



Facultad de Ciencias

Técnicas de aprendizaje automático profundo para la asignación de momento a muones altamente energéticos en el experimento CMS del LHC

(Deep learning techniques for muon momentum assignment with the CMS experiment at LHC)

TRABAJO DE FIN DE MÁSTER
PARA ACCEDER AL
**MÁSTER EN
CIENCIA DE DATOS**

Autor : Pedro José Fernández Manteca
Director : Pablo Martínez Ruíz del Árbol
Co-director : Alicia Calderón Tazón

Septiembre - 2020

Para los agradecimientos
Para los agradecimientos

Índice

| | |
|---|-----------|
| 1. Introducción | 4 |
| 2. El experimento CMS | 7 |
| 2.1. Tubos de deriva: DTs | 8 |
| 2.2. Cámaras de tiras catódicas: CSCs | 8 |
| 3. Asignación de momento transverso en el experimento CMS | 11 |
| 3.1. Reajuste TPFMS (<i>Tracker-plus-first-muon-station</i>) | 11 |
| 3.2. Reajuste Picky | 11 |
| 3.3. Reajuste DYT (<i>Dynamic truncation</i>) | 11 |
| 3.4. Asignación final de momento transverso: el algoritmo TuneP | 12 |
| 4. Método propuesto | 13 |
| 4.1. Herramientas utilizadas para el análisis | 13 |
| 4.2. Muestra de simulación utilizada | 14 |
| 4.3. Selección de muones y segmentos | 14 |
| 4.4. Procesado y construcción de variables | 16 |
| 4.5. Distribuciones de control | 19 |
| 5. Entrenamiento de la DNN | 22 |
| 5.1. Variables de entrenamiento | 22 |
| 5.2. Conjunto de datos y arquitectura de la red | 22 |
| 5.3. Resultados del entrenamiento | 24 |
| 6. Resultados | 27 |
| 7. Conclusiones | 28 |

Resumen

El objetivo de este trabajo es aplicar técnicas de aprendizaje automático profundo supervisado para la asignación de momento transverso a muones altamente energéticos en el experimento CMS (*Compact Muon Solenoid*) del LHC (*Large Hadron Collider*) mediante regresión, teniendo como principal meta mejorar los resultados del procedimiento actual, y de ser así plantear incluir este tipo de metodologías de manera oficial en la futura toma de datos del experimento (Run 3).

Para el entrenamiento se ha utilizado una muestra de simulación Monte-Carlo, donde una partícula masiva se desintegra a muones de las características buscadas, de manera que el algoritmo predictivo se alimenta de la información que estos muones dejan a su paso por el detector CMS.

FILL ME: RESULTADOS

Palabras clave: aprendizaje automático profundo, muones altamente energéticos, CMS

Resumen

The aim of this work is to apply supervised deep machine learning techniques for the transverse momentum assignment to highly energetic muons in the CMS (*Compact Muon Solenoid*) experiment at LHC (*Large Hadron Collider*) collider through regression, with the main goal of improving the results of the current procedure, and if so, consider including this kind of methodologies for the future data taking of the experiment (Run 3).

For the training, a Monte-Carlo simulation sample has been used, where a massive particle decays into muons of the desired characteristics, so that the predictive algorithm feeds on the information that these muons leave as they pass through the CMS detector.

FILL ME: RESULTADOS

Key words: deep learning, highly energetic muons, CMS

1. Introducción

El objeto principal de estudio de este trabajo es el muón, que se trata de una partícula elemental cargada con spin 1/2 y con masa aproximadamente 200 veces mayor que el electrón. Además, siendo además una partícula inestable con un tiempo de vida de $2.2 \mu s$, que es elevado en comparación con otras partículas que poseen esta característica.

La mayor parte de los muones medidos por el detector CMS (del inglés *Compact Muon Solenoid*) situado en el gran colisionador de hadrones LHC (del inglés *Large Hadron Collider*) provienen típicamente de desintegraciones de quarks top, de hadrones o de desintegraciones leptónicas de bosones Z o W [1]. Estos muones se caracterizan normalmente por tener un momento transverso o p_T menor de 200 GeV y se categorizan como muones de bajo momento (*low- p_T muons*).

Por otra parte, los muones de alto momento (*high- p_T muons*) pueden tener como origen procesos físicos atípicos como la desintegración de partículas exóticas fuera del Modelo Estándard de la física de partículas [2] como bosones Z' o W' [3, 4] con masas en la escala del TeV, cuyo descubrimiento sería un indicativo directo de nueva física, por lo que medir las propiedades de estos muones de la manera más precisa posible en el detector es de vital importancia.

Experimentalmente, la medida del momento de los *high- p_T muons* plantea varias dificultades. Primero, hay que tener en cuenta que la resolución de la medida de p_T a partir de la traza empeora cuando el momento del muón aumenta.

En presencia de un campo magnético uniforme B , y con un radio de curvatura de la traza R , el momento transverso p_T de un muón con carga q se puede expresar a través de la ecuación de Lorentz como:

$$p_T[\text{GeV}] = 0,3 \times B[\text{T}] \times R[\text{m}] \times q \quad (1)$$

El campo magnético dentro del solenoide de CMS es prácticamente uniforme y conocido con gran precisión ($B = 3.8 \text{ T}$), mientras que el radio de curvatura se relaciona con la longitud del arco L y la distancia sagitta s definida en la Figura 1 a través de:

$$R[\text{m}] \approx L[\text{m}]^2 / 8s[\text{m}] \quad (2)$$

Esta aproximación es válida para $L/R \ll 1$.

Combinando las ecuaciones (1) y (2) se obtiene:

$$s[\text{m}] \approx (0,3B[\text{T}]L[\text{m}]^2/8)(q/p_T[\text{GeV}]) = (0,3BL^2/8) \times (q/p_T), \quad (3)$$

Se observa que s es inversamente proporcional al momento transverso, por lo que para mejorar la resolución en la medida del p_T en los casos con sagittas pequeñas, las trazas de los muones en CMS se reconstruyen en distintos subdetectores separados a varios metros del punto de colisión, como se detallará en la Sección 2, para así tener trazas de mayor longitud y por consiguiente mayores valores de s .

Sin embargo, aunque esto mejora considerablemente la resolución en la medida de p_T , sagitta sigue siendo pequeña para muones de muy alto momento en la escala del TeV, y además la medida de s se puede ver afectada por fallos en el alineamiento de las señales que componen la traza.

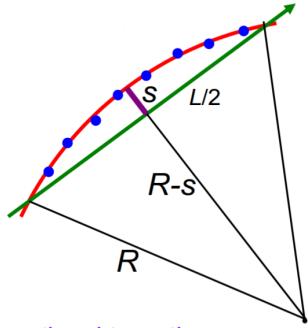


Figura 1: Definición de la distancia saggitta, s , a partir de la longitud de la traza reconstruida L y de su radio R .

Por otra parte, cuando los muones de alto momento atraviesan el hierro, las pérdidas de energía por radiación tales como producción de pares, Bremsstrahlung, o interacciones fotonucleares no son despreciables comparadas con la pérdida de energía por ionización. En la Figura 2 se muestra la dependencia de la pérdida de energía del muón por unidad de distancia dE/dx al atravesar distintos medios como función de su energía.

Se observa que la energía crítica para el hierro, E_c^{iron} , donde la energía de ionización (en marrón) es igual a la suma de todas las pérdidas radiativas (en morado) ocurre aproximadamente a 300 GeV. Como consecuencia, la principal fuente de pérdida de energía para un muón con $E > E_c^{iron}$ que viaja por hierro a través de los distintos subdetectores de CMS es debida a radiación electromagnética fruto de la producción de electrones y fotones.

Esta radiación electromagnética se manifiesta en el detector como una cascada de partículas (*muon shower*) que producen señales adicionales en los detectores, y puede incluso cambiar la dirección de la trayectoria del muón, afectando negativamente a la reconstrucción de su traza y degradando por consiguiente la medida de su momento.

Los algoritmos de reconstrucción de la traza que son utilizados para la asignación del momento transverso de los muones en CMS evitan lidiar con casos de emisión de cascadas, como se explicará con más detalle en la Sección 3. De esta manera, por ejemplo, si se encuentran varios impactos en un subdetector que puedan ser indicativo de que una cascada ha tenido lugar, típicamente se ignoran estas señales a la hora de reconstruir la traza del muón.

