

*Facultad
de
Ciencias*

**Caracterización de fondos en la
desintegración del bosón de Higgs**
(Characterization of backgrounds
in Higgs boson decay)

Trabajo de Fin de Grado
para acceder al

GRADO EN FÍSICA

Autora: Andrea Sainz Bear

Director: Jónatan Piedra Gómez

Co-Director: Jesús Manuel Vizán García

Septiembre - 2024

Resumen

En este trabajo se ha realizado un estudio basado en simulaciones de Monte Carlo de colisiones protón-protón a la energía de centro de masas de 13 TeV, correspondientes a la toma de datos de CMS del año 2018. Se han considerado como señal (S) del análisis tanto la producción del bosón de Higgs como la producción de dos bosones W, considerándose el resto de procesos físicos como fondo o contaminación (B). Son múltiples los procesos físicos que contaminan este análisis, sucesos que en su mayoría se pueden estimar por simulaciones de Monte Carlo. Un proceso diferente, estudiado con detalle en este trabajo, es el fondo instrumental llamado *Fake*, formado por una identificación errónea de electrones o muones. El objetivo principal de este estudio es maximizar la figura de mérito $S/\sqrt{S+B}$, inicialmente en un valor de 59.5. Este análisis ha conseguido aumentar la figura de mérito hasta un valor de 61.5, reduciendo además el fondo *Fake* en un 43 % y manteniendo el 95 % de la señal (Higgs y WW).

Palabras clave: bosón de Higgs, fondo instrumental, figura de mérito, Monte Carlo.

Abstract

This paper presents a study based on Monte Carlo simulations of proton-proton collisions at a centre-of-mass energy of 13 TeV, corresponding to the CMS data acquisition of the year 2018. Both the production of the Higgs boson and the production of two W bosons have been considered as the signal (S) of the analysis, and the rest of the physical processes are considered as background or contamination (B). There are many physical processes that contaminate this analysis, most of which can be estimated by Monte Carlo simulations. Another process studied in detail in this work is the so-called instrumental *Fake* background, which is formed by a misidentification of electrons or muons. The main objective of this study was to maximise the figure of merit $S/\sqrt{S+B}$, initially at a value of 59.5. This analysis has succeeded in increasing the figure of merit to a value of 61.5, while reducing the background *Fake* by 43 % and maintaining 95 % of the signal (Higgs and WW).

Keywords: Higgs boson, instrumental background, figure of merit, Monte Carlo.

Índice

1	Introducción	8
1.1	Motivación del trabajo	8
1.2	Objetivos	9
1.3	Estructura	9
2	Marco teórico	10
2.1	Modelo estándar de la física de partículas	10
2.2	El bosón de Higgs	11
3	Marco experimental	15
3.1	Organización Europea para la Investigación Nuclear	15
3.2	Gran colisionador de hadrones	16
3.3	Experimento CMS	17
3.3.1	Partes del detector	18
3.3.2	<i>Triggers</i>	19
3.3.3	Reconstrucción de objetos	20
4	Análisis	23
4.1	Procesos físicos	23
4.2	Variables básicas	25
4.3	Selección básica	26
4.4	Optimización de la señal	29
4.5	Cortes secuenciales	29
4.6	Fondo instrumental <i>Fake</i>	32
4.7	Automatización del proceso	38
5	Conclusiones	43

Índice de figuras

1	Las tres familias de fermiones (leptones y quarks)	13
2	Diagramas de Feynman de la producción del bosón de Higgs	14
3	Diagramas de Feynman de la desintegración del bosón de Higgs	14
4	Esquema de aceleradores del CERN y sus detectores	17
5	Estructura del detector CMS	19
6	Categorías iniciales del análisis	28
7	Histograma del momento transverso de dos leptones	30
8	Histograma de la diferencia del ángulo azimutal de los dos leptones y del ángulo azimutal de la energía transversa perdida	31
9	Gráfica donde se muestra la evolución de la figura de mérito en función de los valores de la variable $p_T^{\ell\ell}$	32
10	Evidencia de la relación entre los vectores de leptones y de muones	35
11	Histograma de las variables $I_\mu^{charged}$ y mvaTTH para ambos leptones	37
12	Diferencia en la variable $\Delta\phi_{\ell\ell}$ entre el inicio del proceso y el final	38
13	Evolución de la figura de mérito en función de la variable mpmet y su correspondiente histograma	39
14	Gráfica donde se muestra la comparación de la figura de mérito entre el proceso manual y el proceso automático	40
15	Diferencia en la variable mpmet entre el inicio del proceso y el final	41
16	Sistema de coordenadas del experimento CMS	49
17	Primera selección básica de procesos	51
18	Primeros cortes	51
19	Ejemplo de ejecución del código para el calculo de la figura de mérito	52
20	Código para calcular la figura de mérito	53
21	Ejemplo de salida del código de automatización del proceso de cortes secuenciales	55
22	Código para obtener, de forma automática, el valor del corte óptimo	57

Índice de tablas

1	Modos y fracciones de desintegración para el quark <i>top</i> , bosón Z y bosón W .	25
2	Selección preliminar aplicada a todos los procesos	27
3	Valor de la figura de mérito inicial	29
4	Cortes secuenciales en las variables iniciales	31
5	Cortes secuenciales en las variables de muones	36
6	Reducción de los fondos y la señal durante el proceso	36
7	Cortes secuenciales automatizados en las variables iniciales	40

Agradecimientos

Todos los finales, al igual que los comienzos, dan miedo y generan mucha incertidumbre. Así es como me encuentro en este momento, al finalizar una etapa tan importante en mi vida académica. Sin embargo, este final está lleno de gratitud hacia todas las personas que han sido parte de este viaje.

En primer lugar, quiero expresar mi más sincero agradecimiento a todos los profesores y al personal que hace posible la existencia de esta universidad. A lo largo de estos años, he tenido la suerte de aprender de algunos de los mejores profesores e investigadores, cuyas enseñanzas y ejemplos han dejado una gran huella en mí. Vuestra dedicación, pasión por la física y la investigación, y vuestra capacidad para transmitirnos entusiasmo a través de la docencia han sido una fuente constante de inspiración. Gracias por compartir no solo vuestra sabiduría, sino también vuestro tiempo, esfuerzo y apoyo.

Quisiera hacer una mención especial a mis tutores de este trabajo de fin de grado, Jónatan y Jesús. A Jónatan, gracias por tu inmensa paciencia y por estar siempre dispuesto a ayudarme con todos los contratiempos que surgieron en el camino. Tu apoyo constante me ha permitido superar los desafíos que parecían insuperables. A Jesús, gracias por ofrecerme tu guía cuando más lo necesité; tus consejos y orientaciones fueron cruciales para que pudiera completar este trabajo con éxito.

Finalmente, quiero expresar mi más profundo agradecimiento a mi familia. A mis padres, por vuestro amor incondicional, por enseñarme el valor del esfuerzo y la perseverancia, y por estar siempre a mi lado, apoyándome en cada paso de este camino. A mi hermano, por ser mi compañero de vida, ofreciéndome siempre tu constante apoyo. Y a mi pareja, gracias por tu infinita paciencia, comprensión, y por ser mi refugio en los momentos más difíciles. Sin vuestro amor, apoyo y comprensión, este logro no habría sido posible.

1. Introducción

1.1. Motivación del trabajo

La física de partículas es un campo muy amplio que busca entender la naturaleza fundamental del universo a través del estudio de las partículas elementales y sus interacciones. La teoría más aceptada para describir estas interacciones es el modelo estándar, que engloba a todas las partículas fundamentales conocidas y sus fuerzas (excepto la gravedad). En este modelo destaca el bosón de Higgs, un bosón que tiene un mecanismo asociado el cual proporciona masa a otras partículas fundamentales.

El bosón de Higgs fue propuesto teóricamente en la década de 1960 por Peter Higgs y otros investigadores, pero no fue hasta 2012 que se confirmó su existencia experimentalmente en el Gran Colisionador de Hadrones (LHC, por sus siglas en inglés) del CERN. Este descubrimiento supuso un hito significativo para el modelo estándar y la física de partículas en general, ya que completó una de las piezas clave del modelo. Sin embargo, el descubrimiento del bosón de Higgs no significa que se haya resuelto todo; más bien, abre la puerta a nuevas preguntas y la necesidad de mejorar la precisión en nuestras medidas.

El modelo estándar, aunque es exitoso, tiene limitaciones y no puede explicar algunos fenómenos, como la materia oscura y la gravedad. Además, algunas medidas recientes, como la discrepancia entre la masa del bosón W medida en Fermilab y la masa predicha por el modelo [1], y el momento anómalo del muon [2], sugieren posibles desviaciones del modelo estándar, lo que podría ser indicativo de nueva física.

Aumentar la precisión de las medidas permite encontrar discrepancias que abran la puerta a nuevas teorías que completen el modelo estándar y que permitan buscar respuestas a preguntas como qué es la materia oscura, qué pasó con la antimateria después del *Big Bang* o por qué hay tres generaciones de quarks y leptones con escalas de masa tan dispares.

1.2. Objetivos

El objetivo de este trabajo es el estudio en profundidad del bosón de Higgs mediante el análisis de simulaciones de colisiones protón-protón en el LHC a 13 TeV detectadas por CMS. El foco del trabajo radica en la optimización de la relación entre señal y fondo, mediante la reducción de los sucesos de fondo. Más concretamente, los objetivos principales de este trabajo son:

1. Profundizar en el conocimiento del bosón de Higgs, sus procesos de producción y desintegración así como sus principales características.
2. Profundizar en el conocimiento de CMS, investigando sobre las diferentes partes que lo forman y los procesos de reconstrucción de partículas que utiliza.
3. Aprender a manejar grandes volúmenes de sucesos (colisiones protón-protón) usando el sistema de colas para la ejecución de programas de CMS.
4. Conocer y comprender la cinemática, propiedades y diferencias de los diferentes procesos, mediante el estudio de numerosas variables habitualmente utilizadas en el análisis de física de partículas.
5. Maximizar el valor de la figura de mérito seleccionada.
6. Estudiar el fondo *Fake* e investigar y aplicar mecanismos para su eliminación.

1.3. Estructura

Este proyecto está dividido en cuatro partes principales.

- **Marco teórico:** en este apartado se explica el modelo estándar y el bosón de Higgs.
- **Marco experimental:** en este apartado se habla en detalle de la historia del CERN y de las características del colisionador LHC y del experimento CMS. Además, se explica los procesos de cribado de datos y de reconstrucción de objetos físicos realizados en CMS.
- **Análisis:** este apartado engloba varios subapartados donde se detallan las simulaciones usadas, las estrategias de análisis seguidas y los resultados obtenidos.
- **Conclusiones:** en este último apartado se hace una reflexión y una comparación de los resultados obtenidos mediante dos estrategias distintas. También se dan ideas de continuación o mejora para este trabajo.

2. Marco teórico

En esta sección se va a introducir el modelo estándar de la física de partículas, así como las partículas elementales y las fuerzas (salvo la gravedad) que se pueden dar entre ellas, haciendo especial hincapié en el conocimiento científico actual sobre las características del mecanismo de Higgs.

2.1. Modelo estándar de la física de partículas

Las teorías y descubrimientos de decenas de físicas y físicos en el siglo 20 han dado lugar a una extraordinaria comprensión de la estructura fundamental de la materia. Se ha demostrado que todo lo que se conoce del universo está formado por unos pocos componentes básicos, conocidos como partículas fundamentales. Estas partículas se rigen por cuatro fuerzas fundamentales: gravitatoria, electromagnética, nuclear fuerte y nuclear débil.

El Modelo Estándar (SM, por sus siglas en inglés), cuyas bases se establecieron en la década de los años 50, ha sido un pilar fundamental en la física de partículas. Desde entonces, ha logrado explicar con éxito casi todos los resultados experimentales y ha permitido predecir con gran precisión una amplia variedad de fenómenos. Con el paso del tiempo y gracias a numerosos experimentos, el SM se ha consolidado como una teoría bien probada en física [3].

El SM fue desarrollado gracias al trabajo colaborativo de muchos físicos teóricos y experimentales. Se construyó sobre descubrimientos como el principio de conservación de la paridad y la observación de que las partículas pueden clasificarse según sus propiedades de carga, espín y otras características. El desarrollo del modelo estándar culminó en la década de 1960 con el trabajo de Sheldon Glashow, Abdus Salam y Steven Weinberg, quienes recibieron el Premio Nobel en 1979 por su contribución al desarrollo de la teoría electro-débil, que unificó la fuerza electromagnética y la débil [4].

El SM está fundamentado en la mecánica cuántica y la teoría cuántica de campos, y combina las simetrías y principios de la relatividad especial para ofrecer un marco coherente que unifica: la fuerza electromagnética, la fuerza nuclear fuerte y la fuerza nuclear débil. Estas simetrías, que son esenciales para la estructura del modelo, están conectadas con el trabajo de Emmy Noether, cuyo Teorema de Noether establece que cada simetría en la naturaleza corresponde a una ley de conservación, proporcionando una base matemática crucial para entender cómo las fuerzas interactúan a nivel fundamental [5].

El SM describe dos tipos principales de partículas fundamentales: fermiones y bosones. Los fermiones, con espín semientero, son las partículas que constituyen la materia y se dividen en dos categorías: quarks y leptones. Los quarks se combinan para formar hadrones, como protones y neutrones, mientras que los leptones incluyen leptones cargados y neutrinos. Los bosones, con espín entero, son las partículas mediadoras de las fuerzas fundamentales. Estos incluyen el fotón para la fuerza electromagnética, los gluones para la fuerza nuclear fuerte y los bosones W y Z para la fuerza nuclear débil.

Uno de los aspectos más significativos del modelo estándar es el mecanismo de Higgs, que proporciona una explicación de cómo las partículas adquieren masa. Propuesto en la década de 1960, este mecanismo implica la existencia de un campo, conocido como el campo de Higgs, y una partícula asociada llamada bosón de Higgs. El campo de Higgs interactúa con ciertas partículas, confiriéndoles masa a través del fenómeno de ruptura de simetría espontánea.

2.2. El bosón de Higgs

El bosón de Higgs ha sido el centro de atención en física de partículas durante décadas, ya que su existencia era una predicción fundamental del modelo estándar, aunque no se confirmó experimentalmente hasta 2012, cuando fue observado por los experimentos ATLAS[6] y CMS [7] del CERN. Este descubrimiento fue un hito importante que validó el modelo estándar.

El bosón de Higgs es una consecuencia del mecanismo que tuvo lugar en el universo primitivo menos de un picosegundo después del *Big Bang* cuando el campo de Higgs, que permeaba todo el espacio, causó la ruptura espontánea de simetría. Esta ruptura espontánea de simetría es la que activó el mecanismo de Higgs, también conocido en el SM como el mecanismo de Brout-Englert-Higgs (BEH), el cual introduce un campo escalar complejo (espín 0) y provoca que los bosones que interactúan con él, como los bosones W y Z, adquieran masa. El bosón de Higgs descubierto es la manifestación más simple del mecanismo de Brout-Englert-Higgs. Otras teorías más allá del SM (teorías BSM) también predicen otros tipos de bosones de Higgs [8].

El mecanismo BEH se propuso por primera vez en 1964 en los trabajos de Robert Brout, François Englert, Peter Higgs, Gerald Guralnik, Carl Richard Hagen y Tom Kibble. Peter Higgs presentó más detalles del mecanismo en 1966 y Tom Kibble en 1967. En 1967, Steven Weinberg y Abdus Salam, ampliando el trabajo de Sheldon Glashow de 1961, propusieron el uso del mecanismo BEH para una teoría de la unificación de las interacciones electromagnética y débil, denominada interacción electrodébil. El elemento clave de este trabajo fue la conjetura de que la naturaleza posee una simetría electrodébil, descrita matemáticamente por el lagrangiano de la teoría, que se rompe espontáneamente, otorgando masa a los bosones W y Z [3].

Una característica adicional de este modelo es que proporciona un mecanismo para otorgar masa también a los fermiones, a través de las llamadas interacciones de Yukawa. Así, las partículas elementales que interactúan con el campo BEH adquieren masa.

Salam y Weinberg habían conjeturado además que el modelo que proponían podría ser renormalizable (es decir, dar respuestas finitas). En 1971, Gerardus 't Hooft y Martinus Veltman demostraron que esta teoría era renormalizable [3].

En el SM, el bosón de Higgs es una partícula escalar elemental, un tipo que nunca antes se había observado. Las partículas escalares fundamentales están sujetas a correcciones cuánticas que pueden ser tan grandes como la escala de la física más allá del SM (BSM).

El mecanismo BEH no predice la masa del bosón de Higgs, pero una vez fijada la masa, todas sus demás propiedades quedan definidas con gran precisión. El bosón de Higgs, una vez producido, se desintegra directamente a las partículas elementales más pesadas permitidas. Sin embargo, también pueden producirse desintegraciones a partículas sin masa a través de bucles cuánticos. En el LHC, la producción de bosones de Higgs está dominada por la fusión gluon-gluon (ggH) a través de un bucle virtual de quarks *top*. A una masa de alrededor de 125 GeV¹, el bosón de Higgs se desintegra con mayor probabilidad en un quark b y su antiquark.

