

Búsqueda del Bosón de Higgs en el
Canal $H \rightarrow ZZ^{(*)} \rightarrow 4\mu$ en CMS
Empleando un Método de Análisis
Multivariado.

A. Alonso Díaz

Toda correspondencia en relación con este trabajo debe dirigirse al Servicio de Información y Documentación, Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas, Ciudad Universitaria, 28040-MADRID, ESPAÑA.

Las solicitudes de ejemplares deben dirigirse a este mismo Servicio.

Los descriptores se han seleccionado del Thesaurus del DOE para describir las materias que contiene este informe con vistas a su recuperación. La catalogación se ha hecho utilizando el documento DOE/TIC-4602 (Rev. 1) Descriptive Cataloguing On-Line, y la clasificación de acuerdo con el documento DOE/TIC.4584-R7 Subject Categories and Scope publicados por el Office of Scientific and Technical Information del Departamento de Energía de los Estados Unidos.

Se autoriza la reproducción de los resúmenes analíticos que aparecen en esta publicación.

Catálogo general de publicaciones oficiales
<http://www.060.es>

Depósito Legal: M -14226-1995
ISSN: 1135 - 9420
NIPO: 654-07-061-6

Editorial CIEMAT

CLASIFICACIÓN DOE Y DESCRIPTORES

S72

HIGGS BOSONS; CERN; CERN LHC; MULTI-CHANNEL ANALYZERS; HIGH ENERGY PHYSICS; MULTIVARIATE ANALYSIS

**Búsqueda del Bosón de Higgs en el Canal $H \rightarrow ZZ^{(*)} \rightarrow 4\mu$ en CMS
Empleando un Método de Análisis Multivariado.**

Alonso Díaz, A.

20 pp. 10 figs. 8 refs.

Resumen:

En esta nota se presenta un análisis para la búsqueda del bosón de Higgs en el detector CMS del acelerador LHC (CERN, Ginebra, Suiza) en el canal $H \rightarrow ZZ^{(*)} \rightarrow 4\mu$ empleando un método multivariado. El análisis se basa en construir un likelihood dependiente de la masa del bosón de Higgs a partir de variables discriminantes. Este análisis supone una mejora significativa en el potencial de descubrimiento del bosón de Higgs en un amplio rango de masas respecto al análisis publicado por CMS, basado en simples cortes ortogonales independientes de la masa del bosón de Higgs.

**Search for the Higgs Bosón in the $H \rightarrow ZZ^{(*)} \rightarrow 4\mu$ Channel in CMS
Using a Multivariate Analysis.**

Alonso Díaz, A.

20 pp. 10 figs. 8 refs.

Abstract:

This note presents a Higgs boson search analysis in the CMS detector of the LHC accelerator (CERN, Geneva, Switzerland) in the $H \rightarrow ZZ^{(*)} \rightarrow 4\mu$ channel, using a multivariate method. This analysis, based in a Higgs boson mass dependent likelihood, constructed from discriminant variables, provides a significant improvement of the Higgs boson discovery potential in a wide mass range with respect to the official analysis published by CMS, based in orthogonal cuts independent of the Higgs boson mass.

Índice

1. Introducción	1
2. Mecanismo de Higgs en el Modelo Estándar	1
3. Dispositivo Experimental	2
3.1. El acelerador LHC	2
3.2. El detector CMS	3
4. Producción y desintegración del bosón de Higgs en LHC	4
4.1. Estudio del canal $H \rightarrow ZZ^{(*)} \rightarrow 4\mu$	5
5. Análisis	6
5.1. Análisis independiente de m_H	6
5.2. Análisis empleando métodos multivariados	8
5.3. Análisis del canal $H \rightarrow ZZ^{(*)} \rightarrow 4\mu$ con el método de likelihood	9
5.4. Potencial de descubrimiento del bosón de Higgs	10
6. Conclusiones	14

1. Introducción

En 1964 Peter Higgs propuso una solución para explicar cómo adquieren masa las partículas en el Modelo Estándar [1]. Esta solución, conocida como el Mecanismo de Higgs [2], requiere de la existencia de una partícula escalar: el bosón de Higgs, cuya masa no está predicha por la teoría. Desde los años 80 su búsqueda ha sido prioritaria en gran parte de los experimentos de altas energías. Estas búsquedas, junto con medidas de precisión del Modelo Estándar, han permitido fijar límites inferiores y superiores a su masa. Los dos experimentos de propósito general del acelerador LHC [3], CMS y ATLAS, permitirán realizar búsquedas directas hasta masas en torno a $1 \text{ TeV}/c^2$.

Uno de los canales más limpios para descubrir el bosón de Higgs en LHC es $H \rightarrow ZZ^{(*)} \rightarrow 4\mu$, en el que el bosón de Higgs se desintegra en dos bosones Z, cada uno de los cuales da lugar a una pareja muon-antimuon ($\mu^+\mu^-$). Este canal producirá sucesos con cuatro muones aislados, de alto momento transverso, originados en un mismo vértice, y cuya masa invariante tendrá una estructura resonante en torno al valor de la masa del bosón de Higgs, m_H . El diseño de CMS [4] está optimizado para la detección y medida precisa del momento/energía de leptones (e, μ, τ) y fotones (γ) aislados.

En este trabajo se presenta una estrategia para mejorar, respecto a los estudios realizados hasta la fecha en CMS [5, 6], el potencial de descubrimiento del bosón de Higgs usando el canal $H \rightarrow ZZ^{(*)} \rightarrow 4\mu$. En estos estudios previos se aísla la señal de los procesos de contaminación a través de criterios ortogonales de selección aplicados a las variables discriminantes medidas con el detector. Estos criterios de selección no están optimizados para ningún valor concreto de m_H .

El estudio que aquí se presenta utiliza métodos multivariados de análisis que combinan las variables discriminantes en una distribución de probabilidad, que permite clasificar los sucesos en función de su compatibilidad con la señal. Esta distribución de probabilidad es diferente para cada valor de m_H y tiene en cuenta las correlaciones entre las variables discriminantes, posibilitando la optimización del análisis para cada hipótesis de masa del bosón de Higgs. La elección óptima de variables discriminantes en cada rango de m_H y la evaluación del incremento del potencial de descubrimiento del bosón de Higgs en función de m_H son los puntos centrales de este trabajo.

Esta memoria comienza con una breve reseña sobre el mecanismo de Higgs. Luego describe el acelerador LHC y el detector CMS, para comprender el dispositivo experimental de producción y detección de sucesos. A continuación entra en el estudio de las características cinemáticas de los procesos de señal y contaminación, así como de los distintos métodos estadísticos empleados para su discriminación. La memoria finaliza presentando la mejora del potencial de descubrimiento que supone la utilización del método de análisis multivariado aquí propuesto.

2. Mecanismo de Higgs en el Modelo Estándar

El Modelo Estándar de partículas elementales describe las interacciones electromagnética, débil y fuerte entre los constituyentes fundamentales de la materia. Se trata de una teoría cuántica de campos consistente con la mecánica cuántica y la relatividad especial. Principalmente se compone de dos teorías, la cromodinámica cuántica (QCD) y las interacciones electrodébiles, y se basa en que las interacciones entre partículas vienen dadas por el intercambio de bosones. Para la interacción fuerte estos bosones son los gluones, para la débil los bosones Z y W^\pm y la electromagnética está mediada por fotones.

Una parte clave del modelo es explicar el origen de las masas de las partículas. En 1964 Peter Higgs introduce un nuevo campo escalar, el campo de Higgs, que al interactuar con las partículas las dota de masa. Este campo rompe la simetría electrodébil mediante el mecanismo de ruptura espontánea de simetría, introduciendo términos de masa en el lagrangiano del Modelo Estándar. El acoplo del campo de Higgs al resto de partículas es directamente proporcional a la masa de éstas. Aparece además el bosón de Higgs, que es la excitación del campo de Higgs. Su masa no está predicha por la teoría y por tanto su descubrimiento es una pieza fundamental para la verificación del Modelo Estándar.

Desde que se postuló la existencia del bosón de Higgs, multitud de experimentos han tratado de

encontrarlo proporcionando una serie de restricciones teóricas y experimentales a la masa que ha de tener. Las limitaciones teóricas provienen de argumentos de consistencia para el Modelo Estándar [7]. Los valores máximos y mínimos aceptados por la teoría dependen del factor Λ , que es la escala de renormalización. Este hecho tiene su origen en que la masa del bosón de Higgs está afectada por las correcciones radiativas de las masas de fermiones y bosones. El límite superior se obtiene haciendo que el acoplo de los quarks con el potencial de Higgs permanezca finito. El límite inferior corresponde a la región a partir de la cual el acoplo de los quarks con el potencial de Higgs se hace negativo obteniendo un vacío inestable. Los experimentos de LHC pondrán a prueba las predicciones del Modelo Estándar hasta la escala de energía del orden de 1 TeV.

Desde el punto de vista experimental, la búsqueda directa más sensible se ha realizado en el acelerador LEP [8], dando un límite inferior de $m_H > 114.4 \text{ GeV}/c^2$ al 95% de nivel de confianza. Por otro lado, medidas de alta precisión de parámetros electrodébiles llevadas a cabo en LEP, junto con medidas de la masa del quark *top* realizadas en Tevatron, proporcionan un límite superior [9] de $m_H < 225 \text{ GeV}/c^2$ al 95% de nivel de confianza.