El objetivo de este trabajo, por contrapartida, es recopilar toda la información posible de la trayectoria de los muones a su paso por CMS (incluyendo los impactos provenientes de cascadas electromagnéticas) utilizando para ello muones simulados altamente energéticos con p_T generado conocido, y entrenar posteriormente una red neuronal profunda o DNN (del inglés *Deep Neural Network*) [6] que haga regresión

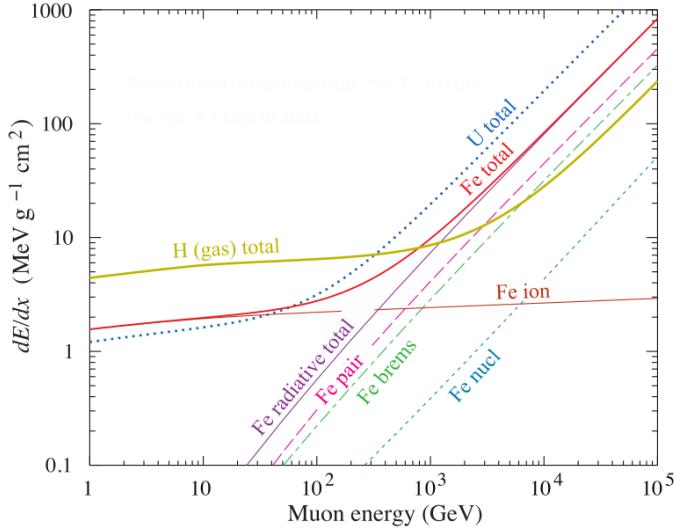


Figura 2: Media de la pérdida de energía por ionización y por radiación de un muón en hidrógeno, hierro y uranio como función de su energía. Average energy loss from ionization and radiative loss of a muon in hydrogen, iron and uranium as a function of its energy. En el caso del hierro, se separan las contribuciones a dE/dx para producción de pares, Bremsstrahlung e interacciones fotonucleares. Figura tomada de [5].

al p_T reconstruido, de manera que se consiga una asignación de momento transverso mejor que el proporcionado por los algoritmos centrales de CMS.

Para cuantificar la calidad de la asignación del momento transverso, se hará uso de la resolución R , definida como:

$$R = \frac{|p_T^{GEN} - p_T^{RECO}|}{p_T^{RECO}} \quad (4)$$

Donde p_T^{GEN} hace referencia al momento generado del muón y p_T^{RECO} al momento reconstruido a través de su traza en el detector.

La presente memoria está estructurada en las siguientes secciones. En la Sección 2 se describirá brevemente el dispositivo experimental utilizado: el detector CMS. Posteriormente, en la Sección 3 se presentarán los algoritmos actuales utilizados para la asignación de momento. En la Sección 4 se describirá el método propuesto, es decir, cómo se realiza la selección de muones y hits sobre la muestra de simulación utilizada. En la Sección 5 se detallará cómo se construyen las variables de entrenamiento, la arquitectura de la red utilizada, y los resultados del entrenamiento. Por último, en las Secciones 6 y 7 se mostrarán los resultados finales y las conclusiones del trabajo.

2. El experimento CMS

El detector CMS, representado de manera esquemática en la Figura 3, se localiza en uno de los puntos del acelerador LHC donde se hacen colisionar los haces de protones. Está compuesto, de la zona más interna a la más externa, por un tracker de píxeles y tiras de silicio para la detección de partículas cargadas con gran resolución espacial, un calorímetro electromagnético de cristal de tungstato (ECAL) para la medida de electrones y fotones principalmente, un calorímetro hadrónico constituido de material denso y absorbente (HCAL) especializado en la medida de hadrones, y finalmente, en la parte más externa se encuentran las cámaras de muones. Entre el HCAL y las cámaras de muones se tiene un imán superconductor que alcanza un campo magnético de 3.8 T, suficiente para curvar partículas cargadas y permitir una buena resolución en la medida del momento de las mismas.

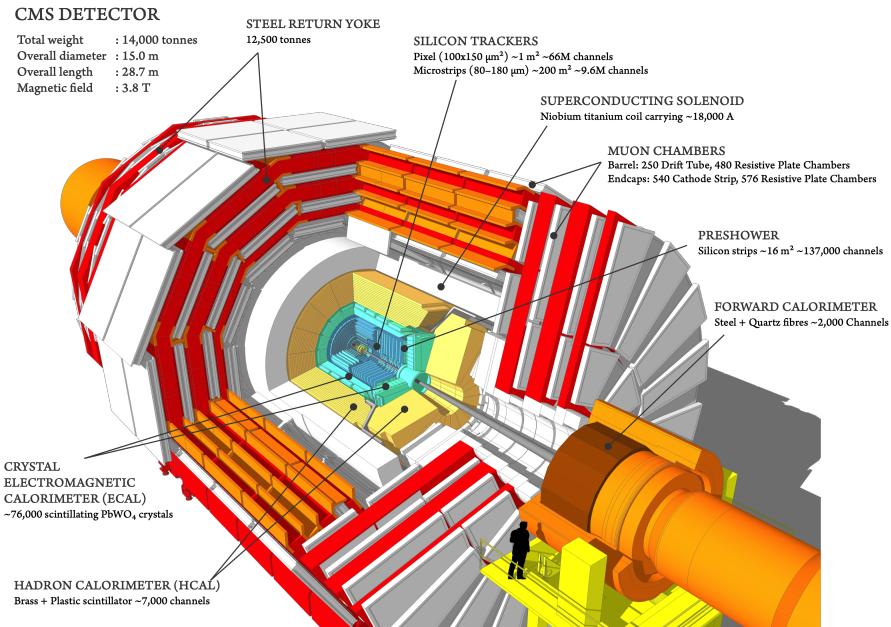


Figura 3: Representación gráfica de las distintas partes del detector CMS. Imagen tomada de [7].

En cuanto a su geometría, el sistema de coordenadas aceptado tiene como origen el punto de colisión, el eje y apunta verticalmente hacia arriba, el eje x radialmente desde el origen, y el eje z recorre la dirección del haz (ver Figura 3). El ángulo azimutal ϕ se mide a partir del eje x en el plano x-y transverso al haz, mientras que el ángulo polar θ se mide desde el eje z en el plano x-z. Otra variable angular importante que será utilizada en el análisis por ser invariante bajo transformaciones de Lorentz en el eje z, es la pseudorapidez η , que se define en función del ángulo polar como:

$$\eta = -\ln \left(\tan \frac{\theta}{2} \right) \quad (5)$$

De esta manera, se suelen usar variables definidas en el plano transversal a la dirección del haz de partículas como el momento transverso p_T o la energía transversa E_T .

En este trabajo nos centraremos en la medida de los muones, que al ser partículas cargadas dejan señal en el tracker interno, no interactúan apenas con el material denso de los calorímetros, y llegan a las cámaras de muones externas, situadas a unos cuatro metros del punto de colisión.

En las siguientes subsecciones se describirán dos de los subdetectores que componen las cámaras de muones que serán utilizados en el análisis: los tubos de deriva o DTs (del inglés *Drift Tubes*), y las cámaras de tiras catódicas o CSCs (del inglés *Cathode Strip Chambers*).

2.1. Tubos de deriva: DTs

Los tubos de deriva [8] cubren el barril de CMS ($|\eta| < 0.9$). Cada tubo se compone de un hilo colector de carga cargado positivamente y está lleno de gas, de forma que una partícula cargada a su paso por el tubo arranca electrones de los átomos del gas, que son atraídos eléctricamente y recolectados por el hilo conductor. De esta manera, se obtienen las coordenadas del paso del muón por el tubo a partir de la posición del hilo donde los electrones impactan y de la distancia del muón al hilo, que se calcula multiplicando la velocidad de deriva del electrón en el tubo por el tiempo de viaje hasta el hilo.

El subdetector completo consta de 250 cámaras, cada una de ellas con unas dimensiones en promedio de 2×2.5 m, formada por tres capas con unos 60 tubos cada una repartidos en cuatro subcapas.

En la Figura 5 se representa gráficamente la posición espacial de las distintas cámaras que forman las DTs, que se denotan como MBZ/N/S, donde Z=-2...+2 se corresponde con el número de rueda a lo largo del eje z, N=1...4 hace referencia al número de estación concéntrica en el plano xy, y S=1...12 al número de sector circular.

2.2. Cámaras de tiras catódicas: CSCs

Las cámaras de tiras catódicas [9] cubren las tapas de CMS ($1.2 < |\eta| < 2.4$) y su funcionamiento es similar al de las DTs. En este caso se tienen cables de carga positivamente (ánodos) cruzados con tiras de cobre cargadas negativamente (cátodos) dentro de un volumen de gas. Cuando los muones atraviesan la cámara arrancan electrones de los átomos de gas, produciéndose una avalancha de electrones que se dirigen a los cables del ánodo, mientras que los iones positivos se alejan del cable y se dirigen al cátodo de cobre, lo que también induce un pulso de carga en las tiras (en dirección perpendicular al ánodo). Debido a que las tiras y los cables son perpendiculares, se obtienen dos coordenadas de posición para cada partícula que atraviesa la cámara.

El subdetector completo de las CSCs contiene 540 cámaras en total, y está compuesto de anillos de cámaras trapezoidales de hasta 3.4 m de largo y 1.5 m de ancho,

colocadas en ocho discos, cuatro en cada tapa.

