



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

Reconstrucción del espectro de rayos cósmicos
con el observatorio HAWC

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Físico

PRESENTA:

Daniel Giovanni Salazar Santiago

TUTOR

Hermes León Vargas





Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



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.

1. Datos del Alumno

Salazar

Santiago

Daniel Giovanni

55 60 71 96 00

Universidad Nacional Autónoma de México

Facultad de Ciencias

Física

310316291

2. Datos del Tutor

Dr.

Hermes

León

Vargas

3. Datos del Sinodal 1

Dr.

Ernesto José María de la Salette

Belmont

Moreno

4. Datos del Sinodal 2

Dr.

Nissim Illich

Fraija

Cabrera

5. Datos del Sinodal 3

Dr.

Ángel

Sánchez

Cecilio

6. Datos del Sinodal 4

Dr.

David

García

Gudiño

7. Datos del trabajo escrito

Reconstrucción del espectro de energía de rayos cósmicos
con el observatorio HAWC

119 p.

2020

Para mi papá Mariano, mi mamá Cristina y mi hermana Brenda. Ustedes son parte importante de mi crecimiento personal y académico. Para Pamela, quien me inspira y apoya para ser mejor cada día.

Daniel Giovanni Salazar Santiago

Agradecimientos

A mi asesor Hermes por su apoyo, paciencia y conocimientos brindados en la realización de mi tesis. También agradezco a mis sinodales por su tiempo y disposición para leer y corregir este trabajo.

Agradezco el apoyo al proyecto CONACyT 254964 y el Fondo Sectorial de Investigación para la Educación.

Resumen

El observatorio HAWC (High Altitud Water Cherenkov) fue construido para realizar estudios sobre rayos gamma de alta energía provenientes del cosmos. Sin embargo, y debido a las características de su funcionamiento, es sensible también para la detección de rayos cósmicos. El presente trabajo tiene como propósito estudiar los rayos cósmicos mediante una de sus características más importantes que es la energía que poseen, de tal forma que el objetivo principal es llevar a cabo una reconstrucción de su espectro de energía.

Los estudios relacionados con la estructura del espectro de energía son de suma importancia ya que dan un indicio sobre las dinámicas por las que se aceleran los rayos cósmicos, el medio por el que viajan e incluso pueden dar información sobre la transición entre fuentes galácticas y extragalácticas. Por esta razón es muy importante llevar a cabo reconstrucciones del espectro de energía confiables, aprovechando las características del detector con que se quieran llevar a cabo. Para el caso de HAWC su altura y el área que cubre hacen viable realizar detecciones en el rango de energía de TeV .

Sin embargo la energía no es detectada directamente debido a que HAWC está diseñado para hacer detecciones indirectas de las partículas secundarias generadas en cascadas atmosféricas, por lo que deben ser aplicados algoritmos de reconstrucción de energía que toman variables medidas de las partículas secundarias (como ángulo de llegada y el tiempo de llegada) para poder estimar la energía de la partícula primaria que generó la cascada.

Actualmente se han desarrollado, por la colaboración de HAWC, dos principales algoritmos de reconstrucción: *Neural Network* y *Ground Parameter*. Ambos métodos están diseñados y optimizados para reconstruir la

energía de los rayos gamma. A pesar de esto, en este trabajo se han decidido implementar para el estudio de rayos cósmicos ya que los resultados pueden establecer un precedente para la creación de estimadores de energía especializados en rayos cósmicos.

Por último, en esta tesis se busca hacer énfasis en la importancia del proceso de *unfolding* dentro de la física experimental para la obtención de resultados más confiables. Este proceso puede llegar a omitirse cuando un fenómeno se estudia solamente de forma teórica, por lo que al introducirse a la física experimental puede resultar confuso. Por esta razón, en el presente trabajo se intenta explicar de manera general y cualitativa en qué consiste dicho proceso, así como la manera en que se implementó mediante un algoritmo llamado *TUnfold*.

Índice general

Resumen	v
1. Rayos cósmicos	1
1.1. Breve historia sobre los rayos cósmicos	1
1.2. Espectro de energía	4
2. Cascadas atmosféricas	7
2.1. Tipos de cascadas	9
2.1.1. Cascadas electromagnéticas	9
2.1.2. Cascadas hadrónicas	12
2.2. Estudio experimental de las cascadas atmosféricas	14
2.3. Estimadores de energía	16
2.3.1. Ground Parameter (GP)	16
2.3.2. Neural Network (NN)	22
3. El observatorio HAWC	27
3.1. Efecto Cherenkov	28
3.1.1. Determinación del ángulo Cherenkov	30
3.2. Tubos fotomultiplicadores (PMT)	33
3.3. Sistema de adquisición de datos (DAQ)	35
3.4. Software utilizado	36
3.4.1. Externals	36
3.4.2. AERIE	36
3.4.3. ROOT	37
3.4.4. CORSIKA	38
3.4.5. GEANT4	38

3.5.	Archivos de simulación y de datos reales	41
3.5.1.	Tipos de datos	41
3.5.2.	Archivos de simulación	42
3.5.3.	Archivos de datos reales	42
3.5.4.	Cortes generales utilizados	44
3.5.5.	Corte extra para <i>Ground Parameter</i>	45
4.	Reconstrucción del espectro de rayos cósmicos	47
4.1.	<i>Unfolding</i>	47
4.1.1.	Problema inverso	48
4.2.	Discretización	48
4.2.1.	Álgebra lineal	49
4.3.	<i>TUnfold</i>	50
4.3.1.	Consideraciones	51
4.3.2.	Uso de <i>TUnfold</i>	52
4.3.3.	Ilustración del procedimiento	53
4.4.	Esquema de bins	57
4.5.	Flujo diferencial de los rayos cósmicos	58
4.6.	Distribución de <i>Bias</i>	61
4.7.	<i>Unfolding</i> sencillo	63
4.7.1.	Conclusiones sobre el caso de <i>unfolding</i> sencillo	68
4.8.	Análisis con <i>TUnfold</i>	70
5.	Resultados	79
5.1.	Índices espectrales obtenidos	79
5.2.	Cambio en los resultados respecto a la cantidad de datos analizados	80
5.3.	Comparación con los resultados de HAWC	84
6.	Discusión y conclusiones	87
A.	Mínimos cuadrados	91
A.1.	Ejemplo introductorio	91
A.2.	Caso general	92
A.3.	Herramientas de álgebra lineal	93
A.4.	Deducción de la pseudo-inversa	96

B. Selección de bins	101
B.1. Errores en la selección de bins	101
C. Comparación de matrices	103
C.1. Cambio en la estadística de una matriz debido al estimador utilizado y al corte extra	103
D. Propagación de errores	107
D.1. Cálculo del error asociado a la curva de área efectiva	107
D.1.1. Evaluación del área efectiva y sus errores asociados en el centro de cada <i>bin</i> de energía	109
D.2. Cálculo del error asociado al flujo diferencial de rayos cósmicos	110
 Bibliografía	 113

Capítulo 1

Rayos cósmicos

1.1. Breve historia sobre los rayos cósmicos

Una de las primeras observaciones realizadas en el campo del electromagnetismo (electrostática en este caso), fue el hecho de que cargas eléctricas del mismo signo se repelen y cargas de signo opuesto se atraen. La interacción entre estas dos cargas fue descubierta en el siglo XVIII por Coulomb y este fenómeno es el principio de la larga carrera en el estudio de los rayos cósmicos. Se presenta a continuación algunos momentos históricos relevantes en el inicio del estudio de los rayos cósmicos.

El dispositivo con el cual comenzó la investigación sobre los rayos cósmicos fue el electroscopio. Éste consiste en dos placas metálicas delgadas que cuelgan de una barra. Al acercar un objeto cargado eléctricamente, el exceso de carga se distribuye en las láminas, las que al cargarse con el mismo signo sufren una repulsión eléctrica [1].

Debido a la conservación de carga eléctrica, un electroscopio aislado debería permanecer cargado, sin embargo en 1785 Coulomb observó que el electroscopio se descargaba lentamente asociando el fenómeno a la interacción de las moléculas de aire con las láminas del electroscopio. En 1836, Faraday explicó que la descarga del electroscopio se debía a partículas cargadas que se encontraban en el aire, pero no explicaba el origen de estas [1].

Una vez descubierta la radiactividad en 1896, Robert Duncan observó que una fuente radiactiva podía descargar un electroscopio. Por su parte, Rutherford propuso que la fuente de partículas cargadas, propuesta por Faraday, eran rocas de la Tierra que con sus diferentes elementos producían radiactividad natural emanada desde el suelo. Para comprobar esta hipótesis se propuso cargar un electroscopio y medir el tiempo de descarga dependiendo de la altura a la que se lleve a cabo la medición. Según lo esperado, a mayor altura debería haber menos radiación y el tiempo de descarga sería mayor [1].

En 1901 C. T. R. Wilson desarrolló la cámara de ionización (o cámara de Wilson) que permitía medir la pérdida de potencial entre dos electrodos con ayuda de un electrómetro y deducir el número de iones formados en el proceso [2]. Posteriormente T. Wulf y W. Kohlhörster desarrollaron instrumentación más avanzada.

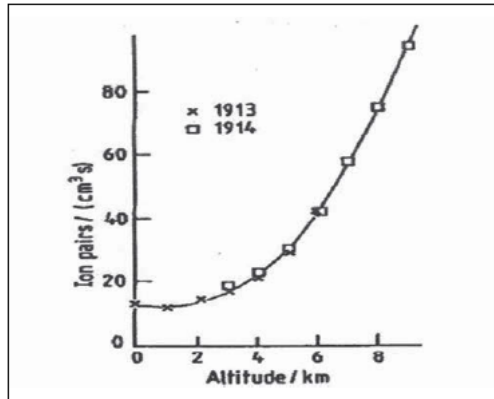
En 1910 A. Gockel realizó las primeras ascensiones en globo hasta 2800 m, sin embargo sus mediciones mostraban fluctuaciones que no permitían extraer conclusiones fiables. Aún con esto, Gockel fue el primero en proponer que la radiación debía proceder de la atmósfera o de un cuerpo celeste [2].

Finalmente en 1912 Victor Hess [3] hizo un vuelo en globo (tras varios vuelos a menores alturas) hasta alcanzar 5350 m y pudo concluir que la radiación provenía del espacio (ver Figura 1.1a). Además de tener un alto poder de penetración, ya que lograba ionizar incluso después de atravesar aproximadamente 10 km de atmósfera, esta radiación no se veía afectada por el día y la noche, y tampoco podía provenir del Sol pues se realizó un ascenso durante un eclipse de Sol y no se mostraron efectos [4].

Los resultados de Hess fueron confirmados en 1914 por Kolhörster [5] quien consiguió ascender hasta 9600 m [2]. Los resultados que obtuvo y su comparación con los resultados de Hess se muestran en la Figura 1.1b.



(a) V. Hess a bordo de un globo sonda (1912)



(b) Medidas de ionización en función de la altitud realizadas por Kolhörster

Figura 1.1: Imágenes históricas representativas del descubrimiento de la radiación cósmica [2].

Hoy se sabe que la composición y abundancia de los rayos cósmicos consta de, aproximadamente, 98 % núcleos atómicos y 2 % de electrones. Del porcentaje de núcleos atómicos el 87 % son protones, 12 % núcleos de helio y el restante 1 % lo componen núcleos pesados [2]. Con el estudio de estas propiedades se obtiene información sobre el origen y propagación de los rayos cósmicos. En la Figura 1.2 se muestra una comparación entre las abundancias relativas de rayos cósmicos galácticos y los elementos presentes en el sistema solar.

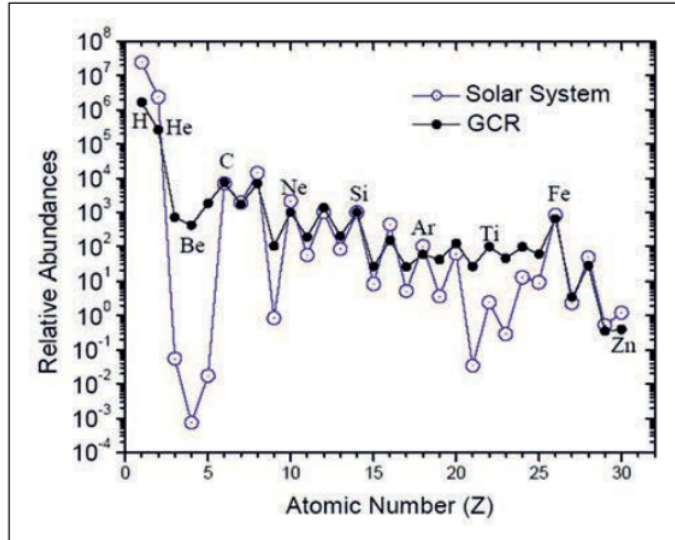


Figura 1.2: Composición y abundancia de los rayos cósmicos galácticos. Las gráficas mostradas han sido normalizadas con la abundancia del elemento $Si = 10^3$. La composición cambia dependiendo de la energía y se conoce relativamente bien hasta el orden de 1 TeV [2]

1.2. Espectro de energía

Otra forma de caracterizar a los rayos cósmicos es midiendo su energía la cual se extiende desde cientos de MeV ($\sim 10^8 \text{ eV}$) hasta 10^{21} eV [6]. El estudio del espectro de energía de los rayos cósmicos se centra en analizar la estructura del mismo ya que las variaciones que presenta pueden dar información sobre la dinámica de los ambientes en donde son acelerados los rayos cósmicos así como del medio que atraviesan [7]. Los cambios en la estructura del espectro pueden ayudar comprender la transición entre fuentes galácticas y extragalácticas [8].

El flujo que incluye a todas las componentes nucleares presentes en los rayos cósmicos se le conoce como *espectro de todas las partículas* y se ha encontrado que puede ser parametrizado por una ley de potencias $N(E)dE \propto E^{-\gamma}dE$, que va cambiando en el valor del índice en distintos

intervalos de energía. En la región de TeV se reporta un índice de $\gamma \approx 2.7$ en M. Aguilar [2], $\gamma \sim 2.65$ en A. D. Erlykin y A. W. Wolfendale [9] y $\gamma = 2.63 \pm 0.01$ en R. Alfaro et al. [7].

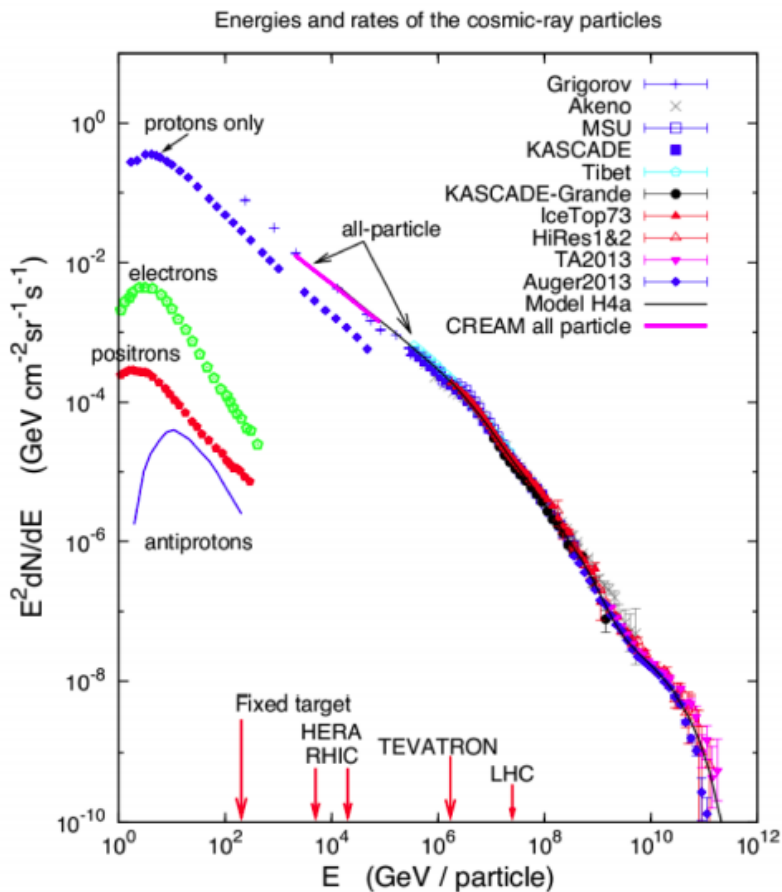


Figura 1.3: Espectro de energía de todas las partículas presentes en los rayos cósmicos [10].

El primer cambio en la estructura sucede alrededor de $3 \times 10^{15} \text{ eV}$ en un punto denominado *knee* (rodilla) en donde el índice cambia a $\gamma \approx 3.0$. El segundo cambio sucede alrededor de $2 \times 10^{17} \text{ eV}$ y se le conoce como

second knee (segunda rodilla) en donde el índice cambia a $\gamma \approx 3.1$ mientras que en $3 \times 10^{18} \text{ eV}$ sucede el tercer cambio denominado *ankle* (tobillo) con $\gamma \approx 2.7$ [2]. En Pasquale Blasi [10] se puede encontrar una explicación sobre los diferentes cambios que tiene el espectro conforme la energía aumenta. En la Figura 1.3 se muestran los espectros de energía medidos por diferentes experimentos.

Los diferentes experimentos destinados a la medición de rayos cósmicos tienen una mejor precisión en ciertos rangos de energía dependiendo de sus características (ver Figura 1.4).

