

*Facultad
de
Ciencias*

**Estudio de la desintegración del bosón de
Higgs en dos bosones W en el experimento
CMS**

**(Study of the Higgs boson decay into two W
bosons in the CMS experiment)**

Trabajo de Fin de Grado
para acceder al

GRADO EN FÍSICA

Autor: Santiago Pascual Sopena

Director: Jónatan Piedra Gómez

Co-Director: Jesús Manuel Vizán García

Julio 2025

Resumen

En este trabajo se ha realizado un análisis de datos basados en simulaciones 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 ha considerado como señal (S) del análisis los sucesos de producción de Higgs, mientras que el resto de sucesos se han considerado fondo (B). El objetivo principal del análisis es maximizar la relación entre la señal y el fondo, formado por diferentes procesos físicos que se pueden estimar por simulaciones Monte Carlo. La mejora de la relación entre la señal y el fondo es cuantificada por medio de una figura de mérito, $\frac{S}{\sqrt{S+B}}$. El objetivo es eliminar la mayor cantidad de fondo sin afectar a la señal de forma considerable. Se ha iniciado el análisis con un valor de la figura de mérito de 4.11, lo que se ha logrado aumentar hasta 8.58.

Palabras clave: LHC, CMS, bosón de Higgs, optimización.

Abstract

This paper presents an analysis of data based on Monte Carlo simulations of proton-proton collisions at a center of mass energy of 13 TeV, corresponding data acquired in the year 2018. The production of Higgs boson has been considered as signal (S), whereas the rest of the physical processes are considered background (B). The main goal of the study is to maximise the proportion of signal against the background, formed by different physical processes that can be estimated with Monte Carlo simulations. The improvement of the signal is defined by a figure of merit, $\frac{S}{\sqrt{S+B}}$. The main goal is to delete the most background possible without reducing the signal. The analysis began with a figure of merit of 4.11, and it increased to a value of 8.58.

Key words: LHC, CMS, Higgs boson, optimization.

Índice

1. Introducción	8
1.1. Motivación del trabajo	8
1.2. Objetivos	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	14
3.1. Organización Europea para la Investigación Nuclear	14
3.2. Gran colisionador de hadrones	15
3.3. Experimento CMS	16
3.3.1. Estructura del detector	17
3.3.2. Selección de eventos	18
3.3.3. Reconstrucción de eventos	19
4. Análisis	21
4.1. Procesos físicos	21
4.2. Variables básicas	22
4.3. Selección básica	23
4.4. Figura de mérito	24
4.5. Herramienta de análisis	24
4.6. Optimización de la señal	25
5. Resultados	30
6. Conclusiones	38
Apéndices	42

Índice de figuras

1.	Resumen de las partículas elementales conocidas, diferenciando los bosones de los fermiones, los que a su vez se dividen en leptones y quarks [4].	10
2.	Esquema del complejo de aceleradores del CERN, donde se distinguen los diferentes aceleradores con sus respectivos experimentos, destacando ATLAS, CMS, ALICE y LHCb en el LHC [18].	16
3.	Estructura del detector CMS [23].	18
4.	Histogramas correspondientes a la variable $p_T^{\ell_1}$ en el canal $e\mu$ al aplicar la selección básica.	26
5.	Histograma de curvas de la variable $p_T^{\ell_1}$ al aplicar la selección básica y representando únicamente los dos fondos principales.	27
6.	FOM frente a la variable m_T^H . Cada punto azul corresponde a el valor de la figura de mérito al aplicar el corte $m_T^H < x$, siendo x el valor del eje X en GeV.	28
7.	FOM frente a la variable m_T^H . En el caso de este variable, la FOM del corte óptimo se ha obtenido tras cortar en la variable $m_{\ell\ell}$	29
8.	Histogramas de curvas de las variables potencialmente discriminantes. Los histogramas del resto de variables están en el apéndice 1.	31
9.	FOM frente a las distintas variables discriminantes.	33
10.	Gráficas correspondientes a las distintas variables discriminantes frente a la figura de mérito obtenidas con el programa optimizacion.py.	35

Índice de tablas

1.	Canales de desintegración del bosón de Higgs y sus <i>branching ratio</i> correspondientes. .	13
2.	Cortes correspondientes con la selección básica de referencia.	23
3.	Aportación a la señal de los distintos modos de producción del Higgs en el LHC. . . .	28
4.	Valores de la señal, fondo y FOM al aplicar la selección inicial.	30
5.	Valores obtenidos al aplicar los cortes en las variables discriminantes con el primer método de análisis.	32
6.	Figura de mérito total correspondiente a los datos obtenidos con el primer método de análisis.	32
7.	Valores obtenidos al aplicar los cortes en las variables discriminantes con el segundo método de análisis.	34
8.	Figura de mérito total correspondiente a los datos obtenidos con el segundo método de análisis.	34
9.	Valores obtenidos al aplicar los cortes en las variables discriminantes con el tercer método de análisis.	36
10.	Figura de mérito total correspondiente a los datos obtenidos con el tercer método de análisis.	36
11.	Figura de mérito total correspondiente a los datos obtenidos con los tres métodos de análisis.	37
12.	Figura de mérito total calculada a partir del valor máximo de la FOM en cada canal. .	37

Agradecimientos

En primer lugar, me gustaría dar las gracias a todo el personal de la universidad, que me han acompañado desde el día que llegué. Durante estos años he podido aprender de los mejores profesores y profesoras, los cuales se han encargado de transmitir no solo enseñanzas, sino la pasión que tiene cada uno por su campo de conocimiento. Estoy muy agradecido por el tiempo, la dedicación y el entusiasmo que pone cada uno de ellos en la docencia, permitiendo al alumno disfrutar el proceso de aprendizaje.

Debo hacer una especial mención a mis tutores de este trabajo de fin de grado, Jónatan y Jesús. A Jónatan, gracias por estar siempre dispuesto a guiarme cuando no sabía cómo avanzar. Tu paciencia y tus ganas de enseñar me han impulsado más lejos de lo que me creía capaz de llegar. A Jesús, gracias por tu apoyo y ayuda. Tus consejos, siempre disponibles, me han ayudado a completar este trabajo.

Por último, quiero agradecer de corazón a mi familia por su apoyo. A mi madre, por su cariño y ánimos en los momentos que más lo necesitaba. A mi padre, por apoyarme en todo momento y por su disposición a ayudarme cuando me hacía falta. Y a mi hermanas, por animarme a seguir siempre que han tenido la oportunidad.

1. Introducción

La estructura de la materia y sus interacciones han sido un misterio durante gran parte de la historia, hasta finales del siglo XIX e inicios del siglo XX, cuando llegaron los avances científicos necesarios para poder estudiar los componentes elementales de la materia, naciendo de este modo la rama de la física que hoy se conoce como física de partículas o de altas energías. Actualmente, la teoría que mejor describe esta rama de la física es el modelo estándar de partículas, el cual predice el comportamiento de las partículas elementales al interactuar por medio de la fuerza electromagnética, la interacción fuerte y la interacción débil.

Uno de los mayores avances de la física de partículas en los últimos tiempos es el descubrimiento del bosón de Higgs, predicho teóricamente en 1964 por Peter Higgs, Robert Brout y François Englert. Este bosón es una de las piezas clave del modelo, ya que su descubrimiento en 2012 permite confirmar el mecanismo de Brout-Englert-Higgs y la existencia del campo de Higgs, ambos necesarios para la unificación de las fuerzas débil y electromagnética en una única teoría de campos. A pesar de esto, es necesario continuar estudiando las características del bosón de Higgs y mejorar la precisión en sus medidas para un mejor entendimiento de la física relacionada con el mismo [1].

1.1. Motivación del trabajo

El modelo estándar, incluso con todos los avances recientes, sigue siendo una teoría incompleta incapaz de describir diferentes fenómenos, como la materia oscura, la energía oscura o la asimetría entre materia y antimateria. Además, existen algunas discrepancias en los acoplamientos del bosón de Higgs, que podrían abrir la puerta hacia nueva física. Las diferencias entre los resultados de los experimentos y de las predicciones teóricas sugieren desviaciones respecto al modelo estándar y física más allá del mismo.

La precisión en las medidas de los experimentos es un parámetro fundamental para poder encontrar discrepancias, las cuales pueden ser el camino hacia nueva física o completar las teorías actuales. Cuanta más alta sea la precisión de los experimentos, así como de la teoría, más cerca se estará de descubrir los misterios del universo, como qué es la materia oscura, la energía oscura o por qué el universo tiene mucha más materia que antimateria.

1.2. Objetivos

El objetivo de este trabajo es el estudio de la desintegración del bosón de Higgs en dos bosones W mediante el análisis de colisiones protón-protón en el acelerador de partículas LHC a 13 TeV detectadas por el detector CMS. El trabajo se centra en optimizar la relación entre la señal (S), que se trata de los sucesos de producción del bosón de Higgs y el fondo (B), que agrupa todos los sucesos que se producen en las colisiones que no son considerados señal. De forma más detallada, los objetivos del trabajo se pueden dividir en tres categorías.

En primer lugar, el aprendizaje detallado sobre el bosón de Higgs. Se trata de comprender las características del bosón de Higgs y su importancia en la física actual, así como los procesos de producción y desintegración del mismo, centrándonos en los canales de desintegración estudiados en el trabajo.

En segundo lugar, la familiarización con el entorno de trabajo. Se trata de aprender cómo trabaja el detector CMS y los distintos procesos de reconstrucción de sucesos, al mismo tiempo que se aprende a trabajar con el gran número de datos que se obtienen de este detector.

Por último, el objetivo del proyecto es maximizar la relación entre la señal y el fondo, lo cual se va a realizar por medio de una figura de mérito. Para ello se busca reducir el fondo lo máximo posible sin afectar considerablemente a la señal. Este proceso se hará de forma manual y automática, para poder comprobar qué método es el más eficaz y, por lo tanto, más conveniente desde un punto de vista práctico.

2. Marco teórico

En esta sección se va a explicar el modelo estándar de partículas, la teoría actualmente más aceptada para describir las partículas elementales y las fuerzas de interacción entre ellas, destacando las características del bosón de Higgs.

2.1. Modelo estándar de la física de partículas

Los avances de la comunidad científica desde la década de 1930 han permitido entender la estructura fundamental de la materia. Todo está formado por unas unidades básicas llamadas partículas elementales, las cuales están sujetas a cuatro fuerzas fundamentales: la fuerza gravitatoria, la electromagnética, la interacción fuerte y la interacción débil. El modelo estándar agrupa tres de las cuatro fuerzas fundamentales, excluyendo la gravedad, debido a que no es posible de momento observar el efecto de esta fuerza a niveles cuánticos, consiguiendo de esta manera explicar y predecir la mayoría de interacciones entre las partículas elementales. Con el paso de los años y después de una gran cantidad de experimentos, el modelo estándar se considera una teoría demostrada en física [2].

Las partículas elementales se dividen en dos grupos: fermiones y bosones. Los fermiones son las partículas que forman la materia, tienen espín semientero y se dividen a su vez en dos subgrupos: quarks y leptones. Los quarks pueden formar hadrones (bariones y mesones), mientras que los leptones existen como partículas aisladas. Por otra parte, los bosones son las partículas mediadoras o portadoras de las fuerzas fundamentales, con espín entero. El fotón es el mediador de la fuerza electromagnética, los bosones Z y W de la nuclear débil y el gluón de la interacción fuerte [3]. Las partículas elementales y los bosones conocidos se muestran en la Figura 1:

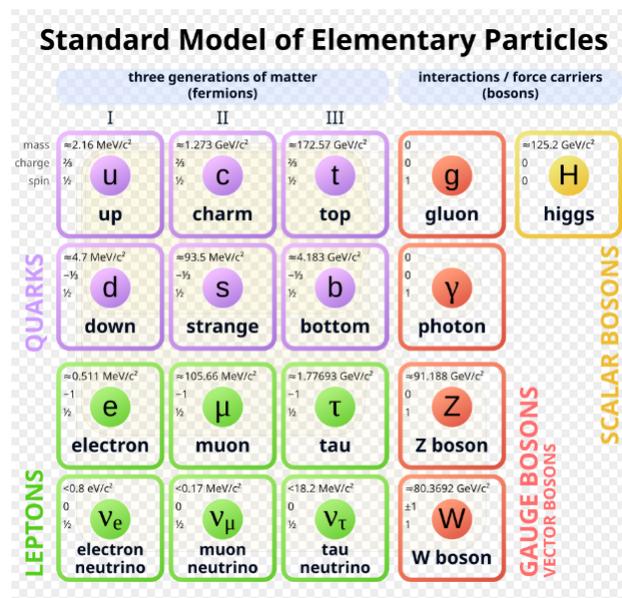


Figura 1: Resumen de las partículas elementales conocidas, diferenciando los bosones de los fermiones, los que a su vez se dividen en leptones y quarks [4].

El éxito del modelo estándar a la hora de agrupar la fuerza nuclear fuerte, nuclear débil y electromagnética en una misma teoría se debe en gran medida a los esfuerzos de la comunidad científica en unificar las teorías electromagnética y débil, en lo que se conoce como teoría electrodébil, impulsada por Sheldon Lee Glashow, Abdus Salam y Steven Weinberg en la década de 1960 y por la que recibieron un premio Nobel en 1979 [5]. Uno de los problemas que supuso la teoría electrodébil era la ruptura de las simetrías del sistema si los bosones W y Z tenían masa, como se observaba experimentalmente. La solución a este problema llegó en 1964 de la mano de Peter Higgs [6], Englert y Brout [7].

El mecanismo de Englert-Brout-Higgs tiene dos bases fundamentales: un nuevo campo cuántico llamado campo de Higgs y una ruptura espontánea de simetría. Este mecanismo describe el campo de Higgs como un campo simétrico e inestable inicialmente, que colapsa a una configuración estable sin romper la teoría y describiendo con precisión la realidad física. Las partículas, como los bosones W y Z, adquieren masa al interactuar con el campo de Higgs por medio del bosón de Higgs.

2.2. El bosón de Higgs

El descubrimiento del bosón de Higgs ha sido uno de los objetivos principales de la física de partículas durante los últimos 50 años. Su descubrimiento en 2012 en los experimentos ATLAS [8] y CMS[9] fue un avance muy importante, ya que permitió validar uno de los pilares sobre los que se asienta el modelo estándar.

El modelo estándar está basado en el modelo quark, la teoría electrodébil y la cromodinámica cuántica, las cuales se agrupan en una única teoría fundamentada en la teoría cuántica de campos, la relatividad especial y las transformaciones *gauge*. Las interacciones entre partículas se describen con un grupo *gauge* $SU(2)_L \times SU(1)_Y \times SU(3)$, correspondiendo los dos primeros términos a la teoría electrodébil y el tercero a la cromodinámica cuántica. En un principio, el grupo $SU(2)_L \times SU(1)_Y$, que describe la teoría electrodébil, era incompatible con la masa de los bosones W y Z, porque la simetría del sistema solo era conservada si los bosones no tenían masa.