En este caso, las cámaras CSCs se denotan espacialmente como $ME \pm S/R$, donde el signo indica en qué tapa de CMS se encuentra, $S=1\dots4$ hace referencia al número de estación, paralelas en el eje z (ver Figura 5), y R se corresponde con el número de anillo, concéntricos en el plano xy. En la Figura 4 se muestra una imagen real de la rueda ME-1, con sus tres anillos concéntricos en el plano xy.

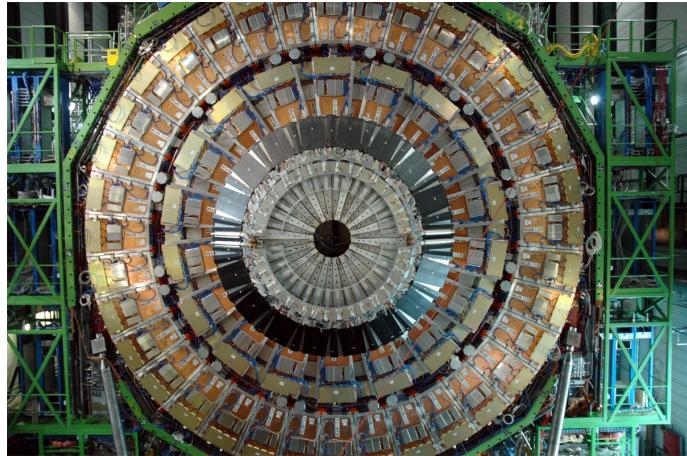


Figura 4: Foto completa de la estación ME-1. Imagen tomada de [10].

Cada una de las cámaras de DTs y CSCs se compone de varias capas, y las señales o impactos que dejan los muones a su paso se reconstruyen en cada una de ellas. A partir de estos impactos, se construyen trazas rectas denominadas segmentos uniendo las señales encontradas en las distintas capas dentro de cada cámara DT o CSC. De esta manera, como se detallará en las sucesivas secciones, se utilizarán los segmentos que los muones dejan a su paso por las distintas cámaras para el análisis.

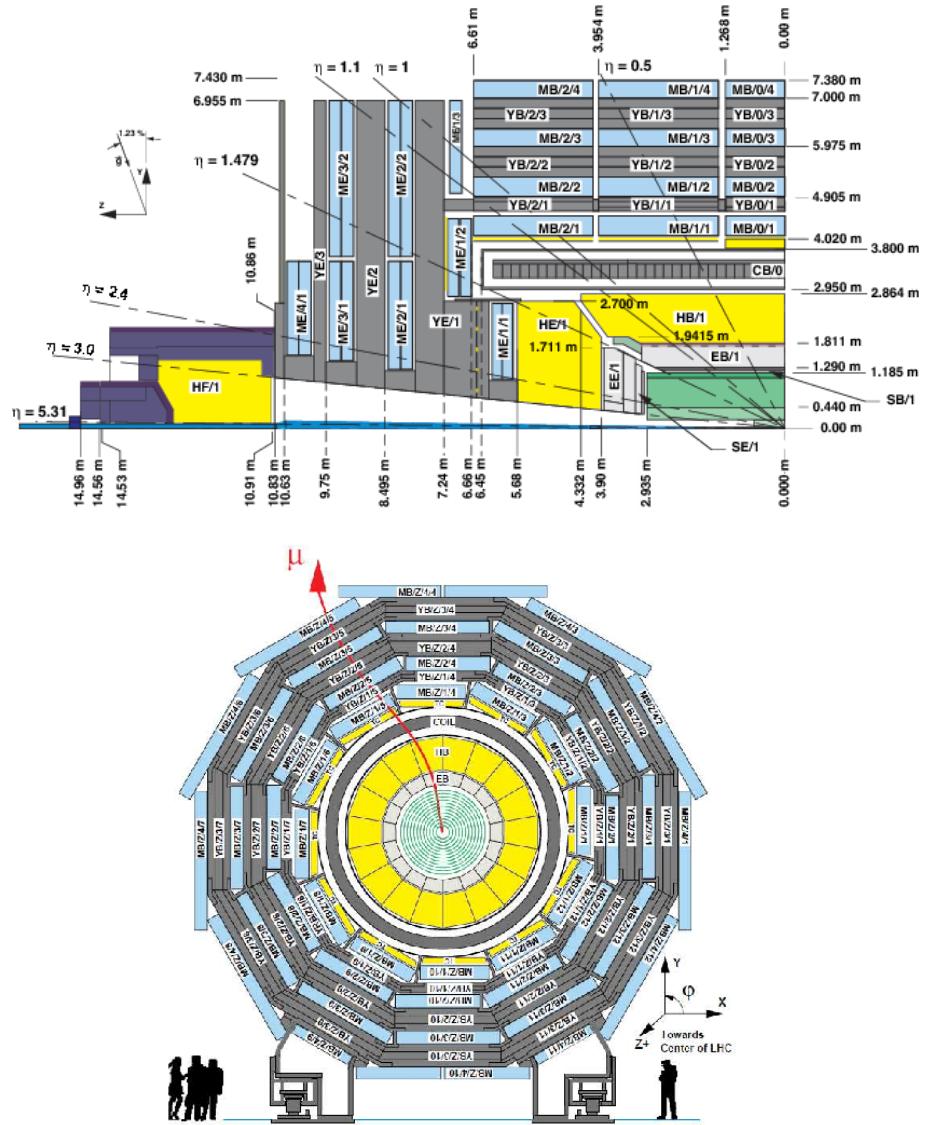


Figura 5: Vista esquemática del detector CMS. Arriba: vista longitudinal de un cuarto del detector. Abajo: vista transversal en $z = 0$. Ambas figuras han sido tomadas de [8].

3. Asignación de momento transverso en el experimento CMS

El momento de un muón *low- p_T* es extraído a partir de la reconstrucción de su trayectoria a través de las señales que deja en el tracker junto con los impactos en el sistema de muones mediante un ajuste por filtro de Kalman [11].

En el caso de muones *high- p_T* , partículas adicionales generadas en las cascadas electromagnéticas dejan típicamente impactos y segmentos adicionales en las distintas cámaras, provocando que estas señales adicionales sean probablemente utilizadas en el ajuste por filtro de Kalman en lugar de las señales que realmente provienen del muón que se quiere medir, o equivalentemente, un número elevado de impactos en una cámara en particular puede dificultar seriamente la propia reconstrucción. Es por esto que los muones de alto momento requieren un tratamiento más cuidadoso de la información encontrada en el sistema de muones, lo que se conoce como reajustes, que seleccionan qué impactos van a utilizarse en el proceso de reconstrucción de la traza y cuáles no.

En las siguientes subsecciones se da una descripción general de los distintos reajustes para muones *high- p_T* que actualmente se llevan a cabo en CMS, así como del algoritmo utilizado para la asignación final del momento transverso.

3.1. Reajuste TPFMS (*Tracker-plus-first-muon-station*)

El primer reajuste, que es el más sencillo en cuanto a su implementación, es el TPFMS, que selecciona únicamente las señales en el tracker y en la estación más interna del sistema de muones que contiene impactos. De esta manera se pretende desechar aquellos impactos de estaciones más lejanas y eleminar así la posible contaminación proveniente de las cascadas electromagnéticas.

3.2. Reajuste Picky

El reajuste Picky tiene como objetivo encontrar posibles cascadas y eliminar sus señales adicionales del ajuste de la traza. De esta manera, si más de n impactos se encuentran dentro de un cono en torno a un impacto concreto, la estación en la que se encuentran dichas señales se marca como contaminada. Así, a la hora de hacer el ajuste a la traza, si su χ^2 está por encima de un cierto umbral, se eliminan del mismo aquellos impactos que se encuentren en cámaras contaminadas y se repite el ajuste de nuevo.

Los parámetros del algoritmo n y χ están optimizados en base a estudios sobre simulaciones.

3.3. Reajuste DYT (*Dynamic truncation*)

Cuando un muón pierde una gran fracción de su energía durante su trayecto, su dirección puede cambiar y las señales encontradas en las siguientes estaciones pueden

ser inconsistentes con la trayectoria inicial (ver Figura 6), provocando problemas a la hora de reconstruir la traza y consecuentemente al tratar de asignar un valor acertado a su momento transverso. En estos casos, el reajuste DYT considera más conveniente parar el filtro de Kalman una vez que un cambio en la trayectoria del muón es detectado.

En el reajuste DYT se define un operador E que da idea de la compatibilidad de un determinado segmento en una cámara con la extrolación de la traza interna a dicha cámara, de manera que si este operador supera un valor predeterminado, el filtro de Kalman se detiene y no tiene en cuenta aquellos impactos en las cámaras posteriores al cambio de trayectoria detectado.

