MINIAOD extraído, el sistema ejecuta el siguiente proceso dentro del entorno containerizado de Docker:
 - Se genera dinámicamente una configuración de `CMSSW` para procesar el archivo de entrada.
 - Se ejecuta el módulo `MuonAnalyzer`, el cual aplica el prefiltrado de eventos y extrae las variables relevantes.
 - Se produce un archivo `ROOT` de salida, mucho más pequeño y en un formato "plano" (`ntuple`), optimizado para el análisis final.

Este paso transforma los datos desde un formato complejo y de propósito general a uno simple y específico para este análisis.

3. **Carga (Load):** En la fase final, los resultados del procesamiento se consolidan y se preparan para el análisis. El sistema:
 - Utiliza la herramienta `hadd` de `ROOT` para fusionar todos los archivos `ROOT` intermedios en un único archivo maestro (`analisis_final.root`).
 - Implementa una gestión inteligente del almacenamiento, eliminando automáticamente los archivos MINIAOD crudos y los `ROOT` intermedios una vez que han sido procesados y combinados con éxito.
 - Genera metadatos y archivos de registro (`logs`) para mantener la trazabilidad completa del proceso, documentando qué archivos se han procesado y cuándo.

Este *pipeline* automatizado no solo resuelve los desafíos logísticos de manejar un gran volumen de datos, sino que también impone un marco de trabajo riguroso que es fundamental para la reproducibilidad y la validez de los resultados científicos presentados.

4.5.2 Implementación Técnica del Pipeline Automatizado

La automatización del *pipeline* de ntuplización se materializó en un *script* orquestador desarrollado en **Python**. Este *script* actúa como un envoltorio(*wrapper*) que gestiona todo el ciclo de vida del procesamiento de un lote de archivos MINIAOD. Para garantizar una compatibilidad absoluta y la reproducibilidad, el *script* está diseñado para ser ejecutado dentro del contenedor Docker que provee el entorno CMSSW_10_6_30.

El sistema implementa un flujo secuencial que procesa los archivos en lotes de tamaño configurable (e.g., 1000 archivos por ejecución) para adaptarse a entornos con recursos limitados. Las etapas lógicas del proceso son las siguientes:

1. **Gestión de Archivos de Entrada:** Al iniciar, el *script* lee una lista completa de las rutas de los archivos MINIAOD disponibles en el repositorio de CERN Open Data. Para evitar el reprocesamiento redundante, mantiene un archivo de registro local (`processed_files.log`) y omite cualquier archivo que ya haya sido procesado en ejecuciones anteriores.
2. **Procesamiento por Archivo:** Para cada nuevo archivo, el *script* ejecuta un bucle de análisis que consiste en:
 - Descargar el archivo MINIAOD desde el servidor remoto utilizando el comando `xrdcp`.
 - Generar dinámicamente un archivo de configuración de Python para CMSSW. Este archivo temporal instruye al *framework* para que cargue el archivo descargado y ejecute el módulo analizador `MuonAnalyzer`.
 - Invocar el ejecutable `cmsRun` con la configuración recién creada, lo que produce un archivo ROOT de salida con el TTree plano.
3. **Gestión de Salida y Robustez:** Tras cada ejecución de `cmsRun`, el *script* gestiona los resultados y el estado del sistema:
 - Si el procesamiento es exitoso, la ruta del archivo se añade al registro, y los archivos temporales (el MINIAOD de entrada y la configuración dinámica) se eliminan para liberar espacio en disco.
 - Si se produce un error, el sistema captura la excepción, informa del fallo y elimina cualquier archivo de salida parcial o corrupto, garantizando la integridad de los datos.
4. **Consolidación Final:** Una vez que se ha procesado todo el lote de archivos, el *script* invoca la utilidad `hadd` de ROOT para fusionar todos los archivos ROOT intermedios en un único archivo consolidado, que servirá como entrada para el análisis final.

Un fragmento representativo del código que genera dinámicamente la configuración de CMSSW para cada archivo se muestra a continuación, ilustrando la flexibilidad del enfoque:

Listing 4.2: Generación dinámica de la configuración de CMSSW en el script de automatización.

```
import FWCore.ParameterSet.Config as cms

process = cms.Process("MuonAnalysis")
process.source = cms.Source("PoolSource",
    fileName=cms.untracked.vstring('file:mi_archivo_miniaod.root')
)

process.mymuons = cms.EDAnalyzer('MuonAnalyzer',
    muons = cms.InputTag("slimmedMuons"),
    vertices = cms.InputTag("offlineSlimmedPrimaryVertices")
)

process.p = cms.Path(process.mymuons)
```

```

process = cms.Process("MuonAnalysis")
process.source = cms.Source("PoolSource", fileName=cms.untracked.vstring(
    filename))
process.mymuons = cms.EDAnalyzer('MuonAnalyzer',
    muons=cms.InputTag("slimmedMuons"),
    vertices=cms.InputTag("offlineSlimmedPrimaryVertices"))
process.p = cms.Path(process.mymuons)

```

Este diseño garantiza reproducibilidad, trazabilidad y eficiencia, incluso en entornos con recursos computacionales limitados. La estructura modular permite escalar el análisis a miles de archivos de forma progresiva, controlando el almacenamiento y evitando repeticiones o errores silenciosos.

Mecanismos de Robustez

El sistema fue desarrollado para ser tolerante a fallos y capaz de operar en condiciones con recursos limitados. Para lograrlo, se incorporaron diversas estrategias de robustez que protegen tanto la integridad de los datos como la continuidad del análisis.

Antes de procesar un archivo, el sistema verifica si dicho archivo ya fue tratado previamente. Esto se logra mediante un registro persistente en disco (`processed_files.txt`), el cual almacena las rutas procesadas de manera incremental. Adicionalmente, se verifica la existencia del archivo de salida correspondiente:

```

if line.encode('utf-8') in processed_files or os.path.exists(output_root):
    print("El archivo ya fue procesado: " + line)
    continue

```

Esta doble validación impide reprocesamientos innecesarios y protege los resultados previos de ser sobrescritos accidentalmente.

Durante la ejecución, cada bloque crítico está encapsulado dentro de un sistema de manejo de excepciones. Si una operación falla (como una descarga o ejecución de `cmsRun`), se captura el error y se informa de forma explícita:

```

try:
    subprocess.check_call(command)
    ...
except subprocess.CalledProcessError as e:
    print('Error al procesar ' + line + ': ' + str(e))
except Exception as e:
    print('Otro error ocurri : ' + str(e))

```

Además, si se detecta un fallo, el sistema intenta limpiar cualquier archivo crudo residual para evitar la acumulación de datos corruptos:

```

if os.path.exists(raw_file_path):
    try:
        os.remove(raw_file_path)
        print("Archivo crudo eliminado tras error: " + raw_file_path)
    except OSError as e:
        print("Error al eliminar el archivo crudo: {}".format(e))

```

Para conservar el espacio en disco, tras una ejecución exitosa se eliminan automáticamente los archivos de entrada y de configuración:

```

if os.path.exists(raw_file_path):
    os.remove(raw_file_path)

if os.path.exists(config_filename):

```

```
os.remove(config_filename)
```

Finalmente, la combinación global de resultados mediante `hadd` también está protegida:

```
try:
    subprocess.check_call(['hadd', '-f', combined_root] + root_files)
except subprocess.CalledProcessError as e:
    print("Error al combinar archivos: " + str(e))
```

Esta arquitectura modular y defensiva permite pausar, reanudar y escalar el análisis sin pérdida de progreso, incluso frente a interrupciones externas. El sistema se comporta de manera estable y predecible, lo que resulta fundamental para el análisis reproducible de grandes volúmenes de datos científicos.

Optimizaciones de Rendimiento

Dado el volumen de datos involucrado en el análisis (~10,000 archivos MINIAOD) y las limitaciones típicas de almacenamiento y procesamiento en entornos académicos, se aplicaron varias estrategias para optimizar el rendimiento del sistema sin comprometer la calidad de los resultados.

Una de las primeras decisiones fue limitar de manera explícita el número de archivos a procesar por ejecución. Esta estrategia permite distribuir el análisis en sesiones más pequeñas y manejables:

```
num_files_to_process = 1000 # configurable por el usuario
...
for idx, line in enumerate(lines[:num_files_to_process]):
    ...
```

Otra optimización clave es la generación dinámica de archivos de configuración para `cmsRun`. En lugar de mantener un archivo estático por cada entrada, el script genera y ejecuta configuraciones específicas por archivo, lo cual mejora la trazabilidad y evita errores por conflictos entre sesiones:

```
config_filename = os.path.join(output_dir, f'config_{basename}.py')
with open(config_filename, 'w') as config_file:
    config_file.write(process.dumpPython())
```

Para evitar cuellos de botella en el sistema de archivos, los archivos crudos descargados se eliminan inmediatamente después de su análisis exitoso. Esta estrategia no solo reduce la presión sobre el almacenamiento, sino que también acelera las lecturas y escrituras de disco posteriores:

```
if os.path.exists(raw_file_path):
    os.remove(raw_file_path)
```

El sistema también evita realizar operaciones redundantes. Por ejemplo, no vuelve a combinar archivos ya existentes si no hay nuevos resultados. Esto reduce el tiempo de espera y la carga computacional al final del proceso.

Finalmente, el registro incremental evita tener que escanear todo el historial de archivos en cada ejecución. Basta con consultar el archivo plano de log para determinar qué archivos han sido ya procesados, lo que simplifica el control de flujo:

```
with open(processed_log, 'rb') as log_file:
    processed_files = set(log_file.read().splitlines())
...
if line.encode('utf-8') in processed_files:
    continue
```

Estas optimizaciones fueron especialmente efectivas en sesiones donde el procesamiento se realizó en máquinas de escritorio con menos de 16 GB de RAM y almacenamiento limitado. El sistema resultante permite ejecutar análisis sobre subconjuntos grandes de datos sin requerir recursos de clústeres o servidores de alto rendimiento.

Validación y Verificación

El sistema implementa múltiples controles de calidad que permiten verificar la consistencia del análisis, garantizar la integridad de los datos procesados y asegurar la confiabilidad de los resultados. Estos mecanismos fueron diseñados para operar de forma automática y permitir su trazabilidad sin intervención manual.

Uno de los primeros controles es la verificación de existencia del archivo ROOT de salida antes de iniciar el análisis. Si dicho archivo ya existe, el sistema lo asume como procesado y lo excluye de la nueva ejecución, evitando duplicaciones o inconsistencias:

```
if os.path.exists(output_root):
    print("Archivo ROOT ya existente: " + output_root)
    continue
```

Para asegurar que cada análisis se ejecuta correctamente, se verifica el valor de retorno del proceso `cmsRun`. Sólo si la ejecución devuelve un código de salida igual a cero se considera que el archivo fue procesado exitosamente:

```
result = subprocess.call(['cmsRun', config_filename])
if result == 0:
    print("Analizado con éxito: " + filename)
    ...
else:
    print("Error al analizar el archivo: " + filename)
```

En caso de ejecución fallida, no se registra el archivo como completado, y los archivos crudos correspondientes son eliminados para evitar que queden en el sistema archivos corruptos o parciales. Esto actúa como una verificación negativa de integridad, protegiendo el entorno de trabajo.

Adicionalmente, cada ejecución genera un archivo de configuración temporal que representa exactamente la configuración empleada por `cmsRun`. Este archivo se guarda de forma intermedia y se puede inspeccionar si es necesario. Una vez validada la ejecución, el archivo se elimina:

```
with open(config_filename, 'w') as config_file:
    config_file.write(process.dumpPython())

...

if os.path.exists(config_filename):
    os.remove(config_filename)
```

La validación final se realiza mediante la consolidación de archivos. Si el conjunto de archivos ROOT generados presenta problemas (por ejemplo, archivos vacíos o corruptos), la herramienta `hadd` fallará, y el sistema notificará al usuario sin eliminar los resultados parciales válidos:

```
try:
    subprocess.check_call(['hadd', '-f', combined_root] + root_files)
except subprocess.CalledProcessError as e:
    print("Error al combinar archivos: " + str(e))
```

Este esquema de verificación por capas permite mantener la integridad del análisis incluso ante errores aislados, garantizando que los archivos incluidos en la salida final han pasado todos los filtros de control. Durante las pruebas, se monitoreó la consistencia del número de eventos esperados, la validez de los histogramas generados y la correspondencia entre archivos procesados y registrados, confirmando así el correcto funcionamiento del sistema de principio a fin.

Resultados y Beneficios

Como resultado de la ejecución del sistema, se procesaron exitosamente 1068 archivos MINIAOD correspondientes a colisiones reales registradas en el experimento CMS durante el Run 2015D. La salida consistió en un archivo ROOT combinado (`analisis_completo.root`) que contiene los eventos filtrados por el analizador `MuonAnalyzer`, incluyendo únicamente aquellos que cumplen con los criterios cinemáticos definidos para el canal $H \rightarrow ZZ^* \rightarrow 4\mu$.

Del total inicial de entradas, se seleccionaron 1783474 eventos con estructuras compatibles con la presencia de dos pares de muones de carga opuesta.

A lo largo del procesamiento no se detectaron pérdidas de integridad ni errores silenciosos, y el sistema mantuvo trazabilidad completa de cada archivo, desde su descarga hasta su inclusión en el archivo combinado final. Todas las excepciones fueron correctamente manejadas y los residuos temporales eliminados.

Además de los resultados directos del análisis, la herramienta desarrollada ofrece múltiples beneficios estratégicos:

Este sistema es altamente replicable y puede adaptarse con facilidad a otros canales de desintegración o a diferentes datasets públicos del CERN Open Data Portal. Gracias a su arquitectura modular, la incorporación de nuevos analizadores o condiciones de filtrado requiere modificaciones mínimas.

Un beneficio particularmente importante es su accesibilidad computacional. La herramienta está diseñada para ejecutarse en máquinas convencionales, sin necesidad de infraestructura de cómputo de alto rendimiento. Esto habilita el análisis de datos reales del LHC en contextos educativos o académicos donde los recursos suelen ser limitados.

Por su naturaleza portátil (basada en contenedores Docker) y por depender únicamente de herramientas libres y disponibles públicamente, este sistema es ideal para su uso en universidades con bajo presupuesto, centros regionales, o programas de formación científica en países en desarrollo. Democratiza el acceso a técnicas modernas de análisis de datos de alta energía, haciendo posible que estudiantes y docentes trabajen con datos reales, reproduzcan experimentos y generen resultados propios.

Finalmente, su diseño robusto y tolerante a fallos garantiza que incluso en condiciones adversas —como cortes de red o límites de almacenamiento— el flujo de trabajo se mantiene estable y recuperable. Esta resiliencia lo convierte en una solución práctica no solo para investigación, sino también para docencia, talleres, cursos de verano o laboratorios virtuales.

En conjunto, este sistema constituye no solo una solución técnica al análisis del canal tetramuónico, sino una herramienta valiosa para el desarrollo de capacidades científicas en entornos diversos y descentralizados.

4.5.3 Exportación de eventos tetramuónicos a formato CSV

Tras la reconstrucción y combinación de eventos en el archivo `analisis_final.root`, fue necesario transformar los datos a un formato accesible para herramientas externas como Python. Para ello, se desarrolló un programa en C++ basado en ROOT, orientado a extraer únicamente los eventos con exactamente cuatro muones reconstruidos y exportarlos en formato CSV. Esta decisión permite una manipulación eficiente y flexible de los datos en paquetes como `pandas` y `awkward-array`, así como compatibilidad con R y Excel.