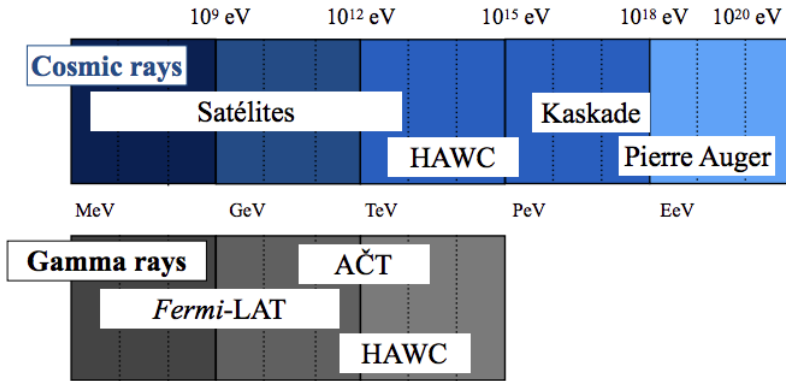


Figura 1.4: Rangos de energía cubiertos por algunos experimentos para rayos cósmicos y rayos gamma [11].

El observatorio de rayos gamma HAWC (*High Altitude Water Cherenkov*) se ubica en un punto medio entre los experimentos con satélites y globos sonda que miden energías por debajo de los TeV y los grandes experimentos como *KASCADE-Grande* (cubre un área de 0.5 km^2) o el *Pierre Auger Observatory* (cubre un área de 3000 km^2) los cuales miden energías por arriba de los PeV . La buena respuesta de HAWC para cascadas hadrónicas permite hacer mediciones por encima de los 10 TeV , en parte debido a que el tamaño de las cascadas producidas por rayos cósmicos en el rango de TeV es comparable con el tamaño del área del observatorio así como la altura del arreglo de tanques se aproxima al máximo de las cascadas a estas energías [7].

Capítulo 2

Cascadas atmosféricas

A los rayos cósmicos que llegan a la Tierra y entran en la atmósfera se les conoce como *primarios*. Si tienen energía suficiente (energías mayores a 100 GeV [12]) interactúan con las moléculas de aire produciendo nuevas partículas llamadas *secundarias* que a su vez son capaces de producir más partículas secundarias en un proceso llamado cascada atmosférica (*extensive air showers*) [13].

El descubrimiento de las cascadas atmosféricas comenzó con el invención de la técnica de coincidencias desarrollada por Walther Bothe en los años 1920's. El acoplamiento de esta técnica con los contadores Geiger-Müller, los circuitos de coincidencias desarrollados por Bothe (resolución en tiempo 1.4 ms) y mejorados por Rossie (resolución en tiempo 0.4 ms) para coincidencias triples, permitieron la detección de eventos cósmicos extraños [14]. En 1933, haciendo medidas de la absorción de rayos cósmicos a un máximo de 101 cm de plomo, Rossie pudo observar y concluir que partículas secundarias eran producidas por los rayos cósmicos que entraban en el material. El crédito por el descubrimiento de las cascadas atmosféricas se asocia a Auger y sus colaboradores (1939), gracias a los desarrollos electrónicos de Maze quien logró crear un circuito de coincidencias con resolución temporal de 5 μ s. Más allá de poder separar detectores hasta 300 m, el gran logro de Auger fue poder estimar que la energía primaria de los rayos cósmicos debía estar alrededor de 10^{15} eV [14]. En la Figura 2.1 se muestra un ejemplo de cascada de partículas.



Figura 2.1: Cascada creada por un rayo cósmico (protón) de aproximadamente 10 GeV dentro de una *spark chamber*. Las características observadas, excepto la escala, son similares a las de una cascada creada por una partícula de alta energía al entrar a la atmósfera terrestre [14].

Las cascadas atmosféricas se caracterizan por tener un frente de partículas secundarias con forma de disco delgado extendido lateralmente en forma curva. Las características principales de una cascada como el contenido de partículas secundarias y la forma de la cascada (qué tan extendida es) son determinadas por el tipo de partícula primaria y su energía. El disco de partículas secundarias se mueve a velocidades cercanas a la de la luz, tiene un ancho y es de forma curva. Debido a que la partícula primaria llega a la Tierra con un *ángulo cenital* específico, la dirección del *eje de la cascada* estará definida por dicho ángulo. A la proyección del eje sobre el suelo se le conoce como *núcleo* de la cascada. Cerca del núcleo, el frente de la cascada es compacto mientras que se hace más delgado y poco denso al extenderse lateralmente. Estas características se pueden observar en el diagrama de la Figura 2.2.

2.1. Tipos de cascadas

Las cascadas atmosféricas pueden ser de dos tipos según el tipo de partícula primaria que las inicia: electromagnéticas o hadrónicas. A continuación se describen las características de estos dos tipos de cascada así como su mecanismo de creación.

2.1.1. Cascadas electromagnéticas

La partícula primaria es electromagnética, es decir, un fotón (rayos gamma), electrones o positrones y durante el proceso se generan fotones, electrones y positrones. De las interacciones que suceden en estas cascadas las más relevantes son:

- *Producción de pares.* Es el proceso predominante a altas energías (arriba de MeV). Consiste en la creación de una partícula y su antipartícula, siendo el caso más común la creación de un par electrón positrón por el paso de un fotón en el campo eléctrico del núcleo de un átomo. A su vez este par puede aniquilarse para dar lugar a un fotón energético (otro rayo gamma) [13].
- *Bremsstrahlung.* Radiación emitida por un electrón (positrón) cuando este es desviado de su trayectoria original (desaceleración de una partícula cargada) por la fuerte atracción eléctrica de otra partícula cargada (usualmente un núcleo atómico). La energía cinética perdida es proporcional a $(\frac{E}{mc^2})^4$ por lo que el proceso es importante para partículas poco masivas como los electrones y positrones, pero no para muones, piones o protones.

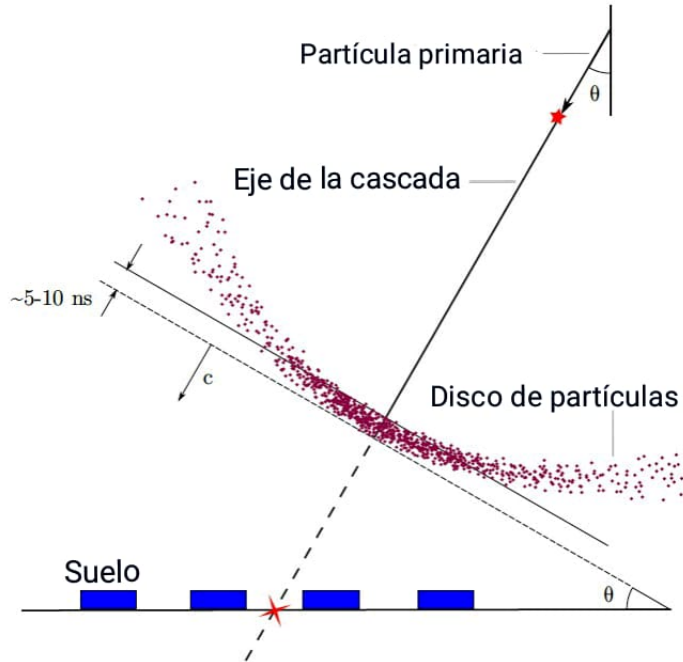


Figura 2.2: Diagrama de una cascada atmosférica (EAS). La estrella de cuatro picos representa la ubicación del núcleo mientras que la estrella de cinco picos representa la primera interacción de la partícula primaria con la atmósfera [12]

Otros posibles procesos físicos.

- *Efecto Compton.* En este proceso un fotón aumenta su longitud de onda (pierde energía) después de ser dispersado por un electrón. Esto sucede a energías menores con las que la probabilidad de que ocurra producción de pares es muy baja.
- *Efecto fotoeléctrico.* Un fotón es absorbido por un átomo mientras un electrón es expulsado de una de sus capas. Este proceso es dominante a energías menores a 1 MeV.
- *Radiación Cherenkov.* Cuando una partícula se mueve más rápido que la velocidad de la luz en un medio, ésta emite un fotón. En las

cascadas atmosféricas se tiene este efecto en el aire, sin embargo para este trabajo se utiliza la radiación Cherenkov producida en agua, la cual se explicará mejor en los capítulos siguientes.

- *Ionización.* Pérdida de energía de los electrones debido a la ionización de un átomo. Para energías por debajo de $\sim \frac{600}{Z} MeV$ (Z número atómico) este proceso toma importancia.

La forma general en que se crea una cascada electromagnética es la siguiente:

1. Entrada del rayo gamma primario a la atmósfera para dar lugar a la producción de pares electrón-positrón. Si la partícula primaria es un electrón o un positrón, el proceso se toma desde el siguiente paso (interacción de los electrones o positrones), sin embargo es muy difícil detectar que la partícula primaria no sea un rayo gamma ya que al ser un proceso iterativo la cascada resultante sería prácticamente igual a la iniciada por un fotón.
2. Interacción de los electrones o positrones.
 - a) Bremsstrahlung.
 - b) Aniquilación electrón-positrón.
3. Repetición del proceso. Además de tener pérdidas de energía por efecto Compton.

Es posible que en la producción de pares se creen muones pero la probabilidad de que esto suceda es muy baja [15].

El modelo de *Heitler* (publicación original [16]) describe las características generales de las cascadas electromagnéticas (ver Figura 2.3). Asumiendo que la creación de nuevas partículas ocurre a intervalos de *longitud de interacción* λ , se tiene que la cantidad de partículas se duplica en cada λ con la energía de la partícula creadora distribuida equitativamente entre las dos partículas producidas. Si la energía del rayo gamma primario es E_0 , después de n longitudes de interacción, el número de partículas es $N = 2^n$ cada una con energía $E = E_0/2^n$. La cascada sigue creciendo hasta llegar a un máximo de producción de partículas en donde tiene una energía crítica E_c (que

depende del medio: 84 MeV para electrones en aire). La cascada alcanza su profundidad máxima a $\lambda \log_2 \frac{E_0}{E_c}$ longitudes de interacción con un número máximo de partículas $N_{max} = \frac{E_0}{E_c}$ [12]. A pesar de ser un modelo simple, este modelo relaciona la energía primaria con el número de partículas en un estado de desarrollo dado de la cascada electromagnética.

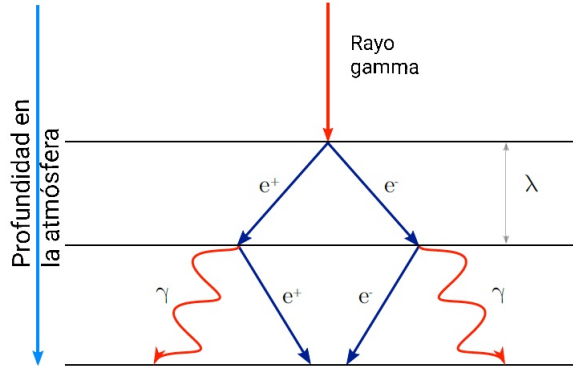


Figura 2.3: Diagrama de una cascada electromagnética [12].

2.1.2. Cascadas hadrónicas

1. El proceso comienza con una colisión nuclear entre un rayo cósmico hadrónico (suele ser un protón) y un núcleo de la atmósfera. Típicamente se generan más de 50 partículas secundarias en la primer colisión siendo la mayoría piones cargados y neutros [17].

A partir de esta primera interacción se pueden identificar tres componentes diferentes de la cascada: la componente electromagnética, la componente muónica y la componente nucleónica.

- a) Electromagnética. Los piones neutros π^0 tienen un vida media muy corta ($[8.52 \pm 0.18] \times 10^{-17}$ s [18]) por lo que decaen en un par de fotones antes de poder interactuar con núcleos en la atmósfera. Debido a la alta energía, estos fotones son rayos gamma por lo que comienza el proceso de cascada electromagnéticas descrito en la Sección 2.1.1.

- b) Muónica. Al igual que los piones neutros, los piones cargados π^\pm son inestables. Sin embargo tienen una vida media mucho mayor ($2.6033 \pm 0.0005 \times 10^{-8}$ s [18]) y esto permite que puedan colisionar con otros núcleos antes de decaer. Después de varias generaciones, los piones cargados pierden energía hasta finalmente llegar a su principal modo de decaimiento:

$$\pi^+ \rightarrow \mu^+ + \nu_\mu$$

$$\pi^- \rightarrow \mu^- + \bar{\nu}_\mu$$

- c) Nucleónica. Consiste en los fragmentos de las moléculas de aire, protones y neutrones. La interacción que causa pérdida de energía en los protones es la electromagnética mientras que los neutrones pasan de largo a través de la atmósfera hasta llegar al suelo [13].

Las cascadas hadrónicas (ver Figura 2.4) se desarrollan cualitativamente de forma similar a las cascadas electromagnéticas sin embargo las diferencias que presentan hacen que no se puedan estudiar exactamente de la misma forma. En [17] se menciona que las cascadas hadrónicas pueden ser consideradas como una superposición de cascadas electromagnéticas producidas por los decaimientos de π^0 y alimentadas por interacciones hadrónicas por lo que tiene sentido tratar de estudiar este tipo de cascadas con un modelo de Heitler modificado similar al de las cascadas electromagnéticas [12].

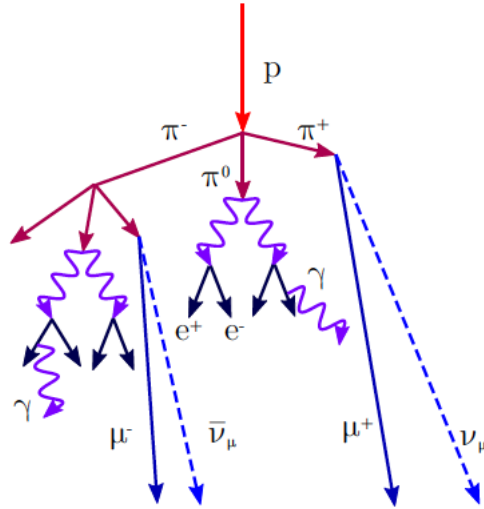


Figura 2.4: Diagrama de una cascada hadrónica [12].

2.2. Estudio experimental de las cascadas atmosféricas

Para llevar a cabo el estudio de las cascadas atmosféricas se deben idear experimentos con la capacidad de obtener algunos parámetros observables.

- *Desarrollo longitudinal.* Describe el número de partículas cargadas en función de la profundidad atmosférica. La masa y energía de la partícula primaria determina el perfil longitudinal de la cascada [19]. En experimentos de detectores de superficie como el *Pierre Auger Observatory* se estudia este parámetro, haciendo uso de detectores de fluorescencia, con el fin de resolver la composición de las partículas primarias [20].
- *Función de distribución lateral (LDF).* El desarrollo lateral describe el número de partículas cargadas en función de la distancia radial desde el eje de la cascada. Luego la *LDF* es una función que refleja el desarrollo lateral al nivel del suelo [19] [21]. En experimentos como

HAWC se utiliza un módulo del software AERIE llamado *LatDist* el cual toma los eventos detectados (cascadas) y para cada *hit* (señal registrada en un detector producida por una partícula cargada perteneciente a una cascada) calcula su distancia al núcleo de la cascada para luego realizar un ajuste de esta distribución con una función *NKG* (Nishimura-Kamata-Greisen) (ver Ecuación 2.1). Esto resulta útil para estimar la energía de la partícula primaria [22]

- *Parámetro de edad*. La función *NKG* que se ajusta a la LDF tiene la siguiente forma:

$$\begin{aligned} \log_{10}(NKG) = & \log_{10}(A) + s \left[\log_{10}\left(\frac{r}{R_m}\right) + \log_{10}\left(1 + \frac{r}{R_m}\right) \right] \\ & - 3 \log_{10}\left(\frac{r}{R_m}\right) - 4.5 \log_{10}\left(1 + \frac{r}{R_m}\right) \end{aligned} \quad (2.1)$$

en donde r es la distancia radial al eje de la cascada, R_m es el Radio de Molire que incorpora el efecto de la dispersión de Coulomb [12] y tiene un valor de $\simeq 79$ m en el nivel del mar y 124.21 m a la altura de *HAWC* [21]. Los únicos parámetros libres son: A , que es un factor de normalización (de tal forma que al integrar sobre todos los valores de r de la función *NKG* se obtiene el número total de partículas cargadas en la cascada) y s , el *parámetro de edad* de la cascada el cual da una idea del desarrollo de la cascada. $s = 0$ indica el inicio de la cascada, $s < 1$ indica que la cascada es "joven", $s = 1$ es el máximo de la cascada, para $1 < s < 2$ la cascada ya ha alcanzado su máximo y va decreciendo, por último $s = 2$ indica que la cascada se ha extinguido [22], [17].

Estas características están descritas para cascadas electromagnéticas, sin embargo ofrecen una buena descripción para cascadas hadrónicas a energías $E \lesssim 50$ TeV [23]. En el observatorio *HAWC* se sigue trabajando para tratar de ampliar el rango de energía en el que se obtiene una buena descripción para cascadas hadrónicas [24].