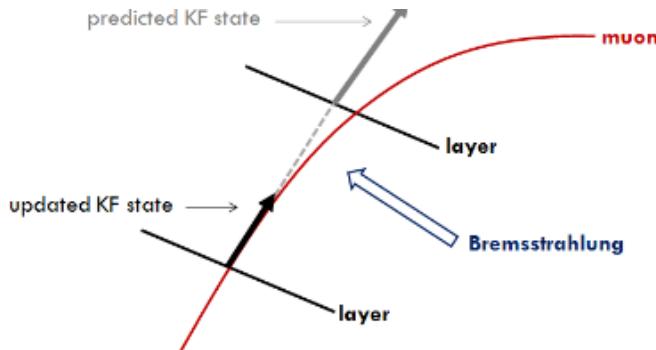


Figura 6: Representación gráfica del cambio en la trayectoria del muón tras una gran pérdida de energía por Bremsstrahlung.

3.4. Asignación final de momento transverso: el algoritmo TuneP

Finalmente, la recomendación central de la colaboración CMS para la asignación del momento transverso de los muones *high- p_T* se corresponde con el proporcionado por el algoritmo TuneP.

El objetivo de este algoritmo es simplemente elegir cuál es la mejor reconstrucción de la traza posible de entre la traza reconstruida únicamente en el tracker y las trazas obtenidas por los distintos reajustes (TPFMS, Picky y DYT). Dicha elección se hace teniendo en cuenta conjuntamente el χ^2/ndof y el σ_{p_T}/p_T de las trazas consideradas.

4. Método propuesto

Dado que el objetivo del trabajo es mejorar la asignación de momento para muones *high-p_T*, especialmente para los casos con alta probabilidad de emisión de cascadas electromagnéticas, el primer aspecto que se ha de tener en cuenta es recopilar toda la información posible del muón (incluyendo los impactos que provienen de las propias cascadas), en contraposición a la estrategia de los distintos reajustes descritos en la Sección 3, que tratan de encontrar aquellas cámaras donde es probable que haya tenido lugar una cascada para después eliminar la información de dichas cámaras al de hacer la reconstrucción de la traza del muón.

De esta manera, el procedimiento que se va a seguir para recoger toda la información posible es extrapolar la traza del muón medida en el tracker interno a las cámaras de muones, para después almacenar todo el conjunto de segmentos que se encuentren en torno al punto extrapolación, siendo capaces así de guardar todas las señales que genere la posible cascada.

En la presente sección se detallarán las herramientas y metodología utilizadas para el tratamiento de los datos.

4.1. Herramientas utilizadas para el análisis

Las herramientas utilizadas en el proceso de análisis de los datos son las siguientes:

- MadGraph [12] es un programa que tiene como objetivo proporcionar todos los elementos necesarios para la simulación de la fenomenología de procesos físicos tanto dentro del Modelo Estándard de la física de partículas como fuera del mismo.

Se utilizará MadGraph para generar una muestra representativa de muones de alto momento para su posterior análisis. Las macroinstrucciones usadas para la producción de la muestra se pueden encontrar en el repositorio [13].

- CMSSW [14] es una colección de software abierto utilizado principalmente para la simulación, calibración, alineamiento, y reconstrucción de objetos físicos para el posterior análisis de los datos.

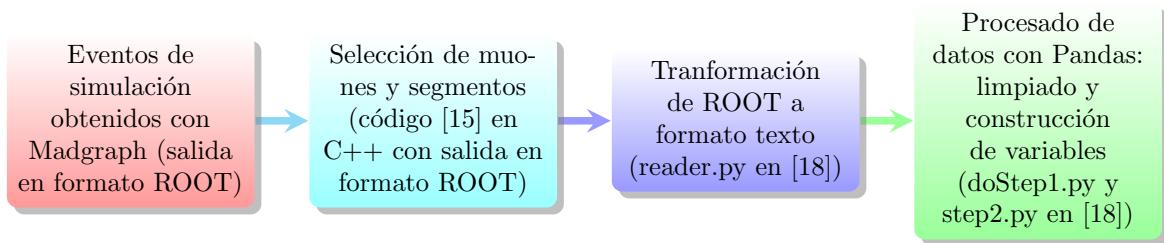
Para este trabajo se ha creado un código en C++ que puede encontrarse en [15]. Este código usa los módulos de CMSSW para realizar la selección apropiada de muones y segmentos que se describirá en la subsección 4.3.

- ROOT [16] es el marco de trabajo más utilizado física de altas energías para el procesado de datos, para análisis estadístico, y proporciona además herramientas de visualización y almacenamiento de los mismos. ROOT está basado en programación orientada a objetos y escrito en C++, y se usará como formato de los datos de entrada y salida del código de selección [15], almacenando la información recogida como objetos de ROOT.

- Pandas [17] (del inglés *Python Data Analysis Library*), es una librería de Python que ofrece gran rendimiento en el manejo de estructuras tabulares de datos.

En este caso, se usará Pandas, previa transformación de los datos en formato ROOT a formato de texto, para el procesado de los segmentos seleccionados (operaciones de limpiado de los datos, agregaciones, construcción de las variables de entrenamiento...etc). Los módulos creados para el procesado pueden encontrarse en [18].

El proceso lógico de trabajo en cuanto a las herramientas utilizadas y al formato de los datos se muestra en el siguiente diagrama de fases:



4.2. Muestra de simulación utilizada

Se han generado un total de 100000 eventos de Z' con $m_{Z'} = 5000$ GeV fruto de la colisiones protón-protón a una energía de centro de masas de 13 TeV¹ mediante simulación de Monte-Carlo utilizando el programa MadGraph5 (ver [13]).

Se impone que las partículas Z' generadas se desintegren a un par de muones de carga opuesta $\mu^+\mu^-$, y se simula el paso de dichos muones por el detector CMS con el paquete Geant4 [19].

Con esta muestra de simulación se tiene una estadística razonable de muones altamente energéticos con p_T generado conocido que se usará para el entrenamiento supervisado de la red neuronal. En la Figura 7 se muestran las distribuciones del p_T generado y el p_T proporcionado por el algoritmo TuneP de todos los muones de la muestra que pasan la selección detallada en la subsección 4.3.

4.3. Selección de muones y segmentos

El proceso de selección de los muones y segmentos (ver código [15]) de la muestra de simulación Monte-Carlo para el posterior análisis consta de las siguientes partes:

1. Lectura de datos: en los datos de entrada, los muones y segmentos reconstruidos por evento se almacenan en colecciones de ROOT, que funcionan a modo de contenedor de información. Por tanto, el primer paso es leer las colecciones para poder iterar sobre ellas.
2. Se recorre la colección de muones y se seleccionan aquellos que cumplen con los requisitos recogidos en la Tabla 1. Esencialmente se pide que los muones tengan traza reconstruida en el tracker con $p_T > 200$ GeV, y que esta traza reconstruida esté próxima a la traza generada del muón dentro de un cono de radio $\Delta R = \sqrt{(\Delta\eta^2 + \Delta\phi^2)} < 0.3$.

¹condiciones actuales de funcionamiento del acelerador LHC

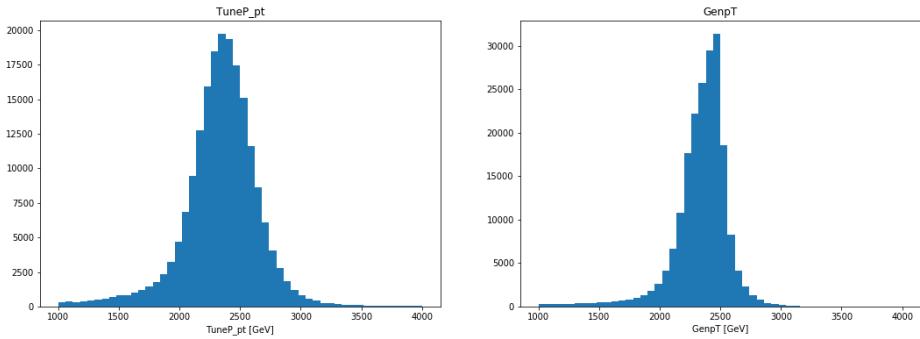


Figura 7: Distribuciones del momento transverso de los muones de la muestra de simulación utilizada. Izquierda: p_T generado. Derecha: p_T proporcionado por el algoritmo TuneP.

| Objeto | Método y selección |
|--------------|--------------------------------|
| TrazaTracker | isNonnull() == 1 |
| TrazaTracker | pt() > 200 |
| MuónReco | $\Delta(\text{MuónGen}) < 0.3$ |

Cuadro 1: Selección aplicada sobre la colección de muones usando los métodos de CMSSW, donde la distancia angular $\Delta R = \sqrt{(\Delta\eta^2 + \Delta\phi^2)}$ se calcula entre la colección de muones reonstruidos y la colección de muones generados.

