

Mayo de 2019

Facultad de Ciencias, Universidad de Oviedo.

TRABAJO DE FIN DE GRADO DE FÍSICA

PRODUCCIÓN ELECTRODÉBIL DE WZ (VBS) EN EL
EXPERIMENTO CMS (LHC)

Tutores

Carlos Francisco Erice Cid

Isidro González Caballero

Autor

Mario González Carpintero

uo250981@uniovi.es

Índice

Introducción	1
1. Marco teórico: El Modelo Estándar	3
1.1. Las tres interacciones fundamentales	3
1.2. Quarks, leptones y bosones en el Modelo Estándar	4
1.3. Producción de WZ, WZ VBS y WZ QCD	7
1.4. Diagramas de Feynman a LO, NLO y NNLO	11
2. El detector CMS en el LHC	14
2.1. El sistema de coordenadas de CMS	15
2.2. La estructura de capas de CMS	17
2.3. Identificación y reconstrucción de partículas	20
2.4. El momento en el eje z. El problema de los neutrinos.	22
2.5. <i>Pile Up</i> en el detector CMS	24
2.6. El trigger de CMS	26
2.7. Introducción al análisis computacional	27
2.7.1. Herramientas informáticas	28
2.7.2. Datos y Monte Carlo	29
3. La sección eficaz en física de partículas	31
3.1. El concepto de sección eficaz	31
3.2. El concepto de la <i>fuerza de la señal</i>	33
4. Análisis físico de los procesos de producción WZ y WZ VBS	37
4.1. El concepto de masa invariante	37
4.2. Rasgos físicos de la producción WZ VBS y WZ QCD	38
4.3. Procesos de fondo. La estructura de los histogramas	41
4.4. Regiones de Control	44
4.4.1. La CR del proceso de producción de dos bosones Z	47
4.4.2. La CR de procesos con quarks top	49
4.5. La región de señal para la producción WZ	51

4.5.1.	La normalización del Monte Carlo de señal	51
4.5.2.	Selección de sucesos	52
4.5.3.	Resultados en la región de señal WZ	60
4.6.	La región de control para la producción WZ	62
4.7.	La región de señal para la producción WZ VBS	64
4.7.1.	Selección de sucesos	64
4.7.2.	Resultados en la región de señal VBS	78
5.	Medida de la fuerza de la señal para la producción WZ y WZ VBS	80
5.1.	Fuentes de incertidumbre	80
5.1.1.	Incertidumbres estadísticas	80
5.1.2.	Incertidumbres sistemáticas	80
5.2.	Propagación de las incertidumbres a la región de señal	83
5.3.	Propagación de incertidumbres a la fuerza de la señal	85
5.4.	Fuerza de la señal para la producción WZ	87
5.5.	Fuerza de la señal para la producción WZ WBS	90
6.	Conclusiones	94

Introducción

El trabajo que aquí se presenta está centrado en el estudio de de producción electrodébil de WZ VBS, utilizando los datos recogidos en el detector CMS del LHC durante el año 2016 correspondientes con una luminosidad de 35.9 fb^{-1} y una energía en el centro de masas de 13 TeV.

Este trabajo se enmarca dentro del área de conocimiento del grupo de investigación en Física Experimental de Altas Energías de la Universidad de Oviedo. Los contenidos de este trabajo están muy ligados con los de la asignatura *Física Nuclear y de Partículas Elementales*, así como con *Física de Altas Energías y Aceleradores*. Además, ha sido de utilidad el conocimiento adquirido en *Métodos Numéricos Aplicados a la Física*.

El presente documento comienza con una introducción al Modelo Estándar (SM, *Standard Model*), en la que se describen las características básicas tanto de las interacciones como de las partículas enmarcadas en esta teoría, tras contextualizarse históricamente.

Dentro de esta primera sección, se describe también el proceso que da nombre a este trabajo así como su principal fondo. Se motiva su estudio destacando dos aspectos de la física de partículas, tanto dentro como fuera del Modelo Estándar, a cuyo entendimiento contribuye la caracterización de este proceso. Se describen también algunas herramientas teóricas adicionales que serán de utilidad a lo largo del resto del trabajo.

La segunda sección está dedicada a la descripción del detector CMS, con el que se han tomado los datos. Además de su estructura, se describirá el modo en el que cada partícula del SM es detectada en cada uno de los subdetectores que lo componen, así como la necesidad de filtrar el flujo de datos que entra al detector antes de almacenar la información externamente. Se describen también los recursos utilizados en el análisis tanto a nivel de *software* como de *hardware*.

En la tercera sección, se describe brevemente la *sección eficaz*, y a partir de ella la *fuerza de la señal*, un observable físico que se medirá en este trabajo.

Seguidamente, en la cuarta sección, se estudia desde el punto de vista físico el proceso de producción tanto de WZ como de WZ VBS. Se caracterizan por separado los dos mayores fondos de los procesos de interés, y se aplican criterios de selección físicamente argumentados con el fin aislar en la medida de lo posible las producciones WZ y WZ VBS de todos sus principales fondos.

Como último punto del trabajo, en la quinta sección, se describen todas las fuentes de incertidumbre relevantes de la *fuera de la seña* así como su propagación a la medida. Esta propagación se calcula junto al valor central para las producciones WZ y WZ QCD.

1. Marco teórico: El Modelo Estándar

Aunque la existencia de los átomos ya fue postulada por Demócrito en la era de los presocráticos, no fue hasta la primera década del siglo XX cuando quedó demostrado que estos objetos son esencialmente huecos.

Más tarde, ya en la segunda mitad del siglo XX, comenzaron a construirse los primeros aceleradores de partículas, con los que fueron descubiertas multitud de partículas nuevas. Estas partículas parecían *demasiadas* como para ser *todas* fundamentales, aunque no tardó demasiado en demostrarse la existencia de quarks. Estos simplificarían enormemente el modelo dando información sobre la estructura de las nuevas partículas a su nivel más fundamental.

La física de partículas del siglo XX comenzó también a entender las interacciones fundamentales como interacciones a distancia en las que una partícula se intercambia entre aquellas que interaccionan. Estas partículas transmisoras son conocidas como *bosones gauge*.

1.1. Las tres interacciones fundamentales

Protones y neutrones forman núcleos, en torno a los cuales orbitan electrones para dar lugar a átomos. La interacción que liga a los electrones con el núcleo es el electromagnetismo, descrito de forma teórica por la electrodinámica cuántica (QED, *Quantum Electrodynamics*). La interacción que mantiene unidos los componentes del interior del núcleo es la fuerte, y el formalismo teórico que lo describe la cromodinámica cuántica (QCD, *Quantum Chromodynamics*).

Tanto la interacción *fuerte* como la *débil* juegan un papel fundamental en la formación de estructuras y en la desintegración de partículas a niveles subatómicos, respectivamente. El estudio de estas dos interacciones a escalas inferiores al fermi ($1 \text{ fm} = 10^{-15} \text{ m}$), ha culminado en la elaboración del *Modelo Estándar* (SM), que da una descripción teórica completa tanto de los bosones gauge como de las tres interacción fundamentales¹. La consistencia del SM ya ha sido comprobada experimentalmente en el LHC y en otros experimentos de física de partículas. El fin último de muchos de estos experimentos es el de poner a prueba este modelo buscando evidencias de física más allá de él (BSM, *beyond the Standard Model*).

El mundo en el que vivimos está esencialmente gobernado, a escala macroscópica, por la gravedad y el electromagnetismo. La interacción débil aparece cuando se estudian desintegraciones

¹El SM no describe la interacción gravitatoria. La gravedad es, a escalas subatómicas, completamente imperceptible en comparación con el electromagnetismo, la interacción fuerte o la débil.

radiactivas, y la fuerza fuerte se manifiesta al estudiar la interacción entre quarks que da lugar a estructuras hadrónicas.

De hecho, para que dos partículas interactúen según una fuerza, tienen que tener la carga asociada a la interacción. Así, dos objetos sin carga eléctrica no interactuarán electromagnéticamente, y por tanto no formarán estructuras gobernadas por esta interacción. Dos objetos sin color, como los electrones, tampoco podrán interactuar de forma fuerte, y no formarán por tanto parte de los núcleos en átomos. Dos electrones no quedarán ligados por mucho que se les acerque.

Por último, aunque tanto la interacción fuerte como la electromagnética forman estructuras (hadrones y átomos), no ocurre lo mismo con la interacción débil, cuya intensidad no es suficiente (a las escalas a las que actúa), como para formar un sistema ligado.

1.2. Quarks, leptones y bosones en el Modelo Estándar

Las partículas elementales que componen la materia pueden dividirse en dos grandes grupos: quarks y leptones. Dentro de estos dos grupos, distinguimos también tres generaciones, caracterizadas por la masa y la vida media de las partículas que las componen. Tanto los quarks como los leptones tienen espín $1/2$, y son por tanto fermiones.

	1ª gen	2ª gen	3ª gen
QUARKS	u <i>up</i> $m \sim 2.2 \text{ MeV}$ $q = 2/3$ $\tau = \infty$	c <i>charm</i> $m = 1.275 \text{ GeV}$ $q = 2/3$ $\tau = 1.1 \cdot 10^{-12} \text{ s}$	t <i>top</i> $m = 173.0 \text{ GeV}$ $q = 2/3$ $\tau = 4.2 \cdot 10^{-25} \text{ s}$
	d <i>down</i> $m \sim 4.7 \text{ MeV}$ $q = -1/3$ $\tau = \infty$	s <i>strange</i> $m = 95 \text{ MeV}$ $q = -1/3$ $\tau = 1.24 \cdot 10^{-8} \text{ s}$	b <i>bottom</i> $m = 4.18 \text{ GeV}$ $q = -1/3$ $\tau = 1.3 \cdot 10^{-12} \text{ s}$
LEPTONES	e <i>electron</i> $m = 0.511 \text{ MeV}$ $ q = 1$ $\tau = \infty$	μ <i>muon</i> $m = 105.7 \text{ MeV}$ $ q = 1$ $\tau = 2.2 \cdot 10^{-6} \text{ s}$	τ <i>tau</i> $m = 1777 \text{ MeV}$ $ q = 1$ $\tau = 2.9 \cdot 10^{-13} \text{ s}$
	ν_e <i>Neutrino del electrón</i> $m < 2 \text{ eV}$ $q = 0$	ν_μ <i>Neutrino del muón</i> $m < 0.19 \text{ MeV}$ $q = 0$	ν_τ <i>Neutrino del tau</i> $m < 18.2 \text{ MeV}$ $q = 0$

Figura 1: Quarks y leptones del Modelo Estándar clasificados por generaciones. Los valores de esta tabla se han tomado de la base de datos del *Particle Data Group*[2], que contiene los resultados más recientes y precisos en el campo de la física de partículas. *Figura propia.*

Las cargas de las partículas en la figura [1] vienen dadas en términos de $|e|$.

Tanto para dar las masas como para dar las cargas, se ha utilizado en la figura [1] el sistema de unidades naturales, en el que $\hbar = 1$ y $c = 1$. Esta elección permite medir masas de partículas en términos de su energía en reposo.

Sabiendo que $E^2 = \sqrt{m^2c^4 + p^2c^2}$, para una partícula en reposo se cumplirá

$$m = E/c^2$$

Si hacemos ahora $c = 1$ estaremos midiendo masas en unidades de energía. Este convenio se utilizará a lo largo de todo el trabajo.

Las partículas de la primera generación forman por sí solas toda la materia observada. A partir de los quarks u y d se forman protones (uud) y neutrones (udd), y con ellos todos los núcleos conocidos.

El electrón, por su parte, es indudablemente la partícula fundamental más estudiada, y cuyas propiedades se conocen con mayor precisión. Además de ser estable, es la fuente más elemental de campo electromagnético y de esencialmente toda la electrónica que conocemos.

Las partículas de la segunda generación son idénticas a las de la primera salvo en la masa, inversamente relacionada con su vida media. Las partículas de la tercera tienen aún más masa que las de la segunda, y por tanto menor vida media.

En concreto, el quark top tiene una masa de 173 GeV (cinco órdenes de magnitud por encima de la del electrón). Su vida media es muy pequeña, del orden de los 10^{-25} segundos, inferior al tiempo característico de actuación de la interacción fuerte. La fuerza fuerte es la responsable de la interacción entre quarks que da lugar a la formación de hadrones, por lo que la corta vida media del top provoca que este se desintegre antes de hadronizarse.

El resto de quarks pueden interaccionar fuertemente entre ellos para dar lugar a mesones (agrupaciones de dos quarks) o bariones (agrupaciones de tres quarks). Ni quarks ni gluones existen aislados en libertad, y cuando estos se producen en el estado final de algún proceso físico darán lugar a *chorros* de partículas, en un proceso al que en adelante nos referiremos como *hadronización*.

Solo los quarks y gluones tienen carga de color, y por tanto solo ellos responden a la interacción fuerte. Aunque los quarks también tienen carga eléctrica, a las distancias típicas del

interior de un núcleo, menores al fermi, la interacción fuerte es claramente dominante. Los quarks además interaccionan débilmente, por lo que frecuentemente veremos desintegrarse uno pesado en otros más ligeros.

Los leptones, por otra parte, suelen dividirse en leptones *cargados* (electrón, muón y tau) y neutrinos (leptones neutros con una masa ínfima). Ninguno de ellos tiene carga de color, y solo interaccionan electromagnética y débilmente. Aunque el electrón es estable, el muón se desintegra débilmente a electrones y neutrinos en el 100 % de los casos. El tau también se desintegra, aunque su abanico de modos de desintegración es mucho más amplio.

Los neutrinos solo interaccionan de forma débil. Aunque no todos los procesos débiles involucren neutrinos, sí serán débiles todos los procesos en los que aparezcan neutrinos.

Además de leptones y quarks, el Modelo Estándar incluye cuatro bosones gauge asociados a las tres interacciones fundamentales. Estos son el fotón (electromagnetismo), los gluones (interacción fuerte) y los bosones W y Z (interacción débil). Estos cuatro bosones tienen espín 1.

Como único bosón de espín 0 dentro del SM encontramos el Higgs, responsable de la existencia del *campo de Higgs*, que da masa a todas las partículas masivas del Modelo Estándar (incluyendo al propio bosón de Higgs).

La evidencia experimental que soporta el Modelo Estándar viene de la mano de la física de altas energías y de los colisionadores de partículas. Cuanto menor sea la distancia a la que uno quiera observar, o cuanto más pequeña sea la partícula que *se quiere romper*, más energético tendrá que ser el acelerador que se destine a ello.

El acelerador de partículas más energético construido hasta la fecha es el LHC. Actualmente, por él circulan protones de 6.5 TeV que colisionan con otros en sentido contrario a una energía en el centro de masas de 13 TeV. Por contextualizar estos valores, 1 mg de protones a 6.5 TeV tiene una energía total de 10^{14} Julios, suficiente como para llevar al punto de ebulición 10^6 m³ de agua a 0 °C.

Una descripción más exhaustiva de los rasgos generales del Modelo Estándar puede leerse en el primer capítulo de la referencia [1].

1.3. Producción de WZ, WZ VBS y WZ QCD

Nos referiremos con *producción WZ* a la producción de estos dos bosones en el estado final, que en su versión más sencilla viene descrita por el diagrama de Feynman de la figura [2].

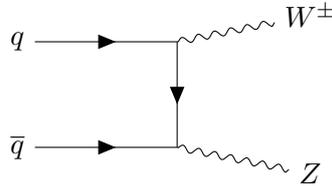


Figura 2: Diagrama de Feynman más simple para la producción WZ

En física de partículas, todos los procesos con el mismo estado final se denominan *irreducibles*. Sin embargo, no todos los diagramas irreducibles contribuyen del mismo modo al proceso que describen. La transición desde el estado inicial al final puede entenderse como una combinación lineal de todos los diagramas de Feynman que ilustran el proceso, pesados por unos coeficientes que hacen más o menos grande su contribución.

Análogamente, esta transición podría entenderse también como la descrita por *únicamente* uno de entre todos los diagramas irreducibles implicados. La probabilidad de cada diagrama de suceder estaría también cuantificada por estos mismos pesos.

Los coeficientes que pesan cada diagrama, y que por tanto dan información sobre la probabilidad de que el proceso que describen tenga lugar, son calculables y dependen de las “constantes de acoplamiento” que aparecen en cada diagrama.

Estas constantes de acoplamiento juegan un papel en cada vértice, y dependen del tipo de interacción que en él tenga lugar (*fuerte* o *electrodébil*) y de la energía a la que ocurra.

Las constantes son esencialmente dos: La constante de acoplamiento fuerte, α_s y la constante de acoplamiento electrodébil, α_{ew} . La constante fuerte decrece con la energía, mientras que α_{ew} crece ligeramente, aunque siendo siempre $\alpha_s > \alpha_{ew}$. La constante fuerte es a bajas energías *mucho* mayor que α_{ew} (~ 1 frente a $1/137$) como también lo es la fuerza fuerte respecto al electromagnetismo o a la fuerza débil.

En el diagrama de la figura [2] encontramos únicamente dos vértices, que involucran quarks y bosones débiles. Diremos entonces que este diagrama es de orden dos en α_{ew} , o simplemente de orden $\mathcal{O}(\alpha_{ew}^2)$, por tener únicamente dos vértices mediados por interacción electrodébil.

Además, este es el diagrama más sencillo que puede construirse para ilustrar la producción de dos bosones W y Z. Diremos entonces que ilustra este proceso *a primer orden*, entendiendo por *diagrama de primer orden* aquel que contribuye *a orden más bajo* al estado final de (en este caso) la producción WZ. El diagrama a primer orden para la producción de un estado final representa también el modo de producción *más probable* para ese mismo estado.

Estudiamos ahora la producción WZ VBS. Nos referimos con WZ VBS (*Vector Boson Scattering*) al proceso en el que los bosones W y Z se producen en conjunción con dos quarks, tras sufrir lo que podría entenderse como un proceso de dispersión elástica entre los dos bosones.

La producción WZ VBS es puramente electrodébil (EW, *Electroweak*). Los bosones que interaccionan son radiados por dos quarks provenientes de dos protones que colisionan en sentidos opuestos. El diagrama de Feynman más sencillo que describe el proceso es el siguiente:

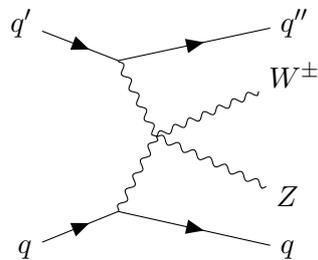


Figura 3: Diagrama de Feynman de la producción electrodébil de WZ VBS.

En el diagrama de la figura [3], encontramos en primer lugar dos vértices electrodébiles en los que intervienen dos quarks y un bosón electrodébil (W, Z o γ). En el centro del diagrama, cuatro quarks se acoplan en el mismo punto dando lugar a otros dos vértices electrodébiles. Diremos entonces que este diagrama representa una contribución de orden $\mathcal{O}(\alpha_{ew}^4)$ a su estado final.

El fondo irreducible más importante para el proceso WZ VBS es precisamente la producción mediada por QCD del mismo estado final. En este caso, ilustrado en la figura [4], un quark de un protón se aniquila con un antiquark de otro produciendo las partículas del estado final. Cada quark radía un gluón, y además se producen un W^\pm y un Z tales que la carga se conserve.

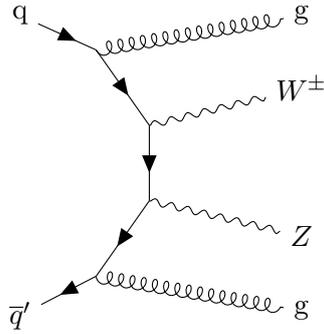


Figura 4: Diagrama de Feynman de la producción QCD de WZ

Este fondo es en realidad una variación del diagrama de producción de WZ a primer orden, en el que los dos quarks del estado inicial han radiado dos gluones.

El diagrama de la figura [2] representa el proceso a primer orden, y por tanto el más sencillo y probable, por el que dos bosones W y Z pueden producirse. El diagrama de la figura [4] tiene dos vértices fuertes más, y será por tanto considerablemente menos probable. El proceso de producción WZ, tenga o no gluones en el estado final, será aislado y medido en las secciones 4.5 y 5.4 respectivamente. Conviene entender este proceso, pues será el principal fondo irreducible de la producción WZ VBS, en la que se centra este trabajo.

Los diagramas de la figuras [3] (VBS) y [4] (QCD) tienen el mismo estado final a ojos del detector. Tanto quarks como gluones se hadronizan dando lugar a chorros de partículas indistinguibles, y del W y el Z se medirán únicamente los leptones producto de su desintegración. Ambos diagramas tienen un interés particular, pues ambos aportan sensibilidad a las secciones electrodébil (diagrama VBS) y fuerte (diagrama QCD) del Modelo Estándar.

En el caso QCD, los quarks del estado final radían dos gluones que contribuyen con dos vértices fuertes al orden total del diagrama. Los bosones W y Z se producen de modo electrodébil, y por tanto el orden total de este diagrama es $\mathcal{O}(\alpha_s^2 \alpha_{ew}^2)$.

El orden de los diagramas VBS y QCD no da información sobre el número de veces que encontraremos cada diagrama en el estado final, pero sí que son útiles para comparar la frecuencia de producción WZ VBS con la de producción WZ QCD.

Concretamente,

$$\frac{N_{QCD}}{N_{VBS}} \sim \frac{\alpha_s^2 \alpha_{ew}^2}{\alpha_{ew}^4} = \frac{\alpha_s^2}{\alpha_{ew}^2} \quad (1)$$

La constante fuerte es a 13 TeV mucho mayor que la electrodébil, y por tanto así lo será la sección eficaz de la producción de WZ QCD frente a la electrodébil. Medir y comparar las secciones eficaces para estos dos procesos es una manera directa de obtener información sobre este cociente a 13 TeV en el centro de masas.

Cabe remarcar que la relación entre las constantes de acoplamiento dada en (1) representa una aproximación a primer orden, pues solo se están incluyendo en ella los diagramas *a primer orden* tanto para la producción VBS como para la QCD. Otros diagramas con más vértices contribuirán en menor medida tanto al numerador como al denominador de este cociente, y supondrán correcciones a un orden mayor.

Volviendo al proceso en el que se centra este trabajo, cabe señalar que el “scattering de dos *bosones vectoriales*” no se produce únicamente con bosones W y Z. Diversos estudios con estados finales de $Z\gamma$, $W^\pm W^\pm$, $W^\pm\gamma$ y ZZ han sido ya también realizados por las colaboraciones ATLAS y CMS[4].

A diferencia de la búsqueda en el canal WZ , la búsqueda en WW permite obtener una mejor relación señal/fondo en el estado final, compuesto por dos leptones cuyos sabores se exigen ser independientes (un electrón y un muón). La desventaja en este caso son las dificultades y limitaciones en el análisis que supone trabajar con dos neutrinos, en lugar de uno solo.

La ventaja clara del canal ZZ , adicionalmente, es el hecho de que su estado final sea completamente reconstruible, formado por cuatro leptones y, de modo teórico, sin energía faltante.

El canal WZ representa un estado intermedio entre los dos anteriores. Además, una de las particularidades de este modo de producción es la carga del estado final. El hecho de que este estado final esté cargado podría permitir contribuciones de nueva física en la que interviniesen bosones cargados, y en concreto H^\pm . El higgs del Modelo Estándar es un bosón neutro, y el H^\pm no pertenece al SM.

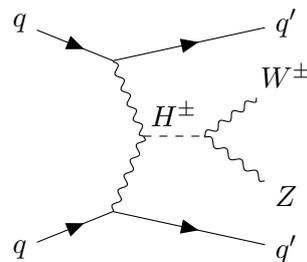


Figura 5: Diagrama de Feynman para un modelo de acoplo de los bosones W y Z a un Higgs cargado.

La búsqueda de este higgs cargado supone por tanto otra motivación física a la hora de estudiar este proceso. Si este diagrama contribuyese al mismo estado final que la producción VBS electrodébil, veríamos variaciones en los observables físicos del proceso. Por supuesto, para ello necesitaríamos minimizar enormemente todas las contribuciones a la incertidumbre de las medidas.

Tanto en el diagrama [3] como en el [5], así como en el resto de procesos VBS, los bosones que “emiten” los quarks son *los mismos* que encontramos en el estado final, y podemos ver el diagrama como un proceso de “scattering” elástico. En otras ocasiones, sin embargo, estos dos bosones se fusionarán para dar un único bosón en el estado final (como ejemplo, un W^+ y un W^- produciendo un Z). Estos procesos, que comparten muchas características con los VBS, son conocidos como VBF (*Vector Boson Fusion*).

1.4. Diagramas de Feynman a LO, NLO y NNLO

El diagrama de la figura [4] *no* es el único que contribuye a la producción WZ, como tampoco es el de la figura [3] el único que contribuye a la producción WZ VBS.

Sin embargo, estos diagramas representan las contribuciones *más sencillas* a ambos procesos. Aquí, entendemos por *más sencillo* aquel diagrama con menos vértices, y por tanto de menor orden.

Para calcular de forma *exacta* la sección eficaz de un proceso, tendríamos que considerar los diagramas de Feynman que contribuyen al mismo *a todos los órdenes*. Sin embargo, el número de contribuciones a un orden n crece muy deprisa con n , por lo que calcular una sección eficaz de modo teórico y de forma exacta es, a efectos prácticos, imposible.

Ilustraremos este hecho atendiendo al scattering de dos electrones ilustrado en la figura [6]:



Figura 6: A la izquierda, diagrama de Feynman de dos vértices para el scattering de dos electrones. A la derecha, una de las contribuciones con cuatro vértices al mismo proceso.

Todos los vértices en la figura superior son vértices electrodébiles. Cada uno de estos vértices, contribuye al diagrama en cuestión con un factor multiplicativo de α_{ew} , con $\alpha_{ew} \sim 1/137$. Este factor disminuye ligeramente con la energía en el centro de masas, aunque considerarlo constante supone en este caso una buena aproximación.

La contribución del diagrama derecho en de la figura [6] (orden 4) es un factor $(1/137)^2$ más pequeña que la del diagrama izquierdo (orden 2), con solo dos vértices. Si uno de los electrones radiase, en cualquier punto del diagrama izquierdo, un fotón, tendríamos un diagrama de orden tres.

Si un fotón fuese intercambiado entre los dos electrones en la parte superior o inferior del diagrama izquierdo en (6), o los dos electrones sufriesen scattering intercambiando dos fotones en lugar de uno, tendríamos dos casos que contribuirían de la misma manera que el diagrama con el “loop” de la derecha en (6).

En cualquier caso, nos referiremos con LO (*Leading Order*) a la contribución más sencilla a un proceso determinado. El diagrama izquierdo en (6) representa el la contribución a LO del scattering de dos electrones.

Si añadimos un vértice, habremos formado un diagrama a NLO (*Next to Leading Order*). Si el diagrama tiene dos vértices más, diagrama derecho en la figura [6], la contribución será a NNLO (*next to next to leading order*).

Cuando el vértice es fuerte (involucra gluones o quarks), este va pesado por la constante de acoplamiento α_s , que tienen una dependencia con la energía mucho más acentuada que α_{ew} .

Cuando añadimos órdenes en teoría de perturbaciones al cálculo de la sección eficaz de

un proceso estamos *sumando* diagramas a órdenes más altos, de forma que $\sigma_{NLO} > \sigma_{LO}$, $\sigma_{NNLO} > \sigma_{NLO}$ y sucesivamente. Aunque esto es cierto en general, podría no cumplirse para procesos mediados por QCD en los que un tratamiento perturbativo no constituya una buena aproximación (no es nuestro caso).

2. El detector CMS en el LHC

La evidencia experimental que soporta el Modelo Estándar viene de la mano de la física de altas energías y de los colisionadores de partículas. Cuanto menor sea la distancia a la que uno quiera observar, o cuanto más pequeña sea la partícula que *se quiere romper*, más energético tendrá que ser el acelerador que se destine a ello.

El “gran colisionador de hadrones” (LHC, *Large Hadron Collider*), es el acelerador de partículas más energético construido hasta la fecha. Está compuesto por 27 km de imanes superconductores formando un anillo, dentro del cual circulan, por dos tubos distintos, dos haces de protones en sentidos opuestos.

Varias cavidades de radiofrecuencia crean diferencias de potencial para acelerar los protones en zonas concretas del anillo. El fin de los imanes es curvarlos y mantener el haz focalizado (que tiende a abrirse por repulsión electromagnética), de modo que los protones no colisionen con las paredes del tubo.

Los dos haces se cruzan en cuatro puntos del anillo, donde se encuentran los detectores CMS, ATLAS, ALICE y LHCb. Antes de cruzarse se colocan otra serie de imanes que comprimen el haz con el fin de aumentar las posibilidades de interacción entre protones en cada cruce.

ALICE y LHCb son detectores específicos, diseñados expresamente para estudiar la física de iones pesados y la física de quarks b respectivamente.

CMS y ATLAS, por otro lado, son dos detectores de propósito general que permiten llevar a cabo una gran cantidad de estudios. Ambos tienen las mismas capas de subdetectores, aunque tanto la estructura como la tecnología utilizada en cada una de ellas es muy diferente en los dos experimentos.

La motivación de disponer de dos detectores de propósito general es comprobar que los resultados de uno pueden reproducirse, y por tanto reafirmarse en el otro.

El detector CMS (*Compact Muon Solenoid*), cuyos datos recogidos en 2016 han sido utilizados en este trabajo, es un detector cilíndrico de 15 m de diámetro, 29 de largo y un peso de 14000 toneladas (figura [7], referencia [26]). Tiene simetría axial (en torno al haz) y respecto al plano transversal que corta al haz en el punto de colisión.

A continuación daremos una visión global de este detector, así como de sus subdetectores

fundamentales. Una descripción más detallada puede encontrarse en [3]. Además, para cada subdetector existe un TDR (*Technical Design Report*) en el que se detalla su estructura y funcionamiento de manera muy completa y precisa.