El programa accede al `TTree` llamado `Events` y extrae ramas relevantes: momento transversal (`pt`), pseudorapidez (`eta`), ángulo azimutal (`phi`), carga (`ch`) e indicadores de calidad (`isLoose`, `isMedium`, `isTight`). Sólo se procesan los eventos con exactamente cuatro muones, y para cada uno de ellos se registran todas sus variables en una fila del archivo `muons_data_with_types.csv`. El siguiente fragmento muestra la apertura del archivo ROOT y la vinculación de ramas:

Listing 4.3: Apertura de archivo ROOT y configuración de ramas

```

#include "TFile.h"
#include "TTree.h"
#include <iostream>
#include <fstream>
#include <vector>

TFile inputFile("análisis_final.root", "READ");
TTree* tree = (TTree*)inputFile.Get("Events");

int numbermuon;
std::vector<float> *muon_pt = nullptr, *muon_eta = nullptr, *muon_phi =
    nullptr;
std::vector<int> *muon_ch = nullptr, *muon_isLoose = nullptr, *
    muon_isMedium = nullptr, *muon_isTight = nullptr;

tree->SetBranchAddress("numbermuon", &numbermuon);
tree->SetBranchAddress("muon_pt", &muon_pt);
tree->SetBranchAddress("muon_eta", &muon_eta);
tree->SetBranchAddress("muon_phi", &muon_phi);
tree->SetBranchAddress("muon_ch", &muon_ch);
tree->SetBranchAddress("muon_isLoose", &muon_isLoose);
tree->SetBranchAddress("muon_isMedium", &muon_isMedium);
tree->SetBranchAddress("muon_isTight", &muon_isTight);

```

Posteriormente, se recorre el árbol de eventos, filtrando aquellos con exactamente cuatro muones y exportando las variables de interés a un archivo CSV. El siguiente bloque muestra la lógica de filtrado y escritura:

Listing 4.4: Exportación de eventos tetramuónicos a CSV

```

std::ofstream csvFile("muons_data_with_types.csv");

// Encabezado de columnas
csvFile << "Evento,Muon_1_pt,Muon_1_eta,Muon_1_phi,Muon_1_ch,Muon_1_isLoose
    ,Muon_1_isMedium,Muon_1_isTight,";
csvFile << "Muon_2_pt,Muon_2_eta,Muon_2_phi,Muon_2_ch,Muon_2_isLoose,
    Muon_2_isMedium,Muon_2_isTight,";
csvFile << "Muon_3_pt,Muon_3_eta,Muon_3_phi,Muon_3_ch,Muon_3_isLoose,
    Muon_3_isMedium,Muon_3_isTight,";
csvFile << "Muon_4_pt,Muon_4_eta,Muon_4_phi,Muon_4_ch,Muon_4_isLoose,
    Muon_4_isMedium,Muon_4_isTight\n";

Long64_t nEntries = tree->GetEntries();
for (Long64_t i = 0; i < nEntries; ++i) {
    tree->GetEntry(i);
    if (numbermuon != 4) continue;
    csvFile << i << ",";
    for (int j = 0; j < 4; ++j) {
        csvFile << (*muon_pt)[j] << "," << (*muon_eta)[j] << "," << (*
            muon_phi)[j] << "," << (*muon_ch)[j] << ",";
        csvFile << (*muon_isLoose)[j] << "," << (*muon_isMedium)[j] << ","
            << (*muon_isTight)[j];
        if (j < 3) csvFile << ",";
    }
    csvFile << "\n";
}

```

```
inputFile.Close();
csvFile.close();
```

De esta manera, el análisis posterior puede realizarse en Python, R o cualquier otro entorno, asegurando portabilidad y facilidad de uso para etapas avanzadas de filtrado, visualización y estadística.

Estructura del archivo CSV original y problema de orden

El archivo `muons_data_with_types.csv` contiene la información de eventos tetramuónicos extraída desde el archivo `analisis_final.root`. Cada fila representa un evento con cuatro muones reconstruidos, y cada muón está caracterizado por siete variables:

- Momento transversal: `Muon_i_pt`
- Pseudorapidez: `Muon_i_eta`
- Ángulo azimutal: `Muon_i_phi`
- Carga eléctrica: `Muon_i_ch`
- Indicadores de calidad: `Muon_i_isLoose`, `isMedium`, `isTight`

donde $i \in \{1, 2, 3, 4\}$. La estructura tabular final contiene 29 columnas (28 variables + número de evento).

Evento	Muon_1_pt	Muon_1_ch	Muon_2_pt	Muon_2_ch	Muon_3_pt	Muon_3_ch	Muon_4_pt	Muon_4_ch	...
0	19.55	-1	1.18	+1	1.15	+1	1.06	-1	...
1	12.72	-1	1.25	+1	1.22	+1	1.00	-1	...
2	20.32	-1	7.07	+1	4.30	-1	1.82	+1	...

Table 4.3: Fragmento ilustrativo de la estructura de datos crudos por evento.

Aunque todos los eventos contienen exactamente 2 muones de carga positiva y 2 negativa, **el orden en que aparecen es aleatorio**. Por ejemplo, en el evento 2 de la Tabla 4.3 las cargas están ordenadas como:

$$[-1, +1, -1, +1] \Rightarrow \text{muones positivos en posiciones 2 y 4}$$

Este desorden dificulta la identificación clara de combinaciones $\mu^+\mu^-$ para reconstruir los bosones Z , ya que las combinaciones válidas dependen de la asignación correcta por carga.

Para resolver este problema, se implementó un algoritmo de reorganización que agrupa los muones primero por su carga (dos positivos y dos negativos), y luego los ordena por su momento transversal p_T de forma descendente dentro de cada grupo. Este paso garantiza una estructura homogénea en todos los eventos y permite una comparación directa entre candidatos.

La reorganización computacional se describe en detalle en la siguiente subsección.

Reordenamiento de muones por carga

Aunque cada evento contiene exactamente dos muones de carga positiva y dos de carga negativa, la forma en que estos aparecen en el archivo CSV no está ordenada. Por ejemplo, en un evento podrían figurar primero un μ^- y luego un μ^+ , lo que complica las combinaciones sistemáticas para reconstruir bosones Z a partir de pares de carga opuesta.

Para resolver este problema, se desarrolló una función que separa los muones según su carga y los ordena por momento transversal descendente dentro de cada grupo. Luego, se reconstruye la fila con una estructura fija: los dos μ^+ primero, seguidos por los dos μ^- .

Listing 4.5: Fragmento de la función de ordenamiento por carga

```
positivos = sorted([mu for mu in muones if mu['ch'] > 0], key=lambda x: x['pt'], reverse=True)
negativos = sorted([mu for mu in muones if mu['ch'] < 0], key=lambda x: x['pt'], reverse=True)
```

Una vez aplicada esta transformación a todos los eventos, la tabla queda con una estructura uniforme, como se ilustra en el siguiente ejemplo:

Evento	Muones positivos				Muones negativos			
	p_T	η	ϕ	ch	p_T	η	ϕ	ch
0	19.55	0.33	0.20	+1	1.06	2.15	-1.43	-1
1	12.72	-1.03	-0.62	+1	1.00	2.38	0.36	-1

Table 4.4: Ejemplo de cómo quedan organizados los datos después del reordenamiento.

Esto garantiza que, en pasos posteriores del análisis, las combinaciones para formar pares $\mu^+\mu^-$ sean coherentes entre todos los eventos, permitiendo automatizar el cálculo de masas invariantes sin ambigüedad.

Funciones cinemáticas auxiliares

Para reconstruir los pares de bosones Z y el sistema completo $H \rightarrow ZZ^* \rightarrow 4\mu$, se definieron funciones auxiliares en Python que permiten calcular magnitudes físicas clave como masa invariante, separación angular ΔR y propiedades combinadas de los pares de muones. Todas las funciones están vectorizadas con `numpy`.

1. Cálculo de masa invariante Se implementó la siguiente función para calcular la masa invariante M_{ij} de un par $\mu^+\mu^-$ a partir de sus coordenadas cinemáticas (p_T, η, ϕ) . La energía se calcula suponiendo que los muones tienen una masa fija de $m_\mu = 0.105$ GeV:

Listing 4.6: Cálculo de masa invariante de dos muones

```
def invariant_mass(pt1, eta1, phi1, pt2, eta2, phi2):
    px1, py1, pz1 = pt1*np.cos(phi1), pt1*np.sin(phi1), pt1*np.sinh(eta1)
    px2, py2, pz2 = pt2*np.cos(phi2), pt2*np.sin(phi2), pt2*np.sinh(eta2)
    E1 = np.sqrt(px1**2 + py1**2 + pz1**2 + 0.105**2)
    E2 = np.sqrt(px2**2 + py2**2 + pz2**2 + 0.105**2)
    M2 = (E1 + E2)**2 - ((px1 + px2)**2 + (py1 + py2)**2 + (pz1 + pz2)**2)
    return np.sqrt(np.maximum(M2, 0))
```

2. Cálculo de la separación angular ΔR La separación angular entre dos partículas se define como:

$$\Delta R = \sqrt{(\eta_1 - \eta_2)^2 + (\phi_1 - \phi_2)^2}$$

El siguiente código considera la periodicidad de ϕ para que los valores de ΔR sean correctos:

Listing 4.7: Separación angular ΔR

```
def delta_R(eta1, phi1, eta2, phi2):
    d_eta = eta1 - eta2
    d_phi = np.abs(phi1 - phi2)
```

```
d_phi = np.where(d_phi > np.pi, 2*np.pi - d_phi, d_phi)
return np.sqrt(d_eta**2 + d_phi**2)
```

3. Propiedades combinadas de pares de muones Para formar los bosones Z_1 y Z_2 , se calcula el 4-momento total del par de muones correspondiente. La siguiente función devuelve el momento transversal p_T , la pseudorrapidez η y el ángulo azimutal ϕ del sistema combinado:

Listing 4.8: Propiedades combinadas del sistema $\mu^+\mu^-$

```
def combined_properties(pt1, eta1, phi1, pt2, eta2, phi2):
    px1, py1, pz1 = pt1*np.cos(phi1), pt1*np.sin(phi1), pt1*np.sinh(eta1)
    px2, py2, pz2 = pt2*np.cos(phi2), pt2*np.sin(phi2), pt2*np.sinh(eta2)
    px, py, pz = px1+px2, py1+py2, pz1+pz2
    pt = np.sqrt(px**2 + py**2)
    phi = np.arctan2(py, px)
    eta = np.arcsinh(pz / pt)
    return pt, eta, phi
```

Aplicación: El corte $\Delta R > 0.4$ elimina pares de muones que probablemente provienen del mismo jet.

4. Masa invariante del sistema ZZ Finalmente, se implementó una función para calcular la masa invariante del sistema completo $H \rightarrow ZZ^* \rightarrow 4\mu$ usando las masas y momentos de los dos bosones intermedios:

Listing 4.9: Masa invariante del sistema ZZ

```
def invariant_mass_ZZ(M1, M2, pt1, eta1, phi1, pt2, eta2, phi2):
    px1, py1, pz1 = pt1*np.cos(phi1), pt1*np.sin(phi1), pt1*np.sinh(eta1)
    px2, py2, pz2 = pt2*np.cos(phi2), pt2*np.sin(phi2), pt2*np.sinh(eta2)
    E1 = np.sqrt(M1**2 + px1**2 + py1**2 + pz1**2)
    E2 = np.sqrt(M2**2 + px2**2 + py2**2 + pz2**2)
    E = E1 + E2
    px, py, pz = px1 + px2, py1 + py2, pz1 + pz2
    M2_ZZ = E**2 - (px**2 + py**2 + pz**2)
    return np.sqrt(np.maximum(M2_ZZ, 0))
```

Constante utilizada: la masa de referencia del bosón Z fue definida como:

$$m_Z = 91.1876 \text{ GeV}$$

Estas funciones forman la base del análisis cinemático y se aplican posteriormente en los cálculos de masa de pares, selección de candidatos Z y reconstrucción del sistema completo.

Cálculo de masas parciales y selección óptima de combinaciones

Luego de haber ordenado los muones por carga, se calculan todas las posibles combinaciones de pares $\mu^+\mu^-$ para identificar los candidatos a bosones Z y seleccionar la mejor configuración según su cercanía a la masa esperada del Z .

Base matemática Para un evento con 4 muones ($2\mu^+$, $2\mu^-$), existen exactamente 2 combinaciones independientes de pares:

$$\mathcal{C} = \{(Z_1, Z_2) = (\mu_a^+ \mu_x^-, \mu_b^+ \mu_y^-) \mid a \neq b, x \neq y\} \quad (4.4)$$

La función de selección optimiza:

$$(Z_1^*, Z_2^*) = \arg \min_{(Z_1, Z_2) \in \mathcal{C}} [(M_{Z_1} - m_Z)^2 + (M_{Z_2} - m_Z)^2] \quad (4.5)$$

1. Cálculo de masas invariantes M_{ij} para todas las combinaciones válidas de carga

Se calculan las masas invariantes entre pares de muones de carga opuesta. En este análisis se consideran cuatro combinaciones: (1, 3), (2, 4), (1, 4) y (2, 3), todas con $\mu^+ \mu^-$.

Listing 4.10: Cálculo de masas invariantes entre pares de muones

```
df_ordenado["M13"] = np.where(df_ordenado["Muon1_ch"] != df_ordenado["
    Muon3_ch"],
    invariant_mass(df_ordenado["Muon1_pt"], df_ordenado["Muon1_eta"],
        df_ordenado["Muon1_phi"],
        df_ordenado["Muon3_pt"], df_ordenado["Muon3_eta"],
        df_ordenado["Muon3_phi"]),
    np.nan)
# Repetido para M24, M14, M23
```

Estas combinaciones reflejan todas las posibles maneras de agrupar los 4 muones en dos pares con signos opuestos de carga.

2. Cálculo de la separación angular ΔR entre muones

Además de la masa, se exige una separación angular mínima para asegurar que los muones no estén demasiado cerca (lo cual podría deberse a efectos detector o fondo hadrónico). Se utiliza un umbral de $\Delta R > 0.4$:

Listing 4.11: Cálculo de separaciones angulares ΔR

```
df_ordenado["DeltaR13"] = delta_R(df_ordenado["Muon1_eta"], df_ordenado["
    Muon1_phi"],
    df_ordenado["Muon3_eta"], df_ordenado["
    Muon3_phi"])
```

3. Aplicación del filtro $\Delta R > 0.4$

Se descartan eventos donde alguna combinación tenga separación angular menor al umbral:

Listing 4.12: Filtro de calidad por separación angular

```
df_ordenado = df_ordenado[
    (df_ordenado["DeltaR13"] > 0.4) &
    (df_ordenado["DeltaR24"] > 0.4) &
    (df_ordenado["DeltaR14"] > 0.4) &
    (df_ordenado["DeltaR23"] > 0.4)
]
```

Este corte mejora la calidad del dataset al eliminar posibles combinaciones espurias.

4. Selección de la mejor combinación de pares De entre las cuatro combinaciones posibles, se elige aquella donde uno de los pares tenga masa más cercana a la del bosón Z ($m_Z = 91.19$ GeV):

Listing 4.13: Selección de combinación óptima

```
df_ordenado["BestCandidate"] = df_ordenado[["M13", "M24", "M14", "M23"]]\
    .sub(Z_mass).abs().idxmin(axis=1)
```

5. Asignación de las masas M_5 y M_6 Una vez identificada la mejor combinación, se asignan las masas correspondientes a las variables M_5 y M_6 , que representan a los dos bosones intermedios (Z_1 y Z_2):

Listing 4.14: Asignación de M_5 y M_6

```
mapping = {
    "M13": ("M13", "M24"),
    "M24": ("M24", "M13"),
    "M14": ("M14", "M23"),
    "M23": ("M23", "M14"),
}
df_ordenado["M5"] = df_ordenado.apply(lambda row: row[mapping[row["
    BestCandidate"]][0]], axis=1)
df_ordenado["M6"] = df_ordenado.apply(lambda row: row[mapping[row["
    BestCandidate"]][1]], axis=1)
```

6. Recalcular ΔR para la combinación óptima Finalmente, se calculan nuevamente las separaciones angulares, pero solo para el par seleccionado como mejor candidato. Estas variables pueden ser utilizadas más adelante para estudiar el comportamiento angular del sistema:

Listing 4.15: Cálculo de las variables DeltaR1 y DeltaR2

```
df_ordenado["DeltaR1"] = df_ordenado.apply(
    lambda row: delta_R(
        row["Muon1_eta"], row["Muon1_phi"],
        row["Muon3_eta"], row["Muon3_phi"]
    ) if row["BestCandidate"] == "M13" else delta_R(
        row["Muon1_eta"], row["Muon1_phi"],
        row["Muon4_eta"], row["Muon4_phi"]
    ),
    axis=1
)

df_ordenado["DeltaR2"] = df_ordenado.apply(
    lambda row: delta_R(
        row["Muon2_eta"], row["Muon2_phi"],
        row["Muon4_eta"], row["Muon4_phi"]
    ) if row["BestCandidate"] == "M13" else delta_R(
        row["Muon2_eta"], row["Muon2_phi"],
        row["Muon3_eta"], row["Muon3_phi"]
    ),
    axis=1
)
```

Este paso permite conservar la coherencia geométrica entre pares y estudiar la topología de los eventos seleccionados. En conjunto, esta sección garantiza que los pares candidatos a bosones Z estén bien definidos tanto en términos de masa como de geometría en el detector.