3. Posteriormente, se recorren las colecciones de segmentos encontradas en el evento, y se almacenan sus coordenadas y los identificadores de los detectores en los que se encuentran si cumplen los requisitos registrados en la Tabla 2. Básicamente, se pide que los segmentos sean válidos², y se hayan medido correctamente sus coordenadas en el detector.

| Objeto | Método y Selección |
|-------------|---------------------------|
| SegmentoDT | isValid()==1 |
| SegmentoDT | hasPhi()==1 |
| SegmentoDT | (hasZ()==1 or station==4) |
| SegmentoCSC | isValid()==1 |

Cuadro 2: Selección aplicada sobre la colección de segmentos usando los métodos de CMSSW.

4. Se selecciona el estado más externo de la traza del tracker y se extrapola (siguiendo la trayectoria de la traza y teniendo en cuenta el campo magnético del solenoide de CMS) a la superficie de cada uno de los detectores registrados donde se ha encontrado al menos un segmento.

²Que son compatibles con la trayectoria inicial, con un χ^2 del ajuste de los impactos que componen el segmento no muy elevado, y además el detector está funcionando correctamente en el momento de la medida

5. Se seleccionan aquellas extrapolaciones que cumplen los cortes de la Tabla 3. En este caso, se pide que las extrapolaciones sean válidas (compatibles con la dirección de la trayectoria inicial), vayan en la dirección del campo magnético impuesto por el solenoide de CMS, y que la distancia entre el centro de la cámara y la extrapolación no exceda el tamaño propio de la cámara. Este último requerimiento es de vital importancia, ya que las superficies a las que se extrae son planos de dimensión infinita (sin delimitar por las dimensiones reales de las cámaras), y una partícula cargada en movimiento sometida a un campo magnético siempre puede cortar un plano de dimensión infinita al curvarse.

| Objeto | Selección |
|-------------|--|
| Propagación | isValid()==1 |
| Propagación | Hacia adelante |
| Propagación | $\text{dist}(\text{centro cámara}) \leq \text{tamaño de cámara}$ |

Cuadro 3: Selección aplicada sobre las propagaciones usando los métodos de CMSSW. “dist” hace referencia a la distancia euclídea.

6. Se guardan todas las variables de interés como objetos de ROOT, es decir, la información sobre la traza del muón en el tracker, las coordenadas de los segmentos recogidos, y las coordenadas de las extrapolaciones.

4.4. Procesado y construcción de variables

Como se ha mencionado en la sección anterior, las propagaciones se hacen a planos de dimensiones infinitas, y a pesar de la selección en distancia de la Tabla 3 aún se tienen algunas extrapolaciones que caen en agujeros entre cámaras o fuera de los límites espaciales de CMS.

En primera aproximación y con el fin de mejorar la calidad y coherencia de los datos, se han definido circunferencias concéntricas en el plano xy, con radio R_{xy} en centímetros, para limpiar la gran mayoría de estas extrapolaciones. De esta manera, una extrapolación es eliminada del conjunto de datos si:

| | |
|---|---------------------|
| $R_{xy} > 810 \text{ cm}$ | si segmento en DTs |
| $R_{xy} > 470 \text{ cm}$ y $R_{xy} < 490 \text{ cm}$ | si segmento en DTs |
| $R_{xy} > 555 \text{ cm}$ y $R_{xy} < 590 \text{ cm}$ | si segmento en DTs |
| $R_{xy} > 670 \text{ cm}$ y $R_{xy} < 690 \text{ cm}$ | si segmento en DTs |
| $R_{xy} > 720 \text{ cm}$ | si segmento en CSCs |

En las Figuras 8 y 9 se muestran las posiciones espaciales de los segmentos y la propagaciones en los planos xy y xz una vez realizado el limpiado.

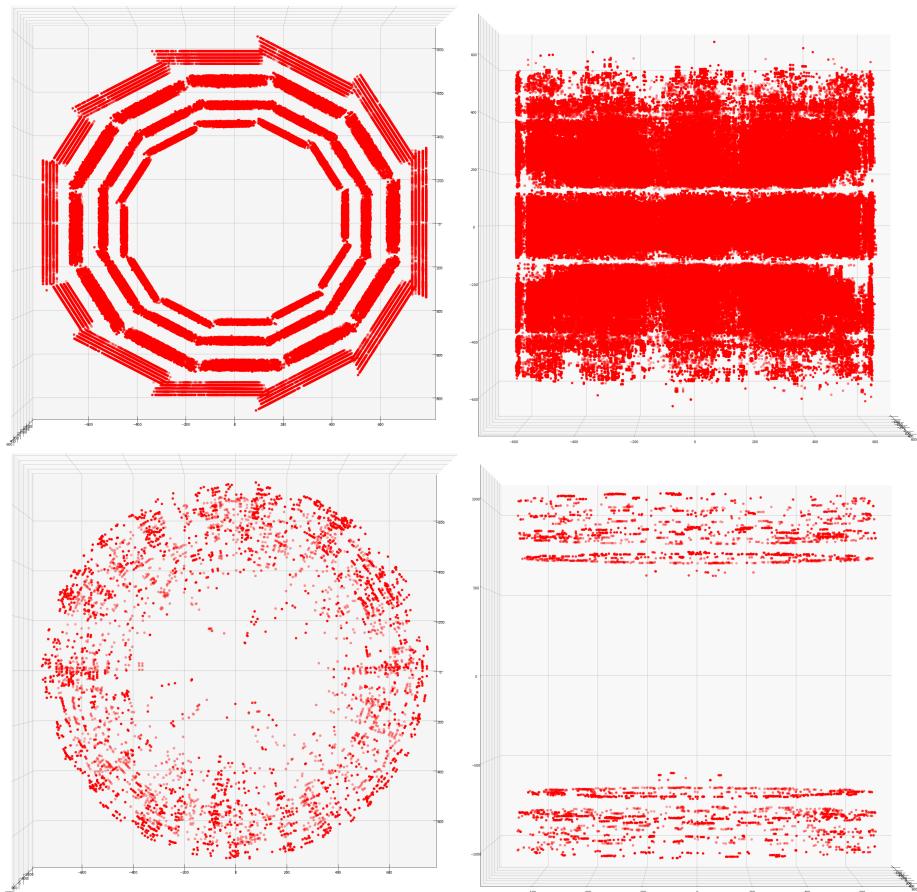


Figura 8: Posiciones geométricas de todos los segmentos seleccionados. Arriba izquierdo: Segmentos en DTs en el plano xy. Arriba derecho: Segmentos en DTs en el plano xz. Abajo izquierdo: Segmentos en CSCs en el plano xy. Abajo derecho: Segmentos en CSCs en el plano xz.

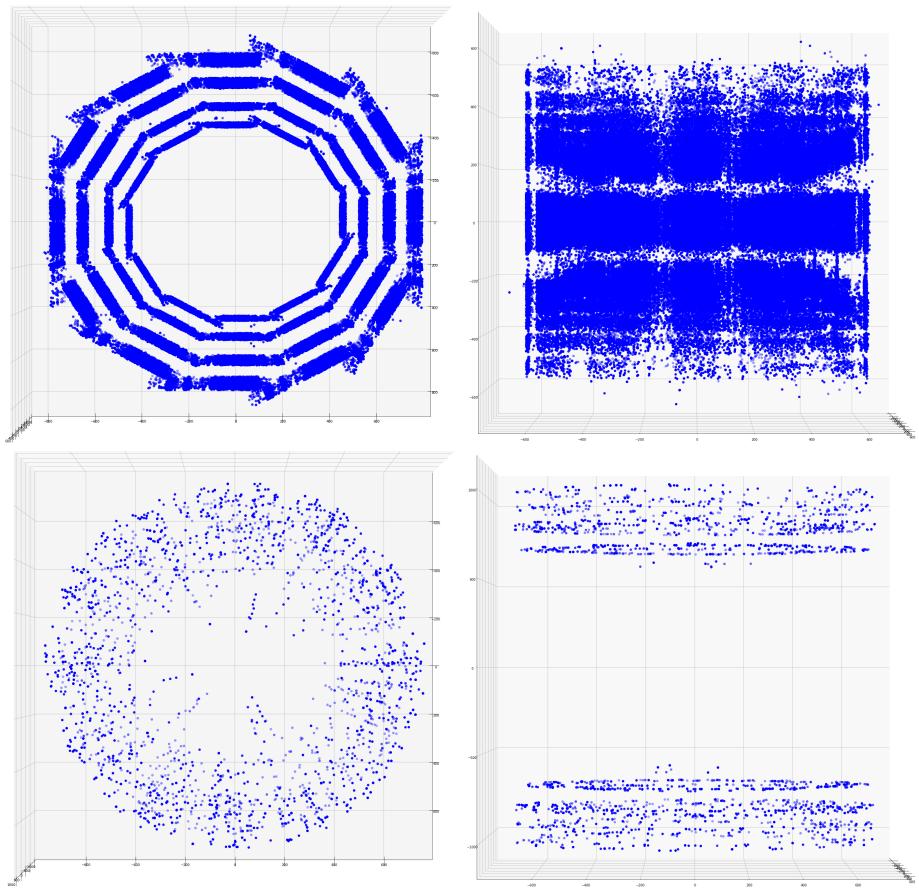


Figura 9: Posiciones geométricas de todos los propagaciones seleccionados. Arriba izquierda: Propagaciones en DTs en el plano xy. Arriba derecha: Propagaciones en DTs en el plano xz. Abajo izquierdo: Propagaciones en CSCs en el plano xy. Abajo derecho: Propagaciones en CSCs en el plano xz.