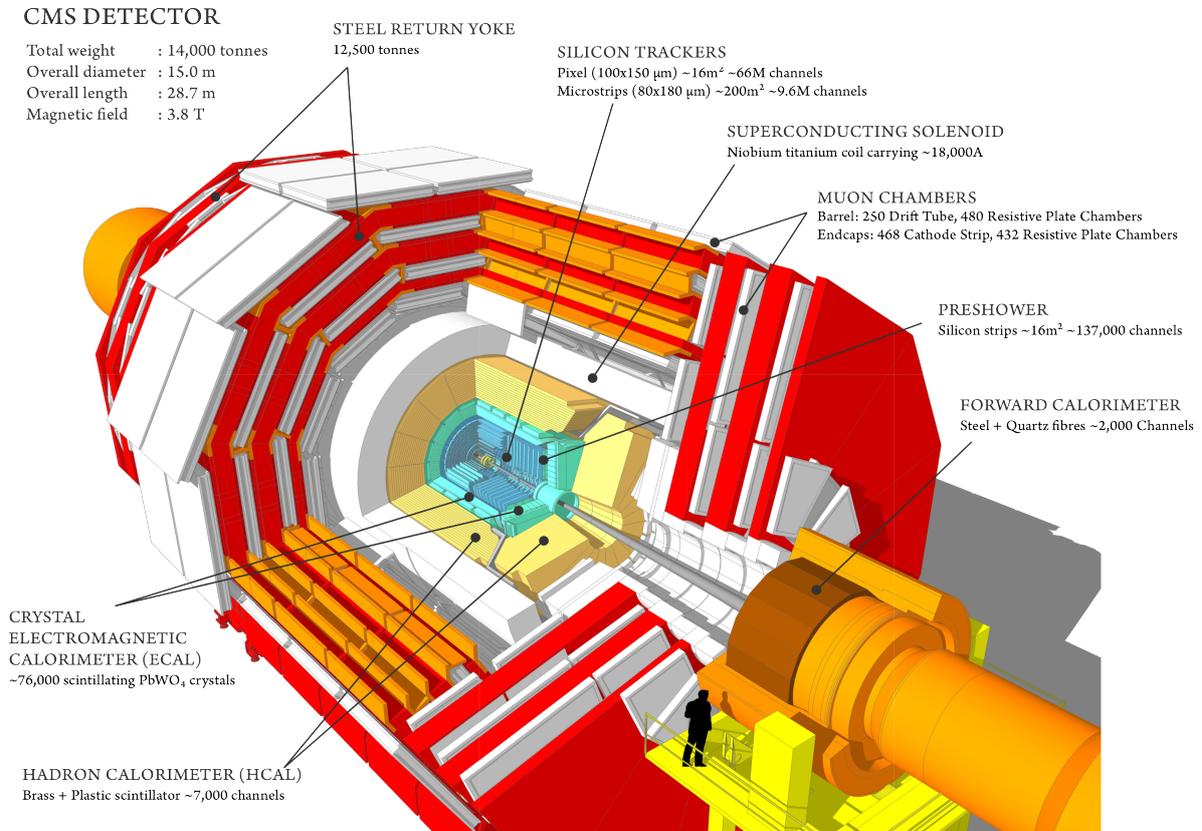


Figura 7: Estructura interna del detector CMS.

Los dos haces de protones se cruzan en el centro de CMS una vez por vuelta. En el momento de la colisión, ambos llevan una energía de 6.5 TeV. Si asumimos que el choque es frontal, la energía de la colisión en el centro de masas será en cada cruce de 13 TeV. Por contextualizar estos valores, 1 mg de protones a 6.5 TeV tiene una energía total de 10^{14} Julios, suficiente como para llevar al punto de ebullición 10^6 m³ de agua a 0 °C.

Cada colisión dará lugar a diversas partículas que, según la magnitud de su vida media, se detectarán directamente o se reconstruirán a partir de los productos de su desintegración.

2.1. El sistema de coordenadas de CMS

Es importante describir el sistema de coordenadas que se utilizará en lo sucesivo para caracterizar cada partícula detectada en CMS. Este sistema se ilustra en la figura [8].

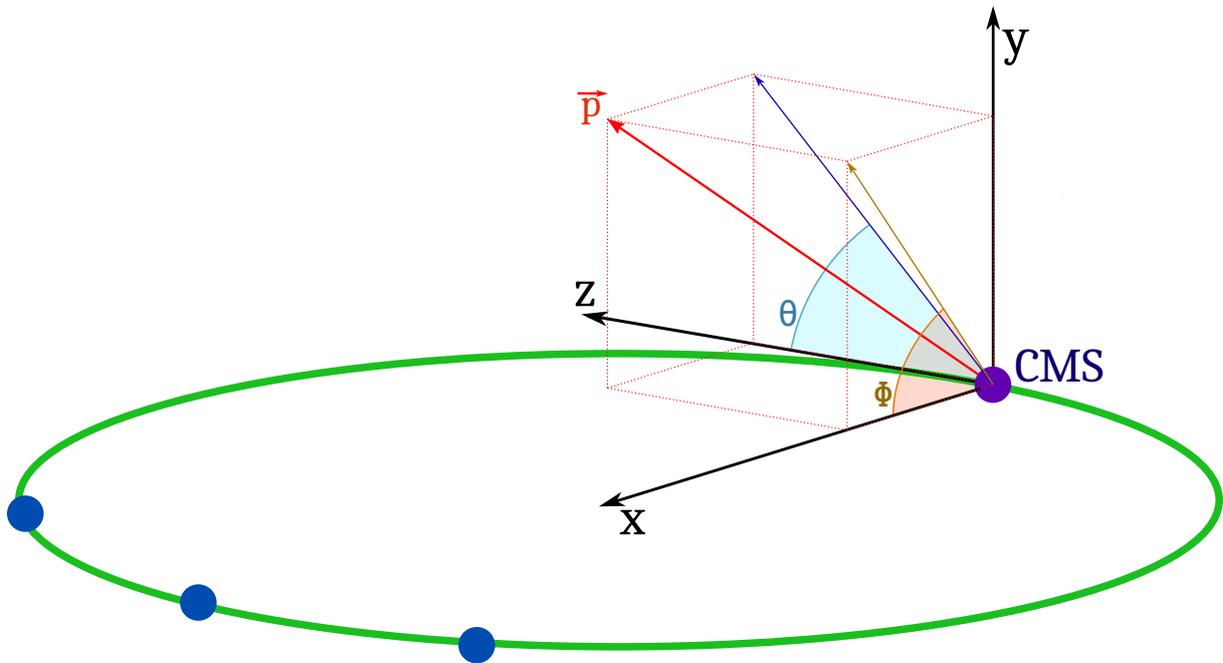


Figura 8: Esquema de coordenadas del detector CMS. ϕ cubre un rango de 2π y θ de π radianes. *Figura propia.*

En adelante, nos referiremos con *plano transverso* al plano xy . Así, el *momento transverso* (p_T) de una partícula será la proyección de su momento lineal en el plano transverso.

La física de cada colisión es simétrica respecto al ángulo ϕ . Sin embargo, no ocurre lo mismo con θ . cuanto más frontal sea el choque, y en general cuanto más energía se ponga en juego, más transversales se producirán las partículas del estado final. Estos sucesos son en general los de mayor interés.

Por este motivo, se define una nueva variable angular, la *pseudorapidez* η , como sustituta a θ . η vale 0 cuando $\theta = \pi/2$ (cuando no hay componente del momento en el eje z), y vale infinito cuando $\theta = 0$. En función del ángulo θ definido en la figura [8],

$$\eta \equiv -\ln[\tan(\theta/2)]$$

La variable η tiene por tanto mejor resolución cuanto más transversa sea la partícula. Para $\theta = 45^\circ$ $\eta = 0.88$; para $\theta = 10^\circ$ $\eta = 2.44$. Para $\theta = 0^\circ$, $\eta = \infty$.

Las piezas cilíndricas de los calorímetros electromagnético y hadrónico cubren aproximadamente hasta $\eta = 1.6$. La sección cilíndrica del detector de muones cubre hasta $\eta = 1.4$. A partir de estos valores, las partículas serán detectadas con las secciones *endcap* (“tapas”) del detector correspondiente, que llegan a cubrir hasta $\eta = 2.4$.

2.2. La estructura de capas de CMS

CMS es un detector compacto con una estructura en capas muy bien diferenciadas. Cada capa está destinada a un propósito diferente, y en concreto cada subdetector está pensado para caracterizar un tipo concreto de partículas.

Colocado en una posición aproximadamente igual a la mitad del radio de CMS se encuentra un imán superconductor que envuelve el detector en un campo magnético de 4 Tesla. Su objetivo es el de curvar las trazas de las partículas cargadas en el interior del detector. Dada la simetría cilíndrica del imán, todas las partículas cargadas se curvarán en un sentido (que depende de su carga y del sistema de referencia) en la zona interna del detector y en el sentido opuesto una vez pasado el imán.

Los subdetectores de CMS están especializados cada uno en un propósito concreto. Desde la parte más interna hasta la más externa, estos subdetectores son[22]:

El detector de trazas

El detector de trazas o *tracker* es la capa más interna de CMS. En él se rastrea la trayectoria de las partículas cargadas que lo atraviesan con el fin de obtener información tanto de la posición de la que provienen como de su momento.

Desde la parte más próxima al punto de interacción y hacia el exterior, el tracker consta de un detector de píxeles de silicio (~ 66 millones de píxeles de $100 \mu\text{m} \times 150 \mu\text{m}$ distribuidos en capas), y un detector de tiras de silicio hasta un radio de 1.1 m.

El tracker de CMS es un instrumento muy preciso, que permite entre otras cosas reconstruir (a partir de sus productos de desintegración) partículas con una vida media lo suficientemente corta como para desintegrarse antes de llegar a él, pero lo suficientemente larga como para alejarse un poco del vértice principal antes de desintegrarse. Concretamente, cada posición registrada en el tracker lleva asociada una indeterminación de apenas $10 \mu\text{m}$.

A lo largo de este trabajo, me referiré con *leptones* únicamente a aquellos que son medidos en el detector: electrones y muones. Por una parte, la vida media del τ (10^{-13} segundos) no le permite entrar en el detector, por lo que este no se mide directamente, sino que se observa a través de los productos de su desintegración. Por otra parte, la sección eficaz (probabilidad de interacción) de los neutrinos con la materia es, de cara a un detector como CMS, esencialmente

nula, por lo que estos objetos tampoco se miden directamente.

La información sobre la proveniencia de cada partícula se obtiene extrapolando su trayectoria calculada en el tracker a la zona de colisión. Con ello se conoce si la partícula rastreada proviene del vértice donde ha tenido lugar la colisión de interés, de un vértice distinto correspondiente a otra interacción protón-protón.

En adelante, nos referiremos con *vértice principal* al punto en el espacio donde se ha producido la colisión de interés. Normalmente este vértice se distingue por ser aquel del que provienen las partículas que computen el mayor valor de $\sum \vec{p}^2$. Los vértices que provengan de otra colisión protón-protón, o aquellos producidos por la desintegración de una partícula cuya vida media le ha permitido desplazarse ligeramente del vértice principal serán referidos como *vértices secundarios*.

En lo que sigue, llamaremos también *bunch* a cada paquete de partículas. En el LHC, circulan al mismo tiempo del orden de 1800 bunches separados entre sí una distancia que viene determinada por las cámaras de radiofrecuencia que los aceleran. Cada cruce de bunches en el punto de colisión será referido como *suceso*.

Además, teniendo en cuenta que una misma partícula cargada bajo la presencia de un campo magnético se curvará más o menos dependiendo de su momento, esta cantidad cinemática también podrá ser calculada a partir del radio de curvatura de las trayectorias medidas en el tracker.

Una descripción mucho más detallada del detector de trazas de CMS (en su primera versión) puede consultarse en su TDR en [5].

El calorímetro electromagnético (ECAL)

En esta segunda capa del detector son medidas por centelleo tanto partículas cargadas como fotones, transformando la energía de estas partículas en intensidad de luz, que es medida. Está compuesto por una serie de cristales que registran con gran precisión la energía de las partículas electromagnéticas que los atraviesan, especialmente fotones y electrones. Los protones o muones, con una masa mucho mayor, apenas dejan deposición en los cristales del ECAL, con lo que lo atraviesan sin gran dificultad y son detectados más adelante.

El TDR conteniendo toda la información sobre la primera versión de este calorímetro puede

consultarse en [6]

El calorímetro hadrónico

El HCAL (*Hadronic Calorimeter*) es el subdetector situado entre el ECAL y el solenoide superconductor (figura [7]). En él se miden protones, neutrones, piones, kaones, y hadrones en general. Su momento, posición, y tiempo de llegada quedan registrados en este detector.

El HCAL también utiliza centelladores, cuyas deposiciones de energía se amplifican y miden en forma de intensidad de luz, a raíz de la cual se calcula la energía de la partícula detectada.

Tanto el HCAL como el ECAL están divididos tres secciones localizadas en tres zonas diferentes del detector. La primera tiene forma cilíndrica y se localiza en la parte central, en torno al haz. Esta sección detectará partículas que abandonen el vértice de la colisión en una dirección relativamente perpendicular al haz (hasta $|\eta| = 1.4$). Las dos secciones restantes actúan como “tapas” de esta primera (también denominadas *endcap*), colocándose en sus extremos, hasta $|\eta| = 3$ (ver figura [7]). El HCAL (y solo el HCAL) tiene además una última sección *forward* que cubre hasta $|\eta| = 5.3$.

El TDR de este calorímetro puede consultarse en [7].

Las cámaras de muones

Los muones tienen una vida media de $2.2 \cdot 10^{-6}$ s en su sistema de referencia. Estos recorrerán, en un tiempo igual a su vida media, una distancia mínima de 600 m (este mínimo se corresponde con $E_\mu = m_\mu = 0.1$ GeV). La energía de los muones producidos en CMS es en general mucho mayor que su masa, y estas partículas son consideradas, de cara al detector (de 15 m de diámetro), estables

La masa de los muones, 200 veces superior a la de electrones y positrones, es la razón por la que estas partículas alcanzan esta última capa (la única más allá del imán superconductor) apenas sin interactuar en los calorímetros.

Del total de las cámaras (1400) que forman el detector de muones, se distinguen tres tipos diferentes, usando tecnologías distintas para detectar y medir muones: Las cámaras de deriva, las cámaras de placas resistivas y las cámaras de tiras de cátodos [8].

Las cámaras de deriva (250) constan a su vez de pequeñas “celdas” llenas de gas, y atravesadas por un hilo de oro cargado que actúa como ánodo (a un voltaje de 3600 V). Cuando un muón (o cualquier otra partícula cargada) las atraviesa, ioniza los átomos del gas, liberando electrones que se son acelerados produciendo una reacción de avalancha al chocar con otros átomos, que se traduce en una medida cuando alcanzan el ánodo. Estas cámaras de deriva se sitúan únicamente en la parte central del detector, cubriendo muones que abandonan el vértice principal con $|\eta| < 1.2$ respecto a la dirección del haz.

Las cámaras de tiras de cátodos (540) se encuentran en las “bases” del cilindro que forma CMS. Cubren todo el rango en $|\eta|$ desde 0.9 hasta 2.4, solapando con las cámaras de deriva en el intervalo $|\eta| \in [0.9, 1.2]$. Las componen un conjunto de hilos (ánodos) atravesados perpendicularmente por otro conjunto de tiras (cátodos) inmersos en un gas. Cuando una partícula cargada los atraviesa, los átomos que golpee se ionizan produciendo tanto electrones (que se miden en los hilos) como iones positivos (que se miden en las tiras). Identificando cada tira e hilo se obtiene una medida de posición en dos dimensiones.

Las cámaras de placas resistivas cubren la región $|\eta| < 1.6$. Estas son detectores con una respuesta temporal muy rápida, que se utilizan en el trigger de muones. Los muones al pasar provocan (también excitando átomos de un gas) una cascada de electrones que es medida en tiras cargadas. Las medidas de todas las tiras son analizadas electrónicamente para estimar el momento del muón y determinar así si el suceso debería ser guardado.

Para mucha más información, consúltese nuevamente el TDR de este detector, en [9].

2.3. Identificación y reconstrucción de partículas

En definitiva, para cada partícula detectada en CMS, se reconstruye su trayectoria utilizando los datos recogidos en el tracker y los tomados en el resto de detectores. La curvatura de cada partícula da información sobre la relación entre su carga y su momento, y el detector donde se haya medido identifica su sabor. La energía se puede estimar a partir de la curvatura de las trayectorias para los muones (siendo su masa conocida), y a partir de las deposiciones en el calorímetro correspondiente para el resto de partículas.

Toda partícula cargada (electrones, muones y hadrones cargados) dejará información en el tracker.

Las partículas neutras (fotones y neutrones) se distinguirán por ser detectadas en el ECAL y

el HCAL respectivamente, sin corresponderse sus deposiciones de energía con ninguna trayectoria en el tracker.

Los neutrinos no se detectarán, y su momento y energía en el plano transversal se calcularán imponiendo la conservación del momento en este mismo plano, en los estados inicial ($\vec{p} = 0$) y final. La figura [9] (referencia [27]) ilustra de manera muy visual el recorrido en el detector de las partículas que éste detecta.

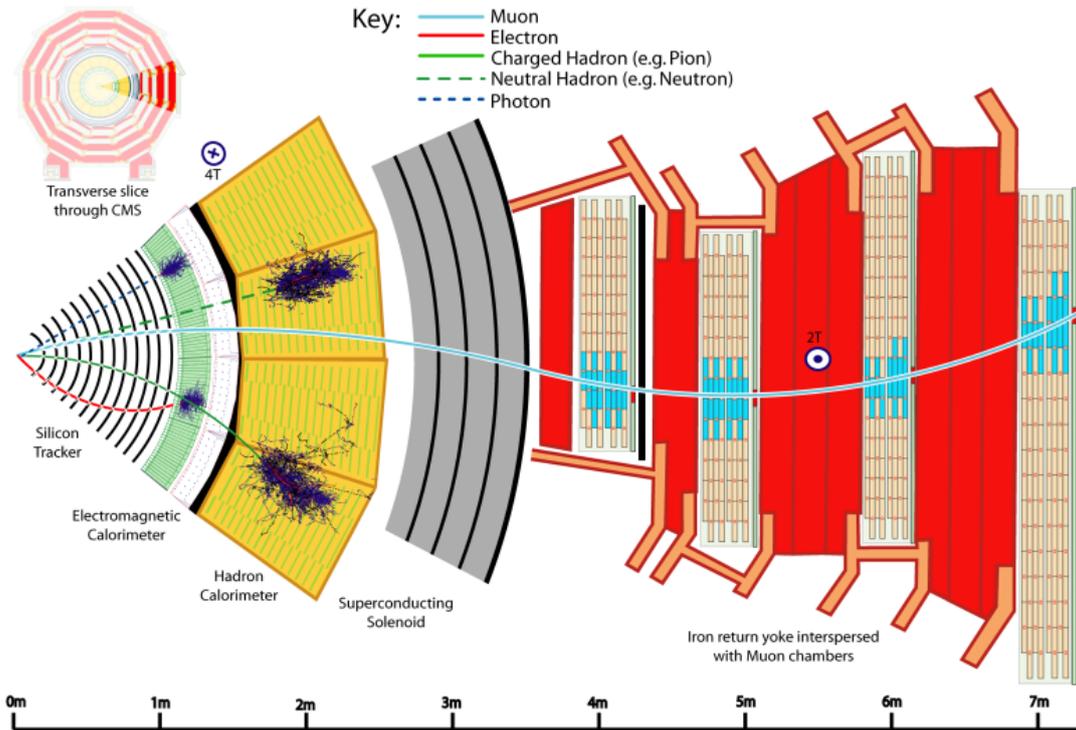


Figura 9: sección circular del detector CMS. Las deposiciones de energía de cada partícula en las sucesivas capas del detector ayudan a caracterizarla por completo.

Un caso especial son los sucesos con quarks o gluones en el estado final. Estos objetos no existen aislados en la naturaleza, y por tanto no se medirán directamente en el detector.

Cuando un quark o un gluón es producido en un estado final, este se hadronizará en un tiempo muy pequeño, que será en general mayor cuanto mayor sea su masa. Esa hadronización está enteramente descrita por la cromodinámica cuántica.

A los “chorros” de partículas provenientes de la hadronización de quarks y gluones nos referiremos en adelante como *jets*.

De cara al detector, los jets son conos de partículas con vértice en el punto en el que el quark se hadronizó. La energía del jet (y por tanto del quark) será la suma de todas las energías de las partículas del interior del cono, y lo mismo se aplica para su momento. Sobre la figura [9],

los jets dejarían varias trayectorias en el tracker y deposiciones en ambos calorímetros.

Los quarks u , d y s (~ 2 MeV, ~ 5 MeV y ~ 93 MeV respectivamente) tienen una vida media muy corta (ver figura [1]), y se hadronizan a efectos prácticos en el punto donde se producen. El quark c (1.27 GeV) viaja una muy pequeña distancia (del orden de la resolución del tracker en la mayoría de los casos) antes de producir jets.

El quark b , (4.18 GeV) tiene una “vida media” ya apreciable (del orden de 10^{-12} s), que le permitirá recorrer una distancia media de unos 2.5 mm antes de hadronizarse. La resolución del tracker es mejor que esta cantidad, y por tanto estos quarks podrán identificarse atendiendo a la posición del vértice del cono del jet que producen.

Supongamos que para un suceso aislado, el tracker identifica un conjunto de partículas cuyas trayectorias extrapoladas se cruzan en un punto separado aproximadamente 2.5 mm del vértice principal. En estos casos, con muy alta probabilidad, se esté identificando un quark b . Sin embargo, los algoritmos de etiquetado de quarks b (*b tagging*) son considerablemente más complejos que esto. En general están basados en cantidades tales como el parámetro de impacto de las partículas cargadas del interior del jet, las propiedades del vértice reconstruido o la presencia o ausencia de un leptón dentro del candidato a b -jet. Una descripción detallada de la identificación de quarks b en el detector CMS puede leerse en [10].

2.4. El momento en el eje z . El problema de los neutrinos.

Ya hemos comentado que, en el plano transversal, el momento total inicial es conocido, y por tanto podemos aplicar su conservación para calcular cantidades en el estado final no medibles directamente.

En el estado inicial, aunque sí es cierto que los haces chocan con un ángulo muy ligeramente menor que 180° (para poder dejar a los protones que no interactúan continuar su camino por tubos distintos en uno y otro sentido), puede asumirse sin cometer un error apreciable que el momento total inicial en las direcciones x e y (definidas en la figura [8]) es cero.

En el estado final, por otra parte, se medirá el momento de todas las partículas medibles, y se impondrá su conservación para calcular el de los neutrinos, invisibles de cara al detector.

En el eje z (el de la dirección del haz, ver la figura [8]) también se conservará el momento, aunque esta cantidad *es desconocida* en el estado inicial.

El LHC acelera actualmente protones a 6.5 TeV. A esas energías, sin embargo, las colisiones se producen entre los componentes del protón, y no entre los protones mismos.

Estos 6.5 TeV de cada protón estarán repartidos entre los partones (quarks y gluones) que lo componen. Cada partón tendrá una energía $E = x \cdot E_p$, con x una cantidad que toma valores entre 0 y 1. La probabilidad de encontrar un partón u otro con una determinada x ha sido calculada a partir de datos experimentales tomados en HERA², ATLAS y CMS [17], y las distribuciones de esta probabilidad se conocen como PDFs (*Parton Distribution Functions*). A “altas” x (del orden de 0.1) lo más probable es encontrar quarks de valencia (quarks u y d). A bajas x , prácticamente todas las colisiones se producirán entre gluones.

Aunque las PDFs sean bien conocidas a 6.5 TeV, estas no dejan de ser distribuciones de probabilidad, y no pueden por tanto utilizarse para conocer el momento inicial en el eje z .

Una explicación más extensa y detallada del significado y la utilidad de las PDFs puede encontrarse en [11].

El desconocimiento del momento inicial en este eje implica que la componente z en el momento lineal del neutrino (o los neutrinos)³ del estado final permanecerá desconocida. Entenderemos entonces por “energía faltante en el plano transversal” (MET, *Missing E_T*) la energía calculada en este plano aplicando la conservación del momento a todos los objetos medidos en el estado final. De modo teórico, la MET coincidirá con la energía de los neutrinos del estado final en el plano transversal.

Suponiendo por simplicidad que solo se produce un neutrino en el vértice principal, el valor de la MET supondrá, de modo teórico, una cota superior para la energía del neutrino, que coincidiría con su energía real si este saliese completamente perpendicular al haz.

Además, por ser su masa esencialmente nula, la energía de los neutrinos coincidirá en todos los casos con su momento (recuérdese $E^2 = m^2 + p^2$). Por tanto, se usará también P_T^{miss} como referencia a la MET.

²HERA fue un acelerador $e^\pm - p$ instalado en DESY (Alemania), que funcionó entre 1992 y 2007 a una energía en el CM en torno a 320 GeV[13]

³Estamos asumiendo que todas las partículas producidas en la colisión pertenecen al SM, y por tanto las únicas partículas invisibles al detector son los neutrinos.

Matemáticamente, la MET se calcula como

$$\text{MET} = \sqrt{\left(\sum_i \vec{p}_{i,x}\right)^2 + \left(\sum_i \vec{p}_{i,y}\right)^2}$$

donde el índice i recorre todos los objetos del estado final que provengan del vértice principal y hayan sido medidos directamente en el detector.

El cálculo de la MET lleva por tanto una de las más grandes incertidumbres asociadas, pues a él se propagan las incertidumbres en las medidas del resto de objetos del suceso. Entre otras cosas, un jet cuyo momento no se haya medido con gran precisión contribuirá a aumentar la incertidumbre de la MET, o un leptón que escape por los huecos del detector será también considerado como MET.

2.5. *Pile Up* en el detector CMS

El hecho de que el número de interacciones protón-protón sea en cada cruce de bunches del orden de 20 (figura 10, referencia [28]) constituye un importante fondo de partículas poco energéticas medidas en el tracker y en los calorímetros. En adelante, llamaremos a este hecho *Pile Up*. El *Pile Up* mide por tanto la cantidad de interacciones por cruce de bunches, y da entonces una buena idea de lo grande que será el fondo de deposiciones en cada subdetector.

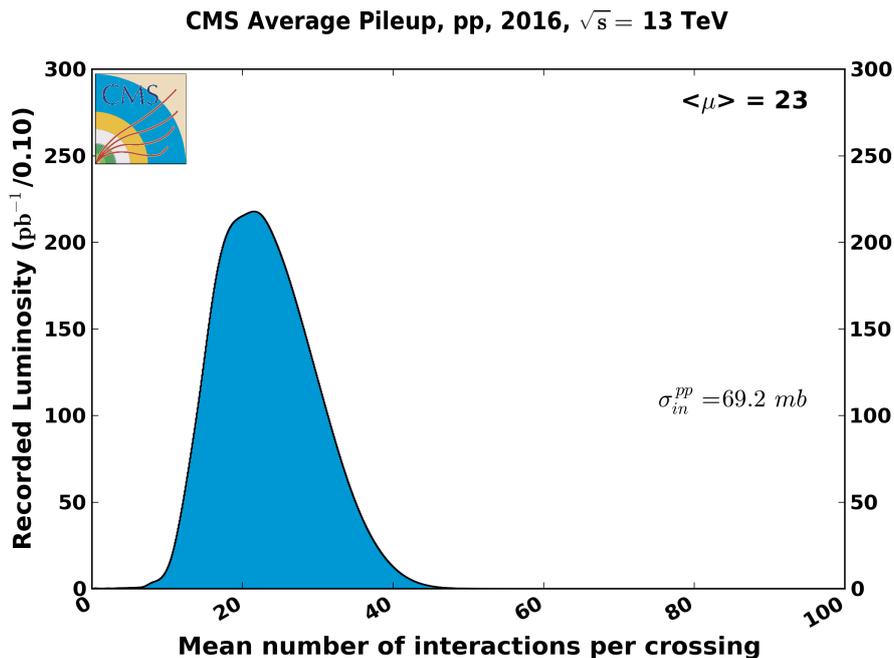


Figura 10: Curva de *Pile Up* recogida en 2016.

Aunque el número de sucesos de Pile Up medio en 2016 fue de 23 (figura [10]), este número no es constante durante toda la toma de datos (que recordemos dura del orden de un año).

El suceso de la figura [11], por ejemplo, tiene un Pile Up muy grande, de 86 vértices.

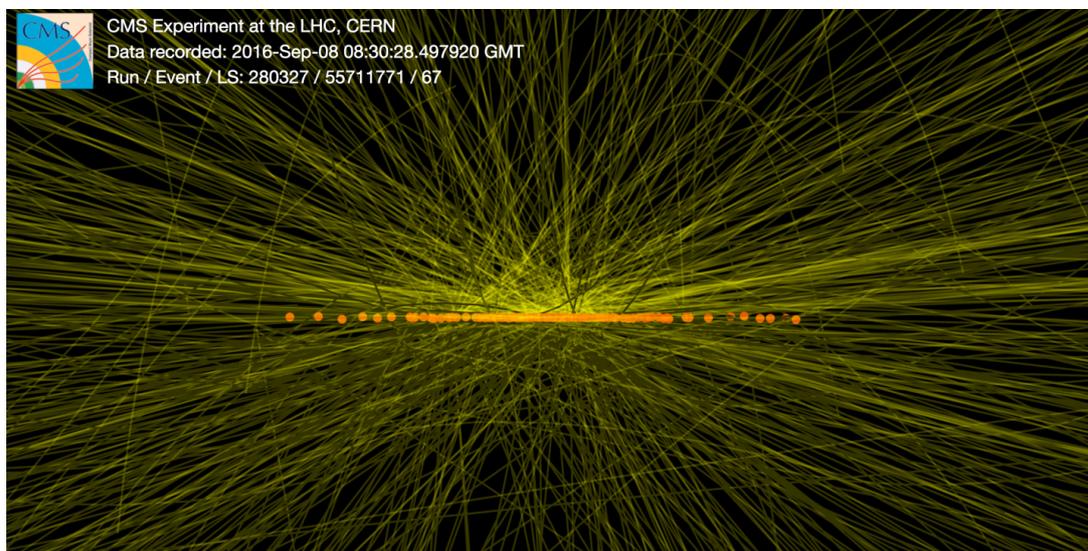


Figura 11: En este suceso de CMS (septiembre de 2016), el detector de trazas ha reconstruido 86 vértices (puntos naranjas). cds.cern.ch

En cada cruce de bunches, sin embargo, se escoge únicamente una entre todas las colisiones que tienen lugar. La elección se basa en cantidades cinemáticas de las partículas del estado final, y también a partir de ellas se construye el vértice principal.

Los productos de las interacciones en vértices secundarios, constituirán un fondo en el detector sobre el que tendrán que distinguirse los objetos con origen en el vértice principal. Esto supone una fuente de indeterminación, en concreto para la caracterización de jets.

Los jets de cara al detector son aglomeraciones (clústeres) de deposiciones de energía dentro de un cono con vértice en el vértice principal. La proyección de esos conos en cada calorímetro no deja de ser un círculo, cuyo radio se escoge en función de las características de cada clúster. La energía de los jets se calcula posteriormente sumando *todas* las deposiciones dentro de cada círculo, tanto las que *realmente* vienen del quark hadronizado como las que vienen de otros vértices secundarios, y restando luego las que se estima que provengan del fondo. Esta cantidad dependerá por tanto en gran medida del conocimiento del Pile Up de cada suceso.

Sin embargo, el Pile Up no es el único problema al que se enfrenta CMS a la hora de recoger datos. A los 6.5 TeV que viaja cada bunch, se producen en el LHC del orden de 40 millones de cruces de paquetes por segundo. Si cada uno de estos cruces lleva asociado una media de 23

sucesos de Pile Up (en 2016), el flujo de datos que recibe el detector por unidad de tiempo es enorme.

2.6. El trigger de CMS

Aunque la resolución temporal de cada detector permite a CMS distinguir entre un suceso y el siguiente (separados en el tiempo apenas 25 nanosegundos), tal cantidad de datos no puede guardarse en discos duros a la velocidad a la que se lee.