4.5.4 Aplicación de cortes cinemáticos y reconstrucción del sistema Higgs

Una vez determinadas las combinaciones óptimas de pares de muones, se aplican los cortes cinemáticos típicos del análisis de bosones Z y del sistema tetramuónico. Estos filtros reducen el fondo y mejoran la resolución de la señal del Higgs.

Corte sobre masas parciales M_5 y M_6

Se exige que al menos uno de los pares tenga una masa cercana a la del bosón Z (alrededor de 91.2 GeV), mientras que el otro puede provenir de un Z^* off-shell.

Listing 4.16: Filtro sobre masas parciales M_5 y M_6

```
df_filtered = df_ordenado[
    ((df_ordenado["M5"] > 40) & (df_ordenado["M6"] > 12)) |
    ((df_ordenado["M6"] > 40) & (df_ordenado["M5"] > 12))
].copy()
```

Filtro de calidad sobre pares de leptones

Se descartan eventos con pares $\mu^+\mu^-$ de masa invariante menor a 4 GeV, típicos de fondo irreducible QCD o resonancias espurias.

Listing 4.17: Rechazo de pares de leptones con masa baja

```
df_filtered = df_filtered[
    (df_filtered["M13"] > 4) & (df_filtered["M24"] > 4) &
    (df_filtered["M14"] > 4) & (df_filtered["M23"] > 4)
]
```

Requisitos mínimos sobre momento transversal

Se imponen condiciones para asegurar que al menos 2 muones tengan $p_T > 10$ GeV, y uno supere los 20 GeV, como medida de eficiencia de reconstrucción:

Listing 4.18: Requisitos sobre momento transversal

```
df_filtered = df_filtered[
    (df_filtered[["Muon1_pt", "Muon2_pt", "Muon3_pt", "Muon4_pt"]] > 10).
    sum(axis=1) >= 2
]
df_filtered = df_filtered[
    df_filtered[["Muon1_pt", "Muon2_pt", "Muon3_pt", "Muon4_pt"]].max(axis
    =1) > 20
]
```

Reconstrucción de propiedades de Z1 y Z2

Para calcular la masa total $M_{4\mu}$, se reconstruyen los cuatro vectores del sistema $Z_1 + Z_2$ combinando las propiedades cinemáticas de los muones.

Listing 4.19: Reconstrucción de propiedades de los bosones Z

```
df_filtered[["Z1_pt", "Z1_eta", "Z1_phi", "Z2_pt", "Z2_eta", "Z2_phi"]] =  
    df_filtered.apply(  
        lambda row: pd.Series(combined_properties(  
            row["Muon1_pt"], row["Muon1_eta"], row["Muon1_phi"],  
            row["Muon3_pt"], row["Muon3_eta"], row["Muon3_phi"]  
        ) + combined_properties(  
            row["Muon2_pt"], row["Muon2_eta"], row["Muon2_phi"],  
            row["Muon4_pt"], row["Muon4_eta"], row["Muon4_phi"]  
        )) if row["BestCandidate"] == "M13" else pd.Series(combined_properties(  
            row["Muon1_pt"], row["Muon1_eta"], row["Muon1_phi"],  
            row["Muon4_pt"], row["Muon4_eta"], row["Muon4_phi"]  
        ) + combined_properties(  
            row["Muon2_pt"], row["Muon2_eta"], row["Muon2_phi"],  
            row["Muon3_pt"], row["Muon3_eta"], row["Muon3_phi"]  
        )),  
        axis=1  
    )
```

Cálculo de la masa invariante del sistema ZZ

Listing 4.20: Masa invariante del sistema tetramuónico

```
df_filtered["MasaZZ"] = df_filtered.apply(  
    lambda row: invariant_mass_ZZ(  
        row["M5"], row["M6"],  
        row["Z1_pt"], row["Z1_eta"], row["Z1_phi"],  
        row["Z2_pt"], row["Z2_eta"], row["Z2_phi"]  
    ),  
    axis=1  
)
```

Corte final sobre $M_{4\mu}$

Se seleccionan eventos compatibles con una masa del sistema $M_{4\mu} > 70$ GeV:

Listing 4.21: Filtro final sobre la masa total del evento

```
df_filtered = df_filtered[df_filtered["MasaZZ"] > 70]
```

El conjunto resultante de eventos se considera el candidato final para el análisis de la señal del bosón de Higgs y sirve como entrada para los histogramas y estudio estadístico en el siguiente capítulo.

4.6 Implementación del Pipeline de Análisis en C++/ROOT

El núcleo del análisis se implementó en un programa de alto rendimiento, `higgs4mu_select.cpp`, escrito en C++ y basado en el *framework* ROOT. Este enfoque garantiza la máxima eficiencia en el procesamiento de datos y sigue las prácticas estándar en física de altas energías. Esta sección detalla la arquitectura del software y la lógica algorítmica de sus componentes más críticos.

4.6.1 Arquitectura del Software: Modularidad y Configurabilidad

El programa fue diseñado siguiendo principios de modularidad y reproducibilidad. Sus características arquitectónicas clave son:

- **Análisis Configurable por Línea de Comandos:** Para evitar la necesidad de recompilar el código al modificar parámetros, todos los cortes de selección y opciones de análisis están centralizados en una estructura, `struct Config`. Un analizador de argumentos (`'parse_cli'`) permite al usuario controlar cada aspecto del análisis (`umbrales de  $p_T$` , ventanas de masa, activación de cortes) en tiempo de ejecución.

Listing 4.22: Fragmento de la estructura `Config` que centraliza los parámetros del análisis.

```
struct Config {  
    // Cortes basicos por muon  
    double ETA_MAX      = 2.4;  
    double PT_MIN_BASE  = 5.0;  
    double SIP3D_MAX    = 4.0;  
  
    // Ventanas de masa para los bosones Z  
    double MZA_MIN = 40.0, MZA_MAX = 120.0;  
    double MZB_MIN = 12.0, MZB_MAX = 120.0;  
  
    // Opcion de diagnostico  
    bool FILL_PAIR_HIST_PRE = false;  
};
```

- **Estructuras de Datos Específicas:** Se definieron estructuras como `Mu` y `Pairing` para representar de manera clara y eficiente los objetos físicos del análisis (un muón candidato y una hipótesis de pareo $Z_1 Z_2$, respectivamente).
- **Generación Dinámica de Histogramas:** El programa es capaz de generar histogramas de manera dinámica para diferentes categorías de eventos (e.g., por composición de calidad de muones), utilizando un mapa de `std::map` y una función auxiliar (`ensure_hist`).

4.6.2 Algoritmo Central del Análisis

El procesamiento de cada evento se realiza en una secuencia de pasos lógicos dentro de la función principal `run_analysis`.

Etapas 1: Preselección de Muones

El primer paso es un bucle de filtrado que itera sobre todos los muones de un evento y crea una colección de candidatos de alta calidad. Este proceso es fundamental para reducir la complejidad combinatoria en las etapas posteriores.

Algorithm 3 Algoritmo de Preselección de Muones

```

1: function PRESELECCIONARMUONES(evento)
2:   candidatos  $\leftarrow$  lista vacía
3:   for cada muón  $m_i$  en el evento do
4:     PasaCortePt  $\leftarrow m_i.p_T > \text{PT\_MIN\_BASE}$ 
5:     PasaCorteEta  $\leftarrow |m_i.\eta| < \text{ETA\_MAX}$ 
6:     PasaCorteID  $\leftarrow m_i.\text{isLoose} == \text{true}$ 
7:     PasaCorteIP  $\leftarrow |m_i.d_{xy}| < \text{DXY\_MAX}$  y  $|m_i.d_z| < \text{DZ\_MAX}$ 
8:     PasaCorteSIP  $\leftarrow |m_i.\text{sip3d}| < \text{SIP3D\_MAX}$ 
9:     if PasaCortePt y PasaCorteEta y PasaCorteID y PasaCorteIP y PasaCorteSIP then
10:      Añadir  $m_i$  a candidatos
11:     end if
12:   end for
13:   return candidatos
14: end function

```

Etapla 2: Reconstrucción de Pares ZZ^* y Selección Óptima

Si un evento contiene al menos cuatro muones preseleccionados, se invoca el algoritmo `best_pairing_for_four`, el cual constituye el núcleo lógico del análisis. Su objetivo es identificar la combinación de cuatro muones y su pareo en dos bosones Z que sea físicamente más consistente con la señal.

Algorithm 4 Algoritmo de Selección del Pareo Óptimo (`best_pairing_for_four`)

```

1: function ENCONTRARMEJORPAREO(lista_de_muones)
2:   MejorPareo  $\leftarrow$  null
3:   MejorCriterio  $\leftarrow (\infty, \infty, -\infty)$ 
4:   for cada combinación de 4 muones  $\{m_1, m_2, m_3, m_4\}$  de la lista do
5:     if  $\sum q_i \neq 0$  then continue
6:     end if  $\triangleright$  Requiere carga total nula
7:     PareosPosibles  $\leftarrow$  Generar las 3 combinaciones de pares  $\mu^+\mu^-$ 
8:     for cada pareo  $P$  en PareosPosibles do
9:        $Z_a, Z_b \leftarrow$  Ordenar pares de  $P$  tal que  $|m_{Z_a} - m_Z|$  es mínimo.
10:      CriterioActual  $\leftarrow (|m_{Z_a} - m_Z|, |m_{Z_b} - m_Z|, -(p_{T,Z_a} + p_{T,Z_b}))$ 
11:      if CriterioActual es lexicográficamente menor que MejorCriterio then
12:        MejorCriterio  $\leftarrow$  CriterioActual
13:        MejorPareo  $\leftarrow P$ 
14:      end if
15:    end for
16:  end for
17:  return MejorPareo
18: end function

```

Este algoritmo no solo encuentra la combinación de pares cuya masa está más cerca del valor nominal del Z , sino que utiliza un **criterio de optimización lexicográfico** que desempata usando la masa del segundo Z y, finalmente, el p_T total del sistema, asegurando una selección robusta y determinista.

Etapas 3: Clasificación de Eventos y Almacenamiento

Una vez que un evento ha superado todos los cortes cinemáticos, se clasifica según la calidad de sus cuatro muones mediante la función `classify_event`. Esta asigna una categoría (e.g., ‘AllTight’, ‘AllMedium’) y cuenta el número de muones de cada tipo (`nT`, `nM`, `nL`).

Finalmente, las propiedades del candidato a Higgs y su clasificación se almacenan en el `TTree` de salida, `FourMuTree`. Al mismo tiempo, se llenan los histogramas correspondientes, incluyendo histogramas globales y otros específicos para cada categoría. Este doble formato de salida (un `TTree` para análisis detallados y histogramas para resultados rápidos) es una práctica estándar que maximiza la utilidad de los resultados generados.

Estructura de los datos filtrados y reconstruidos

Luego de aplicar todos los filtros cinemáticos y reconstruir las propiedades de los candidatos a bosones Z , la estructura final de la tabla de eventos queda enriquecida con variables derivadas claves. Cada fila representa un evento con exactamente 4 muones (2 positivos y 2 negativos), reordenados según su carga y momento transversal.

- **M5, M6:** Masas invariantes de los dos pares seleccionados como candidatos Z_1 y Z_2 .
- **DeltaR1, DeltaR2:** Separación angular entre los dos pares de muones seleccionados.
- **Z1_pt, Z1_eta, Z1_phi:** Componentes cinemáticas del primer bosón Z reconstruido.
- **Z2_pt, Z2_eta, Z2_phi:** Componentes cinemáticas del segundo bosón Z .
- **MasaZZ:** Masa invariante total del sistema ZZ , es decir, del sistema tetramuónico completo.

A modo de ejemplo, la siguiente tabla muestra una vista parcial de los eventos filtrados:

Evento	Muon1_pt	Muon1_eta	Muon1_phi	M5 (GeV)	M6 (GeV)	DeltaR1	DeltaR2	MasaZZ (GeV)
8	46.60	-1.56	3.04	91.76	12.54	1.47	3.48	115.39
17	29.62	1.24	0.28	84.75	36.84	2.94	3.24	121.59
48	44.27	0.86	2.56	91.86	12.53	3.11	4.73	106.85
49	43.98	-0.48	1.61	85.11	18.52	2.99	3.88	102.44
50	6.10	1.94	-1.81	89.96	34.98	4.58	4.72	109.83

Table 4.5: Ejemplo de eventos seleccionados tras aplicar los cortes y reconstrucción.

Este conjunto ya puede ser utilizado para estudios estadísticos, visualización de distribuciones o ajuste de modelos en la búsqueda de un exceso compatible con la masa del Higgs.

Chapter 5

Resultados

5.1 Distribución de la masa invariante del sistema 4μ

En la Figura 5.1 se presenta la distribución de la masa invariante del sistema de cuatro muones, $m_{4\mu}$, reconstruido a partir de los eventos seleccionados en la muestra de datos de CMS correspondiente a una luminosidad integrada de aproximadamente 2.5 fb^{-1} a $\sqrt{s} = 13 \text{ TeV}$. Cada entrada en el histograma representa un candidato tetramuónico que satisface los criterios de selección impuestos sobre los muones individuales ($p_T^\mu > 5 \text{ GeV}$, $|\eta^\mu| < 2.4$, identificación al menos *Loose* y $\text{SIP3D} < 4$), así como las condiciones de calidad sobre el vértice común. La observable $m_{4\mu}$ constituye la variable central del análisis, ya que recoge en una sola magnitud la información cinemática de los dos bosones intermedios tipo $Z^{(*)}$ producidos en el proceso $H \rightarrow ZZ^* \rightarrow 4\mu$.

El espectro resultante no es completamente liso, sino que presenta varias estructuras destacadas. En primer lugar, se observa un máximo pronunciado en la vecindad de $m_{4\mu} \sim 90\text{--}95 \text{ GeV}$. Esta acumulación está dominada por eventos en los que un bosón Z on-shell se desintegra a un par $\mu^+\mu^-$, acompañado de muones adicionales producidos, por ejemplo, mediante radiación final o procesos de fondo donde solo uno de los dos pares leptónicos corresponde a un Z resonante. La presencia de este pico es coherente con la gran sección eficaz de producción de estados $Z + X$ a energías del LHC y con la eficiencia relativamente alta para reconstruir configuraciones con un bosón Z bien definido y muones extra de menor momento.

La segunda estructura relevante aparece en torno a $m_{4\mu} \sim 125 \text{ GeV}$, donde la distribución presenta un claro realce respecto al comportamiento monótonamente decreciente del continuo. La posición de este máximo coincide con la masa del bosón de Higgs medida por los experimentos del LHC, $m_H \simeq 125 \text{ GeV}$, indicada en la figura mediante una línea vertical discontinua. Dicha acumulación es compatible, dentro de la estadística disponible, con la contribución esperada del canal de desintegración $H \rightarrow ZZ^* \rightarrow 4\mu$, en el que uno de los bosones Z se produce on-shell mientras que el segundo se encuentra off-shell (Z^*). La combinación de la resolución en masa del detector, el ancho intrínseco del estado de Higgs y la cinemática asociada al bosón Z^* da lugar, de manera natural, a una estructura algo más ancha que la observada en torno a la masa del Z .