El mecanismo Englert-Brout-Higgs explica una manera de conciliar los bosones masivos W y Z con la teoría de *gauge*. Este proceso se basa en la existencia de un nuevo campo cuántico, llamado campo de Higgs, el cual impregna todo el espacio. En los instantes iniciales del universo, el campo era simétrico, pero inestable. Tras una fracción de segundo, el campo rompe espontáneamente la simetría inicial para situarse en un estado estable, pero asimétrico. De este modo, la descripción matemática del campo es simétrica y es compatible con la teoría gauge, mientras que la simetría rota del campo de Higgs aporta masa a las partículas que interactúan con él.

La teoría cuántica de campos, una de las bases del modelo estándar, describe cada partícula como una perturbación en un campo cuántico. De este modo, la existencia de un nuevo campo significa la existencia de una nueva partícula asociada a este campo, una partícula que haga de mediadora entre el campo y el resto de partículas, un bosón. La partícula asociada al campo de Higgs es el bosón de Higgs.

A pesar de haber logrado unificar las interacciones electromagnética y débil, la teoría electrodébil no cobró una importancia real hasta que Gerardus 't Hooft y Martinus Veltman demostraron que era renormalizable, a pesar de la simetría espontáneamente rota. El trabajo de 't Hooft y Veltman dio el impulso final a los campos de Yang-Mills, esenciales en la unificación de las fuerzas en el modelo estándar, cuyo problema era que no se había demostrado si eran renormalizables.

Una vez desarrollado el modelo estándar, el descubrimiento del bosón de Higgs se convirtió en una prioridad, debido al papel fundamental del mecanismo EBH en esta teoría. Además, el mecanismo no predice la masa del bosón, pero con la masa se pueden obtener el resto de propiedades, lo que incentivó aún más su búsqueda.

La extremadamente corta vida media del bosón de Higgs, aproximadamente de $1.6 \cdot 10^{-22}$ s [8], supone que no se puede encontrar en la naturaleza y es necesario producirlo experimentalmente. Debido a su alta masa, el primer acelerador capaz de producir una cantidad de Higgs apreciable fue el LHC, por lo que hasta 2010, prácticamente 50 años tras su primera mención, no se fue realmente capaz de medirlo experimentalmente.

Tras la producción del Higgs, este decae rápidamente a otras partículas, por lo que no es posible observarlo directamente, sino que es necesario observar los productos de su desintegración. Sin embargo, estas partículas también son generadas en otros procesos en las colisiones en el LHC, por lo que observar los productos de desintegración no significa que hubiera un Higgs, especialmente si se tiene en cuenta que los bosones de Higgs se producen una vez por billón de colisiones.

El proceso para encontrar el bosón de Higgs fue medir la masa invariante de los productos, una cantidad que se puede obtener una vez los productos y sus propiedades han sido medidos. Los productos provenientes de un Higgs tendrán una masa invariante igual a la masa del Higgs, mientras que el resto tendrán masas invariantes diferentes cada vez, aleatorias. Esto produce unos resultados en los que se tiene en la mayoría de los casos un valor aleatorio, pero en ciertos casos un valor fijo. Tras el análisis, se produce un histograma que tendrá un pico en un punto específico.

En este tipo de análisis se tienen que cumplir dos condiciones para poder asegurar que se tiene un pico: el pico y el número total de datos deben ser suficientemente grandes. Para determinar como de grandes deben ser el número de datos y el pico, se definió la probabilidad de que, dado un exceso en un determinado conjunto de datos, aparezca un exceso del mismo tamaño aleatoriamente. Un exceso se cataloga como un descubrimiento cuando dicha probabilidad es una entre 3 millones y medio, lo que corresponde a cinco desviaciones estándar de la media, 5σ .

EL 4 de julio de 2012, los experimentos ATLAS y CMS llegaron al límite de 5σ , demostrando la existencia de una nueva partícula en la región de 125-126 GeV, el bosón de Higgs. Desde su descubrimiento, se han mejorado las medidas para obtener valores más precisos de sus diferentes características [9].

La masa del bosón de Higgs está expresada en unidades naturales, el sistema de unidades utilizado mayoritariamente en la física de partículas. Este sistema se caracteriza por designar la unidad a dos constantes fundamentales de la mecánica cuántica, la constante de Plank, \hbar , y la velocidad de la luz, c . De este modo se tiene $\hbar=c=1$, en vez de $\hbar=1.054\cdot 10^{-34}$ J s y $c=299792458$ m/s [10], lo que proporciona agilidad a la hora de escribir fórmulas. Además, permite escribir la masa, el momento y la energía en unidades de GeV, mientras que la longitud y el tiempo se expresan en GeV^{-1} . [11]

El bosón de Higgs es una partícula elemental que forma parte del grupo de los bosones, partículas con espín entero. El espín del Higgs es 0, lo que se conoce como bosón escalar, y no posee carga eléctrica ni carga de color. Tiene una masa de 125.20 ± 0.11 GeV, una vida media de $1.6\cdot 10^{-22}$ s y una anchura de desintegración media de $3.7_{-1.4}^{+1.9}$ MeV [12].

Actualmente en el LHC, la producción de Higgs se da por medio de diferentes mecanismos en los que participan diferentes partículas, ya que la colisión protón-protón no puede producir Higgs. El mecanismo predominante de producción de Higgs es la fusión de gluones (ggH), en la que dos gluones de los protones interactúan por medio de un bucle de quarks pesados, normalmente quarks top. Este mecanismo produce el 87% de los bosones de Higgs totales.

El segundo mecanismo que más aporta a la producción de Higgs es la fusión de bosones vectoriales, VBF por sus siglas en inglés. En este proceso, dos quarks de los protones emiten un bosón vectorial, W o Z, los cuales se fusionan para producir un Higgs. Los quarks finales producen jets, lo que permite identificar este proceso con mayor facilidad. Este mecanismo produce el 10% de los bosones de Higgs. Otros mecanismos de producción son la producción en asociación con un bosón W o Z, conocida como *Higgs Strahlung*, la producción en asociación con un par $t\bar{t}$ o la producción en asociación con un par $b\bar{b}$.

Una vez producido el bosón de Higgs, decae rápidamente por medio de diferentes canales de desintegración. La probabilidad de que un Higgs decaiga por un determinado canal de desintegración se denomina *branching ratio*, que es la fracción entre la anchura natural del Higgs y la anchura parcial del canal de desintegración. Los diferentes canales están representados en la Tabla 1.

Canal de desintegración	<i>Branching ratio</i> /%
WW	25.7 ± 2.5
ZZ	2.80 ± 0.30
$\gamma\gamma$	$(2.50 \pm 0.20)\cdot 10^{-3}$
$b\bar{b}$	53 ± 8
e^+e^-	$<3.0 \cdot 10^{-4}$
$\mu^+\mu^-$	$(2.6 \pm 1.3)\cdot 10^{-4}$
$\tau^+\tau^-$	6.0 ± 0.8

Tabla 1: Canales de desintegración del bosón de Higgs y sus *branching ratio* correspondientes.

3. Marco experimental

En esta parte se va a detallar la estructura y el funcionamiento del detector CMS. Para ello se va a introducir la historia del CERN y su papel fundamental en la construcción del LHC y, en consecuencia, del CMS.

3.1. Organización Europea para la Investigación Nuclear

La Organización Europea para la Investigación Nuclear fue fundada en 1954 y es un ejemplo de colaboración científica internacional y uno de los centros de investigación más importantes del mundo. Desde su creación, ha tenido como objetivo el avance científico y tecnológico, buscando el beneficio global. Entre sus logros más destacados se encuentran el descubrimiento del bosón de Higgs, la creación de la World Wide Web o los detectores médicos de rayos X [13].

El inicio de esta organización se da en la Europa de la posguerra, cuando más necesario era fomentar la cooperación internacional y la inversión en investigación científica, con campos como la física de partículas con un alto potencial pero un costo inasequible para cualquier país. En una reunión de la UNESCO en París en diciembre de 1951 se acuerda la creación del Consejo Europeo de Investigación Nuclear. Dos meses después se firma un acuerdo estableciendo un consejo provisional bajo el acrónimo CERN, por sus siglas en francés, con el objetivo de producir en 18 meses un primer convenio formal.

En junio de 1953, el borrador final del convenio del CERN fue aceptado y firmado por doce países miembros. El convenio fijaba las maneras por las que los países miembros contribuirían económicamente y establecía las bases de los códigos de ética y organización, las cuales incluían una política de libre acceso y la subdivisión de la estructura interna en direcciones, que con el aumento de tamaño de la organización se han ido dividiendo a su vez en grupos y subgrupos.

El director general Felix Bloch coloca la primera piedra del laboratorio en junio de 1955. Desde entonces, el CERN ha superado las expectativas de los más optimistas, convirtiéndose en un laboratorio internacional donde se han realizado grandes avances científicos y tecnológicos. Actualmente prosigue la misión inicial, lograr entender de qué está hecho el universo y cómo funciona.

El carácter internacional de la organización se ha fortalecido a lo largo de los años con la llegada de nuevos países colaboradores. Hoy en día se cuenta con 24 países miembros y 10 países asociados, con dos de ellos en proceso de convertirse en países miembros. Los países miembros son los encargados principales de la financiación de los programas del CERN, lo que les otorga un puesto en el consejo, donde se toman las decisiones sobre la organización y sus actividades [14].

Desde su creación, el CERN ha sido el protagonista de varios logros científicos muy importantes, la mayoría de ellos centrados en la física de partículas. Además del descubrimiento del bosón de Higgs, y su contribución fundamental a la creación de World Wide Web [15] y los detectores médicos de rayos X [16], esta organización es responsable del descubrimiento de los bosones W y Z, la primera creación de átomos de antihidrógeno o el descubrimiento de una violación directa de CP.

3.2. Gran colisionador de hadrones

El gran colisionador de hadrones, LHC por sus siglas en inglés, es el acelerador de partículas más grande y potente del mundo, capaz de impulsar protones e iones hasta prácticamente la velocidad de la luz. Fue puesto en marcha el 10 de septiembre de 2008 y es la última gran incorporación al complejo de aceleradores del CERN. El LHC consiste en un anillo de 27 km de longitud de imanes superconductores con una serie de estructuras aceleradoras para aumentar la energía de las partículas en su recorrido. El acelerador se construyó en túneles a 100 metros de profundidad, bajo el CERN, en la frontera entre Francia y Suiza, cerca de Ginebra.

El concepto del funcionamiento del acelerador es sencillo: dos haces de partículas muy energéticos son acelerados a velocidades cercanas a la de la luz y se hacen colisionar. Los haces viajan en direcciones opuestas en tubos separados, los cuales se mantienen en un estado de vacío ultra alto. Las partículas se dirigen a través de los tubos por medio de electroimanes superconductores, fabricados con cable eléctrico especial diseñado para ser superconductor, y de este modo transportar la electricidad de manera eficiente sin resistencia ni pérdidas de energía. Para que este cable especial pueda operar como superconductor es necesario que tenga una temperatura de 1.85 K, por lo que la mayoría del acelerador está conectada a un sistema de enfriamiento de helio líquido, que enfría los imanes y otros servicios de suministro.

Miles de imanes de diferentes tipos y tamaños se utilizan para dirigir los haces de partículas. Esto incluye 1232 imanes dipolares de 15 metros de largo que curvan los haces y 392 imanes cuadrupolares, cada uno entre 5 y 7 metros de longitud, para enfocar los haces. En los instantes previos a la colisión, otros imanes juntan las partículas de los haces para aumentar la probabilidad de colisión.

El LHC es el último elemento en el complejo de aceleradores del CERN, donde las partículas se aceleran hasta 6.8 TeV por haz. El acelerador lineal 4, Linac4, se convirtió en la fuente de protones del complejo en 2020. Este acelera iones negativos de hidrógeno hasta 160 MeV, preparándolos para entrar al Propulsor de Sincrotrón de Protones, PSB por sus siglas en inglés. Los iones son separados de sus dos electrones en el paso desde Linac4 hasta PSB, dejando únicamente protones que son acelerados hasta 2 GeV e introducidos en el Sincrotrón de Protones(PS), el cual a su vez los acelera hasta 26 GeV y los manda al Super Sincrotrón de Protones(SPS), donde los protones son acelerados hasta 450 GeV [17].

Los protones acelerados por el SPS son finalmente transferidos a los tubos de vacío del LHC. El proceso de llenado de cada anillo del LHC dura 4 minutos y 20 segundos, mientras que los protones tardan 20 minutos más en acelerarse hasta 6.5 TeV. Los haces circulan varias horas por los tubos del LHC bajo condiciones normales, hasta que se hacen colisionar en 4 detectores distribuidos a lo largo del anillo, donde la energía en el punto de colisión alcanza 13.6 TeV.

El complejo de aceleradores del CERN no sirve exclusivamente al LHC, sino que también forman parte de diversos programas de experimentos. La mayoría de aceleradores son parte de experimentos donde se estudian haces a menores energías, pudiendo practicarse varios experimentos a la vez en cada acelerador. En el caso del LHC, se distinguen cuatro experimentos principales: ATLAS, CMS, ALICE y LHCb; pero existen también otros experimentos como TOTEM o FASER. El complejo de aceleradores se puede observar en la Figura 2.

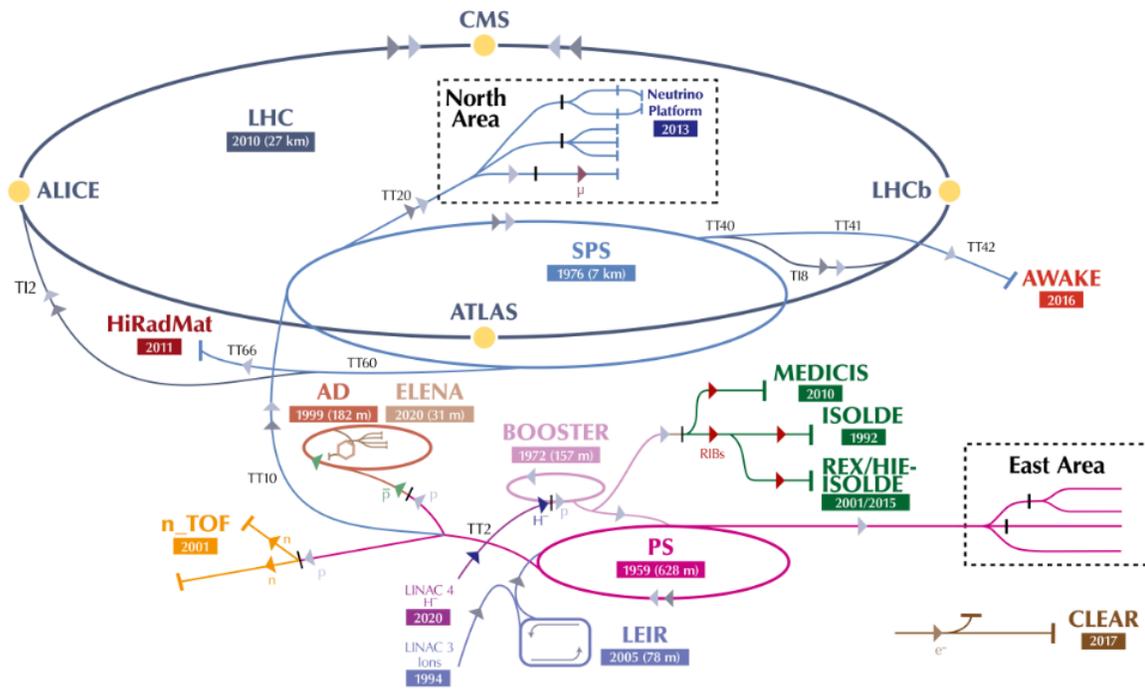


Figura 2: Esquema del complejo de aceleradores del CERN, donde se distinguen los diferentes aceleradores con sus respectivos experimentos, destacando ATLAS, CMS, ALICE y LHCb en el LHC [18].