2.3. Estimadores de energía

La energía que tienen las partículas primarias no puede medirse de forma directa con el observatorio HAWC. Para poder estimar esta energía a partir de las variables obtenidas por el observatorio es necesario hacer un algoritmo de estimación. En la colaboración se ha trabajado para poder obtener un correcto estimador de energía teniendo como primer aproximación la estimación mediante el número de detectores activados. Sin embargo este método se correlaciona débilmente con la energía, ya que deja de lado algunas variables importantes como el ángulo cenital, la distancia del núcleo de la cascada al detector y que tan bien está contenida la cascada en el arreglo del observatorio, entre otras. Además de que a una energía límite (~ 30 TeV), todas las cascadas con energía superior dejarían una señal en cada detector haciendo imposible estimar la energía solo usando el porcentaje de detectores que registran señal [25]. La necesidad de estimadores más ciertos ha llevado al desarrollo de dos estimadores de energía independientes uno del otro: *Ground Parameter* y *Neural Network*. Estos dos métodos han mostrado buenos resultados validados para la nebulosa del Cangrejo, que suele ser el objeto celeste con el que se calibra el observatorio debido a su brillo a altas energías en rayos gamma [26]. Debido a que fueron desarrollados para observar rayos gamma no se puede asegurar con total certeza su eficacia para estimar la energía de rayos cósmicos.

Los dos estimadores fueron desarrollados haciendo uso de las simulaciones estándar Monte Carlo utilizadas en HAWC. Como se ve en la Sección 3.4, las simulaciones constan de cascadas atmosféricas creadas con CORSIKA (v7.4000), la propagación de partículas secundarias a través de HAWC y hasta los PMT's mediante GEANT4 (v4.10.00) y por último DAQsim para el sistema de adquisición de datos.

En este trabajo se utilizarán ambos estimadores para poder comparar los resultados y determinar cual se adapta mejor a lo requerido para la reconstrucción del espectro de energía de los rayos cósmicos.

2.3.1. Ground Parameter (GP)

En el observatorio HAWC los detectores miden la carga depositada en los PMT debido a un evento. Es precisamente esta medición la que ayuda a

estimar la energía mediante el algoritmo GP ya que la carga medida a cierta distancia del eje de la cascada es una forma de estimar la energía primaria [27]. Luego la función de distribución lateral (LDF) describe el número de partículas cargadas en función de la distancia al eje de la cascada (ver la Sección 2.2) por lo que es necesario contar con una apropiada descripción de la LDF.

La función de distribución lateral (LDF)

Históricamente el estudio de las EAS ha sido llevado a cabo tratando de describir las cascadas electromagnéticas mediante el entendimiento de su estructura lateral y angular. Los trabajos de científicos como Molière, Roberg y Nordheim, Eyges y Fernbach, Chartres y Messel aportaron teorías sobre esta investigación, sin embargo las aplicaciones resultaban muy limitadas [28, p. 94]. En [29] Nishimura y Kamata desarrollan un método matemático para dar solución a las ecuaciones de difusión de la función de distribución lateral y angular propuestas por Landau en 1940.

Para realizar su cálculo hicieron uso de las siguientes aproximaciones: la tendencia de las cascadas a extenderse es debido a la deflexión que sufren los electrones por la dispersión de Coulomb al atravesar la materia, las deflexiones debidas a radiación o creación de pares son despreciables (excepto al inicio de la cascada). La teoría tridimensional de cascadas puede ser construida a partir de la teoría en una dimensión más la teoría de dispersión de Coulomb múltiple. Para la teoría de cascadas en una dimensión usaron lo que se conoce como *aproximación B* ([30]) en la cual se asume que los electrones pierden su energía únicamente por procesos de ionización y radiación mientras que los fotones pierden energía por la creación de pares. Otras interacciones electromagnéticas son despreciables. Para la probabilidad de radiación y producción de pares utilizaron la sección transversal de Bethe-Heitler [31]. Se asume que la pérdida por ionización de un electrón es independiente de su energía. Estas suposiciones son válidas para partículas de alta energía y medios con número atómico bajo ya que en estas condiciones los errores debidos a dichas suposiciones no son relevantes [28].

A partir de estas consideraciones Nishimura y Kamata presentan la solución analítica para la función de estructura lateral y muestran los resultados de la evaluación numérica de la misma para diferentes valores del

parámetro de edad [28, Figs. 2-5]. En la Figura 2.5 se muestra uno de sus resultados.

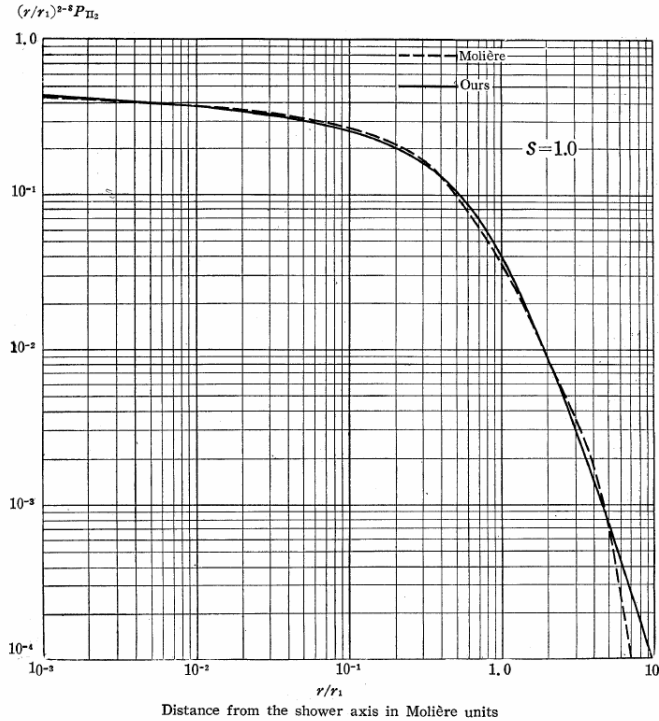


Figura 2.5: Función de estructura lateral normalizada para $s = 1$ en la aproximación B obtenida a partir de resultados numéricos [28]. Se compara con el resultado obtenido por Molière cuya solución era la más utilizada antes de la de Nishimura y Kamata. Se observa que ambas soluciones son congruentes para el máximo de la cascada ($s = 1$).

Debido a que la evaluación de las funciones de estructura requerían de largos y tediosos cálculos numéricos Nishimura y Kamata desarrollaron una fórmula asintótica, a partir de los resultados numéricos, más simple para aplicaciones prácticas (ver [28, p. 140]). Luego Greisen propone una fórmula empírica más simple [32] (ver Figura 2.6).

$$\rho(r/R_m) = c(s) \left(\frac{r}{R_m}\right)^{s-2} \left(1 + \frac{r}{R_m}\right)^{s-4.5}, \quad (2.2)$$

en donde $c(s)$ es un factor de normalización, r es la distancia radial al eje de la cascada, R_m es el Radio de Moliere y s es el parámetro de edad.

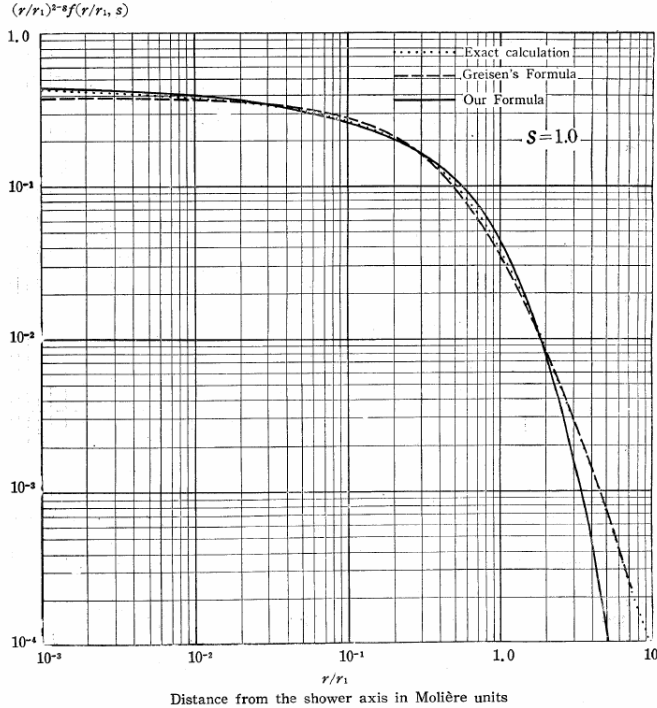


Figura 2.6: Comparación entre los cálculos exactos, la fórmula de Greisen y la fórmula de Nishimura y Kamata para $s = 1$ publicada en K. Kamata y J. Nishimura [28] (r_1 es el radio de Molière en esta publicación).

La Ecuación 2.2 ajusta de manera correcta la distribución de Nishimura y Kamata para $s = 1$ y $s = 1.4$ sobre todo el rango de r/R_m y debido a que la distribución debe variar de forma suave con s se espera que no tenga grandes errores para valores intermedios de s entre 1 y 1.4 e incluso ligeramente más allá de este rango; $0.8 \leq s \leq 1.6$. Este último rango cubre

la región más significativa en los experimentos de EAS [32, p. 27] [33, p. 806].

En P. Grieder [33, p. 199] se muestra una aproximación explícita para $c(s)$, con lo que la Ecuación 2.2 queda como:

$$\rho(N_e, \frac{r}{R_m}) = \frac{N_e}{2\pi r_m^2} \left(\frac{\Gamma(4.5 - s)}{\Gamma(s)\Gamma(4.5 - 2s)} \right) \left(\frac{r}{R_m} \right)^{s-2} \left(1 + \frac{r}{R_m} \right)^{s-4.5} [m^{-2}], \quad (2.3)$$

en donde N_e es el número total de electrones (o partículas cargadas) de la cascada y Γ es la función gamma común. Esta expresión es conocida como la función NKG (Nishimura-Kamata-Greisen). En [28, p. 126] Nishimura y Kamata muestran que la densidad de energía decae $1/r$ veces más que la densidad de partículas por lo que un factor de $1/r$ es introducido a la función NKG. Con un poco de álgebra y escrito en forma logarítmica la Ecuación 2.3 puede ser llevada a la forma:

$$\begin{aligned} \log_{10}(\rho(r/R_m)/r) &= \log_{10}(A) + s [\log_{10}(\frac{r}{R_m}) + \log_{10}(1 + \frac{r}{R_m})] \\ &\quad - 3 \log_{10}(\frac{r}{R_m}) - 4.5 \log_{10}(1 + \frac{r}{R_m}), \end{aligned} \quad (2.4)$$

en donde A es un factor de normalización general. Esta última expresión es utilizada en [22] en el estimador de energía de la partícula primaria.

Radio óptimo

Para el método de GP se utiliza el radio óptimo r_{opt} en el que la carga medida minimiza los errores debidos a la discrepancia con la LDF exacta. Dicho radio debe estar lo suficientemente lejos del eje de la cascada debido a que las grandes fluctuaciones que hay entre una cascada y otra hacen que sea difícil estimar la energía pero también debe estar lo suficientemente cerca tal que el ruido de las señales provocadas por la electrónica del experimento no sean comparables con las señales que se desean medir. Luego para encontrar el radio óptimo se hace una gráfica con la carga efectiva depositada en los PTM dependiendo de la distancia al eje de la cascada. Se le ajusta la función NKG (Ecuación 2.4) y después de obtener el mejor ajuste se varía el parámetro de edad $s \pm 10\%$ para obtener una "banda" de ajustes. El

punto en el que dicha banda es más delgada (ver Figura 2.7) es donde se minimizan las incertidumbres por no conocer la LDF exacta [27].

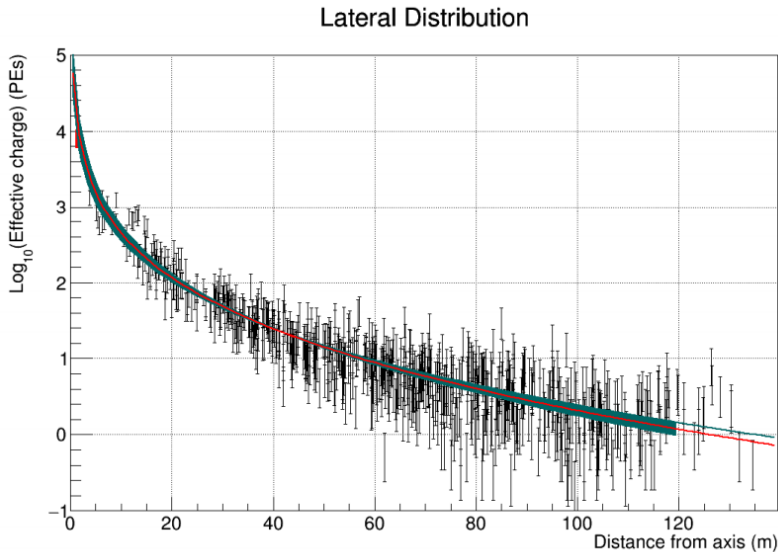


Figura 2.7: Ejemplo de la determinación del radio óptimo para eventos simulados. En este caso el radio óptimo es aproximadamente 41 m [27]. La línea roja representa el mejor ajuste con la función NKG mientras que la banda azul representa la variación del parámetro de edad.

En [34] se muestra que la ubicación del radio óptimo depende sobre todo de la geometría del arreglo del detector y no varía significativamente con el ángulo cenital, por lo que para HAWC se utiliza $r_{opt} = 40\text{ m}$. La Ecuación 2.4 es evaluada en $r = 40\text{ m}$. Este resultado se conoce como $\log_{10}(sig_{40})$ y es luego traducido a energía.

Debido a que la señal medida a nivel de suelo (*ground*) varía con el ángulo cenital de la partícula primaria, para poder ajustar la energía se debe tomar en cuenta tanto la señal medida a $r = 40\text{ m}$ como el ángulo cenital θ : $\log_{10}\hat{E} = f(sig_{40}, \theta)$. La fórmula para ajustar la energía fue desarrollada de forma empírica de tal forma que coincida con la simulación de eventos:

$$\log_{10}\widehat{E} = m(\theta) \log_{10} sig_{40} + c(\theta), \quad (2.5)$$

en donde $m(\theta)$ se escoge para ser una función lineal por partes y $c(\theta)$ una función cuadrática por partes [25][27].

A pesar de que la teoría de EAS fue desarrollada para el caso electromagnético se suele usar la función NKG, ya sea simple o modificada, para la descripción de cascadas iniciadas por hadrones. Esto se justifica por la contribución electromagnética dominante en este tipo de cascadas. Aproximadamente el 90 % del número de partículas presentes en una cascada iniciada por un hadrón consisten en electrones (y positrones) y fotones por lo que el parámetro de edad es una observable que también resulta conveniente para el estudio de cascadas hadrónicas. Estas consideraciones resultan ser adecuadas en muchas ocasiones. En este trabajo se realizan pruebas para comprobar si es adecuado el uso de este estimador de energía de rayos gamma para aproximar la energía de rayos cósmicos.

En [35, p. 194] se muestra una fórmula parecida a la función NKG para cascadas puramente electromagnéticas agregando un factor que la corrige para cascadas hadrónicas:

$$\rho(r/R_m) = \frac{c_1(s)N_e}{2\pi R_m^2} \left(\frac{r}{R_m}\right)^{s-2} \left(1 + \frac{r}{R_m}\right)^{s-4.5} \left(1 + c_2\left(\frac{r}{R_m}\right)^\delta\right). \quad (2.6)$$

En donde Greisen recomienda que $s = 1.25$, $\delta = 1$ y $c_2 = 0.088$ para cascadas de $N_e = 10^6$ partículas al nivel del mar. Esta fórmula no está implementada en HAWC pero podría ser interesante para estudios de rayos cósmicos de origen hadrónico.

2.3.2. Neural Network (NN)

Este algoritmo utiliza una red neuronal (NN) artificial para realizar un estimado de la energía primaria de rayos cósmicos. Esta red neuronal es una complicada función que mapea diversas variables relacionadas a una cascada atmosférica a una variable de salida, que en este caso es la energía de la partícula primaria: $\log_{10} E$. Esta función utiliza 479 parámetros libres llamados *pesos* los cuales son llevados a sus valores óptimos mediante un

proceso de entrenamiento. Los 479 pesos son escogidos para minimizar la función de error:

$$D(\mathbf{w}) \equiv \frac{1}{2} \sum_{i=1}^n u_i \left[\log_{10} \widehat{E}(\mathbf{x}_i; \mathbf{w}) - \log_{10} E_i \right]^2. \quad (2.7)$$

La función anterior es evaluada usando eventos simulados. En donde \mathbf{w} es el vector de pesos para la red neuronal, u_i es la importancia relativa del i -ésimo evento, \mathbf{x}_i es el vector de variables de entrada del i -ésimo evento, \widehat{E} es una función que devuelve la energía estimada para un vector de entradas y vector de pesos, y E_i es la energía verdadera del i -ésimo evento simulado [25].

Variables de entrada

Para poder estimar la energía se utilizan quince variables calculadas por la reconstrucción de eventos de HAWC para describir tres características de las cascadas; *multiplicidad*, *contención* y *atenuación atmosférica* [25], [36].

Multiplicidad

Es una medición aproximada de qué tanta energía de la cascada llegó al detector. Se utilizan tres variables para cuantificar la multiplicidad general de una cascada.

- Fracción de canales activados: se define como el número de canales activados en el evento entre el número total de canales en funcionamiento del detector.
- Fracción de tanques activados: se define como el número de tanques activados entre el número total de tanques en HAWC. Un tanque se considera activado si al menos uno de sus PMTs es activado.
- Parámetro general de normalización obtenido del ajuste del núcleo: el ajuste utilizado para el núcleo no proviene de una función NKG, en su lugar se utiliza una aproximación llamada *Super Fast Core Fit* (SFCF) que en general no es mejor que la misma función NKG pero es suficiente para los fines de ajustar el núcleo. Esta cantidad debería ser en principio un mejor indicador de las

cascadas que quedan fuera del arreglo que la fracción de canales o tanques activados ya que con un ajuste ideal el parámetro de normalización es independiente de la contención de la cascada. En la práctica este parámetro es uno de los menos relevantes para las entradas de la red neuronal.