3. Dispositivo Experimental

3.1. El acelerador LHC

El *Large Hadron Collider* [3] (LHC) es el colisionador protón-protón que está siendo terminado de instalar en el tunel de 27 Km del CERN, en las proximidades de Ginebra. Está previsto que los primeros haces comiencen a circular en Mayo de 2008, esperando que se produzcan las primeras colisiones a mediados del mismo año. Sus características más importantes son una energía en el centro de masas, \sqrt{s} , de 14 TeV y una luminosidad¹ de $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$.

LHC está formado por dos anillos por los que circularán sendos haces de protones en sentidos opuestos. Los haces estarán inmersos en un vacío continuo y guiados por imanes. Para poder confinar partículas de tan alta energía, los imanes tendrán que proporcionar un campo magnético muy alto (8.33 Tesla para $\sqrt{s} = 14 \text{ TeV}$). La única solución posible ha sido emplear imanes superconductores refrigerados a 1.3 K. Cada haz estará compuesto de 2808 paquetes de $15 \mu\text{m}$ de radio y 53 mm de largo los cuales, inyectados a una energía de 450 GeV procentes del SPS, se acelerarán en las cavidades de radio frecuencia hasta llegar a 7 TeV. Tras esto, se mantendrán confinados durante horas para conseguir condiciones estables dando vueltas al anillo a una frecuencia de 40 MHz. En los distintos puntos de interacción de LHC donde están los experimentos, se producirán colisiones protón-protón. En cada cruce de paquetes se producirán en promedio 20 colisiones, por lo tanto se tomará datos a una tasa de 1 GHz. Este nivel elevado de colisiones producto de una luminosidad elevada ($10^{34} \text{ cm}^{-2} \text{ s}^{-2}$) permite la producción de procesos de muy baja sección eficaz, entre ellos la producción del bosón de Higgs.

La sección eficaz de producción del bosón de Higgs aumenta rápidamente con la energía en el sistema en centro de masas de la colisión protón-protón. LHC tendrá una sección eficaz dos órdenes de magnitud superior a Tevatron. Por otro lado, la sección eficaz total en colisiones protón-protón permanece prácticamente constante al aumentar la energía. Esto hace que para la energía de LHC se esté en un caso mucho más favorable que en cualquier otro experimento. Pese a esto, el cociente entre la sección eficaz del Higgs y la total es superior a 9 órdenes de magnitud. De ahí la necesidad de acumular una gran luminosidad para poder obtener suficientes sucesos de Higgs para discriminarlos sobre la contaminación.

Al producirse los cruces de haces a una frecuencia de 40 MHz, y cada suceso ocupar cerca de 1 MB, sería necesario almacenar 40 TB/s. De aquí que se necesite hacer una selección de sucesos en tiempo real mediante un *trigger*, suprimiendo sucesos de física no relevante hasta tomar datos a una tasa de 150 Hz. Pese a ser un valor mucho más bajo que el inicial, si tenemos en cuenta que LHC tomará datos 10^7 segundos por año, serán necesarios 2×10^6 Gigabytes al año y una capacidad de cómputo que requiere

¹Número de colisiones por unidad de tiempo y área efectiva de los haces, $L = f \frac{n_1 n_2}{A}$, donde f es la frecuencia de cruce de los haces compuestos por n_1 y n_2 partículas, y A es el solapamiento de los haces.

decenas de miles de ordenadores personales actuales. Esto ha llevado al desarrollo de un sistema de computación distribuida conocido como Grid.

En LHC hay cuatro experimentos principales situados en puntos de interacción. Dos de ellos, CMS y ATLAS, son multiprósito, con objetivos similares como la búsqueda del bosón de Higgs, supersimetría y dimensiones extra, pero con soluciones técnicas distintas. Otro de los experimentos es ALICE, pensado para el estudio de las colisiones entre iones pesados. El cuarto es LHCb, cuya principal misión será el estudio de la física del quark b y la violación de simetría CP.

3.2. El detector CMS

El *Compact Muon Solenoid* (CMS) [4] es un detector de propósito general. Su característica principal es el empleo de un solenoide superconductor capaz de generar un campo uniforme de 4 T en su interior. Las prioridades de diseño fueron tener un sistema redundante para medir muones, un buen calorímetro electromagnético y un sistema de trazas de gran resolución. La estructura de CMS está basada en simetría cilíndrica. Por el eje del cilindro entran los haces que colisionan en el centro del detector. El cilindro, denominado barril, está cerrado por dos discos, llamados *endcaps*. Las dimensiones totales son 21.6 m de largo y 15 m de diámetro, con un peso total de 14500 toneladas. En la figura 1 podemos ver cada una de las partes que forma el detector, que se describen en detalle a continuación.

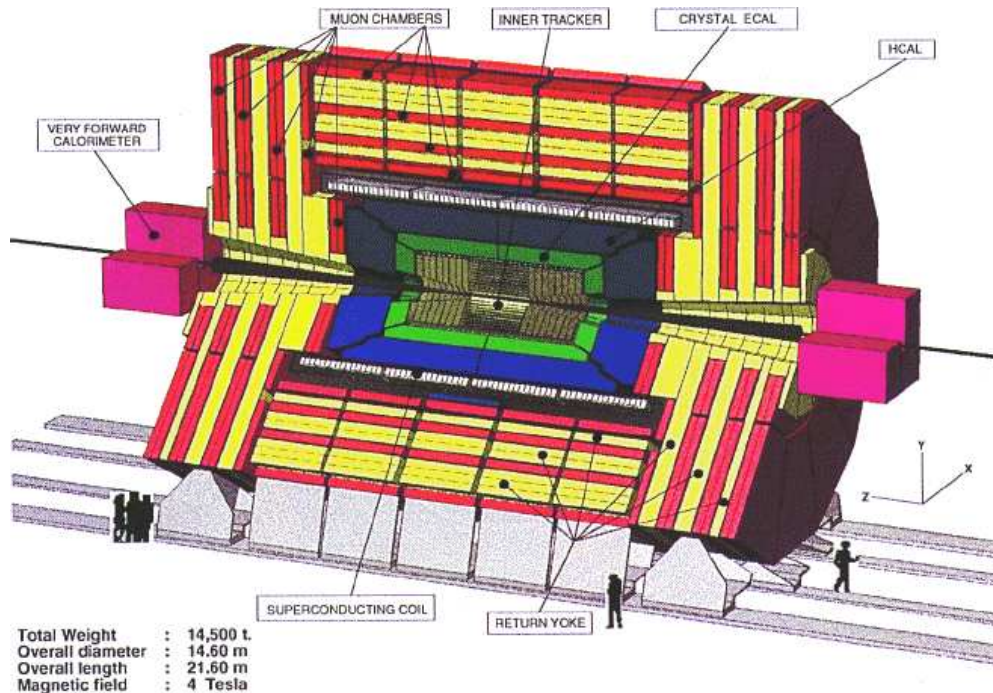


Figura 1: Esquema del detector CMS

El *tracker* está situado alrededor del punto de interacción. Permite medir las trazas de las partículas cargadas. Está optimizado para reconstruir trazas de muones con alto momento transversal, electrones aislados y hadrones cargados, con una resolución para momentos de leptones aislados $\sim 1\%$ y una eficiencia superior al 98%. Permite también la reconstrucción de vértices secundarios cercanos al punto de interacción. Construido por detectores de píxeles de silicio en la parte central (85 millones de píxeles) y por detectores de tiras de silicio en el exterior (11.4 millones de píxeles), con una superficie de detección de más de 215 m².

El calorímetro electromagnético (ECAL) está diseñado para medir con precisión la energía de electrones y fotones. Está formado por 74848 cristales de PbWO₄ que recogen la luz emitida por las partículas incidentes, luz que es medida con fotodiodos de avalancha de silicio.

El calorímetro Hadrónico (HCAL) está concebido para medir la dirección de los *jets* y calcular la energía transversa depositada.

El solenoide superconductor es la pieza clave del detector. Tiene unas dimensiones de 13 m de longitud y 5.9 m de diámetro, lo que le lleva a ser el imán superconductor más grande del mundo. Estas dimensiones le permiten alojar en su interior los calorímetros y el *tracker*. Por sus bobinas circularán 20 kA, que producirán un campo interior de 4 T. Este campo es necesario para poder curvar las trayectorias de las partículas cargadas con alto momento transverso y así obtener una reconstrucción excelente de los momentos. Finalmente, una estructura de 12000 toneladas de hierro se encarga del retorno del campo por el exterior del solenoide.

Las cámaras de muones constituyen el subdetector más extenso de CMS. Están situadas fuera del imán, por lo que aprovechan el campo de retorno de 1.8 T. Su misión es identificar y detectar muones y, junto con el *tracker*, proporcionar una medida precisa del momento. Las cámaras de muones tienen además un papel crucial a efectos de *trigger*, ya que los muones de alto momento transverso proporcionarán señales muy claras y limpias para un amplio rango de procesos físicos. En la región del barril ($0 < \eta < 1.1$)² están construidas con cámaras de deriva, en la zona de los *endcaps* ($0.9 < \eta < 2.4$) por CSC (*Cathode Strip Chambers*) y en la región intermedia por RPC (*Resistive Plate Chambers*).