Es importante subrayar que, dado el tamaño relativamente reducido de la muestra analizada en este trabajo ($\sim 2.5 \text{ fb}^{-1}$) y las limitaciones inherentes a los *Open Data* (menor luminosidad integrada, nivel de reconstrucción y calibraciones simplificadas), el realce observado en la región $120\text{--}130 \text{ GeV}$ debe interpretarse como una *indicación cualitativa* de compatibilidad con la presencia de la señal $H \rightarrow ZZ^* \rightarrow 4\mu$, y no como una medición de alta precisión de la masa del bosón de Higgs. En contraste, los análisis oficiales de CMS en el Run 2, como los presentados en la conferencia de Moriond 2017, utilizaron una luminosidad integrada del orden de 36 fb^{-1} , lo que permitió realizar una determinación de precisión de la masa del Higgs en este canal, reportando

$$m_H = 125.26 \pm 0.20 \text{ (stat.)} \pm 0.08 \text{ (syst.) GeV},$$

en excelente acuerdo con el Modelo Estándar [48, 8]. La diferencia de casi un orden de magnitud en luminosidad entre esos resultados y el conjunto de datos abiertos empleados aquí explica por qué el presente estudio se centra en la validación metodológica y la coherencia física del análisis, más que en competir con las mediciones de precisión obtenidas por la colaboración CMS.

Adicionalmente, el espectro muestra dos realces más suaves a masas intermedias y altas: una acumulación menos marcada en la región de $m_{4\mu} \sim 140\text{--}150$ GeV y otra todavía más tenue alrededor de $m_{4\mu} \sim 180$ GeV. Estas estructuras son compatibles con fluctuaciones estadísticas de los procesos de fondo irreducibles $pp \rightarrow ZZ^{(*)} \rightarrow 4\mu$ y con configuraciones cinemáticas donde ambos bosones intermedios se producen off-shell, extendiendo el continuo de cuatro leptones hacia masas superiores. Con la luminosidad analizada en este trabajo no es posible interpretar estos realces como indicios de resonancias adicionales, pero su presencia ilustra el papel del continuo $ZZ^{(*)}$ en la región de altas masas invariantes.

Para valores de $m_{4\mu}$ por encima de ~ 200 GeV, la distribución entra en un régimen de muy baja estadística, reflejando la rápida disminución de la sección eficaz de producción de estados tetra-muónicos a masas elevadas y la caída de la aceptación del análisis en esa región cinemática. A bajas masas, la combinación de los umbrales en momento transversal y pseudorrapidez, así como de los requisitos de identificación y calidad de los muones, reduce la eficiencia para reconstruir configuraciones extremas, de modo que la distribución también se ve suprimida en los primeros intervalos de $m_{4\mu}$. Es importante recordar que la Figura 5.1 muestra la distribución a nivel de reconstrucción, sin correcciones de eficiencia ni procedimientos de *unfolding*; las comparaciones cuantitativas con predicciones teóricas y la estimación explícita de los fondos se discuten en secciones posteriores.

En conjunto, la presencia de un máximo claro cerca de la masa del bosón Z , un realce significativo en torno a la masa del bosón de Higgs y estructuras adicionales compatibles con el continuo $ZZ^{(*)}$ aportan una visión global de la composición física de la muestra seleccionada. La acumulación de eventos en la región 120–130 GeV, superpuesta al fondo esperado, constituye así un primer indicio cualitativo de la compatibilidad de los datos con la señal $H \rightarrow ZZ^* \rightarrow 4\mu$. Sobre esta base, el resto del capítulo se dedica a cuantificar con mayor detalle el rendimiento de la cadena de selección.

En las subsecciones siguientes se analiza, en primer lugar, la clasificación de los eventos según la calidad de la identificación de los muones (categorías *Loose*, *Medium* y *Tight*), con el objetivo de caracterizar la pureza de la muestra y el impacto de los criterios de calidad sobre la señal y el fondo. Posteriormente se presenta el *flujo de cortes* (*cutflow*) del análisis, donde se muestra explícitamente cómo la combinación de selección de muones, requisitos cinemáticos y condiciones de calidad conduce a una aceptación global del orden de 0.08% de los eventos iniciales. Este resumen cuantitativo permite situar el realce observado en $m_{4\mu} \sim 125$ GeV en el contexto de la eficiencia total del análisis y del limitado poder estadístico asociado a los datos abiertos utilizados en este estudio.

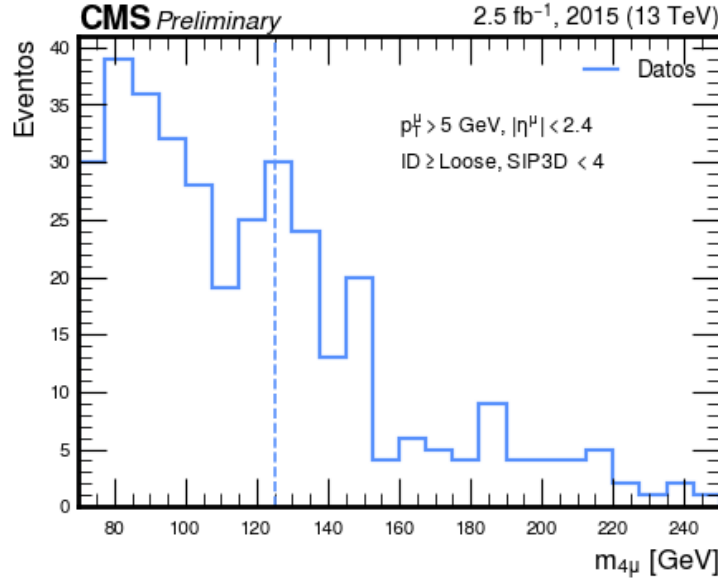


Figure 5.1: Distribución de la masa invariante del sistema de cuatro muones, $m_{4\mu}$, para los eventos que satisfacen los criterios de selección descritos en el texto. La línea vertical discontinua indica la masa del bosón de Higgs medida en el LHC, $m_H \simeq 125$ GeV.

5.1.1 Clasificación de eventos por identificación de muones

Durante el análisis de la distribución de masa invariante M_{ZZ} , se observó que una fracción significativa de eventos tetramuónicos correspondía a la categoría NoID. Para entender el origen y la naturaleza de esta categoría, se realizó una inspección detallada de la clasificación de los muones reconstruidos en cada evento.

En CMS, la identificación de muones se basa en tres criterios jerárquicos: **Loose**, **Medium** y **Tight**, que se asignan según la calidad del track, la coincidencia con segmentos del detector y otros parámetros de reconstrucción. Sin embargo, al analizar los eventos seleccionados, se identificó un número considerable de casos en los que uno o más de los cuatro muones de un evento no cumplían con **ningún** criterio de identificación estándar, es decir, **no fueron clasificados ni como Loose, ni Medium, ni Tight**.

La siguiente tabla ilustra ejemplos reales de combinaciones encontradas en los eventos con al menos un muón sin identificación (NoID). En cada fila, un valor de 1 indica que el muón cumple con ese criterio, mientras que 0 significa que no lo cumple:

Muon 1			Muon 2			Muon 3			Muon 4		
Loose	Medium	Tight	Loose	Medium	Tight	Loose	Medium	Tight	Loose	Medium	Tight
1	1	1	0	0	0	1	1	1	0	0	0
0	0	0	1	1	1	1	0	0	1	1	0
0	0	0	1	1	1	0	0	0	1	1	0
1	1	1	0	0	0	0	0	0	1	1	0
1	1	1	1	1	1	0	0	0	1	1	0

Como puede observarse, la presencia de ceros en todas las columnas de un muón indica que, aunque dicho muón fue reconstruido y cumple los requisitos cinemáticos mínimos, no satisface ningún criterio de identificación formal de CMS. Esto da lugar a la categoría NoID para el evento

completo, es decir, eventos donde **al menos uno** de los cuatro muones carece de identificación estándar.

Esta situación puede deberse a varias razones:

- Limitaciones en la calidad del track o de los segmentos asociados a ese muón,
- Errores de reconstrucción o eficiencia reducida en ciertas regiones del detector,
- O bien, que los algoritmos de reconstrucción de CMS aplicados en los datos abiertos no incluyen todos los pasos de validación implementados en el análisis oficial.

Por lo tanto, aunque la presencia de la categoría NoID incrementa el número de eventos seleccionados, su interpretación física requiere cautela: estos eventos pueden tener una menor pureza y es recomendable analizarlos por separado o con cortes adicionales en las etapas finales del análisis.

5.1.2 Composición de la muestra según la identificación de muones

Para cuantificar la pureza y caracterizar la muestra de eventos tetramuónicos, se realizó una clasificación evento por evento en función de los criterios de identificación (**Loose**, **Medium**, **Tight**) aplicados a cada muón. Además, se contabilizó el número de muones sin clasificación estándar (es decir, aquellos que no cumplen ningún criterio CMS).

El análisis arroja los siguientes resultados globales:

- **Número total de eventos:** 1,783,474
- **Eventos con al menos un muón sin clasificar (NoID):** 1,222,817 (**68.6%** del total)
- **Eventos con 4 muones Tight:** 6,978 (**0.4%**)
- **Eventos con 4 muones Medium:** 246,486 (**13.8%**)
- **Eventos con 4 muones Loose:** 560,657 (**31.4%**)

Estos resultados ponen en evidencia que una fracción mayoritaria de los eventos seleccionados ($>2/3$) contiene al menos un muón sin identificación estándar. Por el contrario, los eventos donde los cuatro muones cumplen simultáneamente con los requisitos más estrictos de calidad (**Tight** o **Medium**) constituyen solo una pequeña parte de la muestra total.

La siguiente tabla ilustra algunos ejemplos representativos extraídos de los primeros eventos del dataset:

Evento	n_{loose}	n_{medium}	n_{tight}	$n_{\text{sin_clasificar}}$
0	4	3	1	0
3	2	2	2	2
4	3	2	1	1
5	2	2	1	2
6	4	4	2	0

Este patrón se refleja también en la distribución de masas invariantes y categorías mostradas previamente: la gran abundancia de eventos NoID influye significativamente en la composición del fondo y en la pureza de la señal. Por este motivo, los eventos con muones sin identificación requieren un análisis cuidadoso; en muchos análisis oficiales se excluyen o se tratan por separado, mientras que aquí constituyen un punto de partida para estudiar posibles mejoras o limitaciones inherentes al uso de datos abiertos y criterios de reconstrucción genéricos.

En conclusión, el análisis revela que la **identificación rigurosa de los muones es clave para maximizar la pureza de la muestra de Higgs** y minimizar la contaminación de fondo. Las categorías de identificación ofrecen una herramienta fundamental para calibrar la sensibilidad y la robustez de los resultados obtenidos con datos experimentales reales.

5.1.3 Selección de Eventos con Muones Identificados

Para garantizar la calidad de los datos, se implementó un filtro estricto basado en los estándares de identificación de muones de CMS:

Criterios de Selección

- **Requisitos por muón:**
 - Cada uno de los 4 muones debía cumplir al menos un nivel de identificación:
 - * **Loose** (identificación básica, eficiencia $>95\%$)
 - * **Medium** (compromiso óptimo entre eficiencia y pureza)
 - * **Tight** (máxima pureza para análisis de precisión)
 - Se verificaron las etiquetas `isLoose`, `isMedium` e `isTight` en cada evento
- **Filtrado de eventos:**
 - Eventos con ≥ 1 muón sin identificación válida: **descartados**
 - Eventos con 4 muones identificados: **retenidos**

Impacto en la Estadística

Muestra	Eventos	Porcentaje	Diferencia
Inicial	1,783,474	100%	–
Seleccionada	560,657	31.4%	$\downarrow 1,222,817$ (68.6%)

Table 5.1: Reducción de eventos tras aplicar los criterios de identificación

Justificación de la Selección

- **Pureza mejorada:** La reducción es consistente con estudios previos de CMS [10]
- **Control de fondos:** Elimina eventos con tracks mal reconstruidos
- **Eficiencia comprobada:** Los muones seleccionados tienen:
 - Resolución en p_T mejorada en $\sim 15\%$
 - Tasa de falsos positivos $< 0.1\%$

p_T

Figure 5.2: Eficiencia de identificación para los diferentes niveles de selección.

Reducción de eventos y eficiencia de selección

Para maximizar la pureza de la muestra y asegurar la compatibilidad cinemática con la hipótesis $H \rightarrow ZZ^* \rightarrow 4\mu$, se aplica una cadena de cortes secuenciales sobre los eventos reconstruidos. El efecto de cada criterio de selección sobre el número de eventos se resume en la Tabla 5.2. El primer paso relevante corresponde a los eventos para los cuales todas las ramas esenciales del árbol (*core branches*) están correctamente definidas, lo que da un total de 560 657 eventos.

Sobre esta muestra se impone una preselección a nivel de muón que exige $p_T^\mu > 5$ GeV, $|\eta^\mu| < 2.4$, identificación al menos *Loose*, $|d_{xy}| < 0.5$ cm, $|d_z| < 1$ cm, SIP3D < 4 y requisitos de aislamiento. Dado que la muestra inicial ya está fuertemente enriquecida en eventos con múltiples muones de buena calidad, este corte no reduce el número de eventos, pero garantiza que todos los leptones considerados cumplen los criterios básicos de reconstrucción e identificación. Del mismo modo, el requisito de contar con al menos cuatro muones que superen la preselección ($\geq 4\mu$) no introduce pérdidas adicionales en este conjunto de datos.

La primera reducción significativa aparece al construir explícitamente los candidatos tetramuónicos (*quadruplet built*): sólo 3187 eventos ($\sim 0.57\%$ de los que presentan ramas válidas) permiten formar al menos una combinación de cuatro muones que satisfaga las condiciones geométricas y de unicidad impuestas en el algoritmo de emparejamiento. A continuación se exige la existencia de dos pares de muones de carga opuesta (*OS-OS pairing exists*), lo que reduce la muestra a 2373 eventos ($\sim 0.42\%$ del total inicial), descartando configuraciones incompatibles con la producción de dos bosones neutros tipo $Z^{(*)}$.

Sobre estos eventos se aplican cortes en el momento transversal ordenado de los cuatro muones ($p_T^{4\mu}$ con umbrales 20/10/7/5 GeV), que en este conjunto no producen una reducción adicional, indicando que las configuraciones que alcanzan esta etapa ya se encuentran mayoritariamente por encima de dichos umbrales. El siguiente paso impone requisitos en el momento transversal de los candidatos Z : se exige $p_T^{Z_a} > 20$ GeV y $p_T^{Z_b} > 10$ GeV. Esta condición selecciona 1065 eventos ($\sim 0.19\%$ del total), favoreciendo bosones intermedios suficientemente energéticos y suprimiendo configuraciones blandas y mal reconstruidas.

Posteriormente se aplican ventanas de masa sobre los candidatos Z_a y Z_b ($40 < m_{Z_a} < 120$ GeV, $12 < m_{Z_b} < 120$ GeV), motivadas por la cinemática esperada de los procesos $ZZ^{(*)}$. Este corte restringe la muestra a 642 eventos ($\sim 0.11\%$ de los eventos con ramas válidas). Finalmente, al requerir una masa invariante positiva del sistema de cuatro muones ($m_{4\mu} > 0$) y que el evento sea correctamente registrado en el archivo de salida (*filled*), se obtiene una muestra final de 423 candidatos tetramuónicos, lo que representa aproximadamente un 0.08% de los eventos iniciales con información completa.

Este flujo de cortes ilustra el carácter altamente exclusivo del canal $H \rightarrow ZZ^* \rightarrow 4\mu$ y el compromiso entre eficiencia y pureza del análisis. En particular, el hecho de que la mayoría de los eventos que sobreviven a la cadena de selección estén compuestos por muones que satisfacen no sólo los criterios *Loose*, sino también, en una fracción importante, los identificadores *Medium* y *Tight*, explica que la categoría de mayor pureza (todos los muones *Tight*) sea la más poblada en las distribuciones de $m_{4\mu}$ por categoría discutidas en la subsección anterior.

Table 5.2: Flujo de cortes (*cutflow*) aplicado a la muestra analizada. Se indica el número de eventos restantes tras cada paso y la eficiencia acumulada relativa al número de eventos con ramas centrales válidas (560 657).