A partir de la observación y medición precisas de los productos de las desintegraciones del bosón de Higgs y de los asociados a su producción, los experimentos son capaces de calcular sus propiedades, incluida la intensidad de su autointeracción y, potencialmente, las desintegraciones en partículas BSM.

En el SM hay diferentes modos de producción del bosón de Higgs, con diferentes secciones eficaces, y diferentes canales de desintegración, cuantificados por las fracciones de desintegración. La tasa de producción del bosón de Higgs viene dada por el producto de la luminosidad instantánea y la sección eficaz. El modo de producción dominante en el SM es ggH, en el que un par de gluones, uno de cada protón incidente, se fusiona, predominantemente a través de un bucle cuántico de quarks *top* virtuales. Esto representa el 87% de la sección eficaz total de producción del bosón de Higgs.

El siguiente modo de producción más importante es la fusión de bosones vectoriales (VBF, del inglés *Vector Boson Fusion*), donde un quark de cada protón irradia un bosón vectorial virtual (W o Z), que se fusionan para formar un bosón de Higgs. Otros procesos, con secciones eficaces más pequeñas, son: la producción en asociación con un bosón vectorial o *Higgsstrahlung* (VH, donde V puede ser W o Z), y la producción en asociación con quarks *top* (tH y ttH) o *bottom* (bbH). Estos modos se encuentran representados en la parte izquierda de la Figura 2.

¹En la física de partículas, y por lo tanto en este trabajo, se suelen usar las unidades naturales, de tal forma que la velocidad de la luz se expresa como $c = 1$.

Con respecto a sus desintegraciones, el bosón de Higgs del SM se acopla a bosones vectoriales, con una amplitud proporcional a la masa del bosón al cuadrado m_v^2 , y a fermiones con una amplitud proporcional a la masa del fermión m_f . Así, por ejemplo, el acoplamiento es más fuerte para la tercera generación de quarks (ver Figura 1) y leptones que para la segunda generación.

		three generations of matter (elementary fermions)		
		I	II	III
QUARKS	mass	$\approx 2.2 \text{ MeV}/c^2$	$\approx 1.28 \text{ GeV}/c^2$	$\approx 173.1 \text{ GeV}/c^2$
	charge	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{2}{3}$
	spin	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
		u up	c charm	t top
		d down	s strange	b bottom
		$\approx 4.7 \text{ MeV}/c^2$	$\approx 96 \text{ MeV}/c^2$	$\approx 4.18 \text{ GeV}/c^2$
LEPTONS	mass	$\approx 0.511 \text{ MeV}/c^2$	$\approx 105.66 \text{ MeV}/c^2$	$\approx 1.7768 \text{ GeV}/c^2$
	charge	-1	-1	-1
	spin	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$
		e electron	μ muon	τ tau
		$< 2.2 \text{ eV}/c^2$	$< 0.17 \text{ MeV}/c^2$	$< 18.2 \text{ MeV}/c^2$
		ν_e electron neutrino	ν_μ muon neutrino	ν_τ tau neutrino

Figura 1: Las tres familias de fermiones (leptones y quarks) junto con sus masas (en MeV/c^2), carga eléctrica y espín [9].

El bosón de Higgs, una vez producido, se desintegra rápidamente en un par de fermiones o en un par de bosones. En el SM, su tiempo de vida es $\tau_H \approx 1,6 \times 10^{-22} \text{ s}$, y su inversa, la anchura natural, es $\Gamma = \hbar/\tau_H = 4.14 \pm 0.02 \text{ MeV}$, donde \hbar es la constante de Planck reducida y con valor uno porque se usan unidades naturales. La anchura natural es la suma de todas las anchuras parciales, y las relaciones entre las anchuras parciales y la anchura total se denominan *branching ratios* y representan las probabilidades de que se produzca ese canal de desintegración. El bosón de Higgs no se acopla directamente a partículas sin masa (como por ejemplo, al gluon o al fotón), pero puede hacerlo a través de bucles cuánticos. Algunos modos de desintegración posibles se encuentran representados en la Figura 3.

En este trabajo nos vamos a centrar en la producción de Higgs, principalmente en el canal de fusión gluon-gluon, y su desintegración a dos bosones W.

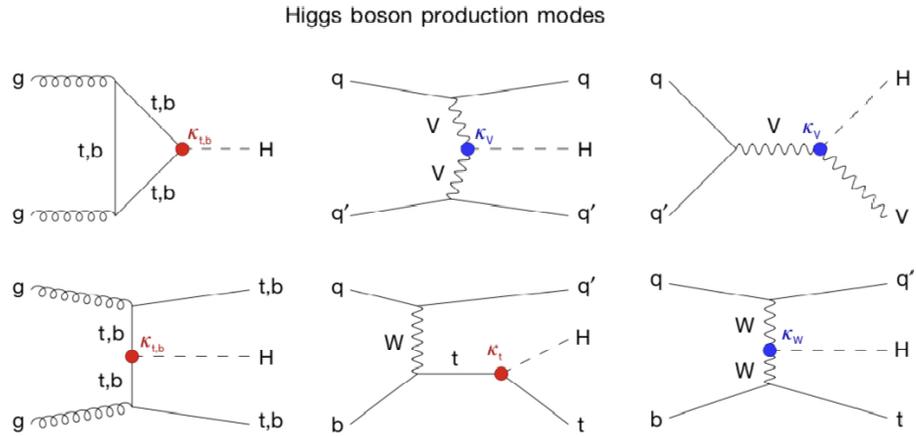


Figura 2: Diagramas de Feynman de la producción del bosón de Higgs. Comenzando en la parte superior izquierda: fusión de gluones o ggH, fusión de bosones o VBF, *Higgsstrahlung*, ttH y bbH, y los dos últimos tHq [3].

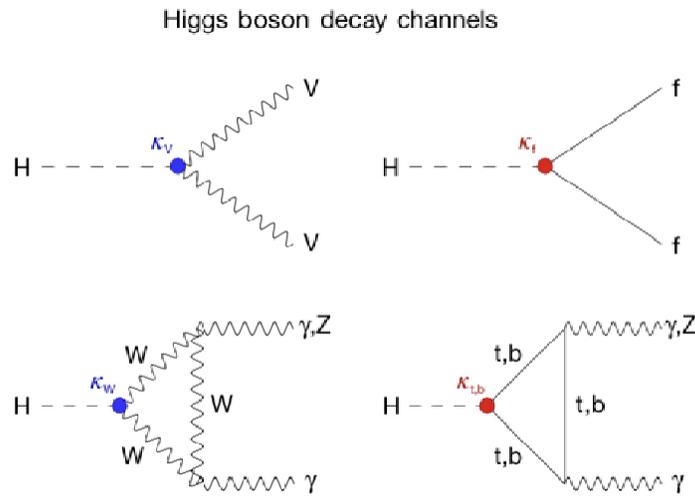


Figura 3: Diagramas de Feynman de la desintegración del bosón de Higgs [3].

3. Marco experimental

En esta sección se va a contar la historia del CERN y se va a detallar la estructura del colisionador LHC y del experimento CMS.

3.1. Organización Europea para la Investigación Nuclear

La Organización Europea para la Investigación Nuclear se originó en un contexto histórico marcado por la posguerra y la necesidad de promover la cooperación internacional en Europa. Tras la Segunda Guerra Mundial, Europa buscaba formas de unir a las naciones en proyectos de paz y reconstrucción. La física de partículas emergía como un campo prometedor, pero requería instalaciones avanzadas y costosas que ningún país podía financiar solo. Por ello, 12 países europeos se unieron para fundar esta organización, con el objetivo de avanzar en el conocimiento científico, desarrollar nuevas tecnologías, fortalecer la ciencia europea y formar a futuras generaciones de físicos, todo mediante la colaboración internacional [10].

En diciembre de 1951, durante una reunión intergubernamental de la UNESCO en París, se adoptó la primera resolución para crear el Consejo Europeo de Investigación Nuclear, conocido en francés como *Conseil Européen pour la Recherche Nucléaire*. Dos meses después, se firmó un acuerdo estableciendo un Consejo provisional, lo que dio origen al acrónimo CERN. Este acuerdo le otorgaba al Consejo un plazo de 18 meses para desarrollar el Convenio formal del CERN. Aunque inicialmente se centraba en la investigación nuclear, la comprensión actual de la materia ha avanzado mucho más allá del núcleo. Hoy en día, el CERN se dedica principalmente a la física de partículas, razón por la cual se le suele denominar Laboratorio Europeo de Física de Partículas.

En junio de 1953, 12 estados acordaron y firmaron el borrador final del Convenio del CERN. Este documento establecía las formas en que los estados miembros contribuirían al presupuesto del CERN, así como las primeras indicaciones sobre ética y organización del CERN, incluyendo la adopción de una política de acceso abierto y la división de su estructura interna en Direcciones. En la actualidad, el tamaño del CERN implica que estas Direcciones se subdividen en departamentos, y a su vez en grupos y secciones.

En julio de 1955, Felix Bloch, primer Director general del CERN, puso la primera piedra del laboratorio. Desde entonces, el CERN ha superado con creces los planes iniciales de aquellos pocos científicos que soñaban con crear un laboratorio internacional para realizar grandes avances en la investigación fundamental y ampliar los límites de nuestra tecnología e imaginación.

Desde que el CERN comenzó su andadura en 1954, se han realizado muchos avances significativos, tanto en física de partículas (como el descubrimiento de las corrientes neutras o el bosón de Higgs) como en tecnologías que han contribuido a mejorar la vida cotidiana (incluida la creación de la World Wide Web) [11]. También se ha conseguido una mayor presencia de mujeres entre el cuerpo de científicos, donde inicialmente había aproximadamente un 2-5 % de científicas y actualmente ya hay una presencia de entre el 20-25 % de mujeres [12, 13].

3.2. Gran colisionador de hadrones

El gran colisionador de hadrones (LHC) es el acelerador de partículas más grande y de mayor energía del mundo que comenzó sus primeras colisiones el 10 de septiembre de 2008 [14]. Esta es la adición más reciente al complejo de aceleradores del CERN. El LHC consiste en un anillo de 27 kilómetros de circunferencia que utiliza imanes superconductores y diversas estructuras aceleradoras para aumentar la energía de las partículas en su trayecto.

Dentro del acelerador, dos haces de partículas de alta energía se mueven casi a la velocidad de la luz antes de colisionar. Estos haces viajan en sentidos opuestas dentro de dos tubos separados, mantenidos en un vacío ultraalto. Se guían alrededor del anillo mediante unos potentes campos magnéticos generados por electroimanes superconductores. Estos electroimanes están hechos con bobinas de un cable especial que funciona en estado superconductor, lo que permite conducir electricidad sin resistencia ni pérdida de energía. Para lograr esto, los imanes deben ser enfriados a 1.85 K, una temperatura más baja que la del espacio exterior (3.15 K). Por esta razón, gran parte del acelerador está conectada a un sistema de distribución de helio líquido, que enfría los imanes y proporciona otros servicios de suministro.

Cientos de imanes de diversos tipos y tamaños se emplean para guiar los haces alrededor del acelerador. Estos incluyen 1232 imanes dipolares de 15 metros de largo, que curvan los haces, y 392 imanes cuadrupolares de entre 5 y 7 metros de largo, que los enfocan. Justo antes de la colisión, se utiliza otro tipo de imán para juntar las partículas, incrementando así las probabilidades de colisión.

Dado el diminuto tamaño de las partículas, 0.85×10^{-15} m para el protón [15] y 2.5×10^{-6} m para un haz de protones [16], hacerlas colisionar es comparable a disparar dos agujas desde 10 kilómetros de distancia con tal precisión que se encuentren en el punto medio. Todos los controles del acelerador, junto con sus servicios e infraestructura técnica, están centralizados en el Centro de control del CERN. Desde esta ubicación, los haces del LHC se hacen colisionar en cuatro puntos del anillo del acelerador, que corresponden a las ubicaciones de cuatro detectores de partículas: ATLAS, CMS, ALICE y LHCb [17]. El esquema general del LHC y sus experimentos puede verse en la Figura 4.

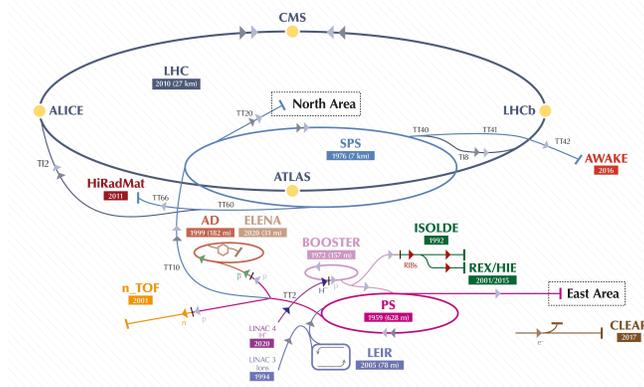


Figura 4: Esquema de aceleradores del CERN y sus detectores [18]. En la imagen destaca el colisionador LHC, cuyos detectores son ATLAS, CMS, ALICE y LHCb. Además, en ella pueden verse los aceleradores PS, SPS, y Booster que han sido reutilizados de anteriores experimentos.

3.3. Experimento CMS

El detector *Compact Muon Solenoid* (CMS) del LHC está diseñado para observar una amplia variedad de partículas y fenómenos generados en colisiones de alta energía. Su estructura en capas cilíndricas detecta diferentes partículas, creando una imagen detallada de los eventos en el centro de la colisión. Los algoritmos que combinan datos de todos los subdetectores para caracterizar los objetos físicos se denominan reconstrucción de partículas. Su nombre, constituido por tres palabras, Compacto, Solenoide y Muon, describe en gran medida su estructura. La palabra Compacto hace referencia a que apenas tiene espacios sin cubrir, la palabra Solenoide hace referencia a su gran campo magnético que permite curvar las partículas cargadas y la palabra Muon realza la especialización de este detector en la detección de muones.

El LHC se construyó para estudiar la ruptura de la simetría electrodébil y el mecanismo de Higgs, y para explorar teorías más allá del SM, como la supersimetría y dimensiones adicionales. La energía del haz y la luminosidad del LHC fueron diseñadas para estudiar la física en la escala de energía del teraelectronvoltio, lo que requirió detectores cuidadosamente diseñados. En condiciones de alta luminosidad, ocurren muchas colisiones simultáneas que complican la identificación de eventos de interés, requiriendo detectores con alta granularidad, buena resolución temporal y resistencia a la radiación, además de algoritmos que distingan entre colisiones de interés (o principales) y colisiones simultáneas (que habitualmente se denominan como *pileup*).

Las características de CMS incluyen buena identificación de muones y resolución de momento, eficiente reconstrucción de partículas cargadas, buena resolución de energía electromagnética, y baja ocupación de canales detectores. Otras características distintivas de CMS son su solenoide de alto campo (4 T), un detector de trazas basado en silicio, y un calorímetro electromagnético de cristales centelladores.

3.3.1. Partes del detector

El material de esta subsección está adaptado de [19], páginas 29 a 33.

El detector CMS del LHC está ubicado a 100 metros bajo tierra cerca de Cessy, Francia, con una longitud de 21.6 metros y un diámetro de 14.6 metros, pesando 12 500 toneladas. Su diseño y estructura están determinados por la necesidad de medir con precisión el momento de los muones, utilizando un solenoide superconductor de 4 T de 13 metros de largo y 6 metros de diámetro interior. Este solenoide proporciona una gran potencia de flexión antes de que el sistema de muones mida el ángulo de curvatura de los muones. El campo de retorno es lo suficientemente grande para saturar 1.5 metros de hierro, integrando 4 estaciones de muones que garantizan potencia y cobertura geométrica completa. Cada estación de muones tiene varias capas de tubos de deriva (DT, del inglés *drift tubes*) de aluminio en la región del barril y cámaras de tiras catódicas (CSC) en la región de las tapas, complementadas por cámaras de placas resistivas (RPC).

En el interior de la bobina magnética se encuentra el detector de trazas (*tracker*) y el calorímetro. El *tracker* de silicio está compuesto por un cilindro de 5.8 metros de longitud y 2.6 metros de diámetro. Para hacer frente a múltiples trazas, CMS consta de 10 capas de detectores microstrip de silicio, que proporcionan la granularidad y precisión requeridas. Además, 3 capas de detectores de píxeles de silicio se sitúan cerca de la región de interacción para mejorar la medición del parámetro de impacto de las partículas cargadas, así como la posición de los vértices secundarios.

El calorímetro electromagnético (ECAL, por sus siglas en inglés) utiliza cristales de tungstato de plomo (PbWO_4) que cubren hasta $|\eta| < 3$ (donde η es la pseudorapidez, ver Apéndice 1). La luz de centelleo se detecta mediante fotodiodos de avalancha de silicio en la región del barril y fototriodos de vacío en la región de las tapas. El ECAL está rodeado por un calorímetro hadrónico de latón (HCAL, por sus siglas en inglés) con una cobertura de hasta $|\eta| < 3$.

La estructura de CMS puede verse en la Figura 5.