4. Producción y desintegración del bosón de Higgs en LHC

En LHC el bosón de Higgs se producirá a través de la interacción de pares de gluones y quarks. En la figura 2 pueden verse los diagramas de Feynman para los principales modos de producción del bosón de Higgs. El mecanismo dominante es la fusión de gluones ($gg \rightarrow H$). Para masas del bosón de Higgs próximas a 1 TeV la contribución de la fusión de bosones vectoriales es igualmente importante ($qq \rightarrow Hqq$).

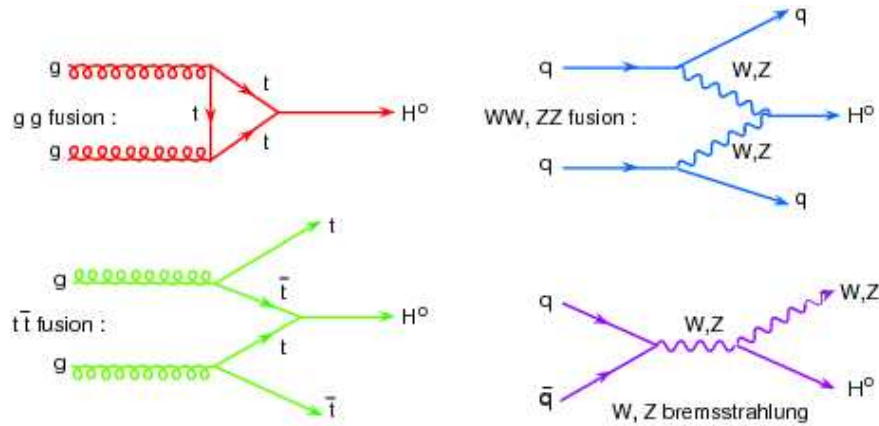


Figura 2: Principales diagramas de Feynman para la producción del bosón de Higgs en colisiones hadrónicas.

En la figura 3 (superior) se presenta la sección eficaz de producción del bosón de Higgs en función de su masa para cada uno de estos procesos. La sección eficaz total varía entre 20 pb para $m_H = 100 \text{ GeV}/c^2$ y 0.2 pb para $m_H = 1 \text{ TeV}/c^2$.

El bosón de Higgs decae en pares de bosones y fermiones, con una constante de acoplo que es proporcional a la masa de sus productos de desintegración. Las anchuras parciales de desintegración del bosón de Higgs están representadas en la figura 3 (inferior) en función de m_H .

Para $m_H > 150 \text{ GeV}/c^2$, los modos de desintegración dominantes son $H \rightarrow ZZ^{(*)}$ y $H \rightarrow WW^{(*)}$. Los canales más limpios para la detección del bosón de Higgs son aquellos con leptones en el estado final, $H \rightarrow WW^{(*)} \rightarrow \ell\nu_\ell\ell\bar{\nu}_\ell$ y $H \rightarrow ZZ^{(*)} \rightarrow \ell^+\ell^-\ell^+\ell^-$, dominando el primero en el rango $130 \text{ GeV}/c^2 <$

²Pseudorapidez $\eta = -\ln(\tan(\frac{\theta}{2}))$, siendo θ el ángulo polar respecto al eje del cilindro de CMS.

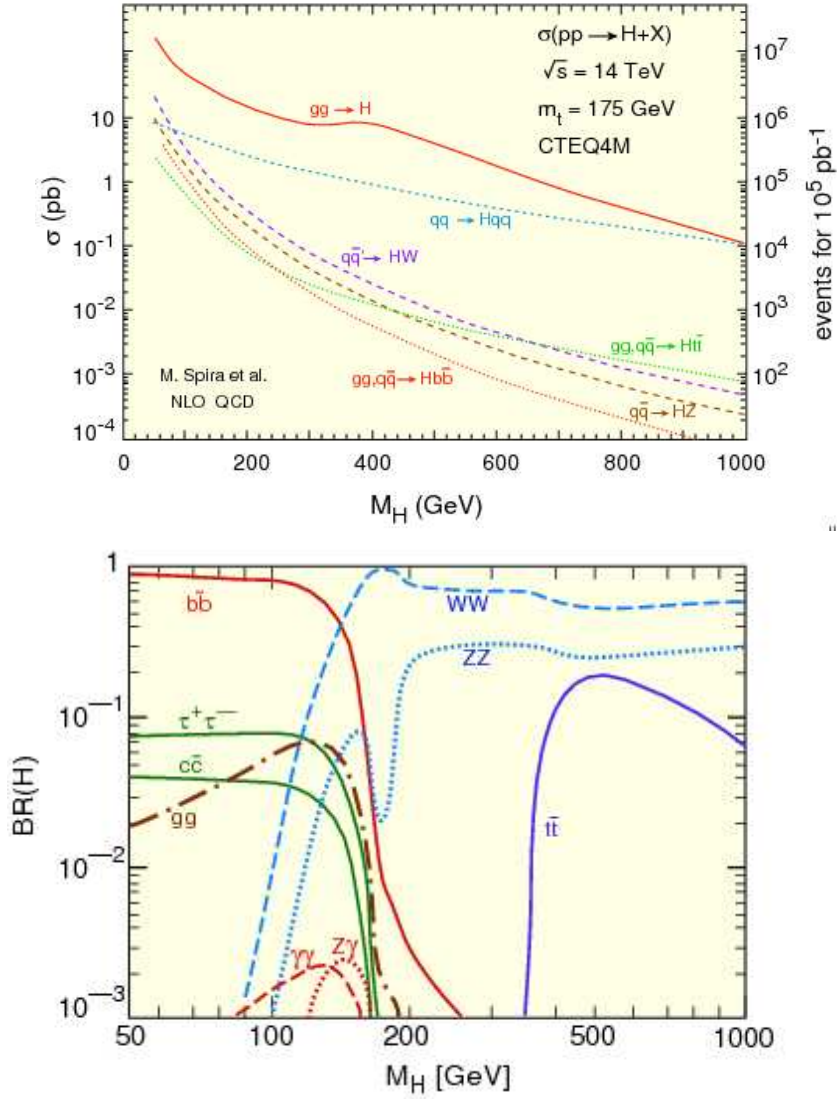


Figura 3: Secciones eficaces de producci3n del bos3n de Higgs para colisiones pp (superior) y anchuras parciales de la desintegraci3n del bos3n de Higgs (inferior).

$m_H < 180 \text{ GeV}/c^2$ (en torno a $2m_W$). En CMS 3stos son dos de los llamados *golden channels* para el descubrimiento del Higgs por la limpieza de sus signaturas en el detector y por la buena resoluci3n de CMS en la medida del momento de leptones aislados.

4.1. Estudio del canal $H \rightarrow ZZ^{(*)} \rightarrow 4\mu$

Este trabajo se centra en el estudio del canal de desintegraci3n $H \rightarrow ZZ^{(*)} \rightarrow 4\mu$. Los sucesos de este tipo contienen cuatro muones en el estado final ($2\mu^+$ y $2\mu^-$) aislados, con alto momento transversal y v3rtice primario com3n. La masa invariante de los dos pares $\mu^+\mu^-$ es aproximadamente m_Z , excepto en el caso $m_H < 2m_Z$, en que uno de los pares $\mu^+\mu^-$ procede de un bos3n Z virtual. La secci3n eficaz del proceso $H \rightarrow ZZ^{(*)} \rightarrow 4\mu$, teniendo en cuenta las anchuras parciales de $H \rightarrow ZZ$ y $Z \rightarrow \mu^+\mu^-$, var3a entre 0.5 fb para $m_H = 120 \text{ GeV}/c^2$ y 5.5 fb para $m_H = 250 \text{ GeV}/c^2$.

En LHC habr3a un elevado n3mero de sucesos, resultado de procesos inel3sticos est3ndar (sin mediar el bos3n de Higgs), que tendr3n cuatro muones en el estado final y supondr3n una contaminaci3n importante. Algunos de estos sucesos poseen caracter3sticas cinem3ticas distintas a la se3al por lo que

son relativamente sencillos de suprimir. Los modos de contaminación que se pueden reducir fácilmente son: $pp \rightarrow t\bar{t} \rightarrow W^+W^-b\bar{b}$ y $pp \rightarrow t\bar{t} \rightarrow Z^{(*)}/\gamma^*b\bar{b}$, con una sección eficaz de 840 pb y 278 pb respectivamente. Sin embargo, los procesos $pp \rightarrow (Z^{(*)}/\gamma^*)(Z^{(*)}/\gamma^*)$ son muy similares a la señal de modo que su discriminación es más complicada. La sección eficaz de este proceso es aproximadamente 175 pb.

En los sucesos de la contaminación $Z^{(*)}/\gamma^*b\bar{b}$, un par $\mu^+\mu^-$ proviene de la desintegración del bosón Z, mientras que el otro par está originado por la cascada hadrónica producida por el par quark-antiquark, lo que hace que en el entorno de las trazas de los muones haya gran cantidad de partículas y no estén aislados. Además, la masa invariante de este último par $\mu^+\mu^-$ no se corresponde con m_Z .