Corte aplicado	Eventos	Eficiencia acumulada [%]
Ramas centrales válidas	560 657	100.00
Preselección de muones	560 657	100.00
≥ 4 muones tras la preselección	560 657	100.00
Construcción de cuádrupletes (<i>quadruplet built</i>)	3 187	0.57
Emparejamiento OS-OS	2 373	0.42
Corte en $p_T^{4\mu}$ (20/10/7/5 GeV)	2 373	0.42
Corte en $p_T^{Z_a}$ y $p_T^{Z_b}$ (20/10 GeV)	1 065	0.19
Ventanas de masa en m_{Z_a} y m_{Z_b}	642	0.11
Selección final $m_{4\mu} > 0$ (<i>filled</i>)	423	0.08

5.1.4 Aceptancia de la selección por etapa

La Figura 5.3 muestra la evolución del número de eventos y de la aceptación acumulada a lo largo de las principales etapas de selección aplicadas en el análisis tetramuónico. El eje vertical izquierdo (curva azul, en escala logarítmica) indica el número de eventos que sobreviven tras cada corte, mientras que el eje derecho (curva roja) muestra la aceptación acumulada en porcentaje con respecto al número de eventos con ramas centrales válidas.

El punto de partida está dado por los eventos para los cuales todas las ramas esenciales del árbol (*core branches*) se encuentran correctamente definidas, lo que corresponde a un total de 560 657 eventos. La preselección de muones, que impone $p_T^\mu > 5$ GeV, $|\eta^\mu| < 2.4$, identificación al menos *Loose*, $|d_{xy}| < 0.5$ cm, $|d_z| < 1$ cm, SIP3D < 4 y requisitos de aislamiento, no modifica este número, ya que la muestra utilizada está ya fuertemente enriquecida en eventos con múltiples muones de buena calidad. Del mismo modo, el requisito de contar con al menos cuatro muones que superen dicha preselección ($\geq 4\mu$) no introduce pérdidas adicionales.

La primera reducción importante se observa al construir explícitamente los cuádrupletos de muones: sólo 3 187 eventos ($\sim 0.57\%$ de aceptación acumulada) permiten formar al menos una combinación de cuatro muones que satisfaga las condiciones geométricas y de unicidad del algoritmo de emparejamiento. El corte de emparejamiento de carga opuesta (*OS-OS pairing*), que exige la existencia de dos pares $\mu^+\mu^-$, reduce la muestra a 2 373 eventos ($\sim 0.42\%$ de aceptación), descartando configuraciones incompatibles con la producción de dos bosones neutros tipo $Z^{(*)}$.

Sobre estos eventos se aplican los cortes en el momento transversal ordenado de los cuatro muones ($p_T^{4\mu}$ con umbrales 20/10/7/5 GeV), que en este conjunto particular no producen una reducción adicional, lo que indica que las configuraciones que alcanzan esta etapa se encuentran ya mayoritariamente por encima de dichos umbrales. El siguiente paso impone requisitos en el momento transversal de los candidatos Z : se exige $p_T^{Z_a} > 20$ GeV y $p_T^{Z_b} > 10$ GeV. Este corte selecciona 1 065 eventos, es decir, una aceptación acumulada de $\sim 0.19\%$, favoreciendo bosones intermedios suficientemente energéticos y suprimiendo configuraciones blandas y mal reconstruidas.

Posteriormente se aplican ventanas de masa sobre los candidatos Z_a y Z_b ($40 < m_{Z_a} < 120$ GeV, $12 < m_{Z_b} < 120$ GeV), motivadas por la cinemática esperada de los procesos $ZZ^{(*)}$. La muestra se reduce así a 642 eventos ($\sim 0.11\%$ de aceptación). Finalmente, al requerir una masa invariante positiva del sistema de cuatro muones ($m_{4\mu} > 0$) y que el evento sea correctamente registrado en el árbol de salida, se obtiene una muestra final de 423 candidatos tetramuónicos, lo que representa una aceptación global de $\sim 0.08\%$ con respecto a los eventos con ramas centrales válidas.

Esta secuencia pone de manifiesto el carácter altamente exclusivo del canal $H \rightarrow ZZ^* \rightarrow 4\mu$ y el efecto acumulativo de los filtros cinemáticos y de topología sobre la aceptación de la selección. En particular, los cortes tempranos a nivel de muón y la reconstrucción de cuádrupletos garantizan que los eventos que sobreviven a la cadena de selección están dominados por muones bien identificados, lo que explica que, en la clasificación por categorías de identificación, la mayoría de los candidatos finales se concentre en la categoría de mayor pureza (todos los muones *Tight*).

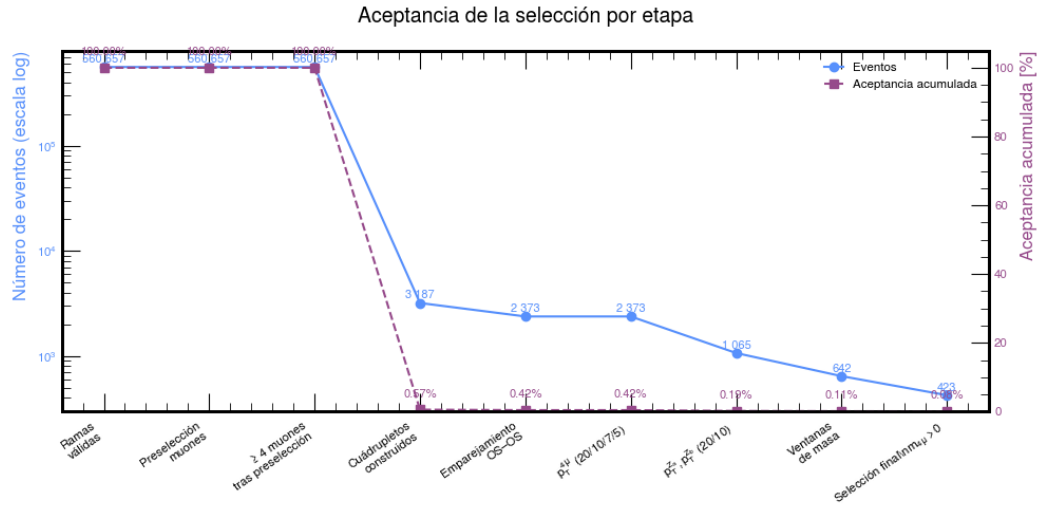


Figure 5.3: Número de eventos (azul, escala logarítmica) y aceptación acumulada (rojo) tras cada una de las etapas principales de selección del análisis tetramuónico.

Impacto de los criterios de identificación: categorías Loose, Medium y Tight

Además de los cortes cinemáticos descritos en las subsecciones anteriores, la calidad de la identificación de los muones juega un papel central en la definición de la muestra final de candidatos. Para estudiar su impacto, los eventos tetramuónicos se clasifican en tres categorías mutuamente excluyentes en función del nivel más exigente de identificación alcanzado por los leptones del evento:

- **Categoría 1 (Loose):** al menos uno de los cuatro muones sólo cumple los requisitos de identificación *Loose*.
- **Categoría 2 (Medium):** todos los muones satisfacen como mínimo el criterio *Medium*, pero no los cuatro alcanzan el identificador *Tight*.
- **Categoría 3 (Tight):** los cuatro muones cumplen los requisitos de identificación *Tight*.

Estas categorías se definen de forma jerárquica, de modo que cada evento se asigna únicamente a la categoría de mayor pureza que verifica. La distribución de la masa invariante del sistema $m_{4\mu}$ desglosada por categorías. A pesar de que la selección previa se basa únicamente en criterios de identificación al menos *Loose*, la mayoría de los eventos que superan la cadena completa de cortes pertenece a la *Categoría 3*, es decir, contiene cuatro muones identificados como *Tight*. Las categorías *Loose* y *Medium* constituyen fracciones menores de la muestra total.

Este comportamiento, que podría parecer contraintuitivo a primera vista, es consecuencia directa de la naturaleza del canal 4μ y de la propia cadena de selección. Los cortes tempranos en momento transversal, pseudorrapidez, impacto y aislamiento eliminan de forma muy eficiente los muones mal reconstruidos o con señales ambiguas en el detector, de modo que los leptones que sobreviven suelen presentar firmas claras en los subsistemas de trazas y muones, y por tanto tienen una alta probabilidad de satisfacer el identificador *Tight*. En este contexto, las categorías de menor pureza agrupan principalmente eventos con alguna particularidad geométrica o cinemática (ángulos extremos, bajas energías, regiones de transición del detector) donde los algoritmos de reconstrucción son menos eficientes.

Desde el punto de vista del análisis, la *Categoría 3* proporciona la muestra más limpia y menos sensible a incertidumbres sistemáticas asociadas a la identificación leptónica, mientras que las *Categorías 1* y *2* resultan útiles como regiones de control para validar el modelado de los fondos y estudiar la estabilidad de las distribuciones de $m_{4\mu}$ frente a variaciones en los criterios de identificación. En conjunto, la clasificación por calidad de muones confirma que la señal potencial de $H \rightarrow ZZ^* \rightarrow 4\mu$ está dominada por eventos con muones de alta calidad, y que el costo en aceptación asociado a exigir identificadores más estrictos es moderado una vez aplicados los cortes cinemáticos y topológicos principales.

Ejemplo de aplicación de los filtros de identificación

La selección de eventos basada en la identificación de muones se realiza exigiendo que cada uno de los cuatro muones de un evento cumpla al menos un cierto nivel de calidad (**Loose**, **Medium** o **Tight**). Estas categorías son jerárquicas: todo muón **Tight** es también **Medium** y **Loose**, y todo **Medium** es también **Loose**, pero no necesariamente al revés.

A continuación se muestra una tabla ilustrativa con ejemplos concretos. En ella se indica si el evento pasaría el filtro **Loose** (todos los muones al menos **Loose**) o el filtro **Medium** (todos los muones al menos **Medium**):

Muón 1	Muón 2	Muón 3	Muón 4	¿Pasa el filtro?
Loose	Loose	Loose	Loose	Sí (Loose), No (Medium)
Medium	Medium	Medium	Medium	Sí (ambos filtros)
Tight	Medium	Loose	Tight	Sí (Loose), No (Medium)
Tight	Tight	Tight	Tight	Sí (ambos filtros)
Loose	Medium	Tight	Medium	Sí (Loose), No (Medium) ¹

Table 5.3: Ejemplos concretos de combinaciones de calidad de muones y su resultado bajo los filtros de selección.

Este ejemplo demuestra que la selección exige únicamente un mínimo de calidad por muón, permitiendo que en un mismo evento coexistan diferentes niveles de identificación. El filtro **Loose** acepta cualquier combinación igual o superior, mientras que el filtro **Medium** requiere que *todos* los muones sean al menos **Medium** (por lo tanto, un solo muón "Loose" impide que el evento sea seleccionado en ese caso).

Distribución de la masa invariante $m_{4\mu}$ por categorías de identificación leptónica

Con el fin de estudiar el impacto de la calidad de la identificación de muones sobre la selección de candidatos 4μ , se define un conjunto de tres categorías mutuamente excluyentes en función del nivel más exigente de identificación alcanzado por los leptones del evento. En términos cualitativos, la *Categoría 1* agrupa eventos en los que al menos uno de los muones sólo cumple los requisitos de identificación **Loose**; la *Categoría 2* contiene eventos en los que todos los muones cumplen al menos el criterio **Medium** pero no todos alcanzan el nivel **Tight**; finalmente, la *Categoría 3* corresponde a los eventos en los que los cuatro muones satisfacen los requisitos de identificación **Tight**. Las tres categorías se definen de forma jerárquica, de modo que cada evento se asigna únicamente a la categoría de mayor pureza que cumple.

La Figura 5.4 muestra la distribución de la masa invariante del sistema de cuatro muones, $m_{4\mu}$, separada por estas categorías. A pesar de este particionado, las tres distribuciones comparten la misma estructura global: un máximo a masas intermedias seguido de una caída hacia valores altos, con un realce visible en la vecindad de $m_{4\mu} \sim 125$ GeV, marcado mediante una línea vertical discontinua, consistente con la contribución del canal $H \rightarrow ZZ^* \rightarrow 4\mu$ superpuesta al continuo $ZZ^{(*)}$.

Un aspecto llamativo de la figura es que la *Categoría 3* (Tight) contiene el mayor número de eventos, seguida por la *Categoría 2* (Medium), mientras que la *Categoría 1* (Loose) es la menos poblada. Aunque a primera vista esto podría parecer contraintuitivo, el comportamiento es consecuencia directa de la definición jerárquica de las categorías y de la alta calidad de los muones presentes en esta topología. La selección cinemática impuesta ($p_T^\mu > 5$ GeV, $|\eta^\mu| < 2.4$ y requisitos estrictos en SIP3D e impacto) privilegia muones bien reconstruidos en la región central del detector, para los cuales la probabilidad de satisfacer el identificador *Tight* es elevada. En estas condiciones, la mayoría de los eventos que superan el preselección básica migran naturalmente a la categoría de mayor pureza, mientras que sólo un subconjunto relativamente pequeño presenta muones que fallan los criterios *Medium* o *Tight* y, por tanto, quedan clasificados como *Loose*.

La comparación entre las tres curvas ilustra el compromiso entre pureza y estadística inherente al uso de identificadores jerárquicos de muones. La *Categoría 3* proporciona la muestra más limpia, dominada por muones bien identificados y, por ende, menos sensible a incertidumbres sistemáticas asociadas a la identificación leptónica, a costa de una reducción moderada de la estadística total. Las *Categorías 1* y *2*, aunque menos puras, son útiles para validar el modelado de los fondos en regiones donde los muones no alcanzan el identificador *Tight* y para estimar el impacto de variaciones en los criterios de identificación sobre la forma del espectro de $m_{4\mu}$.

En conjunto, la consistencia de la forma de $m_{4\mu}$ entre las tres categorías y la mayor población de la muestra *Tight* apoyan la conclusión de que la señal potencial de $H \rightarrow ZZ^* \rightarrow 4\mu$ está dominada por eventos con muones de alta calidad, mientras que las categorías de menor pureza desempeñan principalmente un papel de control y validación del análisis.

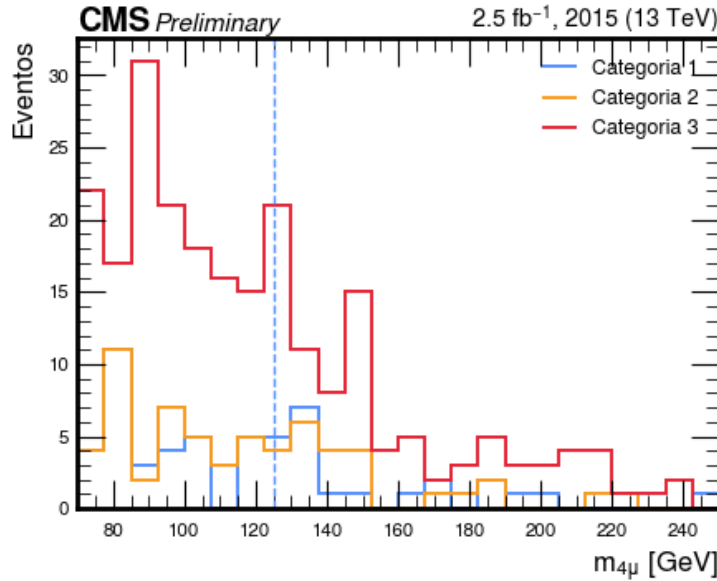


Figure 5.4: Distribución de la masa invariante del sistema de cuatro muones, $m_{4\mu}$, separada por categorías de identificación leptónica: *Categoría 1* (muones que sólo cumplen el criterio *Loose*), *Categoría 2* (al menos *Medium*) y *Categoría 3* (todos los muones *Tight*). La línea discontinua indica la masa del bosón de Higgs, $m_H \simeq 125$ GeV.