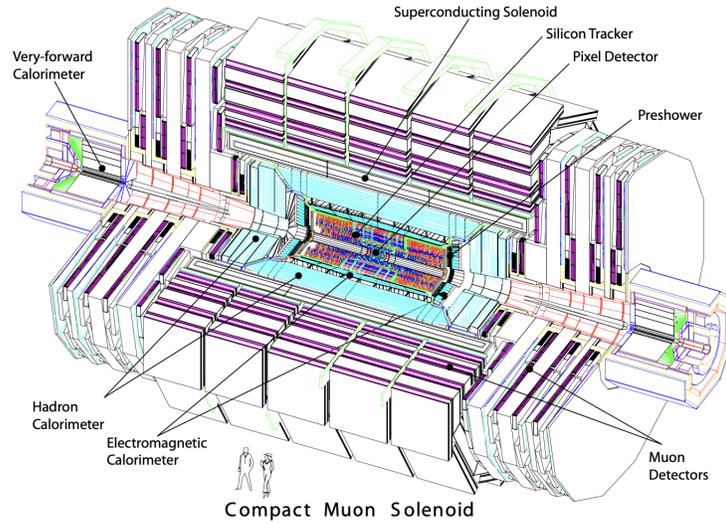


Figura 5: Estructura del detector CMS [19].

3.3.2. *Triggers*

El material de esta subsección está adaptado de [20].

En los experimentos del LHC, el sistema de disparo (*trigger*) es esencial para aceptar o rechazar eventos de colisión en tiempo real. Aunque se producen mil millones de colisiones por segundo, solo una pequeña fracción puede guardarse debido a la falta de espacio de almacenamiento. Ya que la mayoría de las interacciones son de baja energía (*soft*), las colisiones que serán de interés serán las de alta energía. En el detector CMS, el sistema de disparo tiene dos niveles: el Nivel 1 (L1) y el *Trigger* de Alto Nivel (HLT).

El L1 utiliza *hardware* personalizado sincronizado con las colisiones del LHC, tomando decisiones cada 25 ns. Este nivel procesa trazas primitivas de disparo de los subdetectores, calibrando y agrupando datos para identificar objetos candidatos (*jets*, electrones, fotones, etc.). Solo uno de cada 400 eventos se selecciona en L1, reduciendo la tasa de salida a 100 kHz.

El HLT, ejecutado en CPUs comerciales, reconstruye completamente los eventos seleccionados por L1, analizando toda la información del detector. Emplea un enfoque modular para minimizar el uso de recursos, alternando entre construcción y filtrado de eventos. Los algoritmos del HLT analizan los eventos para identificar posibilidades de estado final, emitiendo señales de aceptación o rechazo.

En el análisis que se ha llevado a cabo, un ejemplo de *trigger* en los eventos con muones es la selección de aquellos sucesos donde el muon tenga $|\eta| < 2.4$ y $p_T > 24$ GeV [21, 22].

3.3.3. Reconstrucción de objetos

El material de esta subsección está adaptado de [23], páginas 36 a 43 y de [24], página 3.

Muones

El detector CMS está diseñado para reconstruir muones con alta eficiencia. En los experimentos con colisiones de protones, los muones se reconstruyen primero en el *tracker* interno y en los sistemas de muones de forma independiente. A continuación se explica la caracterización de muones reconstruidos:

- Muones Independientes (*Standalone Muons*): Se reconstruyen usando los detectores de muones, y la reconstrucción se realiza con un método llamado filtro de Kalman [25]. Dentro de cada estación, varios planos detectan impactos de muones, los cuales se combinan en segmentos y luego se ajustan para formar trazas de muones independientes.
- Muones Globales (*Global Muons*): Si la información de las trazas de los muones independientes y las del *tracker* interno coincide en términos de momento, dirección y posición, se combinan y se utiliza nuevamente el filtro de Kalman para ajustar estos datos y obtener una traza global.
- Muones del *tracker* (*Tracker Muons*): Se obtienen extrapolando las trazas del *tracker* interno hacia el sistema de muones y buscando coincidencias con los segmentos DT o CSC. Si al menos un segmento de muon coincide con la traza extrapolada, se clasifica como muon del *tracker*. Estos muones tienen mayor eficiencia en las áreas menos instrumentadas del detector y para muones de bajo momento.

Una vez construida la trayectoria, se realiza un ajuste hacia atrás para evaluar los parámetros de la traza de muones. Debido a la gran cantidad de material que atraviesa la partícula de muon, la resolución del momento medido en las cámaras de muones puede verse afectado por la dispersión múltiple.

Electrones

Otra de las partículas clave son los electrones. Para un electrón de alta energía su trayectoria puede reconstruirse en el *tracker* y la mayor parte de su energía se libera en una pequeña matriz de cristales ECAL alrededor del punto de impacto. El procedimiento de reconstrucción de un electrón puede ser complicado debido a que a menudo el electrón interactúa con el material emitiendo parte o toda su energía a través de fotones de frenado (o *bremstrahlung*). Para los electrones de baja energía en particular, el campo magnético dobla sus trayectorias causando una dispersión de los fotones radiados a lo largo de la coordenada ϕ .

Por tanto, para obtener una medida precisa de la energía y trayectoria del electrón, es esencial tener en cuenta el efecto *bremstrahlung* en ambos subdetectores.

La reconstrucción de electrones en CMS es un proceso de tres pasos en el que se utiliza la información procedente del *tracker* y del ECAL para maximizar la eficiencia de la reconstrucción. En primer lugar, la energía depositada en ECAL se mide con un algoritmo de *clusters*. La medición de la posición y la energía de estos *clusters* se utiliza para predecir la posición de la traza de electrones en la capa más interna del *tracker*. También se construye otro tipo de colección de semillas, las llamadas *tracker-driven seeds*, que permiten extrapolar la información de la trayectoria al ECAL. A continuación, las semillas obtenidas por ambos enfoques se combinan en una traza única. Sobre esta última traza se ejecuta un paso de seguimiento y se construye la trayectoria. En el último paso se aplica una preselección para clasificar los objetos de electrones basándose en criterios de compatibilidad *track-cluster*.

***Jets* y energía perdida**

En las colisiones de protones se producen numerosas partículas, entre ellas quarks b y gluones. Debido al fenómeno de confinamiento del color, estas partículas no pueden existir libremente en la naturaleza y es por eso que se produce la hadronización. En este proceso los quarks b y los gluones forman haces colimados de hadrones, denominados *jets* [26].

En el entorno del LHC, una cantidad significativa de *jets* se producen a través de interacciones QCD (*quantum chromodynamics*) para cada evento. Por lo tanto, la reconstrucción e identificación de *jets* es una de las tareas más importantes y desafiantes en CMS. Para permitir una comparación precisa entre la predicción de la teoría y la observación, se requieren métodos bien definidos de búsqueda de *jets*.

Un tipo de *jets* que suele ser muy interesantes en el estudio de física de partículas son los denominados *b-jets* ya que son originados por un quark b. Para distinguir que *jets* provienen de quark b se utilizan unos algoritmos denominados algoritmos de *b-tagging*. Entre estos algoritmos destacan el algoritmo CSVv2, DeepCSV o DeepJet [27].

La presencia de partículas invisibles para CMS se detecta utilizando la cantidad de energía transversal faltante. La idea principal es buscar un desequilibrio en la energía transversal del suceso, que puede ser causado por partículas neutras como los neutrinos. La probabilidad de que esta energía perdida esté asociada a partículas mal reconstruidas o que caigan fuera de la aceptación del detector es pequeña, debido a la hermeticidad y a la alta eficiencia del CMS.

Algoritmo de flujo de partículas

El algoritmo de flujo de partículas (PF, del inglés *Particle Flow*) es una técnica de reconstrucción de eventos completos. Su objetivo es reconstruir e identificar diferentes partículas en el suceso como son los electrones, muones, fotones, hadrones cargados y hadrones neutros, a partir de una combinación de toda la información del subdetector CMS. De este modo se optimiza la determinación de los tipos, direcciones y energías de las partículas. La lista resultante de partículas reconstruidas con el algoritmo PF representa la mejor descripción del suceso a nivel de partícula y, por tanto, puede utilizarse para mejorar la resolución y la reconstrucción para objetos de nivel superior como *jets*, energía transversal perdida, etc.

Las partículas estables se reconstruyen en el algoritmo PF de la siguiente manera:

- Los fotones se identifican como depósitos de energía ECAL sin ningún historial vinculado de partículas cargadas. Su energía se obtiene directamente de la medición ECAL.
- Los electrones se identifican como un traza primaria de partículas cargadas combinado con depósitos de energía ECAL. También se tiene en cuenta la posibilidad de emisión de fotones *bremstrahlung*. Su energía viene determinada por una combinación del momento de la traza en el punto de interacción, la energía de los *clusters* ECAL y la suma de la energía de todos los fotones *bremstrahlung*.
- Los muones se identifican como una colección de impactos del *tracker* de silicio consistentes con una traza o varios impactos en el sistema de muones. Los muones también se asocian con un déficit de energía en los calorímetros y su energía se obtiene del momento de la traza completa.
- Las partículas cargadas que no se identifican como electrones o muones se consideran hadrones cargados. Su energía se determina a partir de una combinación del momento de la traza y la energía ECAL y HCAL correspondiente.
- Los hadrones neutros se identifican como agrupaciones de energía HCAL no vinculadas a ninguna trayectoria de hadrones cargados, o como excesos de energía ECAL y HCAL con respecto al depósito de energía de hadrones cargados esperado. Su energía se obtiene a partir de la energía ECAL y HCAL correspondiente.

4. Análisis

En esta sección se explican los procesos físicos que han sido estudiados en este trabajo y se expone el proceso seguido para la obtención de resultados.

4.1. Procesos físicos

En este trabajo se van a utilizar únicamente sucesos producidos mediante simulaciones de Monte Carlo (MC). La simulación de Monte Carlo es un tipo de algoritmo que utiliza el muestreo aleatorio repetido para obtener la probabilidad de que se produzca una serie de resultados. La simulación MC es una técnica matemática muy útil para estimar los posibles resultados de sucesos inciertos. Este método fue inventado por John von Neumann y Stanislaw Ulam durante la Segunda Guerra Mundial para mejorar la toma de decisiones cuando no se tiene toda la información necesaria [28].

En CMS se utilizan varios programas para las simulaciones MC, como PYTHIA [29], MadGraph [30], HERWIG [31] y POWHEG [32]. Estos generadores de sucesos permiten recrear a través de la teoría las colisiones que se producen en el acelerador. Por otro lado, también existen programas como GEANT4 [33] que permiten recrear y/o reconstruir la interacción de las partículas con el detector. Estos programas son adaptados y mejorados continuamente para reflejar los cambios que se producen en el detector entre diferentes periodos de toma de datos (*runs*). Estas simulaciones son corregidas, a través de los denominados factores de escala, evaluando selecciones de datos provenientes de las colisiones físicas producidas para corregir las eficiencias. Estas actualizaciones, así como las evaluaciones de los otros programas, garantizan que las simulaciones se mantengan consistentes con el comportamiento real del detector, permitiendo hacer comparaciones más fieles entre los datos simulados y experimentales.

En este trabajo se va a estudiar las colisiones protón-protón simuladas a la energía de centro de masas de 13 TeV, correspondientes a la toma de datos de CMS del año 2018. Los procesos que se van a estudiar en este trabajo se engloban en las siguientes categorías:

- Drell-Yan (DY): en esta categoría se recogen los procesos de DY que son aquellos donde se aniquilan un quark y un antiquark cuando colisionan dos hadrones, para dar un fotón o un bosón Z que a su vez se desintegran en un par de leptón-antileptón [34, 35].
- *Top*: en esta categoría se engloban los procesos en los que se producen un par $t\bar{t}$ o individuales (como tW). En la Tabla 1 puede verse la desintegración del bosón *top*.
- VVV : en esta categoría se engloban los procesos de producción de tres bosones, como son ZZZ, WZZ, WWZ y WWW.

- VZ (donde V es un bosón W o Z): en esta categoría se engloban los procesos donde se desintegra un bosón Z con otro bosón a 2 ó 4 leptones. En esta categoría se van a encontrar desintegraciones del par ZZ en dos leptones y dos quarks, a cuatro leptones o a cuatro quarks. Por otro lado, en este grupo también se contempla la desintegración del par WZ en dos leptones y dos quarks [36]. Las desintegraciones de los bosones W y Z se encuentran en la Tabla 1.
- $V\gamma$ y $V\gamma^*$: en estas categorías se engloban procesos que involucran la producción de un bosón (como W o Z) y un fotón real o virtual. En la categoría $V\gamma^*$ también se ha tenido en cuenta la desintegración de WZ a 3 leptones.
- *Fake* : se denominan procesos *Fake* a aquellos muones o electrones que son *nonprompt* o que están mal reconstruidos, como por ejemplo *jets* que se reconstruyen como electrones debido a que dejan rastro en el ECAL. Los muones o electrones *nonprompt* son aquellos que no se producen en el vértice principal de la colisión, si no que provienen de la desintegración fuerte de hadrones [24]. Los procesos que se estudian pueden ser de tres tipos de combinaciones: leptón *prompt*-leptón *prompt*, leptón *prompt*-leptón *fake*, leptón *fake*-leptón *fake*. El primer caso está limpio (no hay *fakes*) y el tercer caso afecta muy poco al estudio ya que es muy improbable. Es por esto que la mayoría de los procesos de este fondo se encuentran en la segunda categoría. Debido a que el bosón de Higgs y el bosón W se pueden desintegrar en cualquiera de los tres leptones (electrón, muon y τ) dentro de esta categoría también deberían estar incluidos los leptones τ . Sin embargo, debido a que estas partículas pueden desintegrarse en otras, lo que complica mucho su estudio, este trabajo se centra exclusivamente en electrones y muones.
- WW: en esta categoría se engloban la desintegración de un par WW ² a dos leptones cargados ($WW \rightarrow \ell^\pm \nu \ell^\mp \nu$) o por procesos electrodébiles. Además, también se tienen en cuenta la producción del par WW mediante dos gluones o dos quarks.
- Higgs: los procesos Higgs son todos aquellos procesos de producción de Higgs y de desintegración en algunos de sus canales. Ejemplos de estos procesos son: la producción de un Higgs mediante dos gluones y su desintegración en un par WW que se desintegra en dos leptones cargados, la producción de un Higgs mediante dos gluones y su desintegración en un bosón Z y un fotón, que a su vez se desintegra a un par $\tau\bar{\tau}$ o la producción de un Higgs mediante un par $t\bar{t}$.

De todos estos procesos se tomarán como señal los procesos Higgs y WW, y el resto de procesos como fondo. Dentro de todos los procesos que forman el fondo se hará especial énfasis en los procesos *Fake*.

²Si bien es más preciso escribir W^+W^- , es frecuente en física de partículas no escribir las cargas eléctricas de las partículas cuando se sobreentiende por el contexto.

Quark <i>top</i>	Bosón Z	Bosón W
99.8 % $t \rightarrow bW$	70 % $Z \rightarrow q\bar{q}$	67 % $W \rightarrow \text{quarks}$
	30 % $Z \rightarrow \ell\ell$	12 % $W \rightarrow \tau\nu_\tau$
		11 % $W \rightarrow \mu\nu_\mu$
		10 % $W \rightarrow e\nu_e$

Tabla 1: Modos y fracciones de desintegración para el quark *top*, bosón Z y bosón W [36, 37, 38].

4.2. Variables básicas

En esta sección se enumeran y se describen las variables iniciales básicas de leptones que se han utilizado en el análisis. Las definiciones y ecuaciones de estas variables han sido obtenidas de los artículos o publicaciones del grupo de Higgs de CMS [39].

1. eventos: número de sucesos que se han seleccionado.
2. $m_{\ell\ell}$: masa invariante de los dos leptones cargados. Su ecuación:

$$m_{\ell\ell} = \sqrt{(E_{\ell_1} + E_{\ell_2})^2 - (\vec{p}_{\ell_1} + \vec{p}_{\ell_2})^2}$$

donde E es la energía y p es el momento.

3. $\Delta R_{\ell\ell}$: distancia entre los dos leptones. Su ecuación:

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

donde η es la pseudorapidez y ϕ es el ángulo azimutal.

4. $m_T^{\ell\ell+\cancel{E}_T}$: masa transversa. Su ecuación:

$$m_T^{\ell\ell+\cancel{E}_T} = \sqrt{2 \cdot p_T^\ell \cdot \cancel{E}_T \cdot (1 - \cos(\Delta\phi))}$$

donde \cancel{E}_T es la energía perdida.

5. $m_T^{\ell_2+\cancel{E}_T}$: masa transversa calculada entre el segundo leptón y la energía transversa perdida. Su ecuación:

$$m_T^{\ell_2+\cancel{E}_T} = \sqrt{2 \cdot p_T^{\ell_2} \cdot \cancel{E}_T \cdot (1 - \cos(\Delta\phi))}$$

6. $p_T^{\ell\ell}$: una del momento transverso de los dos leptones. Su ecuación:

$$p_T^{\ell\ell} = | \vec{p}_T^{\ell_1} + \vec{p}_T^{\ell_2} |$$

7. $p_T^{\ell z}$: momento transverso de cada leptón. Su ecuación:

$$p_T^{\ell z} = \sqrt{p_x^2 + p_y^2}.$$

8. η_ℓ : pseudorapidez de cada leptón. Su ecuación:

$$\eta_\ell = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right)$$

9. *pupmet*: momento transverso perdido calculado con mitigación de *pileup* [40].