Para la construcción de las variables que serán utilizadas en el entrenamiento se usan agregaciones de Pandas (ver doStep1.py en [18]). Así, para cada muón se agregan los segmentos encontrados en cada estación de DTs y CSCs y se obtiene el número total de segmentos por estación, la media espacial de los segmentos, la desviación estándar de los mismos, la asimetría y la kurtosis.

4.5. Distribuciones de control

Se han realizado varias comprobaciones para asegurarnos de que los datos que se utilizarán en el entrenamiento son coherentes.

En primer lugar, se seleccionan muones con cuatro o más segmentos en las DTs y además se pide que tengan al menos un segmento en cada estación. Posteriormente, se separan los muones que cumplen estas condiciones en dos categorías: aquellos que tienen exactamente cuatro segmentos (uno en cada estación), y aquellos que tienen más de cuatro segmentos, donde los segmentos adicionales pueden provenir de cascadas electromagnéticas.

De esta manera, si para cada muón se selecciona el máximo valor de distancia entre el segmento y la propagación, uno esperaría que la distribución de valores máximos de distancias pique a valores más bajos en la categoría de cuatro segmentos que en la categoría de más de cuatro segmentos, y que se tenga una cola más alargada en la segunda categoría.

Las distribuciones para ambos grupos se muestran en la Figura 10.

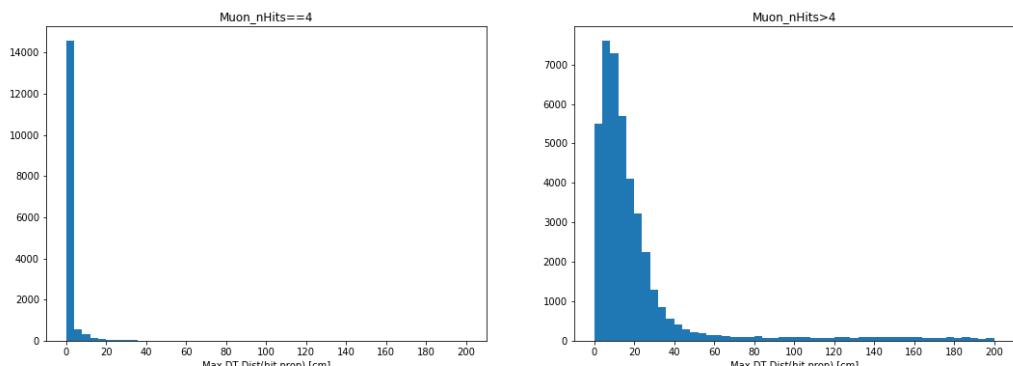


Figura 10: Distribuciones del máximo valor encontrado de distancia entre el segmento y la extrapolación (por muón). Izquierda: Muones con cuatro segmentos en las DTs, uno en cada estación. Derecha: Muones con más de cuatro segmentos en las DTs, y con al menos un segmento por estación.

Por otra parte, en la Figura 11 se muestra la distribución de la media del número de segmentos recogidos por muón en función del p_T de generación. Se observa en este caso que hay una clara tendencia ascendente del promedio del número de segmentos con el p_T , ya que como se indicó en la sección 1, a mayor p_T mayor es la probabilidad de que el muón emita una cascada y se encuentren por tanto más señales en

las cámaras de muones.

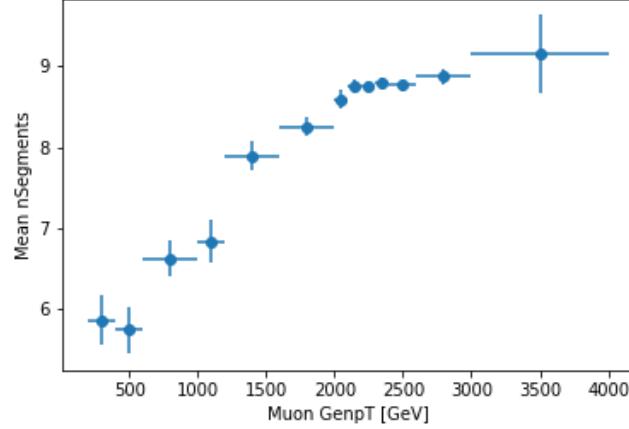


Figura 11: Distribución del promedio de segmentos encontrados por muón en función del momento transverso generado.

Equivalentemente, en la Figura 12 se muestra la distribución de la resolución R por muón en función del p_T de generación, mientras que en la Figura 13 se muestra la distribución bidimensional del momento transverso dado por el algoritmo TuneP en función del momento transverso de generación para cada muón de la muestra.

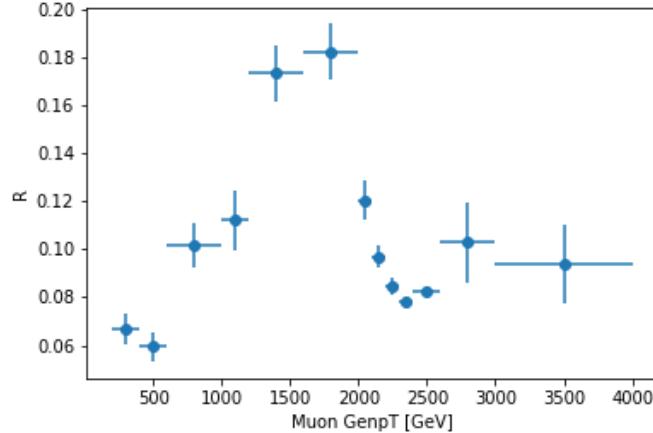


Figura 12: Distribución de la resolución R por muón en función del momento transverso generado.

Por último, teniendo en cuenta que la partícula Z' de la que provienen los muones es muy masiva ($mZ' = 5 \text{ TeV}$) y que esta se produce normalmente en reposo, cabría esperar que los muones fruto de su desintegración viajen en direcciones opuestas. En la Figura 14 se muestra la distancia angular ΔR entre los dos muones de cada

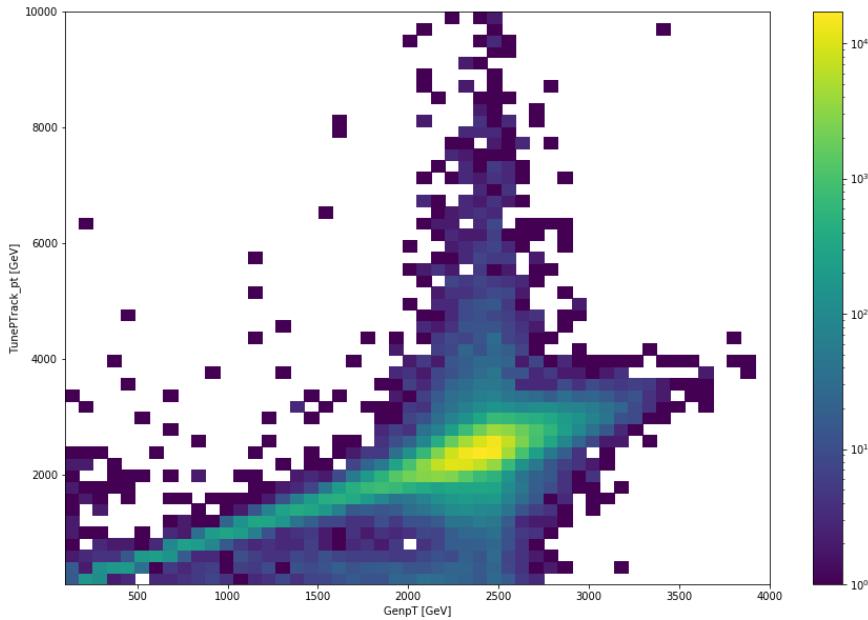


Figura 13: Distribución bidimensional del p_T proporcionado por el algoritmo TuneP en función del p_T de generación para los muones de la muestra.

evento generado.

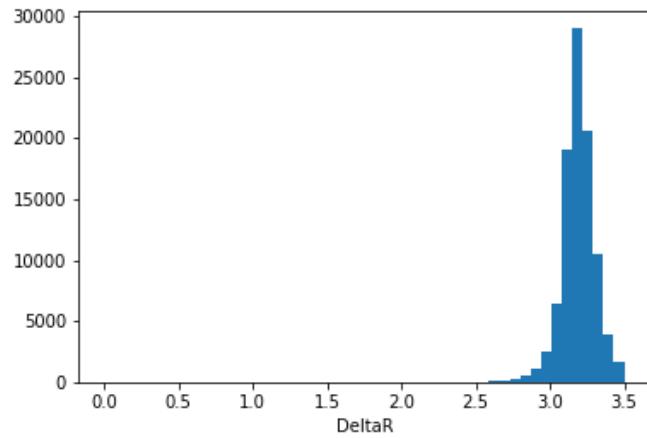


Figura 14: Distribución de la variable $\Delta R = \sqrt{(\Delta\eta^2 + \Delta\phi^2)}$ entre los dos muones de cada evento simulado.

