



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE POSGRADO EN CIENCIAS FÍSICAS
DEPARTAMENTO DE FÍSICA DE ALTAS ENERGÍAS, INSTITUTO DE CIENCIAS
NUCLEARES

ESTUDIO DE LAS INTERACCIONES PARTÓNICAS MÚLTIPLES
EN COLISIONES PROTÓN-PROTÓN USANDO APRENDIZAJE
DE MÁQUINA

TESIS

QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN CIENCIAS (FÍSICA)

PRESENTA

ERIK ALFREDO ZEPEDA GARCÍA

TUTOR

DR. ANTONIO ORTIZ VELÁSQUEZ
(INSTITUTO DE CIENCIAS NUCLEARES)

COMITÉ TUTOR

DR. LUKAS NELLEN FILLA
(INSTITUTO DE CIENCIAS NUCLEARES)

DR. JOSÉ ALBERTO MARTÍN RUÍZ
(INSTITUTO DE CIENCIAS NUCLEARES)

CIUDAD UNIVERSITARIA, CD.MX., ENERO, 2022



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradecimientos

Esta tesis estuvo bajo la dirección del Dr. Antonio Ortiz Velásquez, a quien agradezco profundamente su apoyo, tiempo, paciencia y compromiso durante todo mi trayecto en la maestría. Agradezco que me involucró en el grupo ALICE del ICN-UNAM pues todos aquellos comentarios y discusiones semanales fueron invaluable en mi formación. Le agradezco el impulsarme y confiar en mí para presentar nuestro trabajo en conferencias internacionales. Sin su conocimiento este trabajo no habría sido posible.

Agradezco a la UNAM y particularmente a la Facultad de Ciencias y al Instituto de Ciencias Nucleares, por su riqueza intelectual y cultural a la que nos dan acceso.

Quiero agradecer especialmente a mis padres, hermana y a mis amigos por su apoyo incondicional, a ustedes les dedico cada página.

Agradezco a los miembros del jurado: Dr. Guy Paić, Dr. José David Vergara Oliver, Dr. Genaro Toledo Sánchez y Dr. Hermes León Vargas, sus comentarios y observaciones han hecho de este un mejor trabajo. A los miembros de mi comité tutor: Dr. Lukas Nellen Filla y Dr. José Alberto Martín Ruíz, por estar atentos con mis estudios.

Agradezco a Eduardo Murrieta y Luciano Díaz, por mantener en buen funcionamiento el cluster del ICN y por su soporte técnico.

Quiero agradecer al CONACyT por el apoyo otorgado por medio del proyecto No. A1-S-22917 y durante mis estudios de maestría por medio de la beca nacional. Agradezco al programa UNAM-PAPIIT por el apoyo correspondiente al proyecto No. IN102118.

Índice general

1	Marco teórico	7
§1.1	El modelo estándar	7
§1.1.1	La cromodinámica cuántica	9
§1.2	Colisiones hadrónicas de alta energía	10
§1.2.1	Interacciones partónicas múltiples	11
§1.3	Generadores de eventos Monte Carlo	14
§1.3.1	Modelos de MPI	17
§1.3.2	Modelos de hadronización	18
§1.3.3	Reconexión por color en PYTHIA	20
2	Un Gran Experimento Colisionador de Iones	22
§2.1	El experimento ALICE	22
§2.2	Cámara de Proyección de Tiempo (TPC)	23
§2.3	Sistema de rastreo interno (ITS)	24
§2.3.1	Detector Pixelar de Silicio (SPD)	24
§2.4	El detector V0	25
§2.5	Descubrimiento de fenómenos nuevos en colisiones protón-protón	26
3	Análisis	28
§3.1	Machine Learning en física de altas energías	28
§3.1.1	Árboles de decisión	29
§3.2	Extracción del número de MPI	30
§3.2.1	Entrenamiento y Prueba	30
§3.2.2	Validación Monte Carlo	32
3.2.2.1	Efecto de la distribución de N_{mpi}	34

§3.2.3	Procesamiento de datos y Toy MC	35
4	Resultados	38
§4.1	Número promedio de MPI	38
§4.2	Dependencia de N_{mpi} con la multiplicidad	39
§4.3	Panorama Futuro	44
5	Conclusiones	48
Apéndice A Árboles de decisión en TMVA		50
Apéndice B Clases de multiplicidad		53

Introducción

Las colisiones de iones pesados, producidas en grandes aceleradores de partículas como el Gran Colisionador de Hadrones (LHC), permiten recrear las condiciones similares del universo temprano en un entorno de laboratorio controlado. En el universo temprano, microsegundos después del Big Bang, existió una fase de la materia llamada Plasma de Quarks y Gluones (QGP), su investigación nos permite entender aspectos fundamentales de la interacción fuerte como el origen del confinamiento.

La expansión y flujo del fuertemente interactuante plasma de quarks y gluones (sQGP) son descritos correctamente por las leyes de la hidrodinámica relativista [1], [2]. Su expansión es causada por gradientes de presión [3], los cuales son originados por la alta diferencia de presión entre el medio creado en la colisión y el vacío en los alrededores. La expansión del sQGP es transmitida a los hadrones a través de un boost por la velocidad local de flujo, desplazándolos a momentos más altos, efecto que incrementa en las partículas más pesadas. Este comportamiento se observa en el intervalo de momento transversal 1-10 GeV/ c .

Sorpresivamente, sistemas de colisión pequeños como protón-protón (pp) y protón-núcleo (p-Pb) revelan características presentes en colisiones de iones pesados. Por ejemplo, en estudios de la producción de partículas en colisiones pp se han observado cambios en la forma de los espectros de momento transversal que se vuelve más pronunciado para partículas pesadas [4]. Asimismo, el cociente protón-pión muestra un pequeño pico cuyo máximo se encuentra alrededor de 3 GeV/ c [5], similar al pico observado en colisiones Pb-Pb por debajo de 8 GeV/ c [6]. Estas similitudes entre sistemas de colisión, ha planteado la cuestión de la formación de un sQGP en sistemas de colisión pequeños [7].

Más allá de los modelos hidrodinámicos que describen cuantitativamente la mayoría de los observables en los datos de sistemas pequeños [8] o los modelos de interacción entre cuerdas que también son capaces de simular estos efectos [9], las Interacciones Partónicas

Múltiples (MPI) ofrecen una alternativa para explicar los fenómenos observados. Por ejemplo, el máximo en el cociente protón-pión observado en colisiones pp es cualitativamente bien reproducido por el generador de eventos PYTHIA. El máximo es atribuido al mecanismo de Reconexión por Color (CR) en eventos con un alto número de MPI [10].

El estudio de los efectos de MPI en colisiones pp, ayudan a un mejor entendimiento del comportamiento similar a iones pesados observado en datos de colisiones pp. En esta dirección, en [11] propusimos la extracción de MPI de los datos de colisiones pp por medio de una regresión basada en Machine Learning (ML) y reportamos el número promedio de interacciones partónicas múltiples $\langle N_{\text{mpi}} \rangle$ para $\sqrt{s} = 5.02$ y 13 TeV.

En esta tesis, determinamos la dependencia en la multiplicidad del número de interacciones partónicas múltiples N_{mpi} de los datos de colisiones pp de ALICE disponibles, así como el $\langle N_{\text{mpi}} \rangle$ para $\sqrt{s} = 5.02, 7$ y 13 TeV [12], extendiendo nuestro trabajo [11].

La tesis se encuentra organizada como sigue. El Capítulo 1 está dedicado al marco teórico de las colisiones hadrónicas de alta energía. Adicionalmente, se describen los aspectos más importantes de los generadores de eventos Monte Carlo (MC), así como los modelos de MPI y hadronización implementados en ellos. El Capítulo 2 presenta brevemente el experimento ALICE, principalmente enfocado en los detectores relevantes para el análisis. El Capítulo 3, comienza introduciendo el método de ML utilizado para el análisis y se proporciona una descripción detallada de la extracción de N_{mpi} de los datos de ALICE. El Capítulo 4 está dedicado a la discusión de resultados y el panorama futuro. Finalmente, las conclusiones se presentan en el Capítulo 5.

Para mayor claridad, en el Apéndice A se detalla el método de ML empleado. El Apéndice B facilita las clases de multiplicidad y la densidad de multiplicidad de partículas cargadas promedio de los datos de ALICE.

Marco teórico

Este capítulo proporciona una breve introducción a la física de colisiones hadrónicas de alta energía. Primeramente se presenta el modelo estándar de física de partículas. Luego se introduce la Cromodinámica Cuántica (QCD, por sus siglas en inglés), teoría de la interacción fuerte. Finalmente se discuten los aspectos más importantes de los generadores de eventos MC relevantes para este trabajo; los modelos de MPI y hadronización, así como el modelo de reconexión por color en PYTHIA.

1.1. El modelo estándar

Todo en nuestro universo se encuentra constituido por un conjunto de bloques básicos conocidos como partículas fundamentales, gobernadas por cuatro interacciones fundamentales: la gravitacional, la electromagnética, la nuclear fuerte y la nuclear débil. Estas tres últimas interacciones, así como su relación con las partículas fundamentales son descritas por el modelo estándar de física de partículas.

Las partículas elementales que describe el Modelo Estándar (ME) son fermiones y están clasificados como quarks: *up* (u), *down* (d), *charm* (c), *strange* (s), *top* (t), *bottom* (b), y leptones: *electrón* (e), *muón* (μ), *tau* (τ). A estas tres últimas partículas las acompañan sus correspondientes neutrinos: *neutrino del electrón* (ν_e), *neutrino del muón* (ν_μ),

neutrino del tau (ν_τ). Para cada una de las anteriores partículas existe una antipartícula. Los leptones y quarks del ME se clasifican en tres generaciones de acuerdo a su masa, por ejemplo, la primera generación está constituida por las partículas menos masivas: el electrón e , su neutrino ν_e y los quarks u y d [13]. Las propiedades de las partículas elementales se resumen en la Tabla 1.1. En la parte bosónica del ME se encuentran las partículas mediadoras de las interacciones o bosones de norma y el *bosón de Higgs* (H).

Generación	Leptones			Quarks		
	Nombre	Masa (MeV/ c^2)	Carga	Nombre	Masa (MeV/ c^2)	Carga
1	e	0.511	-1	u	2	2/3
	ν_e	$<2 \cdot 10^{-6}$	0	d	5	- 1/3
2	μ	106	-1	c	1300	2/3
	ν_μ	<0.19	0	s	100	- 1/3
3	τ	1777	-1	t	173000	2/3
	ν_τ	<18.2	0	b	4200	- 1/3

Tabla 1.1 Partículas elementales descritas por el modelo estándar [14], [15]. La carga eléctrica se encuentra en unidades de la carga del protón, que en unidades del SI es $1.60 \cdot 10^{-19}$ C.

La Electrodinámica Cuántica (QED) es la teoría de la interacción electromagnética entre partículas cargadas mediante el intercambio de *fotones* (γ). La interacción débil cargada es mediada por los bosones de norma masivos W^+ y W^- , mientras que la neutra es mediada por el bosón Z^0 . En el ME, el sector electro-débil proporciona una descripción unificada del electromagnetismo y la interacción débil. Cuyas partículas mediadoras son el fotón y los bosones W^+ , W^- y Z^0 . Estas tres últimas partículas adquieren masa por medio del mecanismo de Higgs.

La Cromodinámica Cuántica es la teoría que describe las interacciones fuertes entre quarks, mediada por los ocho bosones de norma no-masivos llamados *gluones* (g). En esta teoría, los quarks tienen una de las tres *cargas de color* (rojo, verde y azul) análogas a la carga eléctrica en QED. En contraste a QED, en QCD los gluones portan carga de color dando lugar a auto-interacciones que no están presentes en los fotones al ser eléctricamente neutros. La tabla 1.2 resume las propiedades de las interacciones fundamentales.

Propiedad	Interacción gravitacional	Interacción débil	Interacción electromagnética	Interacción fuerte
Partículas que la experimentan	Todas	Quarks y leptones	Eléctricamente cargadas	Quarks y gluones
Partículas mediadoras	Gravitón	W^+ , W^- y Z^0	γ	g
Intensidad	10^{-41}	0.8	1	25

Tabla 1.2 Interacciones fundamentales [14]. Todas las interacciones excepto la gravitacional son descritas por el modelo estándar, el gravitón es el bosón de norma hipotético mediador de la interacción gravitacional. Aquí la intensidad mostrada es relativa a la intensidad de la interacción electromagnética entre dos quarks u separados 10^{-18} m.

1.1.1. La cromodinámica cuántica

La parte del ME encargada de describir la interacción fuerte entre quarks y gluones es la QCD. En la década de los 60's Murray Gell-Man [16] e independientemente George Zweig [17] propusieron la existencia de bloques más fundamentales para explicar la composición de los hadrones, al día de hoy los conocemos como quarks. Sin embargo, la composición completamente simétrica de la partícula Δ^{++} (tres quarks up), violaba el principio de exclusión de Pauli. Para distinguir estos quarks era necesario al menos tres números cuánticos distintos, llamados actualmente carga de color.

La evidencia de la existencia de los gluones surge de la dispersión electrón-nucleón, la cual ocurre vía la interacción electromagnética. A bajas energías, el fotón intercambiado solo ve la carga eléctrica del protón, mientras que al aumentar la energía el fotón puede ver la estructura interna del protón, a este proceso se le conoce como dispersión inelástica profunda. Este proceso indica que las partículas participantes en la dispersión y que interactúan débil y eléctricamente llevan alrededor de la mitad del momento del nucleón. Por tanto, los gluones que solo interactúan fuertemente, son identificados como las partículas que llevan el resto del momento [18].

En QCD la intensidad de la interacción entre quarks y gluones está cuantificada por la

constante de acoplamiento α_S , que en términos de una escala energética μ tiene la forma:

$$\alpha_S(\mu) = \frac{g_s^2}{4\pi} \sim \frac{1}{\ln(\mu/\Lambda_{\text{QCD}})}, \quad (1.1)$$

donde $\Lambda_{\text{QCD}} \sim 250$ MeV es la escala energética típica de QCD. La constante de acoplamiento decrece logarítmicamente a altas escalas energéticas μ , esta propiedad es llamada *libertad asintótica* e implica que los quarks y gluones se comporten asintóticamente como partículas libres. Otra propiedad importante en QCD es el *confinamiento de color*, consecuencia de que α_S crece al disminuir la escala μ y es la razón por la cual no es posible observar quarks libres con carga fraccionaria. En su lugar, los observamos unidos en estados de color neutro (sin carga de color) llamados *hadrones*. Los hadrones están agrupados en mesones (pares quark-antiquark) y bariones (estados de tres quarks). Particularmente, las masas de los hadrones no son exactamente la suma de las masas de los quarks que los componen. Suponiendo quarks no-masivos, debido al confinamiento los hadrones aún tienen masa del orden de Λ_{QCD} ¹, de esta manera la fuerte energía de enlace de los quarks y gluones se manifiesta como la masa de los hadrones.

1.2. Colisiones hadrónicas de alta energía

Una colisión pp altamente energética, es tal que las partículas involucradas en la colisión tienen una energía mucho mayor que su masa en reposo. En el momento de la colisión, la transferencia de momento puede ser tan grande que los quarks y gluones (llamados partones de forma genérica) son los participantes en la colisión, a este tipo de procesos se les llama procesos duros.

En los procesos duros, los partones emergentes son de alto momento transversal que radian generando *lluvias de partones*. El estado final de esta lluvia de partones tiene una estructura de haces localizados colineales o *jets* que llevan consigo la mayor parte de la energía, y por el confinamiento de color, estos jets están compuestos principalmente por hadrones. Otras partículas finales se producen debido al evento subyacente (UE, por sus

¹Note que hadrones como el protón y el neutrón están conformados por quarks u y d cuya masa es inferior a Λ_{QCD} .

siglas en inglés), que se conforma por remanentes del haz e interacciones partónicas múltiples.

La QCD ha sido bastante exitosa en describir los procesos con alta transferencia de momento por medio de QCD perturbativa. Sin embargo, las colisiones hadrónicas de alta energía son dominadas por interacciones suaves de partones. Los eventos en los que la producción de partículas es dominada por procesos suaves son llamados eventos *minimum bias* (MB), los cuales están limitados únicamente por la aceptación de los detectores y pretenden ser lo más inclusivos posible. En colisiones pp, la más inclusiva cantidad medible es la sección eficaz total σ_{tot} .

En colisiones hadrónicas, la sección eficaz total es dividida en interacciones elásticas e inelásticas (INEL). En una interacción elástica ambos protones permanecen intactos y no son producidas nuevas partículas. Las interacciones inelásticas son de especial interés por la producción de partículas, y son clasificadas como difractivas y no-difractivas. En las interacciones difractivas la energía transferida entre los protones es pequeña, sin embargo, sólo uno o ninguno de los protones permanece intacto. El momento transversal de las partículas generadas en colisiones difractivas es mucho más pequeño comparado con los procesos no-difractivos, los cuales se caracterizan por una mayor cantidad de partículas producidas en la región de pseudorapidez central.