10. *mpmet*: variable dependiente de la energía transversa perdida, así como de su separación angular (ϕ) con cada uno de los leptones. Para una definición detallada ver Sección 8.6 de [41].

11. $\Delta\phi_{\ell\ell}$: diferencia absoluta del ángulo azimutal (ϕ) entre los dos leptones. Su ecuación:

$$|\Delta\phi_{\ell\ell}| = |\phi_{\ell_1} - \phi_{\ell_2}|$$

12. $\Delta\phi_{\ell\ell, \mathcal{E}_T}$: es la diferencia del ángulo azimutal de los dos leptones y del ángulo azimutal de la energía transversa perdida. Su ecuación:

$$\Delta\phi_{\ell\ell, \mathcal{E}_T} = \phi_{\ell\ell} - \phi_{\mathcal{E}_T}$$

4.3. Selección básica

Los sucesos MC producidos pasan inicialmente por una selección básica para adaptarse al estudio que se quiere realizar. Este trabajo se centra en la señal proveniente de procesos WW y Higgs (de ahora en adelante HWW, ya que el estudio se centra en la desintegración del bosón de Higgs en dos bosones W), donde ambos procesos pueden desintegrarse de diversas formas (distintas combinaciones de quarks y leptones). En este trabajo se va a estudiar la desintegración que contiene menos fondo, que es la que contiene dos leptones cargados (electrón y muon) con carga eléctrica opuesta, donde la energía perdida es causada por los neutrinos producidos en la desintegración de bosones W y con *jets* generados por la hadronización de quarks y gluones.

En este caso, en la configuración inicial (ver Tabla 2) se escogen los sucesos que tengan dos leptones de diferente sabor y carga (sucesos electrón-muon o muon-electrón), lo que reduce en gran medida el fondo DY, y que pasen ciertos criterios de aislamiento definidos, donde el aislamiento de un leptón es la medición de la cantidad de energía y momento total de las partículas dentro de un cono con un cierto radio alrededor de la trayectoria del leptón. En el caso de los muones, hay una configuración inicial definida por el grupo de física de muones de CMS donde se requiere que los muones sean del tipo *tight* [42] y que tengan un cierto aislamiento en un cono de $\Delta R = 0.4$, donde $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ [43]. Al elegir muones *tight* los requisitos son más estrictos a la hora de aceptar que un objeto es un muon.

Esta selección reduce la eficiencia de detección de objetos pero aumenta significativamente la pureza, donde la pureza se define como el cociente entre el número de muones verdaderos reconstruidos y la suma de muones verdaderos y partículas mal reconstruidas como muones.

De forma complementaria a estas selecciones también se pide que el primer leptón tenga un momento transverso mayor de 25 GeV (donde se están usando en todo momento unidades naturales, $c = 1$) y el segundo leptón mayor de 20 GeV, que la masa invariante de los dos leptones sea mayor de 12 GeV y que la energía perdida en el plano transversal sea mayor de 20 GeV. Por otro lado, se define el primer leptón como aquel leptón con mayor momento transversal, siendo el momento transversal definido como:

$$p_T = \sqrt{p_x^2 + p_y^2} \quad (1)$$

donde p_x en la coordenada x del momento y p_y es la coordenada y .

Selección básica
electrón-muon con distinta carga
$p_T^{\ell_1} > 25$ GeV
$p_T^{\ell_2} > 20$ GeV
Si hay más de 2 leptones, $p_T^{\ell_3} < 10$ GeV
$m_{\ell\ell} > 12$ GeV
mpmet > 20 GeV
puppimet > 20 GeV
$p_T^{\ell\ell} > 30$ GeV
$m_T^{\ell\ell + \cancel{E}_T} > 50$ GeV
$\Delta R_{\ell\ell} < 2.5$

Tabla 2: Selección preliminar aplicada a todos los procesos para centrar el análisis en la topología interesante en este trabajo (un par muon-electrón con carga eléctrica opuesta y energía perdida en el plano transversal).

Más detalles de esta configuración inicial están descritos en el Apéndice 3.

Se ha comentado que el hecho de coger sucesos que son del tipo electrón-muon reduce notablemente el fondo DY ya que estos procesos están relacionados con la creación de un leptón cargado y su antileptón, los dos con el mismo sabor. Sin embargo, este fondo sigue teniendo una pequeña contribución en los sucesos deseados ya que la desintegración de leptones τ puede generar la aparición de electrones y muones. Es por este motivo que se aplican los cortes en las variables mpmet, puppimet y $m_T^{\ell\ell + \cancel{E}_T}$. Estas variables también disminuyen la presencia de electrones y muones provenientes de la desintegración de un bosón Z o un fotón en dos leptones $\tau\bar{\tau}$. Por otro lado, debido a que se quiere hacer un estudio de sistemas de electrón-muon, se han realizado varios cortes en diferentes variables que son $p_T^{\ell_1}, p_T^{\ell_2}$ y $p_T^{\ell_3}$. El corte en la variable $m_{\ell\ell}$ suprime sucesos provenientes de resonancias de masa pequeña.

Finalmente, el corte en la variable $\Delta R_{\ell\ell}$ asegura tener sucesos donde se han producido partículas muy energéticas, ya que en estos casos los leptones cargados resultantes tienden a aparecer más cercanos. [41].

Dentro de estas primeras selecciones se crean dos categorías, una categoría de procesos que no tienen *jets* (0j) y otra para procesos con *jets* y un quark b en ellos (1bj). Estas categorías ayudan a diferenciar entre diferentes procesos sin necesidad de eliminar algunos de ellos. La condición para que un suceso esté en la categoría 0j es que no haya ningún *jet* con momento transversal mayor de 30 GeV. Por otro lado, para que un suceso esté en la categoría 1bj tiene que tener un *jet* con momento transversal mayor de 30 GeV, un valor absoluto de pseudorapidez $|\eta|$ inferior a 2.5, y un valor del discriminante de *b-tagging* conocido como DeepCSV [44] mayor a 0.1241. Este umbral de 0.1241 corresponde al punto de trabajo *Loose*. En este trabajo se hará un estudio de varias variables iniciales, que se representarán para ambas categorías. Sin embargo, debido a que la categoría 1bj está enriquecida en procesos *Top*, este trabajo se centrará en tratar de optimizar la señal en la categoría 0j.

A continuación, en la Figura 6 se muestra un ejemplo de la variable de la masa invariante de los dos leptones pintada para cada una de las categorías. En ellas se puede ver como en la categoría 1bj los procesos predominantes son los *Top* mientras que en la 0j el número de estos procesos se reduce bastante.

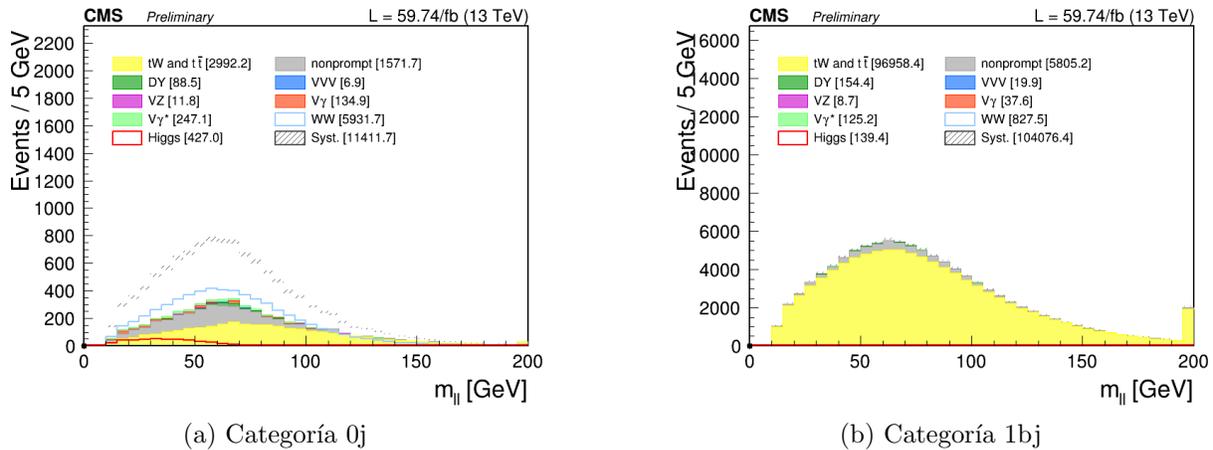


Figura 6: Categorías iniciales del análisis. En ambas gráficas 'Syst.' es la suma de todos los procesos. Los procesos de fondo aparecen apilados unos sobre otros y con color relleno y los dos procesos de señal, Higgs y WW, no aparece apilada y su representación es sin relleno.

4.4. Optimización de la señal

Para poder cuantificar los sucesivos cortes que se realizan sobre los sucesos se usará una figura de mérito (\mathcal{S}) muy típica en análisis de física de partículas, definida como:

$$\mathcal{S} = \frac{S}{\sqrt{B+S}} \quad (2)$$

donde S es el número de sucesos de señal y B es el número de sucesos de fondo. Esta figura de mérito es muy común para señales pequeñas, ya que es una función que crece muy lentamente para señales grandes y que penaliza mucho la pérdida de señal [45].

A continuación, en la Tabla 3 se muestra el \mathcal{S} del que se parte, con la configuración inicial previamente mencionada. Además, se muestra el número de sucesos para los procesos HWW y WW (señal), y el total de los fondos.

Sucesos	Número de sucesos
HIGGS	427
WW	5930
Fondos	5040
\mathcal{S}	59.52

Tabla 3: Valor de la figura de mérito inicial.

4.5. Cortes secuenciales

Antes de empezar el estudio y análisis de las variables específicas del fondo *Fake* se ha realizado un procedimiento de cortes secuenciales en las variables iniciales para tratar de maximizar el \mathcal{S} .

En primer lugar, a partir de los histogramas iniciales, se ha seleccionado las variables donde se apreciaba, visualmente, que las distribuciones del fondo y la señal eran ligeramente diferentes. Un ejemplo de estas variables puede verse en la Figura 7. Cabe recordar que se corresponde con la categoría 0j.

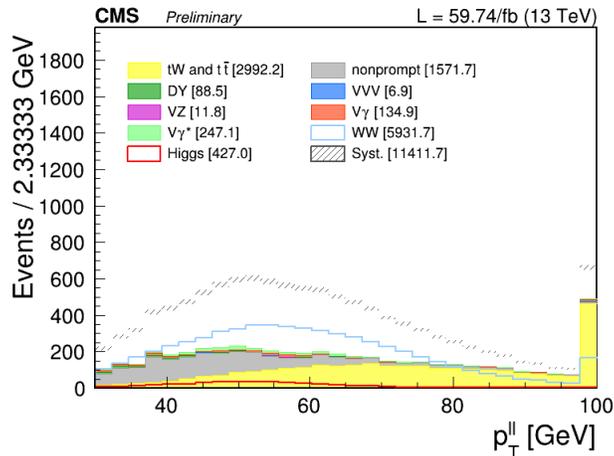


Figura 7: Histograma del momento transverso de dos leptones.

A continuación, se escoge una de las estas variables y se visualiza la zona óptima de corte. Una vez se estima esta zona se realizan diversos cortes en este rango y se escoge el que mejor figura de mérito aporta. Seguidamente se mandan los *jobs* a las máquinas del CERN. Los *jobs* son tareas específicas de procesamiento de datos. En este trabajo, se envían 631 *jobs* cada vez que se ejecuta un nuevo corte, y estos se ejecutan en la *Grid* de computación del LHC, una infraestructura distribuida globalmente para el procesamiento eficiente de grandes volúmenes de datos experimentales.

Por último, se ha programado un *script* de python, referido como *Tabla de sucesos*, en el cual se ejecuta el fichero que te devuelven los *jobs* e imprime en pantalla una tabla con el número de sucesos de cada tipo de fondo y señal, la suma de total de todos los sucesos y el \mathcal{S} para cada valor seleccionado del corte. Este código puede verse en detalle en el Apéndice 4. Finalmente, una vez seleccionado el valor de la variable que aumentaba lo máximo posible el \mathcal{S} se seleccionan los histogramas de ese corte y se vuelve a comenzar el proceso con la siguiente variable.

Estas primeras variables se corresponden con las primeras cinco filas de la Tabla 4. En esta tabla se muestra el nombre de la variable seleccionada, el valor final del corte, la eficiencia de sucesos HWW, WW y fondos y el \mathcal{S} para ese corte. Como ya se ha comentado, la figura de mérito seleccionada tiene un crecimiento muy lento para señales grandes y es por eso que el aumento de \mathcal{S} es muy pequeño de un corte a otro. Además, añadido a esto, la poca pérdida de señal que se produce al realizar los cortes está muy penalizada, lo que limita aún más el crecimiento del \mathcal{S} .

Después de estos primeros cinco cortes secuenciales se han seleccionado de las variables restantes aquellas que tienen una cola más larga, ya sea al comienzo o al final, donde se puede quitar un poco de fondo sin llevarse mucha señal. Un ejemplo de estas variables puede verse en la Figura 8. El procedimiento seguido es idéntico al seguido con las otras variables. Estas variables se corresponden con las últimas cinco filas de la Tabla 4.

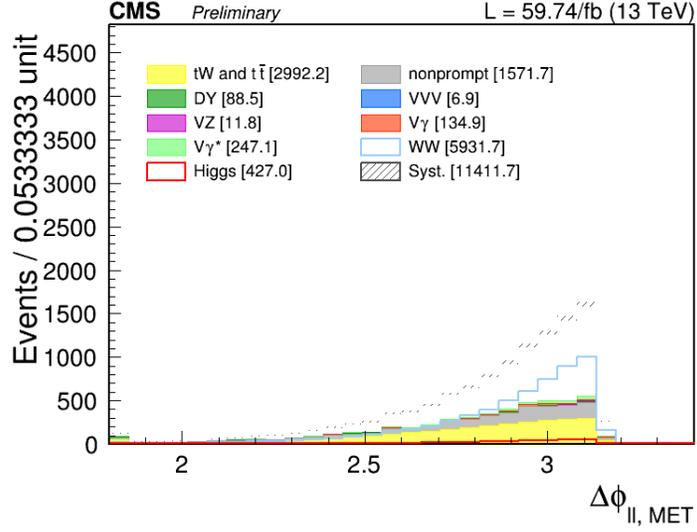


Figura 8: Histograma de la diferencia del ángulo azimutal de los dos leptones y del ángulo azimutal de la energía transversa perdida.

Variable	Corte	HWW	WW	Fondos	\mathcal{S}
Inicial		427	5930	5040	59.52
$p_T^{\ell\ell}$	<95 GeV	99.3 %	96.8 %	87.9 %	59.72
$m_T^{\ell\ell+\cancel{E}_T}$	<207 GeV	99 %	96.6 %	87.9 %	59.75
puppimet	<119 GeV	99 %	96.5 %	87.9 %	59.79
mpmet	<113 GeV	99 %	96.5 %	86.4 %	59.79
$m_T^{\ell_2+\cancel{E}_T}$	<140 GeV	99 %	96.3 %	86.4 %	59.80
$\Delta\phi_{\ell\ell, \cancel{E}_T}$	>1.85	98.1 %	95.9 %	86.2 %	59.86
$p_T^{\ell_1}$	<125 GeV	98.1 %	95.9 %	86 %	59.87
$p_T^{\ell_2}$	<94 GeV	98.1 %	95.9 %	85.5 %	59.88
$m_{\ell\ell}$	<168 GeV	98.1 %	95.8 %	85 %	59.88

Tabla 4: Cortes secuenciales en las variables iniciales.

En la Figura 9 se muestra una gráfica para visualizar la evolución de la figura de mérito en función de la variable $p_T^{\ell\ell}$. Esta gráfica muestra cual sería el valor de la figura de mérito si se realizara el corte en el valor correspondiente de la variable. Se puede observar cómo el valor de \mathcal{S} comienza en 0 y aumenta hasta llegar a un límite. En la parte final, más o menos a partir del valor 85 GeV, hay pequeñas oscilaciones de la figura de mérito que no se perciben debido al rango del eje y. Dentro de este rango es donde se encuentra la mejora de la figura de mérito.