Los cuatro muones procedentes de la desintegración $t\bar{t} \rightarrow W^+W^-b\bar{b}$ tienen características similares a los de los procesos $Z^{(*)}/\gamma^*b\bar{b}$. La contaminación de sucesos $t\bar{t}$ es más fácil de suprimir porque no hay ningún bosón Z en el suceso.

Los sucesos de contaminación ZZ se producen a través de $q\bar{q} \rightarrow ZZ^*$ y $gg \rightarrow ZZ^*$. Ambos pares $\mu^+\mu^-$ son originados por bosones Z, de modo que se obtiene una topología muy similar a la señal, con la salvedad de que en este caso la masa invariante de los cuatro muones, $m_{4\mu}$, no se corresponde con una resonancia, como en el caso del bosón de Higgs.

La figura 4 (izquierda) muestra las distribuciones de $m_{4\mu}$ esperadas para los sucesos de señal para varios valores de m_H y para los procesos de contaminación, después de exigir que los sucesos verifiquen unos criterios de preselección: tener dos muones y dos antimuones de momento transverso superior a 7 GeV/c y que la masa invariante de los dos pares $\mu^+\mu^-$ sea mayor que 12 GeV/c². A este nivel de selección la contaminación se ha reducido en un factor 10⁶, mientras que la señal apenas se ve afectada.

5. Análisis

Este estudio se basa en el análisis de muestras de sucesos simulados por métodos Monte Carlo, tanto de señal como de contaminación. En las producciones oficiales de CMS, se generan muestras para 18 hipótesis de masa, desde 120 GeV/c² hasta 600 GeV/c², y para cada uno de los procesos de contaminación, utilizando los programas estándar de simulación de interacciones entre partículas elementales, PYTHIA [10] y CompHEP [11]. La respuesta del detector se simula usando el software oficial de simulación, basado en GEANT [12], y reconstrucción de CMS.

El número de sucesos simulado está entre uno y dos órdenes de magnitud por encima de lo esperado en CMS durante dos años de toma de datos, para permitir estudios precisos de la señal y de la contaminación. Para realizar el análisis, las muestras de sucesos se pesan para que equivalgan a una luminosidad acumulada de 30 fb⁻¹.

5.1. Análisis independiente de m_H

Como se ha presentado en la sección previa, la señal y las distintas contaminaciones poseen unas signaturas características y diferenciadoras. Este hecho puede ser empleado para, por medio de cortes independientes de la masa del bosón de Higgs, discriminar la señal del fondo [5]. Se utiliza este análisis como referencia para poder cuantificar la mejora obtenida al emplear un análisis multivariado. Además de los cortes de preselección, se aplican los siguientes cortes de selección:

- La masa invariante del par $\mu^+\mu^-$ más próxima al valor nominal de la masa del bosón Z, $m_{\mu^+\mu^-}$, debe verificar $70 \text{ GeV}/c^2 < m_{\mu^+\mu^-} < 100 \text{ GeV}/c^2$. Este valor viene dado por la resolución del detector y es asimétrico respecto a m_Z para tener en cuenta el proceso $Z \rightarrow \mu^+\mu^-\gamma$. Este corte garantiza la existencia de un bosón Z real, eliminando cerca del 50% de $t\bar{t}$, con una eficiencia del 90% para la señal.
- El momento transverso de los dos muones con menor momento transverso debe ser mayor que 12 GeV/c y 8 GeV/c, respectivamente. Esto suprime gran parte de los sucesos $Zb\bar{b}$ y $t\bar{t}$, pues tienen muones procedentes de desintegraciones de quarks b y, por tanto, de momento transverso más bajo

que los producidos en la desintegración de bosones Z. Para los otros dos muones se exige que el momento transversal sea mayor que 15 GeV/c, lo que suprime muones procedentes del halo de los haces de protones o muones ficticios reconstruidos erróneamente. Estos cortes tienen una eficiencia del 90 % para la señal y suprimen en torno al 50 % de $Zb\bar{b}$, 40 % de $t\bar{t}$ y 10 % de ZZ.

- Las variables de aislamiento se definen como la cantidad de energía transversal depositada en el calorímetro y como la suma del momento transversal de las trazas reconstruidas en el *tracker*, dentro de un cono de radio R en torno a cada muon. El radio viene dado por $R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$, siendo η la pseudorapidez y ϕ el ángulo acimutal. La señal se selecciona imponiendo cortes superiores de 5 GeV y 9 GeV en la medida del aislamiento calorimétrico con un radio de 0.24, y de 2.5 GeV/c y 4 GeV/c en el aislamiento del *tracker* para un radio de 0.20, para los dos muones menos aislados. Estos cortes rechazan muchos sucesos $Zb\bar{b}$ y $t\bar{t}$, con dos muones producidos en cascadas hadrónicas. Del mismo modo que para el momento transversal, para los dos muones más aislados se aplican cortes de calidad en 2.5 GeV y 3 GeV para el aislamiento calorimétrico y 2 GeV/c y 2.5 GeV/c para el aislamiento del *tracker*.

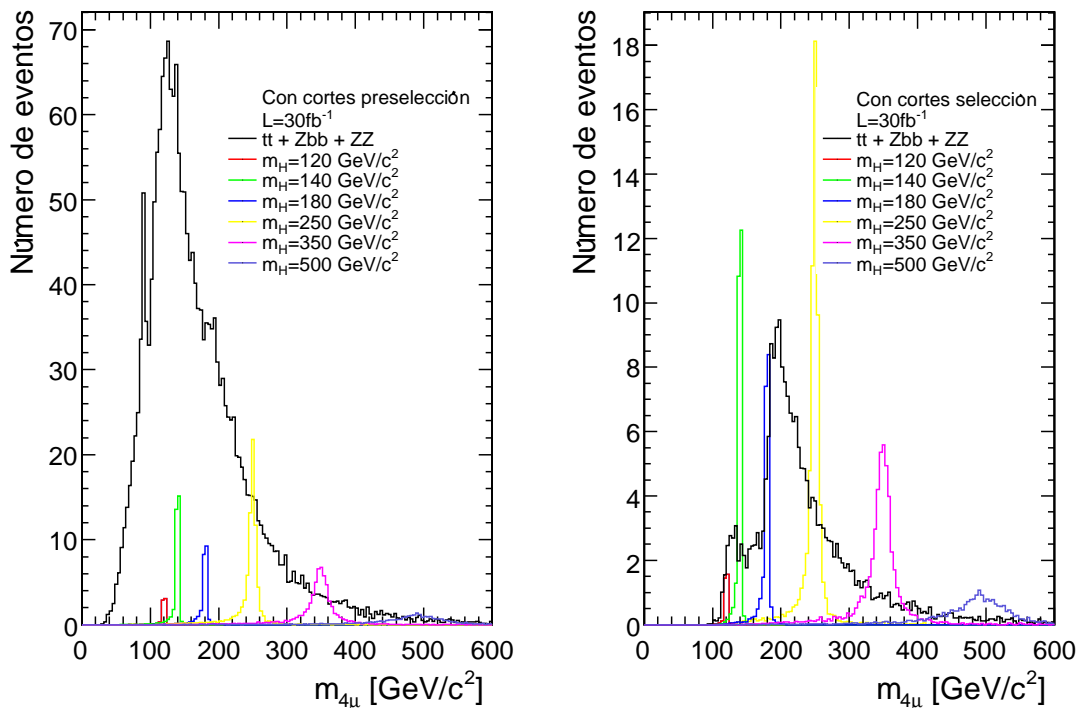


Figura 4: Distribuciones del número de sucesos en función de la masa invariante $m_{4\mu}$ correspondientes a seis hipótesis de masa del bosón de Higgs y la contaminación, aplicando únicamente cortes de preselección (izquierda) y tras los cortes de selección (derecha).

Tras los cortes anteriores, los procesos de contaminación $Zb\bar{b}$ y $t\bar{t}$ son reducidos en un 99.9 %, niveles despreciables en comparación con ZZ, suprimido en un 90 %. La figura 4 (derecha) muestra la distribución de la masa invariante $m_{4\mu}$ de los sucesos de señal para seis valores de m_H y el fondo después de aplicar los cortes de selección. En ella se puede apreciar la importante reducción de contaminación, cuyo remanente es en su práctica totalidad ZZ, así como el leve descenso del número de sucesos de la señal.

Como puede verse, un análisis basado en la selección de sucesos en cierta región del espacio multidimensional de las variables de selección permite reducir el fondo a niveles muy bajos que hacen visible la señal. Sin embargo, este tipo de métodos desprecian una parte importante de la información contenida en

las distintas variables simuladas. Haciendo uso de métodos estadísticos multivariados es posible mejorar la sensibilidad del análisis. En las siguientes secciones, se introduce un método estadístico conocido como *likelihood* que nos permite incorporar al análisis la información contenida en la forma de las distintas variables medidas y la correlación entre ellas, con la consiguiente mejora en la significación estadística de la señal y la reducción de la luminosidad integrada requerida para observarla.