Sin embargo, en la mayoría de sucesos no se producirán colisiones de interés físico, sino que los estados finales estarán constituidos por partículas de baja energía sin ningún valor de cara a los análisis. Teniendo esto en cuenta, surge la necesidad de implementar un mecanismo en el detector que seleccione solo los sucesos interesantes, y no guarde información sobre los demás. Este mecanismo es lo que conocemos como *trigger*.

El trigger es una parte fundamental de la adquisición de datos en un detector como CMS. El trigger de CMS se divide en varios niveles de complejidad que actúan en diferentes etapas de la toma de datos.

El primero, el más básico y rápido, se implementa en el detector *online*, a nivel de *hardware* (electrónica). Es capaz de distinguir en tiempo real sucesos sin interés físico de sucesos con estados finales interesantes, que puedan corresponderse con procesos de baja sección eficaz, con signos de nueva física, etc. A este nivel se conocen las deposiciones en cada subdetector de partículas de cada estado final, pero aún no se ha reconstruido el suceso. Del orden de 100000 sucesos, pasan cada segundo este trigger.

En una etapa posterior, un segundo trigger a nivel *offline* (de *software*), sincroniza la lectura de cada detector para reconstruir cada uno de los sucesos que pasaron el primer nivel del dispositivo de trigger, y un análisis físico más exhaustivo escoge los (del orden de) 1000 sucesos que serán guardados para posteriores análisis específicos, como el que aquí presentamos. Una información mucho más detallada sobre la estructura y el funcionamiento de los triggers de CMS puede consultarse en [12].

2.7. Introducción al análisis computacional

El sistema de adquisición de datos (DAQ, *Data Acquisition*) de CMS constituye una pieza fundamental de la cadena de procesamiento de los datos que salen de este detector.

El enorme flujo de datos del detector motiva que el DAQ sea un sistema tan complejo como eficiente. Una vez salidos los datos del detector, en torno a 3000 ordenadores[18] se utilizan en paralelo para reconstruir cada suceso y descartar aquellos sin interés físico que aun así hayan pasado el primer trigger.

El producto final de todo este primer análisis será incluido en un fichero que contiene todos los sucesos guardados organizados en *N-tuplas* (tuplas de datos de N variables).

ROOT es la librería por excelencia tanto a nivel de análisis como de almacenamiento de datos en física de partículas, desarrollada y actualmente mantenida por el CERN. Está pensada tanto para guardar grandes cantidades de datos en espacios de memoria relativamente reducidos, como para facilitar la implementación de un análisis físico a nivel computacional.

Los archivos `.root` están estructurados en forma de árbol, siendo el “tronco” cada uno de los sucesos, las “ramas” cada cantidad física asociada a ellos (energía faltante, momento de los muones...) y las “hojas” los valores numéricos de tales cantidades.

Tras el primer procesamiento offline, los datos del detector están ya organizados de una forma más intuitiva, y listos por tanto para someterse a análisis⁴. Sin embargo, estas tuplas siguen siendo enormes, pues *todos* los sucesos que han pasado el trigger quedan registrados en ellas.

Por este motivo, es frecuente trocear estos ficheros en N-tuplas más pequeñas en función de algunas variables muy comunes, como puede ser el número de leptones contados en cada suceso. En el análisis que aquí se presenta, se buscarán entre todos los datos sucesos con tres o más leptones en el estado final. Utilizaremos por tanto una tupla de datos en la que esta preselección ha sido llevada a cabo con anterioridad. Esto reducirá considerablemente los tiempos de ejecución sin afectar a la estadística de los resultados.

Además, muchas variables de cada suceso que no vayan a ser utilizadas en un análisis concreto se eliminan de las N-Tuplas también en este punto del procesamiento. Con ello puede reducirse considerablemente el espacio que estos datos ocupan en el disco.

⁴Aunque esto sea cierto como idea básica, en realidad este proceso es mucho más complicado, y los datos pasan por una gran cantidad de filtros y análisis antes de ser guardados en las N-tuplas que utilizamos en este estudio.

El resultado de este análisis será un conjunto de N-tuplas mucho más pequeñas que los de entrada, que una vez producidos utilizaremos para obtener los resultados finales.

2.7.1. Herramientas informáticas

El gran tamaño de las N-tuplas de datos tanto de entrada como de salida del análisis, así como los recursos computacionales requeridos para ejecutarlo, son la razón principal por la que este tipo de análisis, en general, no puede llevarse a cabo en un ordenador personal.

Por tanto, todo el código es ejecutado de forma remota. Los servidores que he utilizado para llevar a cabo este estudio están físicamente localizados en la facultad de geología (8 cores Intel i7, 8 GB de RAM) y en la de ciencias (64 cores Intel Xeon, 126 GB) de la Universidad de Oviedo. La conexión a ambos ordenadores se realiza vía SSH. Todos los servidores en los que he llevado a cabo el análisis, usan GNU/Linux (CERN Centos 7 y Scientific Linux CERN 6) como sistema operativo.

Todos los ficheros de código fuente son editados en un entorno de desarrollo especializado desde mi ordenador personal (con las ventajas que ello conlleva frente a la edición remota), y sincronizados a tiempo real con los mismos archivos en el servidor, vía SSHFS (*SSH Filesystem*). Además, todo el código ha sido sincronizado periódicamente con un repositorio en Github.

El lenguaje de programación protagonista en todo el análisis es C++. Es también el lenguaje en el que está escrito ROOT, pues permite un control muy exhaustivo sobre la memoria y el procesador de cada ordenador, estando al mismo tiempo orientado a objetos.

Adicionalmente se han desarrollado varios scripts en bash. Este lenguaje se ha utilizado para automatizar la ejecución del análisis sobre cada muestra, utilizando más o menos recursos en función del tamaño de cada N-tupla. Además, se ha utilizado un sistema de colas para enviar el análisis al servidor de la facultad de ciencias. He automatizado también este proceso con varios scripts en bash.

El análisis sobre cada muestra se lleva a cabo utilizando PAF (*PROOF Analysis Framework*) [24], un framework desarrollado por el departamento de física de altas energías de la Universidad de Oviedo para este tipo de propósitos. PAF está basado en PROOF (*Parallel ROOT Facility*), una herramienta de paralelización muy utilizada en análisis basados en ROOT, cuando estos requieren de un poder computacional elevado.

2.7.2. Datos y Monte Carlo

De cara a entender y corroborar el comportamiento de las medidas es necesario disponer de un modelo sólido con el que compararlas. Por este motivo se utilizan generadores de Monte Carlo (MC), que emulan tanto la física de cada colisión como la reconstrucción en el detector de las partículas de cada estado final (además otros efectos como el Pile Up). Las N-tuplas que producen estos generadores se someten al mismo análisis que los datos.

Cada N-tupla de Monte Carlo se ha generado con una sección eficaz asociada al proceso específico que modele, que corresponde con la sección eficaz teórica que el SM predice para ese proceso a un orden finito en teoría de perturbaciones.

Existe prácticamente un Monte Carlo por cada proceso físico de interés que se produce en el LHC. Sin embargo, utilizar todos los disponibles (independientemente del proceso que modelen) para un análisis específico supondría una sobrecarga computacional innecesaria en todos los casos.

Supongamos, como caso práctico al párrafo anterior, que queremos estudiar el proceso de producción de pares de tops, que tiene quarks b en el estado final. La sección eficaz de este proceso es del orden de los 800 pb.

Supongamos ahora que disponemos de un MC que modela el proceso WZ VBS (que estudiamos en este trabajo). Su sección eficaz es del orden de 3 fb (1 fb = 0.001 pb), y además este proceso no tiene quarks b en el estado final.

Entonces, incluir el Monte Carlo que modele WZ VBS en el estudio de $t\bar{t}$ no supondrá ninguna ventaja, pues una vez aplicados todos los criterios de selección para aislar el estado final de $t\bar{t}$, el remanente de WZ VBS será completamente despreciable, y quedará completamente absorbido en la incertidumbre estadística de los datos.

Por tanto, no se incluirán en este análisis procesos con un estado final muy diferente o con una sección eficaz mucho menor tanto a los dos procesos que vamos a caracterizar. Discutiremos los MC que se incluirán en nuestro análisis concreto tras describir las peculiaridades físicas del mismo, en la sección 4.2.

En un caso ideal, representar en dos histogramas los datos recogidos y todas las N-tuplas de Monte Carlo debería producir, salvo fluctuaciones estadísticas, un resultado idéntico. En muchas ocasiones, sin embargo, será necesario aplicar correcciones a algún Monte Carlo particular una

vez generado, que por cualquier razón no se ajuste correctamente a los datos.

Usualmente, este desajuste MC - datos se traduce en un pico ligeramente desplazado en el eje horizontal o en un factor de escala en el eje vertical. Las causas más frecuentes son:

- La generación del Monte Carlo con una sección eficaz teórica ligeramente diferente a la real.
- La mala modelización de alguna característica del detector. Los detectores del LHC en general, y CMS en particular, son objetos muy complejos cuya simulación no es trivial en absoluto. Además, pequeñas desviaciones en alguno de sus componentes pueden provocar que las condiciones en las que se toman los datos no sean las mismas a lo largo de todo el periodo de medida.

Otras discrepancias más evidentes (la aparición de dos picos en datos donde el MC solo predice uno, la aparición de un pico en datos que no refleja el MC...) no pueden ser corregidos de forma sencilla, y podrían deberse desde a un error en la generación del Monte Carlo o algún problema en la toma de datos en el detector hasta a evidencias de nueva física.

3. La sección eficaz en física de partículas

Comencemos introduciendo la idea de sección en física de partículas, que permitirá más adelante entender el concepto de *fuerza de señal*. Ambos términos tienen un significado especial en esta ciencia, que conviene tener presente antes de avanzar en nuestro estudio.

3.1. El concepto de sección eficaz

La sección eficaz de un proceso en el contexto de la física de partículas es quizá la cantidad más importante asociada a él. Esta magnitud, medida en unidades de área, es una para cada proceso físico, y se interpreta como la *probabilidad* de que ocurra. Por ejemplo, $\sigma_{pp \rightarrow WZ}$ dará información sobre la probabilidad de encontrar un W y un Z en el estado final cuando el inicial es la colisión de dos protones.

A continuación encontraremos de forma intuitiva y visual una expresión para la sección eficaz en función de cantidades que pueden medirse en un acelerador de partículas. Comencemos entendiéndola en un sencillo experimento de blanco fijo, en el que un haz de partículas choca elásticamente contra los núcleos de una lámina de algún material [14].

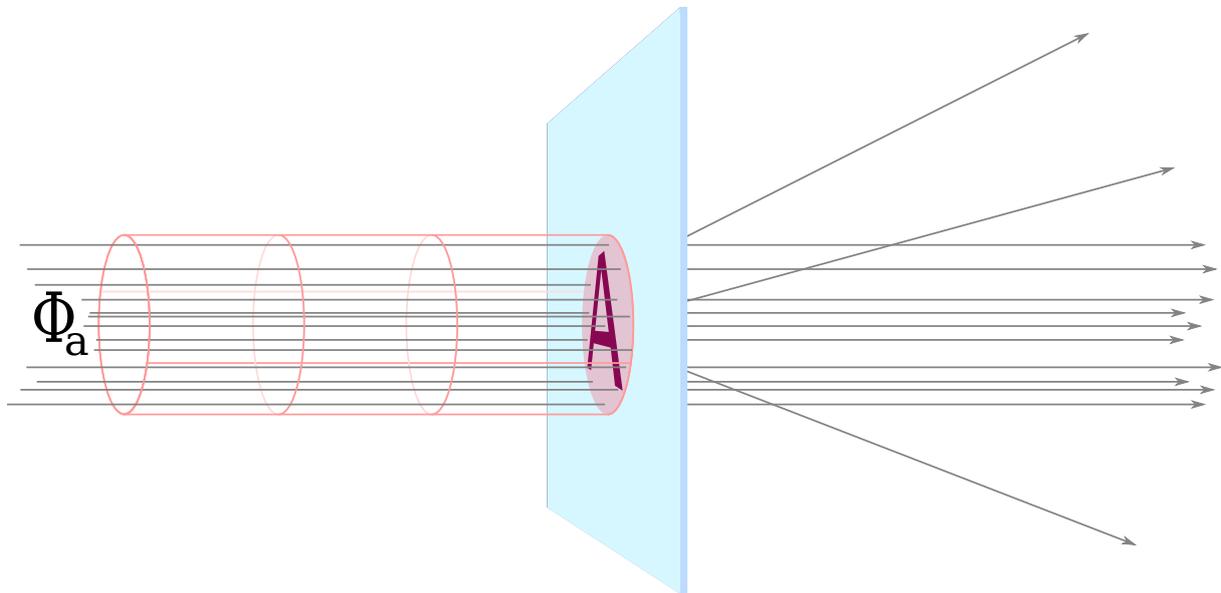


Figura 12: Un haz por el que circulan Φ_a partículas por unidad de área y tiempo colisiona con una fina lámina de material. Solo una pequeña porción de las partículas del haz interactúa y se desvía. *Figura propia.*

El haz incidente de la figura [12] está caracterizado por el número de partículas que entra en el blanco por unidad de área y tiempo, es decir, su flujo: $\Phi_a = \dot{N}_a/A$. En el blanco, dentro

del cilindro de área A contenido en el interior de la placa azul, se encontrarán un total de N_b partículas, que potencialmente podrían reaccionar con las del haz incidente.

Entonces, entendiendo σ como la probabilidad de que una partícula del haz reaccione con otra del blanco para producir un proceso concreto (en la figura [12], desviarse), el número de reacciones por unidad de tiempo podrá calcularse como

$$\dot{S} = \Phi_a \cdot N_b \cdot \sigma$$

donde, suponiendo que nos interesa estudiar la sección eficaz del proceso “desviación de una partícula”:

- \dot{S} es el número de partículas que producen por unidad de tiempo lo que en adelante llamaremos *proceso de señal*. El proceso de señal en este caso particular es *sufrir una desviación*, y \dot{S} será el número de partículas que se desvían por unidad de tiempo.
- $(\Phi_a \cdot N_b)$ da el número máximo de interacciones por unidad de tiempo, y es independiente del proceso concreto que se esté estudiando (que las partículas del haz se desvíen, reboten, queden atrapadas en el material...).
- σ es la probabilidad de que una partícula *se desvíe*.

Entonces, de forma general,

$$\sigma = \frac{\dot{S}}{\Phi_a \cdot N_b} \tag{2}$$

Que igual que en mecánica clásica, tiene unidades de área.

El denominador de la expresión (2) puede interpretarse como una cantidad de normalización que no depende del proceso. Esta cantidad se denomina *Luminosidad*, $\mathcal{L}(t)$, y tiene unidades de inverso de superficie por inverso de tiempo. Frecuentemente, sin embargo, se define la luminosidad integrada como

$$\mathcal{L}_{tot} = \int_0^T \mathcal{L}(t) dt$$

que representa el total de sucesos (independientemente del proceso que haya tenido lugar) que se dan en un periodo de tiempo concreto, y tiene unidades de inverso de superficie.

En un acelerador como el LHC, la luminosidad integrada depende esencialmente de la dura-

ción del periodo de tiempo de adquisición de datos. Todos los datos con los que se trabaja han sido recogidos en un periodo finito de tiempo (un mes, un año...), y por tanto les corresponde una luminosidad integrada determinada. La luminosidad recogida en CMS en el año 2016 se representa en la figura [13] (referencia [29]).

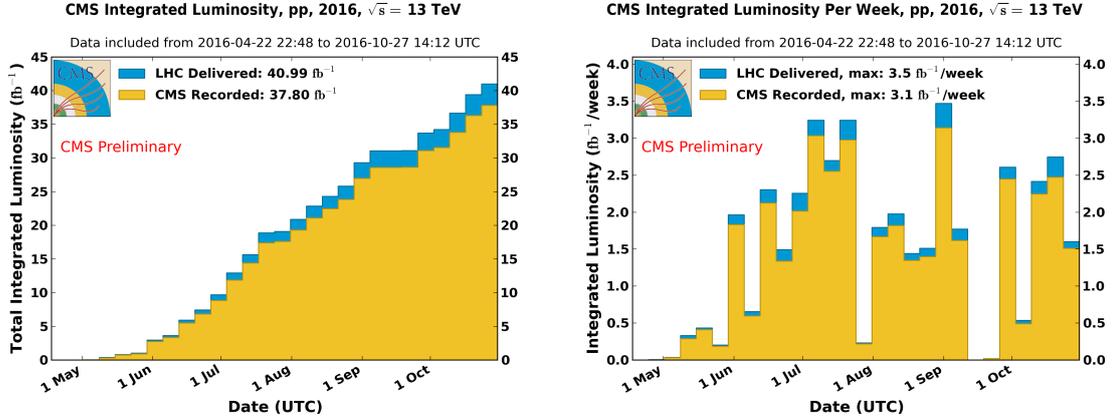


Figura 13: Luminosidad integrada entre mayo y noviembre de 2016. A la izquierda, la luminosidad *acumulada* desde mayo, para cada semana hasta noviembre. A la derecha, la luminosidad registrada cada semana en el mismo periodo de tiempo. La luminosidad que registre CMS será siempre menor o igual que la que suministre el LHC.

Cuando los datos con los que se trabaja se corresponden con una luminosidad integrada determinada, es frecuente modificar la ecuación (2) para calcular la sección eficaz según

$$\sigma = \frac{S}{\mathcal{L}_{tot}} \quad (3)$$

Donde S es el número de veces que contamos el proceso de señal entre el total de los datos tomados en un periodo de tiempo concreto, y \mathcal{L}_{tot} la luminosidad correspondiente a la muestra de datos sobre la que realizamos el análisis, que da información sobre la cantidad de sucesos que han tenido lugar en el mismo periodo de tiempo.

3.2. El concepto de la *fuerza de la señal*

Muchas de las partículas que se producen en el vértice principal no son recogidas en el detector. En primer lugar, aunque estas pueden abandonar el vértice en cualquier dirección, el detector no cubre todo el espacio y muchos objetos por tanto no serán detectados. La aceptación, \mathcal{A} , es un factor multiplicativo menor que 1 que se añade al denominador de (3) para cuantificar

este hecho. Introducirlo implica considerar de entre todos los productos de la colisión, solo aquellos que realmente entran en el rango de acción del detector.

Además, es también necesario tener en cuenta las eficiencias de cada componente del detector. La eficiencia del trigger, la del tracker, la del etiquetado de quarks b o la de la medición de la energía de los jets son cantidades que se añaden también multiplicando a la luminosidad y se tienen en cuenta en el cálculo de σ .

Queda entonces

$$\sigma = \frac{S}{\epsilon \cdot \mathcal{A} \cdot \mathcal{L}_{tot}} \quad (4)$$

con ϵ el producto de todas las eficiencias. Este producto es la cantidad más difícil de determinar al calcular una sección eficaz, pues todos los componentes del detector llevan asociada una ineficiencia, que aunque muy pequeña no es en general despreciable.

La *fuerza de la señal* en física de partículas es una cantidad que da información sobre la distancia entre la sección eficaz medida para un proceso de señal y la sección eficaz teórica que predice para él el Modelo Estándar, o cualquier otra teoría alternativa.

Esta cantidad se define como

$$\mu = \frac{N - B}{S}$$

donde

- $N - B$ es el número de sucesos del proceso de señal calculado como la diferencia entre el número total de sucesos de datos N recogidos por el detector en unas condiciones determinadas, y el número total de sucesos de cada uno de los fondos, B , predicho por la teoría en esas mismas condiciones
- S es el número de sucesos del proceso de señal predicho por la teoría en las mismas condiciones en las que se han tomado los datos.

Si $\mu \sim 1$, $N = B + S$, y los datos estarán incluyendo, además de los fondos, el proceso de señal. Substraer a los datos todos los fondos (que asumimos conocidos y bien modelados por la teoría) dejará un remanente que corresponderá con sucesos del proceso de señal. En un caso ideal, si $\mu = 1$ la realidad será exactamente tal y como la teoría la predice.

Por otro lado, si $\mu = 0$, el número de sucesos de fondo igualará a los datos y no dejará por tanto espacio a la señal. Entonces, la teoría estará prediciendo un proceso físico que no se observa en la realidad. En todo caso se asume que los fondos modelan bien los procesos que describen, y por tanto $B \leq N$ y $\mu \geq 0$. Una μ negativa no tiene sentido físico aunque esté bien definida matemáticamente.

Sin embargo, la utilidad de esta medida es muy limitada si no se acompaña de su incertidumbre. Desde $\mu = 0$ (no hay señal) hasta $\mu \sim 1$ (existe señal y se corresponde con teoría que usamos para describirla), la incertidumbre en μ dará información sobre el punto hasta el que el valor central da un resultado fiable.

En muchas ocasiones, el valor central de la fuerza de la señal se *ciega* a los datos haciéndose independiente de N (para distinguirla de la real, llamaremos a esta fuerza de señal “ciega” μ_c), con el fin comprobar en base a su error como de compatible es el resultado de μ (la medida real) con que realmente haya o no haya señal. Se trabaja entonces con las siguientes dos hipótesis opuestas:

- $\mu_c = 1$: se asume aquí que $N = S + B$, y por tanto que la señal existe. La incertidumbre de este valor dará el nivel de confianza con el que podremos afirmar que la medida de μ asegura la presencia de señal. Este nivel de confianza será mayor cuanto más pequeña sea la indeterminación en μ_c . Si esta indeterminación es superior a 1, y abarca por tanto el caso $\mu_c = 0$, no podremos asegurar que la señal exista, independientemente del resultado que se obtenga con los datos.

Sin embargo, si esta indeterminación es pequeña, el valor de μ calculado con los datos podrá justificar la presencia de señal siempre su margen de incertidumbre no llegue hasta $\mu = 0$.

- $\mu_c = 0$: En este caso se asume que $N = B$ y que la señal no existe. A diferencia del caso anterior, por muy pequeña que sea aquí la incertidumbre nunca podremos afirmar que *no hay señal* donde la estamos buscando, pues siempre podrá haberla siempre y cuando esta absorbida por la incertidumbre de los fondos. Si la incertidumbre en μ fuese mayor que uno, no podríamos en ningún caso afirmar que no se ha medido señal.

Calcular posteriormente la fuerza de señal con los datos medidos indicará con cual de las dos hipótesis es compatible la medida experimental.

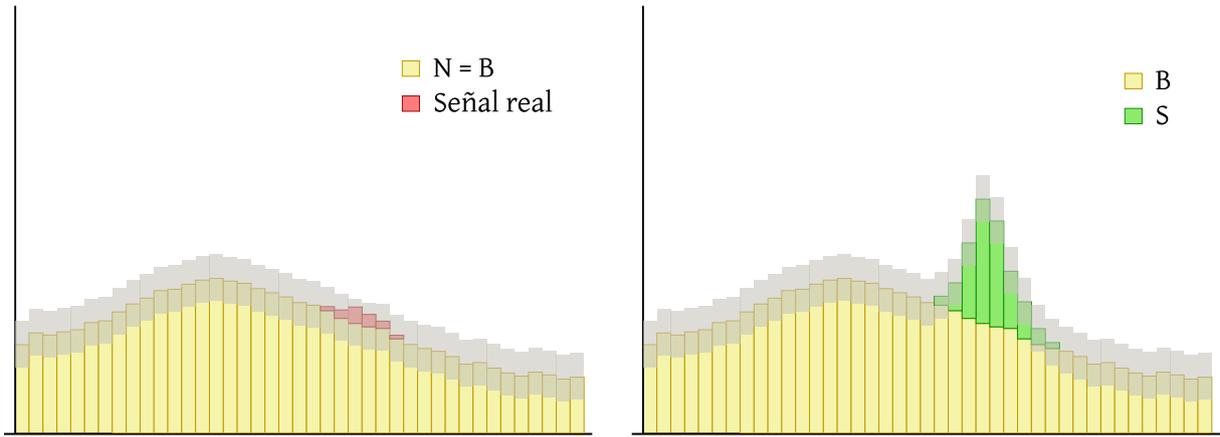


Figura 14: Ilustración gráfica del significado de imponer $\mu = 0$ (izquierda) y $\mu = 1$ (derecha). En gris se ilustra la indeterminación de los MC que modelan fondos (izquierda) y señal + los fondos (derecha). *Figura propia.*

En el histograma de la derecha de la figura [14], si la incertidumbre en gris es lo suficientemente pequeña podremos afirmar que el pico verde (el proceso de señal modelado por la teoría) está reproduciendo fielmente la realidad. En la figura de la izquierda, por muy pequeña que sea la incertidumbre de los fondos, nunca lo será lo suficiente como para afirmar que en la zona en cuestión *no* exista señal. Sí que lo será, sin embargo, para afirmar que en caso de haberla esta será comparable con la incertidumbre de los fondos (zona roja).

Las incertidumbres que contribuyen a esta cantidad se discutirán en la sección 5.1. Posteriormente, la fuerza de señal se calculará para los procesos de producción de WZ y WZ VBS en las secciones 5.4 y 5.5 respectivamente.

4. Análisis físico de los procesos de producción WZ y WZ VBS

Comenzaremos esta sección definiendo la *masa invariante*. Más adelante se describirán los rasgos físicos de los procesos de señal WZ y WZ VBS, así como de sus principales fondos. Se estudiará la validez de la modelización de los principales fondos estudiándolos por separado, y se caracterizarán por último de forma detallada los dos procesos de señal.

4.1. El concepto de masa invariante

La masa invariante (M) es una de las magnitudes más relevantes en cualquier análisis en física de partículas. En particular, es de especial importancia en aquellos análisis que involucren desintegraciones.

La masa invariante de una partícula o un sistema de partículas es una cantidad que *no depende del sistema de referencia*. El valor que se calcule estando el sistema en reposo, será el mismo que el que se obtenga teniendo el sistema un momento \vec{p} arbitrario. La masa invariante, además, se conserva en los estados inicial y final.

Matemáticamente, esta cantidad se define como

$$M^2 = \left(\sum_i \mathbf{P} \right)^2 \quad (5)$$

donde el índice i recorre el conjunto de partículas de un estado, y \mathbf{P} es el cuadrimomento de cada una de ellas.

\mathbf{P} es un vector de componentes $(E, \vec{p}) = (E, p_x, p_y, p_x)$, cuyo cuadrado será $\mathbf{P}^2 = E^2 - p^2$.

Teniendo ahora en cuenta la conocida ecuación de Einstein $E^2 = m^2 + p^2$, podemos señalar que el momento \mathbf{P}^2 de una sola partícula coincide con su masa al cuadrado. Además, de cara al detector, la masa de una partícula producida en el vértice principal coincidirá con la masa invariante calculada con los productos de su desintegración independientemente del desconocimiento del momento total inicial en el eje z

Consideremos como ejemplo un proceso en el que (pongamos) un Z se desintegra a dos leptones:

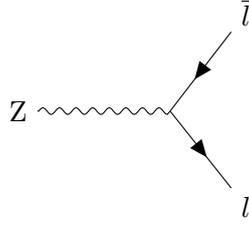


Figura 15: Un bosón Z se desintegra a dos leptones.

En la desintegración del Z ilustrada en la figura [15], la masa invariante del estado inicial coincidirá con la masa invariante del Z. Ahora: esta es una cantidad conservada, por lo que la masa invariante de los dos leptones del estado final *también* coincidirá con la masa del Z (entendiendo la masa del Z como la distribución que es, con su anchura característica y su desviación típica), independientemente de sus momentos lineales.

Para dos leptones, como es el caso, la masa invariante se calcula como

$$\begin{aligned}
 M_Z^2 &= (\mathbf{P}_{l_1} + \mathbf{P}_{l_2})^2 = \mathbf{P}_{l_1}^2 + \mathbf{P}_{l_2}^2 + 2\mathbf{P}_{l_1} \cdot \mathbf{P}_{l_2} = \\
 &= m_{l_1}^2 + m_{l_2}^2 + 2(E_{l_1}E_{l_2} - \vec{p}_{l_1} \cdot \vec{p}_{l_2})
 \end{aligned}
 \tag{6}$$

Donde la energía de los leptones se calcula según $E_l = \sqrt{m_l^2 + p_l^2}$, y por tanto todas las cantidades de la ecuación (6) son masas (conocidas) y momentos lineales (medidos).

En un suceso real, solo vemos los productos del estado final (dos leptones en este caso). Calcular su masa invariante resultará en la masa de la partícula de la que provienen.

Si suponemos ahora que un leptón del diagrama [15] emite un fotón, no tendremos más que incluir este objeto en la ecuación (6) sabiendo que su masa es cero. La masa invariante de los tres objetos seguirá coincidiendo con la masa del Z.

4.2. Rasgos físicos de la producción WZ VBS y WZ QCD

Ya en la sección 1.3 se ha introducido el proceso de producción VBS electrodébil de dos bosones W y Z. Se ha introducido también su principal fondo, la producción de WZ QCD. Recordemos que esta producción es un caso particular de la producción WZ en el que los quarks iniciales han emitido dos gluones.

Describamos ahora los rasgos físicos tanto del proceso de señal WZ VBS como de su principal

fondo (WZ QCD), con el fin de aplicar justificadamente en las secciones posteriores criterios de selección a datos y a Monte Carlos para aislar en la medida de lo posible el proceso de señal.

Comencemos recordando los diagramas de Feynman que representan ambos procesos.

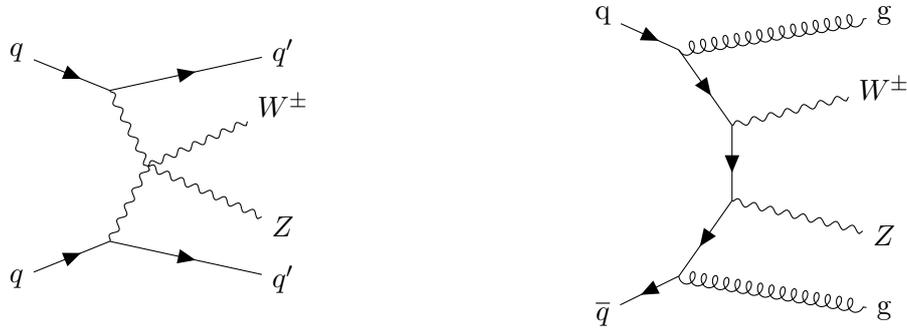


Figura 16: A la izquierda, el diagrama de Feynman del proceso de señal (WZ VBS). A la derecha su principal fondo, la producción QCD del mismo estado final.

En la sección 4.7 describiré los criterios de selección que se aplicarán a las N-tuplas de datos de entrada, para aislar de entre todos sus sucesos aquellos que con mayor probabilidad se corresponden con el proceso de señal. Estos criterios de selección se basarán en las cantidades cinemáticas (momento transverso, pseudorapidez, ...) de los objetos del estado final de WZ VBS. Conviene por tanto tener presente la física que hay detrás, y en particular las señas de identidad de este proceso que lo distinguen de sus fondos principales.