Contención

Esta característica indica cuánto debe ser escalada la multiplicidad. Solo se utiliza una variable para inferir la fracción de cascada contenida dentro del arreglo: la distancia del núcleo reconstruido al centro del arreglo.

Atenuación atmosférica

Está relacionada con la cantidad de energía que pierde la cascada en su paso por la atmósfera hasta el nivel de detección. Once variables son necesarias para introducir la atenuación; una es el ángulo cenital ya que cascadas con mayor ángulo cenital han atravesado una cantidad mayor de atmósfera. Las otras diez variables son conocidas como *variables anulares* y representan la distribución lateral de carga en el frente de la cascada. El propósito de estas variables es estimar la altura a la que sucedió la primer interacción pues esto también indica la cantidad de atmósfera que se ha atravesado. También se puede utilizar el parámetro de edad de la función NKG pero esto no ha demostrado ser relevante para la estimación de energía con la red neuronal. Las variables anulares mapean de una mejor forma empírica la distribución lateral hasta encontrar la altura de la primer interacción. Para una descripción más técnica de estas variables ver S. Marinelli [36, p. 3].

La función que mapea las entradas en energía, resultante del entrenamiento de la red neuronal es de quince dimensiones por lo que para visualizarla se realiza un gráfico de su dependencia con cada una de las variables y se mantienen las demás congeladas en su valor medio calculado con simulaciones. La Figura 2.8 muestra un ejemplo de lo anterior.

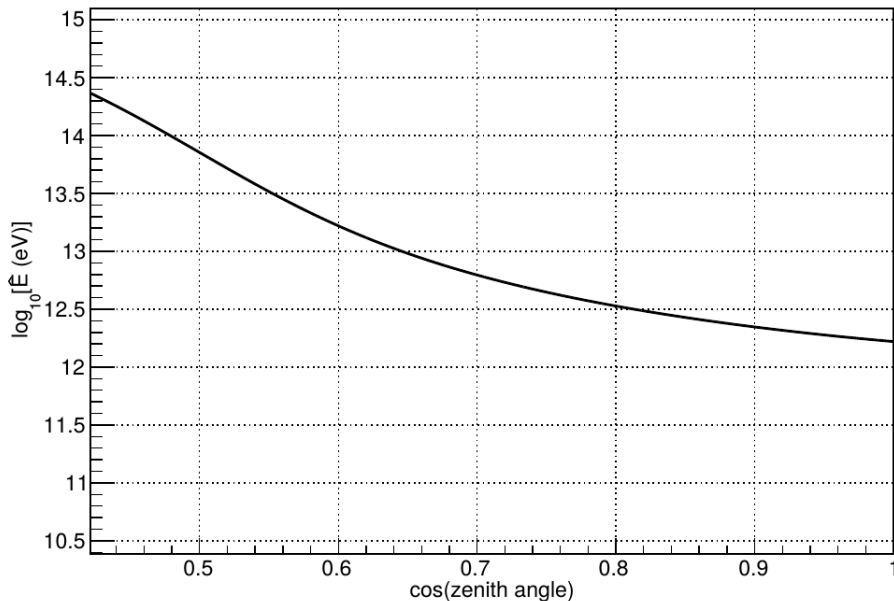


Figura 2.8: Dependencia de la energía estimada por NN con respecto al coseno del ángulo cenital. La dependencia con las otras 14 variables se muestra en S. Marinelli [36].

La dependencia esperada entre la energía y el ángulo cenital puede ser obtenida mediante la fórmula para calcular el número de electrones N_e en cascadas iniciadas por rayos gamma dada en [37, p. 309].

$$N_e \sim \frac{\alpha}{\sqrt{\ln \frac{E}{E_c}}} \exp \left[t \left(1 - \frac{3}{2} \ln s \right) \right], \quad (2.8)$$

en donde $\alpha = 0.31$, E_c es la energía crítica del electrón debajo de la cual domina la ionización sobre el efecto bremsstrahlung, t es la profundidad atmosférica y s es el parámetro de edad. Esta relación puede ser llevada a:

$$\log_{10} E \sim \log_{10} E_c + \frac{\alpha^2 \log_{10} e}{N_e^2} \exp \left[2t \left(1 - \frac{3}{2} \ln s \right) \right],$$

y reescribiendo la profundidad atmosférica en términos del ángulo cenital

$t \equiv k/\cos\theta$ se obtiene:

$$\log_{10}E \sim \log_{10}E_c + \frac{\alpha^2 \log_{10} e}{N_e^2} \exp \left[\frac{2k}{\cos\theta} \left(1 - \frac{3}{2} \ln s \right) \right],$$

que para ajustar el resultado de la dependencia cenital de la energía estimada por la red neuronal se utiliza de la forma:

$$\log_{10}E \equiv A + B \exp \frac{C}{\cos\theta}. \quad (2.9)$$

La ecuación anterior, obtenida de la teoría de cascadas, reproduce de manera cualitativa la relación observada en la Figura 2.8 [36]. La congruencia entre el resultado obtenido por la red neuronal para la dependencia de la energía con el ángulo cenital y el resultado teórico sugiere que la estimación de energía mediante NN es confiable.

Capítulo 3

El observatorio HAWC

El observatorio HAWC (*High Altitude Water Cherenkov*) se encuentra en las laderas del volcán Sierra Negra, Puebla a 4100 m de altura sobre el nivel del mar. Fue diseñado para detectar rayos gamma (RG) de muy alta energía (entre 100 GeV y 100 TeV), sin embargo tiene la capacidad de detectar rayos cósmicos con energías de hasta PeV . El observatorio es un arreglo, a nivel de suelo, de 300 tanques detectores de radiación Cherenkov en agua o WCD (Water Cherenkov Detector) que cubren un área de 22,000 m^2 . Cada WCD mide 5 m de alto por 7.3 m de diámetro y contiene 200,000 litros de agua purificada (hasta 4.5 m de altura) además de contar con 4 detectores de luz llamados tubos fotomultiplicadores (PMT) los cuales permiten transformar la señal de la radiación Cherenkov en datos para procesar y almacenar en el centro de cómputo ubicado en el sitio. Los PMT's son colocados en el fondo del WCD: tres de ellos (de 8 pulgadas de diámetro) son colocados de forma lateral a 1.83 metros de distancia del centro del tanque separados entre sí por 120° , mientras que un cuarto PMT (de 10 pulgadas de diámetro) se coloca en el centro [38], [39].

Los objetivos principales de la colaboración de HAWC son contribuir al estudio sobre el origen de los rayos cósmicos, mediante los rayos gamma provenientes de fuentes astronómicas, como el plano Galáctico y la Nebulosa del Cangrejo, y tratar de entender los mecanismos de aceleración de partículas en los eventos más violentos (energéticos) del Universo. La idea general del observatorio es detectar rayos cósmicos y rayos gamma mediante las cascadas atmosféricas que estos producen. Dichas cascadas no se detec-

tan de forma directa, lo que se detecta en última instancia es la radiación Cherenkov producida por el paso de las partículas cargadas de las cascadas a través del agua. A continuación se hace una descripción más detallada del efecto Cherenkov en agua.

3.1. Efecto Cherenkov

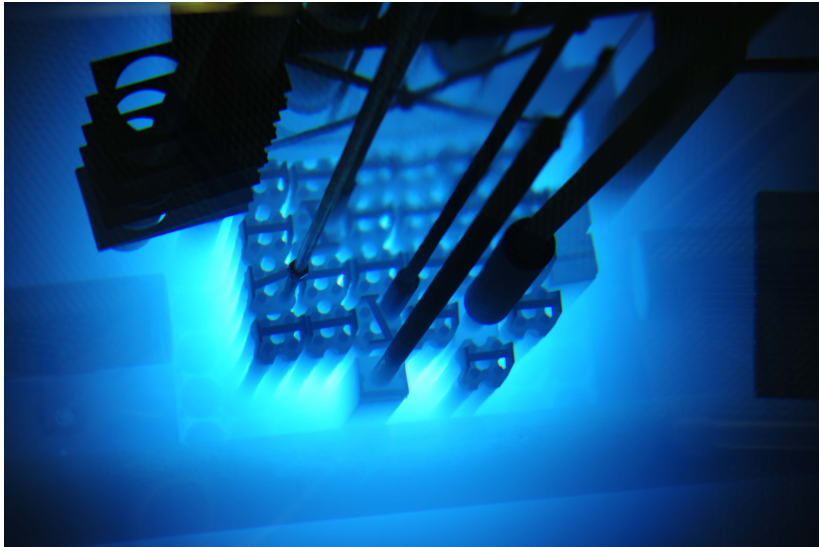


Figura 3.1: Efecto Cherenkov en un reactor nuclear. Imagen tomada de <http://www.philrutherford.com/images>.

Es bien sabido que la teoría de la relatividad especial impide que exista objeto alguno capaz de viajar a velocidades superiores a la de la luz, claro, en el vacío. Y es esta última condición la que permite la existencia del Efecto Cherenkov que ocurre cuando una partícula cargada viaja más rápido que la luz a través de un medio dieléctrico (como el agua, en la cual la velocidad de la luz se reduce a un 75 %). El resultado de este fenómeno es la emisión de luz azul-ultravioleta llamada radiación Cherenkov.

Las primeras investigaciones en este tipo de fenómeno fueron llevadas a cabo en 1889 por Oliver Heaviside. En la publicación titulada *On the*

electromagnetic effects due to the motion of electrification through a dielectric [40, p. 515], se encuentra la sección *The State when the Speed of Light is exceeded* con un análisis sobre si una partícula cargada eléctricamente podría emitir radiación sin acelerar, es decir, solo viajando a una velocidad constante. Sin embargo, debido a la nula investigación y experimentación respecto a este tema en esa época (ni siquiera existía la relatividad especial), las ideas de Heaviside no trascendieron y solo fueron recuperadas hasta 1974 por Kaiser y Tyapkin [41].

Para entender de una mejor forma este efecto, se suele hacer una analogía con el efecto acústico de las ondas de choque producidas por un avión (o un objeto en general) al superar la barrera del sonido en el aire. Esta idea ya había sido concebida de esta forma en 1901 por Lord Kelvin quien propuso que si la velocidad uniforme de un átomo excede la de la luz, por pequeña que sea la diferencia, se produciría un onda cónica de manera similar al efecto de la barrera del sonido (efecto ya ilustrado por Ernst Mach con sus fotografías de proyectiles en pleno vuelo). Kelvin propuso que este efecto podía llevarse a cabo ya fuera en aire o el éter. Las ideas de Kelvin también fueron ignoradas [41].

Arnold Sommerfeld se planteó la misma pregunta sobre una partícula cargada viajando más rápido que la luz y en 1904 encontró que un electrón superlumínico a velocidad constante emitiría radiación pero en 1905 llegó a relatividad especial de Einstein (nada puede moverse a una velocidad mayor que la de la luz) y su estudio tampoco tuvo repercusión [41].

Después de estas ideas empezaron las primeras observaciones experimentales, la primera de ellas llevada a cabo por Marie y Pierre Curie cuando vieron (a principios del siglo XX) un brillo azul en botellas con sales de radio y lo asociaron erróneamente a un efecto de luminiscencia [41].

Como una coincidencia curiosa dentro de la física, en el mismo año en que Sommerfeld pensaba en la idea de partículas más rápidas que la luz, nació el físico soviético Pável Cherenkov quien trabajó su tesis doctoral, bajo la supervisión de Sergei Vavilov, sobre la luminiscencia activada por rayos gamma en líquidos. En sus experimentos observo un resplandor azul que, al igual que los Curie, atribuyó a la luminiscencia. Sin embargo su trabajo fue mucho más lejos al tratar de medir y caracterizar dicha luz azul y junto al trabajo teórico de sus colegas Ilya Tamm e Igor Frank describieron el fenómeno de radiación de Vavilov-Cherenkov el cual, concluyeron, ocurre

cuando una partícula cargada eléctricamente supera la velocidad de la luz en un medio material y pierde energía emitiendo radiación. Conclusión que no contradice a la relatividad especial ya que el límite insuperable es el de la velocidad de la luz en el vacío. Esta investigación en conjunto les valió el Premio Nobel en 1958 a Cherenkov, Tamm y Frank [41] (ver Figura 3.2).



Figura 3.2: De izquierda a derecha: Cherenkov, Frank, Tamm y Vavilov. [41]

3.1.1. Determinación del ángulo Cherenkov

El análisis teórico completo de Frank y Tamm puede ser consultado en J. V. Jelley [42]. Sin embargo puede realizarse un análisis cualitativo más sencillo para determinar el ángulo Cherenkov, que es el ángulo característico del cono de radiación generado.

Supongamos que se tiene un electrón en movimiento a través de un medio dieléctrico (Figura 3.3). Se van a considerar dos casos: en el primero (a) el electrón viaja a velocidades muy bajas comparadas con la de la luz y en el segundo (b) viaja a velocidades comparables con la de la luz en el medio. En ambos casos el electrón parte de **A** hacia **B** siendo el punto **P** un instante durante el recorrido de electrón y **P'** un instante después de **P**.

- **Caso (a)** Los círculos representan los átomos del medio. Estos van a estar en forma esférica no distorsionada hasta que en un instante, en el punto **P**, el campo eléctrico del electrón distorsiona los átomos y debido a que se trata de un material dieléctrico los átomos se polarizan (se comportan como dipolos) alrededor del punto **P**, manteniendo sus polos negativos hacia afuera y viceversa si se tratara de una partícula cargada positivamente. Un instante después, en **P'**, los átomos

alrededor de P vuelven a su forma normal emitiendo un pulso electromagnético breve. Debido a la baja velocidad, para cada punto P se tiene una simetría total (tanto azimutal como a lo largo del eje definido por la trayectoria del electrón de A a B) del campo de polarización alrededor del electrón obteniéndose un campo resultante nulo a distancias mayores y por lo tanto no hay radiación.

- **Caso (b)** Cuando el electrón viaja a velocidades comparables a la de la luz en el medio se mantiene la simetría azimutal pero no la simetría en el eje de la trayectoria del electrón. De la misma forma esto genera pequeños pulsos electromagnéticos a la largo de la trayectoria del electrón y en general estas radiaciones interfieren destructivamente obteniendo una intensidad de campo resultante igual a cero. Sin embargo cuando la velocidad del electrón supera a la de la luz es posible que para cada pulso producido sobre la trayectoria del electrón esté en fase con los demás por lo que el electrón produce un campo a largas distancias.

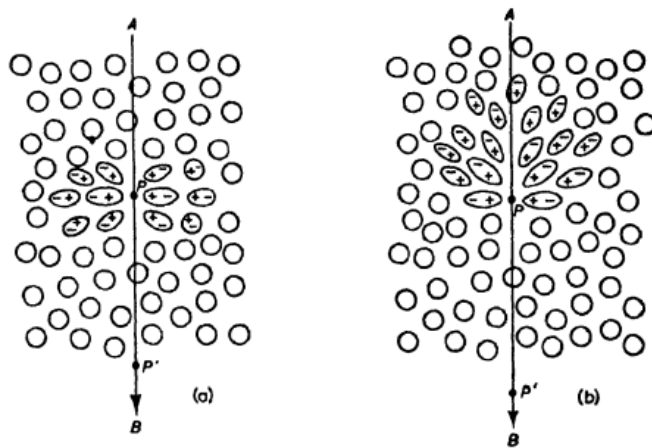


Figura 3.3: Polarización de un medio dieléctrico debido al paso de un electrón. (a) Velocidad baja. (b) Velocidad alta. [41].

Para conocer el ángulo del cono de radiación Cherenkov se puede hacer un análisis geométrico basado en la Figura 3.4.

Dicho ángulo depende de la velocidad de la partícula y del índice de refracción del medio. Si la velocidad de la partícula v , es mayor a la velocidad de la luz c , entonces la partícula viaja una distancia d de A hasta B:

$$d = tv,$$

y se define $\beta = \frac{v}{c}$

$$d = tc\beta. \quad (3.1)$$

Por otra parte, el pulso electromagnético generado viaja a la velocidad de la luz en el medio $\frac{c}{n}$ en donde n es el índice de refracción del medio, por lo que el pulso viaja una distancia d_p de A hasta C:

$$d_p = t\frac{c}{n}. \quad (3.2)$$

Estas distancias forman un triángulo rectángulo, por lo que se puede obtener el ángulo Cherenkov θ_c .

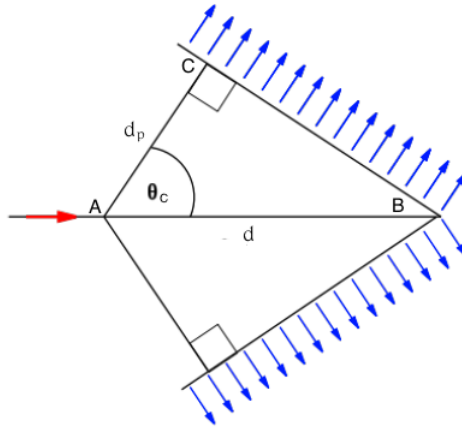


Figura 3.4: Diagrama de la geometría del cono de radiación Cherenkov. Imagen tomada de [43].

$$\cos(\theta_C) = \frac{d_p}{d}. \quad (3.3)$$

Y de la Ecuación 3.1 y la Ecuación 3.2 se obtiene:

$$\cos(\theta_C) = \frac{1}{n\beta}. \quad (3.4)$$

Luego se puede utilizar la siguiente relación para obtener la energía:

$$E = \frac{mc^2}{\sqrt{1 - \beta^2}}, \quad (3.5)$$

en donde m es la masa del electrón. Por lo que la relación entre el ángulo y la energía es:

$$\cos(\theta_C) = \frac{1}{n\sqrt{1 - \left(\frac{mc^2}{E}\right)^2}}. \quad (3.6)$$

Esta última ecuación nos permite obtener el ángulo Cherenkov para el caso más general en que se tenga una partícula cargada con energía E , y masa m a través de un medio con índice de refracción n .