Como se ha mencionado, el momento transversal es una cantidad esencial en las colisiones hadrónicas. Una partícula con energía E descrita por su cuadri-momento $p_\mu = (E, p_x, p_y, p_z)$, con p_z el momento a lo largo del eje del haz, tiene un momento transversal definido por

$$p_T = \sqrt{p_x^2 + p_y^2}. \quad (1.2)$$

1.2.1. Interacciones partónicas múltiples

Debido a la naturaleza compuesta de los hadrones, es posible tener MPI, es decir, eventos en los que dos o más distintas interacciones entre partones ocurren simultáneamente en una misma colisión hadrónica. Esto se espera en colisiones hadrónicas altamente energéticas. Intuitivamente la sección eficaz total hadrón-hadrón debería ser mayor que la sección eficaz

partón-partón. Sorpresivamente, a energías del LHC en la Figura 1.1 podemos observar que

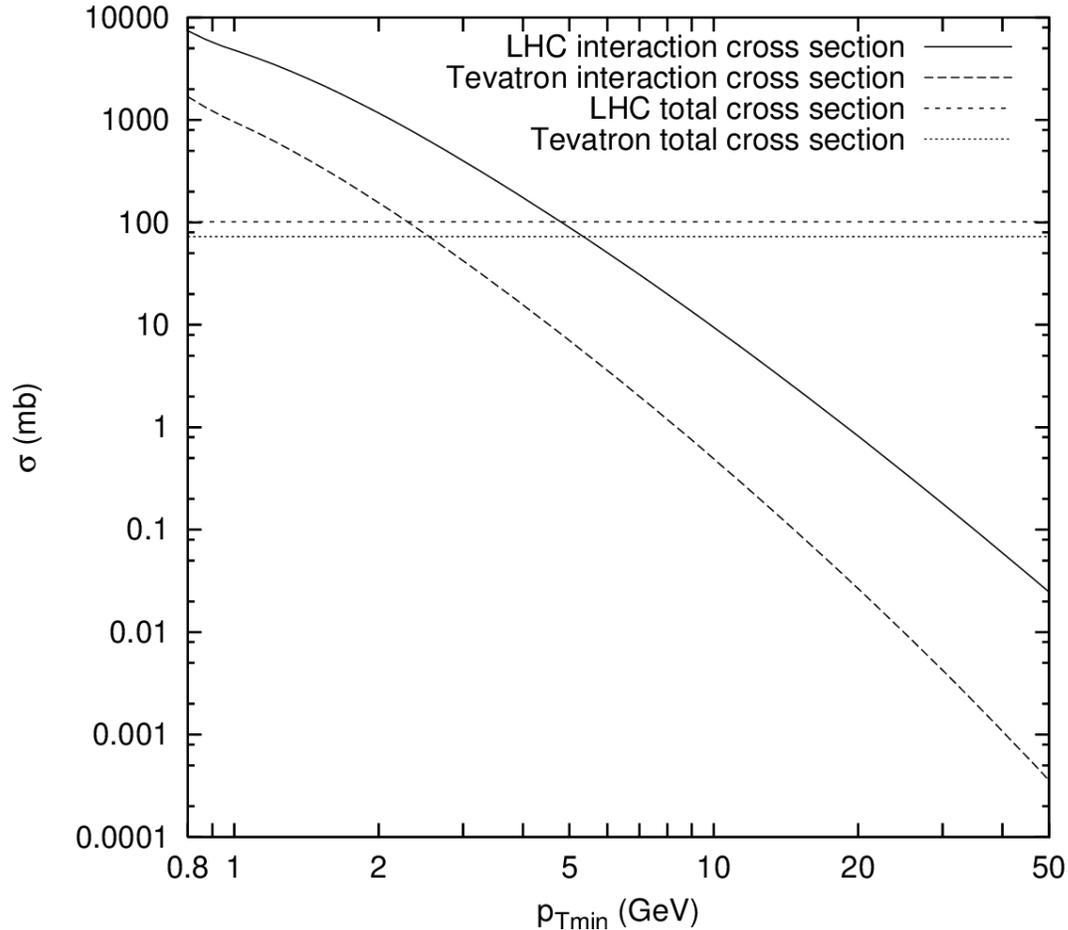


Figura 1.1 Sección eficaz de interacción en función de $p_{T,\min}$ para colisiones protón-antiprotón a 1.8 TeV de Tevatron y para colisiones pp a 14 TeV del LHC. Por comparación, las líneas horizontales representan la respectiva sección eficaz total. Imagen tomada de [19].

la sección eficaz partón-partón también llamada sección eficaz de interacción (entre par de partones) excede la sección eficaz total hadrón-hadrón para valores de $p_{T,\min} \sim 5$ GeV. Esta aparente inconsistencia es resuelta si se supone que cada colisión hadrón-hadrón contiene *varias* colisiones partón-partón, pues si un único evento produce n colisiones partónicas, el evento contribuye n veces a la sección eficaz de interacción.

Los datos sustentan la presencia de MPI en colisiones pp [20]. El estudio de sus efectos en colisiones pp ha atraído recientemente la atención de la comunidad de iones pesados pues proporcionan una alternativa para explicar los fenómenos colectivos observados en

en colisiones pp ha atraído recientemente la atención de la comunidad de iones pesados pues proporcionan una alternativa para explicar los fenómenos colectivos observados en colisiones pp de alta multiplicidad. Un ejemplo de estos fenómenos es en colisiones pp a $\sqrt{s} = 7$ TeV, donde se observa un pequeño máximo en el cociente protón-pión alrededor de $3 \text{ GeV}/c$ [5], similar a lo reportado en colisiones de iones pesados para momento transversal (p_T) menor a $8 \text{ GeV}/c$ [6].

El comportamiento de los espectros generados con CR y MPI es similar a los espectros en colisiones de iones pesados explicados con flujo, esto lo podemos observar en panel superior de la Figura 1.2, la cual muestra el cociente protón-pión en el intervalo $p_T < 6 \text{ GeV}/c$ para simulaciones MB de colisiones pp a $\sqrt{s} = 7$ TeV con PYTHIA 8 Tune 4C. Observamos un pequeño pico en la distribución alrededor de $2.5 \text{ GeV}/c$, cuyo origen se atribuye al mecanismo de CR implementado en PYTHIA. Este pico aumenta si se consideran eventos con un número alto de MPI.

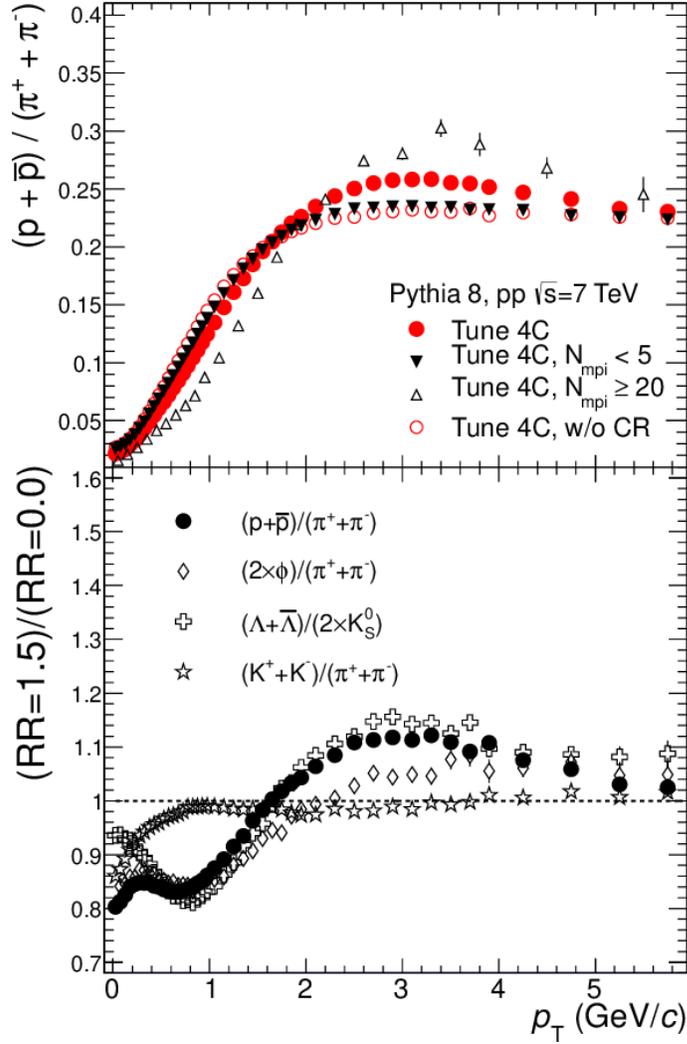


Figura 1.2 Panel superior: Cociente protón-pi3n en funci3n de p_T para colisiones pp simulado con PYTHIA 8. Los eventos con y sin CR corresponden a los c3rculos llenos y c3rculos vac3os respectivamente. Las simulaciones con bajo (tri3ngulos llenos) y alto (tri3ngulos vac3os) N_{mpi} t3ambi3n se muestran. Panel inferior: Cocientes de part3culas en funci3n de p_T para diferentes especies de hadrones. Simulaciones con y sin CR corresponden a $RR=1.5$ y $RR=0.0$ respectivamente. Imagen tomada de [10].

1.3. Generadores de eventos Monte Carlo

El prop3sito de los generadores de eventos Monte Carlo es simular lo que ocurre en una colisi3n de hadrones altamente energ3tica. Estas simulaciones junto con las de detectores proporcionan un estimado realista de la respuesta del detector en la colisi3n, y por tanto ayudan en la planeaci3n de nuevos experimentos, as3 como en crear estrategias para el

análisis de datos.

Dada la complejidad del proceso completo en una colisión de hadrones, en un generador de eventos este proceso se divide en subprocesos que se calculan individualmente, y las probabilidades mecano-cuánticas asociadas a cada resultado son calculadas utilizando métodos Monte Carlo. La estructura de un evento en un generador se puede resumir en los siguientes pasos:

- Inicialmente dos hadrones se encuentran en curso de colisión. Cada hadrón se puede representar como un conjunto de partones.
- Se genera un proceso duro, resultado de la colisión entre dos partones. Las funciones que describen la probabilidad de encontrar un cierto partón que lleva una fracción de la energía del hadrón son las funciones de distribución partónica (PDF).
- En el proceso duro es posible que se generen partículas, como los bosones W^\pm y Z^0 , de muy corta vida media llamadas *resonancias*. Estas y otras emisiones que son asociadas a los partones entrantes son llamadas radiación de estado inicial (ISR).
- Se genera la radiación de estado final (FSR), la cual está asociada con los partones salientes de la colisión.
- Adicionalmente a la colisión entre partones mencionada, debido a la naturaleza compuesta de los hadrones, pueden ocurrir múltiples interacciones entre partones en el mismo evento.
- Cierta fracción de la energía de los hadrones en cuestión es tomada por el proceso duro, y gran parte de la energía restante compone los remanentes del haz.
- Previo a la hadronización, ocurre la reconexión por color, que es controlada por el rango de reconexión RR. Dicho parámetro entra en la que probabilidad de unir partones de bajo y alto momento transversal.
- Debido al confinamiento de color, los partones salientes hadronizan. En PYTHIA el mecanismo de hadronización es por fragmentación de cuerdas.

- Gran parte de los hadrones creados son resonancias y decaen rápidamente. Sin embargo, otros viven lo suficiente para que tanto ellos como sus decaimientos sean detectados.

Existe una gran variedad de generadores de eventos Monte Carlo como PYTHIA 8 [21] y HERWIG 7.2 [22]. Particularmente, opuesto de los modelos hidrodinámicos, en PYTHIA el mecanismo de hadronización está basado en cuerdas. Estos generadores tienen parámetros que pueden ser ajustados para controlar el comportamiento del modelado de los eventos. Un conjunto específico de esos parámetros que se han ajustado para adaptarse mejor a algunos aspectos de los datos es llamado *Tune* o modelo. En PYTHIA y HERWIG se han desarrollado distintos modelos, a continuación se presentan los relevantes para esta tesis:

- PYTHIA Tune 2C [23]: ajustado con los datos de Tevatron ($p\bar{p}$, $\sqrt{s} = 0.3$ hasta 1.96 TeV). Se considera el punto de partida para modelos más sofisticados que incluyan datos del LHC. Este modelo proporciona la peor descripción de los datos del LHC.
- PYTHIA Tune 4C [23]: ajustado con los primeros datos MB y UE del LHC que corresponden a colisiones pp de $\sqrt{s} = 0.9$ y 7 TeV.
- PYTHIA Tune Monash 2013 [24]: ajustado a una gran cantidad de datos del LHC. En contraste con los modelos previos, este modelo comienza con un ajuste más cuidadoso de los datos de LEP².
- HERWIG es un generador que usa el mecanismo de hadronización por clusters, a diferencia de PYTHIA que emplea el mecanismo de fragmentación de cuerdas. En la versión 7 SoftTune de HERWIG: los parámetros del modelo de MPI son ajustados con datos MB del LHC a $\sqrt{s} = 0.9, 7$ y 13 TeV. Este modelo proporcionan una buena descripción de datos MB y UE, desde la más baja energía de Tevatron $\sqrt{s} = 300$ GeV hasta energías del LHC $\sqrt{s} = 13$ TeV [25].

²El gran colisionador electrón-positrón (LEP, por sus siglas en inglés) fue diseñado para proporcionar colisiones electrón-positrón a energías del centro de masa de alrededor de 200 GeV. Funcionó hasta el año 2000 y fue sustituido por el LHC.

Los principales parámetros de los modelos de PYTHIA se resumen en la Tabla 1.3. El parámetro `alphaSvalue` corresponde a la constante de acoplamiento de QCD. Los detalles de los otros parámetros se encuentran en las siguientes secciones.

Parámetro	Tune 2C	Tune 4C	Tune Monash
<code>alphaSvalue</code>	0.135	0.135	0.130
<code>MultipartonInteractions:ecmPow</code> (GeV)	0.21	0.19	0.215
<code>MultipartonInteractions:pt0Ref</code> (GeV)	2.320	2.085	2.28
<code>BeamRemnants:reconnectRange</code>	3.0	1.5	1.8

Tabla 1.3 Principales parámetros de los modelos 2C, 4C y Monash de PYTHIA obtenidos de las referencias [23] y [24] respectivamente.

1.3.1. Modelos de MPI

La evolución de MPI en función de la energía es implementada fenomenológicamente en los generadores de eventos por medio de un corte en el momento trasverso p_{T0} de algunos GeV.

Este corte es introducido para regularizar la sección eficaz partón-partón la cual diverge como $1/p_T^4$. En PYTHIA el corte se introduce por medio del factor $1/(p_T^2 + p_{T0}^2)^2$. La escala p_{T0} evoluciona con la energía del centro de masa mediante una ley de potencia, que en PYTHIA es de la siguiente forma:

$$p_{T0}^2(s) = p_{T0}^2(s_0) \left(\frac{s}{s_0} \right)^b \quad (1.3)$$

donde $\sqrt{s_0} = 7$ TeV es la energía de referencia. Las cantidades $p_{T0}(s_0)$ y b corresponden a los parámetros de la Tabla 1.3, `ecmPow` y `pt0Ref` respectivamente. Un mayor valor en el parámetro b implica un aumento más lento en la actividad hadrónica.

Los parámetros para el modelo de UE en HERWIG 7 controlan la descripción de MPI. El UE en HERWIG 7 es modelado por una combinación de interacciones duras y suaves. Las interacciones de un par de partones salientes con p_T mayor que p_T^{\min} son tratadas como

interacciones duras, de lo contrario son suaves. El umbral p_T^{\min} depende de la energía del centro de masa de la colisión hadrón-hadrón como sigue [26, 27]:

$$p_T^{\min}(s) = p_{T,0}^{\min} \left(\frac{\sqrt{s}}{E_0} \right)^b \quad (1.4)$$

donde $p_{T,0}^{\min}$ es el valor de p_T^{\min} a la energía de referencia $E_0 = 7$ TeV. El parámetro b controla la dependencia en la energía de p_T^{\min} . El valor de los parámetros del modelo de MPI para HERWIG 7.1 SoftTune son: $p_{T,0}^{\min} = 3.502$ y $b = 0.416$.

1.3.2. Modelos de hadronización

En el contexto de los generadores de eventos, hadronización denota el proceso por el cual un conjunto de partones con carga de color es transformado en un conjunto de partículas sin carga de color. En general, los generadores MC realizan este proceso en tres pasos:

1. Mapear el sistema partónico en un continuo de estados hadrónicos masivos (llamados *cuerdas* o *clusters*).
2. Iterativamente mapear estas cuerdas/clusters en un conjunto discreto de hadrones primarios, este proceso es llamado rompimiento de cuerdas/división de clusters o decaimiento de clusters.
3. Finalmente ocurre un decaimiento en partículas secundarias.

El modelo de hadronización en PYTHIA 8 está basado en la fragmentación de cuerdas. El cual se puede explicar considerando un par o dipolo quark-antiquark ($q\bar{q}$), para una separación grande entre los quarks, las líneas de campo se comprimen en una región tubular o cuerda. En este modelo, la tensión en la cuerda aumenta linealmente con la distancia entre los quarks. Finalmente, la cuerda puede romperse mediante la producción de un nuevo par $q'\bar{q}'$, dividiendo el sistema inicial en dos sistemas nuevos $q\bar{q}'$ y $q'\bar{q}$.

HERWIG utiliza el modelo de hadronización por cluster. Al final de la lluvia de partones, todos los gluones se dividen en pares quark-antiquark. Los pares quark-antiquark vecinos forman estructuras de color neutras llamadas clusters. El paso final de este modelo

es el decaimiento del cluster en un par de hadrones. Por ejemplo, para un cluster dado por el par $q_1\bar{q}_2$, un par $q\bar{q}$ es extraído del vacío y se forma el par de hadrones $q_1\bar{q}$ y $q\bar{q}_2$ [27].