La gran diferencia entre ambos diagramas está en que en la producción electrodébil (diagrama izquierdo en la figura [16]), los quarks (jets) producido un bosón. Es de esperar por tanto que los quarks del estado final conserven la mayor parte de la energía inicial, y por consiguiente también el momento. Se esperará entonces que estos quarks (de cara al detector, jets) se desvíen poco de su trayectoria inicial, y salgan por tanto del vértice principal con una pseudorapidez relativamente alta.

En el caso de la producción QCD (diagrama derecho en la figura [16]), los jets del estado final son gluones emitidos por los quarks. Estos *gluones de radiación*, se emitirán en el vértice principal sin ninguna dirección preferencial, y con una energía en general mucho menor a la de los quarks del proceso de señal.

Los dos jets de ambos estados finales tendrán por tanto distribuciones visiblemente diferentes tanto en η como en energía. Estas cantidades supondrán los criterios de selección más decisivos de cara al análisis.

Las distribuciones de E y η de los jets provenientes *de quarks* (y por tanto producidos de

modo electrodébil) pueden explicarse de modo más riguroso atendiendo a la fracción de la energía del protón que estos llevan en promedio antes de la colisión. Como ya hemos discutido en la sección 2.4, los quarks y gluones dentro del protón (y en concreto los quarks *de valencia* u y d) tienen en promedio una fracción ‘ x ’ del total de la energía del protón. Esta x no tiene un valor concreto, sino que sigue una distribución que toma valores entre cero y uno. A diferencia de los gluones, el máximo de esta distribución para los quarks de valencia del protón *no se encuentra en cero*.

Esta idea puede utilizarse para justificar que la distribución de la energía de los jets de la producción WZ VBS electrodébil tendrá un máximo en algún valor *no nulo*. Sin embargo, no podremos atender a las PDFs para explicar las distribuciones de los jets de QCD, pues los gluones que los producen no provienen del interior del protón, sino que son radiados por los quarks del estado inicial. Sin embargo, el proceso de radiación de un gluón (y de cualquier otra partícula) es más favorable cuanto más pequeña sea la energía que cueste producir el estado final. Siendo la masa una constante (en este caso igual a cero), la radiación de un gluón será más energéticamente favorable, y por tanto más probable, cuanto menor sea la energía del gluón emitido.

Una vez entendidas las características de los jets del estado final, comentemos brevemente los rasgos físicos de dos bosones W y Z presentes tanto en WZ VBS como en WZ, y por tanto en WZ QCD.

La corta vida media de estos dos bosones ($\tau \sim 10^{-25}$ s) implica que estos se desintegran a todos los efectos en el mismo vértice en el que se producen. Se reconstruirán por tanto a partir del conjunto de sus productos que tengan una vida media suficiente como para entrar al detector.

El modo de desintegración principal del W es a quarks (jets de cara al detector), y solo en el 11 % de las veces se desintegra a $l + \nu$. Sin embargo, a la energía a la que se producen las colisiones en el LHC, el proceso predominante (con mayor sección eficaz) es con gran diferencia la producción de pares $b\bar{b}$, que de cara al detector son jets. Esto, a efectos prácticos, significa que vamos a tener que distinguir el proceso de señal sobre un fondo enorme de jets. Por esta razón, aunque el modo leptónico del W no sea el predominante, y trabajar en él conlleve tener que lidiar con energía faltante, escogeremos este canal de desintegración por su limpieza en comparación con el canal hadrónico.

El Z es un bosón, en términos del detector, mucho más sencillo de reconstruir que el W. Su modo de desintegración predominante (70 %) es también el hadrónico, y lo ignoraremos igual

que hicimos con el W.

El siguiente modo del Z es a neutrinos (que son invisibles en el detector), con un Branching Ratio (BR, “probabilidad de desintegración”) del 20 %,

Por detrás está el modo a $l\bar{l}$, con un BR del 10 %. Este es el modo en el que, también por su limpieza en el estado final, reconstruiremos el Z.

4.3. Procesos de fondo. La estructura de los histogramas

Esta sección está dedicada a los Monte Carlos que serán utilizados para modelar los fondos principales de los procesos de señal WZ y WZ VBS. Estos Monte Carlos serán sometidos al mismo análisis por el que pasarán la señal y los datos, y serán representados también en los mismos histogramas.

Antes de entrar en detalle, voy a dar unas pequeñas notas para entender los histogramas que se incluirán en lo sucesivo. Todos los histogramas de este trabajo tienen la misma estructura, que queda ejemplificada con los dos siguientes:

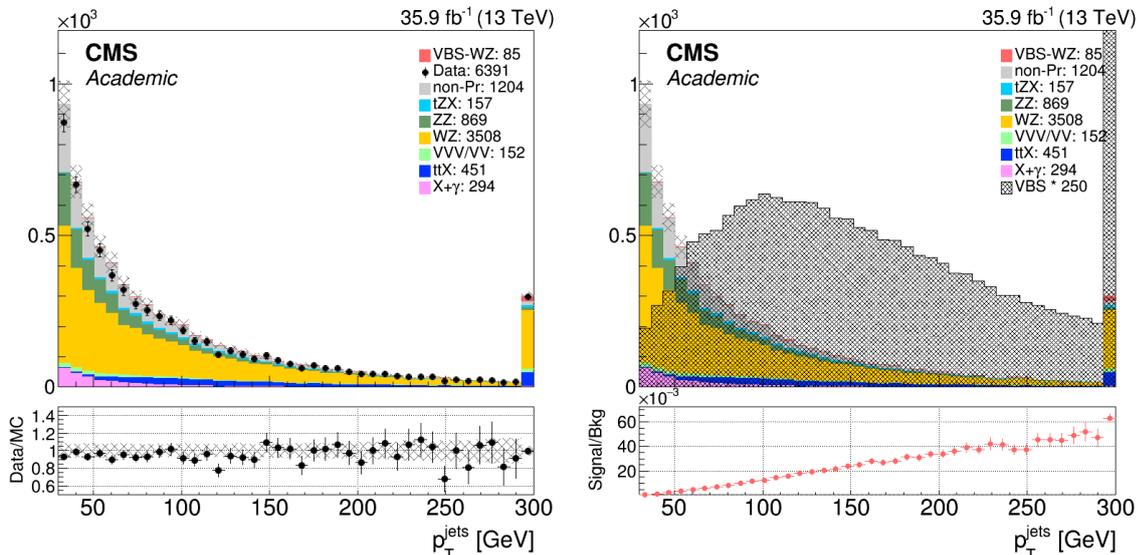


Figura 17: Momento transverso de los jets para todos los sucesos con tres leptones en el estado final.

Los histogramas de la figura [17] se dividen en dos *pads*. En el *pad* superior se representan los histogramas apilados de cada proceso. Cada histograma empieza donde acaba el anterior. Solo el que se representa en la base comienza en cero. El orden de los histogramas en la representación es el mismo que aparece en la leyenda.

Sobre cada columna se representa además con una rejilla gris la suma del total de los errores de todos los Monte Carlos (incluido el que modela el proceso de señal). Estos errores se dividen en estadísticos y sistemáticos. Los comentaremos en detalle más adelante (secciones 5.1.1 y 5.1.2).

Con puntos negros representamos los datos. El histograma de datos empieza a contar desde la base, de manera que en un caso ideal Datos y MC tendrían la misma altura en cada columna. El eje vertical mide número de sucesos, sin unidades. Con una barra vertical se representa la incertidumbre estadística de los datos.

En la esquina superior derecha, se indica qué proceso o procesos físicos representa cada color. Se acompaña este nombre con el total de los sucesos del proceso que han sido representados sobre el histograma. Teóricamente, la suma de todos los colores coincidiría con el valor de “Data”. La señal aparecerá en la parte superior, seguida por los datos y por el resto de procesos. En algunas ocasiones, además, se representará sobre todos los histogramas el MC de señal amplificado un factor que permita situarlo a una altura similar a la del resto. Con esto, que ejemplificamos en el histograma de la derecha en [17], ayudamos a comparar las distribuciones de la señal con la de sus fondos en los casos en los que el fondo sea claramente predominante.

Sobre el pad superior se indica la luminosidad a la que corresponde la toma de datos (común a todos los histogramas), así como la energía en el centro de masas.

El pad inferior es el pad de ratios. En el histograma de la izquierda en la figura [17] representamos con un punto negro el cociente entre los datos y el total de las predicciones para cada columna. Las barras verticales indican la incertidumbre estadística, y las horizontales la anchura de cada columna. Con una rejilla se representa la propagación del total de las incertidumbres de los MC a este cociente. En el de la derecha, con puntos del color del proceso de señal se representa el cociente señal/fondo. A la izquierda, en la parte superior, aparece el factor multiplicativo que escala el eje vertical.

El pad superior y el inferior comparten el mismo eje horizontal. La cantidad que se mida en este eje puede tener o no tener unidades.

Los alias utilizados como referencia a cada proceso físico, que aparecerán en la leyenda de cada histograma, son los siguientes:

- VBS-WZ: Con este color representaremos el Monte Carlo que simula la producción del proceso de señal: WZ VBS.

- Data: Los datos han sido tomados directamente en el detector. El proceso físico al que se refieren es a priori desconocido, aunque en un caso ideal este proceso será uno de entre todos los que se representan como señal o fondos.
- non-Pr: Con “Non Prompt” nos referiremos a todos los leptones que no provienen de un W, de un Z o de un τ . La inmensa mayoría de leptones Non Prompt provendrá del interior de un jet, en el caso en el que este no se identifique como tal pero sí que se mida el leptón de su interior. Otra pequeña parte tendrá origen en vértices secundarios, sin relación con el principal. Tanto los primeros como los segundos serán en general poco energéticos, constituirán por tanto una gran contribución al fondo a energías bajas.

El fondo de Non Prompt depende de multitud de factores complejos de modelar por Monte Carlo. Por tanto, este fondo no se modela, sino que se estima a partir de los datos.

- tZX: Nos referiremos con este alias a todos los sucesos en los que se produce un top, un Z y un quark o un W. Entre los productos de estos procesos también encontraremos leptones y jets, por lo que fácilmente reproducirá el estado final de la señal. A diferencia de la señal, sin embargo, aquí en teoría mediremos quarks b (provenientes del top) en el estado final.
- ZZ: Entrarán aquí todos procesos de producción de dos bosones Z, en el caso de que ambos se desintegren a dos leptones. Estos bosones se producirán en general tras la interacción de un quark con su antiquark. Los jets que encontremos en estos sucesos provendrán de la emisión de un gluón por los quarks del estado inicial. Sin embargo, esta emisión supondrá un orden más en el diagrama de Feynman, haciéndola muy improbable frente al estado final sin jets.

Sin embargo, este proceso sí que será importante al tratarlo como fondo en el estudio de la producción WZ *sin jets* que llevaremos a cabo en la sección 4.5. Este hecho se discutirá en detalle en la sección 4.4.1.

- WZ: Entra aquí la producción de un W y un Z que se desintegran posteriormente a tres leptones. Este proceso tiene un estado final muy parecido a la producción VBS de los mismos bosones, por lo que será el mayor de los fondos en cada uno de los histogramas. Por esta razón lo estudiaremos por separado y en detalle en la sección 4.5.
- VVV/VV: Nos referiremos de esta forma a todos los sucesos en los que se produzcan tres o dos bosones (excluyendo ZZ). En esta categoría entrará la producción de WWW, WWZ, WZZ, ZZZ y WH y ZH. Estos procesos constituirán un fondo pequeño tanto en los análisis de WZ como de WZ VBS.
- ttX: Con esto designamos a todos los procesos de producción de dos tops con un Higgs, un Z o un W. Cada top producirá un b en el 100 % de los casos, que en un alto porcentaje de

ellos (dependiente de la eficiencia de la identificación de quarks b) será etiquetado como tal. Seleccionando únicamente sucesos con ningún b -Jet reduciremos este fondo de manera considerable.

- $X+\gamma$: Entran aquí todos los procesos en los que encontremos un fotón en el estado final, en asociación con tops o bosones W y Z . Recordemos que un fotón muy energético puede producir pares de leptones del mismo modo que lo hace un Z . Si X es un Z , este proceso podría imitar el estado final de la región de señal en los casos en los que uno de los leptones del fotón no se detecte. Si X es un W , este proceso podrá reproducir la señal cuando los dos leptones del fotón se midan en el detector.

Como ya se ha señalado, el hecho de que no incluyamos ningún otro Monte Carlo no significa que no existan más fondos del proceso que se está estudiando. Pero aun habiéndolos, estos serán totalmente despreciables. En las secciones posteriores veremos que, de hecho, muchos de los procesos a los que nos hemos referido en esta sección constituirán también una parte minúscula del total de los fondos una vez hayamos aplicado todos los criterios de selección para aislar la señal.

Cada uno de los sucesos de Monte Carlo que llenan los histogramas de la figura [17] está perfectamente caracterizado. Todas las cantidades físicas asociadas a él son conocidas, así como la partícula de la que provienen los objetos del estado final y el proceso físico por el que se han generado.

De los datos, por otro lado, solo se conocen las cantidades físicas asociadas a las partículas del estado final que han sido medidas en el detector.

4.4. Regiones de Control

En física de partículas, el cálculo de la sección eficaz de un proceso (definida según la ecuación (4)) es *independiente* de lo bien o mal modelado que esté el Monte Carlo del proceso de señal. Intuitivamente tiene que ser así, pues el Monte Carlo de señal se ha implementado utilizando una sección eficaz teórica, que puede coincidir o no con nuestras medidas pero no debería influir en ellas.

Esta independencia, sin embargo, ha de tratarse con cuidado. Solo será cierta en los casos en los que los Monte Carlos modelen perfectamente todos los fondos, y por tanto *todo* el remanente entre ellos y los datos pertenezca al proceso de señal.

Además, modelar la señal adecuadamente es importante a la hora de analizar *la forma* de las distribuciones de cada variable del suceso. En el cálculo de la sección eficaz según la hemos definido en la sección 3.1 se tiene en cuenta únicamente el número de sucesos de datos y fondo, pero no la forma de sus distribuciones. Comprobar que la señal cubre razonablemente bien el espacio entre datos y fondos es por tanto necesario cuando se calcula una sección eficaz de este modo.

En este análisis, sin embargo, calcularemos la *fuerza de la señal* (μ) en lugar de la sección eficaz para cada proceso de señal (WZ y VBS). El Monte Carlo de señal jugará por tanto un papel importante, pues daremos los resultados en función de la sección eficaz con la que este se ha modelado. El valor de la fuerza de la señal dará el número por el que ha de multiplicarse la sección eficaz del MC de señal para adecuarse a la realidad impuesta por los datos. Su indeterminación dará la incertidumbre de esta *normalización*. Hasta que no calculemos μ , por tanto, no asociaremos ningún error al Monte Carlo de señal, con lo que toda la indeterminación sin origen estadístico provendrá de los fondos.

Por este motivo, se aplicarán criterios de selección tanto a datos como a predicciones de los procesos de fondo con el objetivo de definir una región en la que la contribución de los fondos (y por tanto de su indeterminación) sea mínima, y el cociente señal/fondos máximo. A esta región nos referiremos en adelante como *región de señal*.

Los fondos, por otro lado, tienen que ser muy bien conocidos y estar, en la medida de lo posible, muy bien controlados. Todas las indeterminaciones en los fondos se propagarán inevitablemente a la fuerza de la señal.

Cuando un Monte Carlo es generado, se tienen en cuenta *todos* los diagramas de Feynman que contribuyen al mismo proceso *hasta un orden finito* en teoría de perturbaciones, según hemos definido *orden* en la sección 1.3. Todos los Monte Carlos que he utilizado en este análisis tienen en cuenta todas las contribuciones a cada proceso *hasta NLO*. Aún no disponemos de Monte Carlos a NNLO, pues el coste computacional que supone generarlos es enorme.

Ya en la sección 1.4 hemos señalado que sumar diagramas a órdenes más allá de NLO implicará una contribución positiva a la sección eficaz teórica implementada en los Monte Carlos que utilizamos en este análisis. Esta contribución provocará que las predicciones se observen en general un poco por debajo de los datos en los histogramas.

Además de este, sin embargo existen multitud de motivos por los que un Monte Carlo podría no corresponderse con la realidad sobre los histogramas. Incluso aun estando bien generado,

podrían haberse cometido errores o descuidos en el análisis que diesen lugar a resultados erróneos.

Por ello, es conveniente definir mediante criterios de selección regiones en las que cada uno de los fondos sea claramente predominante sobre el resto de Monte Carlos. Si en estas regiones los datos se corresponden con el MC, asumiremos que el proceso en cuestión está bien modelado. Si no fuese así, se buscará una explicación y se corregirá el Monte Carlo pertinentemente. Llamaré a las regiones definidas de este modo *regiones de control*.

Cada una de las regiones de control que defina a continuación tendrá que ser completamente independiente a la región de señal. Queremos utilizar las regiones de control para comprobar si realmente *entendemos* cada proceso del fondo y lo estamos modelando bien. Por definición, el proceso de señal es aquel que pretendemos estudiar, y por tanto aquel hacia el que tenemos un conocimiento más escaso.

Si tuviésemos una contaminación significativa de sucesos de señal en las regiones de control, y observásemos una diferencia apreciable entre datos y Monte Carlo, no podríamos asegurar que esta diferencia se debe a la necesidad de corregir el MC que estamos controlando, y no al desconocimiento sobre el proceso de señal.

En lo que sigue diremos que dos regiones son *ortogonales* cuando un suceso de una de ellas *nunca* pueda pertenecer a la otra. Si definimos la región A exigiendo al suceso (pongamos) una $MET > 30$ GeV, la región B en la que exijamos $MET < 30$ GeV será ortogonal a A, independientemente del resto de criterios de selección implicados en su definición.

Más adelante definiremos dos regiones de señal y tomaremos medidas en ellas para los procesos WZ y WZ VBS. Estos dos procesos tienen un estado final muy similar, y comparten por tanto los mismos fondos.

El mayor fondo en la región de señal de la producción WZ vendrá dado por el proceso ZZ. El estado final de WZ son tres leptones (dos del Z y uno del W) y energía faltante (del neutrino del W). Por tanto en la producción ZZ, todos los sucesos en los que uno de los cuatro leptones no se mida (y se cuente por tanto como MET), contribuirán a este mismo estado final.

Para WZ VBS, el mayor fondo es indiscutiblemente, la producción WZ, siempre y cuando se emitan dos jets durante el proceso. La emisión de dos gluones es relativamente improbable, pero la sección eficaz de WZ es cuatro órdenes de magnitud superior a la de VBS. Esto significa que por cada suceso de VBS se producirán del orden de 10000 sucesos de WZ, y por tanto, por mínima que sea, la contribución de WZ + 2 jets se amplificará considerablemente en la región

de señal de VBS.

Además, todos los procesos que involucren quarks top, a los que en los histogramas me refiero como “tZX” y “ttX”, serán el otro de los fondos no despreciables de WZ VBS. En ambos casos, el top se desintegrará a un b (b-Jet) y a un W (leptón y MET). El jet y los dos leptones restantes provendrán fácilmente de las contribuciones ZX y tX (siendo X, recordemos, un quark, un Z, un Higgs o un W).

A definir la región de señal para WZ hemos dedicado enteramente la sección 4.5. A continuación estudiaremos las regiones de control de los otros dos fondos principales: el proceso ZZ y los procesos con quarks top.

4.4.1. La CR del proceso de producción de dos bosones Z

El estado final de dos bosones Z tiene unas características muy distinguibles e identificables cuando los dos Z caen a leptones, por lo que este es un proceso relativamente fácil de aislar.

Para aislar el proceso ZZ, escogeremos sucesos con exactamente cuatro leptones en el estado final.

En cuanto al p_T de los leptones, pediremos un mínimo de 25 GeV a los más energéticos de cada Z y 10 GeV a los menos energéticos

Este criterio de selección no solapa con las regiones de señal que construiremos más adelante, donde pediremos tres leptones o más, siempre que el cuarto *no supere* los 10 GeV.

Exigiremos también que la masa invariante del par de leptones de cada Z diste como máximo 15 GeV de la masa nominal del Z, aumentando con ello considerablemente la probabilidad de que realmente estemos reconstruyendo dos Z reales.

El resto de criterios de selección coinciden con los que pediremos más adelante para aislar el proceso de producción WZ.

Cuatro de los histogramas que construimos tras aplicar todos los criterios de selección se muestran en la figura [18].

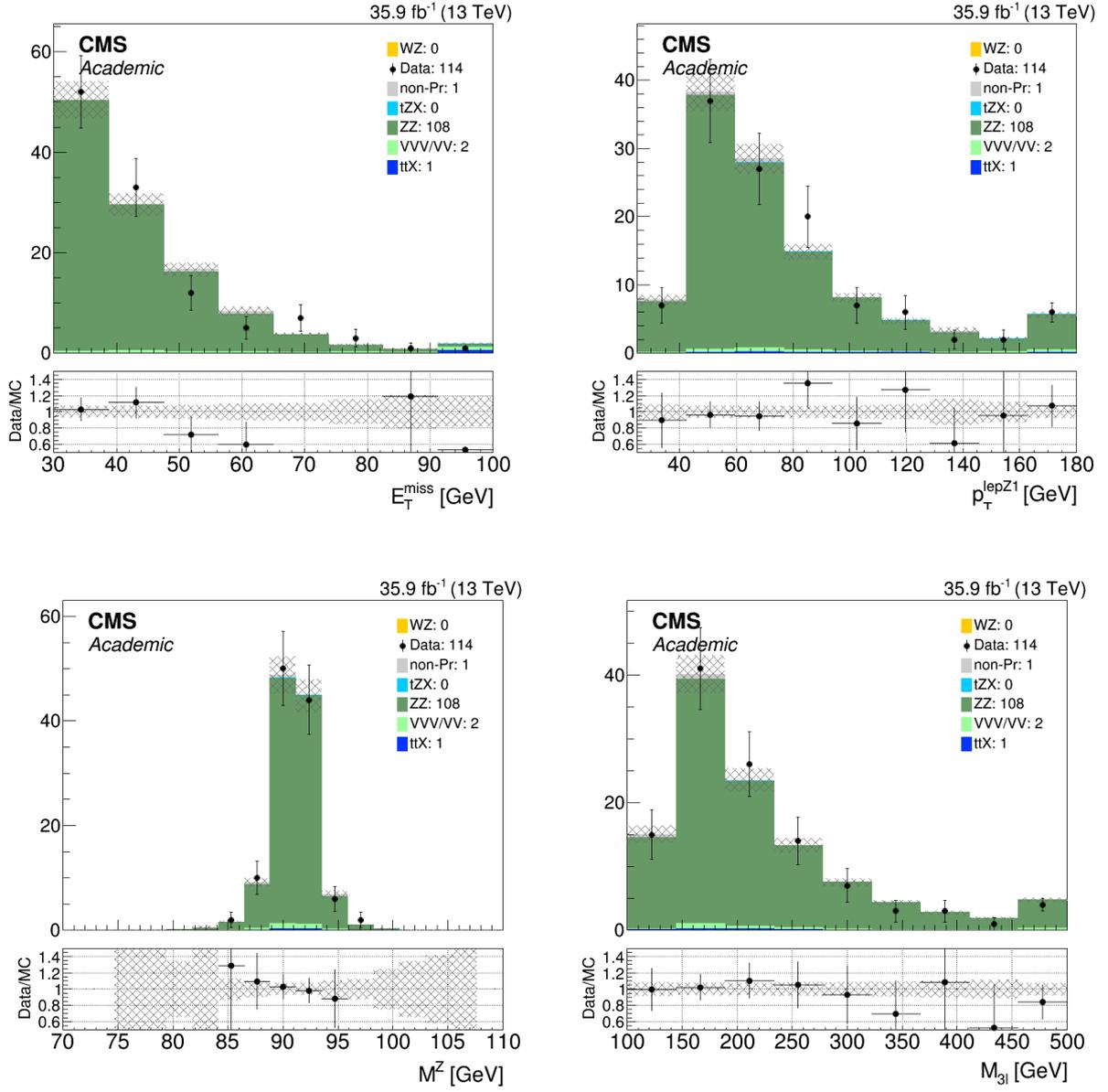


Figura 18: De arriba a abajo y de izquierda a derecha: La energía faltante en el plano transverso, el momento transverso del leptón más energético de cada Z, la masa invariante del par de leptones de los Z y la masa invariante de los tres leptones del estado final.

Los histogramas de la figura [18] muestran que la producción ZZ en el canal tetraleptónico es un proceso muy fácil de aislar. El detector CMS identifica y caracteriza leptones mucho más eficientemente que jets (a demás de medirse mejor, los leptones se producen con mucha menos frecuencia que los jets), y por tanto es mucho más cómodo trabajar con análisis en los que las partículas se desintegren leptónicamente.

El Monte Carlo ZZ representa el 89% del total de los sucesos en los histogramas de la figura [18], y el hecho de que reproduzca bien el comportamiento de los datos en esta región de control nos garantiza que podremos fiarnos de él cuando lo tengamos en cuenta en al región de señal.

Concretamente, a esta región de control han entrado 181 sucesos de datos y 182 de fondos, de los cuales el 89% corresponde al Monte Carlo ZZ. Estos dos números son perfectamente compatibles dentro de la incertidumbre estadística, por lo que en adelante asumiremos que la producción ZZ está bien modelada por MC.

4.4.2. La CR de procesos con quarks top

Los procesos con tops van a constituir un fondo importante cuando estudiemos la SR de WZ VBS, pues igual que la señal, estos procesos también tienen jets en el estado final. Estos procesos tienen estados finales con tres leptones que pasarán los criterios de selección de la región de señal definida para WZ VBS.

En esta ocasión, seleccionaremos únicamente sucesos en los que haya sido identificado un quark b, pues el top se desintegra en el 100% de los casos a un b y a un W. La identificación de un b en el estado final es precisamente la característica que distingue a los procesos con tops de la producción WZ.

Además, a diferencia de nuestra región de señal, excluirémos de este análisis la parte central del pico del Z, imponiendo que el par de leptones identificados como provenientes del Z tenga una masa invariante que diste *como mínimo* (y no como máximo) 15 GeV de la masa nominal del Z. Este criterio de selección es el inverso al que aplicaremos para estudiar ambas regiones de señal, donde a los leptones del Z se los exige computar una masa invariante entre $m_Z - 15$ y $m_Z + 15$ GeV.

Los criterios de selección adicionales son los mismos que utilizaremos en la región de señal de WZ. La definición precisa de todos ellos puede consultarse en la tabla [2].

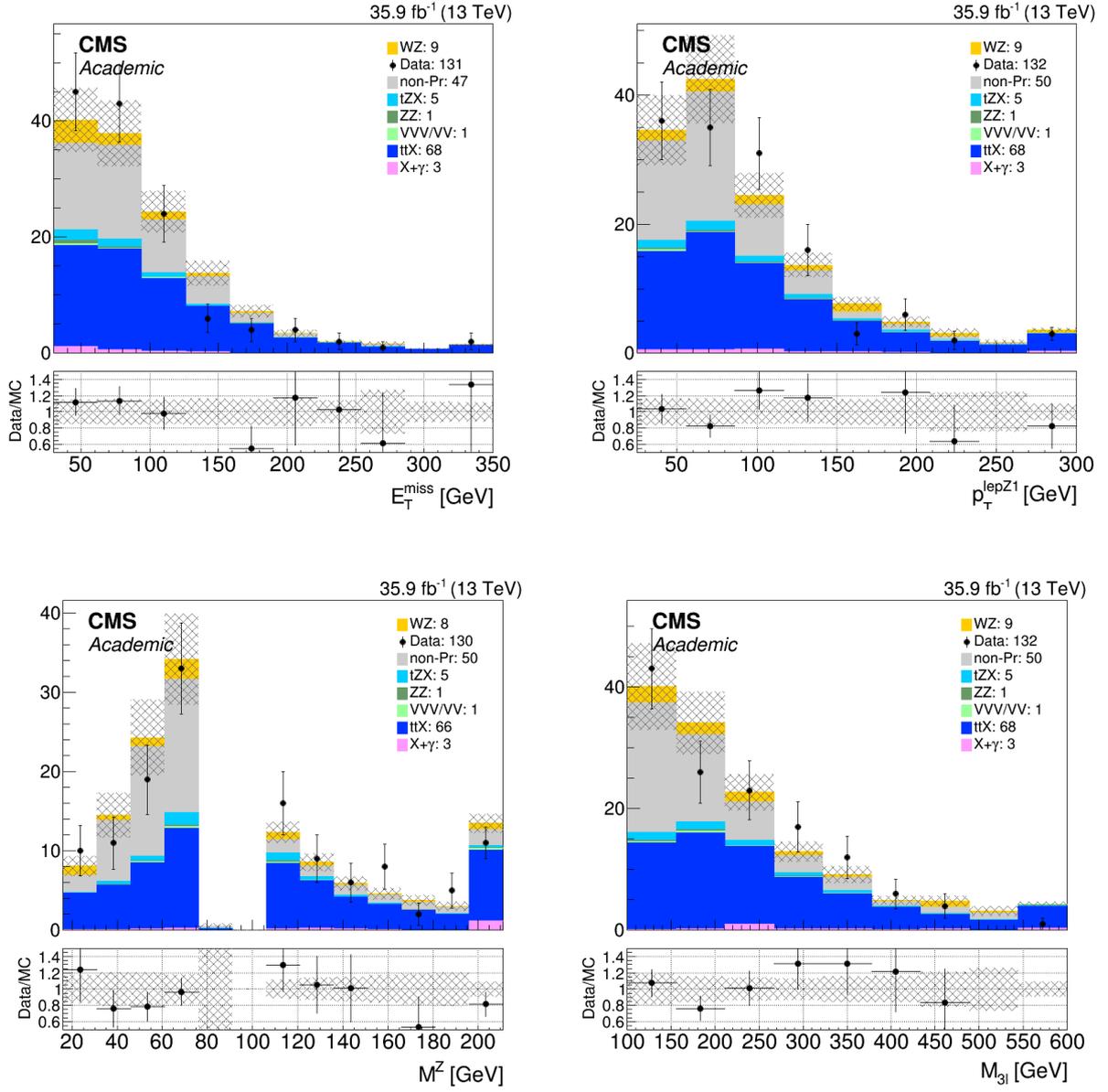


Figura 19: De arriba a abajo y de izquierda a derecha: La energía faltante en el plano transverso, el momento transverso del leptón más energético del Z, la masa invariante del par de leptones del Z y la masa invariante de los tres leptones del estado final. La estadística es bastante reducida en estos histogramas, y la razón principal es haber definido esta CR fuera del pico del Z, donde se concentran gran parte de los sucesos. Pese a ello, no se observa ninguna deriva en los datos que induzca a pensar que haya que aplicarle alguna corrección a los Monte Carols.

Los sucesos Non Prompt (non-Pr) en la figura [19] se estiman a partir de los datos, y podemos asumir por tanto que están bien modelados. El único fondo apreciable de los procesos con tops (azules) es la producción WZ (amarillo). Este proceso es también nuestra señal, que estudiaremos con detenimiento en la sección siguiente.