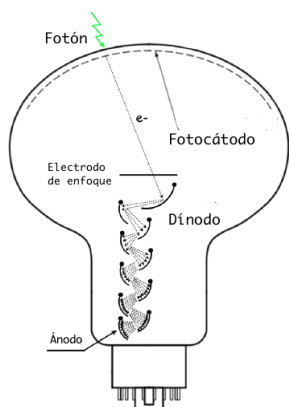
Para el caso de HAWC se tiene como medio dieléctrico el agua por lo que el índice de refracción será $n \approx 1.33$ y las partículas a considerar son las producidas en las cascadas atmosféricas las cuales son muy energéticas, es decir, $\beta \approx 1$. Por lo que sustituyendo en la Ecuación 3.4 y despejando el ángulo Cherenkov se obtiene:

$$\theta_C = 41.25^\circ. \quad (3.7)$$

3.2. Tubos fotomultiplicadores (PMT)

Los tubos fotomultiplicadores o PMTs (*Photo Multiplier Tubes*) son dispositivos detectores de luz capaces de detectar hasta un solo fotón. Cada WCD cuenta con 4 tubos fotomultiplicadores y la ubicación de estos es importante para llevar a cabo la reconstrucción de eventos (cascadas atmosféricas). En HAWC se usan dos tipos diferentes de PMT: el primer tipo es el modelo *Hamamatsu R5912* de 8" (mismos usados en el experimento *MILAGRO*) y se disponen tres en forma de triángulo equilátero a una distancia de 1.8 m del cuarto PMT central, que es del segundo tipo, modelo

Hamamatsu R7081-02 de 10" y de alta eficiencia cuántica QE (probabilidad de que un electrón pueda ser emitido por efecto fotoeléctrico: $\sim 25\%$ para el R5912 y $\sim 40\%$ para el R7081-02) [12, p. 28]. Estos PMTs tienen una mejor detección de señal en longitudes de onda cercanas a los 420 nm , sin embargo su rango de detección va desde 300 nm hasta 650 nm [44] por lo que son los adecuados para la detección de luz proveniente del efecto Cherenkov.



(a) Esquema de funcionamiento del PMT [12].



(b) PMT en el laboratorio antes de ser encapsulado

Figura 3.5: Tubo fotomultiplicador. Recolecta la luz Cherenkov dentro de los tanques de HAWC y devuelve una señal electrónica.

Los PMT funcionan gracias al efecto fotoeléctrico cuando la luz al incidir en un metal, en este caso el fotocátodo del PMT, libera electrones (conocidos como fotoelectrones PE). Estos PEs son acelerados y dirigidos mediante campos eléctricos hasta el primer dínodo en donde más electrones secundarios son emitidos que a su vez son dirigidos a otros dínodos subsiguientes. Los electrones resultantes de esta cascada son recolectados en el *ánodo* en donde ya pueden ser interpretados como una señal de una medición (ver Figura 3.5).

A la relación entre la señal promedio producida por un solo fotón incidente con la carga fundamental del electrón se le conoce como ganancia del PMT. Otra forma de ver este valor es como la amplificación de corriente

que es una medida de qué tanto se multiplicaron los PEs desde el primer dínodo hasta el ánodo. La ganancia depende del número de dínodos y del voltaje total de operación de cada PMT. Para todos los WCDs de HAWC, se calibra la ganancia de los PMTs de tal forma que se asegure una respuesta uniforme al momento de llevar a cabo la adquisición de datos.

Debido a que los PTM son sumergidos en agua, se encapsulan en tubos de PVC y se sellan con silicona para proteger los componentes electrónicos. Se realizan pruebas para verificar que no tengan fugas y así poder ser utilizados en el observatorio.

3.3. Sistema de adquisición de datos (DAQ)

La frontera entre el fenómeno físico y la interpretación de señales eléctricas sucede en el PMT. La señal del PMT se transfiere mediante un cable coaxial a una sala de cómputo ubicada en el sitio del observatorio. Luego la señal es enviada al TDC (*time to digital converter*) en donde se amplifica, se da forma y se pasa por dos discriminadores ($\approx 1/4$ y ≈ 4 PEs) [26],[45]. La cantidad de tiempo que dura la señal después de pasar por estos límites se conoce como *time-over-threshold (ToT)* y es usada para estimar la carga total detectada por un PMT.

En HAWC se utilizan dos sistemas de adquisición de datos (DAQ) para interpretar la señal proveniente del PMT. El sistema principal (*main DAQ*) sirve para registrar eventos individuales de cascadas atmosféricas lo suficientemente grandes como para cubrir una fracción significativa del arreglo de HAWC. Este sistema principal mide el tiempo de llegada y el *ToT* de los pulsos de los PMTs con lo que se reconstruyen el núcleo de la cascada, la dirección de llegada y la *distribución lateral*. Estas variables a su vez sirven para determinar la especie de partícula primaria y su energía [45]. Este sistema produce ≈ 500 MB/s de datos crudos (*raw*) que son reducidos a ≈ 20 MB/s debido a un sistema de procesamiento en línea [46].

Un sistema de conteo secundario (*scaler DAQ*) registra los pulsos (*hits*) en cada PMT. Este sistema es sensible a crecimientos o decrecimientos repentinos en la tasa de conteo de pulsos lo cual es útil para monitorear el estado del detector, estudios sobre la emisión de partículas del Sol, así como

para detectar eventos transitorios, como los GRBs (*Gamma Ray Burst*), a energías que van de $GeV - TeV$ [45].

El uso de ambos sistemas es complementario; mientras que el sistema principal registra el número de eventos aportando información para la reconstrucción de la energía y la dirección de llegada, el sistema secundario cubre el rango de bajas energías (no observadas por el sistema principal) y es capaz de detectar incrementos o decrementos en las tasas de conteo para determinar la significancia de las observaciones.

3.4. Software utilizado

El software que se utiliza en HAWC se divide en dos componentes: los *externals* y *AERIE*. A continuación se dará una breve descripción de estos componentes y de las paqueterías que se usaron principalmente para este trabajo.

3.4.1. Externals

Los paquetes externos incluyen librerías y programas que son necesarios para poder compilar el código base de HAWC. Los externals instalados son: **boost**, **fftw**, **gsl**, **root**, **xerces-c**, **cmake**, **cfitsio**, **healpix-cxx**, **photospline** y **xcdf**. Estos paquetes son instalados haciendo uso de APE (*Auger Package Management*), una herramienta de distribución de software creada para el observatorio Pierre Auger y adoptada por HAWC.

3.4.2. AERIE

Analysis and Event Reconstruction Integrated Environment es el software de HAWC que permite establecer una estructura dedicada a procesar y analizar los eventos registrados, por ejemplo, para generar simulaciones, uso de *aerie-apps* para agregar energías estimadas, hacer reconstrucciones offline, entre otros. Este software está estructurado como un conjunto interdependiente de proyectos en *C++* unido por un núcleo central o *framework*.

Los principales componentes de AERIE son:

1. **HAWCNest Framework.** Es un objeto central que registra e inicializa servicios, no edita datos.

2. **Data Structures.** Una representación de memoria interna de los datos que pueden ser editados por los servicios.
3. **Services.** Código de usuario que puede utilizarse para editar datos en un proceso o proveer cálculos independientes, como generación de números aleatorios.
4. **MainLoop.** Un servicio especial que define el flujo de control para el procesamiento de datos.
5. **Applications.** Un conjunto de programas utilizados para el análisis básico de datos de HAWC tales como creación de mapas o la estimación del espectro de energía.

Los usuarios avanzados escriben sus propios servicios y editan HAWC-Nest framework para tareas especializadas. Los usuarios intermedios pueden interesarse en las estructuras de datos así como en la entrada y salida (I/O) de datos. Los usuarios principiantes suelen estar más interesados en las herramientas que ofrecen las aplicaciones [47].

3.4.3. ROOT

ROOT es un *object-oriented framework* dedicado a resolver los desafíos que representa el análisis de datos de la física de altas energías. Utilizar un *framework* tienes múltiples beneficios como menos código que escribir, códigos de mayor confianza, códigos más consistentes y permite concentrarse en el área de interés a estudiar.

ROOT comenzó como un proyecto para análisis de datos del CERN pero es indispensable su uso para adentrarse en el experimento HAWC. El principal uso que le he dado en este trabajo ha sido el de extraer las variables de los archivos de datos para analizarlas y poder reconstruir histogramas de energía de las partículas.

Para la simulación

Para llevar a cabo la reconstrucción del espectro de energía de los rayos cósmicos y en general para la mayoría de investigaciones experimentales en la actualidad es necesario hacer pruebas simuladas que nos ayuden a

entender de una mejor manera lo que está sucediendo con el experimento. En HAWC se ha desarrollado el software necesario para simular tanto la respuesta del detector como las partículas que pueden ser detectadas por el mismo (rayos cósmicos y gamma).

3.4.4. CORSIKA

COsmic Ray SIMulation for KAScade tiene como objetivo generar una base de datos simulados. Es una herramienta diseñada para simular la interacción de rayos cósmicos y rayos gamma con la atmósfera terrestre (es decir las cascadas atmosféricas). Se obtiene como salida un archivo binario que contiene información sobre las interacciones debidas a rayos gamma y a nueve especies de hadrones (H , He , C , N , O , Ne , Mg , Si , Fe). Los archivos de salida son tomados después por Geant4 [48], [49].

3.4.5. GEANT4

Programa encargado de simular la respuesta del detector. Es usado en HAWC para simular la respuesta del detector a las cascadas atmosféricas tomando como entrada los archivos generados por CORSIKA y devolviendo como salida un archivo que contiene el tiempo de cada PE que interactúa dentro del fotocátodo del PMT. Este programa toma en cuenta todas las interacciones debidas a las partículas de alta energía así como a las interacciones debidas a las partículas secundarias. Guarda información sobre los fotones producidos por la radiación Cherenkov y simula la absorción, dispersión y reflexión de estas partículas con la materia y con el agua. También incluye la simulación completa de todos los PMT del arreglo [48], [50].

Estos dos programas son los principales para obtener los archivos simulados sin embargo también se utiliza un software llamado *DAQsim* que modela el sistema de adquisición de datos [51].

Se obtienen finalmente archivos con las variables simuladas que desde 2014 son reconstruidas de la misma forma en que son reconstruidos los archivos de datos reales por lo que el formato de salida es el mismo [52]. La Figura 3.6 muestra la línea de procesos que se realiza para la simulación de eventos.

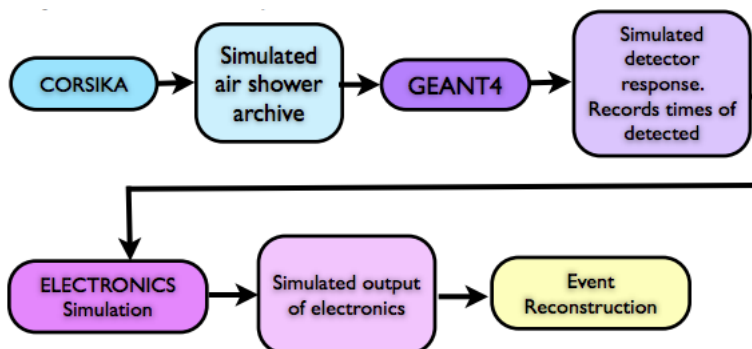


Figura 3.6: Diagrama del proceso de simulación de eventos [52].

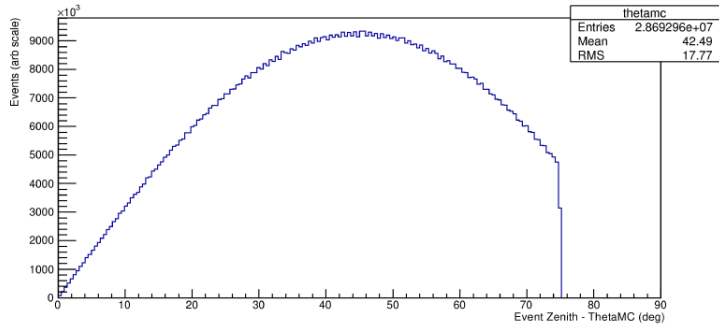
Generación de la simulación

Los eventos simulados son lanzados a partir de tres parámetros que los caracterizan (ver Figura 3.7): la energía del evento E , el ángulo cenital del evento θ y la distancia del evento desde el centro del arreglo r .

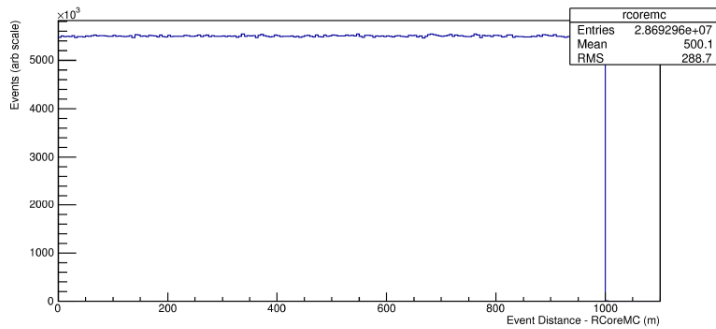
La forma de la distribución cenital es de la forma $\sin(\theta)\cos(\theta)$ y está pensada para modelar algo físico ya que la parte del $\sin(\theta)$ representa el crecimiento del espacio fase a medida que crece θ mientras que el $\cos(\theta)$ representa el hecho de que se esperan menos eventos para ángulos cenitales grandes ya que el campo de visión de HAWC disminuye cuando el ángulo cenital crece (recordar que en HAWC el ángulo cenital es 0° en el cenit y crece hacia el horizonte).

La distribución en r , la distancia del evento al centro del arreglo, no representa algo físico pues se esperarían más eventos conforme r crece y esto debido a que el espacio fase va creciendo junto con r . Sin embargo la distribución mostrada en la Figura 3.7b se hace para aumentar artificialmente el número de eventos cercanos al detector.

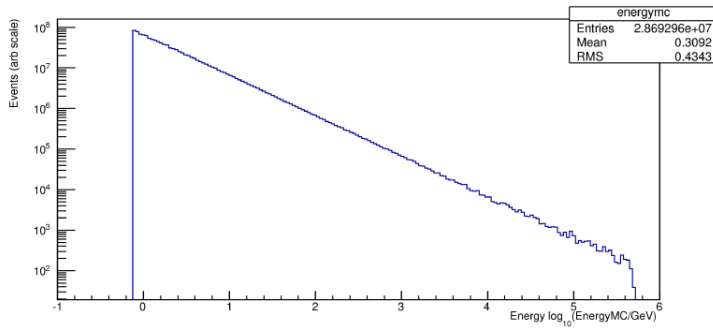
Para simular la energía se utiliza una ley de potencias de la forma E^{-2} .



(a) Distribución para el ángulo cenital.



(b) Distribución para la distancia del centro del arreglo.



(c) Distribución para la energía.

Figura 3.7: Distribuciones para θ , r y E utilizadas para la simulación [53].

La mayoría de los flujos de fuentes astrofísicas presentan distribuciones de energías más blandas que E^{-2} (por ejemplo, el índice espectral de la Nebulosa del Cangrejo medida con el experimento MILAGRO es de -2.29 [54]) lo que implica pocos eventos a grandes energías. Para la distribución simulada se esperan más eventos lanzados artificialmente a grandes energías, esto con la finalidad de poder estudiar la respuesta del detector a estas energías [53].

3.5. Archivos de simulación y de datos reales

Los datos obtenidos por el observatorio HAWC, así como las simulaciones son almacenados en dos clusters: uno en la UNAM (Universidad Nacional Autónoma de México) y otro en UMD (University of Maryland). A continuación se dará una descripción general de los tipos de archivos, las ubicaciones de los que fueron utilizados y los cortes aplicados.

3.5.1. Tipos de datos

El *Data Format* es un documento al que se tiene acceso solo en las páginas internas del sitio web del observatorio HAWC en el cual se explican las variables contenidas en los diferentes tipos de archivos que genera la adquisición y el procesamiento de datos. Los archivos generados pueden ser de tres tipos y son de formato XCDF (*Explicitly Compacted Data Format*) que después han de ser convertidos a formato *ROOT* para poder utilizarlos:

- 1 **raw**. Estos archivos contienen las señales de todos los PMTs del arreglo sin ningún tipo de condición de *trigger* o calibración.
- 2 **trig**. Estos archivos contienen datos de eventos que han disparado (*trigger*) ciertas condiciones para ser considerados cascadas atmosféricas sin embargo la reconstrucción aún no ha sido aplicada.
- 3 **reco**. Estos archivos se obtienen cuando se hace la reconstrucción de los archivos trig dando como resultado los valores de variables importantes para hacer análisis de los rayos cósmicos (por ejemplo el ángulo de llegada de las cascadas así como su núcleo, etc).