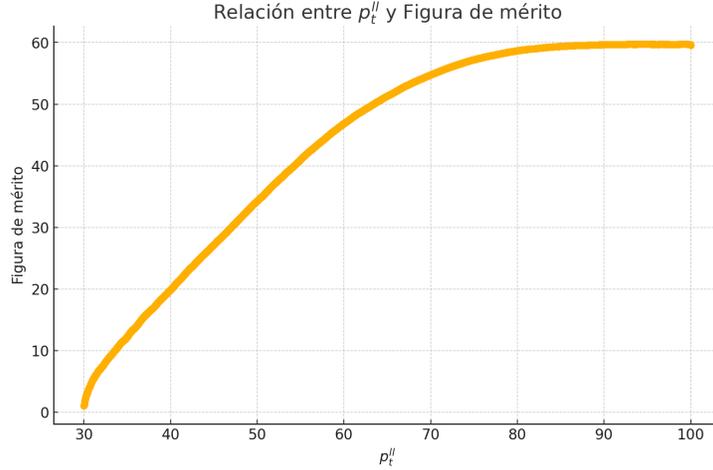


Figura 9: Gráfica donde se muestra la evolución de la figura de mérito en función de los valores de la variable p_T^{ll} .

El procedimiento que se ha seguido tiene una serie de limitaciones.

En primer lugar, es complicado apreciar visualmente la señal del fondo debido a que los procesos se encuentran situados unos sobre otros. Esto se corregirá en los siguientes apartados de análisis, lo que garantizará que se pueda distinguir mejor las diferentes distribuciones. Añadido a esto, el hecho de que las distribuciones de las dos señales, HWW y WW, sean distintas, también dificulta la selección de la zona óptima de corte. Por otro lado, el hecho de que sea difícil la selección visual de la zona óptima de corte, añadido a la penalización del \mathcal{S} a la pérdida de señal, provoca que los cortes sean muy suaves. En muchos casos, como la ganancia de \mathcal{S} es muy pequeña, está ganancia podría deberse a fluctuaciones estadísticas de las colas de las distribuciones. Se han incluido los resultados obtenidos para ilustrar el procedimiento seguido, pero no se aprecia una gran mejoría en estos cortes.

Un paso más avanzado, y que garantizaría una mayor ganancia de la figura de mérito, sería el uso de redes neuronales para estudiar las correlaciones entre las distintas variables, ya que estas correlaciones no se tienen en cuenta en el proceso de cortes secuenciales.

4.6. Fondo instrumental *Fake*

Una vez se ha aumentado todo lo posible el \mathcal{S} con las variables iniciales se va a centrar el estudio en el fondo *Fake*.

El fondo *Fake* no solo está relacionado con la mala clasificación de las partículas en el proceso de reconstrucción, que afecta principalmente a electrones y muones, si no que también está relacionado con los electrones y muones bien clasificados pero que no pertenecen al vértice primario, muones o electrones *nonprompt*. Dentro del primer conjunto estarían todas aquellas partículas que se han clasificado como un electrón o muon sin serlo realmente. Un ejemplo de este grupo serían *jets* que han sido mal reconstruidos como muones o electrones.

Por otro lado, los muones y electrones *nonprompt* son muones y electrones reales que se producen en vértices secundarios. Cuando se produce una colisión protón-protón y se produce un bosón de Higgs este se desintegra muy rápido, tanto que a efectos prácticos esta desintegración ocurre en el punto de colisión o vértice primario. La desintegración interesante en este trabajo es la desintegración del bosón de Higgs en dos bosones W. Este bosón también tiene un tiempo de vida muy pequeño, por lo que también se desintegra en el vértice primario. La desintegración de los bosones W da lugar a dos leptones cargados, aunque en este trabajo se va a centrar el estudio en las desintegraciones en electrones o muones. Estos leptones cargados provenientes de la desintegración de dos bosones W se denominan *prompt* y son los que se toman como señal. Los electrones y muones que provienen de desintegraciones subsecuentes, que no se producen en el vértice primario, son los que se denominan *nonprompt* y los que se toman como fondo. Un ejemplo de este conjunto pueden ser los electrones provenientes de los *jets* que se crean en la desintegración del quark b.

El fondo *Fake* es un fondo difícil de modelar. Para modelar este fondo no se utilizan las simulaciones MC, si no que se utilizan técnicas complejas basadas en los datos recogidos del detector para estimarlo, conocidas como técnicas *Fake Rate* [46]. Dentro de este fondo se va a centrar el análisis en los muones, para lo que se va a utilizar variables específicas de muones. El objetivo de este estudio es encontrar variables que sean discriminantes entre el fondo *Fake* y la señal, con el fin de eliminar el mayor número de sucesos posibles de fondo.

Algunas de las variables seleccionadas hacen referencia al número de segmentos o detectores que se han utilizado para reconstruir los muones, otras variables están calculadas a partir de variables extraídas del proceso de reconstrucción (como el momento, la posición, etc.) y otras son variables calculadas con *machine learning* a partir de una combinación de otras variables más básicas.

A continuación se exponen las variables añadidas junto con su descripción [24]:

- Número de estaciones: número de estaciones del detector de muones que han registrado un paso del muon. Esta variable puede tomar valores entre 0 y 5.
- Compatibilidad de segmentos: compatibilidad del segmento del muon. Es una variable que indica cuán bien los segmentos del detector de muones coinciden con la trayectoria esperada del muon. Esta variable puede tomar valores entre 0 y 1.
- Número de capas del *tracker*: número de capas del *tracker* que han registrado el paso del muón. Esta variable puede tomar valores entre 0 y 20.

- d_{xy}^μ : desplazamiento transversal del muon con respecto al vértice primario de colisión. Esta variable puede tomar valores entre -0.02 y 0.02 y sus unidades son centímetros.
- d_z^μ : desplazamiento longitudinal del muon con respecto al vértice primario de colisión. Esta variable puede tomar valores entre -0.1 y 0.1 y sus unidades son centímetros.
- jetPtRelv2: componente del momento transverso del muon relativa a la dirección del *jet*. Esta variable puede tomar valores entre 0 y 50.
- mvaTTH: evaluación de la identificación de muones basado en un algoritmo de análisis multivariable (MVA) para la identificación de muones y su aislamiento en el contexto del análisis de $t\bar{t}H$. Esta variable puede tomar valores entre -1 y 1.
- mvaId: evaluación de identificación de muones basado en un algoritmo MVA general para la identificación de muones. Esta variable puede tomar valores entre 0 y 4.
- $I_\mu^{charged}$: componente cargada de la variable de mini-aislamiento, donde el mini-aislamiento hace referencia al aislamiento de la partícula en un un radio inversamente proporcional al momento transverso de la partícula de estudio [47]. Esta variable puede tomar valores entre 0 y 0.4.
- jetRelIso: aislamiento relativo del muon en relación con los *jets* del alrededor. Esta variable puede tomar valores entre 0 y 0.4.

Validación del código

Las variables básicas que han sido estudiadas hasta el momento eran variables generales de leptones. Debido a que hay más de un leptón en los sucesos las variables tratadas son vectoriales, donde, por ejemplo, la primera entrada del vector es el valor de esa variable para el primer leptón. Sin embargo, como se están seleccionando procesos que involucran la presencia de un electrón y un muon, para el uso de las variables específicas de muones hay que usar una relación entre el vector de leptones (que indica el tipo de partículas para cada uno de los leptones que hay en el proceso) y los vectores de las variables de los muones. Con este motivo se utiliza la variable `muonIdx`, donde la primera entrada será -1 si el primer leptón es un electrón o un número entre 0 y 3 (el máximo es 3 en este análisis ya que los procesos tienen como mucho 4 muones) si el primer leptón es un muon. Este número entre 0 y 3 indica la entrada en las variables de muones correspondientes al muon del proceso.

Como comprobación de esta relación se han creado dos categorías, complementarias a las categorías iniciales 0j y 1bj que ya se han comentado, denominadas 0j_{mu}, donde se seleccionan procesos sin *jets* donde el primer leptón es un electrón, y 0j_{mue}, donde se seleccionan procesos sin *jets* donde el primer leptón es un muon. De esta manera, para los procesos de la categoría 0j_{mu} se debería observar en los histogramas asociados al 'Muon 2' (donde el 2 hace referencia a que el muón es el segundo leptón) que todos los sucesos están en 0 porque no hay valor para ese muón. Esto puede observarse en la Figura 10.

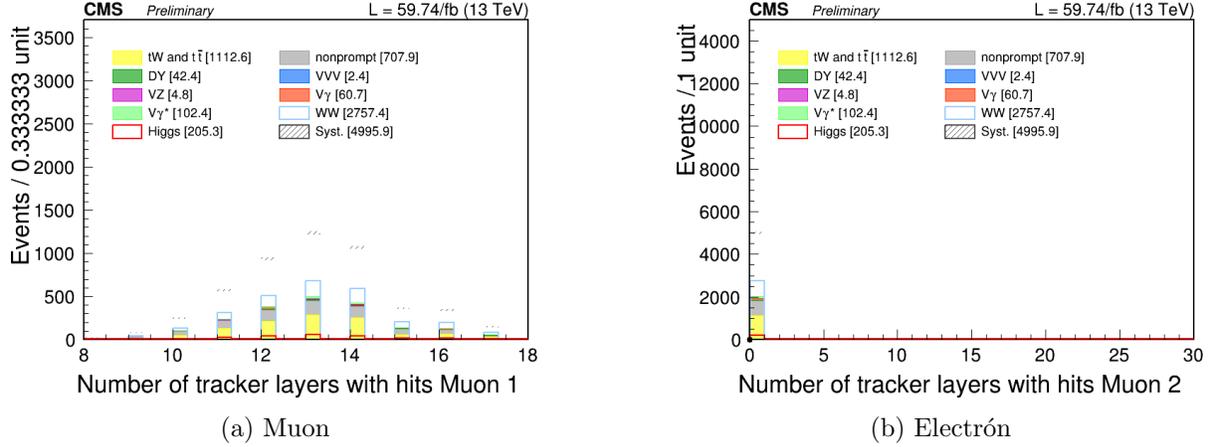


Figura 10: Evidencia de la relación entre los vectores de leptones y de muones.

En estas gráficas puede verse como los valores de la variable Número de capas del *tracker* están entre 9 y 18 para el 'Muon 1' y de 0 para el 'Muon 2'.

Una vez se ha comprobado el correcto funcionamiento del vector de relación entre leptones y muones, se va a modificar el valor 0 al valor máximo o mínimo según interés (para no perder con el corte estos eventos) de la variable específica, para evitar que estos sucesos se amontonen en el valor 0 de la variable y distorsionen las gráficas, dificultando la visualización de la zona óptima de corte.

Resultados del proceso

De todas las variables numeradas anteriormente, las variables $I_{\mu}^{charged}$ y mvaTTH han mostrado ser muy buenas variables de discriminación entre fondo *Fake* y señal. Esto es debido a que es posible visualizar en los histogramas (ver Figura 11) una clara diferenciación entre el fondo *Fake* y la señal.

En la Figura 11 puede verse ambas variables para cada uno de los muones. En la variable $I_{\mu}^{charged}$ se observa como la señal (líneas azul y roja) tienen un mayor número de sucesos hacia la parte izquierda del eje y el fondo *Fake* (gris) tiene un gran número de sucesos hacia la parte derecha del eje. Por otro lado, la variable mvaTTH tiene mayor número de sucesos del fondo hacia la izquierda y más sucesos de señal hacia la derecha. Cabe destacar que estas gráficas están en escala logarítmica debido a la diferencia de ordenes de magnitud entre la cantidad de sucesos de la señal y del fondo.

Para estas dos variables se ha realizado un proceso de cortes, igual que el que se ha realizado con las variables iniciales. En la Tabla 5 pueden observarse las variables donde se han realizado los cortes, el valor seleccionado del corte, la eficiencia de los sucesos HWW, WW, *Fake*, fondos y el valor de \mathcal{S} después del corte.

Variable	Corte	HWW	WW	<i>Fake</i>	Fondos	\mathcal{S}
Inicio del proceso		419	5680	1520	4300	59.88
$I_{\mu_1}^{charged}$	<0.007	99.8 %	99.8 %	94.7 %	95.3 %	60.17
$I_{\mu_2}^{charged}$	<0.008	99.5 %	99.5 %	75.3 %	88.4 %	61.01
mvaTTH 1	<0.05	99.5 %	99.5 %	65 %	86 %	61.14
mvaTTH 2	<-0.05	99 %	99.1 %	63 %	83.7 %	61.51

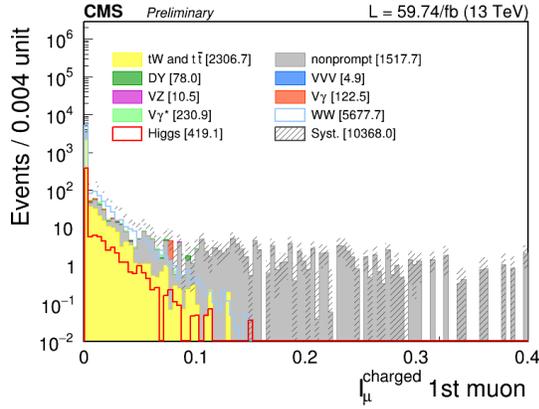
Tabla 5: Cortes secuenciales en las variables de muones.

De estos resultados destaca el aumento significativo del \mathcal{S} comparado con el que se había conseguido con los cortes secuenciales iniciales. Este aumento se debe al gran poder discriminante de estas variables, lo que provoca que al realizar los cortes se elimine muchos sucesos del fondo sin apenas afectar al número de procesos de la señal.

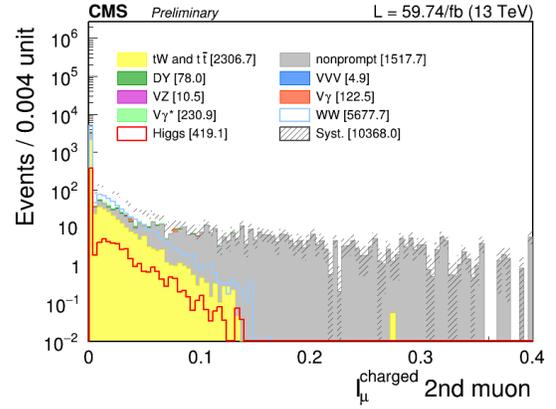
A pesar de que el aumento de la figura de mérito no ha sido muy llamativo, es destacable la reducción del fondo *Fake* obtenida. En la Tabla 6 puede verse la cantidad de sucesos de procesos HWW, WW, señal (que es la suma de HWW y WW) y de todos los procesos de fondo al inicio del estudio y al final del proceso de cortes. Se observa una reducción del 43 % en el fondo *Fake*, mientras se mantiene el 95 % de la señal. Esta significativa reducción del fondo *Fake* es crucial en los análisis de altas energías, ya que estos sucesos pueden imitar procesos físicos importantes, dificultando la identificación de nuevas partículas y fenómenos. Este fondo requiere técnicas complejas de selección y análisis, pues consiste en sucesos donde partículas no leptónicas son erróneamente identificadas como leptones, contaminando las señales de interés.

	HWW	WW	Señal	<i>Fake</i>	<i>Top</i>	DY	VZ	$V\gamma^*$	$V\gamma$	VVV
Inicial	427	5930	6360	1570	2990	89	12	247	135	7
Final	415	5650	6040	900	2290	76	10	226	117	5
Eficiencia (%)	97	95	95	57	76	85	83	92	87	71

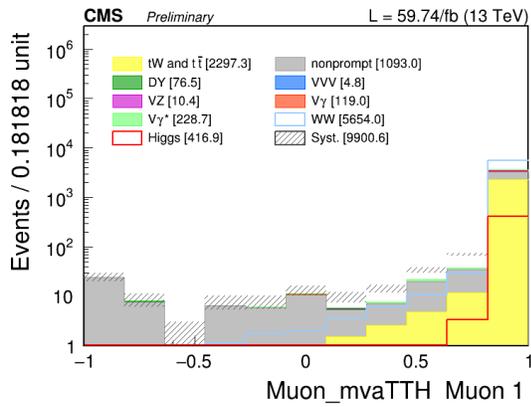
Tabla 6: Reducción de los fondos y la señal durante el proceso.



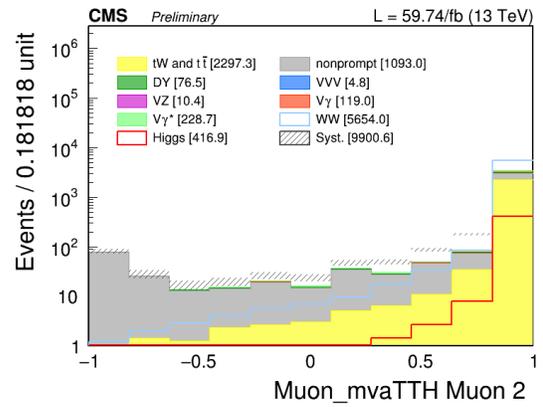
(a) Muon 1.



(b) Muon 2.



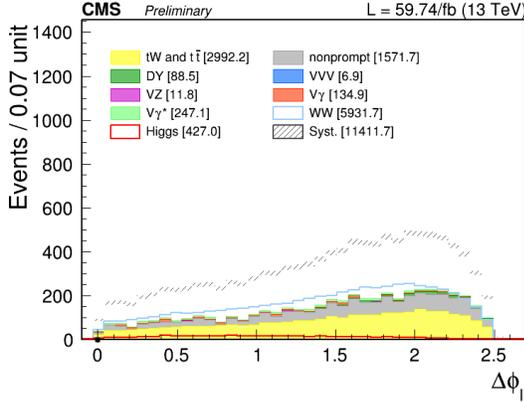
(c) Muon 1.