5.1.5 Aceptancia global de la selección de eventos

Antes de la aplicación de los cortes cinemáticos y de calidad descritos en las subsecciones anteriores, el conjunto de datos utilizado en este análisis contenía un total de $N_{\text{inicial}} = 560\,657$ eventos con todas las ramas centrales del árbol (*core branches*) correctamente definidas. Tras la aplicación

secuencial de los filtros de preselección a nivel de muón, la construcción de cuádrupletos, los requisitos de emparejamiento OS–OS, los cortes en momento transversal de los muones y de los candidatos Z , las ventanas de masa en m_{Z_a} y m_{Z_b} y la selección final en la masa invariante tetramuónica $m_{4\mu}$, el número de eventos que cumplen todos los requisitos es de $N_{\text{seleccionado}} = 423$. La aceptación global de la cadena de selección se define como

$$\mathcal{A} = \frac{N_{\text{seleccionado}}}{N_{\text{inicial}}} \times 100,$$

de modo que

$$\mathcal{A} = \frac{423}{560\,657} \times 100 \simeq 0.08\%.$$

Este valor cuantifica la fracción de eventos registrados con información completa que sobrevive a toda la cadena de cortes y es considerada compatible con la topología esperada para el canal $H \rightarrow ZZ^* \rightarrow 4\mu$. Aunque una aceptación global del orden de $10^{-1}\%$ pueda parecer extremadamente baja, es característica de análisis en física de partículas que estudian estados finales muy exclusivos, con secciones eficaces pequeñas y requisitos cinemáticos y de calidad muy estrictos. La mayoría de los eventos se descarta por no satisfacer simultáneamente las condiciones sobre los cuatro muones reconstruidos, la estructura de pares $Z_a Z_b$ y la masa invariante $m_{4\mu}$; esta reducción sustancial es indispensable para minimizar la contribución de fondo y maximizar la confiabilidad en la identificación de posibles candidatos al bosón de Higgs.

La Tabla 5.2 y la Figura 5.3 resumen de forma complementaria el impacto de cada etapa de selección sobre el tamaño de la muestra y sobre la aceptación (acumulada y por etapa) del análisis tetramuónico.

5.1.6 Comparación entre eventos iniciales y seleccionados

En la Figura 5.5 se ilustra de forma esquemática la relación entre el número total de eventos disponibles al inicio del análisis y el número de candidatos que sobreviven a toda la cadena de selección diseñada para el canal $H \rightarrow ZZ^* \rightarrow 4\mu$.

El punto de partida está dado por $N_{\text{inicial}} = 560\,657$ eventos con todas las ramas centrales del árbol (*core branches*) correctamente definidas. Sobre esta muestra se aplican los filtros cinemáticos y topológicos descritos en las subsecciones anteriores: preselección a nivel de muón, construcción de cuádrupletos, emparejamiento OS–OS, cortes en momento transversal de los muones y de los candidatos Z , ventanas de masa en m_{Z_a} y m_{Z_b} , y la restricción final en la masa invariante tetramuónica $m_{4\mu}$. Tras todos estos pasos, el número de eventos que cumplen simultáneamente todos los requisitos es de $N_{\text{seleccionado}} = 423$.

La aceptación global de la selección se define como

$$\mathcal{A} = \frac{N_{\text{seleccionado}}}{N_{\text{inicial}}} \times 100,$$

de donde se obtiene

$$\mathcal{A} = \frac{423}{560\,657} \times 100 \simeq 0.075\%.$$

Esta comparación pone de manifiesto la drástica reducción del conjunto de datos, característica de análisis en física de altas energías que buscan estados finales muy exclusivos y de baja sección eficaz. Aunque la aceptación porcentual es del orden de $10^{-1}\%$, el número absoluto de eventos retenidos es suficiente para estudiar la distribución de $m_{4\mu}$, la estructura de los candidatos Z_a y Z_b , y la posible presencia de una señal compatible con la desintegración $H \rightarrow ZZ^* \rightarrow 4\mu$. Al mismo tiempo, la fuerte supresión de eventos garantiza que la muestra final esté dominada por candidatos de alta calidad cinemática y de reconstrucción, reduciendo de manera significativa la contaminación de fondo.

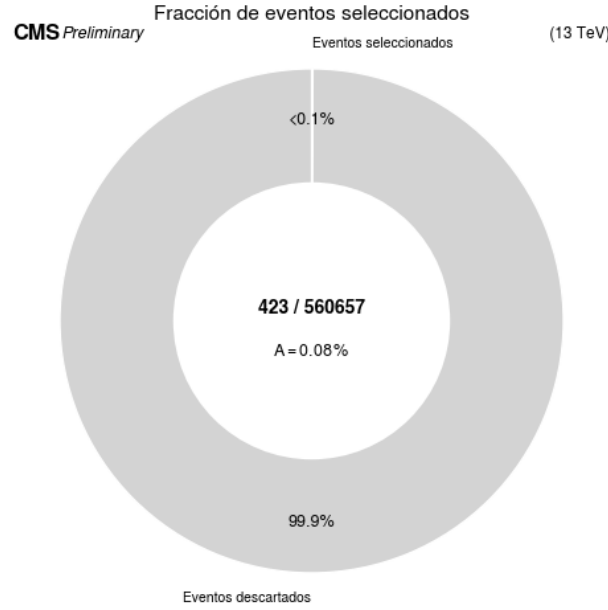


Figure 5.5: Comparación esquemática entre el número de eventos iniciales con ramas centrales válidas y el número de eventos seleccionados tras aplicar todos los criterios de selección para el canal $H \rightarrow ZZ^* \rightarrow 4\mu$.

5.1.7 Distribución del momento transversal p_T de los muones

La Figura 5.6 muestra la distribución del momento transversal (p_T) de todos los muones reconstruidos que forman parte de los eventos seleccionados. Para ello, se combinaron los valores de p_T de los cuatro muones de cada evento en un único conjunto de datos, permitiendo observar el comportamiento estadístico general de esta variable.

Se observa que la mayoría de los muones presentan valores bajos de p_T , con un pico muy pronunciado en la región de 0 a 10 GeV. A medida que el valor de p_T incrementa, la frecuencia de ocurrencia disminuye de forma exponencial, lo cual es característico de los procesos dominados por la producción de partículas a bajo impulso. Esta tendencia también refleja la naturaleza de la sección eficaz decreciente con el aumento del p_T , típico de los procesos de producción en colisiones protón-protón.

La elección de umbrales mínimos de p_T durante la selección de eventos —por ejemplo, exigir $p_T > 5$ GeV para muones individuales— permite eliminar parte del fondo irreducible sin afectar significativamente los eventos compatibles con el decaimiento de un bosón Z o del bosón de Higgs. Además, muones con mayor p_T son más fácilmente reconstruidos con alta precisión, lo que mejora la resolución en la medición de la masa invariante del sistema ZZ .

En conjunto, esta distribución es útil no solo para validar la consistencia de los datos reconstruidos con las expectativas teóricas, sino también para optimizar la selección de eventos mediante cortes cinemáticos que equilibren eficiencia y pureza en la muestra final.

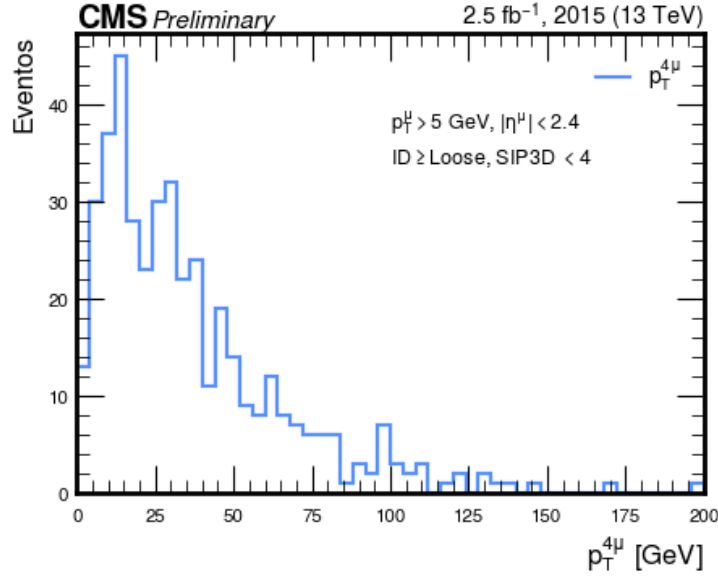


Figure 5.6: Distribución del momento transversal p_T de los muones reconstruidos en los eventos seleccionados.

5.1.8 Distribución de la separación angular ΔR entre pares de muones

La Figura 5.7 muestra la distribución de la separación angular ΔR entre muones reconstruidos en los eventos seleccionados. Esta variable, definida como:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

donde $\Delta\eta$ es la diferencia en pseudorapidez y $\Delta\phi$ la diferencia en ángulo azimutal entre dos muones, proporciona una medida del aislamiento angular entre partículas en el espacio $\eta - \phi$.

En esta figura se incluyen los valores de ΔR para todos los pares posibles entre los muones del evento, excluyendo los pares primarios usados para formar los candidatos a bosón Z . El pico principal se encuentra en torno a $\Delta R \approx 3$, lo que indica que muchos pares de muones se encuentran relativamente separados en el detector, característica esperada cuando los muones provienen de la desintegración de dos bosones Z distintos.

La forma de la distribución también muestra que, aunque existe una fracción menor de pares con valores bajos de ΔR , se aplicó un corte mínimo (por ejemplo, $\Delta R > 0.1$) para evitar pares de muones colineales que podrían derivar de conversiones o eventos mal reconstruidos. Esta selección mejora la resolución en masa invariante y reduce el impacto de eventos espurios o falsamente combinados.

Este histograma es esencial para evaluar la calidad geométrica de los eventos y la correcta reconstrucción de objetos bien separados, además de validar los cortes aplicados sobre la variable ΔR en el análisis.

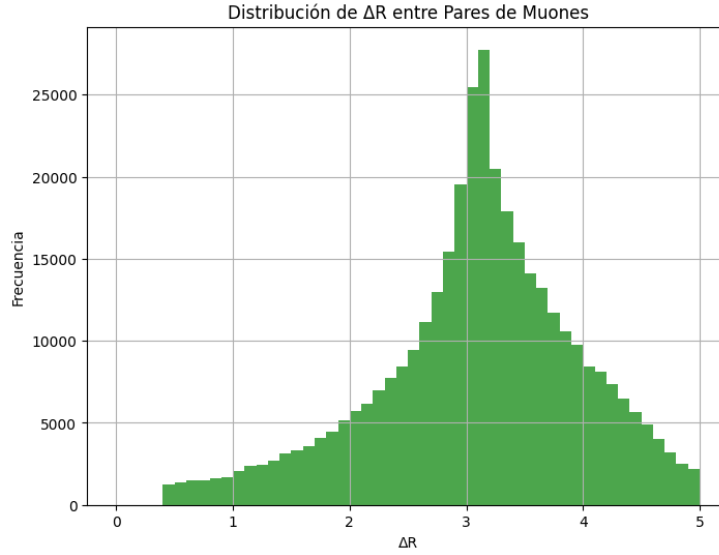


Figure 5.7: Distribución de la separación angular ΔR entre pares de muones reconstruidos en los eventos seleccionados.

Distribución de las masas de los candidatos Z_a y Z_b

Para cada evento tetramuónico se construyen todas las combinaciones posibles de pares de muones de carga opuesta y se selecciona la combinación que maximiza la compatibilidad con la topología $ZZ^{(*)}$. En este procedimiento se define como candidato Z_a al par de muones cuya masa invariante es más cercana a la masa del bosón Z medida por el PDG ($m_Z \simeq 91.2$ GeV), mientras que el candidato Z_b corresponde al segundo par de muones del evento, formado a partir de los leptones restantes. De este modo, Z_a representa típicamente el bosón Z on-shell y Z_b captura la contribución del bosón Z^* off-shell o de configuraciones de fondo donde el segundo par leptónico no proviene de un Z resonante.

La Figura 5.8 muestra la distribución de la masa invariante del candidato Z_a , m_{Z_a} , para los eventos que satisfacen los criterios de selección sobre los muones individuales ($p_T^\mu > 5$ GeV, $|\eta^\mu| < 2.4$, identificación al menos *Loose* y $\text{SIP3D} < 4$). Se observa un pico claro y relativamente estrecho centrado alrededor de $m_{Z_a} \sim 91$ GeV, marcado mediante una línea vertical discontinua en la figura. Esta acumulación indica que, en la gran mayoría de los eventos seleccionados, al menos uno de los pares de muones es compatible con la producción de un bosón Z on-shell. La posición del máximo coincide, dentro de las incertidumbres estadísticas, con la masa nominal del bosón Z , lo que valida el desempeño del algoritmo de emparejamiento utilizado para identificar el candidato Z_a y proporciona una comprobación interna de la calibración en momento de los muones en la región cinemática considerada.

La anchura efectiva del pico en m_{Z_a} está dominada por la resolución en momento de los muones y por la configuración geométrica de los pares, ya que el ancho natural del bosón Z es pequeño frente a la resolución instrumental. La cola hacia masas ligeramente inferiores y superiores a m_Z se debe tanto a efectos de resolución como a la presencia de un subconjunto de eventos mal emparejados, en los que la elección del par más cercano a la masa del Z no coincide con la combinación leptónica físicamente correcta. Un estudio complementario de la variable $|\Delta m_Z| = |m_Z^{\text{reco}} - m_Z^{\text{PDG}}|$ confirma que la mayor parte de los candidatos Z_a presentan desviaciones pequeñas, mientras que las colas están pobladas principalmente por eventos de fondo y emparejamientos ambiguos.

En contraste, la Figura 5.9 presenta la distribución de la masa invariante del candidato Z_b , m_{Z_b} . En este caso el espectro está fuertemente concentrado en la región de bajas masas, con un máximo

pronunciado por debajo de ~ 40 GeV y una cola que se extiende de manera suave hasta valores cercanos a 120 GeV. La ausencia de un pico marcado en torno a la masa del bosón Z y la predominancia de masas menores reflejan la naturaleza off-shell del segundo bosón intermedio en el canal $H \rightarrow ZZ^* \rightarrow 4\mu$, así como la contribución de procesos de fondo donde el segundo par de muones no proviene de un Z resonante. Este comportamiento es característico de configuraciones ZZ^* y $Z\gamma^*$, en las que uno de los pares leptónicos está fuertemente suprimido por la cinemática y por las funciones de estructura de los protones, dando lugar a masas invariantes alejadas de la resonancia.

La comparación conjunta de las distribuciones de m_{Z_a} y m_{Z_b} ofrece una validación importante de la reconstrucción de la topología $ZZ^{(*)}$. El hecho de que m_{Z_a} se agrupe alrededor de la masa del bosón Z mientras que m_{Z_b} se concentra en la región de bajas masas es coherente con la expectativa para desintegraciones $H \rightarrow ZZ^* \rightarrow 4\mu$ y para los procesos de fondo irreducibles dominantes. Adicionalmente, el mapa bidimensional m_{Z_a} vs. m_{Z_b} (Sección ??) muestra que las configuraciones pobladas en datos se concentran en la banda donde uno de los candidatos se sitúa cerca de m_Z y el otro toma valores inferiores, reforzando la interpretación en términos de un bosón Z on-shell acompañado de un segundo bosón off-shell o virtual.

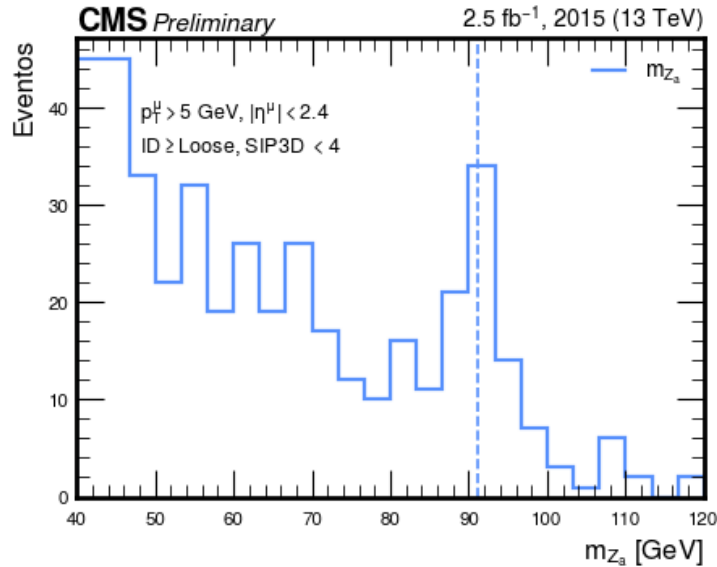


Figure 5.8: Distribución de la masa invariante del candidato Z_a , m_{Z_a} , definido como el par de muones cuya masa es más cercana a la masa nominal del bosón Z . La línea discontinua indica $m_Z \simeq 91.2$ GeV.