3.3. Experimento CMS

El experimento CMS es una de las colaboraciones científicas más grandes de la historia, involucrando a 6126 físicos, ingenieros, técnicos y estudiantes provenientes de 247 institutos en 58 países[19]. El experimento está centrado en el detector que le da nombre, el *Compact Muon Solenoid*, situado en el LHC. Este detector está diseñado para estudiar un amplio espectro de sucesos físicos, desde aquellos descritos por el modelo estándar hasta la materia oscura y la búsqueda de nuevas dimensiones. A pesar de compartir objetivos con el experimento ATLAS, el detector CMS se diferencia en su sistema de imanes, que utiliza un solenoide para generar un gran campo magnético, y en su especial capacidad para detectar muones .

El detector fue construido de acuerdo a los objetivos del LHC, que pretendía estudiar la ruptura de la simetría electrodébil y el mecanismo de Higgs, y explorar la física más allá del modelo estándar. Debido a las condiciones del LHC, donde se producen choques en rangos de energías del teraelectronvoltio con alta luminosidad. Por estos motivos, los detectores deben ser capaces de detectar todo tipo de partículas, tener una buena resolución temporal, sistemas electrónicos capaces de realizar la reconstrucción de sucesos y algoritmos que analicen, filtren y seleccionen los eventos más interesantes, ya que es imposible almacenar todos los sucesos que se dan en cada colisión.

Lo que hace único al detector CMS está indicado en su propio nombre, *Compact Muon Solenoid*. Es un detector especializado en la detección de muones, muy compacto, dejando muy pocos espacios libres en su estructura y cuenta con un potente imán solenoidal que genera un gran campo magnético.

3.3.1. Estructura del detector

Esta sección está basada en [20], [21] y en [22].

EL detector CMS está situado en uno de los puntos de colisión del LHC, a unos 100 metros de profundidad cerca de la ciudad de Cessy, Francia. Tiene una estructura cilíndrica, con un radio de 15 m y longitud de 28.7 m, dividida en capas concéntricas, como en una cebolla. Cada capa está especializada en detectar una propiedad característica de las partículas, que se pueden reconocer observando las señales que han dejado en cada capa.

La estructura de capas del CMS está dividida en los detectores que se encuentran dentro del imán solenoidal y aquellos que están en el exterior. La capa más interior es el detector de trazas de silicio, formado por 75 millones de sensores electrónicos organizados en capas concéntricas. Cuando una partícula interactúa por medio de la fuerza electromagnética con alguno de estos sensores, deja una marca, que es utilizada para reconstruir la trayectoria que ha seguido la partícula.

El detector de trazas se encuentra envuelto en el calorímetro electromagnético de centelleo, que a su vez está en el interior del calorímetro para hadrones. La función de los calorímetros es medir la energía que tienen las partículas procedentes de la colisión, lo cual se logra frenando las partículas por completo. El calorímetro electromagnético es el encargado de frenar y medir la energía de electrones y fotones, mientras que los hadrones lo atraviesan y son frenados por el calorímetro para hadrones.

La siguiente capa es el gran imán solenoidal, formado por fibras superconductoras y con unas dimensiones de 13 m de largo y 6 m de diámetro. El imán genera un campo magnético de 3.8 T que curva la trayectoria de las partículas. El campo generado permite conocer la carga de las partículas, ya que las partículas positivas y negativas curva su trayectoria en direcciones opuestas, y el momento de la partícula, debido a que las partículas con momentos altos siguen una trayectoria con un radio de curvatura menor que aquellas con momentos bajos.

Recubriendo el imán se encuentra una capa de grandes piezas de hierro que se utilizan para aislar el campo magnético producido. Debido a que los muones pueden atravesar bastantes metros de material sin perder energía, los calorímetros del CMS no son capaces de frenarlos y se requieren cámaras de muones para detectarlos. Un total de 1400 cámaras de muones son colocadas entre la piezas de hierro, divididas en dos grupos: 250 tubos de deriva y 540 cámaras de tiras catódicas se encargan de reconstruir la trayectoria del muón a partir de las marcas en los diferentes detectores, mientras que 610 cámaras de placas resistivas y 72 cámaras de gas multiplicadoras de electrones forman un sistema redundante de marcas, que decide si conservar o descartar los datos adquiridos del muón.

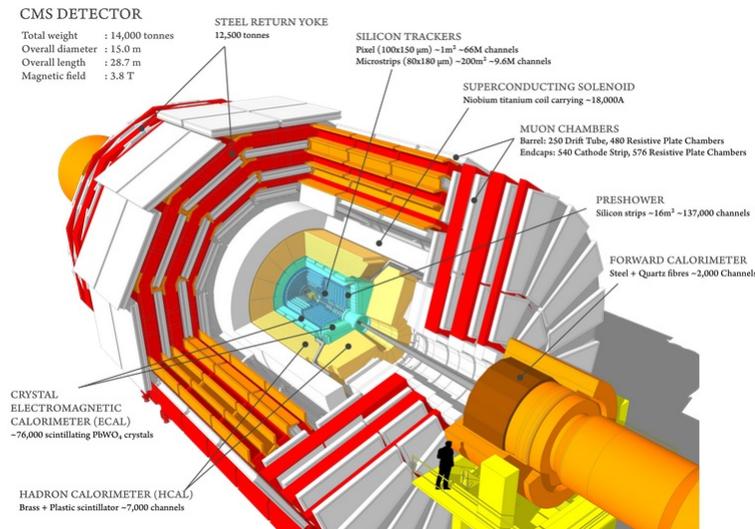


Figura 3: Estructura del detector CMS [23].

3.3.2. Selección de eventos

Esta sección está basada en los capítulos 14 y 15 de [24].

El conjunto de procesos encargados de leer los datos obtenidos en el detector tras una colisión, unirlos bajo una única etiqueta, el evento físico correspondiente, y tomar la decisión de almacenar o descartar dicho evento se conoce como selección de eventos. Esta tarea es realizada por el sistema de *trigger* y adquisición de eventos. En el caso del CMS, este sistema se denomina *TriDAS* y está dividido en dos fases para poder mantener una alta eficiencia: *Level-1 Trigger* y *High-Level Trigger* o HLT.

La primera fase de selección, *Level-1 Trigger*, es realizada por tres subsistemas: el *trigger* del calorímetro, el de primer nivel para muones, y el *trigger* global de primer nivel. Todos los datos que pasan por este proceso de selección son enviados de nuevo al sistema, junto con la decisión de conservarlos o eliminarlos.

La fase HLT del CMS agrupa las fases *Level-2* y *Level-3 trigger*, las cuales sí existen independientemente en otros sistemas de construcción de eventos. Por convención, el *trigger* de segundo nivel normalmente utiliza datos provenientes de los calorímetros y de los detectores de muones, mientras que el de tercer nivel incluye la reconstrucción de la trayectoria de la partícula. Tradicionalmente la reconstrucción de la trayectoria es un proceso muy demandante y largo, por lo que existen *Level-2.5 trigger*, que utilizan una trayectoria parcial.

Las características del CMS han derivado en un proceso de reconstrucción de partículas cargadas menos demandante y largo que lo esperado, por lo que estos algoritmos de selección de eventos es posible que varíen en el futuro. La flexibilidad del CMS en este aspecto es una gran ventaja a la hora de filtrar los sucesos, debido a que permite al sistema utilizar de forma óptima todos los recursos de los que dispone.

3.3.3. Reconstrucción de eventos

Esta sección está basada en los capítulos 3 y 4 de [25].

La estructura de CMS, dividido en diferentes capas especializadas en detectar distintos tipos de partículas, supone que en la primera identificación y reconstrucción de los eventos físicos participa principalmente una parte del detector. Por ejemplo, los electrones y fotones aislados son reconstruidos a partir del calorímetro electromagnético, y los muones a partir de la información de las cámaras de muones. Es posible obtener una descripción de los eventos mejorada utilizando un algoritmo de flujo de partículas (PF, por sus siglas en inglés), que utiliza la información obtenida por cada capa de detectores y combina todas las medidas para reconstruir las propiedades de la partícula.

Este tipo de algoritmo fue utilizado por primera vez en el experimento ALEPH en LEP [26], un acelerador de electrones del CERN previo al LHC en el que se hacían colisionar electrones y positrones [27]. Antes de comenzar a funcionar el LHC se temía que los PF fueran aplicables únicamente en colisiones electrón-positrón, pero en marzo de 2010 el detector CMS demostró que este tipo de algoritmos se podían aplicar a sus datos. En junio de ese mismo año el algoritmo estaba listo para ser aplicado y desde 2011 prácticamente todos los resultados obtenidos en CMS han utilizado algoritmos de *particle flow*.

Una partícula, generalmente, activará varios elementos del PF en distintos subdetectores. Por lo tanto, el primer paso en la reconstrucción de una partícula es identificar y unificar los elementos del algoritmo que ha activado en cada subdetector.

Electrones y fotones

Los electrones se reconstruyen utilizando la información del *tracker* interno y del calorímetro electromagnético. Debido a la gran cantidad de material en el *tracker*, los electrones suelen emitir fotones *bremsstrahlung*, los cuales se convierten en pares e^+e^- que vuelven a emitir fotones. Esto provoca que la identificación de electrones y fotones se realicen de forma similar.

En primer lugar se seleccionan los candidatos a electrón y a fotón. Un candidato a electrón debe tener señal en el calorímetro electromagnético y en el *tracker* interno, pero no puede tener más de tres trayectorias asociadas. Por otra parte, el candidato a fotón debe tener señal en el calorímetro y ninguna trayectoria asociada a dicha señal.

Los candidatos a electrón y fotón no deben superar una energía máxima correspondiente a su señal en las regiones más cercanas del calorímetro de hadrones. Normalmente, la energía máxima permitida en el calorímetro para hadrones es un 10% de la medida en el calorímetro electromagnético. A continuación, se comprueba cuál de las posibles trayectorias, en el caso del electrón, es más consistente con las dos señales identificadas en los dos subdetectores, teniendo en cuenta su energía y momento.

El último paso consiste en comprobar que los candidatos que han superado la primera selección cumplen una serie de parámetros correspondientes a 14 variables, entre las que se encuentra, por ejemplo, la energía irradiada en el *tracker*. La distinción entre electrones y fotones puede variar para selección optimizadas para análisis característicos, por lo que los datos correspondientes a la reconstrucción de electrones y fotones se conserva para permitir diferentes interpretaciones sin tener que aplicar el algoritmo completo de nuevo.

Muones

Los muones son identificados mediante un procedimiento en el que se reconstruyen independientemente las posibles trayectorias obtenidas en el *tracker* y en las cámaras de muones. A partir de las trayectorias se pueden utilizar dos métodos de reconstrucción.

El primer método es la reconstrucción de muones globales. Si una trayectoria del *tracker* coincide con una de las cámaras de muones, se obtiene una trayectoria unificada utilizando el filtro de Kalman. Este método destaca cuando el momento transverso del muon es superior a 200 GeV/c.

El segundo método consiste en extrapolar todas las trayectorias obtenidas en el *tracker* que cumplan unos criterios mínimos de momento transverso y momento total. Las trayectorias extrapoladas se calculan teniendo en cuenta el campo magnético, la pérdida de energía esperada y el *scattering* de Coulomb. Si al menos una trayectoria coincide con un segmento obtenido en las cámaras de muones dicha trayectoria se considera un muon *tracker*.

CMS se caracteriza por su eficiencia a la hora de detectar muones. Se estima que un 99% de los muones producidos en las colisiones protón-protón son reconstruidos por alguno de los dos métodos, y en muchos casos por ambos [28].

4. Análisis

En esta sección se van a describir con detalle el análisis realizado. En primer lugar, se van a describir los procesos físicos que van a ser analizados y las distintas categorías en las que se dividen. En esta primera parte se van a identificar los procesos que se consideran señal (S). A continuación, se van a definir las variables que forman parte del análisis, y se va a introducir la selección básica utilizada. Brevemente se va a describir el apartado técnico del análisis, al que seguirá la definición de la figura de mérito. Por último, se va a realizar una descripción de los diferentes métodos que se han utilizado para mejorar la relación entre la señal y el fondo.

4.1. Procesos físicos

En este proyecto se utilizan sucesos producidos por simulaciones Montecarlo (MC). Este tipo de simulaciones son modelos generados por ordenador que simulan la medición de partículas en un detector. En la física de altas energías se utilizan para teorizar cómo se muestran interacciones entre partículas hipotéticas en un detector. La producción de simulaciones MC consiste en cuatro pasos fundamentales: la generación de eventos, la simulación del detector, el procesado de datos y la reconstrucción de eventos físicos. Algunos de los programas utilizados para producir las simulaciones MC son MadGraph, POWHEG o PYTHIA, los cuales se encargan de la generación de eventos, o Geant4, el cual simula la interacción de todas las partículas con el detector CMS [29].

El estudio va a estar centrado en colisiones protón-protón simuladas, correspondientes a una energía de 13 TeV procedentes de la toma de datos del detector CMS en el año 2018. Los diferentes procesos físicos que se van a estudiar están divididos en las siguientes categorías:

- **Quark Top:** engloba los procesos en los que se produce un quark top, tanto individualmente (tW) como pares ($t\bar{t}$).
- **Non-prompt:** engloba los procesos asociados a partículas que no provienen de la desintegración de una partícula producida en la colisión de protones. Incluye leptones, jets y fotones.
- **VVV:** engloba los procesos de producción de tres bosones, particularmente ZZZ , WZZ , WWZ y WWW [30].
- **WZ:** engloba los procesos en los que un bosón Z se desintegra con un bosón W, dando lugar a 2 o 4 leptones.
- **$V\gamma$:** engloba los procesos de producción de un bosón y un fotón, virtual o real.
- **WW:** engloba la desintegraciones de un par de bosones W a dos leptones cargados.
- **DY:** engloba los procesos Drell-Yan, que son aquellos en los que un quark de un hadrón se desintegra con un antiquark de otro hadrón, produciendo un bosón Z o un fotón, que posteriormente decae a un par de leptones de opuesta carga.

- **Higgs:** engloba todos los procesos de producción de Higgs y desintegración por alguno de sus canales. Esto incluye principalmente la producción de Higgs por medio de la fusión de gluones, aunque también incluye otros procesos en un menor porcentaje, como se indica más adelante en la Tabla 3.

La señal se va a corresponder con los procesos correspondientes a la categoría Higgs, mientras que el resto se considerarán fondo.