En total, contamos en esta CR 132 sucesos de datos y 135 de señal+fondos. La discrepancia es pequeña y queda absorbida por las incertidumbres estadísticas, por lo que tampoco aplicaremos correcciones aquí.

Una vez tenemos controlados los fondos relevantes de cada señal vamos a pasar a estudiar cada una de ellas por separado. Comenzaremos definiendo una SR para la producción WZ, pues esta será más tarde el fondo principal de la producción WZ VBS. El recíproco no es cierto. El hecho de que la sección eficaz de WZ VBS sea tan pequeña hace que este sea un proceso despreciable cuando se estudia WZ.

4.5. La región de señal para la producción WZ

Antes de comenzar con la definición de la región de señal, cabe considerar un aspecto importante sobre el Monte Carlo que la modela.

4.5.1. La normalización del Monte Carlo de señal

Ya hemos comentado que actualmente no disponemos de Monte Carlos hasta NNLO. El generador que se ha utilizado para generar WZ no es una excepción, y constituye la implementación de todos los diagramas de Feynman que contribuyen al proceso hasta NLO (*Next to Leading Order*).

Por otro lado, aun no disponiendo todavía de una implementación a NNLO, $\sigma_{teo,NNLO}$ a 13 TeV ya ha sido calculada teóricamente [19] para el proceso WZ, y ha resultado ser en torno a un 18% mayor que la sección eficaz que utiliza (a NLO) POWGEG, el generador de Monte Carlo que modela WZ en este análisis [20]. Esta variación no es para nada despreciable, y constituye un factor de escala que se le aplicará al Monte Carlo de este proceso antes de comenzar a realizar medidas sobre él.

Concretamente, podemos calcular este factor de normalización dividiendo la sección eficaz a NNLO entre la correspondiente a NLO. El resultado es

$$k = \frac{\sigma_{NNLO}}{\sigma_{NLO}} = \frac{49.98 \text{ pb}}{42.5 \text{ pb}} = 1.18 \quad (7)$$

Este tipo de correcciones es muy común en física de partículas. Implementar y procesar un generador a NNLO es muy costoso en términos de tiempo, con lo que frecuentemente se recurre a este tipo de estrategias. Estos factores de escala (1.18 en este caso) suelen denominarse *k-factors*.

4.5.2. Selección de sucesos

A continuación vamos a detallar la selección de sucesos que se aplica en este análisis, con el fin de obtener la mejor relación posible señal / fondo, siendo la señal el proceso de producción WZ, y el fondo el resto de procesos con un estado final similar, que ya hemos descrito.

Cuando un leptón atraviesa el detector, se registran entre otras cantidades su momento, su energía, su carga y su sabor. Por cada suceso se guarda también el número de leptones.

De los bosones W y Z, por otra parte, se mide la información de sus productos de desintegración. Ya hemos comentado que seleccionaremos por su limpieza los canales leptónicos de ambos bosones.

Del W mediremos por tanto muón o un electrón en el estado final. En el caso del Z encontraremos dos leptones del mismo sabor y con cargas opuestas (el Z no tiene carga eléctrica).

Pediremos por tanto, nada más comenzar el análisis un total de tres leptones en el estado final, entre los cuales podamos identificar al menos un par OSSF (*Opposite Sign Same Flavour*, $e\bar{e}$ o $\mu\bar{\mu}$), que posteriormente asociaremos al Z. Pediremos además que cada suceso de Monte Carlo pase el trigger y un filtro que se aplica en función de la energía faltante, de la misma manera que hicieron los datos en el detector.

En la tabla siguiente, desglosamos el total de los sucesos de MC con este mismo estado final según el proceso del que provengan:

datos		11848
señal	WZ	6480
fondos	non Prompt	2632
	ZZ	1407
	X+ γ	606
	ttX	411
	VVV/VV	184
	tZX	155

Tabla 1: Número de sucesos de datos, señal y fondos con al menos tres leptones y un par OSSF en el estado final, una vez renormalizado el Monte Carlo con las correcciones a NNLO

El objetivo principal de nuestro análisis será, atendiendo a la tabla (1), el de minimizar en la medida de lo posible la contribución de cada fondo, manteniendo razonablemente grande el número de sucesos de señal.

Para ello, aplicaremos en lo sucesivo distintos criterios de selección cinemáticos a las variables de cada suceso, partiendo del hecho de que conocemos bien las características físicas de la producción de WZ, que ya han sido analizadas en la sección 4.2. El número total de sucesos *señal + fondo* en la tabla (1) es de 6516, valor que queda ligeramente por encima de los 6391 sucesos que contamos en datos. Resolveremos esta discrepancia al finalizar esta sección.

Siempre que contemos tres o más leptones y encontremos entre ellos un par OSSF ($\mu\bar{\mu}$ o $e\bar{e}$), tendremos que ser capaces de asociar al Z y al W los leptones que los correspondan, para luego aplicar criterios de selección sobre ellos.

Para ello, entre todos los pares OSSF identificados, se le asociará al Z aquel cuya masa invariante (definida en la sección 4.1) sea más cercana a la masa real del Z. De entre los leptones restantes, se le asociará al W aquel que tenga un mayor momento transversal.

Los criterios de selección que aplicaremos a continuación son los mismos que se aplican a la misma región de señal en el análisis expuesto en [15].

El histograma de la figura [20] ha sido generado en este punto inicial del análisis.

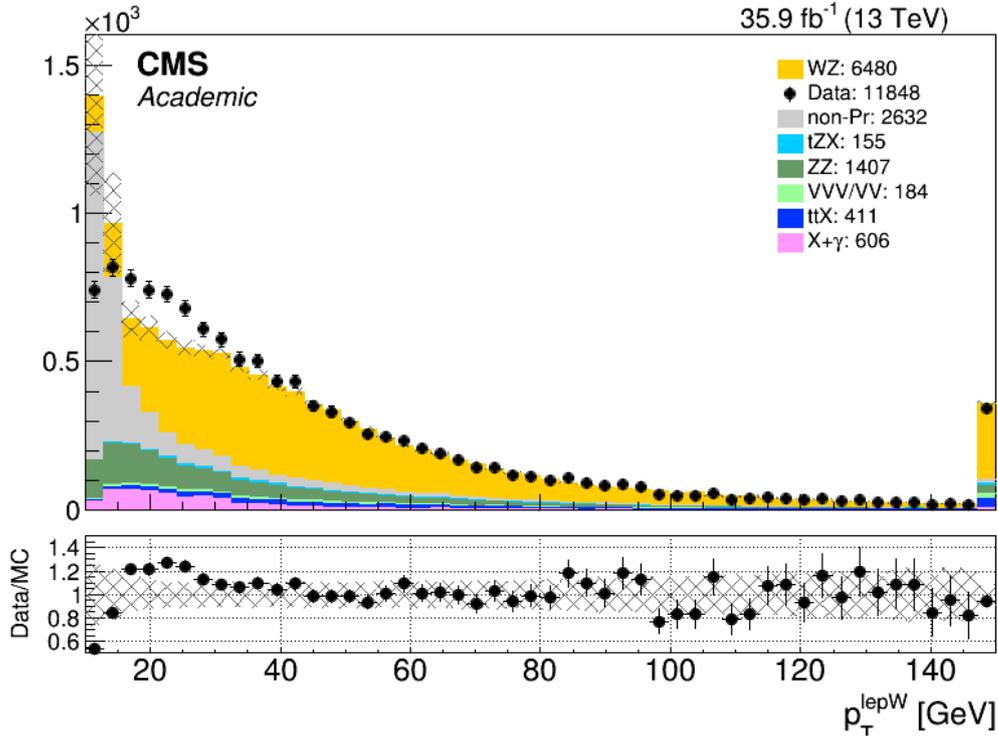


Figura 20: Momento transverso del leptón del W, para todos los sucesos con tres leptones y un par OSSF en el estado final.

El primer objetivo del análisis será eliminar la discrepancia entre datos y predicciones en las primeras columnas del histograma [20].

En este histograma *se han representado los datos* sobre el las predicciones para justificar el criterio de selección en el momento de los leptones, atendiendo a las discrepancias que se observan en las primeras columnas. Sin embargo, en el resto de histogramas que se produzcan a lo largo del análisis hasta el punto final no se representarán los datos. No tiene sentido hacerlo, pues para ellos no conocemos el estado inicial, y por tanto no tenemos forma de saber con absoluta certeza de donde viene cada partícula del estado final.

Esta información sí que se conoce en los Monte Carlos, por lo que los criterios de selección se implementarán, de cara a los histogramas, con el fin de aislar la región amarilla (modelada por el MC del proceso de señal) respecto a todas las demás.

El momento de los leptones del estado final.

Los primeros criterios de selección van a establecer una cota inferior al momento transverso de los leptones del estado final.

Por una parte, con ellos pretendo reducir considerablemente el gran fondo de Non Prompt

que predomina en las primeras columnas de la figura [20]. Por otro lado, excluyendo del análisis las primeras columnas de la figura [20] evitamos trabajar en una región en la que los Monte Carlos modelan considerablemente mal los datos. El desacuerdo datos/predicciones se debe al hecho de la eficiencia del trigger que se utiliza en este análisis decrece considerablemente en la zona de momentos bajos.

Para que un suceso dispare el trigger, y por tanto sea guardado, el estado final tiene que disponer de al menos un leptón (muón o electrón) aislado. En caso de que solo se cuente uno, el suceso será guardado si este es un electrón de más de 27 GeV o un muón de más de 24. En caso de que el estado final disponga de un par de leptones OSSF ($e\bar{e}$ o $\mu\bar{\mu}$), el suceso disparará el trigger si:

- El par es de muones, y los muones más y menos energéticos tienen un momento transversal mínimo de 17 y 8 GeV respectivamente.
- El par es de electrones, y los electrones más y menos energéticos tienen un momento transversal mínimo de 23 y 12 GeV respectivamente.

En caso de que se encuentre en el estado final un par de leptones de diferente sabor, el trigger se activará para sucesos donde estos leptones tengan un mínimo de 23 GeV para el más energético y 8 GeV para el restante.

Todos estos criterios de selección entran dentro de la región de momentos en la figura [20] en la que observamos las desviaciones. En general, un leptón con un momento mucho mayor que el límite establecido por el trigger lo pasará siempre, así como uno mucho menos energético no lo disparará nunca.

Sin embargo, cuando el momento del leptón es similar al límite impuesto por el trigger, la eficiencia del mismo cae considerablemente. Algunos leptones con momento ligeramente menor lo pasarán, y otros por encima del límite no lograrán dispararlo. En estos rangos de energías, emular el trigger del detector en los Monte Carlos se vuelve considerablemente complicado, y se observan por tanto unas discrepancias que impiden hacer análisis en estas zonas del espectro.

Restringir el momento transversal de los tres leptones principales reducirá considerablemente el fondo de Non Prompt y restringirá el análisis a un rango de energías en el que la eficiencia del trigger es próxima al 100 %.

Comenzaremos por tanto el análisis pidiendo un mínimo de 25 GeV al leptón más energético

del Z. Respecto al segundo, nos conformaremos con un mínimo de 15 GeV, que serán suficientes, siempre y cuando este forme un par OSSF con el primero.

Del W solo medimos un leptón, al que imponemos el mismo criterio de selección de calidad que utilizamos para el Z: 25 GeV.

Por supuesto, todos estos criterios de selección podrían hacerse más restrictivos, dejando pasar menos y mejores sucesos, o menos, ganando tanto estadística como incertidumbre.

El resultado de los criterios de selección en el momento de los tres leptones principales se muestra en los histogramas de la figura [21].

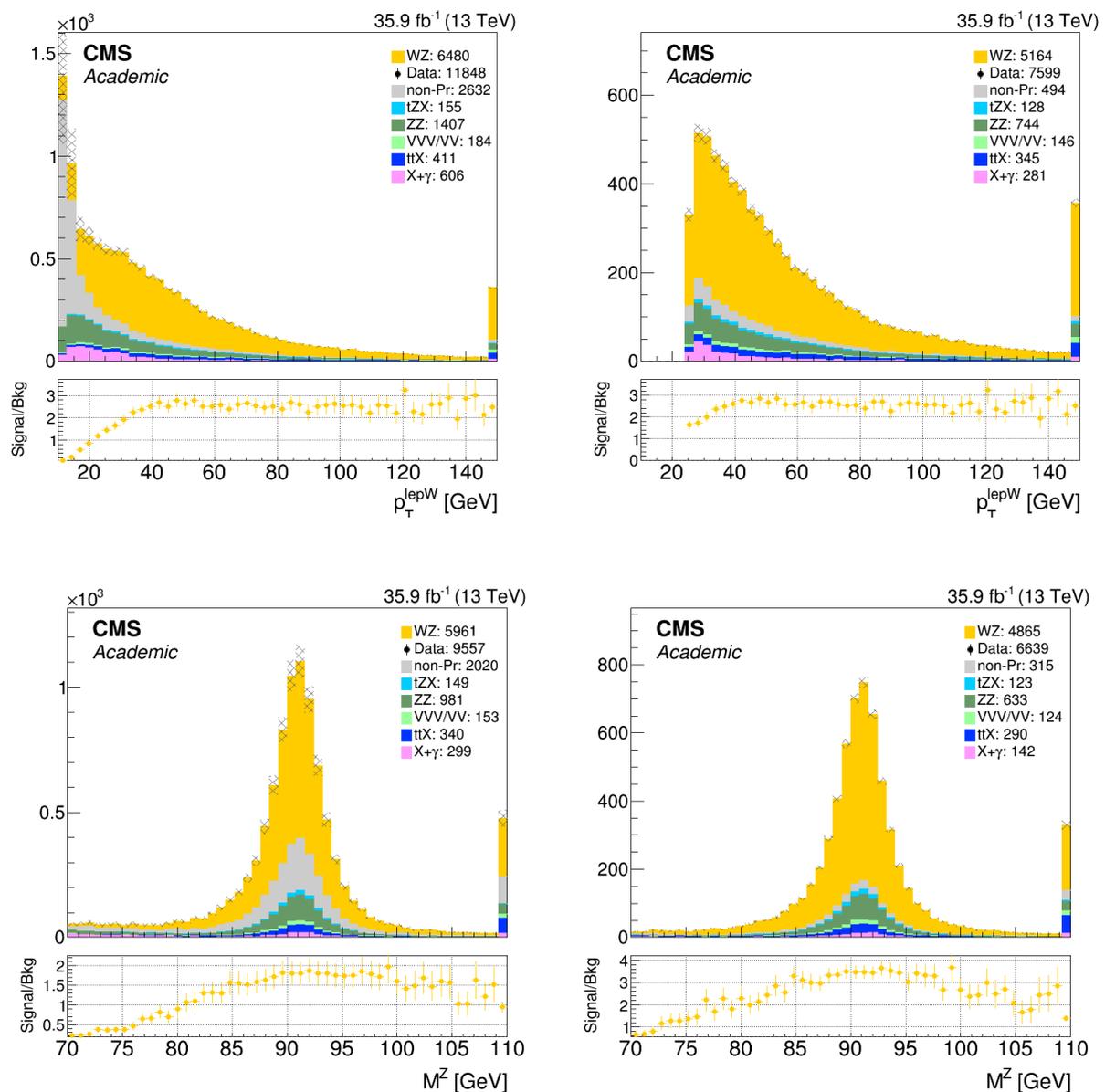


Figura 21: En la fila superior, a la izquierda la distribución del momento transverso del leptón del W antes de aplicar el criterio de selección en su momento transverso. A la derecha, la misma distribución tras el criterio de selección. En la fila inferior, la masa invariante de los dos leptones del Z antes de aplicar el criterio de selección en su p_T (izquierda) y después (derecha).

Con este criterio de selección a sido eliminada una gran parte del fondo de partículas Non Prompt (zona gris en los histogramas). Este fondo se ha reducido en más del 84 %, mientras que de señal apenas hemos perdido el 18 %. Además, hemos excluido con él la zona de energías en la que la eficiencia del trigger es menor.

La masa invariante de los leptones del Z

A continuación, limpiaremos las colas de la distribución del pico del Z (histogramas inferiores en la figura [21]), exigiendo que la masa invariante de los dos leptones que hemos asignado a este bosón se observe dentro del pico del Z, imponiendo que no diste más de 15 GeV de su masa nominal.

A la desintegración del bosón Z le hemos asociado la pareja de dos leptones OSSF con masa invariante más próxima a él. Dos leptones de un suceso que por casualidad tengan carga opuesta y mismo sabor serán asignados al Z *siempre que no haya un par mejor*, independientemente de su masa invariante. Con este último criterio de selección reducimos considerablemente esta incertidumbre y aumentamos las posibilidades de que el par que estamos asignando al Z provenga *realmente* de la desintegración de un Z.

El leptón del W

Estudemos ahora el tercer leptón, al que hemos asignado el W como partícula madre.

Existe la posibilidad de que este leptón no provenga realmente del vértice principal, sino que se dé un suceso de Z a dos leptones (totalmente ajeno a WZ VBS) en el que un fotón radiado por alguno de los leptones del Z produzca un par electrón-positrón, y uno de estos dos nuevos leptones sea lo suficientemente energético como para dejar una deposición apreciable en el calorímetro electromagnético. El otro electrón tendrá una energía por debajo del umbral al que el ECAL puede distinguirlo, y por tanto no se medirá.

La energía mínima que un fotón necesita para producir un par e^+e^- es igual a dos veces la masa del electrón: 1 MeV. el Z tiene una masa de 91 GeV, por lo que los dos leptones que provengan de él tendrán aproximadamente 45 GeV si el Z se desintegra en reposo (más si lo hace en movimiento). Un leptón de 45 GeV tiene energía suficiente para emitir sin problemas un fotón, que en la práctica totalidad de los casos tendrá una energía del orden de los GeV (y recordemos que 1 GeV = 1000 MeV). Todos los fotones emitidos por uno de los leptones del Z tendrán por tanto energía suficiente para formar sin problemas un par e^+e^- .

La característica fundamental de este tipo de *electrones extra* es su baja energía (ahora sí, comparada con la de los leptones que vienen *directamente* del Z). De cara al análisis, la contribución de este tipo de sucesos se reducirá considerablemente aplicando un criterio de selección en la masa invariante mínima de los tres leptones del estado final. De modo teórico, la masa invariante del conjunto de productos provenientes de una partícula X es igual a M_X , independientemente del momento inicial de X y del número de productos del estado final.

En este caso, el tercer leptón proviene del mismo Z que los otros dos leptones, por lo que M_{3l} debería ser del orden de M_Z . La contribución del cuarto leptón es despreciable, pues así lo es su energía si no ha sido siquiera medido en el ECAL.

Si por el contrario (como ocurre en el proceso de señal), el tercer leptón viene del W, la masa invariante de los tres leptones será en general significativamente mayor a M_Z .

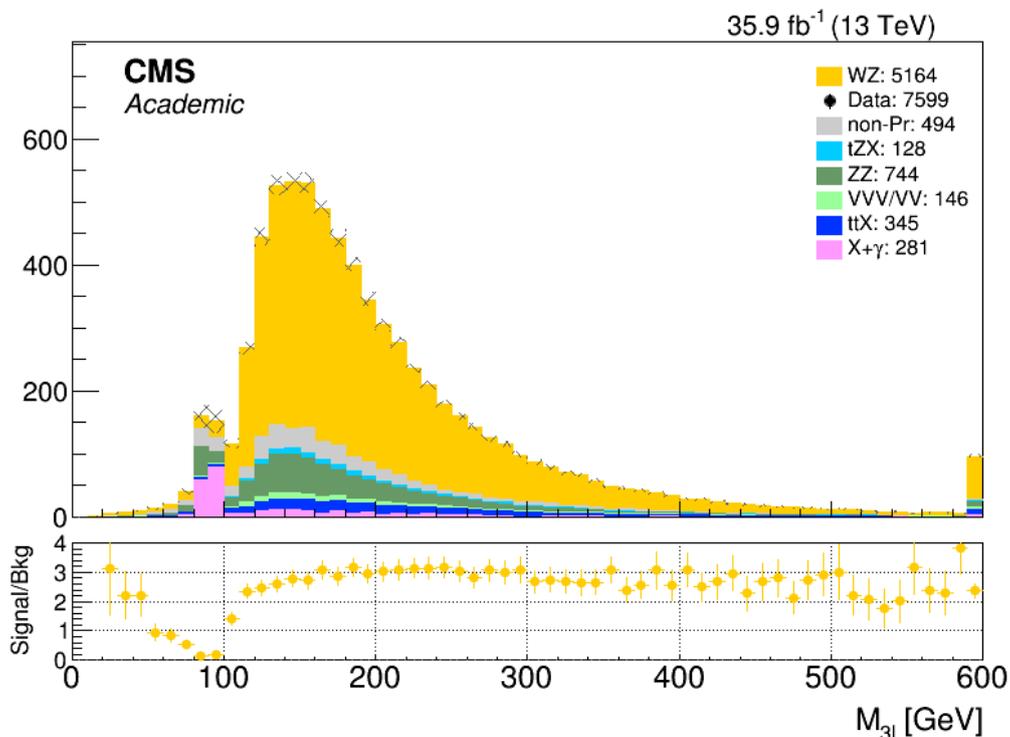


Figura 22: Masa invariante de los tres leptones del estado final una vez se han aplicado los criterios de selección en su momento.

El hecho de que un fotón pueda verse en el detector como un electrón se tiene en cuenta a la hora de generar los Monte Carlos, y es la causa de la aparición de el pico rosa que observamos entre los 80 y 100 GeV aproximadamente en la figura [22].

Por este motivo, seleccionaremos en lo que sigue únicamente sucesos cuya masa invariante de los tres leptones del estado final sea superior a 100 GeV.

La masa mínima de cada par de leptones del suceso

Escogeremos también un límite inferior para la masa invariante de cada pareja de leptones del suceso. Ya hemos restringido los valores de la masa invariante del par del Z, por lo que sabemos que este par tendrá una masa en todos los casos no más de 15 GeV alejada de los 91 del bosón del que proviene.

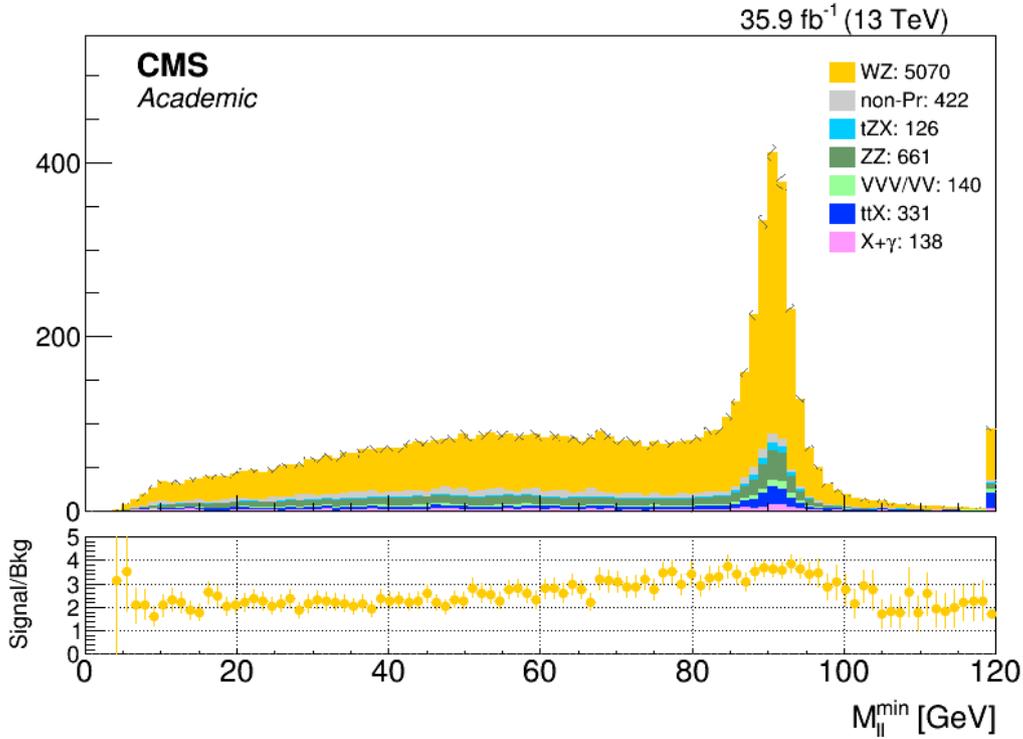


Figura 23: Para cada suceso, la masa invariante del par de leptones que compute la masa invariante mínima, una vez se han aplicado el resto de criterios de selección hasta este punto.

Sin embargo, pediremos además que no pueda formarse con los leptones de cada suceso ningún par con masa invariante inferior a 4 GeV. Este criterio de selección viene principalmente motivado por el hecho de que que el Monte Carlo es menos preciso, y por tanto menos fiable, a bajas energías. Este hecho se deriva de que la teoría de perturbaciones con la que se modelan los procesos QCD pierde precisión a primeros órdenes cuando se aplica a procesos de baja energía.

Criterios de selección adicionales

Atendiendo por otra parte a los productos del W, exigiremos también una energía faltante en el plano transversal mayor a 30 GeV.

La incertidumbre en la MET es, como ya se ha comentado en la sección 2.4, relativamente grande. Es por ello que conviene seleccionar únicamente sucesos con un valor grande en esta

variable, para asegurar en la medida de lo posible que realmente estamos identificando el neutrino del W.

Por último exigiremos, de entre todos los jets del suceso, no encontrar ninguno etiquetado como proveniente de un quark b.

El fondo predominante de cualquier suceso en el LHC son los procesos $b\bar{b}$. Además, muchos otros de gran sección eficaz, como el $t\bar{t}$, también producen b-Jets. Es muy recomendable por tanto asegurar que no encontramos ninguno en una región de señal en la que, de forma teórica, no debería haber.

4.5.3. Resultados en la región de señal WZ

Todos los sucesos que hayan llegado a este punto se considerarán pertenecientes a la región de señal de WZ. En la figura [24] se muestran los histogramas de las variables más significativas en esta región.

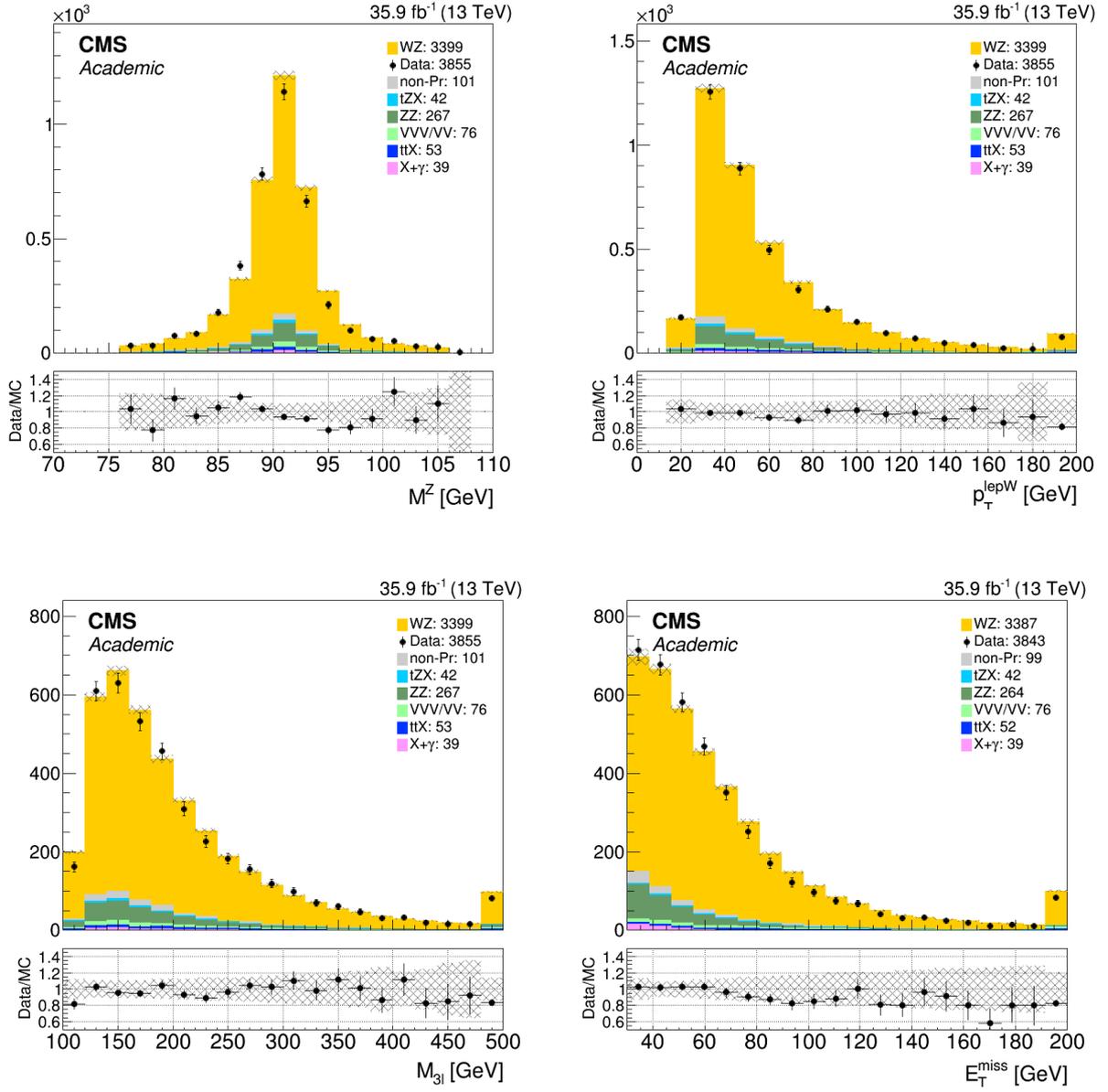


Figura 24: De arriba a abajo y de izquierda a derecha: La masa invariante de los dos leptones del Z, el momento transverso del leptón del W, la masa invariante de los tres leptones del estado final y la energía faltante en el plano transverso. Con la rejilla gris se representa tanto la incertidumbre estadística como la sistemática de los fondos.

Lo primero que observamos en la figura [24] es que, como cabía esperar, la proporción de WZ ha aumentado muy considerablemente sobre la del resto de Monte Carlos. Antes de pasar todos los criterios de selección (tabla [1]), la proporción de sucesos WZ entre sucesos totales de MC era de 0.45. Una vez aplicados los criterios de selección correspondientes a la SR WZ, este valor ha ascendido hasta el 0.82.

Por supuesto, esta limpieza del histograma conlleva una pérdida de estadística. Con todos estos criterios de selección el fondo ha sido reducido un 90 por ciento, aunque la cantidad de señal, el Monte Carlo de WZ, se ha visto también reducida un 48 %.

Se puede observar además que el pico del Z en el histograma superior izquierdo en la figura [24] ha sido modelado en MC ligeramente hacia la derecha respecto a las observaciones en datos (véase la curva que se observa en el pad de ratios). En este trabajo la forma de las distribuciones no jugará ningún papel, y este hecho no será por tanto relevante.