5. Entrenamiento de la DNN

Una vez obtenida toda la información posible del muón a su paso por CMS, el objetivo es entrenar una red neuronal profunda tomando como entrada las propiedades de la traza en el tracker y la información espacial de las señales recogidas en el sistema de muones.

En este caso, la función de pérdida a minimizar será una función que dependa de la diferencia entre el p_T de generación conocido que se quiere predecir y el valor de p_T que devuelve la red, para así aprender las características de los muones (especialmente de aquellos que emitan cascadas) y hacer regresión a su momento transverso.

En cuanto al software, para el entrenamiento se utiliza la librería de Python de código abierto Keras [20], que se caracteriza principalmente por ofrecer sencillez de uso para el usuario, y la red se ejecuta sobre TensorFlow [21].

5.1. Variables de entrenamiento

Las variables utilizadas para el entrenamiento se dividen en dos categorías según su naturaleza:

- Características de la traza interna reconstruida en el tracker: p_T , η , ϕ , carga.
- Información del conjunto de segmentos recogidos en cada estación de DTs y CSCs: número total de segmentos en la estación, media espacial en x, y, z, desviación estándar en x, y, z, asimetría en x, y, z, kurtosis en x, y, z.

En la Figura 15 se muestran las distribuciones de las variables utilizadas en el entrenamiento provenientes de la traza del tracker y de las estaciones de las DTs.

5.2. Conjunto de datos y arquitectura de la red

El conjunto de datos, compuesto por un total de 195777 muones, se ha dividido en un 80 % para el entrenamiento del modelo (del que un 10 % será usado como conjunto de datos de validación), mientras que un 20 % es utilizado para el testeo del mismo.

En cuanto al tipo de red utilizada, se ha entrenado una red de tipo *fully-connected*, donde todos los nodos de cada capa están conectados con los nodos de las capas contiguas, con la siguiente arquitectura:

- Activación: Para las capas 1-9 se utiliza la función de activación ReLu [22], ya que es una de las funciones de activación más apropiadas para problemas de aprendizaje profundo por tener siempre derivada igual a la unidad y así evitar el problema del *vanishing gradient* [23]. En la última capa 10 se utiliza una función de activación lineal para hacer regresión al momento transverso.
- Función de pérdida: MSE (del inglés *Mean Squared Error*), que se define como el promedio de los errores al cuadrado de la siguiente manera:

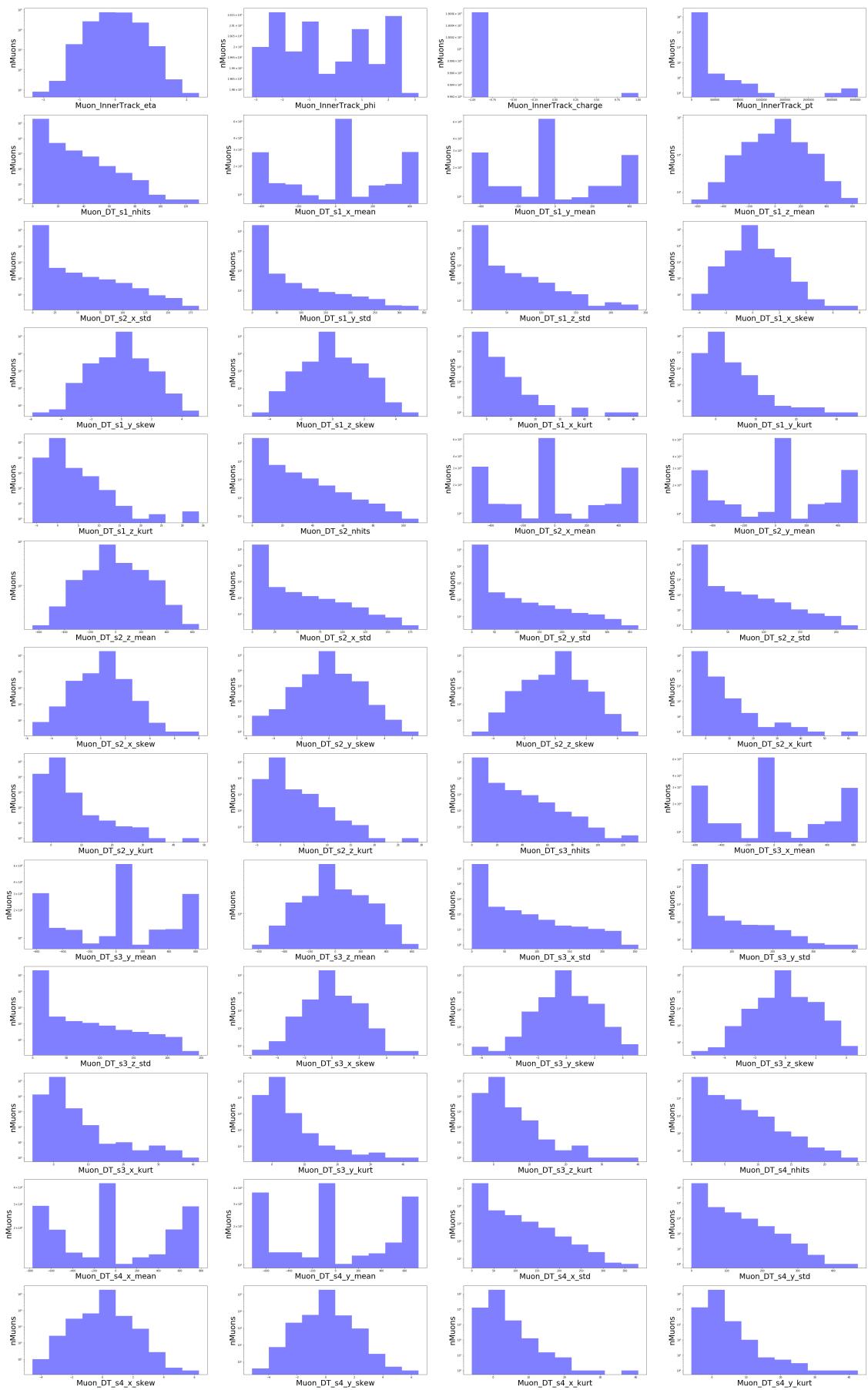


Figura 15: Distribuciones de las variables de entrenamiento

$$MSE = \frac{1}{n} \sum_{i=1}^n \left(p_T^{GEN}_i - p_T^{PRED}_i \right)^2 \quad (6)$$

Donde n se corresponde con el número de muones de la muestra, $p_T^{GEN}_i$ hace referencia al momento transverso generado y conocido del muón i , y $p_T^{PRED}_i$ al momento transverso que predice la red para el mismo.

- Descenso del gradiente: para agilizar el aprendizaje se hace uso de la técnica del *mini-batch gradient descent* [24], que consiste en dividir el conjunto de datos de entrenamiento en fragmentos pequeños denominados *mini-batches*, de tal forma que los parámetros de la red se van actualizando para cada fragmento sin tener que recorrer toda la muestra, mejorando considerablemente la velocidad en el entrenamiento y consiguiendo una convergencia al mínimo de la función de pérdida razonablemente buena. El tamaño del mini-batch elegido ha sido de 2000 muones.
- Como optimizador se usa Adam (del inglés *Adaptive Moment estimation*), que combina el descenso del gradiente con momento [25] y el descenso del gradiente con RMSprop [26], con tasa de aprendizaje = 0.0005 y el resto de hiperparámetros los recomendados por el artículo original de Adam [27].
- Épocas de entrenamiento: 1000.
- Para regularizar se utiliza la técnica de *Early stopping* en la pérdida del conjunto de datos de validación, con una paciencia de 50 épocas. De esta manera, para cada época se obtiene el valor del MSE en el conjunto de datos de validación (que no se utilizan para entrenar), y si el MSE no mejora después de 50 épocas se guarda el modelo encontrado con menor MSE en validación y se detiene abruptamente el entrenamiento para evitar así que el modelo sobreentrene.

El valor de los hiperparámetros se ha tomado tras probar distintas configuraciones aleatorias de los mismos (búsqueda de tipo cuadrícula), y se ha elegido aquella que proporciona un MSE menor en el conjunto de datos de testeo.

5.3. Resultados del entrenamiento

En la Figura 16 se muestra el valor de la pérdida MSE en función de la época de entrenamiento.

El mejor modelo se obtiene después de 184 épocas, con un valor de pérdida para el conjunto de datos de entrenamiento de 49605.3, para el conjunto de datos de validación de 50399.8, y para el conjunto de datos de testeo de 53731.6.