5.2. Análisis empleando métodos multivariados

El método del likelihood se basa en la definición de un nuevo discriminante, y , a partir de las n variables discriminantes $(x_1 \dots x_n)$. Para cada suceso i , el likelihood y viene dado por:

$$y^i(x_1 \dots x_n) = \frac{\mathcal{P}_S^i}{\mathcal{P}_S^i + \mathcal{P}_B^i},$$

donde

$$\mathcal{P}_a^i = \prod_{k=1}^n p_a^k(x_k^i),$$

siendo $a = S, B$, y x_k^i el valor observado de la variable x_k para el i -ésimo suceso. La función $p_a^k(x_k)$ es la densidad de probabilidad de la variable x_k esperada para sucesos de señal ($a = S$) y contaminación ($a = B$). Estas densidades de probabilidad están normalizadas:

$$\int_{-\infty}^{\infty} p_a^k(x_k) dx_k = 1.$$

En general no es posible conocer la forma exacta (esperada) de $p_a^k(x_k)$, por lo que se emplea la parametrización mediante funciones analíticas de las distribuciones obtenidas de la simulación Monte Carlo.

El discriminante y proporciona un valor comprendido entre 0 y 1. Si las variables x_k tienen un cierto poder de discriminación entre señal y contaminación, y toma valores próximos a uno para sucesos similares a la señal esperada y a cero en caso contrario. El mayor grado de discriminación viene reflejado por una significativa diferencia en la forma de las densidades de probabilidad p_a^k para señal y contaminación, que sólo es posible si las variables utilizadas en el análisis son capaces de reflejar la diferente signatura en el detector de los diferentes procesos físicos. Esta transparencia en la interpretación de los resultados, unido a la rapidez del cómputo de y para un número elevado de sucesos, hace este método preferible a otros más complejos y tediosos como las redes neuronales.

Gracias a este método se consigue incorporar al análisis la información contenida en varias variables basándonos en la forma de sus distribuciones, en contraste con el análisis de referencia de la sección anterior. Estas variables se eligen intentando maximizar la diferencia entre las distribuciones de señal y contaminación, al tiempo que se busca minimizar la correlación entre ellas, con el fin de evitar que el resultado del análisis dependa de la exacta simulación de estas correlaciones, en algunos casos difícil de evaluar.

Las funciones de referencia dependen de la señal estudiada por lo que son calculadas independientemente para cada hipótesis de masa del bosón de Higgs, lo que permite optimizar el análisis para cada valor concreto de m_H . La precisión con que se obtienen estas densidades de probabilidad depende la cantidad de sucesos de señal y contaminación simulados, hecho que cobra especial importancia en las colas de las distribuciones.

Una vez haya datos reales disponibles (a partir de Julio de 2008) será crucial ajustar las simulaciones a la respuesta real del detector antes de llevar a cabo análisis de alto nivel como el que aquí se presenta. Este ajuste fino de la simulación se realizará estudiando en detalle procesos estándar ($J/\psi \rightarrow \mu^+\mu^-$, $Z \rightarrow \ell\ell$, $W \rightarrow \ell\nu$, sucesos con jets de alto momento transversal, etc.) que se producirán en gran cantidad desde el principio de la operación de LHC.

5.3. Análisis del canal $H \rightarrow ZZ^{(*)} \rightarrow 4\mu$ con el método de likelihood

El ingrediente fundamental de este análisis es la correcta elección de las variables que se utilizan en la definición del likelihood. Para la señal hay una clara diferencia cinemática entre la región $m_H < 2m_Z$ y $m_H > 2m_Z$, debida al carácter virtual o no del bosón Z con masa más lejana al valor nominal, $m_Z = 91.18 \text{ GeV}/c^2$. Por ello, es conveniente dividir el análisis en dos regiones de masa, separadas por $m_H = 2m_Z$.

Para el caso de masa baja ($m_H < 2m_Z$), una buena variable discriminante para el likelihood es m_{Z_2} , definida como la masa invariante de la pareja $\mu^+\mu^-$ que tiene el valor más lejano a $m_Z = 91.18 \text{ GeV}/c^2$. En el caso del fondo dominante $ZZ^{(*)}$, en esta región en que $m_{4\mu} < 2m_Z$, uno de los dos bosones Z del suceso es real, mientras que el segundo bosón (Z^* o γ^*) está cinemáticamente forzado a ser virtual. La presencia del fotón virtual hace que la distribución de m_{Z_2} sea diferente de la esperada para la señal. Las otras variables empleadas son p_{T_3} y p_{T_4} , donde los índices 3 y 4 hacen referencia a los dos muones de menor momento transversal. Se espera que estas distribuciones sean significativamente distintas para la señal y el fondo, como se vió en la sección anterior.

La figura 5 muestra las distribuciones de las tres variables utilizadas en el análisis para la señal de $m_H = 140 \text{ GeV}/c^2$ (izquierda) y el fondo (derecha). En el cálculo de las funciones de referencia sólo se tienen en cuenta aquellos sucesos de señal y contaminación con una masa invariante de los cuatro muones comprendida en el rango $120 \text{ GeV}/c^2 < m_{4\mu} < 150 \text{ GeV}/c^2$. Esto es debido a que lo importante del método es caracterizar los sucesos de contaminación que están en la región de la señal. Fuera de este rango no hay sucesos de señal y la información que proporcionan las funciones de referencia es superflua. La ventana en torno a $m_{4\mu}$ es diferente para cada hipótesis de m_H y tiene en cuenta el valor de m_H , la anchura natural del bosón de Higgs y la resolución en $m_{4\mu}$.

Si bien las diferencias entre las funciones de referencia individuales no parecen excesivas, su combinación en una distribución de likelihood proporciona una buena separación entre la señal y el fondo. En la figura 6 se presenta el resultado de evaluar y a partir de m_{Z_2} , p_{T_3} y p_{T_4} , para los sucesos simulados de señal y fondo. Para valores próximos a cero los sucesos dominantes proceden del fondo, con una pequeña contribución de sucesos de señal. El número de sucesos de contaminación decrece rápidamente al aumentar el valor del likelihood, mientras que aumenta significativamente el de la señal, tendencia que se hace más acusada para valores próximos a uno. El efecto de aplicar un corte en el likelihood, seleccionando la región de mayor pureza (proporción señal/contaminación) de la muestra, es aumentar la significación estadística de la señal y mejorar el potencial de descubrimiento del bosón de Higgs. En la siguiente sección se da una descripción cuantitativa de esta mejora para cada valor de m_H .

Para $m_H > 2m_Z$, el poder de discriminación del análisis anterior disminuye ya que las distribuciones de m_{Z_2} son muy similares para la señal y el fondo dominante $ZZ^{(*)}$. Esto es debido a que en ambos casos los dos bosones Z son reales. En esta región de m_H se introduce una nueva variable, $p_{T^{4\mu}}$, que corresponde al momento transversal del sistema de los cuatro muones. Para la señal, esta variable denota el momento transversal del bosón de Higgs. Las diferencias en la distribución de $p_{T^{4\mu}}$ para la señal y la contaminación de $ZZ^{(*)}$ tienen su origen en los diferentes mecanismos de producción dominantes, $gg \rightarrow H$ para la señal y $q\bar{q} \rightarrow ZZ^{(*)}$ para el fondo. Por otro lado, las variables p_{T_3} y p_{T_4} siguen siendo discriminantes en este rango de masa. La figura 7 muestra las funciones de referencia de la señal de $m_H = 250 \text{ GeV}/c^2$ (izquierda) y fondo (derecha) empleando sucesos con $100 \text{ GeV}/c^2 < m_{4\mu} < 320 \text{ GeV}/c^2$.

La distribución del likelihood de la figura 8 pone de manifiesto la buena separación entre la señal de $m_H = 250 \text{ GeV}/c^2$ y el fondo, si bien el número de sucesos de contaminación es muy superior al de señal. Al igual que en el caso de $m_H = 140 \text{ GeV}/c^2$, el aplicar un corte en el valor del likelihood aumenta la pureza de la muestra y el potencial de descubrimiento del bosón de Higgs.

Hay que resaltar que en el proceso de selección de variables para las dos regiones de masa se han considerado otras posibilidades. Para $m_H < 2m_Z$, se ha estudiado la variable $p_{T_3} + p_{T_4}$, pues en principio las funciones de referencia eran muy distintas para señal y fondo. La conclusión de este estudio es que $p_{T_3} + p_{T_4}$ no aporta información adicional a la de p_{T_3} y p_{T_4} por separado. Para masa alta, $m_H > 2m_Z$,