La segunda conclusión, no tan evidente, es que los datos aparecen ahora ligeramente por debajo del MC. El hecho de que el principal fondo de la señal, ZZ, haya sido verificado con su propia región de control indica que esta nueva discrepancia es debida al proceso de señal. Este proceso será medido y estudiado en detalle en la sección 5.4.

En la sección 4.6, sin embargo, se calculará el valor central de la corrección que habremos de aplicar al Monte Carlo WZ antes de utilizarlo como fondo en el estudio de la producción WZ VBS. Todas las incertidumbres de esta corrección serán también calculadas en la sección 5.4.

A modo de resumen, y antes de comenzar a estudiar la región de señal de WZ VBS, se indican en la tabla 2 los criterios de selección aplicados en cada una de las regiones de control que hemos definido, así como en la región de señal de WZ.

Región	N_{leps}	N_{OSSF}	MET [GeV]	b Tags	$ M(l_{Z_1}l_{Z_2}) - m_Z $ [GeV]	$p_T\{l_{Z_1}, l_{Z_2}, l_W, -\}$ [GeV]	$\min(M(l_l'))$ [GeV]	$M(l_{Z_1}l_{Z_2}l_W)$ [GeV]
SR WZ	= 3	≥ 1	> 30	= 0	< 15	> {25, 10, 25}	> 4	> 100
CR tops	= 3	≥ 1	> 30	> 0	> 15	> {25, 10, 25}	> 4	> 100
CR ZZ	= 4	≥ 1	> 30	= 0	< 15	> {25, 10, 25, 10}	> 4	> 100

Tabla 2: Criterios de selección aplicados para definir las regiones de control de tops y ZZ, así como la región de señal de WZ. Estos criterios de selección se corresponden con los aplicados en el análisis referenciado en [15].

4.6. La región de control para la producción WZ

Ya se ha definido la región de señal para este proceso, y se ha concluido que aplicar correcciones a este Monte Carlo va a ser algo necesario. Sin embargo, en el momento en el que vemos este proceso *como un fondo* de otro proceso de señal, nos vemos obligados a definir una región *ortogonal* a la SR de WZ VBS para estudiar las correcciones que se le aplicarán a este Monte Carlo.

Aplicaremos esta ortogonalidad en el número de jets del estado final. El proceso WZ VBS tiene en teoría dos jets en el estado final. Como estos son objetos que se miden con una imprecisión relativamente grande, y por tanto comúnmente se identifican en un suceso más de los que

se han producido en realidad, no suele imponerse una cota superior en el número de jets.

En la sección siguiente impondremos para la producción VBS un mínimo de dos jets en el estado final. Aquí, sin embargo, definiremos la región de control de WZ como aquella en la que, además de aplicar todos los criterios de selección de la sección anterior, imponemos también que el número de jets sea en todos los casos menor o igual que uno. De este modo ningún suceso se encontrará a la vez en esta región de control y en la de señal de VBS.

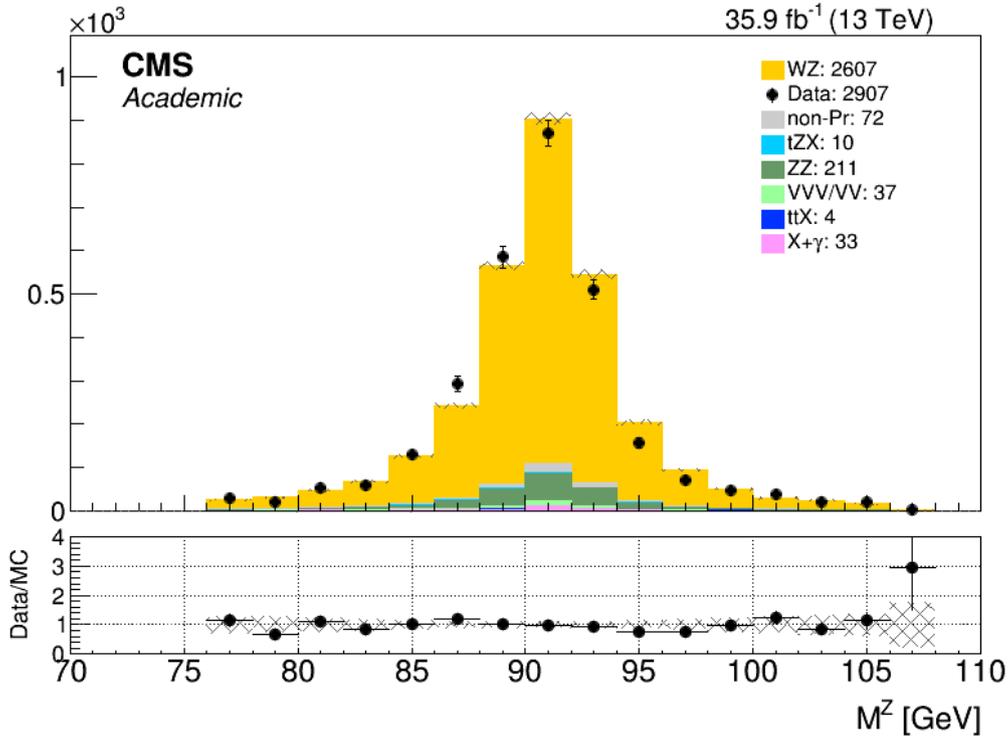


Figura 25: Masa invariante de los leptones del Z en la región de control de la producción WZ.

Por supuesto, el criterio de selección en los jets que define esta CR ha supuesto una pérdida de estadística en esta región respecto a la región de señal. Hemos perdido el 23 % del total de los sucesos de WZ, aunque mayor ha sido la pérdida de fondos, de un 37 %. Estos datos no son relevantes en este punto del análisis, pues en esta ocasión no será necesario calcular ningún error para el k-factor a aplicar a WZ.

Este nuevo *k-factor* tiene como objetivo ajustar la predicción de WZ a la realidad impuesta por los datos en esta región de control. Aunque una vez aplicadas las correcciones a NNLO la discrepancia datos-MC se haya vuelto muy pequeña muy pequeña la tendremos en cuenta en lo sucesivo. Concretamente,

$$k = \frac{N - B}{S} = 0.966$$

Con N el número de sucesos de datos observados en esta región, B el número de sucesos esperados de fondos y S el de sucesos esperados de señal.

4.7. La región de señal para la producción WZ VBS

Una vez entendido y comprobado el modelado de los principales fondos podemos comenzar a definir la región de señal para la producción electrodébil de WZ VBS. Muchos de los criterios de selección que se definirán a continuación son comunes a la SR de WZ, pues en ambos casos encontramos los productos de los dos bosones en el estado final.

4.7.1. Selección de sucesos

En esta sección detallaremos la selección de sucesos desde la misma base de tres leptones que describimos en la sección anterior hasta la definición completa del que ahora es el proceso de señal.

El número de leptones del estado final

Aunque en el estado final teórico de nuestra señal hay exactamente tres leptones, podría contarse en el detector un número superior (provenientes mayoritariamente del interior de un jet).

Estos leptones, que por error han sido identificados como provenientes del vértice principal, tendrán en general un momento transverso relativamente pequeño, y podrán por tanto distinguirse fácilmente.

En vez de seleccionar sucesos con únicamente tres leptones, por tanto, dejaremos entrar en el análisis todos aquellos con tres *o más*, siempre y cuando los leptones extra tengan un momento transverso bajo, nunca superior a 10 GeV. Con esta restricción ganamos algo de estadística frente al criterio de selección estricto en tres leptones que aplicamos al definir la SR para WZ. En esta ocasión la incertidumbre estadística supondrá una contribución mucho más grande que la de muchas otras fuentes de incertidumbres. La estadística por tanto tendrá que ser conservada en la medida de lo posible.

En la figura siguiente se representan todos los sucesos presentes en este punto inicial del análisis.

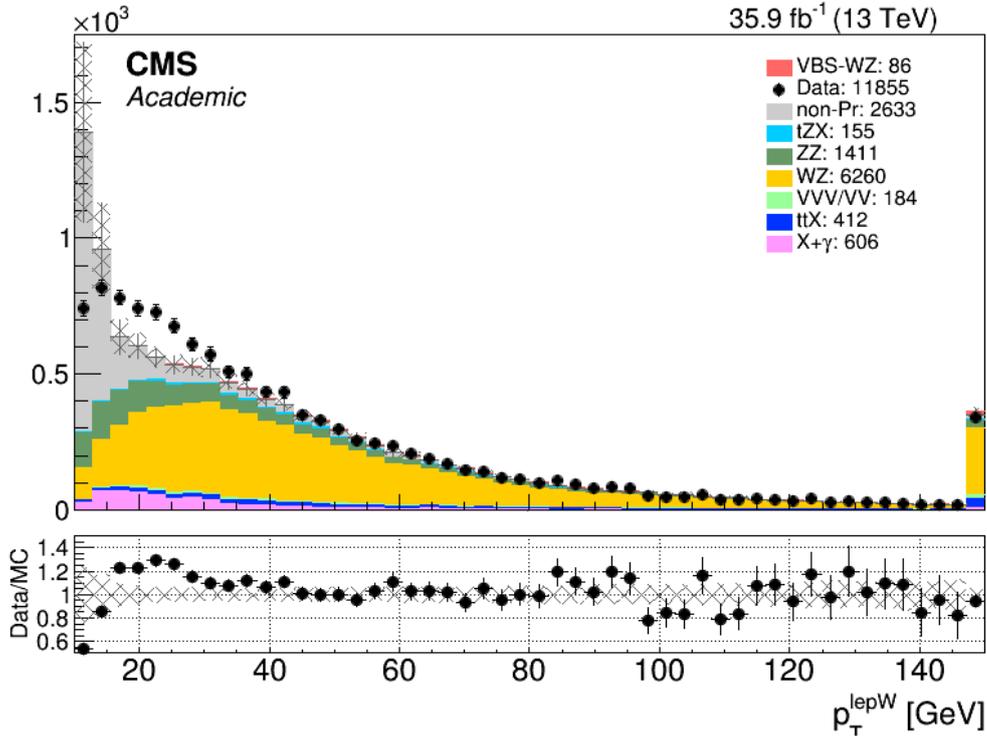


Figura 26: Momento transverso del leptón del W, para todos los sucesos con al menos tres leptones y un par OSSF en el estado final. Si contamos más de tres leptones, los leptones extra tienen un momento transverso menor a 10 GeV.

Los 86 sucesos de WZ VBS que contamos en este punto del análisis apenas representan el 0.7% del total de cuentas. Este valor mejorará significativamente tras aplicar todos los criterios de selección.

Por supuesto, podríamos haber tenido en cuenta también la producción WZ VBS en los histogramas de la sección anterior. No lo hemos hecho porque, como se observa en la figura [26], los sucesos WZ VBS representan una parte despreciable de del total, y quedan totalmente absorbidos por las incertidumbres de los procesos más grandes, representadas con una rejilla gris sobre cada columna.

En todos los histogramas de esta sección, el k factor de 0.966 ya ha sido aplicado sobre la corrección a NNLO del Monte Carlo WZ.

El momento de los leptones del estado final

La motivación de este criterio de selección es la misma por la que aplicamos criterios de selección en esta cantidad al definir la región de señal para WZ. En este caso, la discrepancia datos-MC en las primeras columnas de [26] ya ha sido ligeramente atenuada con el criterio de selección en el excedente de leptones, aunque aún necesita de este criterio para eliminarse por

completo.

Comenzaremos por tanto exigiendo a los dos leptones del Z un momento transverso mayor que 25 GeV para el de mayor p_T y 10 GeV para restante. Recuérdese que el momento transverso es el módulo de la proyección del momento lineal en el plano transverso. Aunque los dos leptones del Z tengan que sumar como mínimo un momento (en módulo) igual a la masa del Z (90 GeV), su proyección en el plano transverso podrá ser arbitrariamente pequeña, por lo que 25 GeV es ya un criterio de selección razonablemente exigente.

Será suficiente con encontrar solo uno de “buena calidad”, siempre y cuando forme un par OSSF con el segundo. Aplicar un criterio de selección en el segundo leptón tan exigente como el primero reduciría considerablemente la estadística del análisis.

Del W solo medimos un leptón, al que imponemos un momento transverso mínimo de 20 GeV.

El resultado de estos criterios de selección es muy similar al que observamos en la figura [21] para la región de señal de WZ. Con él hemos eliminado tanto el fondo de Non Prompt de las primeras columnas como la región de momentos en la que la eficiencia del trigger no permite realizar un análisis preciso.

El leptón del W

Calcularemos también la masa invariante de los tres leptones e impondremos que supere los 100 GeV. Este criterio de selección se implementa en el análisis por la misma razón que lo incluimos al estudiar el proceso de producción WZ. De modo teórico, la masa invariante de los dos leptones del Z vale los 91 GeV que corresponden con la masa del Z. Exigir que este valor aumente un mínimo de 10 GeV al añadir el leptón del W ayuda a asegurar que el tercer leptón viene realmente del W, y no de una fuente menos energética.

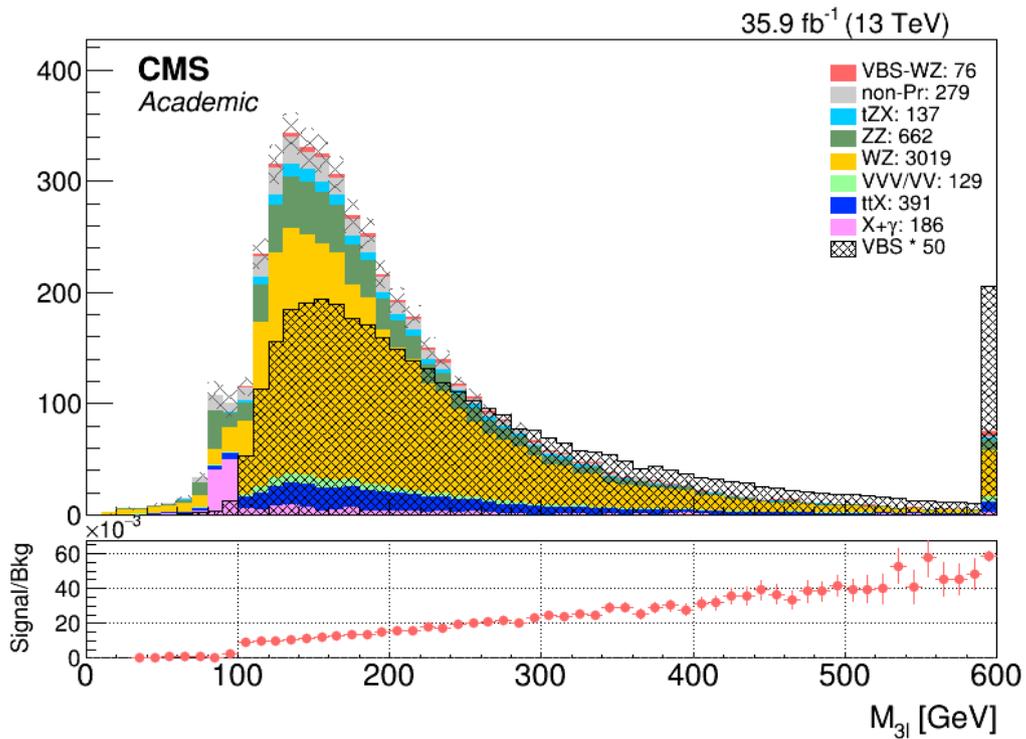


Figura 27: Masa invariante de los tres leptones del estado final una vez aplicado el criterio de selección en su momento transversal. La rejilla negra muestra el Monte Carlo VBS-WZ ampliado un factor 50, con el objetivo de ver de forma clara dónde se concentra la distribución.

En este caso, apenas encontramos sucesos de señal en la región de masas invariantes menores a 100 GeV, con lo que podemos prescindir de ella sin casi perder estadística.

La masa invariante de los leptones del Z

Lo siguiente es exigir para el par OSSF una masa invariante que no diste más de 15 GeV de la masa real del Z, por el mismo motivo en el que nos basamos al definir la SR para WZ.

La masa mínima de cada par de leptones del suceso

La distribución de esta cantidad para el proceso VBS es muy similar a la que observamos para la producción WZ.

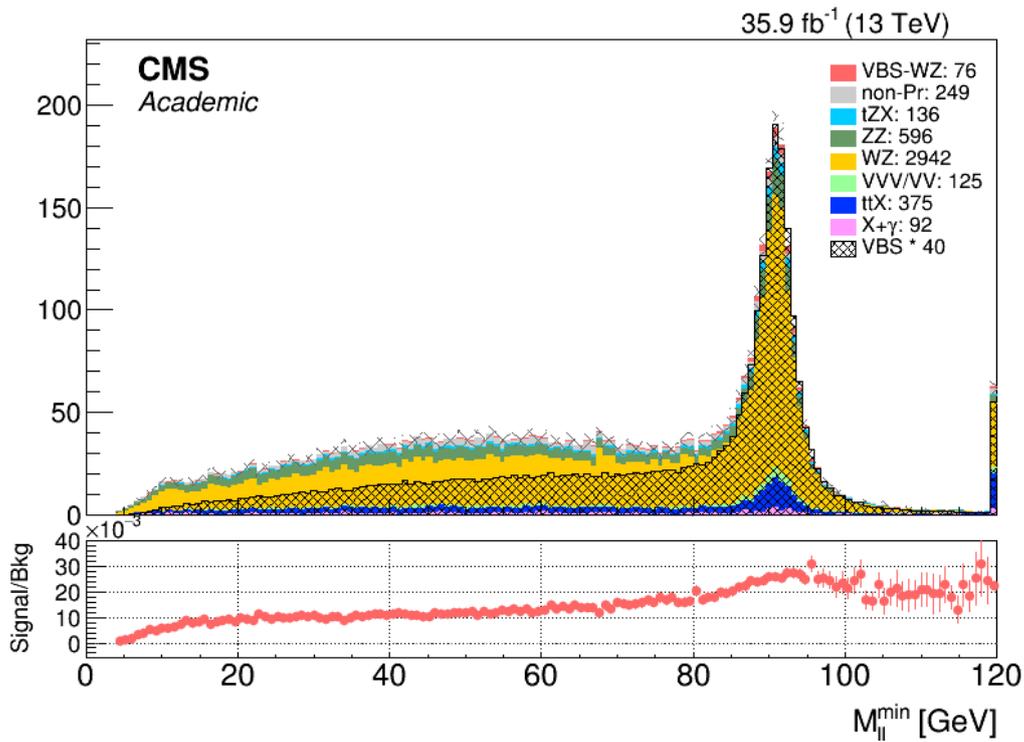


Figura 28: Para cada suceso, la masa invariante del par de leptones que compute la masa invariante mínima. Con una rejilla gris se representa el Monte Carlo del proceso de señal, multiplicado por un factor 40. Todos los criterios de selección hasta este punto del análisis están aplicados en este histograma.

Limitaré inferiormente esta cantidad en 4 GeV, igual que se hizo para WZ. He realizado varios estudios alternativos con criterios de selección más estrictos en esta variable (10 GeV y 20 GeV), y aunque mejore con ellos la relación señal/fondo, pérdida de estadística que supone este cambio introduce en las medidas una incertidumbre demasiado grande.

Limitaremos también la pseudorapidez de los muones y de los electrones. Este criterio de selección no tiene una motivación física, sino que viene inducido por la geometría del detector. En primer lugar, no ha cámaras de muones más allá de $|\eta| = 2.4$, por lo que exigiremos a los muones una pseudorapidez menor que este valor.

El límite del ECAL, está en $|\eta| = 2.4$, y exigiremos por tanto un valor menor para la pseudorapidez de los electrones.

La energía faltante en el plano transversal

Vamos a analizar a continuación la necesidad o no de aplicar un criterio de selección en la energía faltante en el plano transversal. Estamos buscando el estado final en el que el W cae a un leptón y a un neutrino. Entonces, teóricamente en el estado final hay neutrinos, y por tanto energía faltante.

Sin embargo, la presencia de neutrinos no tiene por qué traducirse en energía faltante *en el plano transverso*. Si el leptón del W sale muy paralelo al haz, la proyección de su momento en el plano transverso será muy pequeña, y por tanto lo será también la MET. Recordemos que, para el neutrino,

$$E_{T,\nu} = \sqrt{m_\nu^2 + p_{T,\nu}^2} = p_{T,\nu}$$

que de modo teórico será igual a $p_{T,lepW}$ si no hay más neutrinos en el estado final, como es el caso, y suponiendo que el W se produce con un p_T despreciable.

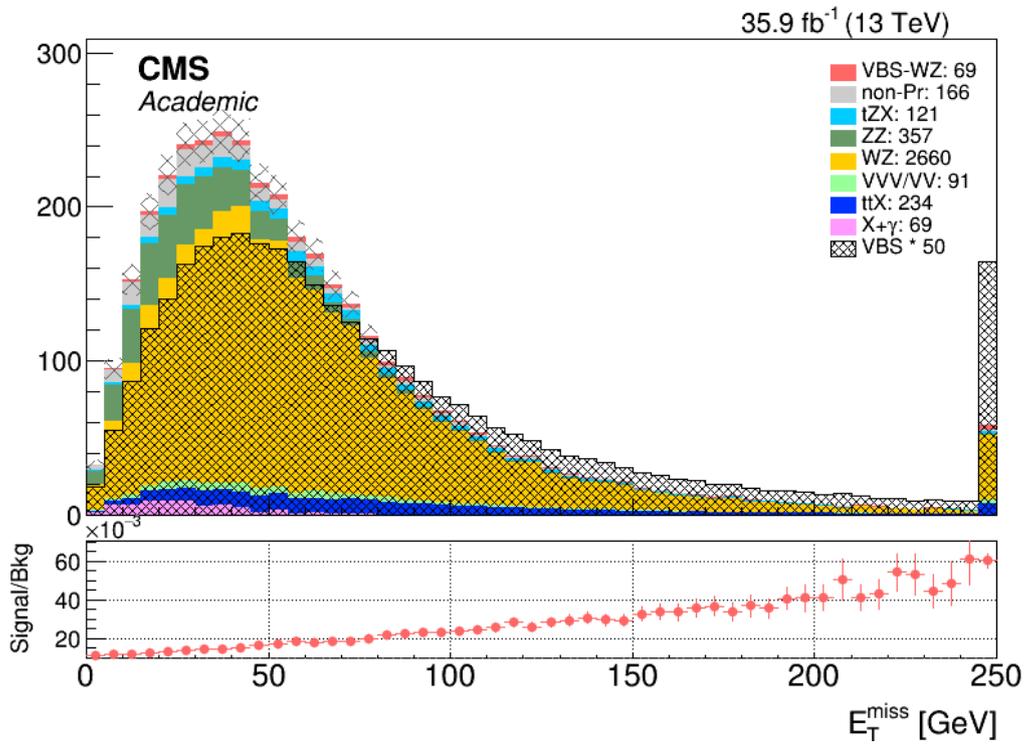


Figura 29: Energía faltante en el plano transverso para todos los procesos modelados por MC. Todos los criterios de selección hasta este punto del análisis están aplicados en este histograma.

La distribución de MET de la figura [29] tampoco propone ningún punto donde seleccionar en esta variable. Como consecuencia, *no aplicaré criterios de selección en la MET*, dejando entrar a la SR cualquier proceso que pase el resto de criterios de selección, independientemente de su energía faltante.

Por una parte, como ya hemos comentado en la sección 2.4, la MET se calcula de cara al detector imponiendo la conservación del momento en el plano transverso. En concreto, la incertidumbre en la energía faltante estará muy relacionada con la incertidumbre en la medida de los jets (mucho mayor que la que conlleva medir muones o electrones). Por tanto, cuanto

menos independiente sea el análisis de cantidades relacionadas con la energía o el momento transverso de los jets, menor será la propagación de su incertidumbre a los cálculos posteriores.

Además, la variación del cociente señal/fondos se mantiene aproximadamente constante en la primera parte del histograma, donde por otra parte se concentra la mayor parte de los sucesos de señal. Un criterio de selección que aumentase este ratio tendría que situarse como mínimo en torno a los 30 - 50 GeV, con la pérdida de estadística y el consiguiente aumento del error estadístico que ello supone.

No imponer ningún criterio de selección en esta variable trae consigo por lo tanto una importante reducción final de las incertidumbres. En la sección 5.5 comprobaremos que, de hecho, las mayores contribuciones a la incertidumbre total de las medidas vienen dadas por la incertidumbre de los jets y la estadística.

El momento transverso de los jets

Ya en la sección 4.2 se ha propuesto para los jets del estado final de nuestra señal una energía relativamente grande. Sin embargo, estos jets se producirán muy pegados al haz, y por tanto gran parte de su momento recaerá en la componente z .

Como el análisis no puede depender de variables medidas en este eje, traduciremos este criterio de selección en la energía a criterios de selección en el momento transverso de los jets (que es conocido con relativa precisión) y en su masa invariante (que no depende del momento del estado inicial).

Comencemos estudiando su momento transverso. Como ya avanzamos en la sección 4.2, esta variable es probablemente la más distintiva entre nuestra señal y los procesos QCD con el mismo estado final.

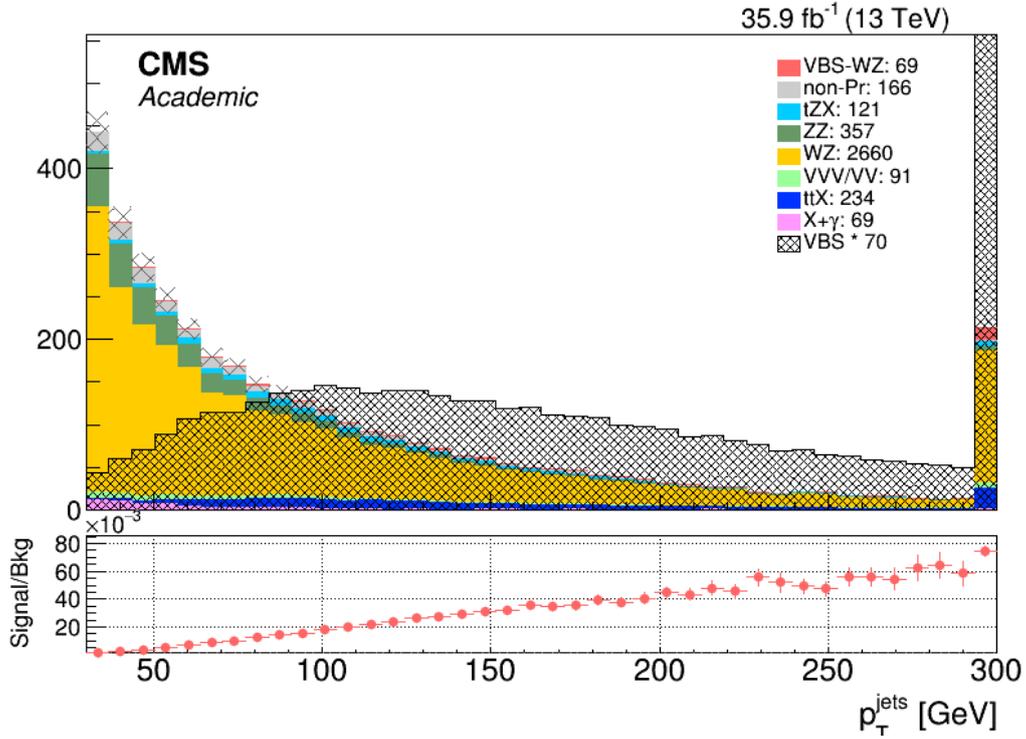


Figura 30: Momento transverso de los jets para cada proceso. Todos los criterios de selección hasta este punto del análisis están aplicados en este histograma.

Con el histograma de la figura [30] comprobamos lo que ya adelantábamos en la sección 4.2. La distribución del p_T de procesos electrodébiles como VBS, o ttX , donde los jets son quarks (dos quarks b en el segundo caso) tiene un máximo sobre los 100 GeV. Sin embargo, para procesos QCD como WZ o ZZ, en los que los jets son gluones, el máximo de la distribución se encuentra en 0, y esta cae uniformemente con p_T .

Para separar los procesos QCD de los electrodébiles (en los que estoy interesado) seleccionaré en la figura [30] sólo sucesos con un p_T superior a 80 GeV.

La pseudorapidez de los jets

Igual que en el momento transverso, se aplicará también un criterio de selección en la pseudorapidez de los jets. Ya se ha justificado que esta será relativamente alta. Su distribución, en comparación con la del resto de predicciones, se muestra en la figura [31]

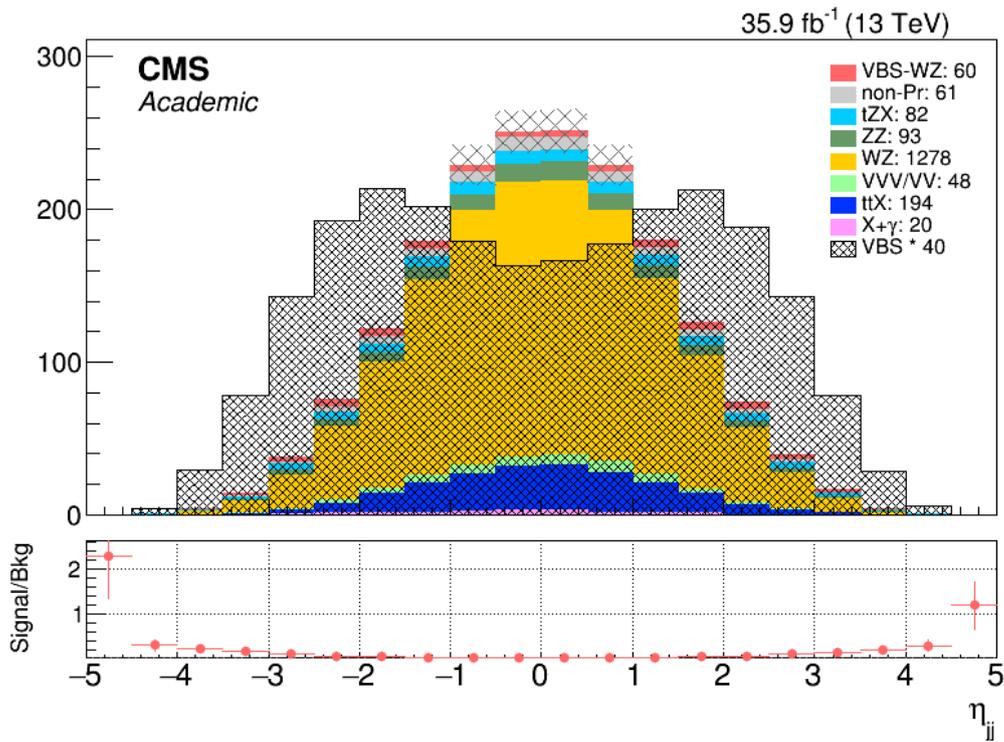


Figura 31: Pseudorapidez de los jets para cada predicci3n. Con la rejilla negra se representa la regi3n de se1al (histograma rojo) amplificada un factor 60.