Es importante señalar que los generadores PYTHIA 8 y HERWIG 7 describen correctamente los datos MB y UE del LHC. La Figura 1.3 muestra el momento transversal promedio ($\langle p_T \rangle$) en función de la multiplicidad (N_{ch}) para los datos de colisiones pp a $\sqrt{s} = 7$ TeV del detector ATLAS.

Los generadores PYTHIA 8 y HERWIG 7 describen mejor los datos en comparación con las versiones PYTHIA 6 y HERWIG++, así como con el generador de eventos Sherpa.

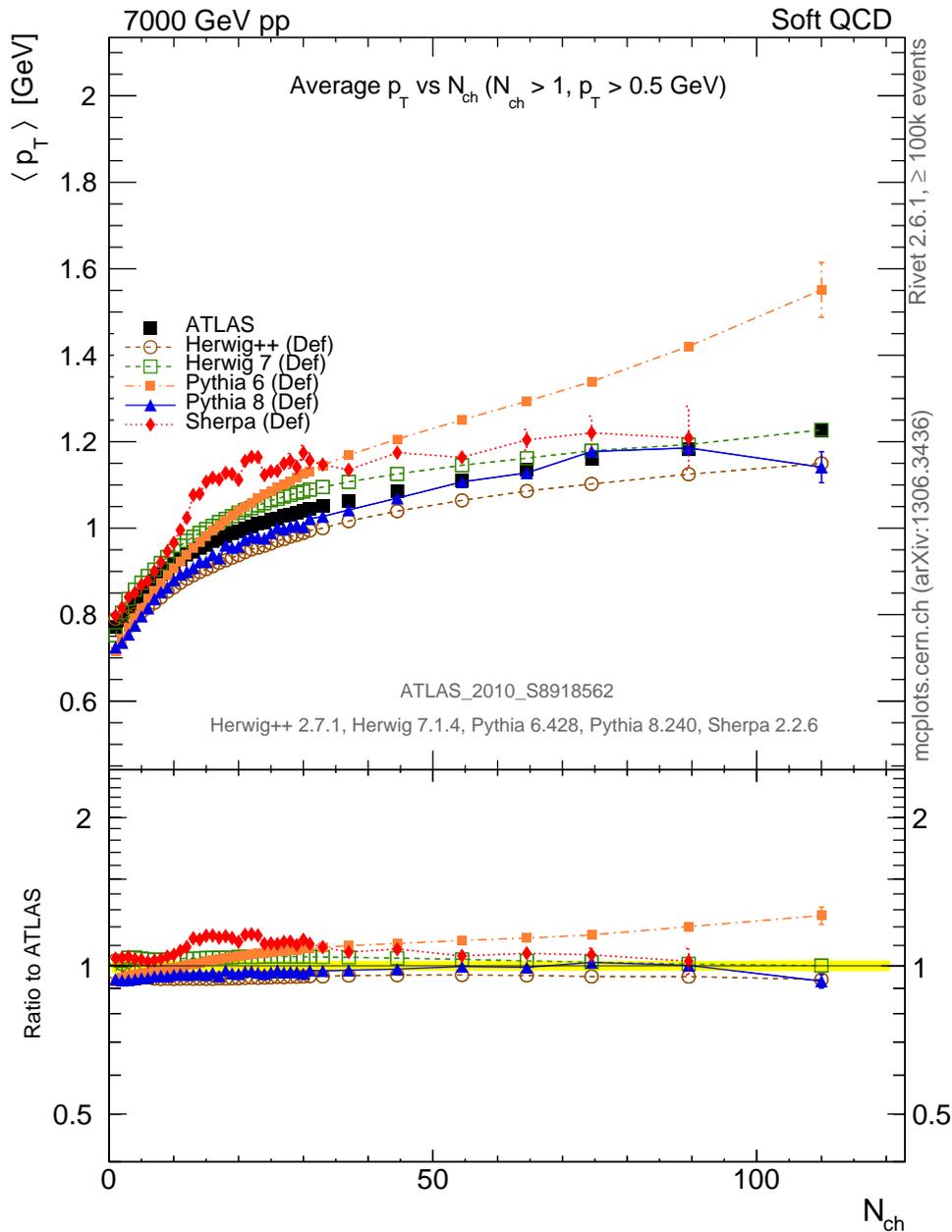


Figura 1.3 Panel superior: Momento transversal promedio en función de la multiplicidad para los datos de colisiones pp a 7 TeV del detector ATLAS (marcadores negros). Los marcadores verdes y azules corresponden a HERWIG 7 y PYTHIA 8 respectivamente. Panel inferior: Cociente entre generadores y los datos. Imagen tomada de [28].

1.3.3. Reconexión por color en PYTHIA

El modelo de CR en PYTHIA está basado en interacciones partónicas múltiples. Empezando con la interacción de menor p_T en un conjunto de MPI, una probabilidad de reconexión es

calculada mediante:

$$P_{rec}(p_T) = \frac{(RR \cdot p_{T0})^2}{(RR \cdot p_{T0})^2 + p_T^2} \quad (1.5)$$

donde RR corresponde al rango de reconexión (`reconnectRange`) de la Tabla 1.3. En la probabilidad de reconexión, los gluones de las interacciones con bajo momento transversal pueden ser añadidos a las cuerdas de los dipolos de las interacciones con alto momento transversal, haciendo que la longitud de la cuerda total se mantenga lo más corta posible. La Figura 1.4 esquematiza CR en el modelo de fragmentación de cuerdas implementado en PYTHIA.

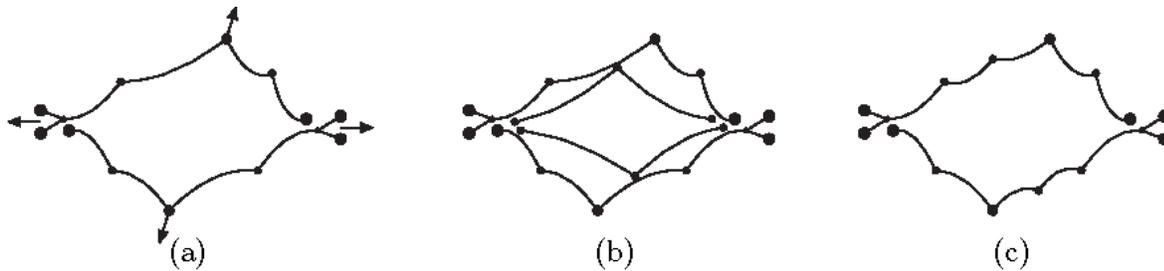


Figura 1.4 (a): La interacción entre partones de mayor p_T se encuentra conectada a los remanentes del haz por medio de una cuerda de color. (b): Una segunda interacción en escala de p_T se produce. (c): Ocurre la reconexión, de tal forma que la longitud de la cuerda total se mantenga mínima. Imagen tomada de [29].

En PYTHIA, el pico en el cociente protón-pión (Figura 1.2) es atribuido a CR. La cuerda que une un par de partones sigue el movimiento de los partones que une. Este movimiento provoca un boost en los hadrones producidos por la fragmentación de la cuerda. Si dos partones procedentes de dispersiones diferentes se reconectan (Figura 1.4), los hadrones producidos por la fragmentación de la cuerda que los une tendrán un mayor boost, por tanto, este efecto aumenta en eventos con un mayor número de MPI.

Un Gran Experimento Colisionador de Iones

El objetivo principal del Gran Experimento Colisionador de Iones (ALICE) es estudiar la materia fuertemente interactuante en condiciones extremas de densidad y temperatura producida en colisiones de iones pesados. El programa de física de ALICE también incluye colisiones pp a las más altas energías del LHC, que recientemente han tomado notoriedad debido al descubrimiento de fenómenos colectivos en colisiones pp de alta multiplicidad.

En esta tesis extraemos el N_{mpi} de los datos disponibles de colisiones pp a $\sqrt{s} = 5.02, 7$ y 13 TeV publicados por ALICE [4] y [30], estos datos utilizan principalmente la información de los detectores V0, SPD y TPC de ALICE. El presente capítulo se centra en su descripción.

2.1. El experimento ALICE

El LHC es el más poderoso acelerador de partículas del mundo, construido por la Organización Europea para la Investigación Nuclear (CERN). Consiste de un anillo superconductor de 27 kilómetros de diámetro, dentro del cual dos haces de partículas viajan a velocidades cercanas a las de la luz para su posterior colisión.

ALICE [31] es uno de los cuatro mayores experimentos del LHC, diseñado para cubrir las condiciones experimentales esperadas en colisiones de iones pesados. El programa de física de ALICE también incluye colisiones protón-núcleo (p-A); así como pp a las más

altas energías del LHC, las cuales complementan el programa de iones pesados.

Los componentes principales del barril central de ALICE son: el sistema de rastreo interno (ITS), la cámara de proyección de tiempo (TPC), el detector de tiempo de vuelo (TOF) y el detector de identificación de partículas de alto momento transversal (HMPID). El detector ALICE también cuenta con un espectrómetro delantero (forward) para medir muones. Una explicación detallada del detector ALICE se encuentra en las referencias [32] y [33].

Cabe mencionar que el sistema de coordenadas global de ALICE, es un sistema cartesiano derecho cuyo origen se encuentra en el punto nominal de interacción. El eje z es paralelo al eje del haz del LHC. El eje x apunta hacia el centro del anillo del LHC y el eje y se elige para definir un sistema derecho, por tanto apunta hacia arriba. De especial interés es el ángulo de emisión de una partícula θ pues se relaciona con la pseudorapidez, definida como:

$$\eta \equiv -\ln\left(\tan\left(\frac{\theta}{2}\right)\right), \quad (2.1)$$

el ángulo θ incrementa del eje z hacia el plano xy .

2.2. Cámara de Proyección de Tiempo (TPC)

La cámara de proyección de tiempo es el principal dispositivo de rastreo de ALICE y está optimizado para proporcionar mediciones de momento de partículas cargadas e identificación de partículas. Cubre el rango de pseudorapidez $|\eta| < 0.9$ y tiene la capacidad de rastrear partículas de momento transversal de alrededor de $0.15 \text{ GeV}/c$ hasta $100 \text{ GeV}/c$.

El detector TPC tiene forma cilíndrica, con un tamaño radial y longitudinal de aproximadamente $85 \text{ cm} < r < 250 \text{ cm}$ y $-250 \text{ cm} < z < 250 \text{ cm}$ respectivamente, y está lleno de una mezcla de gas $\text{Ne-CO}_2\text{-N}_2$ (85.7 %-9.5 %-4.8 %). Los detectores TPC e ITS se encuentran alineados uno respecto al otro, cuya información combinada permite determinar el momento de partículas cargadas en el rango de $115 \text{ MeV}/c < p_T < 100 \text{ GeV}/c$ con una resolución de 1-10 %.

2.3. Sistema de rastreo interno (ITS)

El sistema de rastreo interno consiste de seis capas cilíndricas de detectores de silicio coaxiales al tubo del haz (ver Figura 2.1), ubicado en radios entre 4 y 43 cm. Las dos capas internas son detectores pixelares de silicio (SPD), las dos capas centrales son detectores de arrastre de silicio (SDD) y las dos capas externas son detectores de tiras de silicio (SSD).

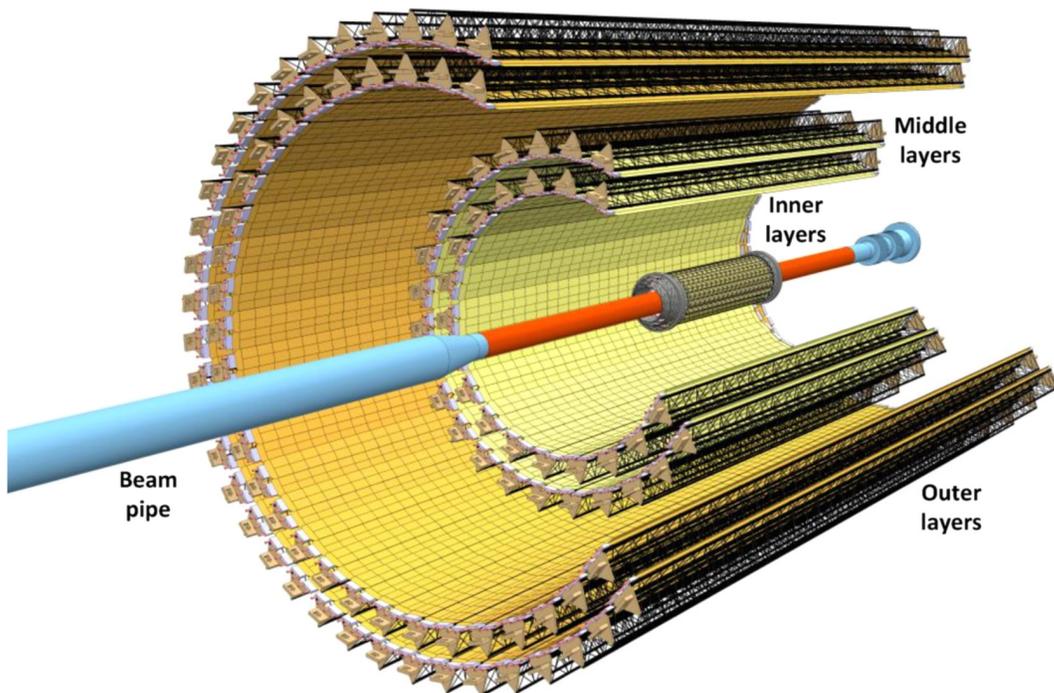


Figura 2.1 Disposición del ITS. Imagen tomada de [34].

El detector ITS tiene como funciones principales: determinar el vértice primario con una resolución mejor que 100 micrómetros, rastrear e identificar partículas de momento menor que 200 MeV/c y mejorar la resolución angular y de momento para partículas reconstruidas por el detector TPC.

2.3.1. Detector Pixelar de Silicio (SPD)

El detector pixelar de silicio se encuentra posicionado a 5 mm del tubo del haz, las capas interna y externa tienen un radio de 3.9 cm y 7.8 cm respectivamente, cubriendo rangos de

pseudorapidez de $|\eta| < 2$ y $|\eta| < 1.4$ respectivamente. Las dos capas de estos detectores que se encuentran en el detector ITS son fundamentales para la determinación de la posición del vértice primario.

El módulo detector básico del SPD consiste en dos submódulos (*ladders*) cada uno con 1200 chips de lectura y 256×160 celdas. En total, el detector SPD incluye 240 ladders y un total de 9.8 millones de celdas. Las capas del detector SPD son las más internas debido a su granularidad, ideal para la alta densidad de partículas esperadas en colisiones de iones pesados.

El detector SPD también es utilizado para reconstruir tracklets, que son segmentos de traza construidos usando la posición del vértice primario reconstruido y dos señales, una en cada capa del detector SPD. El número de tracklets proporciona un excelente estimado de la multiplicidad de partículas cargadas a pseudorapidez media $|\eta| < 0.8$.

2.4. El detector V0

Los detalles técnicos del detector se encuentran en [35]. El detector V0 consiste de dos arreglos de contadores de centelleo llamados V0A y V0C, segmentados en 32 contadores individuales distribuidos en cuatro anillos (ver figura 2.2). Los detectores V0A y V0C están localizados en cada lado del punto de interacción y cubren los rangos de pseudorapidez $2.8 < \eta < 5.1$ y $-3.7 < \eta < -1.7$ respectivamente. El detector V0 principalmente aporta la señal de disparo (trigger) para que los otros dispositivos reconstruyan al evento. También es usado para medir propiedades globales de eventos en colisiones pp y p-Pb, como la multiplicidad de partículas cargadas. En colisiones pp, la eficiencia para la detección de al menos una partícula cargada en ambos arreglos es alrededor de 75 %.

Se le llama estimador V0M a aquel que está basado en la carga total depositada en el detector que cubre las regiones de pseudorapidez de los detectores V0A y V0C. También es llamado estimador de multiplicidad delantera y es comúnmente usado por la colaboración ALICE para minimizar las posibles autocorrelaciones inducidas por el estimador SPD.

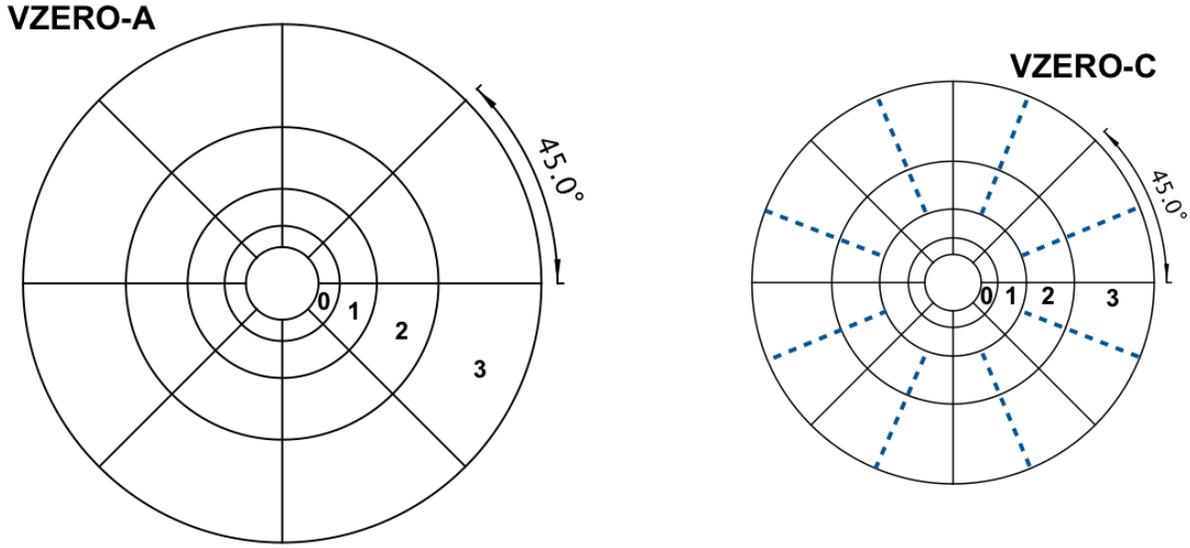


Figura 2.2 Esquemas de los detectores V0A y V0C. Imagen tomada de [36].