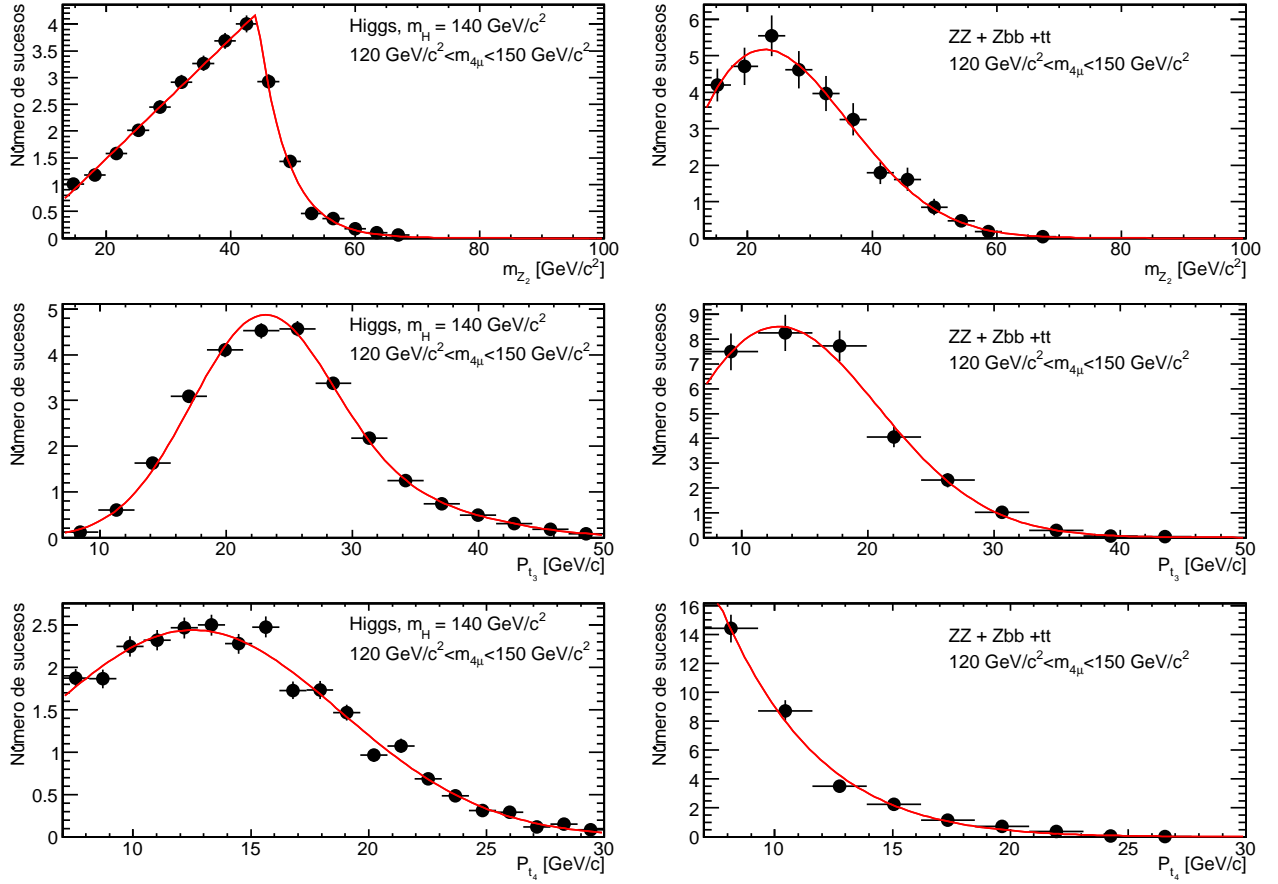


Figura 5: Funciones de referencia usadas como densidades de probabilidad en la definición del likelihood para la hipótesis de masa $m_H = 140 \text{ GeV}/c^2$, tanto para la señal (izquierda) como para la contaminación (derecha), restringidas a la región $120 \text{ GeV}/c^2 < m_{4\mu} < 150 \text{ GeV}/c^2$.

se ha intentado explotar el ángulo formado por los dos bosones Z respecto a la dirección de los haces de protones en el sistema en reposo del bosón de Higgs, θ . Esta variable es sensible al espín del sistema ZZ que, en el caso del bosón de Higgs, se espera que tenga una distribución uniforme. Al realizar las posibles combinaciones con el resto de variables, se vio que la información que aportaba no era relevante.

En esta sección hemos mostrado las funciones de densidad de probabilidad (funciones de referencia) y la distribución de likelihood para dos valores concretos de m_H , $140 \text{ GeV}/c^2$ y $250 \text{ GeV}/c^2$, a modo de ejemplo. El análisis completo se ha realizado para 17 valores de la masa del bosón de Higgs, lo que ha supuesto parametrizar con funciones analíticas las densidades de probabilidad correspondientes, tanto para señal como contaminación. Los resultados del análisis se presentan en la siguiente sección para estos 17 valores de m_H .

5.4. Potencial de descubrimiento del bosón de Higgs

El potencial de descubrimiento se cuantifica en términos de la significación estadística, S_L , de la señal esperada del bosón de Higgs respecto al nivel de contaminación. El método estadístico elegido para evaluar S_L es el llamado método del *log-likelihood ratio* [8, 13], que utiliza como estimador estadístico

$$Q(x) \equiv \frac{\mathcal{L}_{S+B}(x)}{\mathcal{L}_B(x)},$$

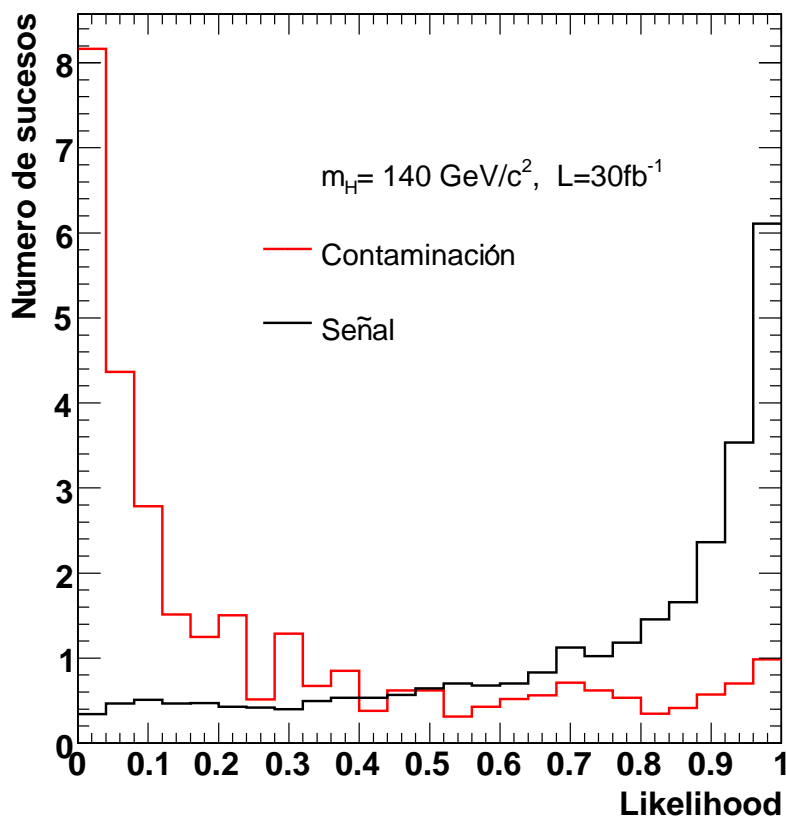


Figura 6: Número de sucesos esperado en función del valor del *likelihood* para el fondo, en rojo, y la señal, en negro, en el análisis correspondiente a $m_H = 140 \text{ GeV}/c^2$, con las variables: m_Z , p_{T3} y p_{T4} .

siendo $\mathcal{L}_{S+B}(x)$ la probabilidad de que la distribución de la variable discriminante x observada en los datos sea compatible con la distribución esperada de x para la hipótesis de señal más fondo, mientras que $\mathcal{L}_B(x)$ es la probabilidad de que la distribución de x sea compatible sólo con la hipótesis de contaminación. Para el caso de funciones discretas Q toma la forma:

$$Q = e^{-s_{tot}} \prod_{i=1}^N \left(1 + \frac{s_i}{b_i}\right)^{n_i},$$

donde el producto se extiende a los N posibles valores discretos, x_i , de la variable x . Por otro lado, s_i y b_i son el número de sucesos de señal y fondo esperados, mientras que n_i es el número de sucesos observados, todo ello para el valor concreto de x_i . En este formalismo, la significación estadística de la señal viene dado por

$$S_L = \sqrt{\langle 2 \ln Q \rangle}$$

donde

$$\ln Q = - \sum_{i=1}^N \left[s_i - n_i \ln \left(1 + \frac{s_i}{b_i}\right) \right].$$

La significación estadística de la señal es más alta en las regiones de x en que el cociente del número de sucesos de señal y contaminación, s_i/b_i , es mayor. Un incremento global del número de sucesos se traduce en un aumento de la significación estadística.