El hecho de que la pseudorapidez de los jets en VBS no tenga el m1ximo en cero es una característica que no comparte su principal fondo (WZ). Aplicar un criterio de selecci3n en esta variable es por tanto algo necesario si nuestro objetivo es maximizar el cociente se1al/fondo. Este, junto con el criterio de selecci3n en la distribuci3n del momento transverso de los jets y en su masa invariante, es uno de los criterios de selecci3n que distinguen con m1s claridad a la producci3n WZ VBS electrod3bil de todos sus fondos.

Sin embargo, el criterio de selecci3n en la pseudorapidez no se aplicar1 directamente sobre la variable η , pues esto supondr1a una importante p3rdida de estadística. Los picos en la figura [31] se encuentran muy pr3ximos, y el solape en la zona central es demasiado grande.

Recordemos que cada suceso de VBS tiene dos jets en el estado final. Cada uno de ellos proviene de uno de los protones involucrados en la colisi3n, con lo que previsiblemente, aunque ambos salgan considerablemente desviados respecto a su trayectoria inicial, conservaran el signo de su velocidad inicial en el eje z.

En otras palabras, los jets del estado final tendran en general valores de η de signo contrario, y perteneceran por tanto uno a cada pico en la figura [31].

Aplicaremos entonces un criterio de selecci3n en la diferencia en pseudorapidez $\Delta\eta$ entre los

dos jets del estado final. Un valor razonable para este criterio de selección, con el que trabajaremos en este análisis, es $\Delta\eta \geq 2.5$.

La masa invariante de los dos jets del estado final

Ya se ha comentado que restringir la energía del sistema de jets es equivalente, en nuestro caso, a aplicar criterios de selección tanto en su momento transversal como en su masa invariante.

Ya hemos analizado el momento transversal de estos objetos, y a continuación aplicaremos el criterio de selección más conveniente en su masa invariante.

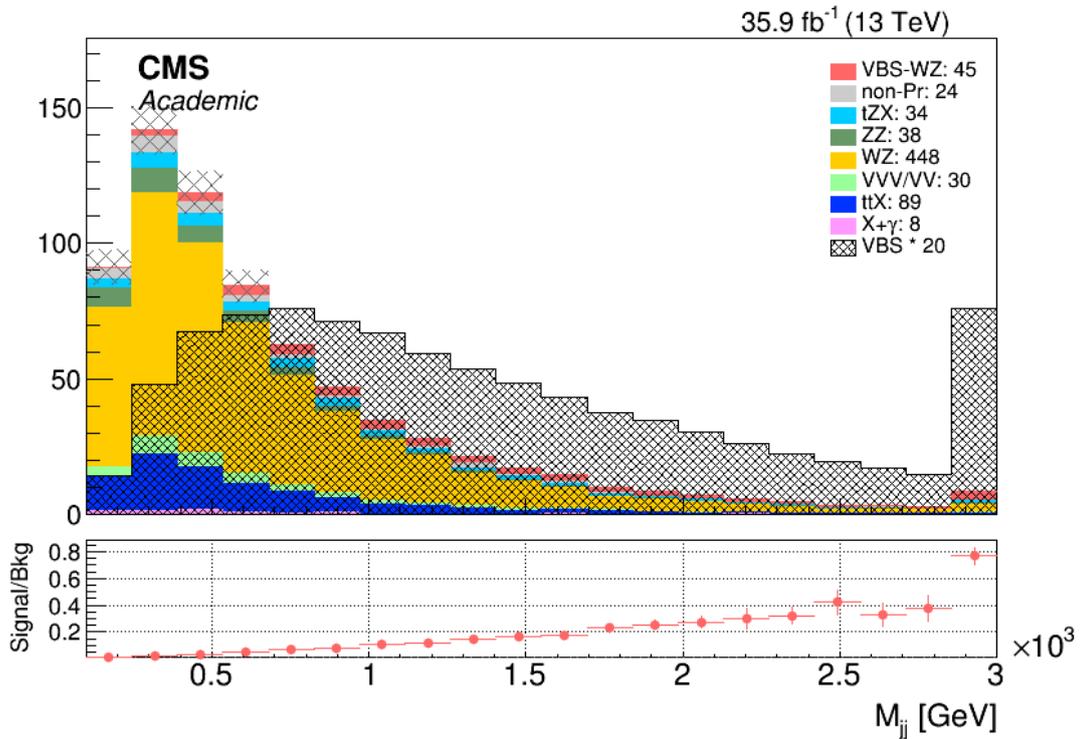


Figura 32: Masa invariante de los jets del estado final. Todos los criterios de selección hasta este punto del análisis están aplicados en este histograma.

Los jets del proceso de señal, como ya avanzamos anteriormente, son en promedio más energéticos que los de sus fondos principales. Basándonos en la distribución de los Monte Carlos de la figura [32], seleccionaremos para nuestra región de señal únicamente sucesos que cumplan la condición $M_{jj} > 500$ GeV.

La variable de Zeppenfeld

Como último estudio relacionado la pseudorapidez de los jets, voy a intentar aplicar otro criterio de selección en η con el fin de asegurar en la medida de lo posible que la pareja de jets del estado final de WZ VBS tiene su origen en el mismo proceso físico que ha dado lugar a los

tres leptones del W y del Z. Por primera vez en el análisis relacionaremos variables de leptones y jets.

Para entender este criterio de selección, comencemos dividiendo los productos del suceso en jets y leptones.

Los leptones provienen de dos bosones que se forman en el mismo punto. Los quarks que originan estos bosones conservarán en el estado final la mayor parte de su energía inicial, y los bosones W y Z se formarán con un momento muy bajo, prácticamente en reposo. Los leptones que provengan de estos bosones tendrán por tanto una energía generalmente mucho menor a la de los jets.

Además, el hecho de formarse prácticamente en reposo implica que, en promedio, los dos leptones del Z y el leptón y el neutrino del W se producirán aproximadamente en sentidos opuestos, y su pseudorapidez media será por tanto casi cero.

Si eliminamos de este promedio el neutrino, los tres leptones tendrán una pseudorapidez que, aunque no nula, será en la mayoría de los casos pequeña.

De modo teórico, si asumimos que el Z y el W se generan en reposo,

$$\bar{\eta}_{3l} = \frac{\eta_{lepZ_1} + \eta_{lepZ_2} + \eta_{lepW}}{3} = \frac{\eta_{lepZ_1} - \eta_{lepZ_1} + \eta_{lepW}}{3} = \frac{\eta_{lepW}}{3}$$

Entonces, cabe esperar en el estado final dos jets con una pseudorapidez muy grande y tres leptones que en promedio se producirán entre los dos jets, con una η pequeña.

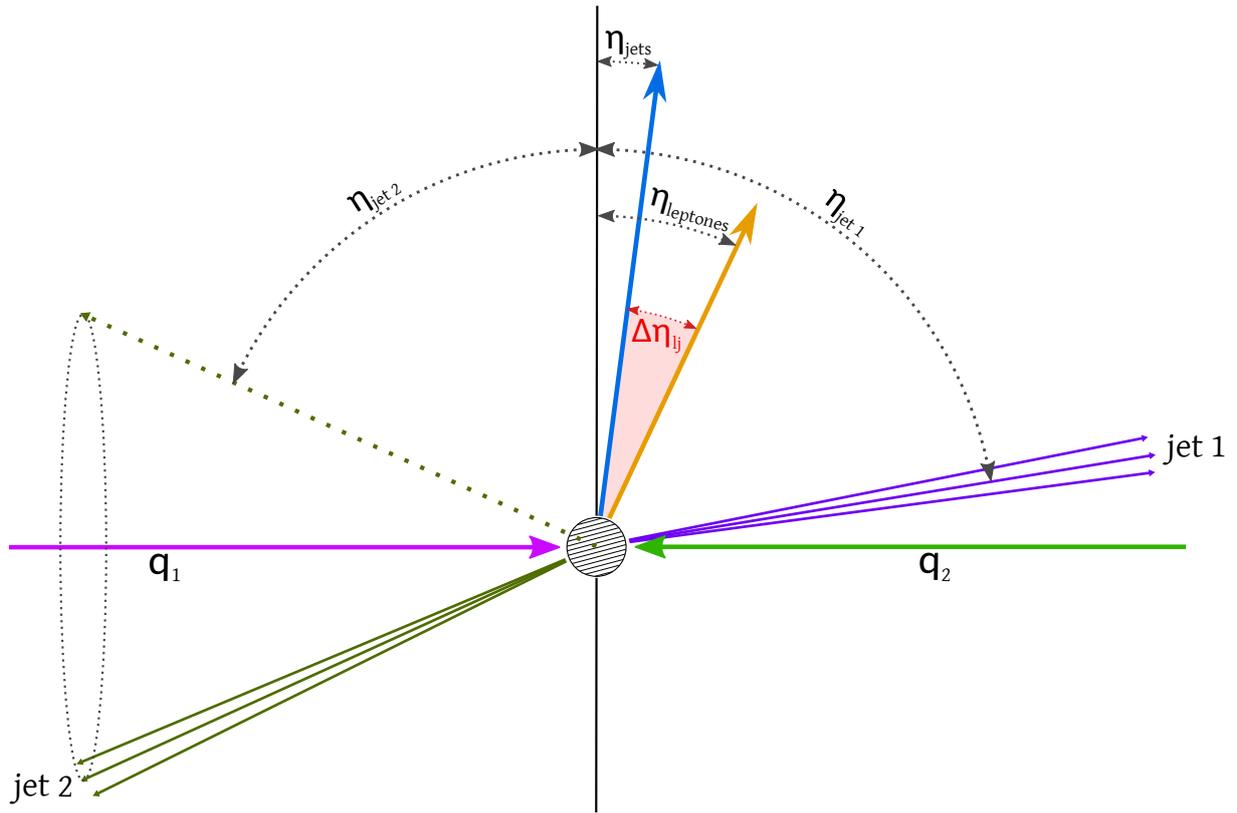


Figura 33: Dos quarks (q_1 y q_2) interactúan en el vértice principal. Ambos quarks emiten un bosón y salen desviados en forma de jets (jet 1 y jet 2). Los dos bosones se desintegran a tres leptones, cuyo momento \vec{p} medio lleva la dirección de la flecha naranja. Con la flecha azul se representa la pseudorapidez media de los dos jets. La separación en η entre los leptones y los jets tendrá un valor $\Delta\eta_j$, que dependerá entre otras cosas del momento del neutrino del W (que de cara al detector es invisible). El momento del neutrino es comparable al de los leptones, y por tanto, en general, su contribución a $\Delta\eta_j$ no será muy grande, y los leptones se producirán aproximadamente entre los dos jets. *Figura propia.*

En la práctica, la variable $\Delta\eta_j$ se define como

$$\Delta\eta_j = \left| \eta_{3l} - \frac{1}{2}(\eta_{j1} + \eta_{j2}) \right|$$

Esta variable, originalmente propuesta en 1996 por D.Zeppenfeld[21] se denomina “variable de Zeppenfeld”, y su uso está muy extendido en análisis VBS y VBF, donde se estudió por primera vez. Representaremos la distribución de esta variable para encontrar la forma de separar aún más el proceso VBS (electrodébil) de su principal fondo, QCD. Esta distribución antes de aplicar criterios de selección en ella es la que se muestra en la figura [34].

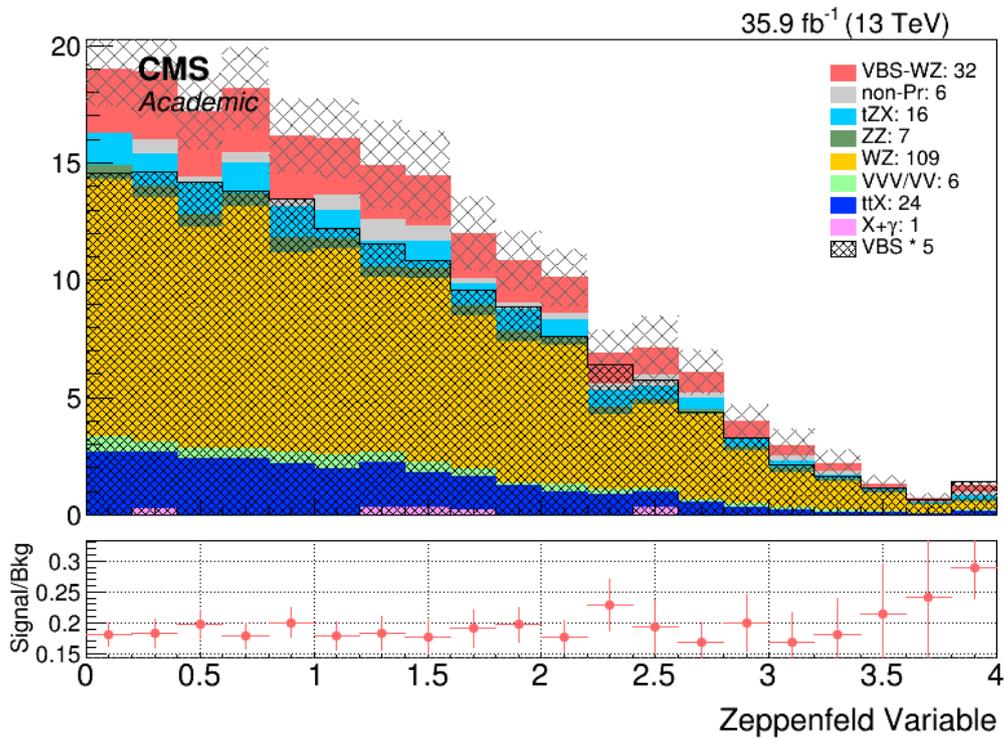


Figura 34: Distribución de la variable de Zeppenfeld. Todos los criterios de selección hasta este punto del análisis están aplicados en este histograma.

Como podemos comprobar, a este nivel del análisis esta variable no permite aplicar en ella ningún criterio de selección. La variable de Zeppenfeld no es independiente de la pseudorapidez de los jets del estado final, sobre la que ya hemos aplicado criterios de selección en la sección 4.7.1. Con el histograma de la figura [34] no obtenemos información nueva, y aplicar un criterio de selección en él reduciría la estadística sin mejorar la relación señal/fondo.

Aun siendo esta una variable típica en análisis de este tipo, en este estudio concreto no resulta de utilidad aplicar en ella criterios de selección. No restringiremos por tanto los valores de la variable de Zeppenfeld.

El criterio de selección en el número de quarks b

En este punto del análisis, los principales fondos de nuestra región de señal son WZ y ttX. El único modo de reducir el fondo de WZ es restringir los criterios de selección en alguna de las variables que ya hemos estudiado. Sin embargo, el fondo de ttX puede eliminarse de modo mucho más sencillo.

El quark top se desintegra en el 100% de los casos a un b y un W. De modo teórico, por tanto, encontraremos siempre un b en el estado final de procesos que involucren tops. Teóricamente, por tanto, los procesos ttX tendrán dos b -Jets en el estado final.

Sin embargo, en nuestro proceso de señal la presencia de quarks b es altamente improbable. Los quarks del estado inicial en la producción VBS serán con una muy alta probabilidad quarks de valencia: u o d . El quark b es el más pesado de entre todos los que podemos encontrar dentro de un protón, y por tanto la probabilidad de tenerlo en el estado inicial de una colisión es realmente baja. En general estos quarks se producen en el LHC por decaimiento de tops o en procesos $q\bar{q} \rightarrow Z \rightarrow b\bar{b}$, donde $q\bar{q}$ son también quarks ligeros.

Dado que, como ya hemos descrito en la sección 2.3, es posible saber si un jet proviene o no de un quark b, podemos conocer el número de bs en cada suceso, y construir con él el siguiente histograma:

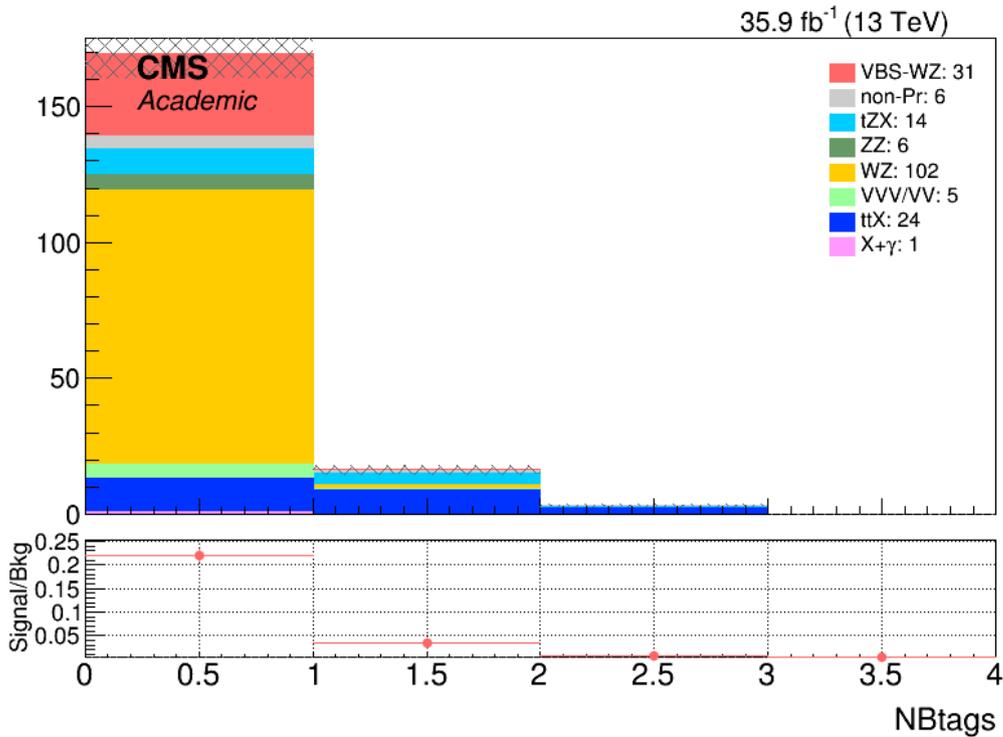


Figura 35: Número de jets identificados como quarks b en el estado final para cada Monte Carlo. Todos los criterios de selección hasta este punto del análisis están aplicados en este histograma.

El criterio de selección en esta variable parece evidente sin más que observar el histograma de la figura [35]. Sin embargo, hay que tener presente que la identificación de quarks b también lleva un error asociado, que se propagará a la región de señal si aplicamos criterios de selección en esta variable.

El algoritmo de identificación de quarks b es muy complejo. El Pile Up del orden de 40 vértices por cruce de bunches, dificulta en gran medida saber si un jet proviene de un b que se generó en el vértice principal o de otro objeto proveniente de un vértice secundario.

Además, aunque la vida media media del b corresponde con una distancia de vuelo de unos

2.5 mm, el valor de τ_b es un valor *medio*, con lo que no todos los b recorrerán exactamente esta distancia antes de desintegrarse.

Estos dos factores provocan que la incertidumbre en el etiquetado de quarks b sea, junto con la de la caracterización de los jets, la más grande del suceso.

Además, este es el único punto del análisis en el que uno podría plantearse restringir los valores de esta variable. Si no lo hacemos aquí, no lo habremos hecho en ningún momento, y por tanto esta incertidumbre no se propagará *en absoluto* a nuestra región de señal. Los resultados numéricos que posteriormente obtengamos de la región de señal serán totalmente independientes a la incertidumbre en el etiquetado de bs si no aplicamos criterios de selección en esta variable.

Tras realizar el estudio pertinente, he llegado a la conclusión de que la incertidumbre que introducen los fondos en la segunda y tercera columna de la figura [35] es mayor a la que introduce la incertidumbre en el etiquetado de quarks b al imponer que el número de ellos en el estado final sea igual a cero. Imponer que no se cuenten b-Jets en el estado final será por tanto el último de los criterios de selección.

4.7.2. Resultados en la región de señal VBS

Con los criterios de selección aplicados hasta ahora ha quedado definida de un modo razonablemente preciso la región de señal de la producción WZ VBS electrodébil. Los histogramas más relevantes en esta región se muestran en la figura [36].

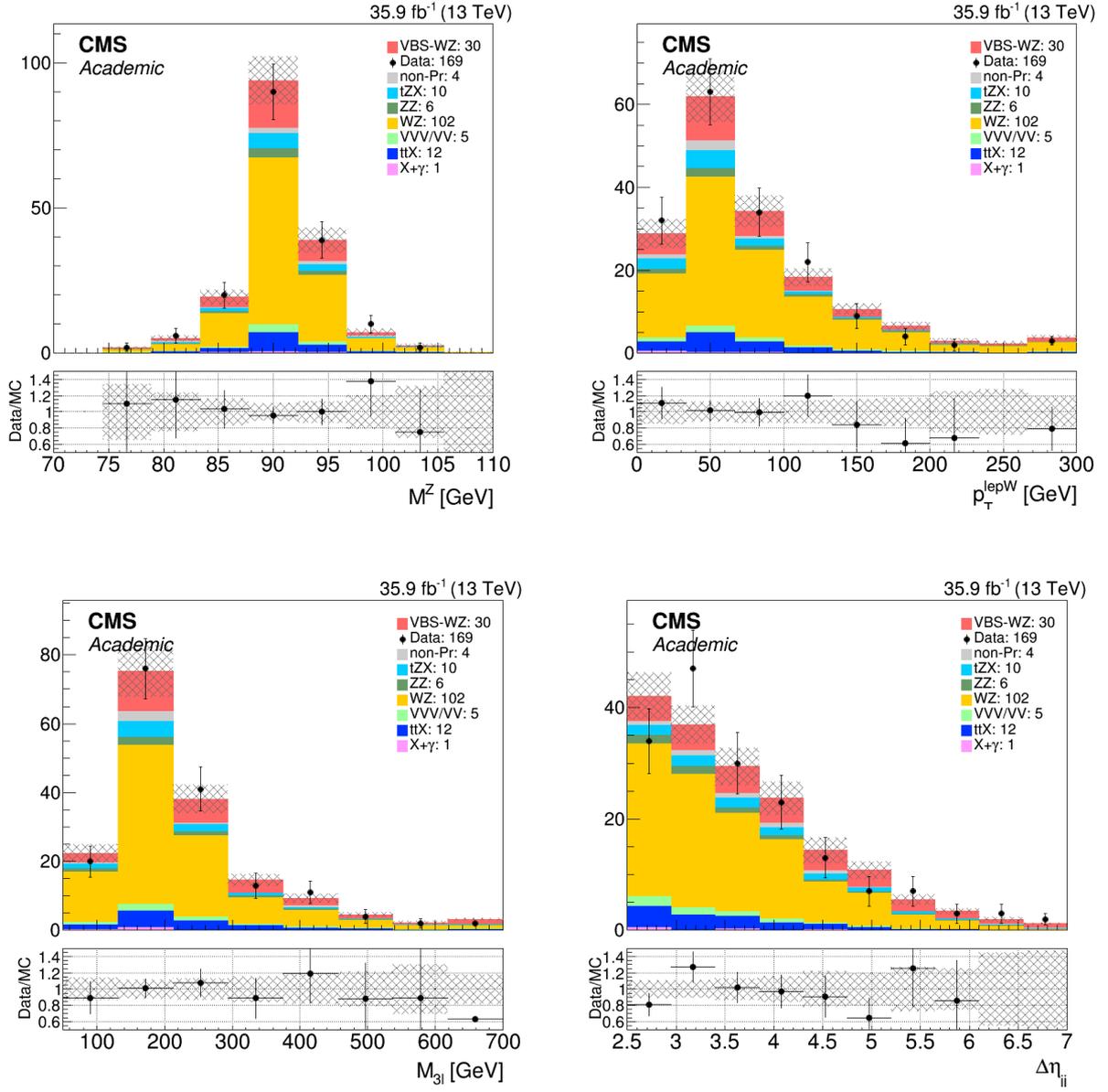


Figura 36: De arriba a abajo y de izquierda a derecha: La masa invariante de los dos leptones del Z, el momento transversal del leptón del W, la masa invariante de los tres leptones del estado final y la diferencia en η de los dos jets del estado final. Se representa con la rejilla gris la incertidumbre total de los Monte Carlos.

En los histogramas de la figura [36], el proceso de señal representa el 18% del total de los Monte Carlos, frente al 0.7% que representaba al inicio del análisis. Por supuesto podemos mejorar esta proporción de forma prácticamente arbitraria, aunque acotar la región de señal conllevaría un importante crecimiento en la incertidumbre estadística de los datos, y por tanto una pérdida de precisión. En la sección 5.5 analizaremos la propagación de todas las incertidumbres a la fuerza de la señal, y comprobaremos que, según hemos definido cada criterio de selección, la contribución a esta incertidumbre de los datos es del orden de la que inducen los Monte Carlos. Intentar mejorar una conllevaría con una muy alta probabilidad un aumento excesivo de la otra.

5. Medida de la fuerza de la señal para la producción WZ y WZ VBS

La *fuerza de la señal* como magnitud física ya ha sido definida en la sección 3.2, donde también se han comentado las características de su valor central. A continuación se estudiarán las principales contribuciones a su incertidumbre, así como su propagación al valor central. Por último, calcularemos esta cantidad tanto para la producción WZ como para WZ VBS.

5.1. Fuentes de incertidumbre

A continuación se describen las principales contribuciones a la incertidumbre de la fuerza de la señal. Dividiré este epígrafe en incertidumbres *estadísticas* y *sistemáticas*.

5.1.1. Incertidumbres estadísticas

Calcularemos todos los errores estadísticos asumiendo una distribución poissoniana:

$$\delta N = \sqrt{N}$$

la incertidumbre estadística dominante es, con gran diferencia, la de los datos. Salvo excepciones puntuales, todos los Monte se generan con un número de sucesos varios órdenes de magnitud superior a los sucesos observados. El objetivo de esta estrategia es minimizar la contribución al error total, haciendo el estadístico de los MC despreciable frente al de los datos.

5.1.2. Incertidumbres sistemáticas

Entenderemos por incertidumbre sistemática toda aquella que no es estadística. Por una parte, todas las incertidumbres provenientes de ineficiencias en el detector (en el etiquetado de jets B, en el trigger de leptones...) se incluyen dentro de esta categoría

Por otro lado, serán también sistemáticas todas las incertidumbres que deriven del hecho de que, en la práctica totalidad de los casos, será necesario aplicar correcciones al Monte Carlo para ajustarlo a los datos, antes de comenzar a extraer conclusiones de su comparación.

Esto es así porque el Monte Carlo nunca reproducirá con fidelidad absoluta el comportamiento del detector. Por cuestiones prácticas (generar un MC completo requiere un largo periodo de tiempo), los Monte Carlos se producen computacionalmente al mismo tiempo que se toman los datos en el detector. Muchas variables que puedan cambiar a lo largo del periodo de toma de datos (perfil de Pile Up, etc), así como imprevistos que alteren puntualmente el funcionamiento del detector supondrán correcciones a los MC, que se les aplicarán una vez generados.

Por tanto, los MC tendrán que ser corregidos para adaptarse a los datos medidos. Estas correcciones llevarán asociadas las siguientes incertidumbres sistemáticas que tendremos en cuenta de cara al análisis.

Jet Energy Corrections (JEC)

Los jets son objetos complejos cuya caracterización lleva asociada una incertidumbre relativamente grande.

Representando en un histograma alguna variable relacionada con los jets, y tomando en ella un pico como referencia y comparando en él las distribuciones de datos y predicciones, se distinguen las siguientes correcciones:

JES, Jet Energy Scale: Me referiré con ellas a las correcciones que son aplicadas al MC con el fin de desplazar un pico en el eje horizontal para centrarlo con las observaciones.

JER, Jet Energy Resolution: Hace referencia a las correcciones aplicadas a un pico para ajustar su anchura a la observada en los datos.

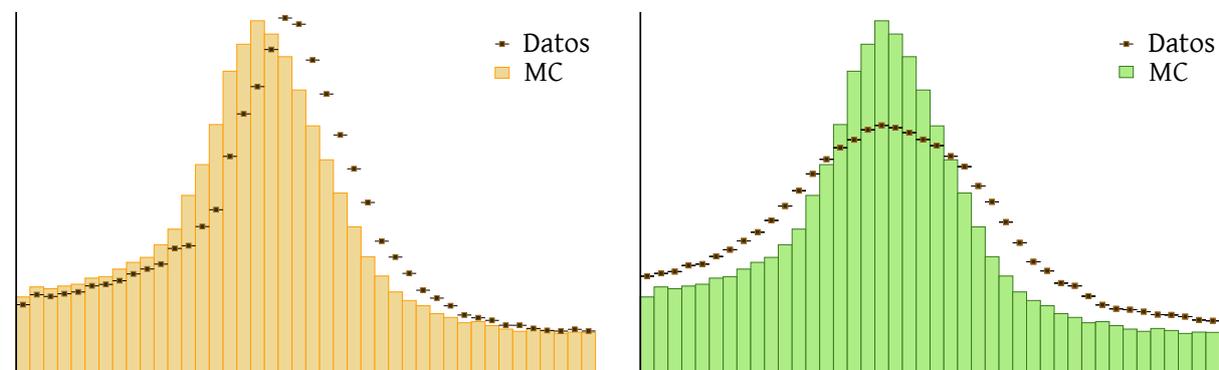


Figura 37: El Monte Carlo del histograma de la izquierda está desplazado hacia la izquierda respecto a los datos, por lo que necesita ser corregido por un factor de escala. El Monte Carlo de la derecha tiene una resolución mucho mayor que la que se ve en datos, y necesitará también ser corregidos para adaptarse a la realidad. *Figura propia.*

En adelante, nos referiremos con JEC exclusivamente a las JES, por ser mucho mayores que las JER, que no se tendrán en cuenta.

Corrección a la eficiencia del etiquetado de quarks b

El etiquetado de quarks b también lleva asociada una de las mayores incertidumbres, que por consiguiente tendrá que tenerse en cuenta.

El b Tag SF se aplica a los Monte Carlos para adaptar la eficiencia de etiquetado de quarks b a la que se observa en datos.

Tanto esta incertidumbre como la que describimos a continuación son incertidumbres de eficiencia que en general son debidas a errores en la modelización en los leptones provenientes del interior de jets.

Factores de escala para electrones y muones

Nos referimos con estos factores de escala a los que se le aplican a los electrones y muones para ajustar el MC a los datos. Generalmente, la calibración se lleva a cabo en el pico del Z . Este es un pico muy bien conocido, que se construye calculando la masa invariante de los leptones en los que se desintegra este bosón.

Estos leptones tienen siempre el mismo sabor, por lo que el mismo histograma puede construirse tanto utilizando sucesos con electrones en el estado final como utilizando sucesos con muones. Los electrones y muones son identificados con precisión en el detector, con lo que el histograma del Z a muones y del Z a electrones puede construirse de forma muy sencilla en datos.

El mismo histograma se construye en ambos canales utilizando Monte Carlos. Se comparan ambas distribuciones con los datos y se le aplican al MC las correcciones necesarias para ajustarlo a la realidad.

Corrección a la distribución del Pile Up (Pile Up Scale Factor, PUSF)

Ya se ha comentado en la sección [2.5] que el Pile Up de cada suceso de datos no es constante a lo largo de todo el periodo de medidas.