Nota: anteriormente se tenía que trabajar con los archivos trig y hacer una reconstrucción agregando la aplicación *aerie-apps-add-energies* para obtener las variables de energía de los dos diferentes estimadores utilizados en HAWC (*Neural Network* y *Ground Parameter*), sin embargo en las reconstrucciones recientes (en este caso en particular eventos desde junio del 2018) ya se han incluido estas variables de energía en los archivos reco.

Para el presente trabajo se hizo uso de de archivos de datos reales reconstruidos que contienen variables de tipo reco y archivos de simulación *Monte Carlo* que contienen variables de tipo *mc* y *reco*.

3.5.2. Archivos de simulación

Los archivos de simulación fueron extraídos de la siguiente ruta dentro del *cluster* de UMD:

```
/data/disk01/home/imc/HAWCROOT/sim/reco/dev/nominal_allspecies/
reco_files/sweets
```

en donde se encuentran archivos simulados para diferentes declinaciones desde -30° hasta 70° con un total de 21 archivos (151.5 GB). Cada uno de estos archivos contiene la información de 32,428,939 eventos simulados con 98 variables por evento de las cuales 79 son *rec* y 15 *mc*.

3.5.3. Archivos de datos reales

Los archivos de datos reales se agrupan por año y mes en los *clusters* de HAWC. Dentro de las carpetas mensuales se encuentra varias carpetas denominadas *run* y corresponden a un día de datos. Dentro de cada *run* hay 692 *subruns* (o archivo de datos) lo que implica dos minutos de datos por cada *subrun*. Es importante mencionar que también se pueden encontrar *runs* con muy pocos *subruns* debido a que los datos podrían tener problemas. Para evitarlo solo he tomado aquellos *runs* con 692 *subruns* (es decir un día completo ininterrumpido de toma de datos).

Los *subruns* utilizados fueron tomados de la siguiente ruta en el cluster de UMD:

```
/data/archive/hawcroot/data/hawc/reconstructed/hawcprod/v2.06.
01/config-44003/reco_xcdf/2018/06
```


Cada archivo de datos tiene un tamaño promedio de 435 MB lo que da como resultado ≈ 301 GB por *run*, contiene la información de $\sim 3.1 \times 10^6$ eventos registrados por HAWC con 82 variables tipo *rec* por cada evento.

En la siguiente lista se muestran las variables que se utilizaron para el análisis del presente trabajo. Las variables de tipo *mc* fueron exclusivamente para las simulaciones mientras que las variables de tipo *rec* fueron utilizadas tanto en las simulaciones como con los datos reales.

- **mc.logEnergy**: Energía verdadera de las partículas primarias simuladas.
- **mc.corsikaParticleId**: Tipo de la partícula primaria.
- **rec.logNNEnergyV2**: Energía de las partículas primarias reconstruida a partir de la respuesta del detector usando el estimador *Neural Network* en su segunda versión.
- **rec.logGPV2**: Energía de las partículas primarias reconstruida a partir de la respuesta del detector usando el estimador *Ground Parameter* en su segunda versión.
- **rec.nHit**: Número de *hits* (señales detectadas por algún PMT) en una ventana de tiempo de $1.5 \mu s$ para ser considerado como un evento.
- **rec.angleFitStatus**: El estado de la reconstrucción angular; 0 =exitoso, 1 =fallido, 2 =desconocido y 3 =no hay convergencia.
- **rec.coreFitStatus**: El estado de la reconstrucción del núcleo: 0 =exitoso.
- **rec.nChAvail**: Número total de PMT's disponibles para la reconstrucción.
- **rec.coreFiduScale**: Distancia del núcleo de un evento al centro del arreglo.
- **rec.zenithAngle**: Ángulo cenital reconstruido.

3.5.4. Cortes generales utilizados

A las variables anteriores del tipo *rec* (excepto las de energía) se les aplican condiciones que han de cumplir por evento para que éste pueda ser considerado dentro de los histogramas de energía. A estas condiciones se les conoce como cortes y sirven para asegurar que los histogramas sean llenados con eventos reales y no con fluctuaciones o ruido.

Utilicé los cortes (*quality cuts QC*) sugeridos en [55] agregando el corte de *rec.nHit* mencionado en [23, Pág. 4]. La lista de cortes es la siguiente:

- $\text{rec.nHit} > 75$. Se pide que cada evento haya activado más de 75 PMT's. Los eventos con pocos PMTs activados pueden ser debido a ruido.
- $\text{rec.angleFitStatus} = 0$. La reconstrucción angular debe ser exitosa.
- $\text{rec.coreFitStatus} = 0$. La reconstrucción del núcleo de la cascada debe ser exitosa.
- $\text{rec.nChAvail} \geq 700$. Al registrarse el evento debe haber al menos 700 PMT's disponibles, es decir el 58 % del arreglo de HAWC.
- $\text{rec.coreFiduScale} \leq 100$. Quita las cascadas cuyo núcleo está más allá de 100 metros, es decir, solo toma en cuenta las que caen dentro del arreglo.
- $\text{rec.zenithAngle} < 0.7854$. Solo toma en cuenta las cascadas cuyo ángulo cenital es menor a 45° (rango del campo de visión cenital de HAWC). También se realizan pruebas para ángulos cenitales menores a 35° y 16° con la finalidad de hacer un estudio sistemático.

En [56, p. 117] se muestra una tabla con todas la partículas que pueden ser simuladas por CORSIKA cada una de ellas con una número de identificación correspondiendo al fotón el número 1. Para las simulaciones se pide que la variable $\text{mc.corsikaParticleId} > 1$ para poder descartar los rayos gamma simulados y solo quedarnos con los rayos cósmicos. Para los datos reales existen algoritmos de separación gamma-hadrón indispensables

para los estudios de rayos gamma, sin embargo para rayos cósmicos se puede prescindir de su uso ya que el flujo de rayos cósmicos se piensa que es aproximadamente el 1% de la radiación cósmica [57].

Después se procede a llenar histogramas de energía solo con los eventos que han pasado todos los cortes establecidos.

3.5.5. Corte extra para *Ground Parameter*

Al análisis del estimador *Ground Parameter* se le agregará un corte extra que toma en cuenta el valor del parámetro de edad s de la función de distribución lateral según la descripción dada en la Sección 2.3.1.

- $0.8 \leq \text{rec.LDFAge} \leq 1.6$

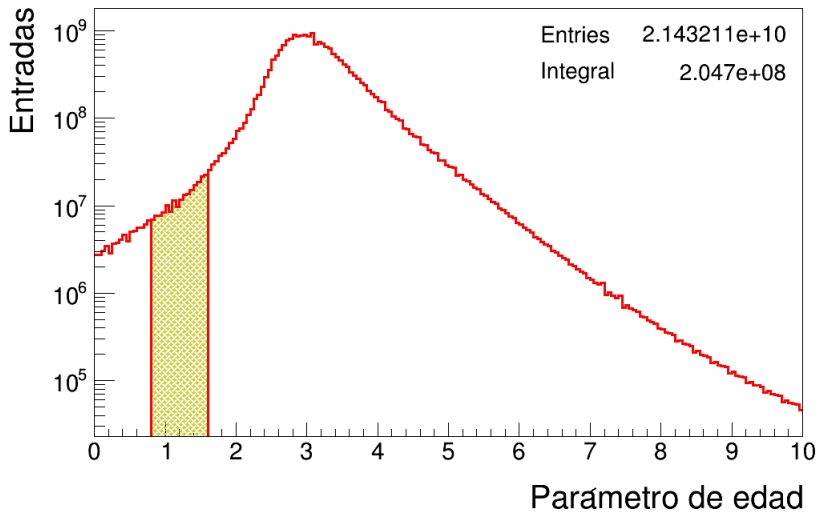


Figura 3.8: Histograma del parámetro de edad.

La Figura 3.8 muestra un histograma del parámetro de edad, resaltando de manera sombreada aquellos eventos que cumplen el corte extra aplicado. Los datos totales utilizados representan 2.14×10^{10} entradas que se reducen

a 2.05×10^8 entradas que pasan el corte sobre el parámetro de edad. Estas entradas se reducen aún más al aplicar todos los demás cortes.

A pesar de recortar de forma considerable el número de eventos disponibles, con este corte se observaron mejoras notables en los resultados de este estimador.

Capítulo 4

Reconstrucción del espectro de rayos cósmicos

4.1. *Unfolding*

En la física experimental intervienen muchos factores para llevar a cabo la detección y clasificación de los eventos que se quieren observar. Debido a esto los eventos observados son diferentes al resultado esperado en el caso de contar con un detector ideal y esto es principalmente causado por alguno de los siguientes factores [58].

- **Efectos del detector:** variables importantes para definir un evento, tales como energía, ángulo de llegada, tiempo de llegada entre otros, son medidos con precisión finita y eficiencia limitada por lo que algunos eventos podrían llegar a ser reconstruidos en un *bin* incorrecto o perdidos debido a los cortes aplicados.
- **Background:** existen señales similares a aquellas debidas a eventos pero producidas por otros procesos.

Al proceso de extraer información sobre el contenido *verdadero* de los *bines* medidos dada una medición observada se le conoce como *unfolding*. Esta transformación para conocer la verdadera distribución de una variable es un tipo de problema lineal inverso.

4.1.1. Problema inverso

Para entender como funciona el *Unfolding* se deben tomar en cuenta los procesos directo e inverso. A continuación se realiza una descripción de como se relacionan ambos procesos en el *Unfolding*.

Supongamos que se quiere medir una variable física E la cual tiene asociada una distribución $f(E)$ y que hasta el momento son desconocidas por quien realiza el experimento. Debido a los factores experimentales se obtiene una variable medida E' con una distribución $g(E')$ relacionada con la distribución verdadera $f(E)$ mediante migraciones, distorsiones y transformaciones.

Usando métodos de Monte Carlo (MC) se puede realizar el proceso directo en el que se asume un modelo de la *distribución verdadera* $f(E)_{MC}$ y mediante la simulación de la respuesta del detector se obtiene su correspondiente distribución medida $g(E')_{MC}$. El proceso inverso, en el cual se tiene una distribución medida y se quiere conocer la distribución verdadera, es difícil de resolver. Si no se realiza con cuidado, pequeñas variaciones en la distribución medida pueden causar grandes cambios en la distribución verdadera.

- **Proceso directo** (a partir de MC) $f(E)_{MC} \implies g(E')_{MC}$
- **Proceso inverso** (unfolding) $g(E') \implies f(E)$

Los procesos directo e inverso son descritos por la siguiente ecuación integral, llamada ecuación de primer tipo de Fredholm.

$$\int_{\Omega} K(E', E) f(E) dE = g(E'). \quad (4.1)$$

En donde $K(E', E)$ es una función Kernel que describe los procesos físicos de medición [59].

4.2. Discretización

Al medir experimentalmente una variable, el rango de valores que ésta puede tener es dividido en intervalos usualmente consecutivos. De esta

forma, cuando se hace una medición, el valor obtenido de la variable es asignado como una entrada que cae en alguno de los intervalos. Después de realizar varias mediciones se obtiene un histograma de la variable medida. A los intervalos definidos para construir un histograma se les conoce como *bines*.

La Ecuación 4.1 describe el caso para distribuciones continuas, sin embargo en lo experimental se tienen cuentas de entradas distribuidas en *bines* por lo que se debe realizar una discretización para poder solucionar el problema inverso de forma numérica. Esta discretización está dada por la siguiente ecuación lineal.

$$Ax = y. \quad (4.2)$$

La Ecuación 4.2 es análoga a la Ecuación 4.1 y se relacionan de la siguiente manera:

distribución verdadera $f(E)$	$\Rightarrow x$	n – vector desconocido
distribución medida $g(E')$	$\Rightarrow y$	m – vector de datos medidos
kernel $K(E', E)$	$\Rightarrow A$	matriz rectangular de respuesta dimensión $m \times n$

(4.3)

4.2.1. Álgebra lineal

Para poder entender un poco mejor el proceso inverso que se lleva se mostrará un breve repaso de los conceptos relevantes de álgebra lineal involucrados en este tema.

El punto crucial de este análisis está centrado en el comportamiento de la matriz de respuesta. Debido a la forma en que se construye la misma, no se puede asegurar que sea una matriz cuadrada (mismos número de *bines* en x y en y) por lo que el problema inverso no es inmediato.

Este problema se puede tratar desde el álgebra lineal para resolver sistemas de ecuaciones. Supongamos que la matriz del sistema de ecuaciones es de $m \times n$ en donde m indica el número de renglones (ecuaciones) y n el

número de columnas (incógnitas), $y \in \mathbb{R}^m$ el vector de resultados y $x \in \mathbb{R}^n$ el vector solución. La idea es tratar de resolver la Ecuación 4.2, y se tienen tres diferentes casos:

- **$m=n$** Si la matriz A es cuadrada e invertible entonces el problema se resuelve de la manera usual, encontrando A^{-1} y resolviendo la ecuación $x = A^{-1}y$.
- **$m < n$** Si la matriz es más ancha que alta implica que el sistema de ecuaciones tiene más incógnitas que ecuaciones por lo que la solución dependerá de una o más incógnitas (similar a una representación paramétrica) obteniéndose así una infinidad de soluciones. Este caso no se va a considerar para tratar problemas experimentales.
- **$m > n$** Si la matriz es más alta que ancha se tiene un sistema de ecuaciones *sobre-determinado* el cual en la mayoría de los casos resulta ser incompatible (sin solución).

El caso que se trata usualmente es el tercero ($m > n$) debido a que la razón de tener más ecuaciones que incógnitas surge de manera natural en problemas experimentales en los que se toman varias mediciones con el fin de minimizar errores. El resultado es un sistema sobre-determinado e inconsistente. Como ejemplo, Gauss resolvió un problema de once ecuaciones con seis incógnitas para determinar la órbita del asteroide Pallas [60].

Para poder dar una solución aproximada se utiliza el *método de los mínimos cuadrados* desarrollado por Gauss y Legendre para resolver problemas de trayectorias en astronomía. En el *Apéndice A* se muestra un ejemplo introductorio sobre mínimos cuadrados así como su caso general además de un repaso sobre conceptos del álgebra lineal y su uso en la resolución del problema inverso de un sistema sobre determinado.

4.3. *TUnfold*

TUnfold [61] es un algoritmo diseñado para obtener las distribuciones medidas por un detector a un nivel verdadero, es decir, corregir los efectos producidos por los factores mencionados en la Sección 4.1. El algoritmo

está basado en un ajuste por mínimos cuadrados agregando un término de regularización [58, p. 6] (ver sección *Regularization methods* en V. Blobel [59]), que busca minimizar las fluctuaciones producidas de un *bin* a otro en el resultado, además de un término opcional que constriñe el área (número total de cuentas) forzando la normalización del resultado para encajar con el número total de eventos. En [61, p. 2] se encuentra una descripción detallada del tratamiento matemático que describe los tres términos descritos antes.

En términos prácticos, *TUnfold* se usa para descomponer una medición y en varias fuentes x dadas las incertidumbres de la medición y una matriz de migraciones A (matriz de respuesta). Tomando en cuenta lo anterior, para correr el código de *unfolding* solo son necesarios dos histogramas [62]. Para más especificaciones técnicas se recomienda ver el manual de uso de *TUnfold* [63]. Los dos histogramas que se utilizan son:

Matriz de respuesta: es el histograma bidimensional construido con simulaciones Monte Carlo para poder predecir la distribución verdadera de una variable a partir de la distribución medida. Para llenar la matriz de respuesta, en el eje x se colocan las entradas correspondientes a la distribución verdadera provenientes de la simulación de eventos y en el eje y se toman las entradas correspondientes a la distribución reconstruida provenientes de la simulación de eventos y de la simulación de la respuesta del detector.

Vector de mediciones: es el histograma uno-dimensional construido a partir de los datos reales, es decir, la distribución medida reconstruida (en este caso, a partir de algún estimador de energía).

4.3.1. Consideraciones

A continuación se muestra una serie de consideraciones que se han tomado en cuenta en la implementación del código de *unfolding* en este trabajo.

En [64] se hace la siguiente advertencia: el algoritmo está basado en una inversión de matriz *estándar*, con las limitaciones conocidas en precisión numérica y costo de cómputo para matrices con grandes dimensiones. Por esta razón el algoritmo no debe ser usado para matrices demasiado grandes; no más de 200 *bines* para el eje x y no más de 1000 *bines* para el eje y .

Al momento de ejecutar el código pueden darse algunos problemas de los cuales los más comunes se muestran en *Common problems when using TUnfold* [65] junto con las recomendaciones para evitarlos, además de una lista con los mensajes de error que puede mostrar el código. A continuación se explican dos recomendaciones que se han usado para este trabajo:

- **Número de *bines*.** El arreglo de *bines* para la variable medida debe ser mas fino (mayor número de *bines*) en comparación con el de la variable verdadera. Se sugiere que se use al menos el doble de *bines* para la variable medida.

Ejemplo: Si se tiene una medición realizada en 20 *bines*, la matriz de respuesta estará conformada por 20×10 *bines* para el eje *y* y para el eje *x* respectivamente. Luego el código devolverá un histograma *unfolded* en 10 *bines*.

- **Uso de los *bines* para *underflow*/*overflow*.** El *underflow* y *overflow* son las cuentas que han quedado fuera del rango en que se definen los histogramas utilizados. El código de *unfolding* asigna dos *bines* extras a los definidos para el *underflow* y *overflow* dándole un significado especial para la variable reconstruida. Se sugiere que, a menos que no se tenga totalmente claro el significado de estos *bines*, se asegure que los histogramas utilizados no tengan entradas en *underflow* y *overflow*. Dado que este fue el primer acercamiento al uso del código se llevó a cabo esta acción durante todo el análisis.