4.2. Variables básicas

Durante el análisis se va a trabajar con una serie de variables que definen diferentes características o propiedades de las partículas detectadas. Estas variables son:

- **events:** número de sucesos seleccionados.
- $m_{\ell\ell}$: masa invariante de dos leptones.
- $\Delta R_{\ell\ell}$: distancia angular entre dos leptones.
- m_T^H : masa transversal del bosón de Higgs.
- $m_T^{W^2}$: masa transversal de un sistema que incluye el segundo leptón y su energía perdida. Esta variable está definida como:

$$m_T^{W^2} = \sqrt{2p_T^{\ell_2} p_T^{\text{miss}} [1 - \cos\Delta\phi(\tilde{p}_T^{\ell_2} \tilde{p}_T^{\text{miss}})]} \quad (1)$$

- $p_T^{\ell\ell}$: momento transversal de dos leptones.
- $p_T^{\ell_1}$: momento del primer leptón.
- $p_T^{\ell_2}$: momento del segundo leptón.
- $p_T^{\ell_3}$: momento del tercer leptón.
- η_1 : pseudorapidez del primer leptón.
- η_2 : pseudorapidez del segundo leptón.
- **puppijet**: energía transversal perdida, calculada con el algoritmo PUPPI y teniendo en cuenta el *pileup* [31].
- **mpmet**: energía transversal perdida proyectada.
- $\Delta\phi_{\ell\ell}$: diferencia en el ángulo azimutal entre dos leptones.

- $\Delta\phi_{\ell\ell met}$: diferencia en el ángulo azimutal entre dos leptones y el vector de energía perdida.
- m_{jj} : masa invariante del sistema formado por dos jets.
- **njet** : número de jets.
- **nbjet**: número de b-jets.

4.3. Selección básica

Los sucesos producidos por simulaciones MC son aplicados una primera selección básica, que consiste en una serie de condiciones que se impone a los datos que se van a analizar para adaptarlos al estudio que se quiere realizar. Este trabajo se centra en la desintegración del bosón Higgs en dos bosones WW, que a su vez se desintegran en dos leptones y dos neutrinos, especialmente en los canales de desintegración en los que participan electrones y muones. Se excluye del estudio a los taus debido a que los electrones y muones son más fáciles de detectar, su reconstrucción es más eficiente y tienen menos fondo.

Se ha utilizado una selección básica de referencia [32], que ha sido utilizada en un estudio en el que solo se incluye el canal $e\mu$, que a su vez incluye $e^+\mu^-$ y $e^-\mu^+$. Este canal es habitual en estudios de este tipo debido a que el fondo DY se reduce considerablemente, ya que este tipo de procesos están asociados a leptones con el mismo sabor. En este trabajo se van a incluir los canales e^+e^- y $\mu^+\mu^-$, para posteriormente valorar la señal proporcionada por estos canales.

La selección básica se trata de una serie de condiciones que se aplica a los sucesos para adaptarlos a este estudio. En primer lugar, se requiere que el momento transverso de los dos leptones sea mayor que 25 y 13 GeV, respectivamente, lo que permite asegurarse que los *triggers* de los electrones y muones son eficientes. Otros candidatos a leptón, si hay alguno, deben tener un momento transverso menor que 10 GeV. Con el objetivo de reducir los fondos correspondientes a los *multijet* correspondientes a la cromodinámica cuántica y los fondos $\tau^+\tau^-$, se requiere que $m_T^{W^2}$ sea mayor que 30 GeV. Por último, la contribución de $\tau^+\tau^-$ al fondo se ve aun más reducida al exigir que la masa invariante de los dos leptones sea mayor que 12 GeV, y que m_T^H y $p_T^{\ell\ell}$ sean mayores que 60 y 30 GeV, para el primer y segundo leptón respectivamente. La selección básica se puede ver en la Tabla 2.

Selección básica
$p_T^{\ell_1} > 25 \text{ GeV}$
$p_T^{\ell_2} > 13 \text{ GeV}$
en caso de haber, $p_T^{\ell_3} < 10 \text{ GeV}$
$m_{\ell\ell} > 12 \text{ GeV}$
$p_T^{\ell\ell} > 30 \text{ GeV}$
$m_T^{W^2} > 30 \text{ GeV}$
$m_T^H > 60 \text{ GeV}$

Tabla 2: Cortes correspondientes con la selección básica de referencia.

Aplicar una selección básica de referencia tiene varios beneficios. En principio, supone una gran ayuda para centrar el trabajo en los canales de desintegración que se van a estudiar. Por otra parte, se pueden comparar los resultados obtenidos con aquellos correspondientes a la selección, y comprobar que no se ha cometido ningún error aplicando los cortes y que las variables utilizadas están definidas de igual manera.

4.4. Figura de mérito

La figura de mérito es un parámetro muy común en los análisis de física de partículas que se utiliza para optimizar señales. Su popularidad se debe a la precisión de la función que describe para señales pequeñas. A partir de la señal (S) y el fondo (B), se puede definir la figura de mérito (FOM, por sus siglas en inglés):

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

En los experimentos de conteo con fondos es muy común el uso de figuras de mérito, pudiendo elegir otros tipos como S/B o S/\sqrt{B} . Sin embargo, la FOM escogida define mejor la significancia estadística de la señal cuando B es pequeño. Además, permite ser maximizada, mientras que otras figuras de mérito como S/\sqrt{B} pueden disminuir considerablemente la eficiencia del experimento al aplicar ese tipo de procesos [33].

El análisis se va a centrar en la optimización de la señal respecto del fondo, lo que significa que se va a intentar aumentar el valor de la FOM. Al mismo tiempo, este parámetro sirve para comprobar que los distintos pasos que se van a dar realmente mejoran la señal, ya que tras cada uno se puede obtener la FOM y ver si ha mejorado o empeorado.

Como el análisis se va a realizar en tres canales, es necesario definir una figura de mérito total, que incluya las FOM correspondientes a cada canal. Definimos la FOM_{total} :

$$\text{FOM}_{total} = \sqrt{(\text{FOM}_{e^+e^-})^2 + (\text{FOM}_{\mu^+\mu^-})^2 + (\text{FOM}_{e\mu})^2} \quad (3)$$

La FOM_{total} permite medir la aportación al resultado final del estudio de los canales con leptones del mismo sabor, los cuales debido al fondo DY tienen una FOM menor.

4.5. Herramienta de análisis

Este apartado está basado en [34].

El análisis se hace con una herramienta habitual en la física de partículas en el CERN, el *framework* de análisis ROOT. Esta herramienta, a pesar de estar escrita en el lenguaje C++, permite trabajar con otros lenguajes de programación, como python.

Se trabaja en una carpeta que contiene todos los programas necesarios para el análisis, en la que principalmente se va a trabajar con el fichero cuts.py. Este fichero se va a modificar cada vez que se quiera introducir un nuevo corte. Una vez el corte ha sido introducido y el cambio en el fichero es guardado, se ejecuta el programa de análisis. Este programa envía los datos que se quieren analizar a la red de procesamiento del CERN, siendo cada tarea específica de procesamiento un *job*. En este trabajo el análisis consiste en 492 *jobs*.

Los datos procesados se envían de vuelta al sistema que los solicitó en forma de archivos de extensión .root. Estos archivos se unen en uno solo, el cual contiene toda la información necesaria para el análisis. A partir de este archivo se pueden pintar los histogramas de las diferentes variables, o acceder al mismo y leer los datos directamente.

4.6. Optimización de la señal

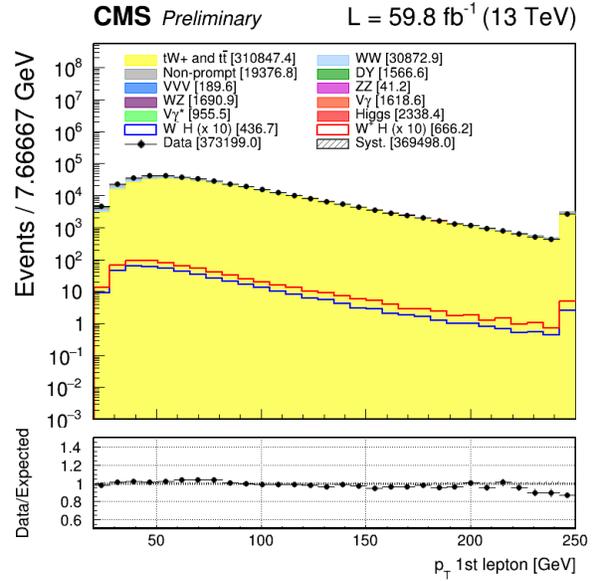
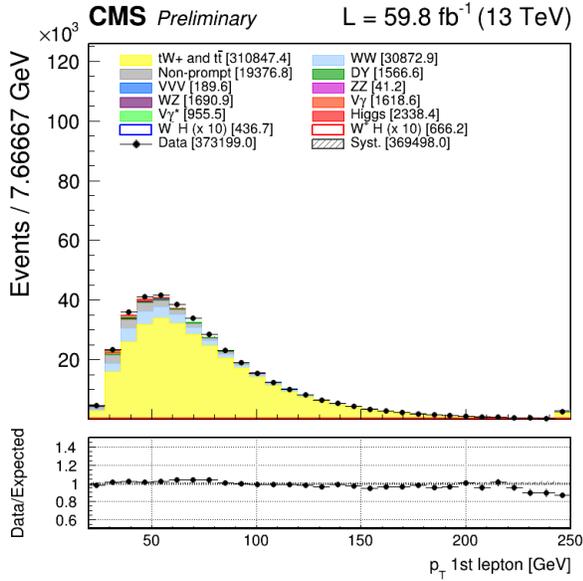
El proceso de optimización de la señal consiste principalmente en aplicar diferentes cortes a las variables iniciales con el objetivo de reducir el fondo lo máximo posible al mismo tiempo que se reduce la señal lo mínimo posible. El análisis que se va a realizar es en base a los datos simulados por MC para evitar sesgos. Los cortes se van a aplicar de forma secuencial, lo que significa que no se tienen en cuenta posibles correlaciones. Esto provoca que al aplicar un corte a algunas variables es posible que no se mejore la FOM, debido a que dicha variable esté muy relacionada con otra variable a la que ya se le ha aplicado un corte anteriormente.

La selección del orden en el que se cortan las variables, por lo tanto, tiene importancia. En primer lugar, se van a cortar en las variables más discriminantes. Estas variables se caracterizan por ser capaces de eliminar una gran cantidad de fondo en la región en la que se está realizando el estudio. Como los cortes aplicados a estas variables son capaces de eliminar mucho fondo, el análisis va a comenzar con el estudio de estas variables, para después continuar con el resto de variables.

La masa invariante de dos leptones (m_{ll}) y la masa transversa del primer leptón y su energía perdida (m_{th}) son las dos variables discriminantes que indica el artículo de referencia que se ha utilizado en la selección básica. Por otra parte, según [35] la energía transversal perdida proyectada ($mpmet$) y la masa transversa del segundo leptón y su energía perdida ($mtw2$) son también variables discriminantes. Sin embargo, sigue siendo necesario realizar un estudio de las variables discriminantes en este trabajo. Este estudio consiste en observar los *plots* de las diferentes variables y determinar en cuáles se puede añadir un corte eliminando la mayor cantidad de fondo posible sin afectar a la señal.

Una vez seleccionadas las variables en las que se va a cortar inicialmente, es el momento de seleccionar el corte óptimo, que es el corte que hace máxima la FOM. En este análisis se van a utilizar tres métodos para seleccionar el corte óptimo.

El primer método consiste en observar los histogramas de las variables y determinar visualmente el corte óptimo. Este primer método, además, sirve para señalar las variables más discriminantes, que serán aquellas en las que se aprecien distintas distribuciones para el fondo y para la señal. Los histogramas por defecto tienen la apariencia que se observa en la Figura 4, en la que se ha utilizado como ejemplo la variable $p_T^{\ell_1}$.



(a) Histograma de ejemplo correspondiente a la variable $p_T^{\ell_1}$ al aplicar la selección básica. (b) Histograma logarítmico de ejemplo correspondiente a la variable $p_T^{\ell_1}$ al aplicar la selección básica.

Figura 4: Histogramas correspondientes a la variable $p_T^{\ell_1}$ en el canal $e\mu$ al aplicar la selección básica.

En primer lugar, en el histograma se observan dos tipos de datos. Por una parte, los puntos negros se corresponden con los datos producidos en el LHC, mientras que los histogramas de colores son diferentes sucesos físicos simulados por MC. La leyenda indica que color corresponde a cada suceso simulado. Por ejemplo, la señal (S) o producción del bosón de Higgs se corresponde con Higgs en la leyenda y se pinta en color rojo. Los fondos principales son los correspondientes a los sucesos relacionados con el quark top, en amarillo, y los sucesos WW, en azul claro. Por otra parte, visualmente, este tipo de histogramas no permite diferenciar con claridad diferentes tendencias entre la señal y el fondo, por lo que se van a utilizar *plots* de *shapes*. Este tipo de histogramas representan únicamente la señal (S), en este caso Higgs en rojo, y los dos principales fondos, los sucesos relacionados con el quark top y los sucesos WW. Continuando con el ejemplo de la variable $p_T^{\ell_1}$, la Figura 5 muestra su *plot* de *shapes*.

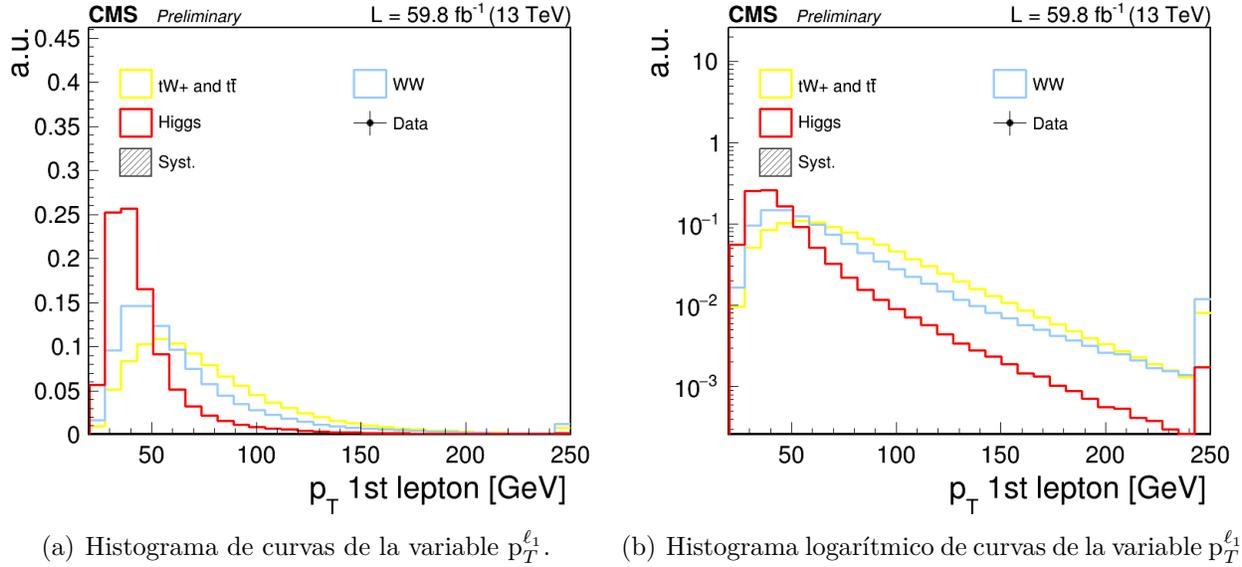


Figura 5: Histograma de curvas de la variable $p_T^{\ell_1}$ al aplicar la selección básica y representando únicamente los dos fondos principales.