Los Monte Carlos tienen que reproducir este hecho, y se generan también con un perfil de Pile Up prefijado, que en un caso ideal será idéntico al recogido en datos. Como esto no ocurre, el

MC tendrá que ser nuevamente repesado para adaptarse al Pile Up registrado tras cada periodo.

Estos pesos constituirán el Pile Up Scale Factor (PUSF), que llevará asociada una incertidumbre sistemática que también tendremos en cuenta en lo sucesivo.

5.2. Propagación de las incertidumbres a la región de señal

Tanto todas las incertidumbres sistemáticas como la estadística tendrán que ser propagadas a la fuerza de la señal. Esta propagación no es trivial y ha sido llevada a cabo de un modo distinto en cada caso.

Incertidumbres de eficiencias y correcciones en los Jets

Las incertidumbres de eficiencias (factores de escala para el Pile Up, para electrones y para muones) se llevan a cabo modificando el peso de cada suceso de Monte Carlo hacia arriba y hacia abajo, dentro de lo que permitan las cotas superiores e inferiores de cada factor de escala con su incertidumbre. Con cada peso variado, se reconstruye la región de señal, y la propagación de estas incertidumbres se obtiene comparándola con la original (construida con el valor central de cada peso).

Propagar las incertidumbres de los jets es algo más complicado. Considerar el error en las correcciones de las JEC (y por tanto en el valor de la energía de los jets) y en el algoritmo de etiquetado de quarks b (y por tanto en el número de b-Jets del suceso) provocará que el número de sucesos en la SR cambie con la variación hacia arriba y hacia abajo de estas correcciones. El análisis se llevará a cabo tres veces por corrección (una con su valor central y otras dos en los extremos del rango de incertidumbre), y de los resultados se obtendrá la propagación de estas incertidumbres a la región de señal.

Incertidumbres de normalización

Las incertidumbres de normalización aparecen como consecuencia de la necesidad de renormalizar los Monte Carlos una vez generados a NLO.

En todos los casos, la sección eficaz de un proceso a NLO es menor que la sección eficaz real de ese mismo proceso. De modo teórico (independiente de las medidas), muchas secciones eficaces son ya conocidas a NNLO, aunque aún no existan Monte Carlos a este orden. La sección eficaz a NNLO de muchos procesos físicos es considerablemente más cercana a la realidad que la misma cantidad calculada a NLO.

Por este motivo, comúnmente se multiplica el peso de todos los sucesos de un Monte Carlo por el cociente $\sigma_{NNLO}/\sigma_{NLO}$. Esto, sin embargo, *no* es equivalente a generar el mismo Monte Carlo a NNLO, y lleva por tanto una incertidumbre asociada. En otras ocasiones, las incertidumbres aparecerán como consecuencia de un mal modelado en una variable concreta, que puede ocurrir aunque la sección eficaz del Monte Carlo sea la que le corresponde.

Cuando se genera un Monte Carlo a NNLO, se están teniendo en cuenta *todas* las contribuciones de *todos* los diagramas de Feynman del proceso hasta NNLO. Sin embargo, la contribución de cada diagrama no es independiente del valor de las variables cinemáticas del suceso. Algunos diagramas tendrán más peso a (pongamos) momentos transversos bajos que a altos valores de p_T , y lo mismo ocurrirá para el resto de variables.

Este tipo de dependencias para los diagramas a NNLO es desconocido hasta que no se genere el Monte Carlo. En general son pequeñas, y multiplicar todos los pesos por el mismo factor es una buena aproximación. Estos factores vienen dados para cada proceso en la tabla [3].

Proceso	Error (%)
WZ	6
Non Pr	30
$X + \gamma$	20
ttX	10
tZX	15
VVV/VV	50
ZZ	7

Tabla 3: Incertidumbres de normalización de los Monte Carlos utilizados en este análisis.

La propagación a la fuerza de la señal dependerá esencialmente del tamaño de cada Monte Carlo en la región de señal. Más adelante estudiaremos en detalle esta propagación.

Otras incertidumbres

Otras incertidumbres presentes en la región de señal son la de la luminosidad (dada por el LHC, independiente al análisis) y la de la eficiencia del trigger. Ambas tienen un valor fijo (2.6 % y 1.6 % respectivamente), que afectan de igual manera tanto a datos como a Monte Carlos.

5.3. Propagación de incertidumbres a la fuerza de la señal

En este punto del análisis nos son ya conocidas todas las incertidumbres sistemáticas y estadísticas en la región de señal, donde mediré la fuerza de la señal.

Concretamente, he adaptado a este análisis una de las funciones del framework utilizado para producir una tabla con la contribución de cada una de las incertidumbres a cada uno de los procesos. En total, estudiaré la propagación del factor de escala del Pile Up, de los factores de escala de electrones y muones, de las JEC y la incertidumbre en el etiquetado de quarks b, y de los factores de normalización y errores estadísticos de cada una de las predicciones. Incluiré además la incertidumbre en la eficiencia del trigger, y en la luminosidad, que son constantes y no dependen de nuestro análisis.

Calcularé por separado cada contribución a la fuerza de la señal. Posteriormente, asumiendo que todas ellas son independientes, calcularé la incertidumbre total sumándolas cuadráticamente.

Denotemos con $\sigma_{src}(\mu)$ a la contribución de la incertidumbre de cualquier fuente a la fuerza de la señal. Aquí, *src* puede sustituirse por cualquiera de las incertidumbres indicadas en el primer párrafo de esta sección. Calcularemos su valor utilizando la expresión general para este tipo de propagaciones:

$$\sigma_{src}(\mu) = \sqrt{\sum_i \sum_j \left(\frac{\partial \mu}{\partial N_i} \right) V_{ij} \left(\frac{\partial \mu}{\partial N_j} \right)} \quad (8)$$

donde los índices i y j recorren todos los Monte Carlos, incluido el de señal. las N_i son cada uno de los procesos (señal + todos los fondos) modelados por Monte Carlo.

$$\mu = \frac{N - \sum_k B}{S}$$

donde el índice k recorre todos los fondos.

V_{ij} son las entradas de la denominada *matriz de covarianzas*. Es una matriz cuadrada de dimensión el número de Monte Carlos utilizados y componentes $V_{ij} = \rho_{ij} \sigma_{src}(N_i) \sigma_{src}(N_j)$, con i y j recorriendo la señal y todos los fondos, y siendo ρ_{ij} el llamado *coeficiente de correlación* entre la incertidumbre del proceso i y la del j .

Si englobamos por simplicidad todos los fondos bajo el símbolo \mathbf{B} , la matriz de correlación

tomará la forma

$$V = \begin{pmatrix} \sigma_{src}^2(S) & \rho \cdot \sigma_{src}(S)\sigma_{src}(\mathbf{B}) \\ \rho \cdot \sigma_{src}(S)\sigma_{src}(\mathbf{B}) & \sigma_{src}^2(\mathbf{B}) \end{pmatrix}$$

donde ρ varía entre -1 y 1 y que indica lo correlacionados que están, en este caso, la propagación de la incertidumbre de la fuente *src* al Monte Carlo de señal y al de los fondos:

- $\rho = 1$: las dos incertidumbres están completamente correlacionadas. Si una aumentase, la otra lo haría de la misma manera.
- $\rho = -1$: las incertidumbres están completamente *anticorrelacionadas*. Aumentar una supondrá reducir la otra de modo proporcional.
- $\rho = 0$: Las dos incertidumbres están descorrelacionadas.

Valores intermedios de ρ se corresponderán con casos intermedios de las tres proposiciones anteriores.

Los valores de $\sigma_{src}(S)$ y $\sigma_{src}(\mathbf{B})$, por otra parte, se obtienen directamente de los histogramas de la SR para cada fuente (b Tag, SF de electrones, normalización del MC ttX...) y para cada Monte Carlo. La excepción son la incertidumbre de la luminosidad y la de la eficiencia del trigger, que son conocidas de antemano.

Todas las incertidumbres relativas $\sigma_{src}(N_i)$ (dadas, por ejemplo, en porcentaje) dependen únicamente de la fuente y no del tamaño del set de sucesos sobre el que se estén midiendo. Se cumplirá entonces, en todos los casos, $\sigma_{src}(\mathbf{B}) = \sigma_{src}(\sum_k B_k) = \sum_k \sigma_{src}(B_k)$, y todos los coeficientes de correlación ρ serán además iguales a 1.

Expandiendo los sumatorios de la ecuación (8) teniendo en cuenta la definición de $\mathbf{B} = \sum_k B_k$ llegamos de modo sencillo a

$$\sigma_{src}(\mu) = \sqrt{\sigma_{src}^2(S) \left(\frac{N - \mathbf{B}}{S} \right)^2 + \sigma_{src}^2(\mathbf{B}) \frac{1}{S^2} + 2\sigma_{src}(S)\sigma_{src}(\mathbf{B}) \frac{1}{S} \frac{N - \mathbf{B}}{S^2}} \quad (9)$$

que se corresponde con la propagación de la incertidumbre con origen en *src* a la fuerza de la señal.

5.4. Fuerza de la señal para la producción WZ

En primer lugar, calcularemos la fuerza de la señal para el proceso de producción WZ. Mediremos este valor en la región de señal definida en la sección 4.5.

El script que he escrito para calcular $\mu \pm \sigma(\mu)$ utiliza como input una tabla producida junto a los histogramas. Esta tabla aporta información sobre el número de sucesos de cada Monte Carlo, y sobre la medida en la que le afecta cada sistemático.

$N = 3855$	PUSF	ElecSF	MuonSF	JEC	b Tag	Estadística	ev. totales
X+γ	0.7	0.9	0.3	5	0	3	39
ttX	0.6	0.6	0.7	0.1	6	0.5	53
VVV/VV	0.7	1	1	0.8	0.4	1	76
ZZ	4	4	3	10	0.3	1	267
tZX	0.3	0.5	0.5	0.3	3	0.3	42
non-Pr	0	0	0	0	0	5	101
WZ	40	50	40	30	3	20	3400

Tabla 4: De izquierda a derecha, las contribuciones del Pile Up, de los electrones y de los muones, las incertidumbres en las correcciones de energía de los jets y en el etiquetado de quarks b y la incertidumbre estadística. En la última columna, el número de sucesos totales para cada Monte Carlo. Todos los valores se dan en número de sucesos. Arriba a la izquierda se indica el número de sucesos de datos contados en esta SR.

A las incertidumbres de la tabla [4] hay que añadir las de normalización (dadas en la tabla [3]), excepto la de normalización de WZ, que es precisamente la que queremos calcular.

El 6% que leemos para WZ en la tabla [3] corresponde con el utilizado en la referencia [15], que se obtuvo experimentalmente de un modo similar al que aquí se plantea, y que no tiene por qué coincidir con el resultado que aquí se obtenga.

Tendremos en cuenta también la de la luminosidad (un 2.6% para todos los sucesos) y la proveniente de la eficiencia del trigger, un 1.6%. Por último, añadiremos la incertidumbre estadística de los datos, calculada como $\sqrt{N} = 62$ sucesos.

Para calcular la incertidumbre estadística de los Monte Carlos, se ha tenido en cuenta que estos están *pesados* por un número que en todos los casos es mucho menor que 1, para adaptarse a la estadística de los datos. El número real de sucesos de cada MC es muy superior al que figura en la última columna de cada fila en la tabla [4].

Una vez disponemos de toda esta información, estamos ya en condiciones de propagar numéri-

camente cada incertidumbre a la fuerza de la señal. Utilizaremos para ello la ecuación [9], calculando previamente los valores de $\sigma_{src}(S)$ y $\sigma_{src}(\mathbf{B})$.

Para cada fuente *src* (factores de escala, correcciones de los jets, estadística de los MC, incertidumbres de normalización, eficiencia del trigger y luminosidad), la incertidumbre sistemática total (en porcentaje) proveniente de ella podrá calcularse como

$$\delta_{src} = \sum_{i \in MC} \frac{\delta_i}{N_{tot}} \cdot 100$$

con δ_i la contribución de *src* al Monte Carlo *i*, y N_{tot} el número de sucesos totales. (la suma de todos los sucesos de todos los Monte Carlos).

Una vez conocidas las incertidumbres asociadas a cada fuente, calcularemos $\sigma_{src}(S)$ y $\sigma_{src}(\mathbf{B})$ como la contribución, medida en número de sucesos, de la incertidumbre proveniente de *src* a la señal y a los fondos:

$$\sigma_{src}(S) = \delta_{src} \cdot \frac{S}{100} \qquad \sigma_{src}(\mathbf{B}) = \delta_{src} \cdot \frac{\mathbf{B}}{100}$$

Y con estos valores, calculamos la propagación a la fuerza de la señal de la incertidumbre de cada fuente, $\sigma_{src}(\mu)$ sustituyéndolos en la ecuación [9].

Si tomamos por *N* el número medido de datos, obtenemos como valor central $\mu = 0.964$. Todas las contribuciones a la incertidumbre de este resultado vienen dadas en la tabla [5]

$\mu_{WZ} = 0.964 \pm$	Leptones	Muones	0.014
		Electrones	0.015
	Pile Up	Pile Up	0.014
	Jets	JEC	0.013
		b tagging	0.0035
	Incertidumbres de normalización	Non Prompt	0.0087
		X+ γ	0.0022
		ttX	0.0015
		tZX	0.0018
		VVV/VV	0.011
		ZZ	0.0053
	Estadísticos	Datos	0.018
		Monte Carlo	0.0084
	Otros	Trigger eff	0.018
		Luminosidad	0.029

Tabla 5: Contribución de cada fuente de incertidumbre a la medida de la fuerza de la señal de la producción WZ.

La contribución más grande viene representada por la incertidumbre en la luminosidad, que viene determinada por el LHC y no depende ni del análisis ni de los Monte Carlo. El resto de contribuciones son aproximadamente del mismo orden.

Con esto ya estamos en condiciones de calcular la fuerza de la señal junto a su incertidumbre, a partir de los datos medidos en el detector y todas las predicciones por Monte Carlo.

Sin embargo, se estudiarán antes las hipótesis $\mu = 1$ ($N = S + B$, la señal es tal y como la predice el MC que la modela) y $\mu = 0$ ($N = B$, no hay señal en absoluto), cuyo significado se ha explicado en la sección 3.2. Este estudio previo se conoce como *análisis ciego*[25].

Comencemos estudiando la hipótesis de ausencia de señal. Se obtiene en este caso $\mu_c = 0.00 \pm 0.01$. Recuérdese que con esto no podemos garantizar que no exista señal, pero sí que si existe será muy pequeña, y quedará contenida en la banda de incertidumbres del resto de procesos.

En la hipótesis opuesta, se obtiene $\mu_c = 1.00 \pm 0.05$. Esta incertidumbre es también razona-

blemente pequeña, y podrá por tanto afirmarse que hay señal si así lo determina el valor real de μ .

Este valor real, calculado utilizando por N el número de sucesos de datos medidos en el detector, devuelve un valor de

$$\mu_{WZ} = 0.96 \pm 0.05$$

Que es compatible con 1, con lo que dentro del margen de incertidumbre verificamos la hipótesis de existencia de señal, y concluimos que el Monte Carlo que la modela reproduce la realidad con una indeterminación del 5% en μ .

La incertidumbre en esta medida es razonablemente pequeña, aunque también la cantidad de estadística del análisis ha sido considerablemente grande. El 5% que obtenemos como indeterminación se utilizará posteriormente como incertidumbre de normalización de WZ, cuando utilicemos este proceso como fondo de la producción VBS.

5.5. Fuerza de la señal para la producción WZ VBS

Como ya sabemos, la sección eficaz de WZ VBS es mucho menor que la de WZ. La región de señal para este proceso se ha definido con mucho cuidado, y con ello se espera obtener una medida razonablemente buena dentro de la estadística de la que disponemos.

La tabla siguiente ilustra la contribución de cada incertidumbre a cada proceso, una vez optimizados y aplicados todos los criterios de selección.

N = 169	PUSF	ElecSF	MuonSF	JEC	b Tag	Estadística	ev. totales
X+γ	0.01	0.02	0.01	0	0	0.56	1.11
ttX	0.15	0.15	0.16	0.69	1.5	0.23	12.01
VVV/VV	0.04	0.07	0.07	0.54	0.05	0.33	5.07
WZ	1.1	1.4	1.4	10	0.45	3	101.53
ZZ	0.03	0.10	0.07	0.65	0.03	0.2	5.72
tZX	0.1	0.13	0.12	0.68	0.84	0.10	9.58
Non Pr	0	0	0	0	0	0.94	4.23
WZ VBS	0.4	0.39	0.41	1.1	0.08	0.14	30.3

Tabla 6: Contribución de cada incertidumbre sistemática a la región de señal de WZ VBS. Todos los valores vienen dados en número de sucesos.

A las incertidumbres de la tabla [6] hay que añadir las de normalización (dadas en la tabla [3]).

La mayor fuente de incertidumbre en este análisis son las correcciones en la energía de los jets. A pesar de no haber aplicado criterios de selección en la MET, tanto el momento de los jets como su masa invariante (dos restricciones imprescindibles en este análisis) se ven muy afectados por su energía, y reducir este error es por tanto muy complicado. Solo estas correcciones suponen 10 sucesos de incertidumbre cuando se aplican a WZ, un 30 % del total de los sucesos de señal.

Las incertidumbres de normalización de los Monte Carlos tampoco representan una contribución despreciable. Renormalizar los Monte Carlos es algo indispensable en cualquier análisis de este tipo. El mejor modo de minimizar esta contribución es reducir los fondos en la medida de lo posible. El 20 % que supone la normalización del MC $X + \gamma$ es despreciable cuando se aplica al único suceso que contamos de este proceso en la región de señal. El 5 % de normalización de WZ representará en cambio una de las principales contribuciones a la incertidumbre total. Aplicado a los 102 sucesos que contamos en la SR, supondrá una incertidumbre de cinco sucesos en la SR, casi un 20 % del total de los sucesos de señal.

Las incertidumbres estadísticas en los Monte Carlos son de un orden aceptable. La más preocupante es el 3 % de WZ. Este es el mayor de los fondos, y por tanto un 3 % de los 102 sucesos de WZ se traduce en tres sucesos de incertidumbre, un 10 % de los 30 sucesos de señal.

Sin embargo, basta únicamente con utilizar un Monte Carlo más grande para reducir esta incertidumbre, aumentando con ello el número de sucesos N_{WZ} y disminuyendo por consiguiente su error estadístico relativo, $\sqrt{N_{WZ}}/N_{WZ}$.

En este caso, el cálculo de μ devuelve un valor central de 0.981. Todas sus incertidumbres están desglosadas en la siguiente tabla, análoga a la que construimos para WZ.

$\mu_{VBS} = 0.981 \pm$	Leptones	Muones	0.072
		Electrones	0.072
	Pile Up	Pile Up	0.06
	Jets	JEC	0.46
		b tagging	0.097
	Incertidumbres de normalización	WZ	0.17
		Non Prompt	0.042
		X+ γ	0.0073
		ttX	0.039
		tZX	0.047
		VVV/VV	0.083
		ZZ	0.013
	Estadísticos	Datos	0.43
		Monte Carlos	0.18
	Otros	Trigger eff	0.089
		Luminosidad	0.14

Tabla 7: Contribución de cada fuente de incertidumbre a la medida de la fuerza de la señal de la producción WZ VBS.

Las principales incertidumbres en la tabla [7] son la proveniente de las correcciones en la energía de los jets y la estadística de los datos, seguidas de la de la luminosidad (irreducible) y la normalización de WZ.

La incertidumbre en las JEC es inevitable en este análisis, pues la definición de la región de señal depende para este proceso en muy buena parte de las características de los jets. La incertidumbre estadística de los datos podría reducirse relajando los criterios de selección, aunque ello conllevaría un aumento de las incertidumbres de normalización en general, y de la de normalización de WZ en particular, que fácilmente compensarían la pérdida de error estadístico.

Imponer $\mu = 1$ devuelve en este caso un error de 0.7, que aunque elevado no llega a incluir el caso $\mu = 0$, permitiendo afirmar que la señal existe si así lo determinasen los datos reales. Si se impone la hipótesis nula, el resultado es $\mu = 0.0 \pm 0.6$.

Calculando μ_{VBS} con el número de sucesos de datos contados en la región de señal cuadráticamente todas las incertidumbres, se obtiene para este proceso de señal

$$\mu_{VBS} = 1.0 \pm 0.7$$

La incertidumbre de el valor de esta fuerza de la señal es demasiado grande como para obtener de él conclusiones claras; no podemos afirmar que el Monte Carlo reproduzca la realidad con la rotundidad con la que pudimos hacerlo para el proceso WZ. Aunque la medida indica que pueda haberla, una medida más precisa es necesaria para corroborar esta hipótesis.

Mejorar esta medida requeriría de una reducción de las indeterminaciones asociadas a ella. Aunque reducir la incertidumbre estadística es sencillo, no lo es tanto mejorar la incertidumbre de las correcciones de la energía de los jets. Estos objetos suponen además una de las señas de identidad de la producción VBS, con lo que prescindir de ellos en el análisis no es una opción viable.

6. Conclusiones

Este trabajo se ha centrado en el estudio de los procesos de producción WZ y WZ VBS. En un primer lugar, se han definido dos regiones de control para la producción de dos bosones Z y de estados finales ricos en quarks top. El objetivo de este estudio previo ha sido el de validar el modelado de estos procesos, de cara a utilizarlo en el análisis de los procesos de señal. Tanto la normalización como la incertidumbre asociadas a ambos fondos ha sido validada al certificarse la compatibilidad con los datos observados en ambas regiones de control.

Toda la selección de sucesos se ha establecido con el objetivo de minimizar el error total en la fuerza de señal. En primer lugar, se ha realizado un estudio del criterio de selección en la energía faltante. Se ha concluido que seleccionar en ella lleva asociado un aumento de la incertidumbre estadística y de la proveniente de las correcciones relacionadas con los jets, que supone un aumento en la incertidumbre total de μ_{VBS} en todos los casos. No se han aplicado por tanto restricciones en esta variable. Seguidamente, se ha llevado a cabo otro estudio específico en el número de b-Jets del estado final que ha determinado, por el contrario, que el aumento en la relación señal/fondos y la disminución en la contribución de incertidumbres de normalización compensa el error sistemático que supone restringir el análisis a sucesos sin b-Jets. Finalmente, se ha analizado también la posibilidad de seleccionar sucesos en la variable de Zeppenfeld (a la que frecuentemente se recurre en análisis VBS), concluyendo que este criterio de selección no supone ninguna ventaja en el estudio concreto que aquí se presenta.

Se ha definido también una región de control para la producción WZ antes de utilizarse como fondo en el estudio de WZ VBS. En este caso se ha observado una ligera discrepancia entre los datos y las predicciones, que ha sido corregida calculando e introduciendo un factor de normalización en el Monte Carlo que modela WZ.

La región de señal para la producción WZ ha sido definida, y se ha calculado en ella una fuerza de señal de $\mu_{WZ} = 0.96 \pm 0.05$. La mayor contribución a la incertidumbre de esta medida ha resultado ser la de la luminosidad, que viene determinada por el LHC. Este resultado representa una medida competitiva, con una incertidumbre equiparable a la de la mejor medida de CMS hasta la fecha para este mismo proceso, publicada en [arXiv:1901.03428](https://arxiv.org/abs/1901.03428).

Se ha definido también la región de señal para la producción WZ VBS, y se ha calculado en ella la fuerza de la señal como parte central del análisis. Definir esta región de señal ha requerido más trabajo que el necesario para la región de WZ. La sección eficaz de WZ VBS es cuatro órdenes de magnitud inferior a la de producción WZ, lo que ha supuesto un importante

aumento de la contribución de la incertidumbre estadística a la medida final. Además, la incertidumbre proveniente de las correcciones a la energía de los jets ha supuesto en este análisis una contribución al total del orden de la estadística. El resultado obtenido para la fuerza de la señal es $\mu_{VBS} = 1.0 \pm 0.7$. El resultado más reciente de CMS para esta medida, publicado en [arXiv:1901.04060](https://arxiv.org/abs/1901.04060) (referencia [16]), ha sido obtenido de un modo más complejo, haciendo un ajuste a una distribución teniendo en cuenta todos los bins de la misma, con el objetivo precisamente de reducir las incertidumbres sistemáticas. Este resultado presenta una incertidumbre del 63% (dominada por la contribución estadística), compatible con el 70% obtenido en este trabajo siguiendo un procedimiento más directo.

Una posible extensión de este trabajo podría consistir en calcular la sección eficaz de la producción electrodébil de WZ VBS utilizando la distribución de una variable sensible a este proceso, en lugar de comprimir los histogramas en una sola columna como se ha hecho aquí. Este procedimiento requiere de técnicas estadísticas más avanzadas que las que aquí han sido necesarias, pero proporciona resultados más precisos que supondrían una reducción importante de la incertidumbre total en una situación como esta, en la que la estadística es reducida.

Podría también estudiarse experimentalmente la relación entre las constantes de acoplamiento fuerte y electrodébil, calculada como el cociente entre los diagramas que representan la producción de WZ VBS y la producción WZ QCD (discutido en la sección 1.3) en el estado inicial de los datos, antes de aplicar absolutamente ningún criterio de selección. Sin embargo, los sucesos de WZ VBS representarían una ínfima parte de todos los del estado inicial, y quedarían absorbidos en la incertidumbre de los fondos. Podría realizarse el cálculo, aunque se requeriría de muestras muy grandes para minimizar la contribución estadística y obtener así conclusiones claras.

Por último, el mismo análisis que se ha utilizado para definir y caracterizar las regiones de control y de señal en este trabajo podría utilizarse para estudiar otros procesos con leptones en el estado final. Como ejemplo, modificando la definición de las regiones de control para la producción ZZ o de sucesos con quarks top, podrían construirse las regiones de señal de estos procesos. Así podrían calcularse las fuerzas de señal en ambos casos, siguiendo un procedimiento muy similar al que se ha utilizado con WZ y WZ VBS.

Referencias

- [1] J R Aitchison y J G Hey, “Gauge Theories in Particle Physics”, ed. IOP Publishing Ltd 2004
Capítulo I: *Quarks and Leptons*, tercera edición.
- [2] M. Tanabashi et al. (Particle Data Group), “2019 Review of Particle Physics”, *Phys. Rev. D* **98**, 030001 (2018). También disponible online en pdglive.lbl.gov
- [3] The CMS Collaboration (2008). “The CMS experiment at the CERN LHC”. *Journal of Instrumentation*. **3**. S08004.
- [4] The ATLAS collaboration, “Measurement of Standard Model VBS/VBF production with the ATLAS and CMS detectors” en *26th International Conference on Supersymmetry and Unification of Fundamental Interactions (SUSY2018)* ATL-PHYS-PROC-2017-165
- [5] The CMS Collaboration, “The CMS tracker system project : Technical Design Report”, 1997, CERN-LHCC-98-006 ; CMS-TDR-5
- [6] The CMS Collaboration, “The CMS electromagnetic calorimeter project : Technical Design Report”, 1997, CERN-LHCC-2011-006. LHCC-P-004
- [7] The CMS Collaboration, “The CMS hadron calorimeter project: Technical Design Report”, 1997, CERN-LHCC-97-031
- [8] The CMS Collaboration, “The performance of the CMS muon detector in proton–proton collisions at $\sqrt{s} = 7$ TeV at the LHC”, [arXiv:1306.6905](https://arxiv.org/abs/1306.6905) [physics.ins-det]
- [9] The CMS Collaboration, “The CMS muon project : Technical Design Reports”, 1997, CERN-LHCC-97-032 ; CMS-TDR-3
- [10] The CMS Collaboration
“Identification of b-quark jets with the CMS experiment”
doi [10.1088/1748-0221/8/04/P04013](https://doi.org/10.1088/1748-0221/8/04/P04013)
- [11] Feltese, J. “Introduction to Parton Distribution Functions”, 2010
Scholarpedia, Volumen 5, doi:[10.4249/scholarpedia.10160](https://doi.org/10.4249/scholarpedia.10160)
- [12] The CMS Collaboration “The CMS trigger system” doi [10.1088/1748-0221/12/01/P01020](https://doi.org/10.1088/1748-0221/12/01/P01020)
- [13] M. Klein, R. Yoshida, “Collider Physics at HERA”, [arXiv:0805.3334](https://arxiv.org/abs/0805.3334) [hep-ex]
- [14] W Herr y B Muratori, “Concept of luminosity” en cds.cern.ch
- [15] The CMS Collaboration, “Measurements of the $pp \rightarrow WZ$ inclusive and differential production cross section and constraints on charged anomalous triple gauge couplings at $\sqrt{s}= 13$ TeV”, [10.1007/JHEP04\(2019\)122](https://doi.org/10.1007/JHEP04(2019)122).
- [16] The CMS Collaboration, “Measurement of electroweak WZ production and search for new physics in pp collisions at $\sqrt{s} = 13$ TeV, 2018, [10.1016/j.physletb.2019.05.042](https://doi.org/10.1016/j.physletb.2019.05.042).
- [17] The NNPDF Collaboration
“Parton distributions for the LHC Run II”
[arXiv:1410.8849](https://arxiv.org/abs/1410.8849) [hep-ph]
- [18] A Oh, “The CMS DAQ and run control system”, doi:[10.1088/1742-6596/110/9/092020](https://doi.org/10.1088/1742-6596/110/9/092020)
- [19] M Grazzini, S Kallweit, D Rathlev y M Wiesemann
“ $W^\pm Z$ production at hadron colliders in NNLO QCD”
[arXiv:1410.8849](https://arxiv.org/abs/1410.8849) [hep-ph]

- [20] M Grazzini, S Kallweit, D Rathlev (DESY) y M Wiesemann (Zurich U.)
“ W^+W^- , WZ and ZZ production in the POWHEG BOX V2”
[arXiv:1311.1365 \[hep-ph\]](#)
- [21] D. L. Rainwater, R. Szalapski, and D. Zeppenfeld, “Probing color singlet exchange in Z + two jet events at the CERN LHC”, *Phys. Rev. D* **54** (1996) 6680–6689
- [22] The CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004
- [23] R Gomez-Ambrosio, “Vector Boson Scattering studies in CMS: The $pp \rightarrow ZZjj$ channel”, 15 de julio, 2018 [arXiv:1807.09634](#)
- [24] I. González et. al., “PROOF Analysis Framework (PAF)”, en *Journal of Physics: Conference Series*, doi:10.1088/1742-6596/664/3/032009
- [25] P K. Sinervo.
“Signal Significance in Particle Physics”, 5 de agosto, 2002, [arXiv:hep-ex/0208005](#)
- [26] The CMS Collaboration.
Figura [7]: Estructura interna del detector CMS [cms.web.cern.ch](#)
- [27] The CMS Collaboration.
Figura [9]: Sección circular del detector CMS [indico.cern.ch](#)
- [28] The CMS Collaboration.
Figura [11]: Curva de Pile Up recogida en 2016. [indico.cern.ch](#) [twiki.cern.ch](#)
- [29] The CMS Collaboration.
Figura [13]: Luminosidad integrada en CMS entre mayo y noviembre de 2016. [indico.cern.ch](#) [twiki.cern.ch](#)