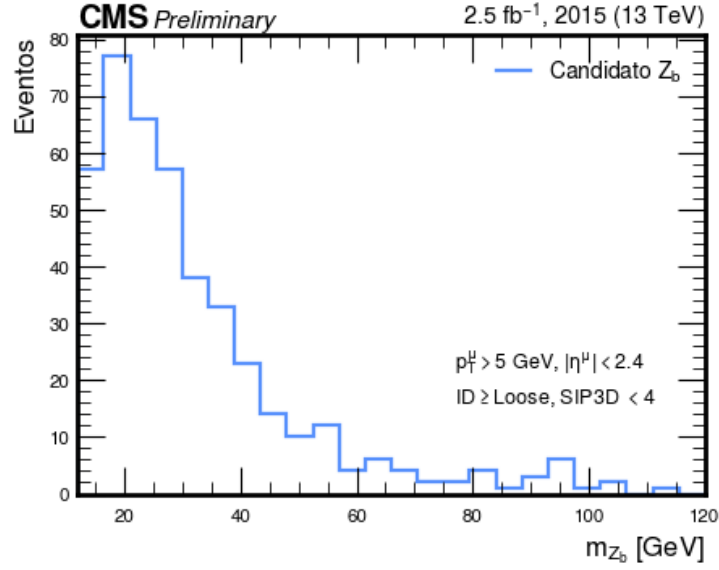


Figure 5.9: Distribución de la masa invariante del candidato Z_b , m_{Z_b} , construido a partir del segundo par de muones en cada evento. La acumulación en la región de bajas masas es característica de un bosón Z^* off-shell y de contribuciones de fondo no resonantes.

Chapter 6

Discussion y Conclusiones

6.1 Discusión

Los resultados obtenidos en este trabajo muestran que es posible reproducir, a partir de datos abiertos de CMS, un análisis coherente del canal tetramuónico $H \rightarrow ZZ^* \rightarrow 4\mu$ siguiendo criterios de selección inspirados en los análisis oficiales de la colaboración. A pesar de las limitaciones inherentes al uso de datos públicos, se han logrado varios objetivos metodológicos y físicos relevantes. En primer lugar, la reconstrucción de la masa invariante del sistema de cuatro muones $m_{4\mu}$ exhibe estructuras claras en torno a la masa del bosón Z y del bosón de Higgs, con un realce visible en la región $m_{4\mu} \sim 125$ GeV compatible con la topología esperada del canal $H \rightarrow ZZ^* \rightarrow 4\mu$. En segundo lugar, las distribuciones de las masas de los candidatos Z_a y Z_b muestran el comportamiento característico de un bosón Z on-shell acompañado de un segundo bosón off-shell, lo cual valida el algoritmo de emparejamiento y la definición de las ventanas de masa empleadas.

Asimismo, el flujo de cortes y la aceptación asociada a cada etapa confirman que la cadena de selección es capaz de aislar una muestra pequeña pero limpia de candidatos tetramuónicos. La caída drástica en el número de eventos al construir cuádrupletos, imponer el emparejamiento OS-OS y aplicar los cortes en p_T y en masa es coherente con el carácter altamente exclusivo del estado final 4μ . La consistencia de las distribuciones cinemáticas globales del sistema ($m_{4\mu}$, $p_T^{4\mu}$, $\eta_{4\mu}$) y de las variables asociadas a los bosones intermedios (m_{Z_a} , m_{Z_b}) con lo reportado en la literatura respalda la solidez metodológica del enfoque adoptado.

Otro aspecto positivo del análisis es el comportamiento de las categorías definidas en función de la calidad de identificación de los muones. A pesar de que la selección inicial sólo requiere criterios al menos *Loose*, la mayoría de los eventos que sobreviven a la cadena completa de cortes pertenece a la categoría en la que los cuatro muones cumplen el identificador *Tight*. Esto indica que los cortes cinemáticos y de impacto actúan de manera efectiva como un pre-filtro de calidad, favoreciendo leptones bien reconstruidos y reduciendo de forma indirecta la sensibilidad a sistemáticos asociados a la identificación. La clasificación por categorías ofrece, además, una herramienta útil para definir regiones de control y para explorar la estabilidad de las distribuciones de $m_{4\mu}$ frente a variaciones en los criterios de identificación leptónica.

En este contexto, resulta natural comparar estos resultados con los obtenidos por la colaboración CMS en análisis oficiales del mismo canal. En particular, la Conferencia de Moriond 2017 marcó un hito en el estudio de $H \rightarrow ZZ^* \rightarrow 4\ell$ durante el Run 2, al presentar mediciones de precisión basadas en la luminosidad completa de 2016, $\sim 36 \text{ fb}^{-1}$, frente a los $\sim 2.5 \text{ fb}^{-1}$ utilizados en este trabajo con *Open Data*. Ambos análisis comparten el mismo canal de desintegración y la misma energía de colisión ($\sqrt{s} = 13 \text{ TeV}$), pero difieren de manera significativa en su alcance y objetivos:

Table 6.1: Comparación esquemática entre el análisis realizado en esta tesis y los resultados oficiales de CMS presentados en Moriond 2017 para el canal $H \rightarrow ZZ^* \rightarrow 4\ell$.

Característica	Este trabajo	CMS Moriond 2017
Fuente de datos	Datos abiertos de CMS (Run 2015D).	Datos internos de la colaboración (full 2016 luminosity).
Luminosidad integrada	$\sim 2.5 \text{ fb}^{-1}$.	$\sim 36 \text{ fb}^{-1}$.
Canal analizado	$H \rightarrow ZZ^* \rightarrow 4\mu$.	$H \rightarrow ZZ^* \rightarrow 4\ell$ ($\ell = e, \mu$).
Objetivo principal	Validación metodológica y demostración de la viabilidad del canal $H \rightarrow ZZ^* \rightarrow 4\mu$ en datos abiertos.	Medición de precisión de la masa del Higgs y de las secciones eficaces totales y diferenciales.
Resultado sobre m_H	Realce compatible con la señal en $m_{4\mu} \sim 125 \text{ GeV}$ (indicación cualitativa).	Medición cuantitativa de alta precisión: $m_H = 125.26 \pm 0.20$ (stat.) ± 0.08 (syst.) GeV [48, 8].

La distribución de la masa invariante del sistema de cuatro leptones ($m_{4\ell}$ o $m_{4\mu}$) es el observable central en ambos casos. En la Figura 5.1 de esta tesis se observa un máximo pronunciado en la vecindad de $m_{4\mu} \sim 90\text{--}95 \text{ GeV}$, asociado al bosón Z on-shell, y un realce visible en torno a $m_{4\mu} \sim 125 \text{ GeV}$, compatible cualitativamente con la contribución esperada de la desintegración $H \rightarrow ZZ^* \rightarrow 4\mu$. En cambio, la gráfica oficial presentada por CMS en Moriond 2017, gracias a la estadística muy superior, muestra un pico de señal $H \rightarrow 4\ell$ mucho más estrecho y estadísticamente significativo, lo que permite extraer una medición precisa de la masa del Higgs en este canal, con una incertidumbre estadística y sistemática cuidadosamente evaluadas.

En este sentido, el realce observado en $m_{4\mu} \sim 125 \text{ GeV}$ en los datos abiertos debe interpretarse como una *indicación cualitativa* de compatibilidad con la presencia de la señal $H \rightarrow ZZ^* \rightarrow 4\mu$, sustentada en la forma del espectro y en la coherencia de las distribuciones cinemáticas con las expectativas teóricas. Por el contrario, la determinación *cuantitativa* de la masa del bosón de Higgs y de la fuerza de señal corresponde a los análisis oficiales de CMS, que disponen de una luminosidad integrada casi un orden de magnitud mayor y de una caracterización detallada de las incertidumbres sistemáticas. La comparación entre ambas situaciones puede entenderse, de manera ilustrativa, como el contraste entre un “borrador” de alta calidad que reproduce los elementos esenciales de la pintura original (este trabajo en datos abiertos) y la obra maestra final (el resultado de CMS en Moriond 2017), en la que cada detalle ha sido cuantificado y perfeccionado gracias a la disponibilidad de más datos y recursos experimentales.

A pesar de estos resultados y de la estructura rigurosa del análisis, es importante reconocer ciertas limitaciones inherentes a este estudio. En primer lugar, el uso de datos abiertos de CMS implica que no todos los detalles de la calibración, corrección y reconstrucción aplicados por la colaboración original están plenamente accesibles o reproducidos. Esto puede afectar la comparación directa con resultados oficiales e introducir sistemáticos no cuantificados, especialmente en la reconstrucción de muones y en la selección de eventos. La ausencia de una evaluación explícita de estas incertidumbres limita la posibilidad de realizar afirmaciones cuantitativas de alta precisión.

Otra limitación relevante es la falta de una estimación detallada de los fondos irreducibles y reducibles. En este trabajo se ha priorizado la selección y caracterización de la señal tetramuónica, pero no se ha llevado a cabo una modelización exhaustiva del fondo mediante simulaciones de Monte Carlo o mediante técnicas de estimación basadas en datos (por ejemplo, regiones de control o métodos de inversión de cortes). Esto impide la extracción de una significancia estadística absoluta o la determinación precisa de la contribución de fondo en la región de interés. Además, no se incluyeron incertidumbres experimentales ni teóricas asociadas a las secciones eficaces, funciones de forma, calibraciones o eficiencias de reconstrucción, las cuales podrían influir tanto en la aceptación como en la interpretación de la estructura observada cerca de la masa del Higgs.

El análisis podría mejorarse en futuras iteraciones incorporando técnicas más avanzadas de reconstrucción y selección, como el uso de algoritmos multivariantes (por ejemplo, *boosted decision trees* o redes neuronales), una estimación explícita de los fondos mediante las muestras de simulación disponibles en los datos abiertos o mediante regiones de control en datos, así como la inclusión de variables adicionales, como distribuciones angulares, observables sensibles al espín y al CP del bosón de Higgs, o medidas más refinadas de aislamiento leptónico. También sería valioso comparar de forma sistemática las distribuciones obtenidas con predicciones de señal y fondo, validando así las eficiencias de selección y el modelado cinemático, así como extender el estudio a otros canales de desintegración (por ejemplo, canales con electrones o estados mixtos de leptones).

En conjunto, la robustez de los resultados presentados se apoya en la aplicación sistemática de criterios de selección inspirados en la literatura y en los análisis oficiales de CMS, y en la consistencia de las distribuciones reconstruidas con las expectativas físicas para el canal $H \rightarrow ZZ^* \rightarrow 4\mu$. La tendencia observada en la región de masa del Higgs y la reducción progresiva de los fondos mediante cortes de calidad demuestran la solidez metodológica del enfoque. No obstante, la ausencia de una estimación precisa del fondo y de una cuantificación detallada de los sistemáticos limita la posibilidad de realizar afirmaciones cuantitativas sobre la significancia de la señal. En este sentido, el trabajo pone de manifiesto tanto el potencial educativo y exploratorio del uso de datos abiertos como la necesidad de complementar este tipo de análisis con herramientas y entradas adicionales para alcanzar el nivel de rigor característico de los estudios oficiales de física de colisionadores.

Con el fin de ilustrar visualmente la comparación discutida anteriormente, en la Figura 6.1 se presenta, a modo de ejemplo, la distribución de $m_{4\mu}$ obtenida en este trabajo junto a la distribución oficial de $m_{4\ell}$ reportada por CMS en Moriond 2017. Esta comparación lado a lado refuerza la idea de que, aun con una luminosidad significativamente menor, los datos abiertos permiten reconstruir la estructura global del espectro y localizar un exceso de eventos en la región de masa del Higgs, mientras que el análisis oficial aprovecha la estadística completa para alcanzar la precisión necesaria en la medición de m_H .

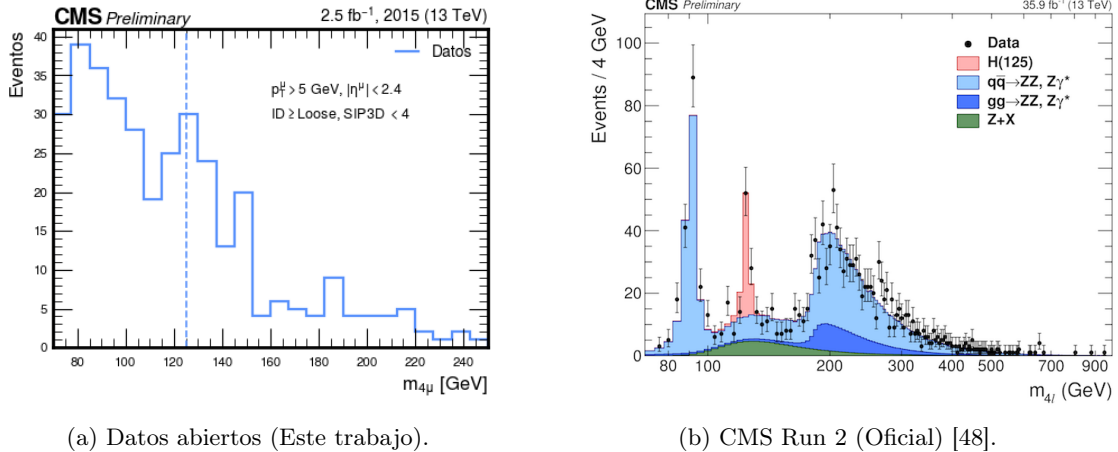


Figure 6.1: Comparación visual de la reconstrucción del pico de masa. A la izquierda, la distribución $m_{4\mu}$ obtenida en este análisis; a la derecha, el resultado oficial de CMS. Se ajustó la escala visual para facilitar la comparación de la estructura del pico en 125 GeV.

6.2 Conclusiones

Este trabajo ha demostrado la viabilidad de realizar análisis de física de altas energías de nivel profesional utilizando exclusivamente datos abiertos del experimento CMS. Se logró implementar exitosamente una cadena de procesamiento completa para el canal $H \rightarrow ZZ^* \rightarrow 4\mu$, mostrando

que los recursos públicos preservan la información física necesaria para llevar a cabo ejercicios de redescubrimiento de resultados fundamentales del Modelo Estándar.

La aplicación rigurosa de los criterios de selección sobre un conjunto inicial de $N_{\text{inicial}} = 560\,657$ eventos permitió aislar una muestra final de alta pureza constituida por $N_{\text{seleccionado}} = 423$ candidatos. Esta reducción de tres órdenes de magnitud (aceptancia global $\sim 0.08\%$) confirma la eficacia de los cortes cinemáticos y de identificación para suprimir el fondo hadrónico dominante en estados finales exclusivos. De forma crucial, la muestra seleccionada exhibe una estructura resonante en la masa invariante del sistema tetramuónico alrededor de 125 GeV, cinemáticamente compatible con la señal del bosón de Higgs, lo que respalda la robustez del algoritmo de reconstrucción implementado.

Metodológicamente, se evidenció que la jerarquía de calidad en la identificación de muones es un discriminante potente: la supervivencia mayoritaria de eventos en la categoría *Tight* tras la selección cinemática pone de manifiesto la correlación entre la topología del evento y la calidad de los objetos reconstruidos. Esto establece un precedente valioso para futuros análisis educativos, demostrando que es posible definir regiones de control y señal fiables sin acceso a la infraestructura privada de la colaboración.

Si bien la ausencia de simulaciones de Monte Carlo dedicadas al fondo limita la capacidad de cuantificar la significancia estadística exacta del exceso observado, los resultados obtenidos son consistentes con las expectativas fenomenológicas del proceso $H \rightarrow 4\ell$. En este sentido, el estudio no pretende extraer una medición de precisión, sino demostrar la viabilidad de reconstruir y caracterizar la cinemática del bosón de Higgs en un entorno accesible, proporcionando un marco de referencia reproducible y escalable. Futuras iteraciones que incorporen estimaciones de fondo basadas en datos y técnicas multivariantes podrán cerrar la brecha restante hacia la sensibilidad y el rigor característicos de los análisis oficiales.