2.5. Descubrimiento de fenómenos nuevos en colisiones protón-protón

Una de las evidencias clave de la formación del fuertemente interactuante sQGP en colisiones de iones pesados son los fenómenos de colectividad [37]. El incremento de partículas extrañas [38] o las correlaciones azimutales de dos partículas [39], en colisiones pp de alta multiplicidad, son ejemplos de mediciones que presentan características previamente observadas en colisiones Pb-Pb.

Las correlaciones azimutales de dos partículas de largo alcance, son parametrizadas por medio del ángulo azimutal relativo de las partículas $\Delta\phi$ y la pseudorapidez relativa $\Delta\eta$. Una característica clave de tales correlaciones es la formación de una estructura alrededor de $|\Delta\phi| = 0$ que se extiende sobre un amplio rango de pseudorapidez relativa ($|\Delta\eta| \approx 4$). Esta estructura es conocida como *cresta*, y fue primeramente observada en colisiones núcleo-núcleo (AA) desde las energías de RHIC a las del LHC, incluyendo sistemas Pb-Pb [40]. Extensos estudios sugieren que el flujo colectivo hidrodinámico de un medio fuertemente interactuante y en expansión [1, 41] es responsable de las correlaciones de largo alcance en

sistemas grandes de iones pesados. Sin embargo, en colisiones pp y p-Pb, donde la región de traslape es comparable con el tamaño del protón, la formación de un QGP no es esperada.

Se piensa que el aumento en la producción de quarks y antiquarks extraños (s y \bar{s}) está relacionada con la creación del QGP, pues al no estar presentes en la materia ordinaria estos son 'recién creados' a partir de la energía cinética de núcleos en colisión. Asimismo, la masa de los quarks s y \bar{s} es del orden de la temperatura o energía a la que los protones y neutrones se disuelven en quarks, por tanto, la abundancia de estos quarks es sensible a las condiciones de desconfinamiento. La producción de pares $s\bar{s}$ favorece la eventual producción de hadrones extraños. El incremento de extrañeza se ha observado en colisiones Pb-Pb a $\sqrt{s_{NN}} = 2.76$ TeV [42].

Sorpresivamente, se ha reportado la observación de un incremento de partículas extrañas en colisiones pp de alta multiplicidad, donde las mediciones están en notable acuerdo con los resultados de colisiones p-Pb. En estos eventos, la producción de partículas extrañas alcanza valores similares a aquellos observados en colisiones Pb-Pb.

Un excelente review de fenómenos colectivos en colisiones pp y p-Pb se encuentra en la referencia [43].

Análisis

Como hemos señalado en los anteriores capítulos, las colisiones pp de alta multiplicidad son interesantes debido al descubrimiento de fenómenos colectivos previamente observados en colisiones de iones pesados. Más allá de los modelos hidrodinámicos, cuya aplicabilidad a sistemas pequeños (pp y p-Pb) puede ponerse en discusión debido a la clara diferencia de tamaño con sistemas grandes (Pb-Pb). Las interacciones partónicas múltiples, proveen una alternativa para explicar el origen de estos fenómenos. En particular, el ya mencionado pico en el cociente protón-pión con máximo en $3 \text{ GeV}/c$.

La extracción del número de interacciones partónicas múltiples de datos de colisiones pp aporta información sobre la presencia de MPI en estos eventos de alta multiplicidad.

En este capítulo se detalla la estrategia para la extracción del N_{mpi} de los datos de colisiones pp de ALICE e introducimos la técnica de aprendizaje de máquina empleada para el análisis.

3.1. Machine Learning en física de altas energías

Verificar el ME y la búsqueda de nueva física son dos de los objetivos principales que tiene la física de altas energías experimental. El tamaño de los eventos, el volumen de datos y complejidad generados por experimentos como el LHC representan un reto en la forma tradicional de analizar los datos, la cual se basa en un conocimiento profundo del problema

e implementación en un software para su posterior aplicación. El aprendizaje de máquina o *Machine Learning*, busca que la computadora aprenda y genere un modelo capaz de generalizar e inferir comportamientos en un conjunto de datos, en lugar de que el humano desarrolle un algoritmo altamente especializado.

El análisis de datos en Física de Altas Energías (HEP), en ocasiones requiere la identificación de señales raras inmersas en ruido. El descubrimiento del bosón de Higgs en el 2012 por las colaboraciones ATLAS y CMS vio el primer uso de una técnica de ML para separar pequeñas señales sobre gran cantidad de ruido [44].

Las técnicas de ML han sido aplicadas a una gran cantidad de problemas en HEP. Por ejemplo, el lector puede encontrar en [45] una lista de publicaciones recientes donde se desarrollan y aplican técnicas de ML en HEP.

3.1.1. Árboles de decisión

El método de ML empleado en esta tesis son los *árboles de decisión impulsados* (Boosted Decision Trees, BDT). Un árbol de decisión tiene una estructura de árbol binario en el cual decisiones son tomadas sobre una variable, de esta forma el espacio fase se divide en distintas regiones cuyos nodos de salida son eventualmente clasificados como señal o ruido. En el caso de un problema de regresión, cada nodo de salida representa un valor específico de la variable objetivo. El término impulsados, extiende el concepto de un árbol de decisión a muchos árboles para formar un bosque [46].

Los BDT se encuentran en el kit de herramientas para el análisis multivariado (TMVA) integrado en el software para análisis de datos científicos ROOT [47]. TMVA proporciona un entorno para el procesamiento, evaluación y aplicación de técnicas multivariadas de regresión y clasificación. En TMVA un típico análisis de regresión o clasificación consiste de dos fases independientes: la fase de entrenamiento, donde los métodos son entrenados, probados y evaluados, y una fase de aplicación, donde los métodos elegidos son aplicados a un problema en concreto de regresión o clasificación. Las técnicas de ML en TMVA pertenecen a la familia de *aprendizaje supervisado*, pues los algoritmos utilizan eventos de entrenamiento en los cuales el resultado deseado es conocido. TMVA está específicamente

diseñado para su aplicación en HEP. Todos los detalles técnicos de TMVA se encuentran en [46].

En el Apéndice A se encuentra la descripción de los BDT en TMVA, así como de las fases de entrenamiento y aplicación.

3.2. Extracción del número de MPI

En general, una regresión es el proceso en el que se estima los parámetros de una función, la cual predice el valor de una variable de respuesta u objetivo en términos de los valores de otras variables (las variables de entrada).

Siguiendo el enfoque que reportamos en [11], la extracción del N_{mpi} se puede considerar como un problema de regresión donde dado un conjunto de variables de entrada se intenta minimizar una función de pérdida. La función de pérdida mide la desviación entre el modelo y la variable objetivo (para mayor detalle ver el Apéndice A). En esta tesis la variable objetivo es el N_{mpi} .

La estrategia para la extracción del N_{mpi} se resume en los siguientes pasos:

1. Entrenamiento y Prueba
2. Validación Monte Carlo
3. Procesamiento de datos

3.2.1. Entrenamiento y Prueba

Para el entrenamiento de los BDT utilizamos eventos simulados de colisiones pp a $\sqrt{s} = 13$ TeV generados por medio de PYTHIA 8 Tune 4C. Al tratarse de un problema de regresión, en este paso se define la variable objetivo para los datos de entrenamiento.

La elección de las variables de entrada se basa en su correlación con N_{mpi} . Estas variables son:

- La multiplicidad de partículas cargadas a pseudorapidez media (N_{ch}).
- El momento transversal promedio ($\langle p_T \rangle$).

En una colisión de hadrones central donde el área de traslape es del tamaño del hadrón, la probabilidad de tener un mayor número de MPI aumenta. Un incremento en MPI se traduce en un mayor número de mini-jets que fragmentan en partículas finales. En particular, los eventos de baja multiplicidad se caracterizan por un bajo número de MPI y con bajo momento transversal [48].

Entrenamos dos conjuntos de BDT. Para partículas cargadas primarias¹ en los rangos de pseudorapidez $|\eta| < 0.8$ y $|\eta| < 0.5$, de acuerdo a las restricciones cinemáticas de los datos de ALICE. En ambos casos el $\langle p_T \rangle$ considera partículas con $p_T > 0.15$ GeV/c.

En TMVA una parte de los eventos simulados son utilizados para probar los métodos. Al finalizar el entrenamiento y prueba, TMVA proporciona por defecto una variedad de gráficos de control y rendimiento. Por ejemplo, las Figuras 3.1 y 3.2 muestran la desviación entre la variable objetivo (N_{mpi}) obtenida por la regresión y el valor verdadero, en función de las variables de entrada ($\langle p_T \rangle$ y N_{ch}). En ambas Figuras, la desviación $N_{\text{mpi}}^{\text{regression}} - N_{\text{mpi}}^{\text{true}}$ es muy cercana a cero en la mayoría de eventos, tanto en el entrenamiento como en la prueba.

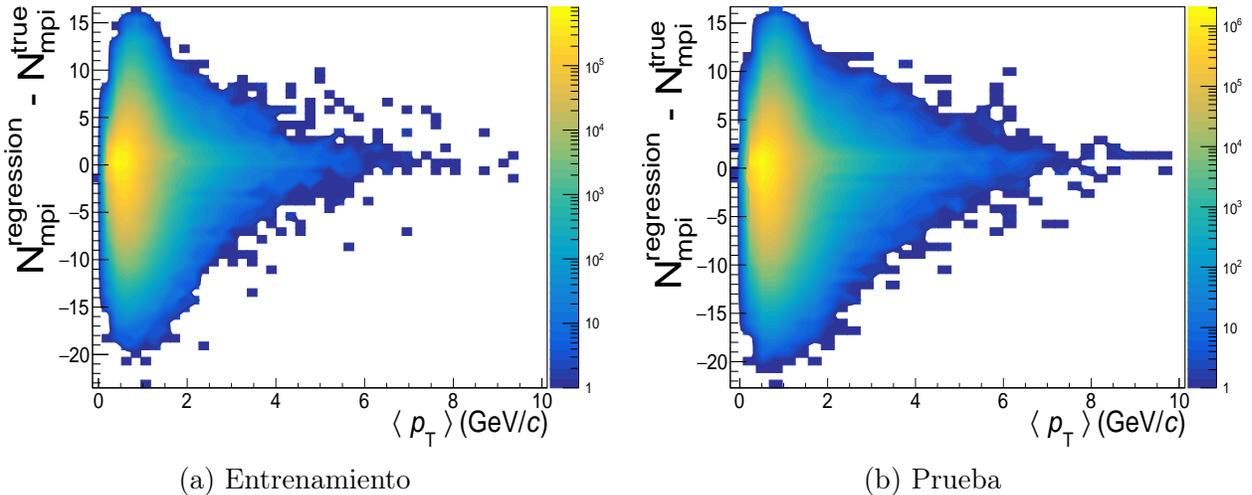


Figura 3.1 Desviación entre la variable objetivo obtenida por regresión y el valor verdadero, en función del momento transversal promedio. Las figuras izquierda y derecha corresponden a resultados de entrenamiento y prueba respectivamente.

¹Una partícula cargada primaria es definida como una partícula cargada con tiempo de vida promedio (τ) mayor que 1 cm/c, la cual es producida directamente en la interacción o por el decaimiento de partículas con τ menor que 1 cm/c [49].

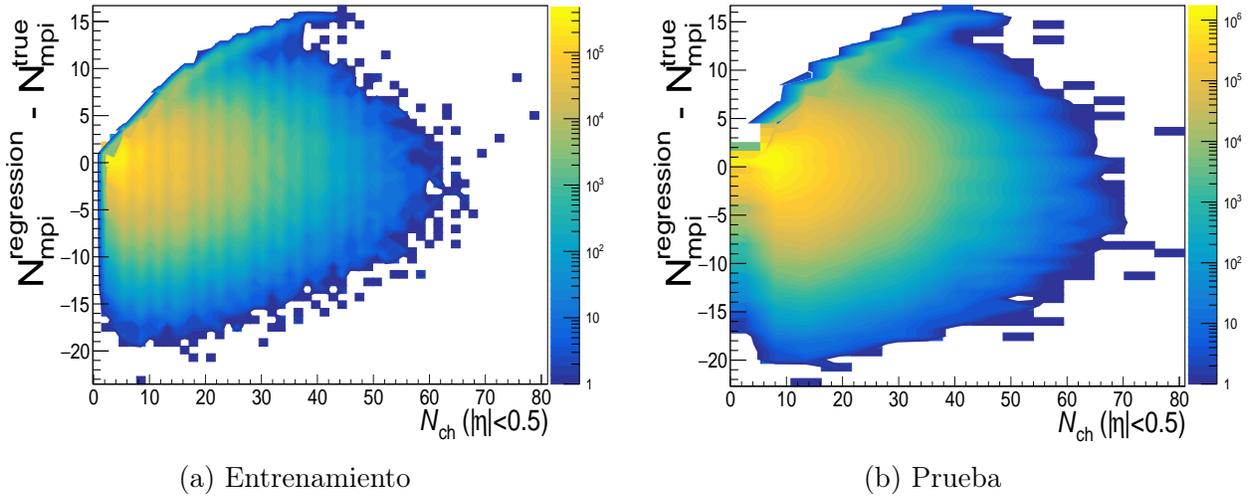


Figura 3.2 Desviación entre la variable objetivo obtenida por regresión y el valor verdadero, en función de la multiplicidad. Las figuras izquierda y derecha corresponden a resultados de entrenamiento y prueba respectivamente.

3.2.2. Validación Monte Carlo

Previo a la aplicación de los árboles entrenados a los datos de ALICE, desarrollamos la validación MC del método con la finalidad de verificar su robustez frente al modelo MC utilizado para el entrenamiento incluyendo el modelo de hadronización. Los BDT fueron entrenados con los modelos 2C, 4C y Monash 2013 de PYTHIA 8, y con HERWIG 7 Soft Tune para su posterior aplicación a simulaciones de colisiones pp a $\sqrt{s} = 5.02, 7$ y 13 TeV generadas por medio de PYTHIA 8 Tune 4C. Debido a que PYTHIA y HERWIG utilizan distintos modelos de MPI y hadronización, sus efectos son considerados en las incertidumbres sistemáticas. La validación consistió en comparar los resultados de los BDT con la información proporcionada por simulaciones de PYTHIA 8 Tune 4C (información verdadera).