El potencial de descubrimiento del bosón de Higgs se evalúa a partir de los sucesos que pasan los cortes de selección descritos en el análisis de referencia 5.1, excepto para las variables p_{T3} y p_{T4} sobre las

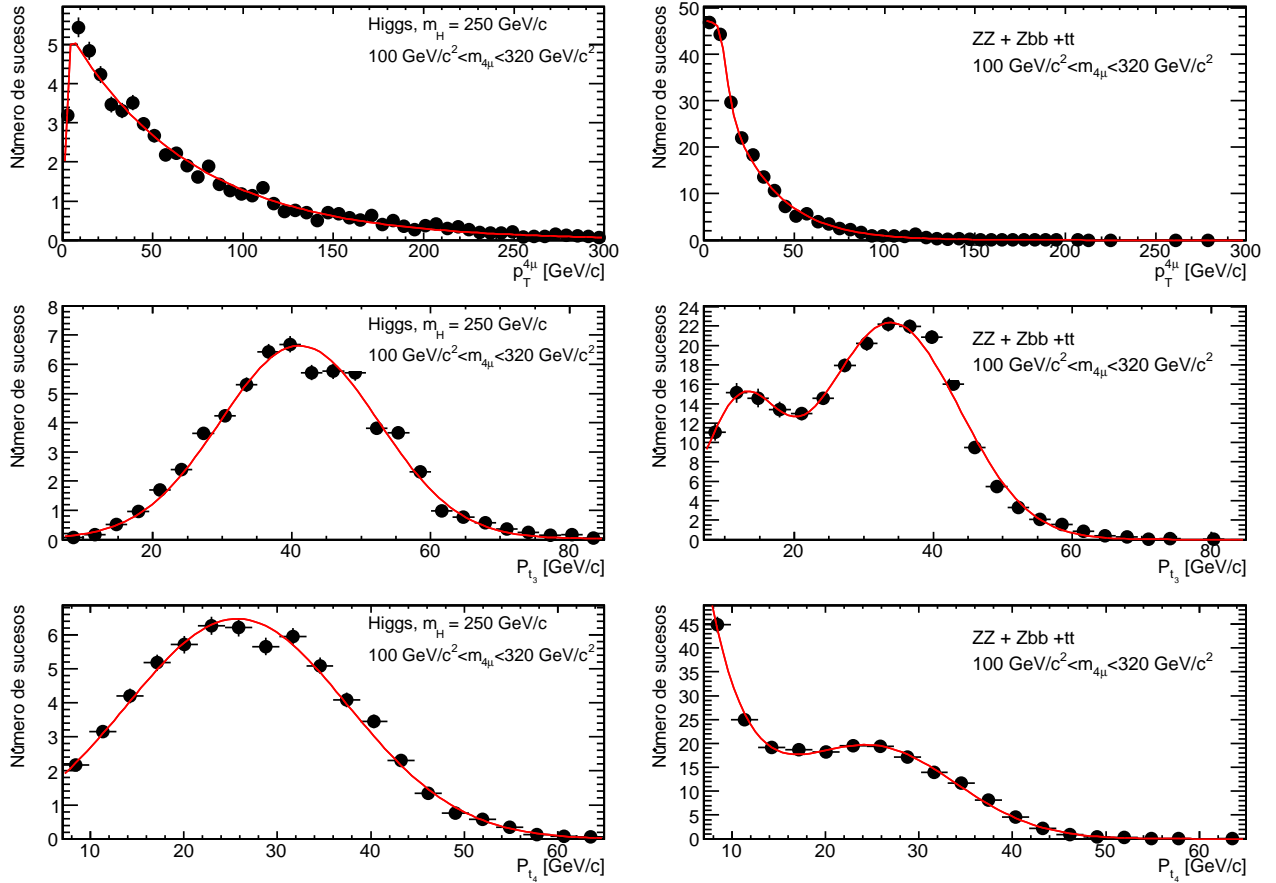


Figura 7: Funciones de referencia usadas como densidades de probabilidad en la definición del likelihood para la hipótesis de masa $m_H = 250 \text{ GeV}/c^2$, tanto para la señal (izquierda) como para la contaminación (derecha), restringidas a la región $100 \text{ GeV}/c^2 < m_{4\mu} < 320 \text{ GeV}/c^2$.

que se aplica únicamente los cortes de preselección ($p_T > 7 \text{ GeV}/c$), ya que éstas pasan a formar parte del likelihood descrito en la sección anterior. Se encuentra que el corte $y > 0.4$ maximiza la significación estadística para el rango de masas $m_H > 140 \text{ GeV}/c^2$. Un corte en y dependiente de m_H sólo aumenta la significación estadística a costa de complicar innecesariamente el análisis.

En este análisis, la significación estadística, S_L , se evalúa a partir de la distribución de $m_{4\mu}$ de aquellos sucesos que han satisfecho todos los criterios de selección. En la figura 9 se muestra el valor de S_L para todo el rango de m_H estudiado, tanto para el análisis de referencia como para el análisis multivariado. Estos valores están calculados para una luminosidad de 30 fb^{-1} .

En las regiones $m_H < 140 \text{ GeV}/c^2$ y $m_H \simeq 160 \text{ GeV}/c^2 \simeq 2m_W$, la fracción de desintegración del canal $H \rightarrow ZZ^{(*)}$ (figura 3) es muy baja lo que se traduce en una significancia estadística reducida, $S_L \sim 2 - 4$. En el resto del rango de m_H , la significación estadística es superior a cinco, salvo para valores de $m_H > 450 \text{ GeV}/c^2$ para los que se produce un descenso de S_L debido a una disminución de la sección eficaz de producción del bosón de Higgs.

En la tabla 1 se muestran los valores numéricos de la figura 9 junto con la mejora porcentual del análisis multivariado respecto al análisis de referencia. Esta mejora está entre el 12% y el 22% para masas superiores a $250 \text{ GeV}/c^2$. En el caso de $m_H < 2m_Z$, la mejora en términos de porcentaje también es importante, entre 8% y 17% para masas entre $160 \text{ GeV}/c^2$ y $180 \text{ GeV}/c^2$. Para valores de m_H ligeramente superiores a $2m_Z$, apenas se aprecia alguna mejora debido a que el poder de discriminación de $p_T^{4\mu}$ no es efectivo todavía. Para masas muy bajas, $m_H < 140 \text{ GeV}/c^2$, el análisis multivariado no introduce ninguna mejora debido a que el corte genérico $y > 0.4$ no es óptimo.

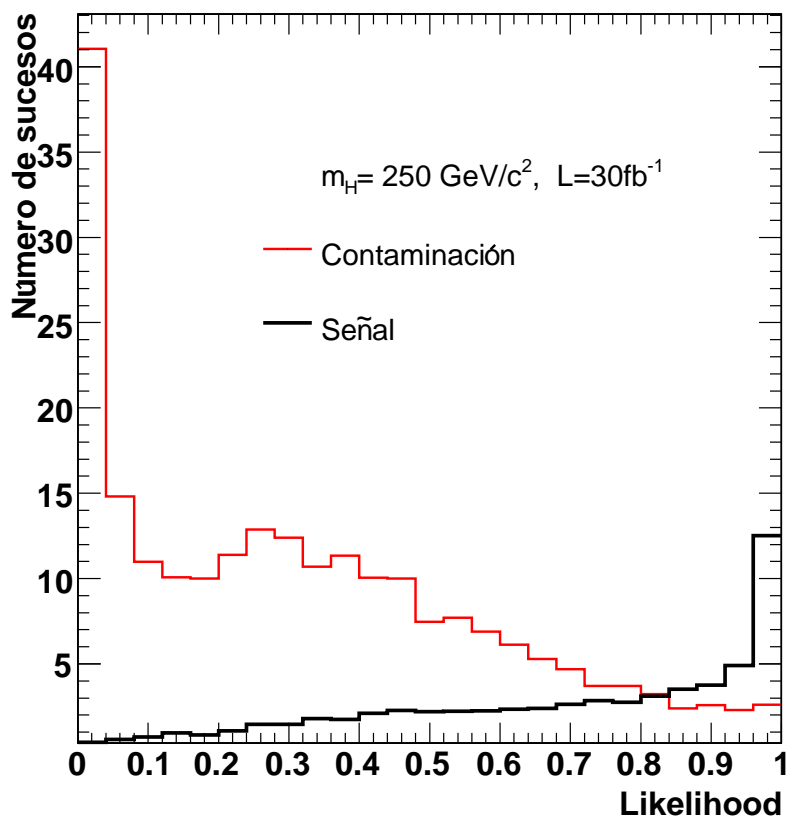


Figura 8: Número de sucesos esperado en función del valor del *likelihood* para el fondo, en rojo, y la señal, en negro, en el análisis correspondiente a $m_H = 250 \text{ GeV}/c^2$, con las variables: $p_T^{4\mu}$, p_{T3} y p_{T4} .

m_H [GeV/c^2]	120	130	140	150	160	170	180	190	200
S_L (ref.)	1.94	5.10	8.40	9.89	5.44	2.71	4.31	10.38	10.49
S_L ($y > 0.4$)	1.82	4.82	8.42	10.44	5.87	3.19	4.70	10.56	10.65
ΔS_L [%]	-6.33	-5.53	0.14	5.51	7.84	17.46	9.06	1.68	1.45

m_H [GeV/c^2]	250	300	350	400	450	500	550	600
S_L (ref.)	8.88	8.16	8.41	8.43	6.94	5.66	4.15	3.40
S_L ($y > 0.4$)	10.01	9.47	10.06	9.91	8.47	6.43	4.97	3.98
ΔS_L [%]	12.66	14.01	19.65	17.47	21.98	13.59	19.85	16.65

Tabla 1: Mejora de la significación estadística, S_L , al aplicar el corte genérico en el likelihood, $y > 0.4$, respecto a los resultados del análisis de referencia.

En el momento que existan datos reales, se considerará que se está ante un descubrimiento cuando la probabilidad de que los datos sean compatibles con una fluctuación del fondo esté por debajo de 2.85×10^{-7} , que corresponde a $S_L = 5$. La figura 10 muestra la luminosidad acumulada requerida para obtener los valores $S_L = 5$ y $S_L = 3$, tanto en el análisis de referencia como en el multivariado. Este último requiere un 15% menos de luminosidad acumulada que el análisis de referencia para alcanzar el nivel de descubrimiento del bosón de Higgs, en un amplio rango de m_H .