El histograma de curvas permite observar con mayor claridad en que zona se produce el cambio de tendencia entre la señal y los fondos principales, dando una primera aproximación de donde podría situarse el corte óptimo. Además, es destacable mencionar la menor diferencia entre tendencias en el histograma logarítmico, lo que supone más dificultad para aproximar visualmente el corte óptimo.

Observando el histograma de la figura 5(a) se puede determinar que el corte óptimo es $p_T^{\ell_1} < 50$ GeV. De este modo se conserva la parte de la distribución donde la señal predomina sobre el fondo, mientras que se descarta la zona en la que la señal cae y el fondo es mayor.

Existen varias formas de calcular la FOM en este tipo de estudios. La más sencilla, aunque también más lenta, consiste en pintar los histogramas, en este caso aplicando la selección básica, y calcular manualmente la FOM a partir de los datos pintados en la leyenda del histograma. Se considera Higgs la señal (S) y el resto de sucesos que aparecen en el histograma son el fondo. El problema de esta forma es que es necesario pintar los histogramas y leerlos cada vez que se quiera comprobar si la figura de mérito ha aumentado tras un corte. Para agilizar este proceso se ha creado un programa en un fichero llamado merito.py, el cual es capaz de leer el fichero que devuelven los jobs y pinta en la pantalla el número de sucesos correspondientes a cada proceso, destacando el número de sucesos de la señal (S), del fondo (B), y la FOM obtenida.

Por otra parte, utilizando este programa se ha podido conocer la aportación de cada modo de producción de Higgs al total. Esto se debe a que el programa pinta en la pantalla el número de sucesos total de Higgs, lo que se llama señal (S), pero también el resto de procesos físicos, permitiendo calcular la aportación de cada modo de producción como un porcentaje del total. Los resultados se observan en la Tabla 3:

Proceso de producción	Aportación a la señal
Fusión de gluones	85.5 %
VBF	7.1 %
Par $t\bar{t}$	4.9 %
Higgs Strahlung	1.5 %
Otros procesos	1.0 %

Tabla 3: Aportación a la señal de los distintos modos de producción del Higgs en el LHC.

El segundo método que se ha utilizado para obtener el corte óptimo se basa en el programa mencionado para calcular la FOM. Este método consiste en aplicar paralelamente diferentes cortes a una misma variable y posteriormente calcular la figura de mérito en cada corte. Los cortes se han aplicado cada 10 GeV, correspondiéndose a variable $<x$, siendo variable la variable a la que se aplica el corte y x el conjunto de múltiplos de 10 GeV. Además, con los resultados obtenidos se puede realizar una gráfica de la FOM frente a la variable discriminante que permita ver el corte óptimo y cómo se comporta la FOM al cortar dicha variable.

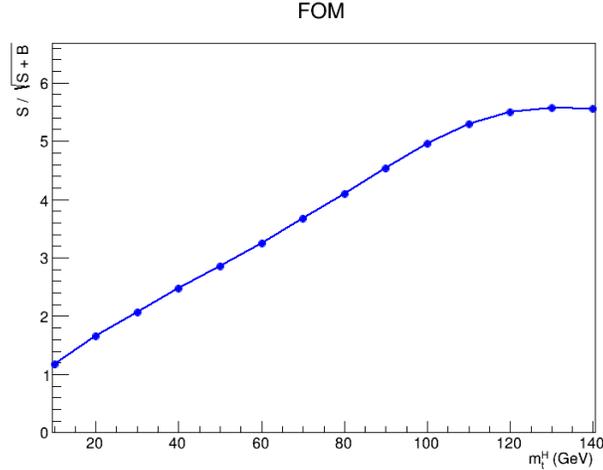


Figura 6: FOM frente a la variable m_T^H . Cada punto azul corresponde a el valor de la figura de mérito al aplicar el corte $m_T^H < x$, siendo x el valor del eje X en GeV.

En la Figura 6 se puede observar como la FOM alcanza un máximo para un corte determinado, en este caso para $m_T^H < 130$ GeV. Este método tiene varias ventajas respecto al primero. Entre estas ventajas se encuentra la capacidad de calcular la FOM sin necesidad de pintar los histogramas y de leer la leyenda para calcularlo manualmente. Además, no depende de una tendencia visual que deba ser observada, si no que el programa pinta el valor de la figura de mérito para cada corte, lo que permite escoger el corte óptimo sin necesidad de pintar la gráfica. A pesar de no ser necesaria para obtener el corte óptimo, la gráfica permite observar la tendencia de la figura de mérito y descartar máximos estadísticos no significativos, los cuales teniendo en cuenta la correlación entre variables es posible que empeoren la figura de mérito al disminuir la señal.

El problema con este método es que los diferentes cortes son discretos, cada 10 GeV como se ha indicado, y en intervalos relativamente amplios, lo que significa que es más probable que el corte óptimo esté situado en el intervalo entre dos cortes en vez de en el corte obtenido o su entorno más cercano. Para solucionar este problema de precisión se crea un programa en un fichero llamado optimización.py el cual calcula la FOM mediante un proceso de optimización bin a bin, que permite obtener el mejor corte posible y no una aproximación como en los casos anteriores.

El tercer método se basa en el uso del programa de optimización para calcular los cortes óptimos de cada variable. El programa permite obtener el corte óptimo sin necesidad de pintar los histogramas, como en el primer método, o de aplicar cortes intermedios, como en el segundo método. Además, el programa calcula una aproximación de la figura de mérito si se aplicase el corte que propone, por lo que es posible comparar cortes antes de aplicarlos. Esto permite aplicar en primer lugar los cortes óptimos en las variables más discriminantes, y una vez aplicados volver a ejecutar el programa para que calcule los cortes óptimos teniendo en cuenta los cortes anteriores.

Para facilitar la visualización del corte se le pide al programa que, una vez hecha la optimización, cree una gráfica con la figura de mérito frente a la variable. La gráfica permite comprender visualmente qué corte se está haciendo, y además permite realizar una doble comparación entre los dos últimos métodos: una primera comparación de los cortes óptimos obtenidos y las gráficas asociadas a los mismos, y una segunda comparación de la FOM obtenida en cada caso.

Un ejemplo de gráfica obtenida con este método se ve en la Figura 6:

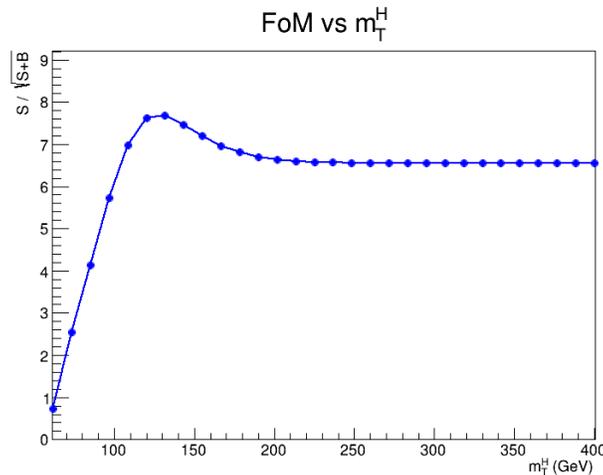


Figura 7: FOM frente a la variable m_T^H . En el caso de esta variable, la FOM del corte óptimo se ha obtenido tras cortar en la variable $m_{\ell\ell}$.

Cada punto azul corresponde a un valor de la figura de mérito en el proceso de optimización bin a bin. En este caso se observa un pico de casi FOM=8 en torno a los 80 MeV, seguido de un pequeño descenso en el valor de la FOM que después se estabiliza en torno a FOM=6.9.

Tras completar los tres métodos de análisis se compararán los resultados obtenidos y se determinará qué método es mejor para cada circunstancia.

5. Resultados

El análisis comienza con la aplicación de la selección básica al conjunto de datos que se van a estudiar. El primer paso es obtener la FOM con la selección básica aplicada para cada una de las categorías, para de este modo saber de que valor se parte y poder juzgar si se está mejorando o no la señal.

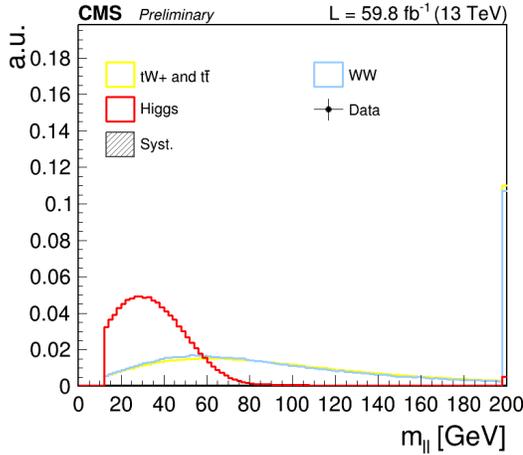
Número de sucesos	Señal	Fondo	FOM
e^+e^-	857.6	1285336	0.76
$\mu^+\mu^-$	1723.7	2112689	1.19
$e\mu$	2338.4	367064	3.86
Total	4919.7	3765089	4.11

Tabla 4: Valores de la señal, fondo y FOM al aplicar la selección inicial.

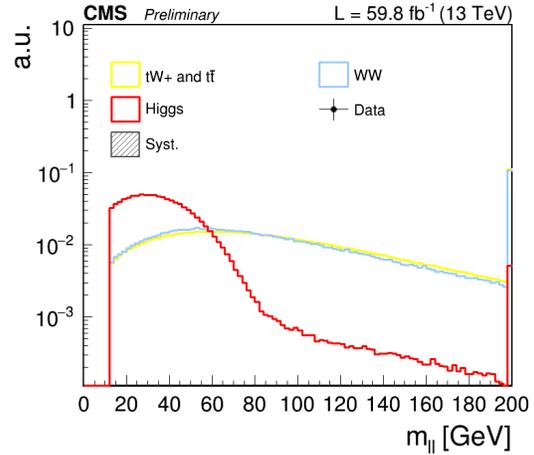
Como se ha introducido en la sección 4.4, la FOM_{total} se utiliza para cuantificar la aportación de los canales e^+e^- y $\mu^+\mu^-$ a la figura de mérito del canal $e\mu$. Tras aplicar la selección inicial, el valor de este parámetro es $FOM_{total}=4.11$. Por otra parte, se observa que la FOM del canal $e\mu$ es mayor que los otros dos canales. Esto se debe principalmente a que la selección inicial aplicada está diseñada para centrar el estudio en este canal, además de no tener una gran cantidad de fondo DY. Por lo tanto, no solo los canales e^+e^- y $\mu^+\mu^-$ tienen más fondo, sino que como la selección inicial no está diseñada para ellos la señal también es menor. Cabe destacar que la señal (S) en $\mu^+\mu^-$ prácticamente es el doble que la de e^+e^- , lo que se puede deber a que los cortes aplicados en la selección inicial favorezcan más a una categoría que a la otra.

Antes de comenzar con el análisis es necesario comprobar que la selección básica ha sido aplicada correctamente. Para ello se van a comparar dos histogramas, uno procedente del artículo de referencia y otro obtenido tras aplicar la selección básica.

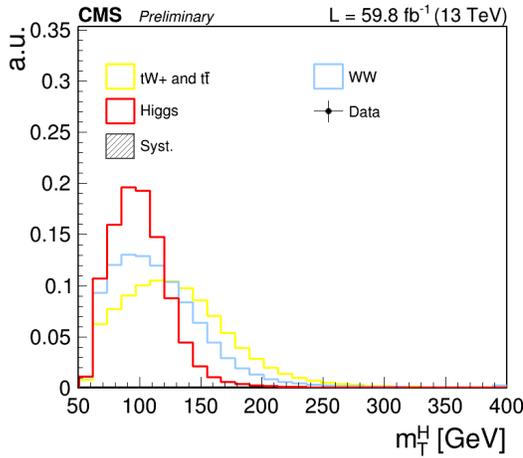
El primer paso señalado en la sección 4.6 es determinar qué variables son las variables discriminantes, aquellas que eliminan más fondo manteniendo la mayor cantidad de señal posible. Las variables discriminantes son las que tengan una mayor diferencia entre la distribución de la señal y de los fondos en los histogramas, fácilmente observable con los histogramas de curvas. En la Figura 8 se pueden observar las variables más discriminantes.



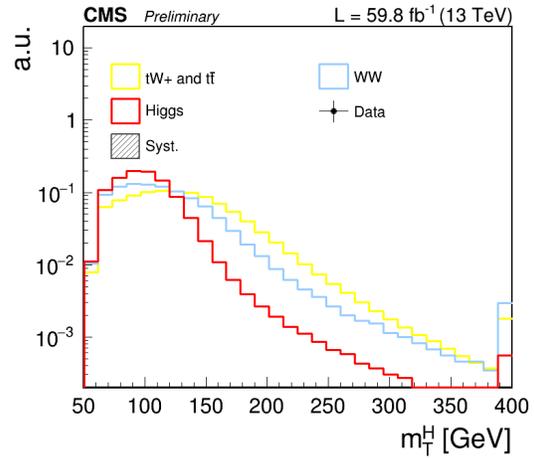
(a) Histograma de curvas de la variable $m_{\ell\ell}$.



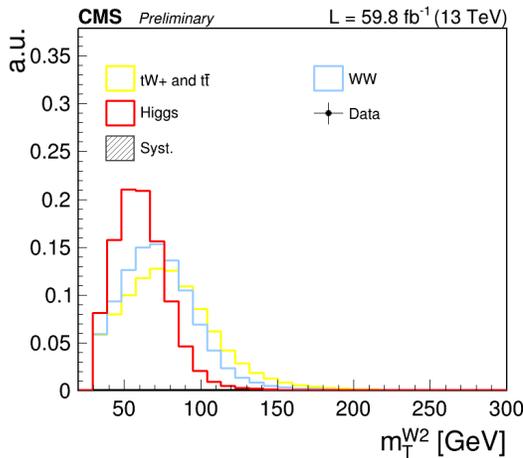
(b) Histograma logarítmico de curvas de la variable $m_{\ell\ell}$.



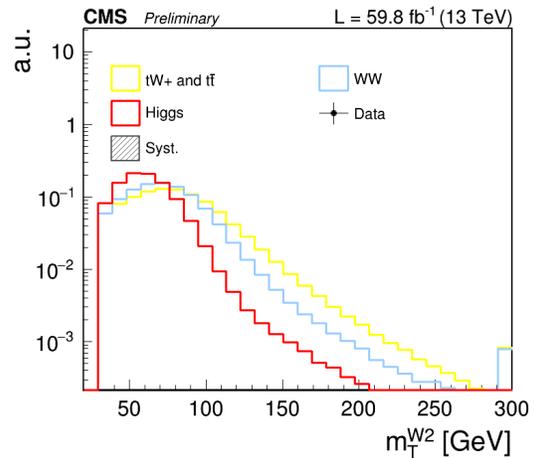
(c) Histograma de curvas de la variable m_T^H .



(d) Histograma logarítmico de curvas de la variable m_T^H .



(e) Histograma de curvas de la variable m_T^{W2} .



(f) Histograma logarítmico de curvas de la variable m_T^{W2} .

Figura 8: Histogramas de curvas de las variables potencialmente discriminantes. Los histogramas del resto de variables están en el apéndice 1.

Se puede observar en la Figura 8 que las variables $m_{\ell\ell}$, $m_T^{W^2}$ y m_T^H se pueden considerar discriminantes, como se esperaba teniendo en cuenta el artículo de referencia para la selección básica.