La Figura 3.3 muestra el número de MPI normalizadas por su promedio ($N_{\text{mpi}}/\langle N_{\text{mpi}} \rangle$) en función de la densidad de multiplicidad de partículas cargadas a pseudorapidez media normalizada por su promedio ($dN_{\text{ch}}/d\eta/\langle dN_{\text{ch}}/d\eta \rangle$) para colisiones pp a $\sqrt{s} = 5.02, 7$ y 13 TeV generadas por medio de PYTHIA 8 Tune 4C. Las líneas amarillas y rojas corresponden a los resultados de los BDT entrenados con los distintos modelos MC, mientras que los marcadores negros corresponden a las simulaciones de PYTHIA 8 Tune 4C. Para altos valores de multiplicidad ($dN_{\text{ch}}/d\eta/\langle dN_{\text{ch}}/d\eta \rangle > 4$), observamos una desviación respecto al

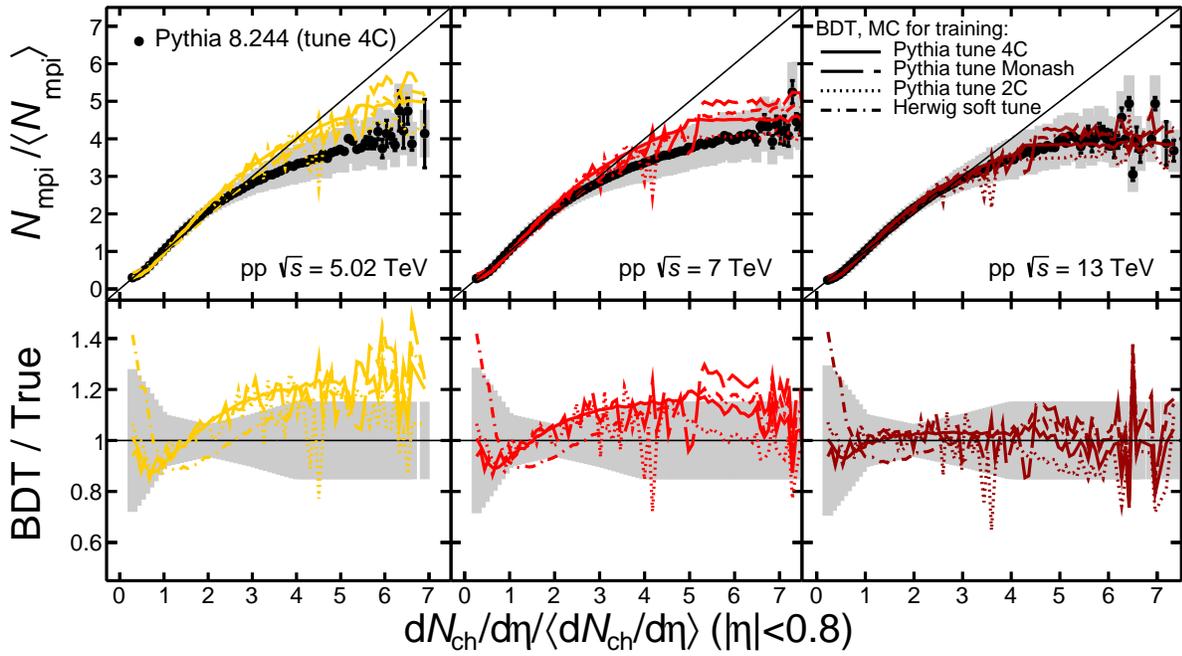


Figura 3.3 Validación Monte Carlo utilizando colisiones pp a $\sqrt{s} = 5.02$ (izquierda), 7 (centro) y 13 TeV (derecha) simuladas con PYTHIA 8. Los paneles superiores muestran el número de MPI normalizadas por su promedio en función de la densidad de multiplicidad de partículas cargadas a pseudorapidez media normalizada por su promedio. Los resultados de PYTHIA corresponden a los marcadores negros; mientras que los resultados de los BDT entrenados con los modelos MC: 4C, Monash, 2C y Soft Tune de Herwig 7 corresponden a la línea sólida, discontinua, punteada y punteada-discontinua respectivamente. Los paneles inferiores muestran el cociente entre los resultados de los BDT y los de PYTHIA. En todos los paneles la banda gris corresponde a la incertidumbre sistemática.

incremento lineal de $N_{\text{mpi}}/\langle N_{\text{mpi}} \rangle$ con la multiplicidad. Este comportamiento es compatible con el estudio realizado en [50]; donde para valores de $z \equiv dN_{\text{ch}}/d\eta/\langle dN_{\text{ch}}/d\eta \rangle > 4$, el porcentaje de eventos que tienen jets con $p_T > 5 \text{ GeV}/c$ es del 80%, es decir, las colisiones pp de alta multiplicidad únicamente pueden ser producidas seleccionando jets de alta multiplicidad. Dentro de las incertidumbres, este comportamiento es completamente reproducido por el método de regresión, así como la pequeña dependencia en la energía.

En la Figura 3.3 las bandas grises corresponden a las incertidumbres sistemáticas, las cuales son del 30% para $dN_{\text{ch}}/d\eta/\langle dN_{\text{ch}}/d\eta \rangle \approx 0$, 11-12% para $dN_{\text{ch}}/d\eta/\langle dN_{\text{ch}}/d\eta \rangle \approx 2-3$ y de 15% para $dN_{\text{ch}}/d\eta/\langle dN_{\text{ch}}/d\eta \rangle > 4$. Esta asignación cubre las variaciones del modelo MC utilizado para el entrenamiento. A pesar de que el algoritmo se entrenó con eventos de

colisiones pp a $\sqrt{s} = 13$ TeV generados con PYTHIA 8 Tune 4C, al aplicarlo a colisiones de bajas energías ($\sqrt{s} = 5.02$ y 7 TeV) y simuladas con otros modelos, dentro de las incertidumbres el método funciona; por ello podemos confiar en aplicarlo a datos.

3.2.2.1. Efecto de la distribución de N_{mpi}

En [11] realizamos un estudio del impacto de la distribución de N_{mpi} utilizada para el entrenamiento, con la finalidad de cuantificar sus efectos en la estimación de $\langle N_{\text{mpi}} \rangle$. La doble gaussiana que describe MPI en PYTHIA fue cambiada por una distribución homogénea ($\sim kN_{\text{mpi}}$, donde k es una constante). La Figura 3.4 muestra la distribución de $N_{\text{mpi}}/\langle N_{\text{mpi}} \rangle$ para colisiones pp (MB) a $\sqrt{s} = 2.76$ TeV simuladas con PYTHIA Tune 4C. La distribución se compara con aquellas obtenidas por la regresión, indicando que cuando los BDT fueron entrenados con una distribución homogénea de N_{mpi} , los valores altos en N_{mpi} son mejor descritos.

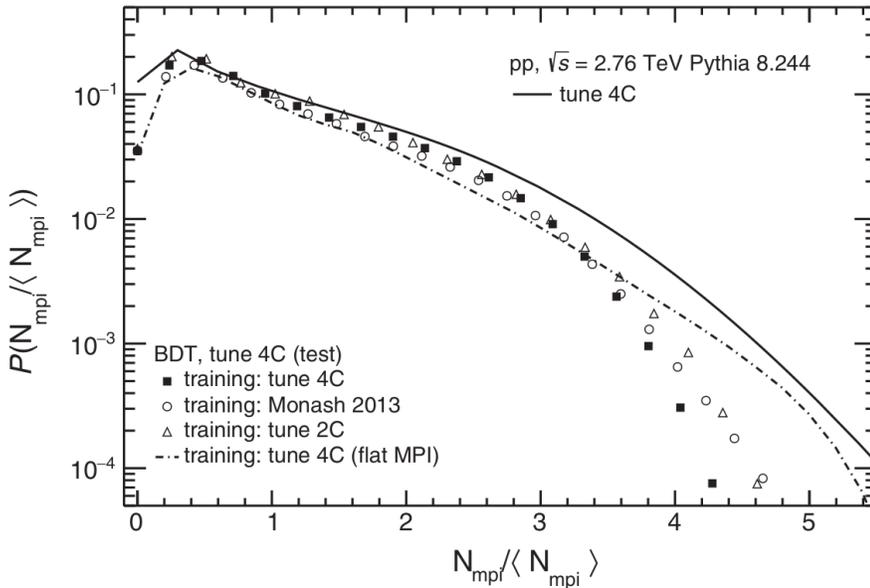


Figura 3.4 La línea sólida corresponde a la distribución de $N_{\text{mpi}}/\langle N_{\text{mpi}} \rangle$ para colisiones pp (MB) a $\sqrt{s} = 2.76$ TeV simuladas con PYTHIA Tune 4C. La distribución obtenida con los BDT cuando fueron entrenados con los modelos de PYTHIA: 4C, Monash y 2C corresponden a los marcadores cuadrados, circulares y triangulares respectivamente. La línea punteada corresponde a los BDT entrenados con una distribución homogénea de N_{mpi} . Imagen tomada de [11].

El efecto del cambio de la distribución de N_{mpi} usada para el entrenamiento se cuantificó y asignó como incertidumbre sistemática en el valor $\langle N_{\text{mpi}} \rangle$ obtenido de la regresión.

3.2.3. Procesamiento de datos y Toy MC

Debido a que no hay datos disponibles que proporcionen correlaciones evento por evento de las variables de entrada N_{ch} y $\langle p_{\text{T}} \rangle$. Construimos un generador Monte Carlo (Toy MC) para determinar la correlación entre esas variables evento por evento, utilizando los datos de ALICE disponibles. Bajo este enfoque, los BDT entrenados fueron aplicados a la información producida por el Toy MC, con la finalidad de extraer la actividad de MPI asociada con los datos.

Los datos de ALICE utilizados en este trabajo son de colisiones pp a $\sqrt{s} = 5.02$, 13 TeV y $\sqrt{s} = 7$ TeV que corresponden a las referencias [30] y [4] respectivamente. Los datos contienen los espectros de momento transverso para distintas clases de multiplicidad². Con la finalidad de estudiar la dependencia en la multiplicidad de las distribuciones de p_{T} , los datos de $\sqrt{s} = 5.02$ y 13 TeV se dividen en clases de eventos cuantificados por los estimadores V0M y SPD, mientras que los datos de $\sqrt{s} = 7$ TeV únicamente están cuantificados por el estimador V0M.

Los datos también proporcionan la densidad de multiplicidad de partículas cargadas promedio ($\langle dN_{\text{ch}}/d\eta \rangle$), así como su contribución a la sección eficaz inelástica. Por ejemplo, para $\sqrt{s} = 13$ TeV, la clase de multiplicidad de eventos I en el estimador V0M contribuye con 0-1 % de la sección eficaz inelástica y le corresponde $\langle dN_{\text{ch}}/d\eta \rangle = 26.6 \pm 1.1$. El Apéndice B proporciona las clases de multiplicidad y su densidad de multiplicidad de partículas cargadas promedio correspondiente.

El Toy MC es un programa en el lenguaje C++ y se construyó dentro de ROOT. A continuación resumimos su funcionamiento:

1. Por simplicidad, cada clase de eventos fue simulada considerando que su espectro de

²Los datos se encuentran públicos en HEPData y pueden ser descargados en formato compatible con el software ROOT (ver <https://www.hepdata.net/record/ins1684320> y <https://www.hepdata.net/record/ins1735345>).

multiplicidad sigue una distribución de Poisson [51], con media dada por la densidad de multiplicidad de partículas cargadas promedio correspondiente. Para los datos de colisiones a $\sqrt{s} = 5.02$ y 13 TeV se simularon $17 \cdot 10^5$ eventos, mientras que para los datos a $\sqrt{s} = 7$ TeV se simularon $9 \cdot 10^6$. Por ejemplo, para $\sqrt{s} = 13$ TeV, la clase de multiplicidad de eventos I en el estimador V0M contribuyó con $17 \cdot 10^3$ eventos.

2. Dentro de un loop de eventos se generan N_{ch} pseudo-partículas, este número de pseudo-partículas es extraído aleatoriamente de los espectros de multiplicidad, y nuevos espectros de momento transversal son llenados aleatoriamente N_{ch} -veces obedeciendo los espectros de p_T reportados por ALICE.
3. En un loop sobre partículas se extrae aleatoriamente el p_T de los espectros de momento transversal llenados en el paso anterior, con la restricción $p_T > 0.15$ GeV.
4. En cada evento el $\langle p_T \rangle$ es calculado. Toda la información generada se guarda en un tipo de archivo llamado *árbol* (TTree de ROOT), que consiste en una lista de columnas o *ramas* independientes. Este tipo de archivos facilita el posterior análisis con los BDT.

El desempeño del Toy MC se resume en la Figura 3.5, la cual muestra el momento transversal promedio en función de la densidad de multiplicidad de partículas cargadas promedio en colisiones pp a $\sqrt{s} = 5.02, 7$ y 13 TeV.

Dentro de las incertidumbres, el Toy MC reproduce por completo la correlación entre $\langle dN_{ch}/d\eta \rangle$ y $\langle p_T \rangle$. Este resultado es cualitativamente compatible con la correlación entre el $\langle p_T \rangle$ y la densidad de multiplicidad de partículas cargadas $dN_{ch}/d\eta$ para colisiones pp a $\sqrt{s} = 13$ TeV reportada en la referencia [30].

Para estimar la incertidumbre debida al Toy MC, consideramos lo siguiente:

- Fijando la densidad de multiplicidad de partículas cargadas promedio, variamos los espectros de momento transversal a su mínimo y máximo valor dado por sus correspondientes incertidumbres.
- Fijando las distribuciones de momento transversal, variamos la densidad de multiplicidad de partículas cargadas promedio de la misma forma que los espectros de momento

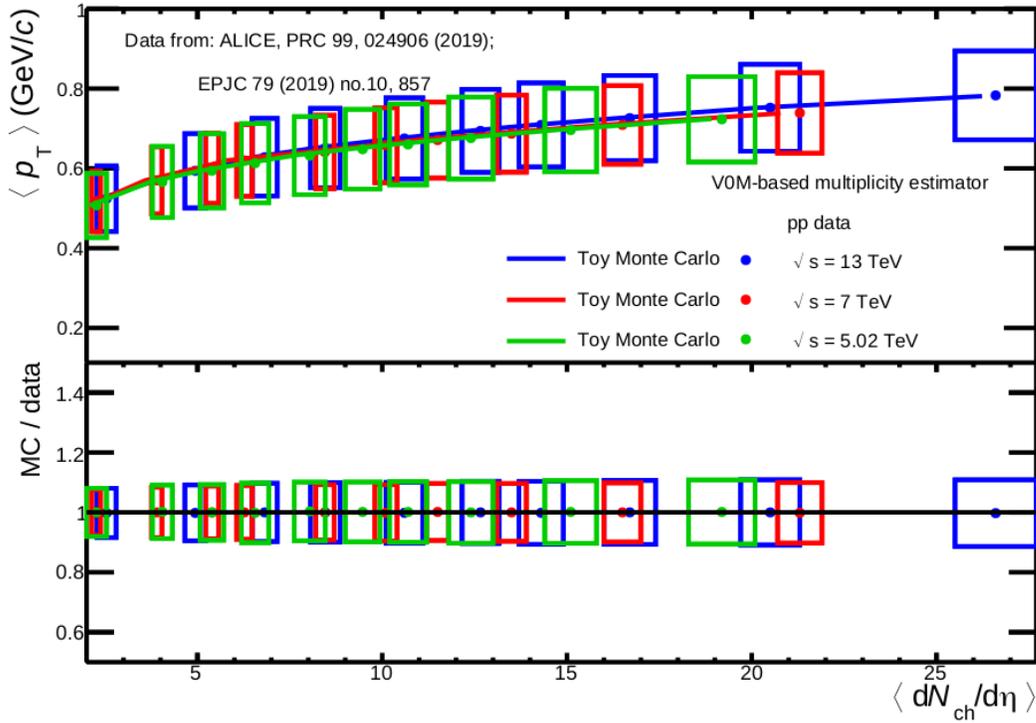


Figura 3.5 Momento transversal promedio en función de la densidad de multiplicidad de partículas cargadas promedio en colisiones pp a $\sqrt{s} = 5.02, 7$ y 13 TeV. En el panel superior, los marcadores corresponden a los datos de ALICE y las líneas sólidas a los resultados del Toy MC. El panel inferior muestra el cociente entre los resultados del Toy MC y los datos. En ambos paneles los rectángulos alrededor de los marcadores corresponden a las incertidumbres sistemáticas.

transverso. Por ejemplo, para $\sqrt{s} = 13$ TeV, la clase de multiplicidad de eventos I en el estimador V0M tiene una $\langle dN_{ch}/d\eta \rangle = 27.7$ y 25.5 como valores máximos y mínimos respectivamente.

La mayor diferencia se encontró para los datos de $\sqrt{s} = 5.02$ TeV, dando una incertidumbre de alrededor 7.5% , cantidad que es despreciable al estar sumada en cuadratura con la incertidumbre dada por el modelo.

Como validación del Toy MC, procesamos simulaciones de colisiones pp a $\sqrt{s} = 5.02, 9$ y 13 TeV generadas por medio de PYTHIA 8 Tune 4C. Las diferencias entre N_{mpi} de regresión con N_{mpi} proporcionado por PYTHIA se encontraron despreciables en comparación con la incertidumbre sistemática debida a la dependencia del modelo.

Resultados

4.1. Número promedio de MPI

Utilizando los BDT, estimamos el número promedio de interacciones partónicas múltiples $\langle N_{\text{mpi}} \rangle$ de los datos de colisiones (INEL > 0) pp a $\sqrt{s} = 7$ TeV de ALICE. Encontramos que $\langle N_{\text{mpi}} \rangle = 3.98 \pm 1.01$. Este resultado complementa aquellos que reportamos en [11], donde $\langle N_{\text{mpi}} \rangle = 3.76 \pm 1.01$ y 4.65 ± 1.01 para los datos de ALICE de colisiones pp a $\sqrt{s} = 5.02$ y 13 TeV respectivamente.

La Figura 4.1 muestra $\langle N_{\text{mpi}} \rangle$ en función de la energía del centro de masa \sqrt{s} , para colisiones pp a $\sqrt{s} = 5.02, 7$ y 13 TeV. La clase INEL > 0 corresponde a una clase de eventos definida por ALICE en la que al menos una partícula cargada es producida en el intervalo $|\eta| < 1$, y corresponde con alrededor del 75 % de la sección eficaz inelástica.

Con el método de ML aplicado a los datos de ALICE encontramos un valor de $\langle N_{\text{mpi}} \rangle$ superior a la unidad. Por tanto, nuestros resultados sugieren la presencia de MPI en colisiones pp.