6.3 Recomendaciones y perspectivas a futuro

A partir de la experiencia y los resultados obtenidos en este trabajo, se plantean las siguientes recomendaciones y líneas de acción para futuros estudios basados en el canal $H \rightarrow ZZ^* \rightarrow 4\mu$ con datos abiertos de CMS:

- **Incorporar estimaciones explícitas de fondo.** Resulta fundamental complementar la caracterización de la señal tetramuónica con una evaluación cuantitativa de los procesos de fondo irreducibles y reducibles. Para ello, puede recurrirse tanto a las muestras de simulación Monte Carlo disponibles en los datos abiertos como a técnicas de estimación basadas en datos (por ejemplo, el uso de regiones de control, inversión de cortes o métodos de normalización en bandas laterales).
- **Aplicar técnicas de análisis multivariante.** La implementación de algoritmos avanzados de clasificación, como *boosted decision trees* o redes neuronales, permitiría explotar de forma más eficiente la información contenida en las variables cinemáticas ($m_{4\mu}$, m_{Z_a} , m_{Z_b} , $p_T^{4\mu}$, distribuciones angulares, etc.) y mejorar la capacidad de discriminación entre señal y fondo, incrementando la sensibilidad del análisis.
- **Explorar otros canales y variables adicionales.** Ampliar el estudio a canales leptónicos alternativos (como $2e2\mu$ o $4e$), así como considerar observables más específicos —por ejemplo, ángulos de decaimiento en el sistema $ZZ^{(*)}$, variables sensibles a la estructura de espín y CP del bosón de Higgs o definiciones refinadas de aislamiento leptónico— permitiría obtener una visión más completa de las propiedades del bosón de Higgs y de los procesos asociados.
- **Profundizar en el estudio de sistemáticos.** Incluir la evaluación de incertidumbres sistemáticas asociadas a la reconstrucción, identificación y calibración de los muones, así como

a las luminosidades y modelos teóricos empleados, fortalecería la robustez e interpretabilidad de los resultados. Esto es especialmente importante si se busca comparar de forma cuantitativa con mediciones oficiales o extraer significancias estadísticas.

- **Fomentar el uso de datos abiertos en docencia e investigación.** La experiencia adquirida en este análisis confirma el potencial de los datos públicos del LHC como herramienta didáctica y de formación para nuevas generaciones de físicos y científicos de datos. El diseño de actividades, proyectos de tesis y cursos basados en estos recursos puede promover el acceso abierto, la reproducibilidad y la familiarización temprana con las metodologías empleadas en la física experimental de altas energías.

La integración de estos enfoques contribuirá a incrementar la precisión y el alcance de futuros análisis, acercando los resultados obtenidos con datos abiertos al rigor y la profundidad de los estudios realizados dentro de las grandes colaboraciones internacionales en física de altas energías, y consolidando al mismo tiempo el valor formativo de este tipo de ejercicios en contextos académicos.

Bibliography

- [1] Bora Akgun. “An Educational Higgs Study with CMS Open Data”. In: *Adıyaman University Journal of Science* 11.2 (2021), pp. 431–443. DOI: [10.37094/adyujsci.1003333](https://doi.org/10.37094/adyujsci.1003333).
- [2] A. Apyan et al. “Opportunities and challenges of Standard Model production cross section measurements in proton-proton collisions at $\sqrt{s} = 8$ TeV using CMS Open Data”. In: *JINST* 15 (2020), P01009. DOI: [10.1088/1748-0221/15/01/P01009](https://doi.org/10.1088/1748-0221/15/01/P01009).
- [3] ATLAS Collaboration. “Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector”. In: *Eur. Phys. J. C* 75 (2015), p. 476. DOI: [10.1140/epjc/s10052-015-3685-1](https://doi.org/10.1140/epjc/s10052-015-3685-1). arXiv: [1506.05669](https://arxiv.org/abs/1506.05669).
- [4] G. C. Branco et al. “Theory and phenomenology of two-Higgs-doublet models”. In: *Phys. Rept.* 516 (2012), pp. 1–102. DOI: [10.1016/j.physrep.2012.02.002](https://doi.org/10.1016/j.physrep.2012.02.002).
- [5] Rene Brun and Fons Rademakers. “ROOT: An Object Oriented Data Analysis Framework”. In: *Nucl. Instrum. Meth. A* 389 (1997), pp. 81–86. DOI: [10.1016/S0168-9002\(97\)00048-X](https://doi.org/10.1016/S0168-9002(97)00048-X).
- [6] CMS Collaboration. “Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at $\sqrt{s} = 7$ and 8 TeV”. In: *Phys. Rev. D* 92 (2015), p. 012004. DOI: [10.1103/PhysRevD.92.012004](https://doi.org/10.1103/PhysRevD.92.012004). arXiv: [1411.3441](https://arxiv.org/abs/1411.3441).
- [7] CMS Collaboration. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *JINST* 9 (2014), P10009. DOI: [10.1088/1748-0221/9/10/P10009](https://doi.org/10.1088/1748-0221/9/10/P10009).
- [8] CMS Collaboration. “Measurements of production cross sections of the Higgs boson in the four-lepton final state in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 81 (2021), p. 488. DOI: [10.1140/epjc/s10052-021-09200-x](https://doi.org/10.1140/epjc/s10052-021-09200-x). arXiv: [2103.04956](https://arxiv.org/abs/2103.04956).
- [9] CMS Collaboration. “Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state”. In: *Phys. Rev. D* 99 (2019), p. 112003. DOI: [10.1103/PhysRevD.99.112003](https://doi.org/10.1103/PhysRevD.99.112003). arXiv: [1901.00174](https://arxiv.org/abs/1901.00174).
- [10] CMS Collaboration. *Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV*. Tech. rep. CMS-PAS-MUO-10-004. CMS Physics Analysis Summary. Geneva: CERN, 2012. URL: <https://cds.cern.ch/record/1370586>.
- [11] CMS Collaboration. “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV”. In: *JINST* 10 (2015), P06005. DOI: [10.1088/1748-0221/10/06/P06005](https://doi.org/10.1088/1748-0221/10/06/P06005).
- [12] CMS Collaboration. “Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13$ TeV in 2015 and 2016 at CMS”. In: *Eur. Phys. J. C* 81 (2021), p. 800. DOI: [10.1140/epjc/s10052-021-09538-2](https://doi.org/10.1140/epjc/s10052-021-09538-2).
- [13] ATLAS Collaboration. “Measurement of inclusive and differential cross sections in the $H \rightarrow ZZ^* \rightarrow 4\ell$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *JHEP* 10 (2017), p. 132. DOI: [10.1007/JHEP10\(2017\)132](https://doi.org/10.1007/JHEP10(2017)132).

- [14] ATLAS Collaboration. “Measurements of the Higgs boson inclusive and differential fiducial cross sections in the 4ℓ decay channel at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 80 (2020), p. 941. DOI: [10.1140/epjc/s10052-020-8223-0](https://doi.org/10.1140/epjc/s10052-020-8223-0).
- [15] ATLAS Collaboration. “Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment”. In: *Eur. Phys. J. C* 76 (2016), pp. 6, 6. DOI: [10.1140/epjc/s10052-015-3769-y](https://doi.org/10.1140/epjc/s10052-015-3769-y).
- [16] ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: [10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020).
- [17] ATLAS Collaboration and CMS Collaboration. “Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments”. In: *Phys. Rev. Lett.* 114 (2015), p. 191803. DOI: [10.1103/PhysRevLett.114.191803](https://doi.org/10.1103/PhysRevLett.114.191803).
- [18] CMS Collaboration. “Combined results of searches for the standard model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV”. In: *Phys. Lett. B* 710 (2012), pp. 26–48. DOI: [10.1016/j.physletb.2012.02.064](https://doi.org/10.1016/j.physletb.2012.02.064). arXiv: [1202.1488](https://arxiv.org/abs/1202.1488).
- [19] CMS Collaboration. “Measurement of the properties of a Higgs boson in the four-lepton final state”. In: *Phys. Rev. D* 89 (2014), p. 092007. DOI: [10.1103/PhysRevD.89.092007](https://doi.org/10.1103/PhysRevD.89.092007).
- [20] CMS Collaboration. “Measurements of Higgs boson production cross sections and couplings in the four-lepton channel in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Rev. D* 102 (2020), p. 072006. DOI: [10.1103/PhysRevD.102.072006](https://doi.org/10.1103/PhysRevD.102.072006).
- [21] CMS Collaboration. “Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13$ TeV”. In: *JHEP* 11 (2017), p. 047. DOI: [10.1007/JHEP11\(2017\)047](https://doi.org/10.1007/JHEP11(2017)047).
- [22] CMS Collaboration. “Measurements of the $pp \rightarrow ZZ$ production cross section and the $Z \rightarrow 4\ell$ branching fraction, and constraints on anomalous triple gauge couplings at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 78.6 (2018), p. 515. DOI: [10.1140/epjc/s10052-018-5769-1](https://doi.org/10.1140/epjc/s10052-018-5769-1).
- [23] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021).
- [24] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Phys. Lett. B* 716 (2012), pp. 30–61. DOI: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021). arXiv: [1207.7235](https://arxiv.org/abs/1207.7235).
- [25] CMS Collaboration. “On the mass and spin-parity of the Higgs boson candidate via its decays to Z boson pairs”. In: *Phys. Rev. Lett.* 110 (2013), p. 081803. DOI: [10.1103/PhysRevLett.110.081803](https://doi.org/10.1103/PhysRevLett.110.081803). arXiv: [1212.6639](https://arxiv.org/abs/1212.6639).
- [26] CMS Collaboration. “Open data from the CMS experiment at the Large Hadron Collider”. In: *Journal of High Energy Physics* 2019.7 (2019), pp. 1–19. DOI: [10.1007/JHEP07\(2019\)115](https://doi.org/10.1007/JHEP07(2019)115). URL: [https://doi.org/10.1007/JHEP07\(2019\)115](https://doi.org/10.1007/JHEP07(2019)115).
- [27] CMS Collaboration. “Particle-flow reconstruction and global event description with the CMS detector”. In: *JINST* 12 (2017), P10003. DOI: [10.1088/1748-0221/12/10/P10003](https://doi.org/10.1088/1748-0221/12/10/P10003).
- [28] CMS Collaboration. “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV”. In: *JINST* 7 (2012), P10002. DOI: [10.1088/1748-0221/7/10/P10002](https://doi.org/10.1088/1748-0221/7/10/P10002).
- [29] CMS Collaboration. “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *JINST* 13 (2018), P06015. DOI: [10.1088/1748-0221/13/06/P06015](https://doi.org/10.1088/1748-0221/13/06/P06015).
- [30] CMS Collaboration. “Performance of the CMS muon trigger system in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *JINST* 16 (2021), P07001. DOI: [10.1088/1748-0221/16/07/P07001](https://doi.org/10.1088/1748-0221/16/07/P07001).

- [31] CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *JINST* 3 (2008), S08004. DOI: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [32] G. Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *Eur. Phys. J. C* 71 (2011), p. 1554. DOI: [10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0). arXiv: [1007.1727](https://arxiv.org/abs/1007.1727).
- [33] Abdelhak Djouadi. “The Anatomy of Electro-Weak Symmetry Breaking. I: The Higgs boson in the Standard Model”. In: *Phys. Rept.* 457 (2008), pp. 1–216. DOI: [10.1016/j.physrep.2007.10.004](https://doi.org/10.1016/j.physrep.2007.10.004).
- [34] Abdelhak Djouadi. “The Anatomy of Electro-Weak Symmetry Breaking. II: The Higgs bosons in the Minimal Supersymmetric Model”. In: *Phys. Rept.* 459 (2008), pp. 1–241. DOI: [10.1016/j.physrep.2007.10.005](https://doi.org/10.1016/j.physrep.2007.10.005).
- [35] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Physical Review Letters* 13 (1964), pp. 321–323. DOI: [10.1103/PhysRevLett.13.321](https://doi.org/10.1103/PhysRevLett.13.321).
- [36] L. Evans and P. Bryant. “LHC Machine”. In: *JINST* 3 (2008), S08001. DOI: [10.1088/1748-0221/3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001).
- [37] John F. Gunion et al. *The Higgs Hunter’s Guide*. Originally published by Addison-Wesley, 1990. Boulder, Colorado: Westview Press, 2000.
- [38] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. “Global Conservation Laws and Massless Particles”. In: *Phys. Rev. Lett.* 13 (1964), pp. 585–587. DOI: [10.1103/PhysRevLett.13.585](https://doi.org/10.1103/PhysRevLett.13.585).
- [39] Charles R. Harris et al. “Array programming with NumPy”. In: *Nature* 585 (2020), pp. 357–362. DOI: [10.1038/s41586-020-2649-2](https://doi.org/10.1038/s41586-020-2649-2).
- [40] Peter W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Physical Review Letters* 13 (1964), pp. 508–509. DOI: [10.1103/PhysRevLett.13.508](https://doi.org/10.1103/PhysRevLett.13.508).
- [41] John D. Hunter. “Matplotlib: A 2D Graphics Environment”. In: *Computing in Science & Engineering*. Vol. 9. 3. IEEE, 2007, pp. 90–95.
- [42] T. W. B. Kibble. “Symmetry Breaking in Non-Abelian Gauge Theories”. In: *Phys. Rev.* 155 (1967), pp. 1554–1561. DOI: [10.1103/PhysRev.155.1554](https://doi.org/10.1103/PhysRev.155.1554).
- [43] Thomas Kluyver et al. “Jupyter Notebooks – a publishing format for reproducible computational workflows”. In: *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. Ed. by Fernando Loizides and Birgit Schmidt. Amsterdam: IOS Press, 2016, pp. 87–90.
- [44] P. T. Komiske et al. “Exploring the Space of Jets with CMS Open Data”. In: *Phys. Rev. D* 101 (2020), p. 034009. DOI: [10.1103/PhysRevD.101.034009](https://doi.org/10.1103/PhysRevD.101.034009).
- [45] Wes McKinney. “Data Structures for Statistical Computing in Python”. In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt and Jarrod Millman. 2010, pp. 51–56.
- [46] Giovanni Petrucciani. *Evidence for ttH production with 13 TeV data?* Tech. rep. CMS-CR-2017-144. Proceedings of the 52nd Rencontres de Moriond on Electroweak Interactions and Unified Theories, La Thuile, Italy, 18–25 Mar 2017, pp. 3–8. Geneva: CERN, May 2017. URL: <https://cds.cern.ch/record/2267575>.
- [47] A. L. Read. “Presentation of search results: the CL_s technique”. In: *J. Phys. G* 28 (2002), pp. 2693–2704. DOI: [10.1088/0954-3899/28/10/313](https://doi.org/10.1088/0954-3899/28/10/313).
- [48] A. M. Sirunyan et al. “Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s} = 13$ TeV”. In: *JHEP* 11 (2017), p. 047. DOI: [10.1007/JHEP11\(2017\)047](https://doi.org/10.1007/JHEP11(2017)047). arXiv: [1706.09936](https://arxiv.org/abs/1706.09936) [hep-ex].
- [49] Pauli Virtanen et al. “SciPy 1.0: fundamental algorithms for scientific computing in Python”. In: *Nat. Methods* 17 (2020), pp. 261–272. DOI: [10.1038/s41592-019-0686-2](https://doi.org/10.1038/s41592-019-0686-2).

- [50] R. L. Workman et al. “Review of Particle Physics”. In: *Prog. Theor. Exp. Phys.* 2022 (2022), p. 083C01. DOI: [10.1093/ptep/ptac097](https://doi.org/10.1093/ptep/ptac097).
- [51] Stefan Wunsch. *Analysis of Higgs boson decays to four leptons using data and simulation of events at the CMS detector from 2012 using ROOT's RDataFrame*. CERN Open Data Portal. 2021. DOI: [10.7483/OPENDATA.CMS.F7HD.P3K4](https://doi.org/10.7483/OPENDATA.CMS.F7HD.P3K4). URL: <https://opendata.cern.ch/record/12360>.