4.3.2. Uso de *TUnfold*

En esta sección se presenta el código de *unfolding* utilizado así como un ejemplo introductorio de su uso. En [66] se encuentran algunos *test* realizados para *TUnfold* los cuales también están disponibles e incluidos, junto con los archivos necesarios para correr el *unfolding*, en los *externals* del software de HAWC (ver Sección 3.4.1). Este trabajo está basado en el archivo llamado *testUnfold3.C* el cual es una prueba para la clase *TUnfoldDensity* que forma parte de la paquetería de *TUnfold* y muestra algunas mejoras respecto a la clase principal también llamada *TUnfold* (ver [63, p. 3]). El archivo *testUnfold3.C* es por sí mismo un ejemplo del uso de *TUnfold* el

cual se resume una serie de pasos para poder realizar un *unfolding* (ver código [67]). Los pasos llevados a cabo en este trabajo son:

- 1.- Definir los *bines* y declarar los histogramas.
- 2.- Generar la distribución de datos.
- 3.- Generar las distribuciones simuladas.
- 4.- Realizar el *unfolding*.
- 5.- Recuperar los resultados del *unfolding*.
- 6.- Dibujar los histogramas.

4.3.3. Ilustración del procedimiento

El siguiente ejemplo tiene como objetivo tener una idea intuitiva del funcionamiento práctico del código de *unfolding*, por lo que los *bines* utilizados no son los mismos que en el análisis además de manejar una variable general (solo llamada variable) en unidades arbitrarias que en el caso del análisis corresponde a la energía primaria de los rayos cósmicos.

- **Definir los *bines* y declarar los histogramas.** Para ver de forma más clara el efecto de la matriz de respuesta sobre el histograma de datos reales se ha decidido usar el mismo número de *bines* tanto en el eje x como en el eje y que en este caso, y sin pérdida de generalidad, son 30 *bines*.

Los histogramas de datos reales y la matriz de respuesta se han declarado para analizar la variable en el rango de 10 a 30 unidades arbitrarias (u.a.).

- **Generar la distribución de datos.** Se ha construido de manera artificial una distribución gaussiana, con valor medio de 20 y sigma de 5, llenada de forma aleatoria con 10^7 entradas para representar la distribución medida (datos reales).

- **Generar las distribuciones simuladas (matriz de respuesta).**
En este ejemplo se comparan dos matrices de respuesta para ver los efectos de la misma sobre el histograma de datos.

La primer prueba se hará para el caso más trivial, es decir, la matriz identidad. Esta matriz es creada de manera artificial siguiendo una distribución uniforme llenada aleatoriamente con 10^4 entradas (ver Figura 4.1).

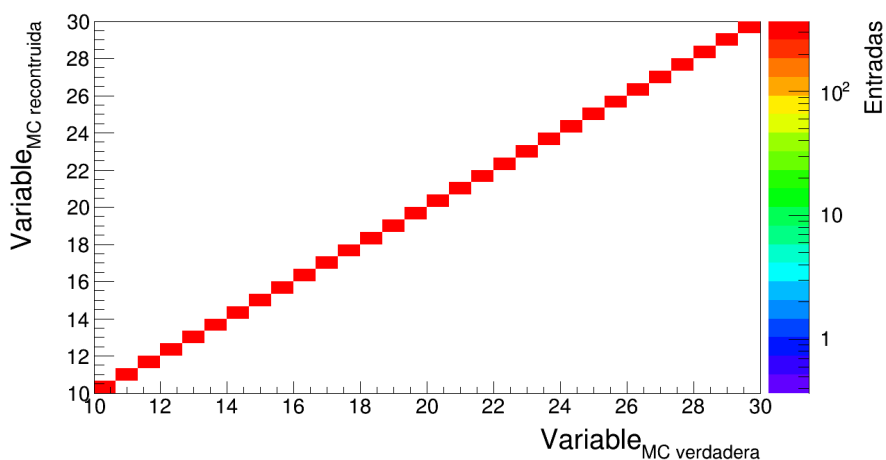


Figura 4.1: Matriz identidad.

Para el segundo caso, la matriz identidad es alterada agregando en su llenado una distribución gaussiana, con valor medio de 30 y sigma de 5 lo cual hará que la matriz identidad registre una mayor número de entradas para *bines* más altos. Después, para no tener una matriz identidad, la distribución gaussiana agregada se multiplica por otra distribución gaussiana (valor medio de 1 y sigma de 0.5 para no alterar demasiado la matriz) que la varía aleatoriamente 5.5% dando como resultado una matriz con valores fuera de la identidad pero no separados de esta (ver Figura 4.2). El proceso anterior agrega 5×10^5 entradas a la matriz de respuesta.

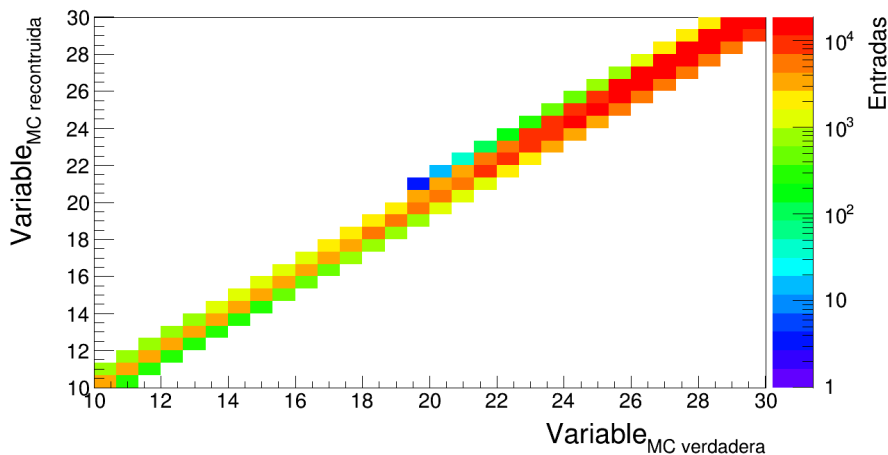


Figura 4.2: Matriz modificada para presentar más entradas en *bines* más altos y valores fuera de la identidad.

- **Distribuciones medida y verdadera (histograma unfolded)**

Las Figuras 4.3 y 4.4 muestran, respectivamente, los histogramas *unfolded* que resultaron de aplicar la matriz identidad y la modificada sobre el mismo histograma de entrada.

Como se esperaba, la matriz identidad no tiene influencia alguna sobre la distribución medida. En ningún caso realista se espera trabajar con una matriz identidad por lo que esta prueba solo fue para verificar que el código del *unfolding* estuviera ejecutándose de manera correcta.

El segundo caso representa, de manera muy sencilla, un caso más realista pues se espera tener entradas en diversos *bines* con una tendencia centrada en la identidad. Esto es debido a que este histograma dos dimensional relaciona una distribución verdadera con una reconstruida y en general, si la reconstrucción es buena, se espera que no difieran mucho los resultados. En los resultados del segundo caso se observa que de 10 a 20 la distribución verdadera y la medida son muy similares mientras que de 20 a 30 se obtiene una distribución que se aleja de la medida. Esto es debido a la forma en que se construyó la

matriz modificada, dándole más peso a los *bines* más altos.

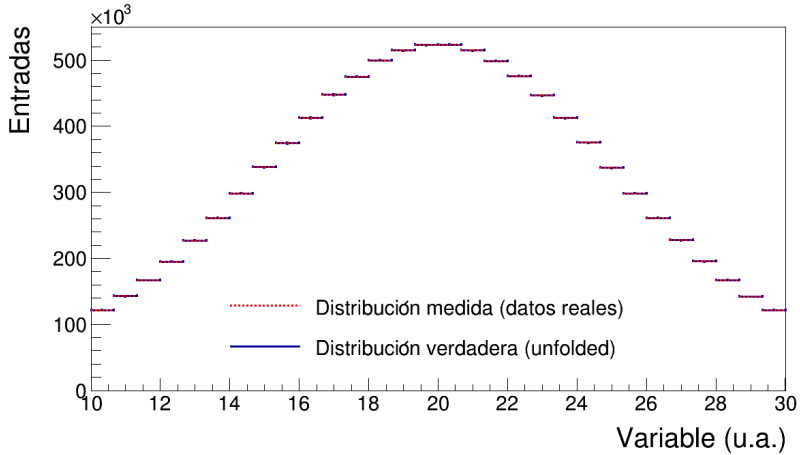


Figura 4.3: Histograma de la distribución medida y la distribución verdadera obtenida al hacer *unfolding* con la matriz identidad.

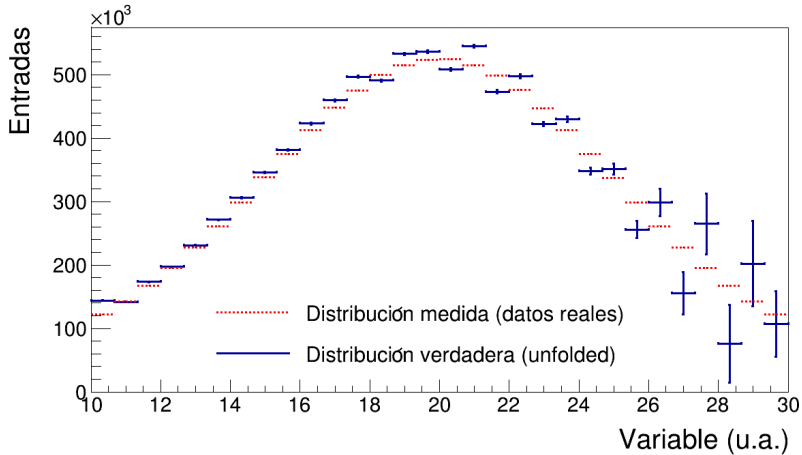


Figura 4.4: Histograma de la distribución medida y la distribución verdadera obtenida al hacer *unfolding* con la matriz modificada.

Con estas dos pruebas sencillas se pudo poner a prueba el código de *unfolding* preparándolo para su uso con los histogramas que se utilizarán en el análisis. También sirvió para ver de manera intuitiva la influencia que tienen las matrices, que representan los efectos del detector, sobre los histogramas.

4.4. Esquema de bins

Para este trabajo se lleva a cabo el análisis en tres casos angulares para el corte en el ángulo cenital (ver Tabla 4.1), el cual puede ir desde cero en el cenit del sitio del observatorio (cascadas verticales) y hasta 90° para eventos horizontales; θ_2 con corte a 45° para tener la máxima apertura angular del observatorio, θ_0 con corte a 16.8° para tomar eventos aproximadamente verticales y un paso intermedio θ_1 con corte a 35.2° .

Caso	Corte
θ_0	$\theta \leq 16.8^\circ$
θ_1	$\theta \leq 35.2^\circ$
θ_2	$\theta \leq 45.0^\circ$

Tabla 4.1: Casos para el corte del ángulo cenital.

El rango de energía va de 10^3 a 10^6 GeV y debe ser distribuido en *bines* de energía. Encontrar estos *bines* no es un trabajo trivial ya que durante las pruebas del código de *unfolding* se encontraron algunos errores al trabajar con cierta selección de *bines* (ver *Apéndice B*), los cuales eran corregidos al elegir otra selección. En A. U. Abeysekara et al [25] se utiliza un arreglo con ancho de *bin* de 0.25 en $\log E$, mientras que en R. Alfaro et al [7] se utiliza un ancho de *bin* de 0.1 en $\log E$ sin embargo no es posible simplemente tomar alguna de estas selecciones de *bines* e implementarla para este trabajo debido a los errores mencionados.

Las primeras pruebas exitosas para definir el arreglo de *bines* fue tomándolo de la siguiente forma:

$$\text{esquema1} = \{1, 2, 5, 10, 20, 50, 100, 200, 500, 1000\} TeV$$

Después, resultó más cómodo trabajar el arreglo de *bines* en su forma logarítmica definiéndolos por potencias de 10, que es aproximadamente:

$$\text{esquema1} = \{10^3, 10^{3.3}, 10^{3.7}, 10^4, 10^{4.3}, 10^{4.7}, 10^5, 10^{5.3}, 10^{5.7}, 10^6\} \log E(\text{GeV})$$

Luego se hicieron diferentes pruebas variando el arreglo anterior para mejorar el esquema de *bines*, en el sentido de disminuir los errores asociados a las entradas de cada *bin* del histograma *unfolded*, con lo que se obtuvo el siguiente arreglo:

$$\text{esquema2} = \{10^3, 10^{3.3}, 10^{3.65}, 10^4, 10^{4.3}, 10^{4.65}, 10^5, 10^{5.3}, 10^{5.65}, 10^6\} \log E(\text{GeV})$$

En la Tabla 4.2 se muestra el esquema de *bines* utilizado (esquema2) con sus valores aproximados en *TeV*.

Bin	Energía mínima (TeV)	Energía máxima (TeV)	Centro de <i>bin</i> (TeV)
1	1	1.99	1.50
2	1.99	4.47	3.23
3	4.47	10.0	7.23
4	10.0	19.9	15.0
5	19.9	44.7	32.3
6	44.7	100	72.3
7	100	199	150
8	199	447	323
9	447	1000	723

Tabla 4.2: Bines de energía.

Es importante recalcar que, tomando en cuenta las consideraciones mencionadas en la Sección 4.3.1, para obtener el doble de *bines* en el eje *y* en la matriz de respuesta, simplemente se dividió entre dos cada intervalo de energía cubierto por los *bines* del esquema2.

4.5. Flujo diferencial de los rayos cósmicos

Las mediciones del espectro de energía realizadas por diferentes experimentos no pueden ser comparadas directamente entre sí debido a que están

sujetas a diversos factores que deben ser tomados en cuenta dentro de una sola cantidad que sí pueda ser comparada. Para presentar los resultados de la medición de la energía se suele usar el flujo diferencial de rayos cósmicos:

$$\mathcal{F}(E) = \frac{N(E)}{T \Omega A_{\text{eff}} \Delta E}, \quad (4.4)$$

en donde $N(E)$ es la distribución de energía de los rayos cósmicos, T es el tiempo de observación, Ω es el ángulo sólido, A_{eff} es el área efectiva del detector y ΔE es el ancho de *bin* de energía centrado en E [7]. A continuación se describe cada uno de los elementos necesarios para la normalización de la distribución de energía para poder obtener el flujo diferencial de rayos cósmicos. Se muestran los valores utilizados para el análisis en este trabajo.

- Tiempo de observación (T). Estimado a partir del número de archivos de datos reales utilizados (ver Sección 3.5.3). Por lo que para calcular el tiempo de observación simplemente se multiplican 120 segundos de cada *subrun* por 692 (número de *subruns* en cada *run*) por 10 (número *runs* o días tomados).

$$T = 120 \text{ s} \times 692 \times 10 = 830,400 \text{ s} \quad (4.5)$$

- Ángulo sólido (Ω). Se obtiene a partir de la siguiente ecuación.

$$\Omega = 2\pi[1 - \cos \theta] \quad (4.6)$$

En donde θ es el ángulo cenital. HAWC tienen un amplio campo de visión instantáneo de aproximadamente el 16 % del cielo ($4\pi \text{ sr}$) por lo que su campo de visión sería de $\approx 2 \text{ sr}$ [45]. Para calcular los ángulos sólidos utilizados durante el análisis solo se introduce en la Ecuación 4.6 los valores de los tres cortes cenitales definidos en la Sección 4.4.

Caso	Corte	Ω
θ_0	$\theta \leq 16.8^\circ$	$2.69 \times 10^{-1} \text{ sr}$
θ_1	$\theta \leq 35.2^\circ$	1.15 sr
θ_2	$\theta \leq 45.0^\circ$	1.83 sr

Tabla 4.3: Ángulo sólido correspondiente a cada caso de corte cenital.

- Área efectiva (A_{eff}). El cálculo del área efectiva correspondiente al análisis realizado no es trivial, por lo que para este trabajo se utilizaron las áreas efectivas calculadas en 2011 en K. Sparks [68] y 2018 en J.C. Arteaga-Velázquez y J. D. Álvarez [69] (ver Figura 4.5). Se ha tomado el área efectiva calculada en 2018 para realizar el cálculo del flujo en todos los casos mientras que las de 2011 se han usado de apoyo en cada caso angular con el fin de obtener un error sistemático a partir de la falta del cálculo del área efectiva específica. Se eligió el resultado de 2018 como principal (a pesar de tener un corte del ángulo cenital a 16.71°) debido a que al ser más reciente presenta cortes que en 2011 no se incluían (como el corte *rec.coreFiduScale* que solo toma los eventos contenidos en el arreglo del observatorio). La forma en que se implementa el área efectiva estará descrita a detalle en la sección del análisis.

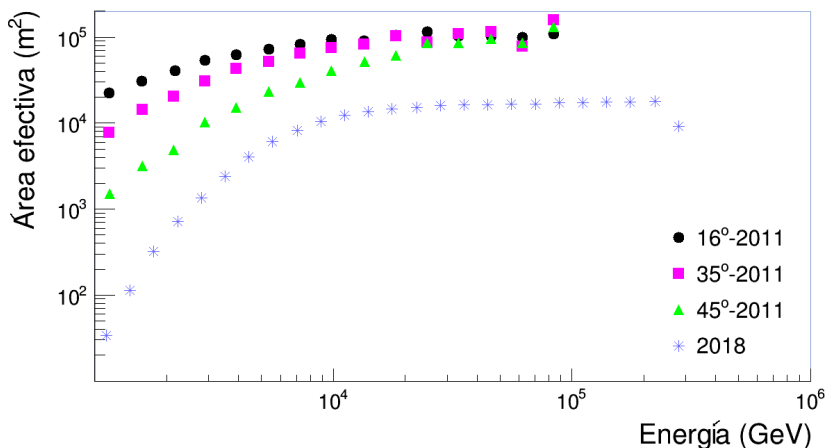


Figura 4.5: Área efectiva calculada por la colaboración HAWC. De estas gráficas se obtendrán los valores de área efectiva utilizados para este análisis. Tomadas de [68] y [69].