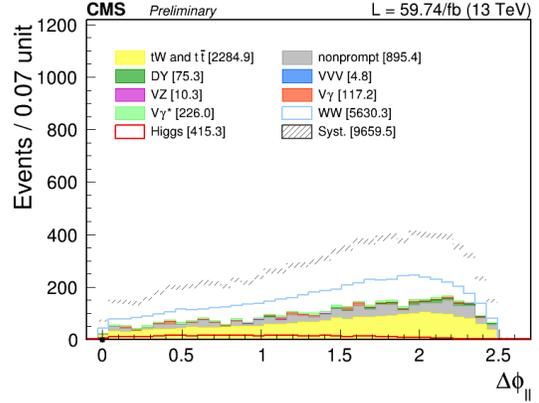
(d) Muon 2.

Figura 11: Histograma de las variables $I_{\mu}^{charged}$ y $mvATTH$ para ambos leptones.

Por último, en la Figura 12 puede observarse en la parte izquierda un histograma solo con la selección básica inicial y el derecha un histograma para la misma variable una vez que se han aplicado todos los cortes secuenciales. En estas gráficas puede verse como en la Figura 12a hay una mayor presencia de color gris (fondo *Fake*) y como se ha reducido al final de todo el proceso, habiendo menos presencia de zona gris en la Figura 12b.



(a) Con los cortes base.



(b) Después de los dos procesos de corte.

Figura 12: Diferencia en la variable $\Delta\phi_{\ell\ell}$ entre el inicio del proceso y el final.

4.7. Automatización del proceso

Uno de los problemas mencionados en los primeros cortes secuenciales es la dificultad de seleccionar la zona óptima de corte, ya que esto se ha realizado visualmente y con gráficas que al principio dificultaban la observación. Por un lado, se modificaron las gráficas para conseguir una mejor visualización de las señales y del fondo. Sin embargo, seguía siendo un poco difícil la selección de un valor óptimo de corte, por lo que había que realizar varios cortes en una misma ejecución para localizar el valor que maximizaba la figura de mérito. Esto se traduce a un gasto de tiempo y de recursos computacionales enormes, provocando que el proceso tardase en realizarse una media de 4 días por variable (un total de un mes y medio teniendo en cuenta las variables donde se han estudiado la posibilidad de hacer cortes secuenciales).

Como alternativa a este proceso manual se ha realizado un *script* en python, referido como **Proceso automático**, que puede verse con detalle en el Apéndice 4, totalmente complementario al *script* **Tabla de sucesos** que calcula la figura de mérito para los cortes seleccionados, donde introduciendo el nombre de la variable donde se quiere realizar el siguiente corte y si se va a utilizar la parte derecha o izquierda de la gráfica, te selecciona el *bin* óptimo de corte del histograma y te da el primer valor de la variable donde comienza el *bin* para meterlo dentro del corte.

En la Figura 15 se muestra a la izquierda la evolución de la figura de mérito en función del valor inicial del *bin* correspondiente en la variable *mpmet*. A la derecha de esta misma figura se observa el histograma de esta variable, en el cual hay que destacar que hay 100 *bines*. Observando la figura de la izquierda se ve que para un valor de 85 GeV (más o menos porque hay pequeñas oscilaciones que no pueden observarse con esta escala) la figura de mérito alcanza un valor límite. Eso significa que el *bin* que tenga el valor 85 GeV como primer valor será el *bin* óptimo de corte del histograma.

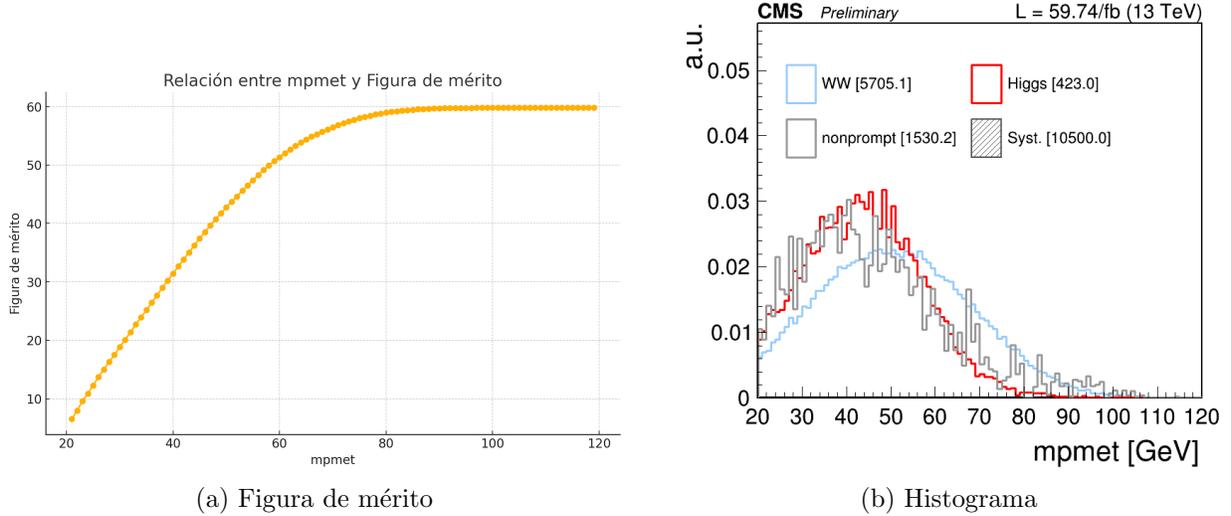


Figura 13: Evolución de la figura de mérito en función de la variable $mpmet$ y su correspondiente histograma.

Este proceso automatizado asegura hacer cortes mucho más optimizados y disminuyendo de una forma muy notoria el tiempo y los recursos computacionales del CERN, reduciendo el tiempo medio de cada corte a un día (el proceso total se ha reducido a aproximadamente 12 días).

El proceso se realiza como sigue. En primer lugar se ejecutan los *jobs* para obtener un primer archivo *root* con esa información. En los datos que se darán a continuación, se ha comenzado con la selección básica, explicada con detalle en el Apéndice 3. Una vez se obtiene el archivo *root*, se ejecuta el *script* **Tabla de sucesos** para conocer la cantidad de sucesos de cada tipo y el valor del optimizador. Una vez conocidos estos valores, se ejecuta el *script* **Proceso automático** y aparece una tabla como la que se puede ver en la Figura 21 del Apéndice 4. Por último, se coge el valor del eje x óptimo y se añade al archivo de cortes, volviendo a comenzar el proceso.

Para ilustrar este nuevo proceso automatizado se van a volver a realizar los primeros cortes secuenciales para poder comparar los resultados obtenidos en ambos casos.

En la Tabla 7 se muestran los resultados obtenidos de cada valor de corte en la variable, así como la cantidad de sucesos HWW, WW, de fondo y la figura de mérito.

Variable	Corte	HWW	WW	Fondos	\mathcal{S}
Inicial		427	5930	5050	59.52
$p_T^{\ell\ell}$	<94.23 GeV	99.1 %	96.6 %	88.1 %	59.74
$m_T^{\ell\ell+\cancel{E}_T}$	<204.1 GeV	99.1 %	96.5 %	87.5 %	59.77
puppijet	<117.4 GeV	99.1 %	96.1 %	86.7 %	59.80
mpmet	<102.7 GeV	99.1 %	96.1 %	86.3 %	59.79
$m_T^{\ell_2+\cancel{E}_T}$	<139.9 GeV	99.1 %	96.1 %	86.3 %	59.80
$\Delta\phi_{\ell\ell,\cancel{E}_T}$	>1.8462	98.1 %	95.6 %	83.3 %	59.88
$p_T^{\ell_1}$	<113.9 GeV	98.1 %	95.5 %	83.3 %	59.89
$p_T^{\ell_2}$	<94.08 GeV	98.1 %	95.5 %	83.2 %	59.89
$m_{\ell\ell}$	<167.6 GeV	98.1 %	95.5 %	83.2 %	59.89

Tabla 7: Cortes secuenciales automatizados en las variables iniciales.

Por otro lado, en la Figura 14 se muestra la comparación entre los valores de la figura de mérito entre los dos procesos. En el eje x se marca el número de paso del proceso, siendo el paso 0 el valor inicial (selección básica), el paso 1 el corte en la variable $p_T^{\ell\ell}$, etc. En esta figura se ve claramente una mejora en el valor de la figura de mérito, a pesar que de nuevo, las limitaciones no permiten maximizar la \mathcal{S} más.

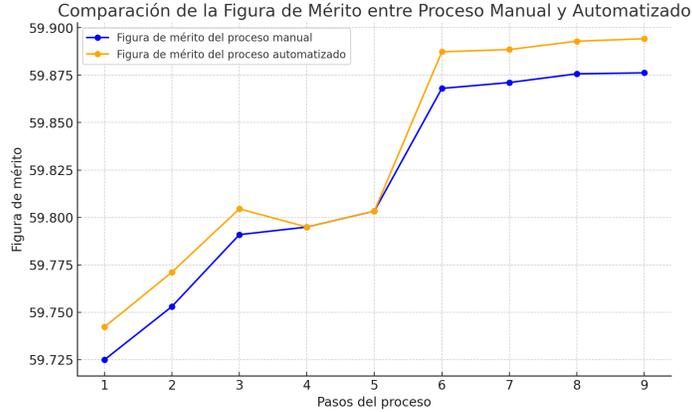


Figura 14: Gráfica donde se muestra la comparación de la figura de mérito entre el proceso manual y el proceso automático.

Comparando estos resultados con los obtenidos con el procedimiento visual y manual se pueden observar como hay cortes muy similares y cortes que se alejan de los elegidos. Por ejemplo, para la variable $m_T^{\ell\ell+\cancel{E}_T}$ hay una pequeña diferencia entre el primer corte seleccionado, que fue el de 207 GeV y el corte seleccionado en el proceso automatizado, que es de 204.1 GeV. Una diferencia un poco mayor a la anterior puede verse en la variable $m_{T^{\text{mpmet}}}$, ya que inicialmente se seleccionó como corte 113 GeV y el proceso automático ha mostrado que era mejor corte el valor 102.7 GeV. Sin embargo, destaca también la buena selección de los valores de cortes en las variables $p_T^{\ell\ell}$, siendo el valor manual de 95 GeV y el automático de 94.23 GeV, y el $m_T^{\ell_2+\cancel{E}_T}$, siendo el valor inicial de 140 GeV y el valor automático de 139.9 GeV.

Cabe destacar que en el proceso automático se puede jugar con la cantidad de *bines* de se añaden para cada variable, haciendo que la selección del valor óptimo sea tan precisa como quiera. En este caso los *bines* se han seleccionado para que estuvieran entre 1000 y 10000, en función de la variable, haciendo que se mire las décimas o las centésimas en cada caso. Sin embargo, al igual que pasaba en el proceso manual, este proceso puede ser sensible a las fluctuaciones estadísticas en las colas.

Centrando el foco en la figura de mérito, se observa que la mejora no es llamativa, ya que se ha pasado de tener una figura de mérito como el proceso manual de 59.87 a una figura de mérito de 59.89 con el proceso automático. A pesar de no ha sido una gran mejora, hay que destacar que el objetivo principal de esta automatización es generar un archivo de código que facilite la labor de búsqueda de cortes óptimos y que reduzca, muy notablemente en este caso, el tiempo y coste computacional en la red del CERN.

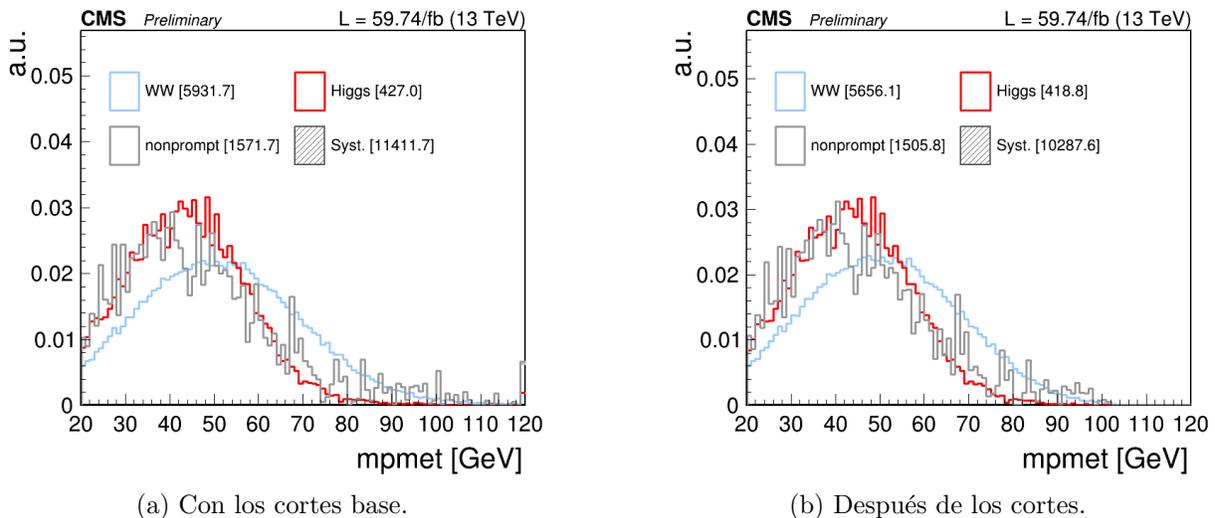


Figura 15: Diferencia en la variable m_{pmet} entre el inicio del proceso y el final.

La Figura 15 muestra el cambio en la variable $mpmet$ entre el histograma obtenido con los cortes básicos iniciales y con el último corte del proceso automático. Se puede observar una diferencia entre ambas gráficas a partir del valor aproximado de 100 GeV. Estas gráficas sirven también para ilustrar los cambios realizados en las gráficas para mejorar el problema visual de cortes que se tenía en el proceso manual. Estas gráficas solo muestran los procesos HWW , WW y $Fake$ de forma independiente y normalizada, haciendo más sencilla la búsqueda de la zona óptima de corte. Tanto estas gráficas, como las que se pueden observar en la subsección de Fondo $Fake$ han sido usadas como alternativa y mejora de los primeros histogramas mostrados.

5. Conclusiones

En este trabajo se ha hecho un análisis de simulaciones Monte Carlo en el detector CMS. En este análisis se han definido como señal los eventos englobados en los procesos del Higgs y WW, denominando el resto de procesos como fondo.

De todos estos fondos se ha centrado el estudio en el fondo *Fake*, caracterizado por eventos con muones *nonprompt* y con partículas no leptónicas mal identificados como leptones. Este fondo es de gran importancia porque, a pesar de no ser el fondo principal (el fondo *Top* tiene un mayor número de eventos), es un fondo instrumental que requiere de técnicas avanzadas para su caracterización y reducción, ya que no es fondo relacionado con un proceso físico como podría ser el fondo DY.

Para cuantificar el proceso seguido se ha usado una figura de mérito muy usada en análisis de altas energías, cuando las señales son pequeñas. Esta figura de mérito penaliza mucho la pérdida de señal, aunque sea pequeña, lo que tiene sus ventajas y puntos débiles. Por un lado, esta penalización asegura mantener la máxima señal posible, pero sin embargo, obliga a realiza cortes muy suaves evitando algunas veces la eliminación de sucesos de fondo. Con la configuración inicial, la figura de mérito tiene un valor de 59.52. Con el primer procedimiento de cortes secuenciales en las variables iniciales se ha conseguido una mejora del 0.5%. Este aumento prácticamente nulo es debido a la dificultad de apreciar visualmente las zonas óptimas de corte y a la penalización de la figura de mérito con la pérdida de señal.

Al centrarse el estudio en el fondo *Fake*, y más concretamente en los muones *nonprompt*, se han conseguido representaciones de variables con un gran poder discriminante entre el fondo *Fake* y la señal. Estas variables son $I_{\mu}^{charged}$ y mvaTTH. Al realizar el procedimiento de cortes secuenciales en estas variables se ha conseguido un aumento muy pequeño, de alrededor del 3%. A pesar de haber conseguido un aumento de tres veces el conseguido con las variables de leptones, sigue siendo un aumento muy pequeño.

Este pequeño aumento tiene múltiples razones, siendo principalmente causado por la penalización de la figura de mérito. Por otro lado, debido a que el fondo *Fake* no es el fondo principal y el fondo *Top* tiene una enorme cantidad de eventos, a pesar de conseguir una gran eliminación del fondo *Fake* la figura de mérito apenas se ve afectada.

Los resultados importantes obtenidos en este trabajo es la reducción de un 43% de los procesos del fondo *Fake*, con la reducción de tan solo un 5% de la señal. Estos datos garantizan una señal mucho más limpia de fondo *Fake* para futuros análisis de características intrínsecas del bosón de Higgs.