Una vez definidas las variables discriminantes se procede a obtener los cortes óptimos. Como es un procedimiento visual es importante determinar un criterio que se pueda replicar para las cuatro variables. En este estudio el criterio va a ser fijarse donde la distribución de la señal (S), una vez comienza a decaer desde su máximo, corte con la distribución de al menos uno de los fondos. Aplicando este criterio se obtienen los cortes:

- $m_{\ell\ell} < 60$ GeV
- $m_T^H < 120$ GeV
- $m_T^{W^2} < 80$ GeV

Los cortes óptimos obtenidos se aplican a los datos y se observa su efecto en la señal, el fondo y la FOM. Los resultados se recopilan en la Figura 5

Corte aplicado	e^+e^-			$\mu^+\mu^-$			$e^+\mu^-$		
	Señal	Fondo	FOM	Señal	Fondo	FOM	Señal	Fondo	FOM
Sin cortes	857.6	1285336	0.76	1723.7	2112589	1.19	2338.4	367064	3.86
$m_{\ell\ell} < 60$ GeV	85.3 %	11.5 %	1.90	87.9 %	13.4 %	2.85	90.8 %	29.6 %	6.44
$m_T^H < 120$ GeV	69.1 %	9.1 %	1.74	73.4 %	10.6 %	2.67	74.5 %	14.3 %	7.61
$m_T^{W^2} < 80$ GeV	65.9 %	8.8 %	1.68	70.4 %	10.3 %	2.59	71.3 %	12.6 %	7.74

Tabla 5: Valores obtenidos al aplicar los cortes en las variables discriminantes con el primer método de análisis.

La Tabla 5 indica varios aspectos sobre el análisis por medio del primer método. En primer lugar, el primer corte es muy bueno, ya que reduce desde un 70.4 % hasta un 88.5 % del fondo al mismo tiempo que reduce la señal entre un 9.2 % y un 24.7 %. El segundo y tercer cortes aumentan la FOM del canal $e\mu$ mientras que empeoran los otros dos canales. A pesar de ello, es posible que los cortes mejoren el total del análisis si la FOM_{total} aumenta.

	Sin cortes	$m_{\ell\ell} < 60$ GeV	$m_T^H < 120$ GeV	$m_T^{W^2} < 80$ GeV
FOM_{total}	4.11	7.29	8.25	8.33

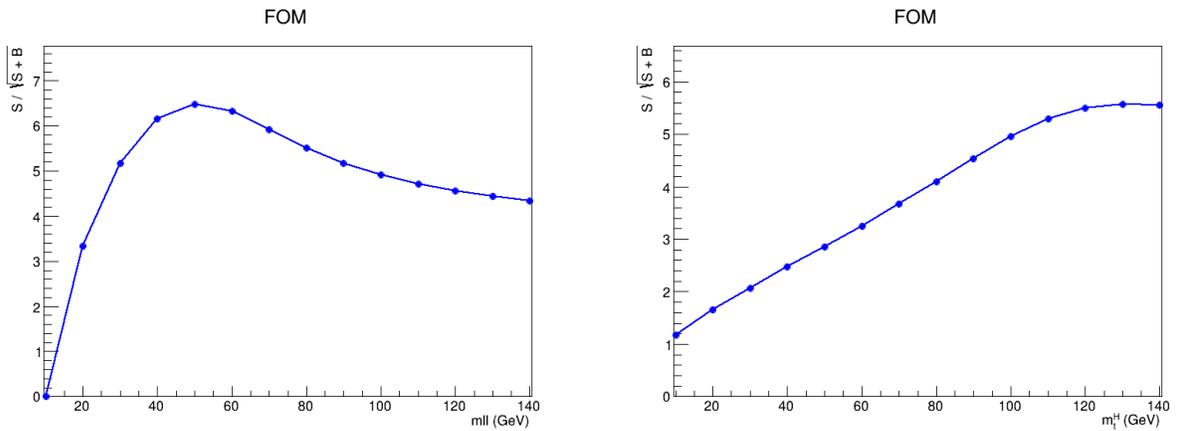
Tabla 6: Figura de mérito total correspondiente a los datos obtenidos con el primer método de análisis.

La tabla 6 demuestra que el segundo y tercer cortes mejoran la figura de mérito total, por lo que a pesar de perjudicar dos categorías su impacto es positivo en el total del análisis.

La segunda parte del análisis añade el uso del programa en el fichero merito.py, el cual calcula la figura de mérito sin necesidad de pintar los histogramas. Además de obtener la FOM, pinta un *plot* para visualizar el máximo de la figura de mérito para cada variable. Cada *plot* se ha generado a partir de la FOM calculada para cada variable al aplicar cortes en intervalos de 10 GeV desde 0 hasta 140 GeV. Los cortes óptimos obtenidos con el programa son:

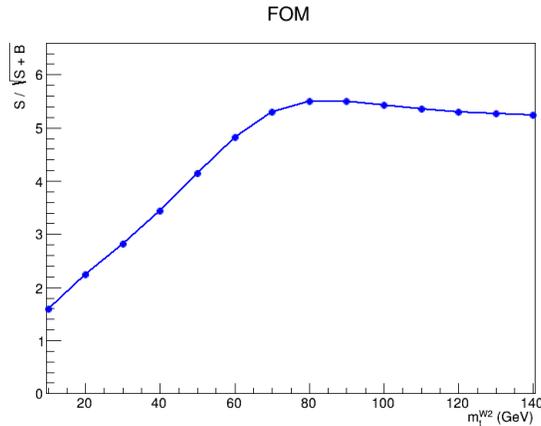
- $m_{\ell\ell} < 50$ GeV
- $m_T^H < 130$ GeV
- $m_T^{W^2} < 90$ GeV

Los *plots* que permiten visualizar estos cortes son los siguientes:



(a) FOM frente a la variable $m_{\ell\ell}$.

(b) FOM frente a la variable m_T^H .



(c) FOM frente a la variable $m_T^{W^2}$.

Figura 9: FOM frente a las distintas variables discriminantes.

Aplicando los cortes obtenidos con el segundo método de análisis se han obtenido los siguientes resultados:

Corte aplicado	e^+e^-			$\mu^+\mu^-$			$e^+\mu^-$		
	Señal	Fondo	FOM	Señal	Fondo	FOM	Señal	Fondo	FOM
Sin cortes	857.6	1285336	0.76	1723.7	2112689	1.19	2338.4	367064	3.86
$m_{\ell\ell} < 50$	74.9 %	9.7 %	1.82	78.0 %	11.4 %	2.73	80.2 %	21.9 %	6.63
$m_T^H < 130$ GeV	66.8 %	8.2 %	1.76	70.9 %	9.8 %	2.68	72.1 %	12.6 %	7.84
$m_T^{W^2} < 90$ GeV	62.6 %	8.0 %	1.70	66.9 %	9.8 %	2.59	68.1 %	11.3 %	7.90

Tabla 7: Valores obtenidos al aplicar los cortes en las variables discriminantes con el segundo método de análisis.

El primer corte mejora la figura de mérito en los tres canales por lo que no hay duda de que es un buen corte. El segundo y tercer corte se comportan del mismo modo que los cortes análogos de la primera parte: mejoran el canal $e\mu$ mientras que empeoran los otros dos canales. De nuevo se va a utilizar la FOM_{total} para comprobar si su aportación total al análisis es positiva o negativa.

	Sin cortes	$m_{\ell\ell} < 50$ GeV	$m_T^H < 130$ GeV	$m_T^{W^2} < 90$ GeV
FOM_{total}	4.11	7.40	8.47	8.48

Tabla 8: Figura de mérito total correspondiente a los datos obtenidos con el segundo método de análisis.

La tabla 8 muestra que tanto el segundo como el tercer corte son buenos, ya que mejoran la figura de mérito total. Cabe destacar que la mejora entre el segundo y el tercer corte es mínima, por lo que sería interesante tener en cuenta la correlación entre variables para recalcular el corte óptimo. Por el momento se observa una mejora respecto a la primera parte del análisis, lo que tiene sentido teniendo en cuenta los métodos utilizados en cada parte.

Por último, la tercera parte del análisis. En esta parte se va a utilizar un programa, `optimizacion.py`, para obtener los cortes óptimos mediante una optimización bin a bin. Además, se va a tener en cuenta la correlación entre variables y tras cada corte se volverá a calcular el corte óptimo de cada variable, para así poder comparar los resultados con las otras dos partes del análisis.

El proceso de análisis consiste en comenzar con la selección básica y obtener el corte óptimo de las variables discriminantes con el programa `optimizacion.py`. Cada corte óptimo tiene asociado el valor de la FOM al aplicar dicho corte, por lo que se pueden comparar unos con otros. El corte que maximice más la FOM será aplicado, el programa de análisis ejecutado y con el corte aplicado se volverá a obtener el corte óptimo del resto de variables con el programa de optimización. Este proceso se repetirá con las variables discriminantes hasta obtener primer valor final de la FOM.

Los cortes obtenidos por este método son los siguientes:

- $m_{\ell\ell} < 50$ GeV
- $m_T^H < 131.67$ GeV
- $m_T^H < 76$ GeV

Como se ha mencionado anteriormente, cabe destacar que los cortes de las variables m_T^H y m_T^{W2} no se han obtenido al mismo que el corte óptimo para $m_{\ell\ell}$. Al aplicar el programa a la selección básica el corte óptimo que más mejoraba la FOM era el correspondiente a masa invariante de los leptones. Una vez aplicado el primer corte, se ejecutó de nuevo el programa, variando los cortes óptimos de las otras dos variables. El mismo proceso se repitió para el último corte. De este modo se está teniendo en cuenta, al menos parcialmente, las correlaciones entre variables. Las gráficas correspondientes a los cortes obtenidos se muestran a continuación:

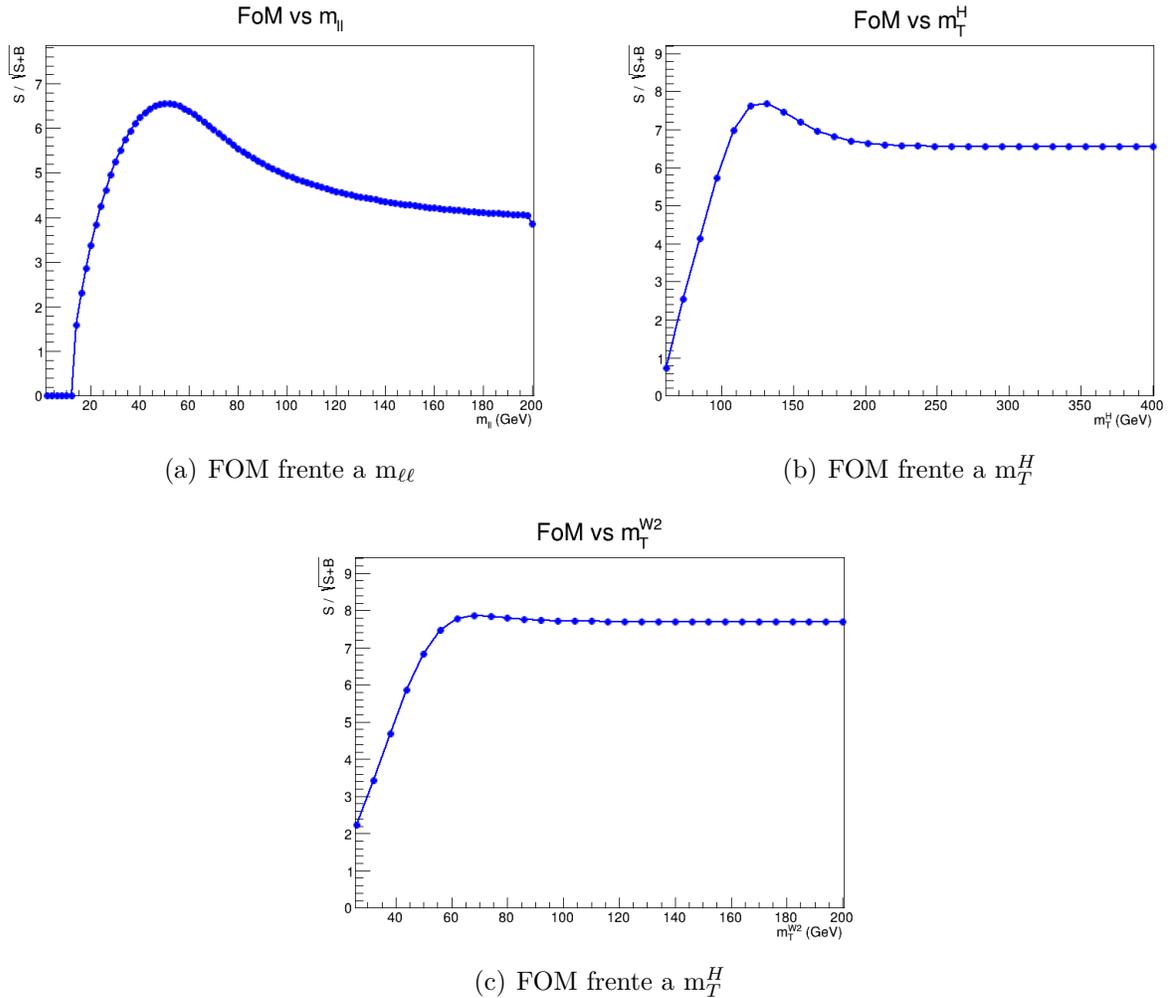


Figura 10: Gráficas correspondientes a las distintas variables discriminantes frente a la figura de mérito obtenidas con el programa optimizacion.py.

Los resultados obtenidos de la señal, el fondo y la FOM al aplicar los cortes están recogidos en la figura 9.

Corte aplicado	e^+e^-			$\mu^+\mu^-$			$e^+\mu^-$		
	Señal	Fondo	FOM	Señal	Fondo	FOM	Señal	Fondo	FOM
Sin cortes	857.6	1285336	0.76	1723.7	2112689	1.19	2338.4	367064	3.86
$m_{\ell\ell} < 50$ GeV	74.9 %	9.7 %	1.82	78.0 %	11.4 %	2.73	80.2 %	21.9 %	6.63
$m_T^H < 131.67$ GeV	67.6 %	8.3 %	1.78	71.6 %	9.8 %	2.70	72.9 %	12.9 %	7.86
$mtw_2 < 76$ GeV	60.5 %	7.5 %	1.67	64.7 %	9.2 %	2.52	71.2 %	11.5 %	7.93

Tabla 9: Valores obtenidos al aplicar los cortes en las variables discriminantes con el tercer método de análisis.

Al igual que en los dos métodos anteriores, el primer corte mejora notablemente la FOM en los tres canales, pero el segundo y tercer cortes mejoran únicamente el canal $e\mu$. Por este motivo es necesario comparar la figura de mérito total para poder determinar si los cortes aplicados mejoran o no el análisis.

	Sin cortes	$m_{\ell\ell} < 50$ GeV	$m_T^H < 130$ GeV	$m_T^{W^2} < 90$ GeV
FOM_{total}	4.11	7.40	8.50	8.49

Tabla 10: Figura de mérito total correspondiente a los datos obtenidos con el tercer método de análisis.