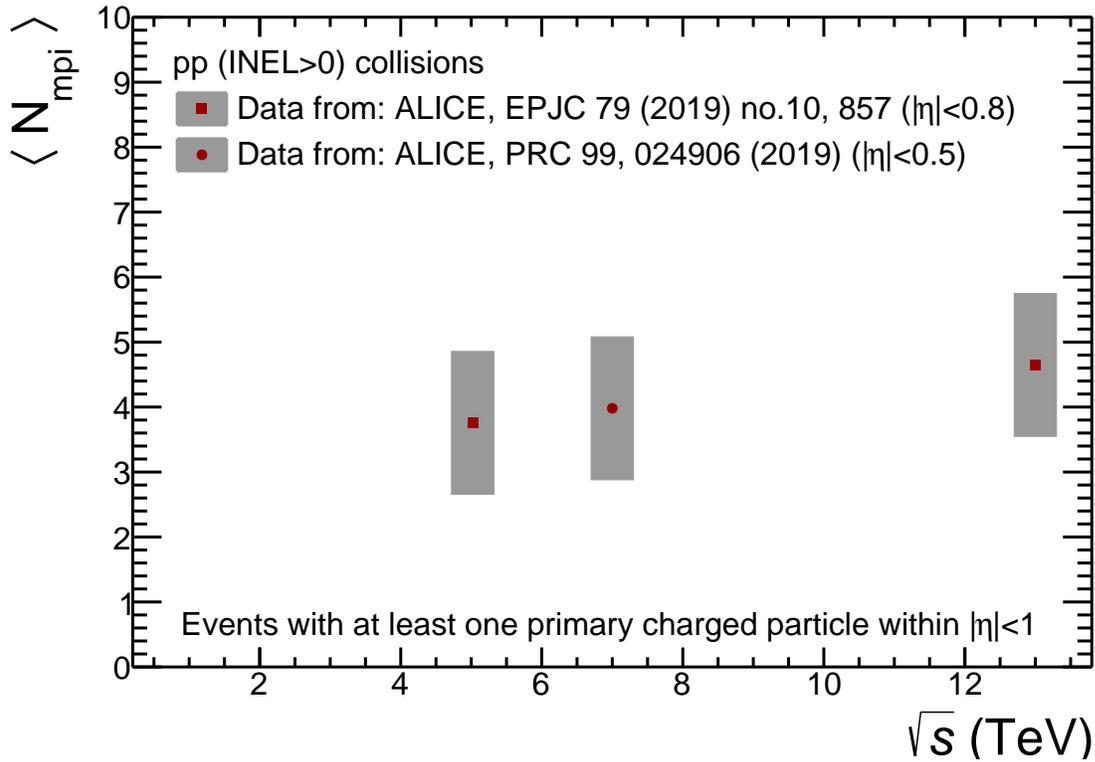


Figura 4.1 Número promedio de MPI en función de la energía del centro de masa. El resultado de los BDT entrenados aplicados a los datos de ALICE de $\sqrt{s} = 7$ TeV (marcador circular) es comparado con los resultados de colisiones pp a $\sqrt{s} = 5.02$ y 13 TeV que reportamos en [11] (marcadores cuadrados). Los rectángulos alrededor de los marcadores corresponden a las incertidumbres sistemáticas.

En los resultados de la Figura 4.1 también observamos una ligera dependencia en la energía, el valor de $\langle N_{\text{mpi}} \rangle$ aumenta levemente con el incremento en la energía del centro de masa. Este comportamiento es similar a lo predicho por PYTHIA 8 Tune 4C reportado en [11].

4.2. Dependencia de N_{mpi} con la multiplicidad

En este trabajo, extendemos el estudio que reportamos en [11] extrayendo la dependencia de N_{mpi} con la multiplicidad de los datos de ALICE disponibles. La Figura 4.2 muestra el número de MPI normalizadas por su promedio ($N_{\text{mpi}}/\langle N_{\text{mpi}} \rangle$) como función de la multiplicidad de partículas cargadas a pseudorapidez media normalizada por su promedio ($N_{\text{ch}}/\langle N_{\text{ch}} \rangle$) en colisiones pp a $\sqrt{s} = 5.02, 7$ y 13 TeV de los datos de ALICE. Los resultados

se dividen para los distintos estimadores de multiplicidad, tal y como son divididos los datos de ALICE. Las cajas alrededor de los marcadores indican las incertidumbres sistemáticas, que son del 30 % para $N_{\text{ch}}/\langle N_{\text{ch}} \rangle \approx 0$, 11-12 % para $N_{\text{ch}}/\langle N_{\text{ch}} \rangle \approx 2-3$ y de 15 % para $N_{\text{ch}}/\langle N_{\text{ch}} \rangle > 4$.

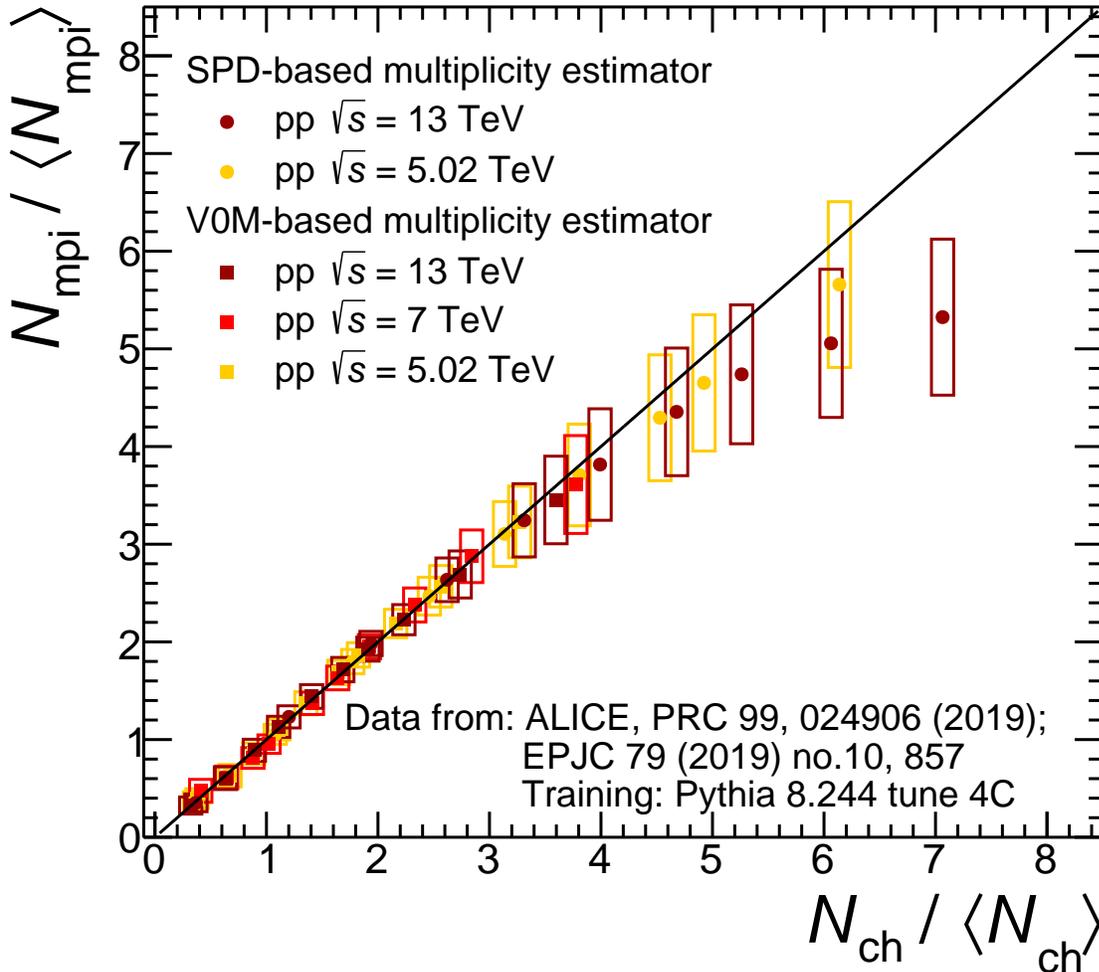


Figura 4.2 Correlación entre el número de MPI normalizadas por su promedio y la multiplicidad de partículas cargadas a pseudorapidez media normalizada por su promedio determinada por medio de la aplicación de los BDT entrenados a los datos de ALICE de colisiones pp a $\sqrt{s} = 5.02, 7$ y 13 TeV. Las cajas alrededor de los marcadores indican las incertidumbres sistemáticas.

Para $N_{\text{ch}}/\langle N_{\text{ch}} \rangle < 3$, $N_{\text{mpi}}/\langle N_{\text{mpi}} \rangle$ incrementa linealmente con la multiplicidad, mientras que para altas multiplicidades ($N_{\text{ch}}/\langle N_{\text{ch}} \rangle \gtrsim 4$) observamos una desviación en $N_{\text{mpi}}/\langle N_{\text{mpi}} \rangle$ respecto a la tendencia lineal. Este comportamiento es cualitativamente consistente con la correlación proporcionada por simulaciones de PYTHIA 8 (Figura 3.3), y no muestra una

dependencia significativa en la energía.

Es importante resaltar como nuestro resultado se compara con el estudio de ALICE reportado en [52], en el cual se define la cantidad llamada *uncorrelated seeds* ($N_{\text{uncorrelated seeds}}$), que interpretada en el contexto de PYTHIA proporciona información sobre el número de interacciones partónicas semi-duras por evento. El análisis involucra la medición de pares de partículas en el rango de pseudorapidez $|\eta| < 0.9$, una partícula 'trigger' y una 'asociada' cuyos rangos de momento trasverso son $p_{\text{T, trig}} > 0.7 \text{ GeV}/c$ y $p_{\text{T, assoc}} > 0.4(0.7) \text{ GeV}/c$ respectivamente.

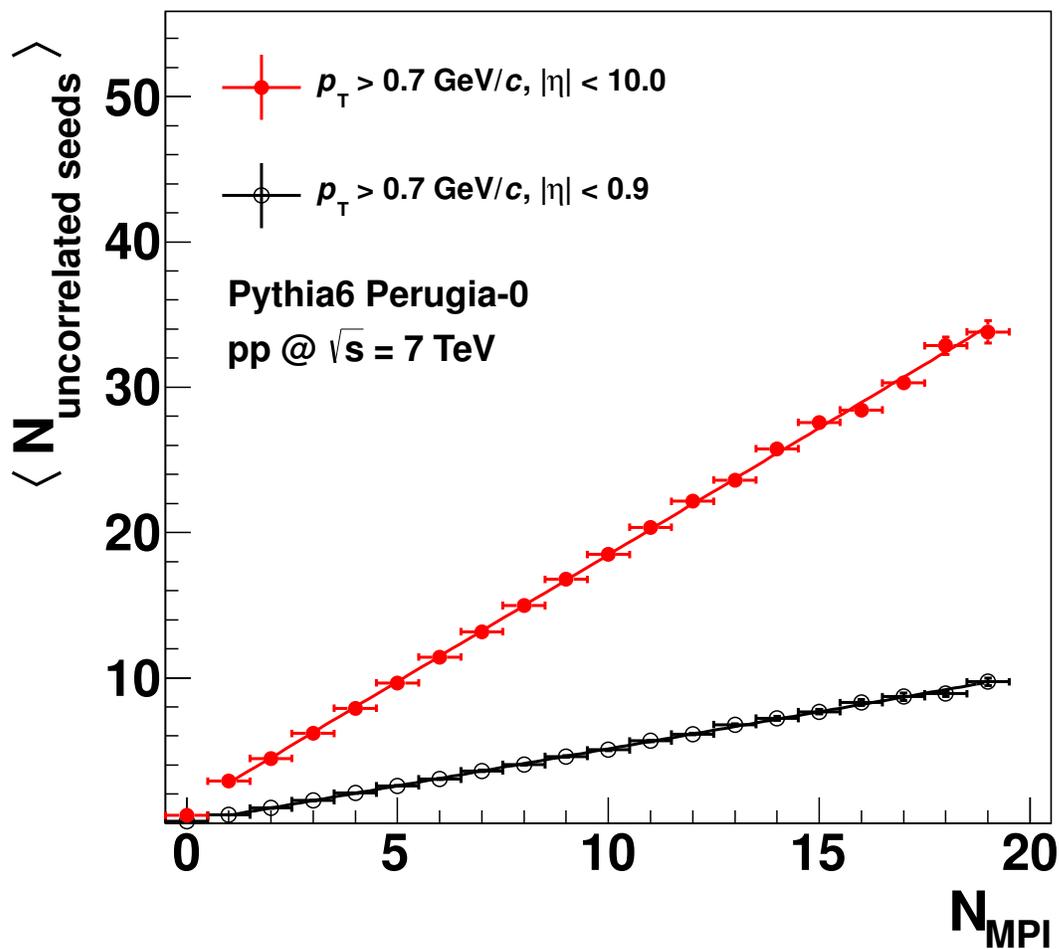


Figura 4.3 Dependencia lineal entre $\langle N_{\text{uncorrelated seeds}} \rangle$ y N_{mpi} . Imagen tomada de [52].

La Figura 4.3 muestra $\langle N_{\text{uncorrelated seeds}} \rangle$ en función de N_{mpi} por medio de simulaciones de PYTHIA 6 Tune Perugia-0 de colisiones pp a $\sqrt{s} = 7$ TeV, en diferentes rangos de pseudorapidez. Observamos una clara dependencia lineal de $\langle N_{\text{uncorrelated seeds}} \rangle$ con N_{mpi} .

Por otra parte, la Figura 4.4 muestra el número promedio de uncorrelated seeds ($\langle N_{\text{uncorrelated seeds}} \rangle$) en función de la multiplicidad de partículas cargadas N_{ch} (medida en los cortes cinemáticos $|\eta| < 0.9$ y $p_{\text{T}} > 0.2$ GeV/c), para los datos de ALICE de colisiones pp a $\sqrt{s} = 7, 2.76$ y 0.9 TeV.

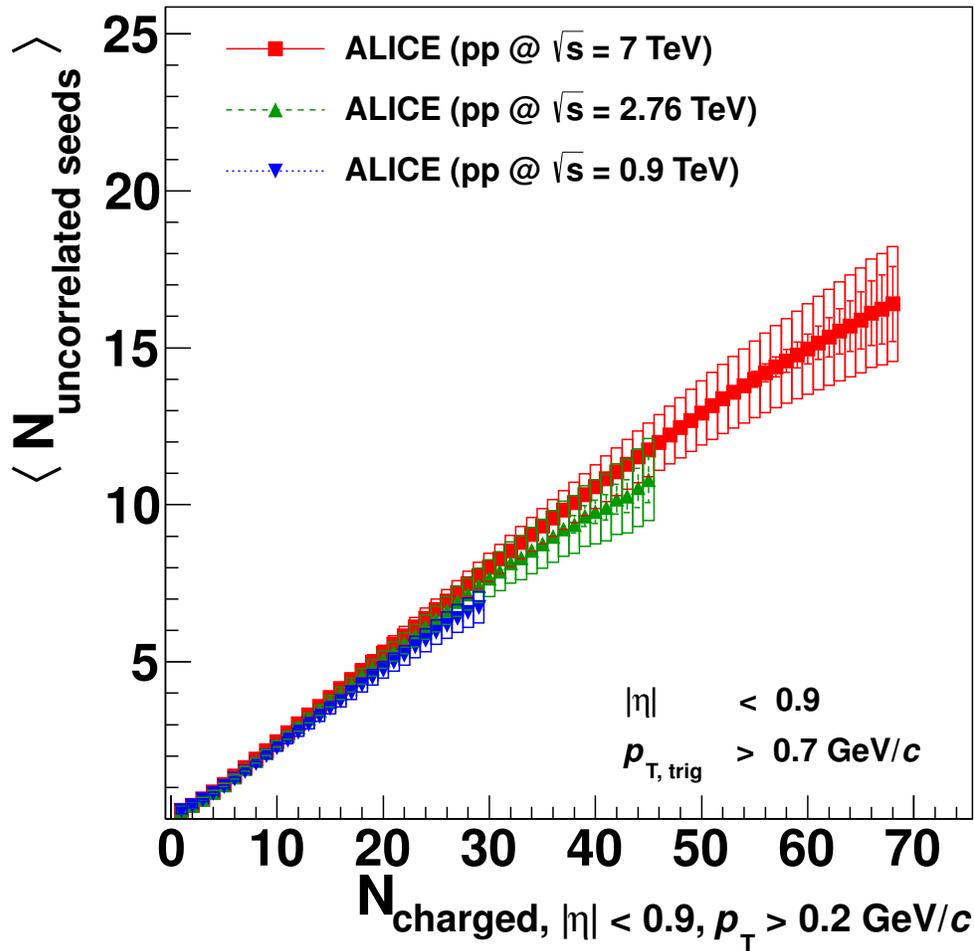


Figura 4.4 Número promedio de uncorrelated seeds en función de la multiplicidad de partículas cargadas para los datos de ALICE de colisiones pp a $\sqrt{s} = 7, 2.76$ y 0.9 TeV. Imagen tomada de [52].

Observamos que $\langle N_{\text{uncorrelated seeds}} \rangle$ incrementa linealmente con N_{ch} . Sin embargo, para valores altos en multiplicidad, $\langle N_{\text{uncorrelated seeds}} \rangle$ se desvía respecto a la tendencia lineal. Este resultado no muestra una dependencia importante en la energía.

Las anteriores conclusiones son completamente compatibles con nuestros resultados. Y sugieren que a altas multiplicidades (a pseudorapidez central) un incremento en el número de interacciones partónicas múltiples es sumamente improbable, en su lugar, alta multiplicidad de partículas cargadas es únicamente alcanzable seleccionando eventos con jets de alta multiplicidad. Esta conclusión es similar al análisis de la producción de jets en función de la multiplicidad de eventos en colisiones pp [50]. En tal estudio, para valores de $z \equiv dN_{\text{ch}}/d\eta / \langle dN_{\text{ch}}/d\eta \rangle > 4$ el porcentaje de eventos que tienen jets con $p_{\text{T}} > 5 \text{ GeV}/c$ es del 80 %. Esta característica es explicada como un sesgo de selección que surge cuando la multiplicidad del evento y el momento transversal de las partículas son medidos en intervalos de pseudorapidez central ($|\eta| < 1$ o menor).