Finalmente, se ha elaborado un *script* de Python que ha conseguido reducir el tiempo del procedimiento de cortes secuenciales en un 75 %, el coste computacional y ha conseguido obtener cortes mucho más precisos. A pesar de que la mejora de la figura de mérito obtenida finalmente ha sido tan solo del 0.03 %, esta automatización del proceso ha sido un éxito, ya que se ha generado código que puede ser usado en otros estudios que requieran de un proceso de cortes similar al que se ha utilizado en este trabajo.

Una posible mejora de este procedimiento podría involucrar el uso de algoritmos de *machine learning*, como el *Random Forest*, para la selección de cortes en las diferentes variables. Estos algoritmos pueden manejar grandes cantidades de datos y seleccionar automáticamente los puntos de corte óptimos en múltiples variables simultáneamente, lo que podría aumentar la eficiencia de la discriminación entre señal y fondo. El *Random Forest*, al combinar la salida de múltiples árboles de decisión, es especialmente robusto contra el sobre ajuste y puede proporcionar una selección de variables más refinada y precisa.

Además, los algoritmos de *machine learning*, como las redes neuronales, pueden utilizarse para la creación de nuevas variables que sean combinaciones óptimas de las existentes. La implementación de estas técnicas podría resultar en una variable compuesta que maximice la separación entre los eventos de señal y fondo, mejorando significativamente la sensibilidad del análisis y la figura de mérito \mathcal{S} .

Referencias

- [1] CDF Collaboration. “High-precision measurement of the W boson mass with the CDF II detector”. En: *Science* 376.6589 (2022), págs. 170-176. DOI: [10.1126/science.abk1781](https://doi.org/10.1126/science.abk1781). URL: <https://www.science.org/doi/abs/10.1126/science.abk1781>.
- [2] Fermilab Collaboration. *Muon g-2 doubles down with latest measurement, explores uncharted territory in search of new physics*. [Online; accedido Agosto-2024]. 2023. URL: <https://www.washington.edu/news/2023/08/10/muon-g-2-2023/>.
- [3] The CMS Collaboration. “A portrait of the Higgs boson by the CMS experiment ten years after the discovery.” En: *Nature* 607 (2022), 60–68. DOI: <https://doi.org/10.1038/s41586-022-04892-x>.
- [4] cerncourier. *The Nobel path to a unified electroweak theory*. [Online; accedido Agosto-2024]. URL: <https://cerncourier.com/a/the-nobel-path-to-a-unified-electroweak-theory/>.
- [5] Max Bañados e Ignacio Reyes. “A short review on Noether’s theorems, gauge symmetries and boundary terms”. En: *International Journal of Modern Physics D* 25.10 (ago. de 2016), pág. 1630021. ISSN: 1793-6594. DOI: [10.1142/S0218271816300214](https://doi.org/10.1142/S0218271816300214). URL: <http://dx.doi.org/10.1142/S0218271816300214>.
- [6] ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. En: *Physics Letters B* 716.1 (oct. de 2012), 1–29. ISSN: 0370-2693. DOI: [10.1016/j.physletb.2012.08.020](https://doi.org/10.1016/j.physletb.2012.08.020). URL: <http://dx.doi.org/10.1016/j.physletb.2012.08.020>.
- [7] CMS collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. En: *Physics Letters B* 716.1 (oct. de 2012), 30–61. ISSN: 0370-2693. DOI: [10.1016/j.physletb.2012.08.021](https://doi.org/10.1016/j.physletb.2012.08.021). URL: <http://dx.doi.org/10.1016/j.physletb.2012.08.021>.
- [8] *The Standard Model*. [Online; accedido Mayo-2024]. URL: <https://home.cern/science/physics/standard-model>.
- [9] Wikipedia. *Fermión — Wikipedia, La enciclopedia libre*. [Online; accedido Agosto-2024]. 2024. URL: <https://es.wikipedia.org/w/index.php?title=Fermi%C3%B3n&oldid=160226102>.
- [10] *La UNESCO y el CERN: historia de una cooperación*. [Online; accedido Mayo-2024]. URL: <https://courier.unesco.org/es/articles/la-unesco-y-el-cern-historia-de-una-cooperacion>.
- [11] *WHAT IS CMS?* [Online; accedido Mayo-2024]. URL: <https://cmsexperiment.web.cern.ch/news/what-cms#:~:text=CMS%20is%20a%20particle%20detector,the%20heart%20of%20the%20collision>.
- [12] *Gender diversity on the right track at CERN*. [Online; accedido Agosto-2024]. URL: <https://home.cern/news/news/cern/gender-diversity-right-track-cern>.

- [13] Alessandra Ciocio Lucie de Nooij Monica Dunford Paula Eerola Despina Hatzifotiadou Kerstin Jon-And Danièle Lajust Chiara Mariotti Morna Robillard Sarah Seif El Nasr Storey Claudia-Elisabeth Wulz Pauline Gagnon (editor) Doris Chromek-Burckhart y Chiara Zampolli. *Women in science through the decades*. [Online; accedido Agosto-2024]. URL: <https://cerncourier.com/a/women-in-science-through-the-decades/>.
- [14] ¿Qué es el Gran Colisionador de Hadrones y por qué cambió la historia de la física? [Online; accedido Agosto-2024]. URL: <https://iberomx.com/prensa/que-es-el-gran-colisionador-de-hadrones-y-por-que-cambio-la-historia-de-la-fisica#:~:text=Enterrado%20en%20un%20t%C3%BAnel%20de,marcha%20hasta%20noviembre%20de%202009..>
- [15] Wikipedia. *Protón* — *Wikipedia, La enciclopedia libre*. [Online; accedido Agosto-2024]. 2024. URL: <https://es.wikipedia.org/w/index.php?title=Prot%C3%B3n&oldid=162082356>.
- [16] *LHC report: full house for the LHC*. [Online; accedido Agosto-2024]. URL: <https://home.cern/news/news/accelerators/lhc-report-full-house-lhc>.
- [17] *The Large Hadron Collider*. [Online; accedido Mayo-2024]. URL: <https://home.cern/science/accelerators/large-hadron-collider>.
- [18] Wikipedia. *Gran colisionador de hadrones*. [Online; accedido Agosto-2024]. 2024. URL: https://es.wikipedia.org/wiki/Gran_colisionador_de_hadrones.
- [19] S. Chatrchyan et al. “The CMS Experiment at the CERN LHC”. En: *JINST* 3 (2008), S08004. DOI: [10.1088/1748-0221/3/08/S08004](https://doi.org/10.1088/1748-0221/3/08/S08004).
- [20] Pallabi Das y on behalf of the CMS Collaboration. “An overview of the trigger system at the CMS experiment”. En: *Physica Scripta* 97.5 (abr. de 2022), pág. 054008. DOI: [10.1088/1402-4896/ac6302](https://doi.org/10.1088/1402-4896/ac6302). URL: <https://dx.doi.org/10.1088/1402-4896/ac6302>.
- [21] CMS collaboration. “Measurements of the Higgs boson production cross section and couplings in the W boson pair decay channel in proton-proton collisions at $\sqrt{s} = 13\text{ TeV}$ ”. En: *The European Physical Journal C* 83.7 (jul. de 2023). ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-023-11632-6](https://doi.org/10.1140/epjc/s10052-023-11632-6). URL: <http://dx.doi.org/10.1140/epjc/s10052-023-11632-6>.
- [22] CMS collaboration. “The CMS trigger system”. En: *Journal of Instrumentation* 12.01 (ene. de 2017), P01020–P01020. ISSN: 1748-0221. DOI: [10.1088/1748-0221/12/01/p01020](https://doi.org/10.1088/1748-0221/12/01/p01020). URL: <http://dx.doi.org/10.1088/1748-0221/12/01/p01020>.
- [23] E. Brondolin. “Track reconstruction in the CMS experiment for the high luminosity LHC [Dissertation, Technische Universität Wien].” En: *repositUM*. (2017), 36–43. DOI: <https://doi.org/10.34726/hss.2017.30764>.
- [24] CMS. *Identification of prompt and isolated muons using multivariate techniques at the CMS experiment in proton-proton collisions at $\sqrt{s} = 13\text{ TeV}$* . Inf. téc. Geneva: CERN, 2023. URL: <http://cds.cern.ch/record/2859395>.

- [25] Mark Matthewman. *Kalman filter for muon reconstruction in the CMS Phase-2 endcap calorimeter*. Inf. téc. Geneva: CERN, 2023. URL: <https://cds.cern.ch/record/2889010>.
- [26] CMSb Collaboration. *Jets at CMS and the determination of their energy scale*. [Online; accedido Agosto-2024]. URL: <https://cms.cern/news/jets-cms-and-determination-their-energy-scale/>.
- [27] Imobach García Ibañez. “Etiquetado de b-jets en el experimento CMS del LHC”. En: (feb. de 2023). URL: <https://repositorio.unican.es/xmlui/handle/10902/29597>.
- [28] IBM. *What is Monte Carlo Simulation?* [Online; accedido Julio-2024]. URL: <https://www.ibm.com/topics/monte-carlo-simulation>.
- [29] “CMS pythia 8 colour reconnection tunes based on underlying-event data”. En: *The European Physical Journal C* 83.7 (jul. de 2023). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-023-11630-8. URL: <http://dx.doi.org/10.1140/epjc/s10052-023-11630-8>.
- [30] *MadGraph Interface to CMSSW*. [Online; accedido Agosto-2024]. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMadGraphInterface>.
- [31] “Development and validation of HERWIG 7 tunes from CMS underlying-event measurements”. En: *The European Physical Journal C* 81.4 (abr. de 2021). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-08949-5. URL: <https://hdl.handle.net/10495/176625>.
- [32] *Scripts for Produce and Run a Precompiled POWHEG Tarball*. [Online; accedido Agosto-2024]. URL: <https://twiki.cern.ch/twiki/bin/view/Main/YuanChaoPowHeg>.
- [33] *CMS Physics Analysis project - Geant4*. [Online; accedido Agosto-2024]. URL: <https://twiki.cern.ch/twiki/bin/view/Main/CmsGeant4Cascade>.
- [34] Mark Thomson. *Modern Particle Physics*. Cambridge University Press, 2013.
- [35] Alexander Feike. “Drell-Yan Process in the Standard Model and Beyond”. En: (sep. de 2021). URL: https://www.uni-muenster.de/imperia/md/content/physik_tp/theses/klasen/feike_bsc.pdf.
- [36] *Producción de dos Bosones vectoriales débiles, $W+W^-$, en el LHC*. [Online; accedido Agosto-2024]. URL: <https://ifca.unican.es/es-es/Paginas/Produccion-de-dos-Bosones-vectoriales-debiles-W+W--en-el-LHC.aspx#:~:text=%C2%B7E1%20bos%C3%B3n%20W%20se%20desintegra,y%20sus%20compa%C3%B1eros%20los%20neutrinos%20..>
- [37] Agnieszka Leyko. “Measurement of the top quark decay branching fraction ratio $R_b = B(t \rightarrow Wb)/B(t \rightarrow Wq)$ and of the inclusive production cross section $pp \rightarrow tt$ in the dilepton channel at ATLAS”. En: (2020). URL: <https://cds.cern.ch/record/2790810/files/CERN-THESIS-2020-375.pdf>.

- [38] The DELPHI Collaboration. “Measurement of the W-pair production cross-section and W branching ratios in e^+e^- collisions at $\sqrt{s}=161\text{-}209$ GeV”. En: *Eur. Phys. J. C* 12.34 (mayo de 2004), 127–144. ISSN: 1748-0221. DOI: <https://doi.org/10.1140/epjc/s2004-01709-5>. URL: <https://link.springer.com/article/10.1140/epjc/s2004-01709-5#citeas>.
- [39] Emanuele Di Marco Adinda de Wit. *CMS Higgs Physics Results*. [Online; accedido Julio-2024]. URL: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsHIG>.
- [40] Daniele Bertolini et al. “Pileup per particle identification”. En: *Journal of High Energy Physics* 2014.10 (oct. de 2014). ISSN: 1029-8479. DOI: [10.1007/jhep10\(2014\)059](https://doi.org/10.1007/jhep10(2014)059). URL: [http://dx.doi.org/10.1007/JHEP10\(2014\)059](http://dx.doi.org/10.1007/JHEP10(2014)059).
- [41] Nicolò Trevisani. “Búsqueda de materia oscura producida junto a un bosón de higgs en el canal de desintegración a dos bosones W+W en colisiones de protones a $s=13$ TeV de energía del centro de masa en el LHC con el experimento CMS.” En: (mayo de 2019). URL: <http://hdl.handle.net/10902/16360>.
- [42] *Baseline muon selections for Run-I*. [Online; accedido Agosto-2024]. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideMuonId#Tight_Muon_selection.
- [43] CMS Collaboration. “Measurement of W^+W^- inclusive and differential cross sections in pp collisions at $\sqrt{s} = 13.6$ TeV with the CMS detector”. En: (mar. de 2024). DOI: [CMSPASSMP-24-001](https://doi.org/10.1007/JHEP03(2024)001). URL: <https://cms-results.web.cern.ch/cms-results/public-results/preliminary-results/SMP-24-001/index.html>.
- [44] Markus Stoye y on behalf of the CMS collaboration. “Deep learning in jet reconstruction at CMS”. En: *Journal of Physics: Conference Series* 1085.4 (2018), pág. 042029. DOI: [10.1088/1742-6596/1085/4/042029](https://doi.org/10.1088/1742-6596/1085/4/042029). URL: <https://dx.doi.org/10.1088/1742-6596/1085/4/042029>.
- [45] Laura V. Trujillo Taborda. *BÚSQUEDA DE PARTÍCULAS DE LARGA VIDA EN EL LHC*. [Online; accedido Julio-2024]. URL: <https://digital.csic.es/bitstream/10261/234675/1/TFM-LauraVTrujilloTaborda.pdf>.
- [46] *Background Modelling Techniquel*. [Online; accedido Agosto-2024]. URL: <https://cms-opendata-guide.web.cern.ch/analysis/backgrounds/>.
- [47] ANTONIO MÁRQUEZ GARCÍA. “Redes neuronales profundas en búsquedas relacionadas con la producción de pares de quarks top en el CMS”. En: (). URL: https://www.hep.uniovi.es/jfernand/TFG/TFG2019_Antonio_Version9.pdf.
- [48] Izaak Neutelings. *CMS coordinate system*. [Online; accedido Julio-2024]. URL: https://tikz.net/axis3d_cms/.

Apéndices

Apéndice 1

El sistema de coordenadas adoptado por CMS tiene el origen centrado en el punto de colisión dentro del experimento, el eje y apuntando verticalmente hacia arriba, y el eje x apuntando radialmente hacia el centro del LHC. Así, el eje z apunta a lo largo de la dirección del haz.

El ángulo acimutal ϕ se mide desde el eje x en el plano xy y la coordenada radial en este plano se denota por r . El ángulo polar θ se mide desde el eje z . La pseudorapidez se define como $\eta = -\ln \tan(\theta/2)$. Así, el momento y la energía transversales a la dirección del haz, denotados por p_T y E_T , respectivamente, se calculan a partir de las componentes x e y . El desequilibrio de energía medido en el plano transversal se denota por \cancel{E}_T [19].

En la Figura 16 puede verse un esquema de las coordenadas utilizadas.

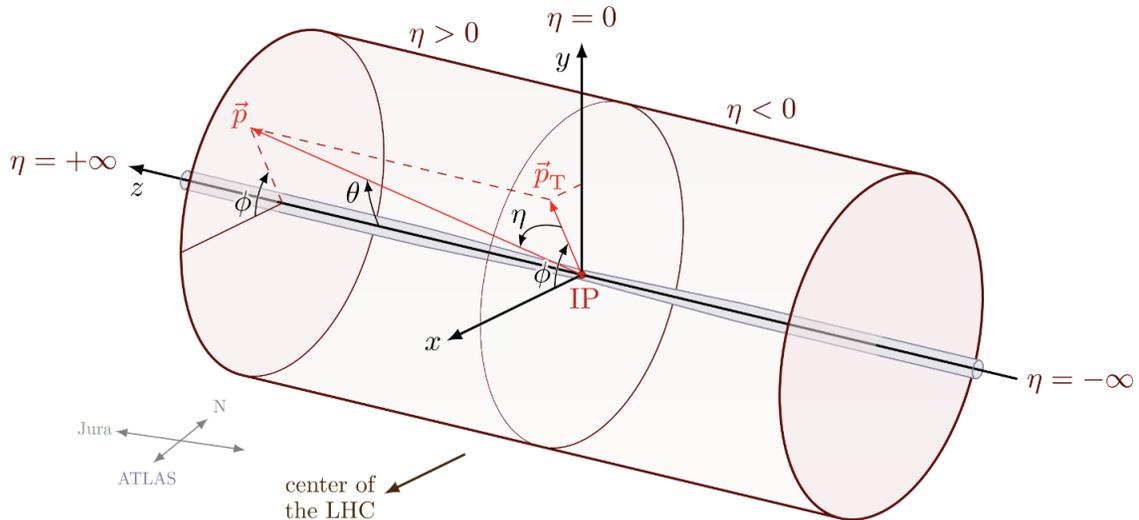


Figura 16: Sistema de coordenadas del experimento CMS [48].