El segundo corte mejora la FOM_{total} , por lo que es considerado un buen corte. Por otra parte, el tercer corte empeora ligeramente la figura de mérito total, a pesar de hacer la FOM del canal $e\mu$ máxima. Esto se debe probablemente a que el corte que hace máxima la FOM para un canal elimina demasiada señal de los otros dos canales, de modo que el empeoramiento de la FOM en dos canales tiene más impacto que el aumento en solo uno.

Por último, es interesante comparar los resultados obtenidos utilizando los tres métodos de análisis. Antes de proceder con la comparación, es importante tratar de predecir los resultados en función de que procesos se han llevado a cabo en cada parte del análisis. La primera parte del análisis ha consistido en una aproximación visual del corte óptimo a partir de los *plots* de *shapes*. Los cortes han mejorado al FOM en el total de los tres canales. A continuación, se ha utilizado un programa que calcule la figura de mérito para una determinada variable para una serie de valores y se ha pintado una gráfica para apoyar visualmente el resultado. En este caso, en vez de observar directamente una serie de distribuciones y aplicar un criterio visual para decidir el corte, se han obtenido una serie de datos que indican cuál es el corte óptimo, por lo que se espera más precisión a la hora de obtener el mejor corte y, por lo tanto, mejores resultados. Por último, se ha creado un programa que calcule el corte óptimo por un proceso de optimización, al que se le ha añado el hecho de que los cortes sucesivos se han calculado teniendo en cuenta los cortes que ya han sido calculado. De este modo, la tercera parte debería obtener mejores resultados que la segunda y, en consecuencia, que la primera.

Se va a realizar una primera comparación de los datos obtenidos en cada parte del análisis, sin realizar ningún cambio para optimizar los resultados.

	Sin cortes	$m_{\ell\ell}$	m_T^H	$m_T^{W^2}$
Primera parte	4.11	7.29	8.25	8.33
Segunda parte	4.11	7.40	8.47	8.48
Tercera parte	4.11	7.40	8.50	8.49

Tabla 11: Figura de mérito total correspondiente a los datos obtenidos con los tres métodos de análisis.

Comprensiblemente, en las tres partes del análisis se ha obtenido el mismo resultado antes de aplicar los cortes. Aunque pueda parecer una primera comparación ilógica, permite asegurar que los tres métodos han partido de la misma FOM, algo que no sería posible si no se hubiera calculado esta FOM en cada parte. El primer corte es bastante similar en las tres partes, idéntico en la segunda y tercera parte, por lo que tiene sentido que se obtenga el mismo valor. Continuando con el corte en m_T^H , se observa un valor más bajo en la primera parte, un valor intermedio en la segunda parte y un valor mayor en la tercera parte. Como se ha descrito previo a la tabla 11, este tipo de comportamiento en los resultados es lo que se espera de un análisis como el que se ha realizado, en el que cada método mejora al anterior. Por último, el corte en $m_T^{W^2}$ muestra la misma estructura, con un valor menor para la primera parte y un valor mayor para la tercera parte.

Es interesante realizar otro tipo de comparación en la que se compare el valor máximo posible de la FOM para cada método. Esto se realiza escogiendo el corte que haga máxima la FOM en cada canal independientemente y calculando la figura de mérito total a partir de los tres valores máximos. De este modo se obtiene:

	Primera parte	Segunda parte	Tercera parte
FOM_{total}	8.46	8.55	8.58

Tabla 12: Figura de mérito total calculada a partir del valor máximo de la FOM en cada canal.

De nuevo se observa la tendencia de que la primera parte tenga un valor menor y la tercera parte tenga un valor mayor, pero destaca la cercanía entre los tres valores, habiendo únicamente 12 centésimas entre el mayor y menor valor.

6. Conclusiones

Se ha realizado un análisis de datos procedentes de simulaciones Monte Carlo en el detector CMS. Se ha considerado como señal (S) los procesos de producción de bosones de Higgs, mientras que el resto de procesos conforman el fondo (B). El análisis se ha centrado en la maximización de la relación entre la señal y el fondo, y ha sido aplicado en el canal de desintegración del bosón de Higgs que corresponde a la desintegración del Higgs en dos bosones W, los cuales se desintegran en cuatro leptones, dos cargados y dos neutrinos. Se han seleccionado para este estudio los canales $e\mu$, e^+e^- y $\mu^+\mu^-$, y se han descartado los canales que incluyen leptones tau, debido a que su reconstrucción es menos precisa y tienen más fondos. Para cuantificar el progreso en cada paso del análisis se ha utilizado una figura de mérito, un parámetro muy habitual en el análisis de física de partículas cuando la señal es pequeña.

En el análisis se han utilizado datos de simulaciones Monte Carlo, los cuales han superado una etapa previa de filtros durante el proceso de selección y reconstrucción de eventos. Además, se les ha aplicado una selección básica para que estuvieran mejor adaptados al estudio realizado. El análisis ha comenzado con una figura de mérito total con un valor de 4.11. Con el objetivo de mejorar este parámetro y, consecuentemente, mejorar la relación entre la señal y el fondo, se ha procedido por tres métodos distintos. El primer método consistía en obtener visualmente los cortes óptimos para cada variable a partir de los histogramas pintados en el análisis. Se ha representado la señal frente a los dos fondos principales para lograr una mayor precisión a la hora de determinar qué corte era el mejor. Este método ha logrado aumentar la figura de mérito hasta 8.33, y hasta 8.46 si se hace uso de los cortes óptimos en cada canal.

El segundo método consiste en el uso de un programa que calcula la figura de mérito. Haciendo uso de este programa, es posible cortar periódicamente una variable, en este caso cada 10 GeV, para posteriormente ejecutar el programa y que calcule la figura de mérito en cada corte. De este modo se puede seleccionar el corte que maximice la figura de mérito, además de que es posible pintar una gráfica de la FOM frente a la variable correspondiente utilizando todos los valores calculados. Este método ha logrado aumentar la figura de mérito hasta un valor de 8.48, y hasta 8.55 si se calcula con los cortes óptimos de cada canal.

El tercer método se basa en el uso de un programa que realiza una optimización bin a bin de una variable determinada para obtener el corte óptimo y la FOM que se obtendría si se realizara dicho corte. Conocer la FOM antes de realizar el corte permite seleccionar la mejor variable, de modo que si se repite el proceso tras realizar el corte en la mejor variable es posible optimizar la figura de mérito teniendo en cuenta parcialmente las correlaciones entre variables. Utilizando este método se ha logrado una figura de mérito con un valor de 8.49, y hasta de 8.58 haciendo uso del corte óptimo para cada canal.

Cada método se trata de una mejora del método anterior, por lo que los resultados son coherentes. En primer lugar, el primer método aplica cortes óptimos obtenidos visualmente, lo que aumenta la FOM un 102.7%. Sin embargo, este método depende de criterios estrictamente visuales que conllevan varios errores. Esto se debe a que es prácticamente imposible mantener la misma precisión a la hora de elegir el corte óptimo al elegirlo visualmente. Además, todos los cortes se determinan en base a la selección básica, por lo que no se tiene en cuenta las correlaciones entre variables.

El segundo método mejora uno de los problemas mencionados anteriormente. Al utilizar un programa en el que los distintos cortes se calculan en intervalos continuos de 10 GeV, se logra estabilidad en las medidas para diferentes variables, ya que se ha aplicado el mismo proceso a todas. De este modo el segundo método es capaz de mejorar la FOM inicial un 106.3%. Sin embargo, este método no tiene en cuenta las correlaciones entre variables y, además, al utilizar intervalos constantes en todas las variables es posible que no tenga tanta precisión como se espera.

Por último, el tercer método es capaz de mejorar la precisión del segundo, al hacer la optimización bin a bin en vez de en segmentos arbitrarios, y además tiene en cuenta, al menos parcialmente, las correlaciones entre variables. Esto se debe a que cada vez que se obtiene un corte óptimo, se aplica el corte en cuestión y se vuelve a obtener el mejor corte para el resto de variables. Esto le permite mejorar la FOM inicial un 106.6%.

La mejora en la relación entre la señal y el fondo es considerable, pero eso no significa que no exista un margen de mejora. En primer lugar, y como en la práctica totalidad de los experimentos de física de partículas, y de la física en general, tener acceso a más datos y con más precisión permitiría mejorar los resultados obtenidos. En segundo lugar, es importante centrarse en el trabajo hecho y los pasos que podrían ser mejorados con el objetivo de optimizar la relación entre la señal y el fondo. Observando las imperfecciones de los distintos métodos utilizados es posible determinar que caminos se podrían seguir para mejorar este trabajo. Uno de ellos podría consistir en aumentar la precisión de la optimización por medio de un aumento de los bins en los histogramas originales. Este proceso permite obtener una mayor sensibilidad en las distribuciones de señal y de fondo, lo que permite observar con más claridad los picos. Además, una mayor cantidad de bins significa que el programa de automatización bin a bin tiene más datos a partir de los cuales calcular el corte óptimo, lo que mejoraría la precisión de su resultado. A pesar de las ventajas de aumentar los bins por histograma, este proceso tiene algunas desventajas, como el aumento de la incertidumbre estadística, al reducirse el número de eventos por bin, y un mayor consumo de recursos computacionales.

El otro proceso de mejora en el análisis es tener en cuenta la correlación entre las variables. A pesar de que es un concepto que se introduce en el tercer método del análisis, se hace de una manera muy simple y sin tener en cuenta varios factores implicados en la correlación de variables. La mejor manera de realizar el análisis incluyendo la correlación entre variables es crear una red neuronal. Las redes neuronales son capaces de combinar múltiples variables en una sola variable discriminante, que permite considerar correlaciones no lineales entre las distintas variables que se han englobado. Además, una red neuronal es menos dependiente de bins específicos, ya que asigna a cada evento un parámetro basado en varias variables.

Referencias

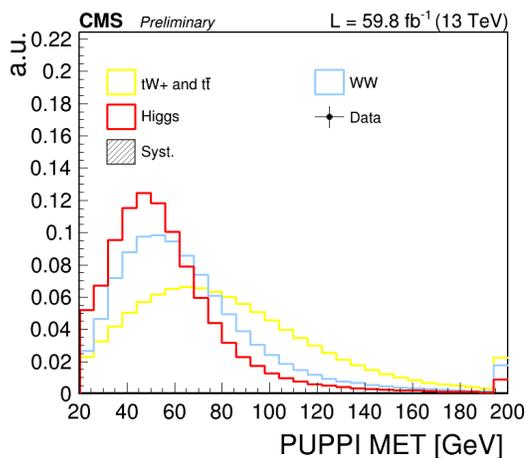
- [1] *What's so special about the Higgs boson?* URL: <https://home.cern/science/physics/higgs-boson/what>.
- [2] *The Standard Model*. URL: <https://home.cern/science/physics/standard-model>.
- [3] Tipler Paul A. y Mosca Gene. *Física para la ciencia y la tecnología. Física Moderna*. Editorial Reverté, 2024.
- [4] *Standard Model of Elementary Particles*. URL: https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.
- [5] *Nobel Prize in Physics 1979*. URL: <https://www.nobelprize.org/prizes/physics/1979/summary/>.
- [6] Peter Higgs. “Broken symmetries and the masses of the gauge bosons”. En: *Physical Review Letters* (1964).
- [7] F. Englert y R. Brout. “Broken symmetry and the mass of the gauge vector mesons”. En: *Physical Review Letters* (1964).
- [8] The CMS collaboration. “A portrait of the the Higgs boson by the CMS experiment ten years after the discovery”. En: *Nature* 607 (2022). DOI: <https://doi.org/10.1038/s41586-022-04892-x>.
- [9] *CERN experiments observe particle consistent with long-sought Higgs boson*. CERN. URL: <https://home.cern/news/press-release/cern/cern-experiments-observe-particle-consistent-long-sought-higgs-boson>.
- [10] Bureau International des Poids et Mesures. *The International System of Units(SI)*. 2019.
- [11] Francis Halzen y Alan D. Martin. *Quarks and Leptons: An Introductory course in Modern Particle Physics*. 1984.
- [12] S. Navas et al.(Particle Data Group). “Higgs boson”. En: *Phys. Rev. D* 110 (2025). URL: https://pdg.lbl.gov/2025/listings/contents_listings.html.
- [13] Unión Europea y cooperación Ministerio de asuntos exteriores. URL: <https://www.exteriores.gob.es/RepresentacionesPermanentes/oficinadelasnacionesunidas/es/Organismo/Paginas/Convenios-y-otras-00II/CERN.aspx%5D>.
- [14] *Where did it all begin?* URL: <https://home.cern/about/who-we-are/our-history>.
- [15] Guillaume Grallet. “Le Web a été inventé... en France !” En: *LePoint.Fr* (2012).
- [16] Aquae Fundación. *Medipix y sus aplicaciones en la salud y el arte*. URL: <https://www.fundacionaquae.org/medipix-y-sus-aplicaciones-en-la-salud-y-el-arte/#:~:text=El%20chip%20Medipix1%20fue%20un%20proyecto%20que,de%20N%C3%A1poles%20y%20la%20Universidad%20de%20Pisa.&text=En%20el%20a%C3%B1o%201999%20nace%20el%20Medipix2,utilizando%20un%20proceso%20CMOS%20de%200%2C25%20%CE%BCm..>
- [17] *CERN's accelerator complex*. URL: <https://home.cern/science/accelerators/accelerator-complex>.
- [18] *The CERN accelerator complex, layout in 2022*. URL: <https://cds.cern.ch/images/CERN-GRAPHICS-2022-001-1>.

- [19] *The CERN Experimental Programme*. URL: <https://greybook.cern.ch/experiment/detail?id=CMS>.
- [20] *Detector*. URL: <https://cms.cern/detector>.
- [21] *Detecting Muons*. URL: <https://cms.cern/detector/detecting-muons>.
- [22] *Acercándonos al LHC: CMS*. URL: https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.cms/.
- [23] *Detector*. URL: <https://cms.cern/detector>.
- [24] Sergio Cittolin, Attila Rácz y Paris Sphicas. “CMS The TriDAS Project. Technical Design Report, Volume 2: Data Acquisition and High Level Trigger”. En: *CMS* (2002). URL: <https://cds.cern.ch/record/578006/?ln=es>.
- [25] A.M. Sirunyan et al.(The CMS Collaboration). *Particle-flow reconstruction and global event description with the CMS detector*. 2017. DOI: <https://doi.org/10.1088/1748-0221/12/10/P10003>. URL: <https://iopscience.iop.org/article/10.1088/1748-0221/12/10/P10003/pdf>.
- [26] D. Buskulic et al.(ALEPH collaboration). *Performance of the ALEPH detector at LEP*. 1995.
- [27] *El Gran Colisionador de Electrones y Positrones*. URL: <https://home.cern/science/accelerators/large-electron-positron-collider>.
- [28] The CMS Collaboration. *Measurement of Tracking Efficiency*. 2010. URL: <https://cds.cern.ch/record/1279139>.
- [29] The CMS Collaboration. *Performance of missing transverse momentum reconstruction in proton-proton collisions at $s = 13$ TeV using the CMS detector*. 2019. URL: <https://iopscience.iop.org/article/10.1088/1748-0221/14/07/P07004/pdf>.
- [30] Jónatan Piedra. *HEP/Full2018_v9*. URL: https://github.com/piedraj/HEP/blob/main/Full2018_v9/samples.py.
- [31] The CMS Collaboration. *The Phase-2 Upgrade of the CMS Data Acquisition and High Level Trigger*. 2021. URL: <https://cds.cern.ch/record/2759072?ln=es>.
- [32] The CMS Collaboration. *Measurement of the inclusive and differential Higgs boson production cross sections in the leptonic WW decay mode at $s = 13$ TeV*. 2021. DOI: [https://doi.org/10.1007/JHEP03\(2021\)003](https://doi.org/10.1007/JHEP03(2021)003).
- [33] Giovanni Punzi. *Sensitivity of searches for new signals and optimization*. 2003. DOI: [10.48550/ARXIV.PHYSICS/0308063](https://arxiv.org/abs/physics/0308063). URL: <https://arxiv.org/abs/physics/0308063>.
- [34] *About ROOT*. URL: <https://root.cern/about/>.
- [35] 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*. 2019.