Recordemos que los datos de ALICE con los que trabajamos corresponden a eventos inelásticos, los cuales están conformados por interacciones difractivas y no-difractivas. Las interacciones difractivas son en general suaves, mientras que las no-difractivas son duras. De acuerdo con mediciones de la colaboración ALICE, alrededor del 68 % de las colisiones inelásticas son no-difractivas [53]. Siguiendo la definición de pseudorapidez:

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right), \quad (4.1)$$

las interacciones duras producen jets en regiones de $|\eta| \sim 0$, pues el mayor intercambio de momento se produce cuando $\theta \sim 90^\circ$ donde θ es el ángulo de producción de la partícula. De esta manera, la probabilidad de encontrar jets en otras regiones de pseudorapidez es muy baja, es decir, altas multiplicidades se encuentran seleccionando eventos con jets de alta multiplicidad.

4.3. Panorama Futuro

Con la finalidad de determinar si la extracción del N_{mpi} mejora evento por evento, proponemos incluir mayor información en el entrenamiento de los BDT, determinando la multiplicidad del evento en rangos de pseudorapidez delantera.

Entrenamos tres conjuntos de BDT con datos de colisiones pp a $\sqrt{s} = 13$ TeV generados con PYTHIA 8 Tune 4C. En todos los casos el $\langle p_{\text{T}} \rangle$ fue calculado evento por evento en el rango de pseudorapidez media $|\eta| < 0.8$ y considerando partículas con $p_{\text{T}} > 0.15$ GeV/ c :

- ALICE Run2 Mid: El primer conjunto de BDT fue entrenado calculando la multiplicidad en el rango de pseudorapidez $|\eta| < 0.8$, de acuerdo con la capacidad del barril central de ALICE para la corrida 2.
- ALICE Run2 V0A+V0C: El segundo conjunto de BDT fue entrenado calculando la multiplicidad en los rangos de pseudorapidez forward $2.8 < \eta < 2.5$ y $-3.6 < \eta < -1.7$ que cubren los arreglos V0A y V0C respectivamente. Estos arreglos componen al detector forward V0 que formaba parte de la corrida 2 de ALICE.
- ALICE Run3 MFT+V0+: El tercer conjunto de BDT fue entrenado calculando la multiplicidad en los rangos de pseudorapidez forward $-3.6 < \eta < -2.45$ y $2.2 < \eta < 5.1$ que cubren los detectores MFT y VZERO+ respectivamente. Los nuevos detectores VZERO+ y el Muon Forward Tracker (MFT) forman parte de las actualizaciones para la corrida 3 de ALICE que está planeada de 2021 a 2023 [54].

Cada conjunto de BDT fue posteriormente aplicado a datos de colisiones pp a $\sqrt{s} = 13$ TeV generados con PYTHIA 8 Tune 4C.

La Figura 4.5 muestra el número promedio de interacciones partónicas múltiples determinado con los BDT ($\langle N_{\text{mpi}} \rangle$ (Reg)) en función del número verdadero de interacciones partónicas múltiples (N_{mpi} (True)) en colisiones pp a $\sqrt{s} = 13$ TeV simulado con PYTHIA.

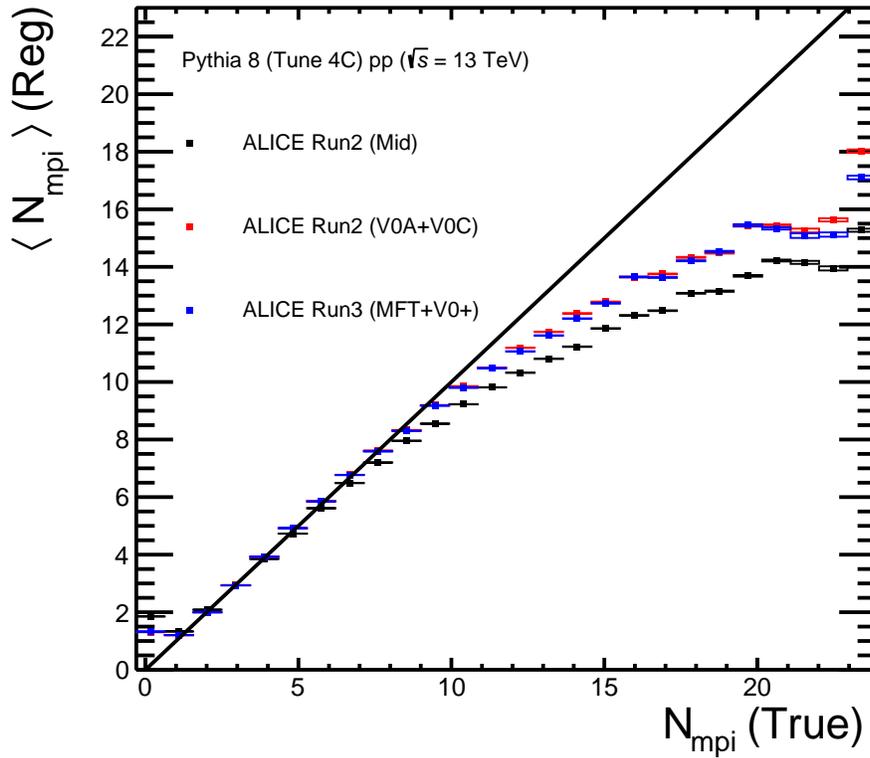


Figura 4.5 Número promedio de interacciones partónicas múltiples determinado con los BDT ($\langle N_{\text{mpi}} \rangle (\text{Reg})$) en función del número verdadero de interacciones partónicas múltiples proporcionado por simulaciones de PYTHIA ($N_{\text{mpi}} (\text{True})$). La correlación se muestra cuando la multiplicidad es determinada en el rango de pseudorapidez media (marcadores negros), al igual que cuando es determinada en la región forward definida por los detectores V0A+V0C (marcadores rojos) y MFT+V0+ (marcadores azules).

Observamos que para valores de $N_{\text{mpi}} (\text{True}) \approx 16$, cuando la multiplicidad ha sido determinada en el rango de pseudorapidez forward (ALICE Run2 V0A+V0C y ALICE Run3 MFT+V0+), la desviación del cociente $\langle N_{\text{mpi}} \rangle (\text{Reg}) / N_{\text{mpi}} (\text{True})$ con respecto al caso cuando la multiplicidad es determinada en el rango de pseudorapidez media (ALICE Run2 Mid), es de aproximadamente 13%. Por lo tanto, la extracción del N_{mpi} mejora al incluir mayor información en el entrenamiento.

El resultado anterior abre la posibilidad de extraer el número de MPI evento por evento y de esta manera estudiar la producción de partículas en función de MPI en lugar del estimador de multiplicidad que define cada experimento. Esta idea se exploró en la referencia [11], donde la producción de partículas es estudiada en función del número verdadero de MPI (N_{mpi}), aquel obtenido por regresión ($N_{\text{mpi}}^{\text{reg}}$), y la densidad de multiplicidad

de partículas cargadas a pseudorapidez media ($dN_{\text{ch}}/d\eta$), en colisiones pp a $\sqrt{s} = 2.76$ TeV. Los paneles superiores de la Figura 4.6 muestran la cantidad llamada R_{pp} , definida como:

$$R_{\text{pp}} = \frac{d^2 N_{\pi}^{\text{mpi}} / (\langle N_{\text{mpi}} \rangle dy dp_{\text{T}})}{d^2 N_{\pi}^{\text{MB}} / (\langle N_{\text{mpi,MB}} \rangle dy dp_{\text{T}})} \quad (4.2)$$

donde, $d^2 N_{\pi}^{\text{MB}}/dy dp_{\text{T}}$ y $d^2 N_{\pi}^{\text{mpi}}/dy dp_{\text{T}}$ cuantifican la producción de piones cargados para una muestra MB y para una sub-muestra definida por su actividad de MPI, respectivamente. La cantidad R_{pp} define la producción de piones por dispersión partónica semidura en colisiones pp con un alto o bajo número de MPI normalizada al de las colisiones pp MB.

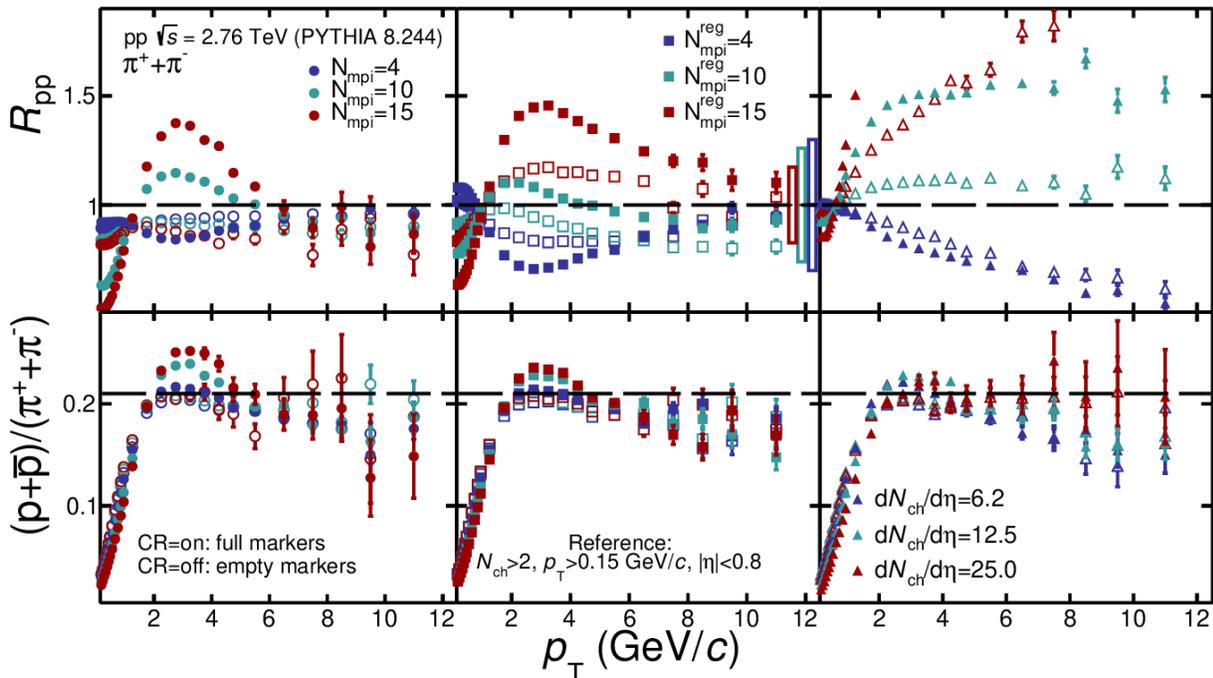


Figura 4.6 Producción de partículas en función de clases de eventos basados en el N_{mpi} (izquierda), $N_{\text{mpi}}^{\text{reg}}$ (centro) y $dN_{\text{ch}}/d\eta$ (derecha), en colisiones pp a $\sqrt{s} = 2.76$ TeV. Los resultados de simulaciones con y sin reconexión por color corresponden a los marcadores rellenos y vacíos respectivamente. Imagen tomada de [11].

En el panel superior izquierdo de la Figura 4.6, la característica más importante a notar, es que R_{pp} tiene un pico a p_{T} intermedio (1-8 GeV/c), el cual es atribuido a reconexión por color [10]. Esta característica es consistente con las simulaciones sin reconexión por color, pues R_{pp} permanece casi plano e independiente de N_{mpi} . El panel superior central de la Figura 4.6 muestra R_{pp} cuando el número de MPI es obtenido por regresión, resultado que

recupera cualitativamente la dependencia de R_{pp} con el verdadero N_{mpi} . Los resultados cuando el análisis es desarrollado en función de $dN_{ch}/d\eta$ se muestran en el panel superior derecho de la Figura 4.6, y no presentan las características previamente discutidas.

Los paneles inferior izquierdo y central de la Figura 4.6 muestran el cociente protón-pi3n para las clases de eventos con N_{mpi} y N_{mpi}^{reg} respectivamente, y al igual que lo reportado en la referencia [10], el cociente tiene un pico a p_T intermedio que aumenta al incrementar N_{mpi} y N_{mpi}^{reg} respectivamente. Tambi3n observamos que las simulaciones sin CR son independientes de N_{mpi} y N_{mpi}^{reg} . El panel inferior derecho de la Figura 4.6 muestra el análisis en funci3n de $dN_{ch}/d\eta$, y no presenta el efecto observado en los casos anteriores.

Conclusiones

En esta tesis, determinamos la dependencia en la multiplicidad del número de interacciones partónicas múltiples N_{mpi} de los datos de ALICE disponibles de colisiones pp a $\sqrt{s} = 5.02, 7$ y 13 TeV así como el número promedio de interacciones partónicas múltiples $\langle N_{\text{mpi}} \rangle$ para $\sqrt{s} = 7$ TeV. A continuación se resaltan los puntos más importantes:

- Utilizando los datos de ALICE que consisten en espectros de momento transversal en función de la multiplicidad de eventos para colisiones pp a $\sqrt{s} = 7$ TeV, reportamos la extracción del $\langle N_{\text{mpi}} \rangle$ por medio de un método de aprendizaje de máquina basado en árboles de decisión, y encontramos $\langle N_{\text{mpi}} \rangle = 3.98 \pm 1.01$. Comparando este resultado con los que reportamos en [11], donde $\langle N_{\text{mpi}} \rangle = 3.76 \pm 1.01$ y 4.65 ± 1.01 para los datos de ALICE de colisiones pp a $\sqrt{s} = 5.02$ y 13 TeV respectivamente, nuestro resultado indica una pequeña dependencia en la energía, consistente con las predicciones de PYTHIA. Este resultado proporciona evidencia experimental de MPI en colisiones hadrónicas.
- Utilizando los datos de ALICE disponibles de colisiones pp, reportamos $N_{\text{mpi}}/\langle N_{\text{mpi}} \rangle$ en función de $N_{\text{ch}}/\langle N_{\text{ch}} \rangle$ para las energías $\sqrt{s} = 5.02, 7$ y 13 TeV. Nuestros resultados indican que para $N_{\text{ch}}/\langle N_{\text{ch}} \rangle < 3$, $N_{\text{mpi}}/\langle N_{\text{mpi}} \rangle$ incrementa linealmente con la multiplicidad, mientras que para altas multiplicidades ($N_{\text{ch}}/\langle N_{\text{ch}} \rangle > 4$) observamos una desviación de $N_{\text{mpi}}/\langle N_{\text{mpi}} \rangle$ respecto a la tendencia lineal. Resultado que es consistente

con el análisis de la colaboración ALICE donde una cantidad sensible a MPI es determinada en función de la multiplicidad. Por lo tanto, eventos de alta multiplicidad (a pseudorapidez media) únicamente pueden ser alcanzados seleccionando eventos con jets de alta multiplicidad.

- Por medio del método de aprendizaje de máquina aplicado a datos de colisiones pp a $\sqrt{s} = 13$ TeV generados con PYTHIA, encontramos que la extracción del N_{mpi} mejora al incluir mayor información en el entrenamiento de los BDT, calculando la multiplicidad en la región delantera. Esto abre la posibilidad de extraer el número de MPI evento por evento y de esta manera estudiar la producción de partículas en función de MPI en lugar del estimador de multiplicidad que define cada experimento.
- Basándonos en las verificaciones realizadas con los generadores de eventos Monte Carlo, y en el acuerdo de nuestros resultados con las mediciones de la colaboración ALICE. Confirmamos que nuestro enfoque es robusto y puede ser utilizado por experimentos con el fin de estudiar la producción de partículas en función de MPI. Esto contribuiría a comprender las características similares a iones pesados observadas en los datos de colisiones pp.

Árboles de decisión en TMVA

En general, un típico análisis de regresión en TMVA se divide en dos fases:

- **Fase de entrenamiento:** en esta fase el usuario se comunica con el programa para especificar los conjuntos de datos de entrenamiento y prueba, para registrar las variables objetivo, y para definir los métodos multivariados. A continuación, el programa entrena, prueba y evalúa los métodos seleccionados. Al finalizar esta fase se crea un archivo de *pesos* para cada método.
- **Fase de aplicación:** los métodos entrenados son aplicados a un conjunto de datos con variables objetivo por determinar. Al iniciar, el usuario registra las variables de entrada y especifica los métodos que encontró más apropiados después de la fase de entrenamiento. El archivo de pesos proporciona para cada método una configuración completa y consistente de acuerdo a los resultados del entrenamiento.

En los BDT, el entrenamiento es el proceso que define el criterio de división de cada nodo. El entrenamiento comienza con un nodo principal, donde el criterio de división inicial para el entrenamiento completo es determinado. En problemas de regresión, el criterio de división lo provee el error cuadrático medio:

$$\frac{1}{N} \sum (y - \hat{y})^2 \quad (\text{A.1})$$

donde y es la variable objetivo de cada evento en el nodo, \hat{y} es el valor promedio de la variable objetivo sobre todos los eventos en el nodo y N el número de eventos. La división de los nodos se desarrolla en la variable que proporciona la disminución máxima en el error cuadrático medio. Respecto a este trabajo, los datos de entrenamiento se encuentran en un archivo `TTree` que contiene ramas con las distribuciones de N_{ch} , $\langle p_{\text{T}} \rangle$ y N_{mpi} . Se utiliza un código en C++ llamado `TMVAREgression.C` donde se definen las variables de entrada y la variable objetivo, así como el número de eventos para el entrenamiento y prueba.