Apéndice 2

En este apéndice se va a exponer los diferentes conocimientos adquiridos a lo largo del proceso de creación de este trabajo.

En primer lugar, para el acceso a los datos simulados y las máquinas del CERN ha sido necesario aprender comandos básicos de Linux. Dentro de estos comandos cabe destacar los comandos de cambio de carpetas (**cd *path*** para entrar en la carpeta, **cd ..** para moverte a la carpeta anterior), el comando de eliminación de ficheros **rm** (**rm *** elimina todo lo que haya en la carpeta donde estas situado, **rm *path*** elimina todo lo que haya dentro de la carpeta a la que se acceda con ese *path* y **rm *condición*** elimina los archivos que en su nombre tengan la condición específica) y el comando de cambio de nombre o ubicación de un fichero **mv** (**mv *fichero fichero1*** cambia el nombre del fichero por el que se ponga en *fichero1*, **mv *fichero path*** mueve el fichero especificado al lugar que indique el *path*).

Por otro lado, para el estudio de los ficheros donde se encuentran las simulaciones es necesario aprender unos comandos básicos de root. Estos comandos incluyen el de apertura de un fichero (**root -l *fichero***), el de acceso a un árbol (**(TTree*)_file0->Get(" *Arbol*")**), el comando para pintar una variable (suponiendo que el archivo tiene un árbol llamado Events, **Events->Draw("variable", *condición*)**), si en el comando anterior utilizamos Scan en vez de Draw obtendremos una tabla con los valores de la variable y por último, el comando para obtener todas las variables que se encuentran en un archivo root (**Events->MakeClass(" *Archivo*")** donde *Archivo* es el nombre del archivo donde se van a escribir las variables).

Una vez ya se han explorado los archivos con los que se va a trabajar y se tienen integrados los comandos básicos de Linux se procede a la generación de gráficas y la selección de cortes. Para este paso ha sido necesario estudiar y entender el código proporcionado para la selección de procesos y creación de las figuras, ya que es necesario hacer cambios en él para adaptarlo a las visualizaciones y cortes que se requieran.

Por último, para la selección de cortes es necesario conocer y estudiar las variables que se tienen disponibles en los ficheros y entender la naturaleza de las gráficas.

Apéndice 3

Con el fin de maximizar los procesos relacionados con la señal y minimizar el fondo se van a ir realizando cortes a una serie de variables seleccionadas.

Los primeros cortes que vienen ya seleccionados en el código aparecen en la Figura 17.

```
'Lepton_pdgId[0]*Lepton_pdgId[1] == -11*13',
'Lepton_pt[0] > 25.',
'Lepton_pt[1] > 20.',
'(nLepton >= 2 && Alt(Lepton_pt,2, 0) < 10.)',
'mll > 12',
'mpmet > 20.',
'PuppiMET_pt > 20.',
]
```

Figura 17: Primera selección básica de procesos.

La primera línea de la selección restringe los procesos a aquellos que tengan un muón y un electrón de carga contraria. Esta selección minimiza el fondo de Drell-Yan al evitar procesos de creación de dos leptones de mismo sabor. La segunda y tercera línea selecciona procesos donde el momento transversal del primer leptón sea mayor de 25 y del segundo leptón mayor de 20 respectivamente. La cuarta selección restringe los procesos a aquellos que tengan o solo dos leptones o si tienen más, que el momento transversal del tercer leptón sea menor de 10. La siguiente línea solo permite el paso de procesos en los que la masa invariante de los dos leptones sea mayor de 12. La ante última y ultima línea son variables relacionadas con la energía perdida, eligiendo procesos con energía transversa perdida (mpmet) mayor de 20 y con momento transversal perdido calculado con mitigación de *pileup* (puppiMET) mayor de 20. El término *pileup* hace referencia a choques múltiples que se producen en el detector y que producen ruido en las medidas de energía perdida, por lo que el algoritmo *Pileup Per Particle Identification* (PUPPI) tiene en cuenta este ruido en el calculo de la energía perdida en las variables Puppi.

Añadido a esta selección, se hace unos cortes extra cogiendo la suma del momento transversal de los dos leptones (p_T^l) mayor de 30, que la masa transversa ($m_T^{l+\cancel{E}_T}$) sea mayor de 50 y que la distancia entre los dos leptones (ΔR_{ll}) sea mayor de 2.5.

Estos cortes pueden verse en la Figura 18.

```
cuts['dhww2l2v_13TeV_sr'] = {
  'expr': 'ptll>30 && mth > 50 && drll < 2.5',
}
```

Figura 18: Primeros cortes.

Apéndice 4

Se ha creado un *script* de python, referido como **Tabla de sucesos**, para agilizar el proceso de selección de los valores de los cortes.

El código completo del *script* **Tabla de sucesos** puede verse en la Figura 20.

Este código mezcla los lenguajes de python (para la creación de la función, los bucles, etc.) con lenguaje Root. La función definida, denominada 'archivo_root', coge la ruta del archivo que se quiere leer, lo abre y lee todos los directorios contenidos en él. De estos directorios solo entra en aquellos que contengan '0j' en su nombre, ya que son los directorios correspondientes a la categoría 0j que es de interés en este trabajo.

Dentro de estos directorios se accede al árbol de eventos, donde se encuentran los diferentes procesos con la cantidad de eventos producidos de cada uno. Este código imprime en pantalla cada proceso con su cantidad de eventos y calcula, además, la suma total de todos los procesos y el \mathcal{S} .

En la Figura 19 se muestra un ejemplo de ejecución de este código.

```
(myenv) [asainzbe@lxplus945 Full2018_v7]$ python tabla_histograma.py
Abriendo archivo: rootFiles__darkHiggs2018_v7/mkShapes__darkHiggs2018_v7.root
El archivo se ha abierto correctamente.
Se ha encontrado el directorio: dhww2l2v_13TeV_sr_nstation1_0j
Se ha accedido al directorio 'events' en: dhww2l2v_13TeV_sr_nstation1_0j
Directorio: dhww2l2v_13TeV_sr_nstation1_0j
Nombre histograma      Total Sucesos
histo_DY                75.32276108754961
histo_top               2282.641371085601
histo_WW                4990.207244895232
histo_WWewk             5.080242050787206
histo_ggWW              629.5479349335488
histo_Vg                117.1668435436127
histo_VgS               225.9763125804583
histo_VZ                10.261216236338893
histo_VVV               4.7794566298142085
histo_Higgs             414.7815186450584
histo_Fake              893.0031367987394
histo_DATA              16230.0
WWTOTAL 5624.8354218795685
SISTEMATICO            9648.768038486742
OPTIMIZADOR            61.48560655014881
```

Figura 19: Ejemplo de ejecución del código para el calculo de la figura de mérito.

```

import ROOT
import math

def archivo_root(file_path):
    # Abrimos el archivo root
    print(f"Abriendo archivo: {file_path}")
    file = ROOT.TFile.Open(file_path)
    if not file or file.IsZombie():
        print("Error al abrir el archivo.")
        return
    print("El archivo se ha abierto correctamente.")

    # Sacamos todos los directorios que hay dentro del archivo
    keys = file.GetListOfKeys()

    # Recorremos todos los directorios
    for key in keys:
        dir = key.ReadObj()

        # Miramos si el nombre del directorio contiene "_0j" pero no "0jemu" o "0jmue" (no nos interesan)
        dir_name = dir.GetName()
        if "_0j" in dir_name and "0jemu" not in dir_name and "0jmue" not in dir_name:
            print(f"Se ha encontrado el directorio: {dir_name}")

            # Entramos al directorio Events, donde está la información de los histogramas
            events_dir = dir.Get("events")
            if not events_dir:
                print(f"Error al acceder a 'events' en el directorio principal {dir_name}")
                continue
            print(f"Se ha accedido al directorio 'events' en: {dir_name}")

            # Obtenemos los nombres de los histogramas del directorio Events
            hist_keys = events_dir.GetListOfKeys()

            # Inicializamos las variables que nos van a interesar
            WWTOTAL = 0.0
            SISTEMATICO = 0.0
            DATA = 0.0

            # Diccionario para guardar todos los histogramas del directorio Events
            hist_totals = {}

            # Recorremos todos los histogramas
            for hist_key in hist_keys:
                hist = hist_key.ReadObj()
                if not hist or not isinstance(hist, ROOT.TH1):
                    print(f"Error en el histograma {hist_key.GetName()}")
                    continue

                # Obtener la suma de sucesos (integral) de cada histograma
                total_sum = hist.Integral()
                hist_name = hist.GetName()
                hist_totals[hist_name] = total_sum

                # Calculamos el total de sucesos WW
                if hist_name in ["histo_WW", "histo_WWewk", "histo_ggWW"]:
                    WWTOTAL += total_sum

                # Calculamos el total de sucesos entre fondo y señal
                if hist_name != "histo_DATA":
                    SISTEMATICO += total_sum

                # Calculamos nuestro OPTIMIZADOR
                HIGGS = hist_totals.get("histo_Higgs", 0.0)
                OPTIMIZADOR = (HIGGS + WWTOTAL) / math.sqrt(SISTEMATICO) if SISTEMATICO > 0 else 0.0

                # Mostramos la tabla en pantalla
                print(f"Directorio: {dir_name}")
                print("Nombre histograma\tTotal Sucesos")
                for hist_name, total_sum in hist_totals.items():
                    print(f"{hist_name}\t{total_sum}")
                print(f"WWTOTAL\t{WWTOTAL}")
                print(f"SISTEMATICO\t{SISTEMATICO}")
                print(f"OPTIMIZADOR\t{OPTIMIZADOR}")
                print()

            # Cerramos el archivo root
            file.Close()

if __name__ == "__main__":
    archivo_root("rootFiles__darkHiggs2018_v7/mkShapes__darkHiggs2018_v7.root")

```

Figura 20: Código para calcular la figura de mérito.

Por otro lado, se ha creado un *script* de python, referido como **Proceso automático**, para automatizar el proceso de selección de los valores de los cortes y reducir el coste computacional.

El código completo del *script* **Proceso automático**, puede verse en la Figura 22.

Este código, al igual que el *script* **Tabla de sucesos**, mezcla los lenguajes de programación Python y Root. La función definida en este archivo realiza los mismos pasos iniciales que en el *script* **Tabla de sucesos**, abre el archivo definido por la ruta de entrada, mira los directorios disponibles y solo entra en los directorios que contengan 0j en su nombre (sin entrar en los que contengan 0jmue o 0jemu). A partir de aquí las dos funciones realizan pasos distintos.

En este caso, una vez esta dentro del directorio, entra en el histograma `histo_DATA` para obtener de él, el número de *bines* que tiene la variable seleccionada, y que viene especificada en la entrada de la función como 'var'. Seguido a esto, se inician las variables que son interesantes para la optimización, como son el número de bin óptimo, el valor del eje x, el valor de la figura de mérito, la cantidad de sucesos WW, la cantidad de sucesos Higgs y la cantidad de sucesos totales.

A continuación, se inicializa un bucle para ir recorriendo todos los *bines* disponibles. Una vez dentro del bucle, la función entra dentro de cada histograma disponible y calcula la cantidad de procesos para ese proceso integrando, o bien desde el *bin* 0 hasta el *bin* que indique el bucle (si en la entrada de la función se ha indicado como 'lado' la izquierda, `izq`) o bien del *bin* indicado hasta el último *bin* (si se ha indicado otra cosa en la entrada 'lado'). Esta especificación nos sirve para indicar si vamos a hacer un corte para quedarse con la parte izquierda o derecha de la gráfica.

Cuando ya se han recorrido todos los histogramas se termina de calcular la cantidad de procesos WW, Higgs y totales y se calcula la figura de mérito. Si esta figura de mérito es mayor que la que había anteriormente entonces se modifican las variables iniciales antes del bucle de *bines* con los valores actuales. En caso contrario se mantienen los valores que hubiera anteriormente.

Finalmente, cuando se termina el bucle de *bines*, se imprime en pantalla una tabla con los valores del *bin* óptimo, el valor del eje x correspondiente a ese *bin* y la cantidad de sucesos de los procesos Higgs, WW y totales, así como el valor de la figura de mérito.

Un ejemplo de lo que imprime esta función puede verse en la Figura 21.

```
[(myenv) [asainzbe@lxplus906 Full2018_v7]$ python proceso_auto.py
Abriendo archivo: rootFiles__darkHiggs2018_v7/mkShapes__darkHiggs2018_v7.root
El archivo se ha abierto correctamente.
Se ha encontrado el directorio: dhww2l2v_13TeV_sr_0j
Se ha accedido al directorio ptll en: dhww2l2v_13TeV_sr_0j
Se han contado 700 bines para la variable ptll
Bin: 642
Valor eje x: 94.10000000000001
HIGGS 423.3067875114143
WW 5726.1369023956
TOTAL 10595.119626919448
OPTIMIZADOR 59.742431839911006
```

Figura 21: Ejemplo de salida del código de automatización del proceso de cortes secuenciales.

```

import ROOT
import math

def archivo_root(file_path, var, lado):
    # Abrimos el archivo root
    print(f"Abriendo archivo: {file_path}")
    file = ROOT.TFile.Open(file_path)
    if not file or file.IsZombie():
        print("Error al abrir el archivo.")
        return
    print("El archivo se ha abierto correctamente.")

    # Sacamos todos los directorios que hay dentro del archivo
    keys = file.GetListOfKeys()

    # Recorremos todos los directorios
    for key in keys:
        dir = key.ReadObj()

        # Miramos si el nombre del directorio contiene "_0j" pero no "0jemu" o "0jmue" (no nos interesan)
        dir_name = dir.GetName()
        if "_0j" in dir_name and "0jemu" not in dir_name and "0jmue" not in dir_name:
            print(f"Se ha encontrado el directorio: {dir_name}")

            # Entramos al directorio de la variable que queremos, donde está la información de los histogramas
            var_dir = dir.Get(var)
            if not var_dir:
                print(f"Error al acceder a {var} en el directorio principal {dir_name}")
                continue
            print(f"Se ha accedido al directorio {var} en: {dir_name}")

            # Obtenemos la cantidad de bins que hay para esta variable entrando en un histograma que conocemos, histo_DATA:
            hist = var_dir.Get("histo_DATA")
            if not hist:
                print(f"Error al obtener el histograma histo_DATA en el directorio {dir_name}")
                continue
            n_bins = hist.GetNbinsX()
            print(f"Se han contado {n_bins} bins para la variable {var}")

            # Inicializamos las variables que nos interesan
            valor_bin = -1
            valor_x = -9.999
            valor_opti = 0
            valor_ww = 0
            valor_higgs = 0
            valor_total = 0

            # Recorremos todos los bins para calcular el número de sucesos y el optimizador:
            for bin in range(1, n_bins + 1):
                # Obtenemos los nombres de los histogramas del directorio (hay que recorrer todos para cada bin)
                hist_keys = var_dir.GetListOfKeys()

                # Inicializamos las variables del bucle
                WWTOTAL = 0.0
                TOTAL = 0.0
                HIGGS = 0.0

                # Diccionario para guardar todos los histogramas del directorio
                hist_totals = {}

            # Recorremos todos los histogramas

```

```

for hist_key in hist_keys:
    hist = hist_key.ReadObj()
    if not hist or not isinstance(hist, ROOT.TH1):
        print(f"Error en el histograma {hist_key.GetName()}")
        continue

    # Obtener la suma de sucesos (integral) de cada histograma hasta el número de bin seleccionado
    if lado == "izq":
        total_sum = hist.Integral(0, bin)
    else:
        total_sum = hist.Integral(bin, n_bins)
    hist_name = hist.GetName()
    hist_totals[hist_name] = total_sum

    # Calculamos el total de sucesos WW
    if hist_name in ["histo_WW", "histo_WWewk", "histo_ggWW"]:
        WWTOTAL += total_sum

    # Calculamos el total de sucesos entre fondo y señal
    if hist_name != "histo_DATA":
        TOTAL += total_sum

    # Calculamos el número de sucesos Higgs
    if hist_name == "histo_Higgs":
        HIGGS = total_sum

# Calculamos nuestro OPTIMIZADOR
OPTIMIZADOR = (HIGGS + WWTOTAL) / math.sqrt(TOTAL) if TOTAL > 0 else 0.0

if OPTIMIZADOR > valor_opti:
    valor_bin = bin
    valor_x = hist.GetBinLowEdge(bin)
    valor_opti = OPTIMIZADOR
    valor_ww = WWTOTAL
    valor_higgs = HIGGS
    valor_total = TOTAL

# Mostramos la tabla en pantalla si encontramos valores
if valor_bin != -1:
    print(f"Bin: {valor_bin}")
    print(f"Valor eje x: {valor_x}")
    print(f"HIGGS\t{valor_higgs}")
    print(f"WW\t{valor_ww}")
    print(f"TOTAL\t{valor_total}")
    print(f"OPTIMIZADOR\t{valor_opti}")
    print()

# Cerramos el archivo root
file.Close()

if __name__ == "__main__":
    archivo_root("rootFiles__darkHiggs2018_v7/mkShapes__darkHiggs2018_v7.root", "pt11", "izq")

```

Figura 22: Código para obtener, de forma automática, el valor del corte óptimo.