En la Figura 17 se comparan las distribuciones de la resolución R, definida en (4), calculadas sobre el conjunto de datos de testeo a partir del p_T que da el algoritmo TuneP y a partir del p_T que predice la red neuronal respectivamente.

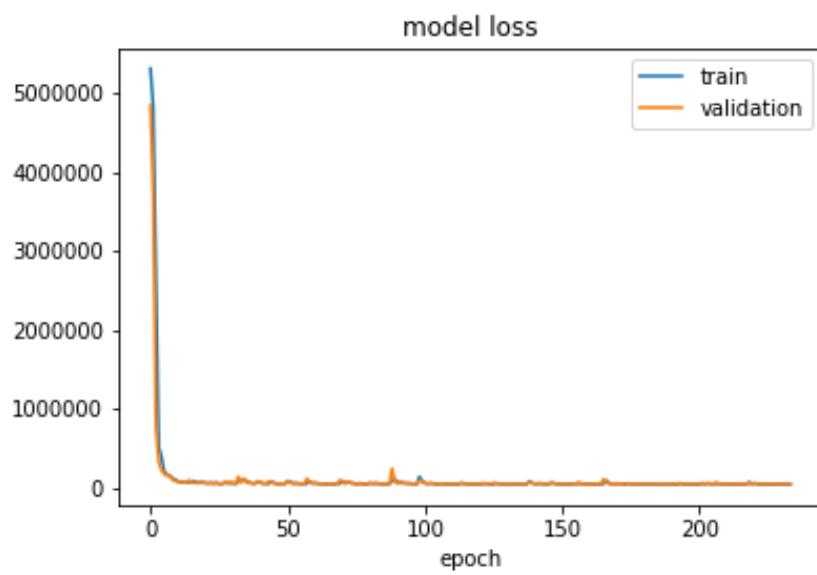


Figura 16: Valor del MSE en función de la época de entrenamiento para el conjunto de datos de entrenamiento (en azul) y para el conjunto de datos de validación (en naranja).

Por último, en la Figura 18 se muestra la distribución bidimensional del momento transverso de predicho por la red en función del momento transverso de generación para los muones del conjunto de datos de testeo.

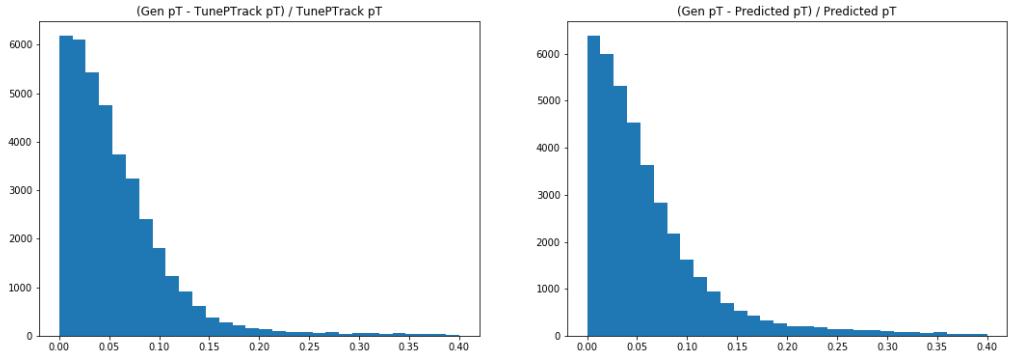


Figura 17: Distribución de la resolución R dada por (4) sobre el conjunto de datos de testeo. Izquierda: tomando el p_T que da el algoritmo TuneP. Derecha: tomando el p_T predicho por el modelo de regresión al p_T .

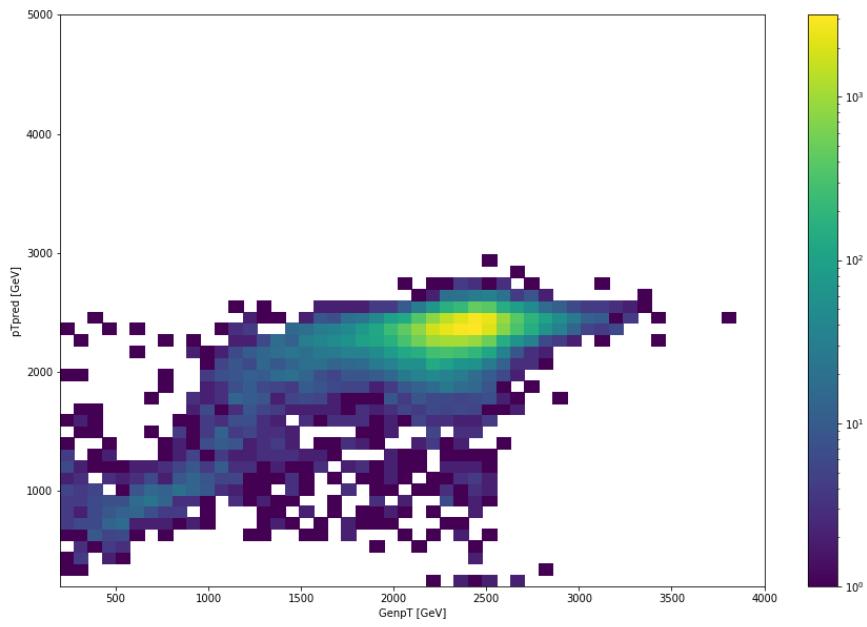


Figura 18: Distribución bidimensional del p_T predicho por la red neuronal en función del p_T de generación para los muones del conjunto de datos de testeo.

6. Resultados

FILL ME

7. Conclusiones

FILL ME

Referencias

- [1] M. et al. Tanabashi. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018.
- [2] Mary K Gaillard, Paul D Grannis, and Frank J Sciulli. The standard model of particle physics. *Reviews of Modern Physics*, 71(2):S96, 1999.
- [3] The CMS Collaboration. Search for a narrow resonance in high-mass dilepton final states in proton-proton collisions using 140 fb^{-1} of data at $\sqrt{s} = 13 \text{ TeV}$. Technical Report CMS-PAS-EXO-19-019, CERN, Geneva, 2019.
- [4] The CMS Collaboration. Search for heavy gauge w bosons in events with an energetic lepton and large missing transverse momentum at $s=13 \text{ tev}$. *Physics Letters B*, 770:278 – 301, 2017.
- [5] M. Tanabashi et al. Review of Particle Physics. *Phys. Rev.*, D98(3):030001, 2018.
- [6] J. Schmidhuber. Deep Learning. *Scholarpedia*, 10(11):32832, 2015. revision #184887.
- [7] Tai Sakuma. Cutaway diagrams of CMS detector. May 2019.
- [8] The CMS Collaboration. Performance of the CMS Drift Tube Chambers with Cosmic Rays. *JINST*, 5:T03015, 2010.
- [9] The CMS Collaboration. Performance of the CMS Cathode Strip Chambers with Cosmic Rays. *JINST*, 5:T03018, 2010.
- [10] R Breedon and D Wenman. Images of the CMS Muons - Cathode Strip Chambers (CSC). CMS Collection., Nov 2008.
- [11] R. E. Kalman. A new approach to linear filtering and prediction problems”transaction of the asme journal of basic. 1960.
- [12] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.
- [13] Repositorio: scripts para la generación de la muestra. https://github.com/fmanteca/Zprime_production. Accessed: 2020-07-27.
- [14] Cmssw. <https://github.com/cms-sw/cmssw>. Accessed: 2020-07-27.
- [15] Código: selección de muones y segmentos. <https://github.com/fmanteca/RECOAnalysis/blob/master/plugins/RECOAnalysis.cc>. Accessed: 2020-07-27.
- [16] Fons Rademakers and René Brun. Root: an object-oriented data analysis framework. *Linux Journal*, 1998:6, 1998.

- [17] Wes McKinney. Data structures for statistical computing in python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 51 – 56, 2010.
- [18] Código: procesado de datos. https://github.com/fmanteca/HighPt_DNN/tree/master/processing. Accessed: 2020-07-27.
- [19] S. Agostinelli et al. GEANT4—a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.
- [20] François Chollet et al. keras, 2015.
- [21] Martín Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [22] Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.
- [23] S. Hochreiter. Untersuchungen zu dynamischen neuronalen Netzen. Diploma thesis, Institut für Informatik, Lehrstuhl Prof. Brauer, Technische Universität München, 1991.
- [24] Michael P Perrone, Haidar Khan, Changhoan Kim, Anastasios Kyrillidis, Jerry Quinn, and Valentina Salapura. Optimal mini-batch size selection for fast gradient descent. *arXiv preprint arXiv:1911.06459*, 2019.
- [25] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural networks : the official journal of the International Neural Network Society*, 12(1):145151, January 1999.
- [26] Tijmen Tieleman and Geoffrey Hinton. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning*, 4(2):26–31, 2012.
- [27] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2015.