En este paso también es posible modificar las opciones de los BDT, respecto a esta tesis se consideraron los siguiente parámetros (para mayor detalle consultar [46]):

- `NTrees = 2000`: corresponde al número de árboles.
- `Shrinkage = 0.1`: el radio de aprendizaje para el Gradient Boost. Un Shrinkage pequeño puede mejorar significativamente la precisión en la predicción.
- `MaxDepth = 4`: profundidad máxima permitida de los árboles.

La Figura A.1 muestra esquemáticamente como luce uno de los árboles de decisión entrenados. En el nodo principal, la variable que proporciona la disminución máxima en el error cuadrático medio ($R = 0.1$) es la multiplicidad. A pesar de que en algunos nodos finales el error cuadrático medio es nulo en este ejemplo, en general, los árboles están sujetos a fluctuaciones estadísticas en los datos de entrenamiento.

Con la finalidad de mejorar el desempeño de la regresión e incrementar la estabilidad con respecto a dichas fluctuaciones, en TMVA se desarrolla un algoritmo de *Boosting*. Para tareas de regresión el algoritmo empleado es el *Gradient Boost*, el cual consiste en considerar la función $F(x)$ como una suma de los árboles de decisión que son añadidos iterativamente intentando minimizar una función de pérdida $L(F, y)$. La función de pérdida mide la desviación del modelo $F(x)$ con el valor verdadero de la variable objetivo y obtenido de la muestra de entrenamiento. La función de pérdida empleada por TMVA en estas tareas es la función de Huber [55].

La contribución de cada árbol al modelo puede ser ponderada por el Shrinkage, ralentizando las iteraciones. Un valor pequeño de Shrinkage favorece un mayor número de árboles

añadidos al modelo, lo que se traduce en una mejora en la precisión de la predicción.

La información procesada por el Toy MC se encuentra en un archivo TTree con las distribuciones de N_{ch} y $\langle p_T \rangle$. En un código llamado `TMVAREgressionApplication.C` se especifica el archivo de pesos con la configuración de los BDT. Al finalizar la aplicación, el código arroja un archivo ROOT con la distribución de N_{mpi} .

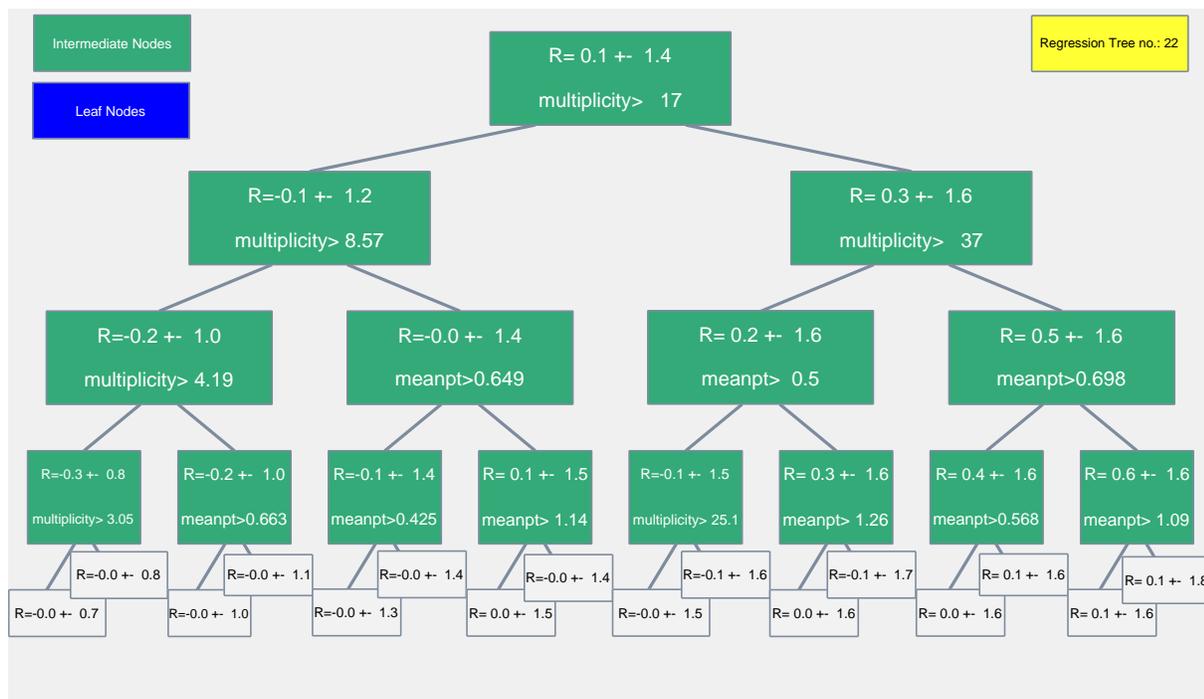


Figura A.1 Ejemplo de un árbol de decisión después del entrenamiento. La profundidad del árbol es 4. En el nodo principal, la variable que proporciona la disminución máxima en el error cuadrático medio es la multiplicidad.

Clases de multiplicidad

Los datos de ALICE se dividen en clases de eventos cuantificados por los estimadores V0M y SPD. A cada clase le corresponde un valor particular de $\langle dN_{\text{ch}}/d\eta \rangle$. Las tablas presentadas a continuación corresponden a los datos de colisiones pp a $\sqrt{s} = 5.02, 13$ TeV y $\sqrt{s} = 7$ TeV de las referencias [30] y [4] respectivamente.

Colisiones pp a $\sqrt{s} = 13$ TeV, V0M					
Clase	I	II	III	IV	V
$\langle dN_{\text{ch}}/d\eta \rangle$	26.6 ± 1.1	20.5 ± 0.8	16.7 ± 0.7	14.3 ± 0.6	12.6 ± 0.5
Clase	VI	VII	VIII	IX	X
$\langle dN_{\text{ch}}/d\eta \rangle$	10.6 ± 0.5	8.46 ± 0.40	6.82 ± 0.34	4.94 ± 0.28	2.54 ± 0.26

Tabla B.1 Clases de multiplicidad de eventos y $\langle dN_{\text{ch}}/d\eta \rangle$, cuantificadas por el estimador V0M.

Colisiones pp a $\sqrt{s} = 5.02$ TeV, V0M					
Clase	I	II	III	IV	V
$\langle dN_{\text{ch}}/d\eta \rangle$	19.2 ± 0.9	15.1 ± 0.7	12.4 ± 0.6	10.7 ± 0.5	9.47 ± 0.47
Clase	VI	VII	VIII	IX	X
$\langle dN_{\text{ch}}/d\eta \rangle$	8.04 ± 0.42	6.56 ± 0.37	5.39 ± 0.32	4.05 ± 0.27	2.27 ± 0.27

Tabla B.2 Clases de multiplicidad de eventos y $\langle dN_{\text{ch}}/d\eta \rangle$, cuantificadas por el estimador V0M.

Colisiones pp a $\sqrt{s} = 13$ TeV, SPD					
Clase	I'	II'	III'	IV'	V'
$\langle dN_{\text{ch}}/d\eta \rangle$	54.1 ± 2.7	44.6 ± 2.2	38.9 ± 1.9	34.1 ± 1.7	29.3 ± 1.5
Clase	VI'	VII'	VIII'	IX'	X'
$\langle dN_{\text{ch}}/d\eta \rangle$	24.5 ± 1.3	19.5 ± 1.2	14.4 ± 0.9	9.03 ± 0.58	2.91 ± 0.29

Tabla B.3 Clases de multiplicidad de eventos y $\langle dN_{\text{ch}}/d\eta \rangle$, cuantificadas por el estimador SPD.

Colisiones pp a $\sqrt{s} = 5.02$ TeV, SPD					
Clase	-	II'	III'	IV'	V'
$\langle dN_{\text{ch}}/d\eta \rangle$	-	34.6 ± 1.8	29.9 ± 1.5	26.2 ± 1.3	22.4 ± 1.1
Clase	VI'	VII'	VIII'	IX'	X'
$\langle dN_{\text{ch}}/d\eta \rangle$	18.5 ± 1.0	14.6 ± 0.9	10.6 ± 0.7	6.58 ± 0.43	2.21 ± 0.24

Tabla B.4 Clases de multiplicidad de eventos y $\langle dN_{\text{ch}}/d\eta \rangle$, cuantificadas por el estimador SPD.

Colisiones pp a $\sqrt{s} = 7$ TeV, V0M					
Clase	I	II	III	IV	V
$\langle dN_{\text{ch}}/d\eta \rangle$	21.3 ± 0.6	16.5 ± 0.5	13.5 ± 0.4	11.5 ± 0.3	10.1 ± 0.3
Clase	VI	VII	VIII	IX	X
$\langle dN_{\text{ch}}/d\eta \rangle$	8.45 ± 0.25	6.72 ± 0.21	5.40 ± 0.17	3.90 ± 0.14	2.26 ± 0.12

Tabla B.5 Clases de multiplicidad de eventos y $\langle dN_{\text{ch}}/d\eta \rangle$, cuantificadas por el estimador V0M.

Bibliografía

- [1] Ulrich Heinz and Raimond Snellings. Collective flow and viscosity in relativistic heavy-ion collisions. *Annual Review of Nuclear and Particle Science*, 63(1):123–151, Oct 2013.
- [2] Paul Romatschke and Ulrike Romatschke. Relativistic fluid dynamics in and out of equilibrium – ten years of progress in theory and numerical simulations of nuclear collisions. 2019.
- [3] J. D. Bjorken. Highly relativistic nucleus-nucleus collisions: The central rapidity region. *Phys. Rev. D*, 27:140–151, Jan 1983.
- [4] S. Acharya et al. Multiplicity dependence of light-flavor hadron production in pp collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev. C*, 99:024906, Feb 2019.
- [5] A. Ortiz. Production of pions, kaons and protons at high p_T in $\sqrt{s_{NN}} = 2.76$ TeV Pb-Pb collision. *Nuclear Physics A*, 904-905:763c–766c, May 2013.
- [6] B. Abelev et al. Centrality dependence of pions, kaons, and production in Pb-Pb collisions $\sqrt{s_{NN}} = 2.76$ TeV. *Physical Review C*, 88(4), Oct 2013.
- [7] Future high-energy pp programme with ALICE. Jul 2020.
- [8] Piotr Bożek. Small systems – hydrodynamics. *Nuclear Physics A*, 956:208–215, 2016. The XXV International Conference on Ultrarelativistic Nucleus-Nucleus Collisions: Quark Matter 2015.
- [9] Christian Bierlich et al. Collectivity without plasma in hadronic collisions. *Physics Letters B*, 779:58–63, 2018.

-
- [10] A. Ortiz et al. Color Reconnection and Flow like Patterns in pp Collisions. *Phys. Rev. Lett.*, 111(4):042001, 2013.
- [11] A. Ortiz et al. Multiparton interactions in pp collisions from machine learning-based regression. *Physical Review D*, 102(7), Oct 2020.
- [12] Antonio Ortiz and Erik A Zepeda. Extraction of the multiplicity dependence of multiparton interactions from LHC pp data using machine learning techniques. *Journal of Physics G: Nuclear and Particle Physics*, 48(8):085014, jul 2021.
- [13] David Griffiths. *Introduction to elementary particles*. Wiley-VCH, 2nd edition, 2008.
- [14] G. Aubrecht. The newest Standard Model Chart from the Contemporary Physics Education Project. *Latin-American Journal of Physics Education*, 10:3, 2016.
- [15] B. Barish. Tau neutrino physics: an introduction. *Nuclear Physics B Proceedings Supplements*, 98(1):12–25, apr 2001.
- [16] M. Gell-Mann. A schematic model of baryons and mesons. *Physics Letters*, 8(3):214–215, 1964.
- [17] G. Zweig. An SU(3) model for strong interaction symmetry and its breaking. Version 1. 1964.
- [18] O. Miramontes and K. Volke. *Fronteras de la física en el siglo XXI*. CopIt arXives, Mexico, 2013.
- [19] T. Sjöstrand and P. Skands. Multiple interactions and the structure of beam remnants. *Journal of High Energy Physics*, 2004(03):053–053, Mar 2004.
- [20] B. Abelev et al. Multiplicity dependence of the average transverse momentum in pp, p–Pb, and Pb–Pb collisions at the LHC. *Physics Letters B*, 727(4):371–380, 2013.
- [21] T. Sjöstrand et al. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008.
- [22] J. Bellm et al. Herwig 7.2 release note. *Eur. Phys. J. C*, 80(5):452, 2020.

-
- [23] Richard Corke and Torbjörn Sjöstrand. Interleaved parton showers and tuning prospects. *Journal of High Energy Physics*, 2011(3), Mar 2011.
- [24] P. Skands et al. Tuning PYTHIA 8.1: the Monash 2013 Tune. *Eur. Phys. J. C*, 74, Apr 2014.
- [25] S. Gieseke et al. Colour reconnections in Herwig++. *Eur. Phys. J. C*, 72:2225, 2012.
- [26] J. Bellm et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, 76(4):196, 2016.
- [27] M. Bähr et al. Herwig++ physics and manual. *The European Physical Journal C*, 58(4):639–707, Nov 2008.
- [28] A. Karneyeu et al. MCPLOTS: a particle physics resource based on volunteer computing. *The European Physical Journal C*, 74(2), Feb 2014.
- [29] G. Gustafson. Multiple interactions, saturation, and final states in pp collisions and dis, 2009.
- [30] S. Acharya et al. Charged-particle production as a function of multiplicity and transverse sphericity in pp collisions at $\sqrt{s} = 5.02$ and 13 TeV. *Eur. Phys. J. C*, 79(10):857, 2019.
- [31] The ALICE Collaboration. The ALICE experiment at the CERN LHC. *Journal of Instrumentation*, 3:S08002, 08 2008.
- [32] P. Cortese et al. ALICE: Physics performance report, volume I. *J. Phys. G*, 30:1517–1763, 2004.
- [33] C. Fabjan et al. ALICE: Physics performance report, volume II. *J. Phys. G*, 32:1295–2040, 2006.
- [34] A. Szczepankiewicz. Readout of the upgraded ALICE-ITS. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 824:465–469, 2016. Frontier Detectors for Frontier Physics: Proceedings of the 13th Pisa Meeting on Advanced Detectors.

-
- [35] P. Cortese et al. *ALICE forward detectors: FMD, TO and V0: Technical Design Report*. Technical design report. ALICE. CERN, Geneva, 2004. Submitted on 10 Sep 2004.
- [36] The ALICE collaboration. Performance of the ALICE VZERO system. *Journal of Instrumentation*, 8(10):P10016–P10016, oct 2013.
- [37] Sergei A. Voloshin et al. Collective phenomena in non-central nuclear collisions, 2008.
- [38] J. Adam et al. Enhanced production of multi-strange hadrons in high-multiplicity proton–proton collisions. *Nature Physics*, 13(6):535–539, Apr 2017.
- [39] V. Khachatryan et al. Evidence for collectivity in pp collisions at the LHC. *Physics Letters B*, 765:193–220, Feb 2017.
- [40] K. Aamodt et al. Elliptic flow of charged particles in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Physical Review Letters*, 105(25), Dec 2010.
- [41] Charles Gale et al. Hydrodynamic modeling of heavy-ion collisions. *International Journal of Modern Physics A*, 28(11):1340011, Apr 2013.
- [42] B. Abelev et al. Multi-strange baryon production at mid-rapidity in Pb–Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Physics Letters B*, 728:216–227, Jan 2014.
- [43] Kevin Dusling et al. Novel collective phenomena in high-energy proton–proton and proton–nucleus collisions. *International Journal of Modern Physics E*, 25(01):1630002, Jan 2016.
- [44] D. Bourilkov. Machine and deep learning applications in particle physics. *International Journal of Modern Physics A*, 34(35):1930019, Dec 2019.
- [45] Matthew Feickert and Benjamin Nachman. A living review of machine learning for particle physics, 2021.
- [46] A. Hoecker et al. TMVA - Toolkit for Multivariate Data Analysis. *arXiv e-prints*, page physics/0703039, March 2007.
- [47] R. Brun and F. Rademakers. ROOT: An object oriented data analysis framework. *Nucl. Instrum. Meth. A*, 389:81–86, 1997.

-
- [48] E. Cuautle et al. Effects produced by multi-parton interactions and color reconnection in small systems. *Nuclear Physics A*, 956:749–752, Dec 2016.
- [49] The ALICE definition of primary particles. Jun 2017.
- [50] A. Ortiz et al. Revealing the source of the radial flow patterns in proton–proton collisions using hard probes. *Journal of Physics G: Nuclear and Particle Physics*, 44(6):065001, Apr 2017.
- [51] A. Golokhvastov. Independent production and Poisson distribution. *Phys. Atom. Nucl.*, 58:1998–2000, 1995.
- [52] B. Abelev et al. Multiplicity dependence of two-particle azimuthal correlations in pp collisions at the LHC. *JHEP*, 09:049, 2013.
- [53] B. Abelev et al. Measurement of inelastic, single- and double-diffraction cross sections in proton–proton collisions at the LHC with ALICE. *The European Physical Journal C*, 73(6), Jun 2013.
- [54] W. Henryk. New ALICE detectors for run 3 and 4 at the CERN LHC. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 958:162116, 2020. Proceedings of the Vienna Conference on Instrumentation 2019.
- [55] Peter J. Huber. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73 – 101, 1964.