Gracias a esta mejora, se podrá descubrir el bosón de Higgs de masa en el rango $200 \text{ GeV}/c^2 < m_H <$

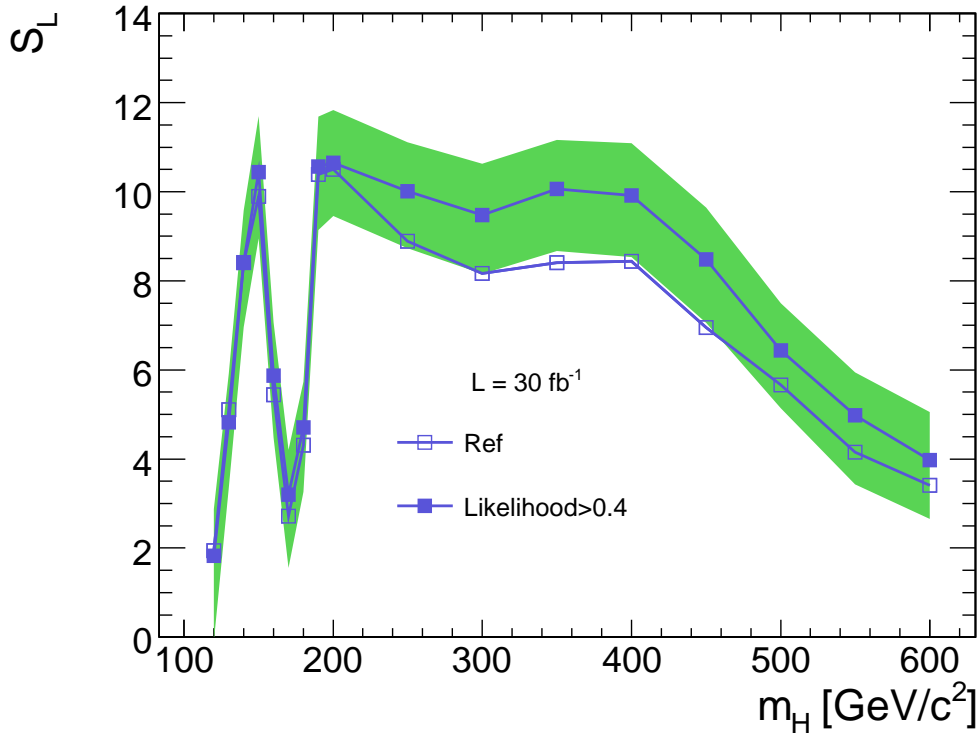


Figura 9: Significación estadística en función de m_H para el análisis de referencia y el análisis basado en el likelihood, después de aplicar el corte $y > 0.4$.

450 GeV/c² con una luminosidad acumulada inferior a 10 fb⁻¹, que es la luminosidad que espera ser acumulada en un año de toma de datos en CMS. Además permitirá observar la evidencia de una señal, $S_L = 3$, con tan solo 4 fb⁻¹ en el mismo rango de masas. Para valores de m_H en torno a 120 GeV/c² y $2m_W$, la luminosidad requerida para el descubrimiento del bosón de Higgs en el canal $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$ es excesivamente alta, por lo que para esas masas son de particular interés los análisis de los canales $H \rightarrow \gamma\gamma$ y $H \rightarrow WW^{(*)}$, respectivamente.

6. Conclusiones

Se ha presentado un análisis para la búsqueda del bosón de Higgs en CMS en el canal $H \rightarrow ZZ^{(*)} \rightarrow 4\mu$ empleando un método multivariado. El análisis se basa en construir un likelihood dependiente de la masa del bosón de Higgs a partir de variables discriminantes. Este análisis supone una mejora significativa en el potencial de descubrimiento del bosón de Higgs en un amplio rango de masas respecto al análisis publicado por CMS, basado en simples cortes ortogonales en las variables discriminantes independientes de la masa del bosón de Higgs.

Agradecimientos

Agradezco a mis supervisores su continua dedicación y apoyo, así como a la División de Física de Partículas del CIEMAT por haberme posibilitado realizar este trabajo.

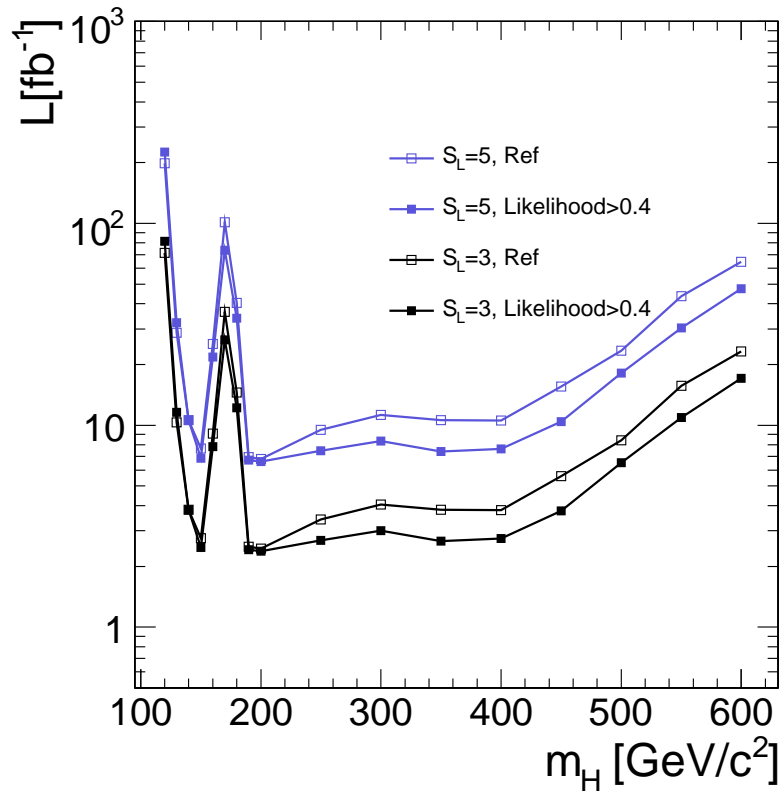


Figura 10: Evolución de la luminosidad necesaria para alcanzar $S_L = 5$ y $S_L = 3$ en función de la masa del Higgs para el análisis de referencia y para el análisis empleando el método likelihood con la condición $y > 0.4$

Referencias

- [1] S.L. Glashow, Nucl. Phys. 22 (1961) 579; S. Weinberg, Phys. Rev. Lett. 19 (1967) 1264; A. Salam, *Elementary Particle Theory*, edited by N. Svartholm (Almqvist and Wiksell, Stockholm, 1968), p. 367.
- [2] P. W. Higgs, Phys. Lett. 12 (1964) 132; P. W. Higgs, Phys. Rev. Lett. 13 (1964) 508; P. W. Higgs, Phys. Rev. 145 (1966) 1156; F. Englert and R. Brout, Phys. Rev. Lett. 13 (1964) 312; G. S. Guralnik, C. R. Hagen and T. W. B. Kibble, Phys. Rev. Lett. 13 (1964) 585.
- [3] The LHC Study Group, *The Large Hadron Collider Conceptual Design*, CERN/AC 95-05 (1995), <http://cern.ch/ab-div/Publications/LHC-DesignReport.html>
- [4] The CMS Collaboration, *Technical Proposal*, CERN/LHCC 94-38, LHCC/P1 (1994).
- [5] CMS TDR 8.2, “CMS Physics Technical Design Report, Volume II: Physics Performance”, CERN/LHCC 2006-021, J. Phys. G: Nucl. Part. Phys. 34 (2007) 995-1579.
- [6] M. Aldaya et al., *Discovery potential and search strategy for the Standard Model Higgs in the $H \rightarrow ZZ^{(*)} \rightarrow 4\mu$ decay channel using a mass-independent analysis*, CMS NOTE 2006/106.
- [7] T. Hambye and K. Riesselmann, *Matching conditions and Higgs mass upper bounds revisited*, Phys. Rev. D 55 (1997) 7255.
- [8] The LEP Collaborations ALEPH, DELPHI, L3 and OPAL, *Search for the Standard Model Higgs Boson at LEP*, Phys. Lett. B565(2003) 61.

- [9] The ALEPH, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, the SLD Electroweak and Heavy Flavour Groups, *Precision Electroweak Measurements on the Z Resonance*, Phys. Rept. 427 (2006) 257.
- [10] T. Sjostrand, L. Lonnblad and S. Mrenna, *PYTHIA 6.2 Physics and Manual*, arXiv: hep-ph/0108264.
- [11] CompHEP collaboration, *CompHEP: A package for evaluation of Feynman diagrams and integration over multi-particle phase space. User's manual for version 33*, arXiv:hep-ph/9908288.
- [12] S. Agostinelli *et al.*, *GEANT4: A simulation toolkit*, NIM A 506 (2003) 250-303; J. Allison *et al.*, IEEE Trans. Nucl. Sci. 53, February 2006 (in press).
- [13] A.L.Read, *Modified Frequentist Analysis of Search Results (The CLs Method)*, CERN-EP/2000-005.