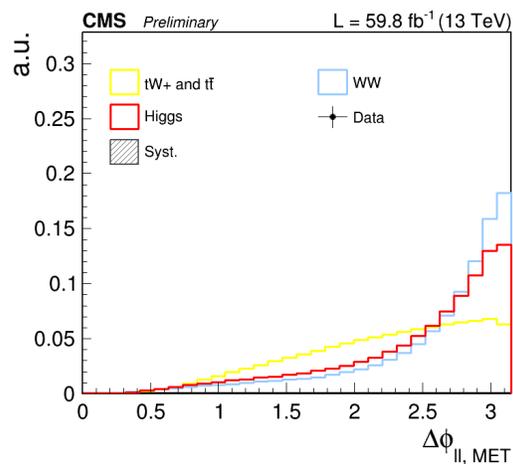
Apéndices

Apéndice 1

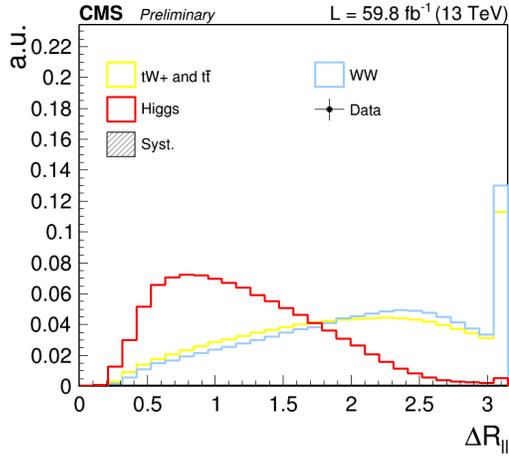
En este apéndice se muestran los *plots* de *shapes* de las variables que finalmente no se han considerado discriminantes. Se puede observar bastante variedad en las distribuciones. En primer, se tienen variables como η_1 o η_2 en las que la distribución de la señal y los dos fondos principales es prácticamente idéntica, por lo que aplicar un corte a estas variables eliminaría una cantidad considerable de señal. Por otra parte, variables como puppimet o $\Delta\phi_{\text{met}}$ muestran dos distribuciones que se comportan de forma similar, siendo estas distribuciones la señal y uno de los fondos, mientras que la distribución del otro fondo es ligeramente diferente. Al intentar aplicar un corte en este tipo de variables se lograría reducir considerablemente un fondo, pero al mantener el otro fondo, ya que su distribución es bastante similar a la señal, no se consigue una gran mejora en la relación entre la señal y el fondo. Por último, hay variables como $\Delta R_{\ell\ell}$ o $p_T^{\ell_1}$ cuya forma podría indicar que son una variable discriminante. Las distribuciones de los fondos son similares entre ellas y distintas a la señal, por lo que sería posible aplicar un corte que eliminara la mayoría de los fondos manteniendo la señal. Sin embargo, no se está teniendo en cuenta la correlación entre variables, y al aplicar cortes a estas variables no se logra una mejora significativa en la FOM, llegando incluso a empeorarla en algunos casos.



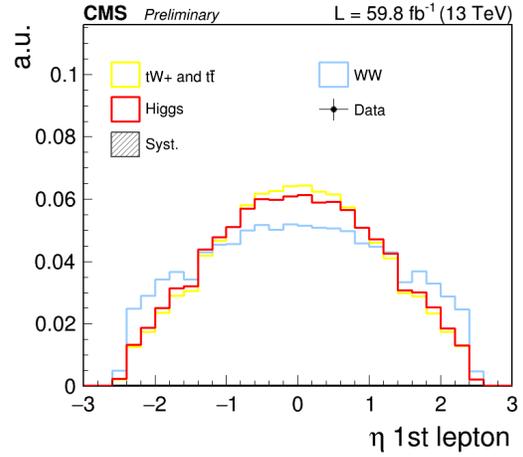
(a) Histograma de la variable puppimet .



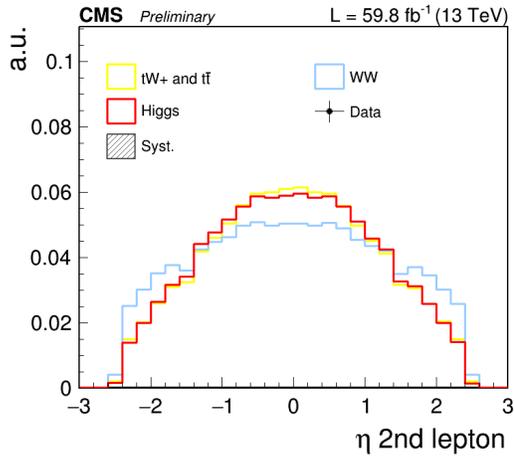
(b) Histograma de la variable $\Delta\phi_{\text{met}}$.



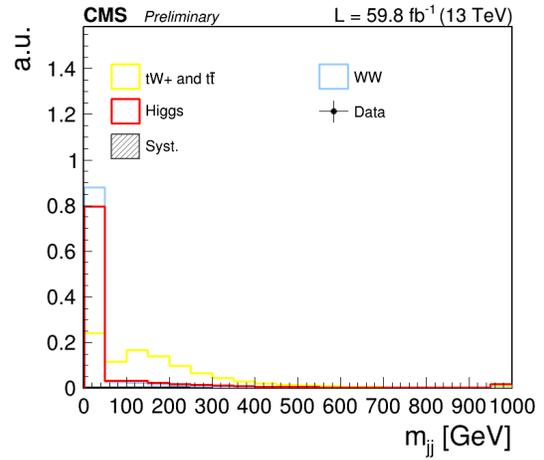
(c) Histograma de la variable $\Delta R_{\ell\ell}$.



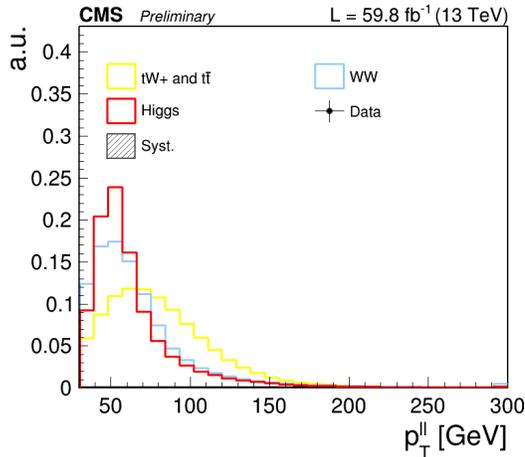
(d) Histograma de la variable η_1 .



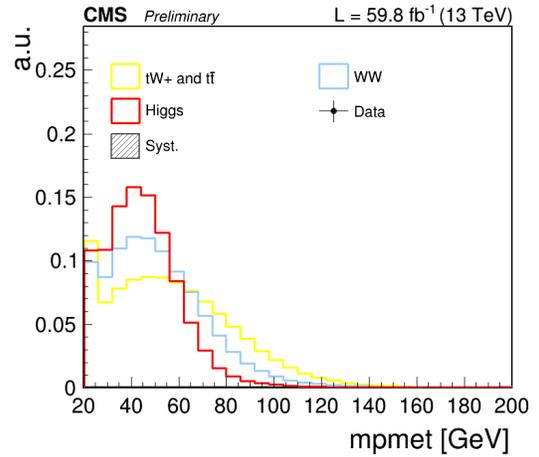
(e) Histograma de la variable η_2 .



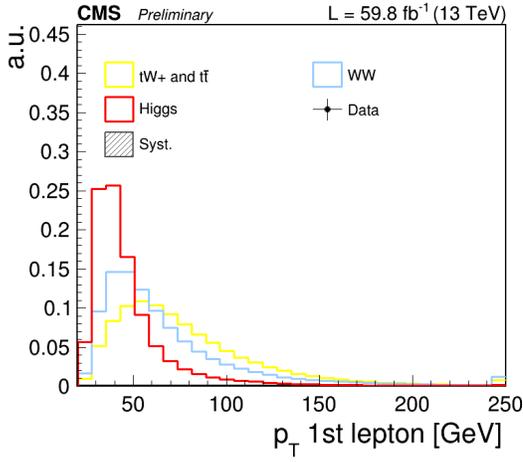
(f) Histograma de la variable m_{jj} .



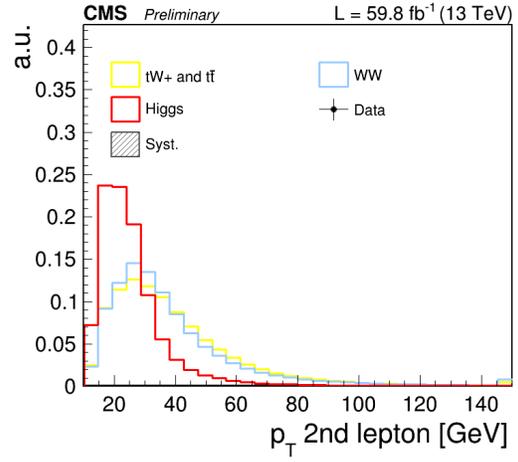
(g) Histograma de la variable $p_T^{\ell\ell}$.



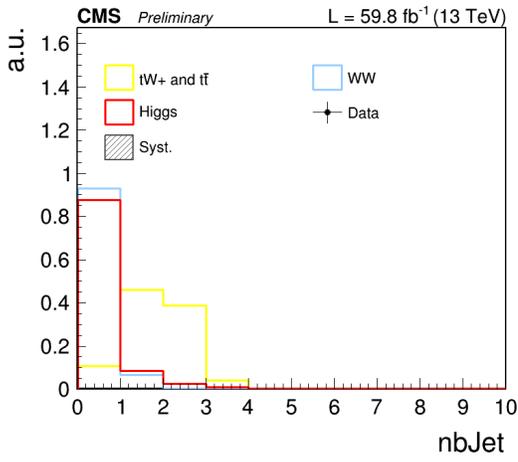
(h) Histograma de la variable m_{pmet} .



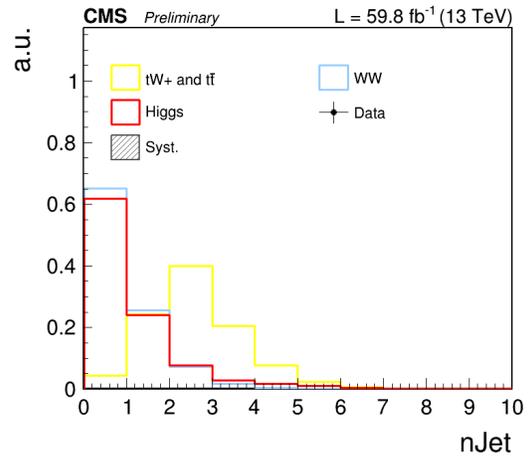
(i) Histograma de la variable $p_T^{\ell_1}$.



(j) Histograma de la variable $p_T^{\ell_2}$.



(k) Histograma de la variable nbjet.



(l) Histograma de la variable njet.

Apéndice 2

En este apéndice se van a mostrar los conocimientos nuevos que se han necesitado para completar este trabajo.

En primer lugar, para operar con los datos simulados y la red de análisis del CERN se ha necesitado un nivel básico de Linux. Esto comprende conocer los comandos básicos para moverse entre carpetas y manipular documentos. Entre estos comandos básicos se encuentran **cd**, **mv**, **rm**, **mkdir** y **scp**, que han sido los más utilizados durante el trabajo. El *framework* de análisis de datos utilizado y desarrollado en el CERN es ROOT, el cual requiere conocimientos en C++ y en python. Algunos de los comandos aprendidos para operar con este entorno de trabajo han sido **root -l fichero** o **Events->MakeClass(Archivo)**. Para editar archivos se ha utilizado el editor de texto emacs, desarrollado en C, para el que se han aprendido nuevos comandos como **Ctrl x u**, **Ctrl x Ctrl c** o **Ctrl x Ctrl s**.

Por otra parte, el paso inicial para comenzar el análisis ha sido aprender a leer los datos provenientes de las simulaciones Monte Carlo de CMS, los cuales están en formato ROOT. Una vez se tiene acceso a los datos, el análisis se ha dividido principalmente en tres procesos. En primer lugar, se ha aprendido a pintar los histogramas con los datos disponibles y a extraer información de ellos. Además, editando el fichero plot.py se ha podido comprender cómo se pintan los histogramas y qué señales se utilizan para pintar cada proceso físico. Una vez se sabe cómo se pintan los histogramas es posible modificar los mismos para pintar sólo los procesos que interesan, como en el caso de los *plots* de *shapes*, en los que sólo se ha pintado la señal y los dos fondos principales. Por otra parte, se ha aprendido a implementar cortes y selecciones modificando el fichero cuts.py. Este fichero está dividido en dos secciones: tmp y cuts. En la sección tmp se ha aplicado la selección básica, ya que esta sección afecta a todos los cortes que se introduzcan en cuts. La sección cuts permite implementar los cortes agrupados en categorías. De este modo, en este análisis se han creado tres categorías, que corresponden a los distintos canales estudiados. En cada categoría se ha aplicado un primer corte que define qué tipo de partícula se requiere, electrones o muones, utilizando el esquema de numeración de partículas Monte Carlo. Tras definir el tipo de partículas, se pueden aplicar diferentes cortes independientes. El hecho de ser independientes permite, en la primera parte del análisis, crear tres cortes con la siguiente estructura: el primer corte se aplica solo a una variable, el segundo corte contiene cortes a dos variables y el tercer corte a las tres variables. De esta forma, se puede observar la evolución de la FOM al aplicar los cortes de forma secuencial. Tras cada modificación del fichero cuts.py es necesario compilar, para aplicar los cambios realizados, y es recomendable realizar el análisis con un número de eventos limitado en el sistema local, para así asegurarse de que todo funciona correctamente. Una vez comprobado que las modificaciones no conducen a errores, se pueden mandar los jobs a la red de análisis del CERN. Por último, se han creado algoritmos para leer los datos de los histogramas y maximizar la FOM. Estos algoritmos han necesitado de conocimientos en ROOT para poder leer los datos, de linux para poder acceder al fichero donde están los datos y de python para generar las instrucciones que debía seguir el algoritmo. Los algoritmos son capaces de leer una categoría exclusivamente e ignorar las otras dos, o incluso leer los datos correspondientes a un corte determinado dentro de una categoría. El programa encargado de obtener la figura de mérito utilizado en la segunda parte del análisis es capaz de acceder únicamente a los datos de los cortes que se le indicara.