- Ancho de *bin* (ΔE). Estos valores están definidos por el esquema de *bines* de energía y se obtienen para cada *bin* al restar el valor de energía mínimo del valor máximo presentados en la Tabla 4.2.

4.6. Distribución de *Bias*

Para conocer el sesgo que existe entre la energía verdadera y la energía reconstruida por algún estimador se realiza una distribución llamada *bias*, definida por la diferencia entre los logaritmos de la energía reconstruida y la energía verdadera.

$$bias = \log E_{\text{reconstruida}} - \log E_{\text{verdadera}} \quad (4.7)$$

Estudiar esta distribución, para cada *bin*, sirve para evaluar si los cortes utilizados y el esquema de *bines* son adecuados.

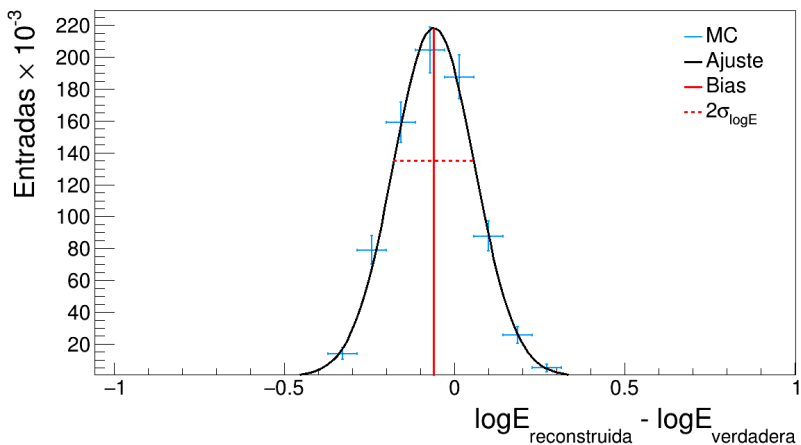


Figura 4.6: Distribución de *bias* para el estimador *NN* en el caso θ_2 para el *bin* 6 (44.7 a 100 TeV). Se muestran las definiciones de *bias* y resolución, los cuales se obtienen de un ajuste gaussiano realizado a la distribución obtenida a partir de simulaciones y de la Ecuación 4.7.

El valor medio de la distribución define el *bias* o sesgo, mientras que el ancho define la resolución de la energía [7, p. 6] (ver Figura 4.6). Estos dos valores cuantifican los efectos generales sobre el estimador de energía debido a la forma en que se modela el desarrollo de la cascada, así como el proceso de reconstrucción [12, p. 64].

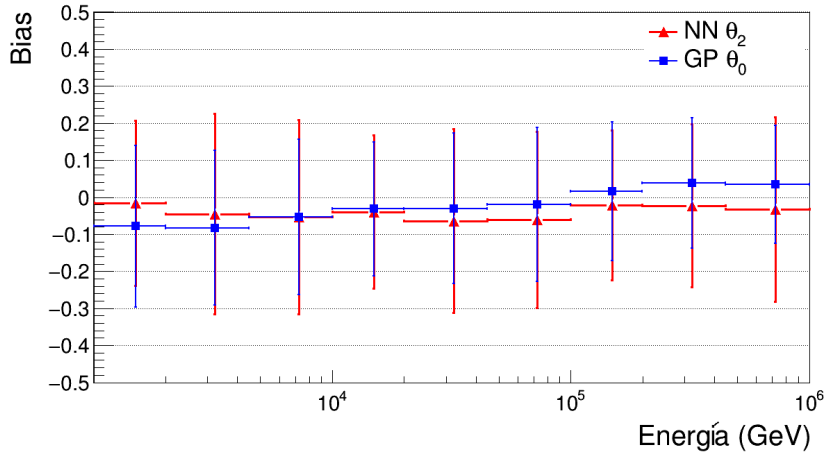


Figura 4.7: Valores de *bias*. La barras verticales representan el ancho de la distribución de *bias* o resolución.

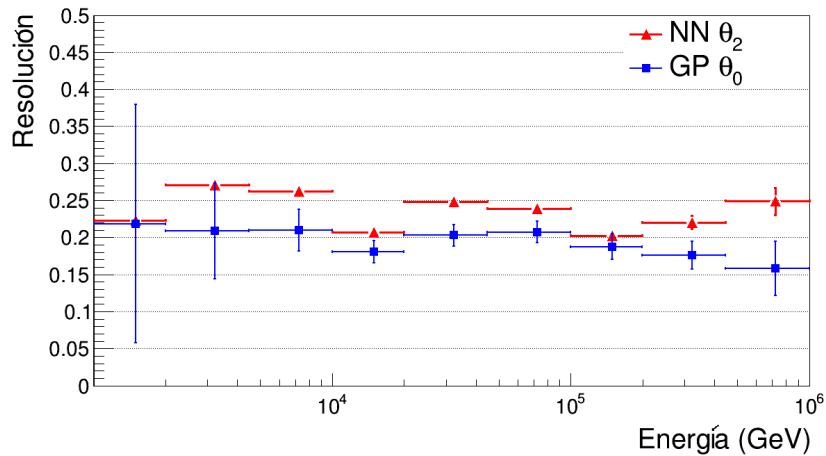


Figura 4.8: Valores para la resolución. Las incertidumbres reportadas en el eje *y* son aquellas calculadas por el ajuste gaussiano.

La Figura 4.7 muestra los Valores de *bias* para todos los *bines* de energía, mientras que la Figura 4.8 muestra los valores de resolución. Ambas figuras son para el caso *NN* θ_2 y para el caso *GP* θ_0 .

4.7. *Unfolding* sencillo

Para aterrizar el concepto del *unfolding* se realizó un ejercicio, para el caso *NN* θ_2 , que consta de obtener la energía de las partículas primarias a partir de una curva de respuesta generada con las simulaciones y aplicada a los datos reales. La curva de respuesta sería la forma análoga uno dimensional de la matriz de respuesta. La descripción de como hice este caso sencillo la pondré a continuación en forma de pasos.

1. Histogramas de la simulación.

Se generan los histogramas de la energía verdadera (variable `mc.logEnergy`) y la energía reconstruida (variable `rec.logNNEnergyV2` para el estimador *Neural Network* o `rec.logGPV2` para el estimador *Ground Parameter*) mostrados en la Figura 4.9.

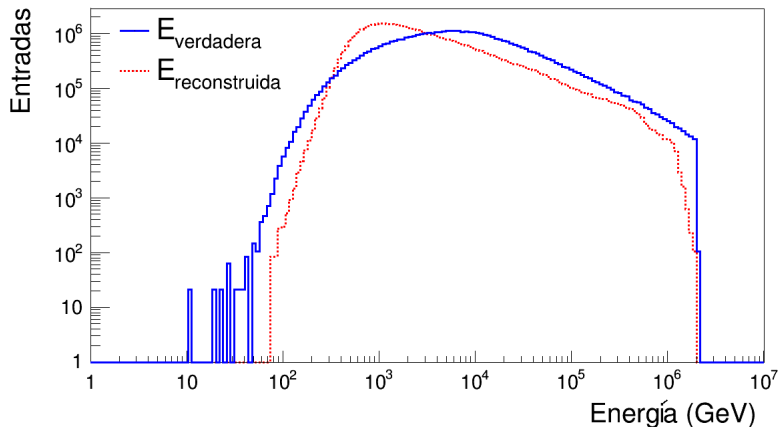


Figura 4.9: Histogramas de energía verdadera y energía reconstruida a partir de simulaciones. Se tienen 4.68×10^7 entradas en total para el caso *NN* θ_2 en todo el rango de energía disponible.

2. Hacer *rebin* en el intervalo 10^3 - 10^6 GeV.

Hacer *rebin* es cambiar el arreglo de *bines* de un histograma tanto en el ancho de *bin* (intervalo de energía que cubre) como en el número de *bines* en un cierto rango de energía.

En este caso de *unfolding* sencillo todos los histogramas deben tener el mismo número de *bines*, a diferencia del caso con *TUnfold* en el que se ha tomado la sugerencia de que el histograma de la energía reconstruida tenga el doble de *bines* que el histograma de la energía verdadera.

Primero se trabajó con 18 *bines*, pues debe ser compatible la curva de respuesta con el histograma de datos reales (el cual ha sido creado en 18 *bines*). Los histogramas obtenidos se muestran en la Figura 4.10.

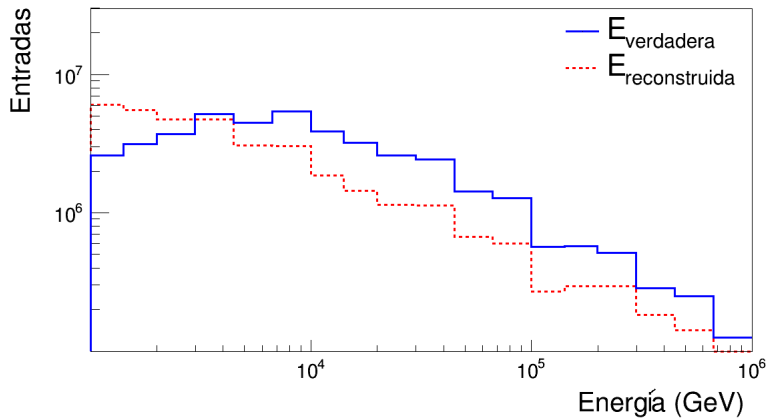


Figura 4.10: *Rebin* basado en la selección de *bines* utilizada en la Sección 4.4.

3. Generar la curva de respuesta.

La curva de respuesta es el factor multiplicativo que relaciona a los histogramas de las energías verdadera y reconstruida de la simulación (ver Figura 4.11). De esta forma se obtiene un histograma que al ser aplicado sobre la energía de los datos reales nos permite recuperar la energía de las partículas primarias utilizando una aproximación muy sencilla.

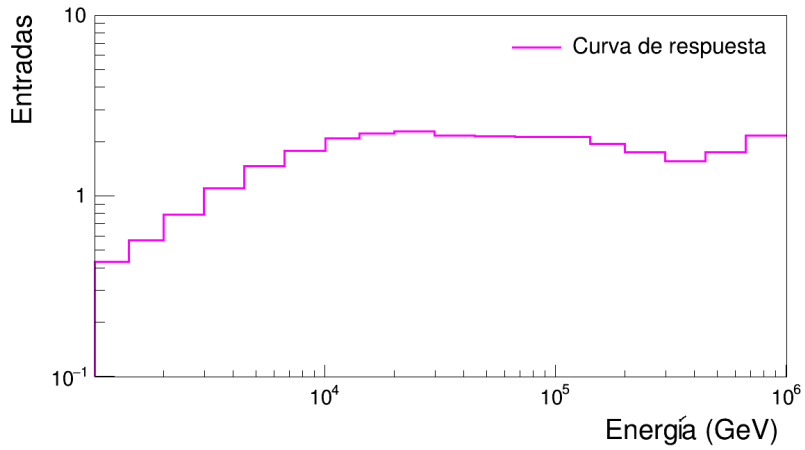


Figura 4.11: Curva de respuesta generada a partir de la simulaciones.

4. Aplicar la curva de respuesta.

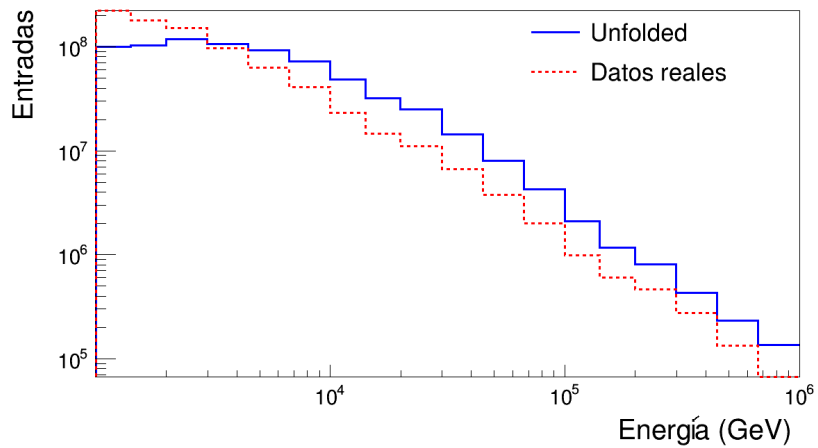


Figura 4.12: Datos reales antes y después de la aplicación de la curva de respuesta.

La curva de respuesta generada puede aplicarse sobre el histograma de datos reales siempre y cuando contengan el mismo número de *bines*. El resultado de este proceso se muestra en la Figura 4.12

5. *Rebin* de 18 a 9 *bines*.

Esto se hace con la finalidad de ser consistente con los *bines* utilizados en ambos casos de *unfolding*, el sencillo como el más complejo, y poder comparar resultados entre ambos métodos (ver Figura 4.13).

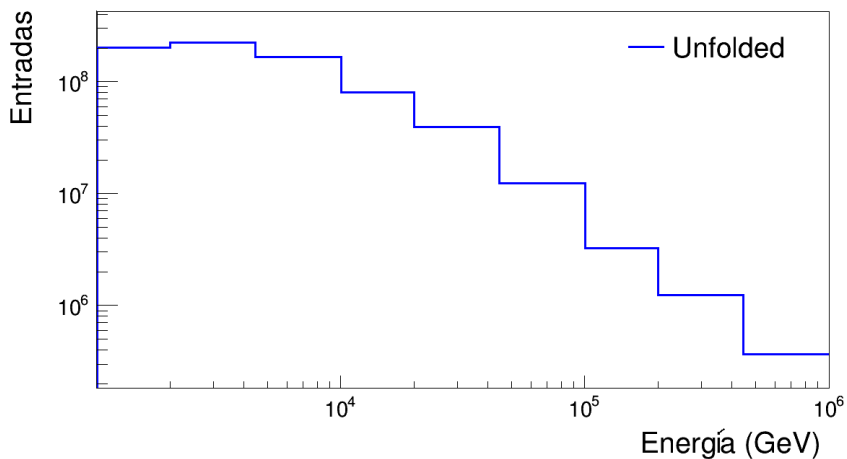


Figura 4.13: Histograma *unfolded* en 9 bins.

6. Normalizar y hacer ajuste.

Para poder comparar el resultado de este *unfolding* sencillo con otros resultados es necesario llevar a cabo una normalización, es decir, presentar la distribución de energía como flujo diferencial de rayos cósmicos tal como se describió en la Sección 4.5. Por último se realiza un ajuste exponencial según lo descrito en la Sección 1.2. En este caso de *unfolding* sencillo no se ha encontrado una forma de definir el rango de energía en el que se ha de

realizar el ajuste, sin embargo se optó por usar el mismo rango utilizado para el caso con *TUnfold* que va de 10 TeV a 447 TeV (del *bin* 4 al 8 en la Tabla 4.2).

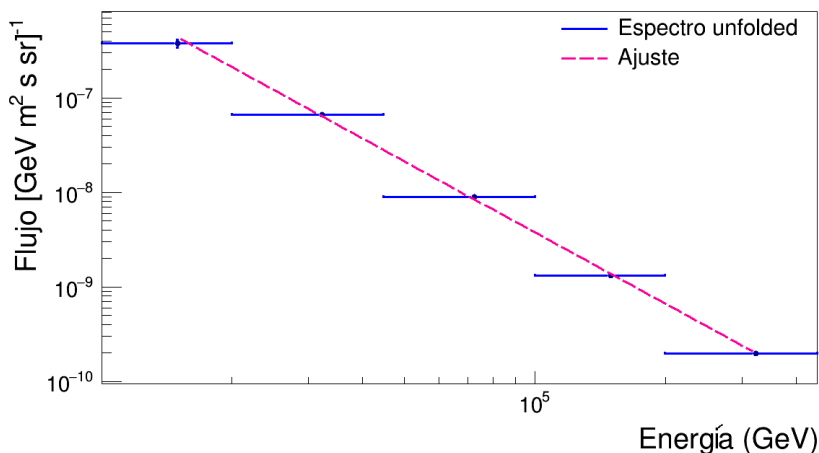


Figura 4.14: Flujo diferencial de energía de rayos cósmicos ajustado por una ley de potencias simple.

En la Figura 4.14 se muestra el resultado de la normalización para este caso de *unfolding* sencillo. Al realizar el ajuste exponencial $\propto E^\gamma$, en el rango de 10 TeV a 447 TeV , se obtiene un índice espectral $\gamma = -2.50 \pm 0.02$.

En los resultados de HAWC, publicados en [7], se muestra el flujo diferencial de rayos cósmicos escalado por un factor de $E^{2.5}$ por lo que para poder comparar el resultado del caso sencillo de *unfolding* con el de HAWC se debe escalar de la misma forma (ver Figura 4.15). El error sistemático se considera a partir del uso de dos diferentes áreas efectivas utilizadas en el análisis (ver el punto sobre área efectiva en la Sección 4.5).

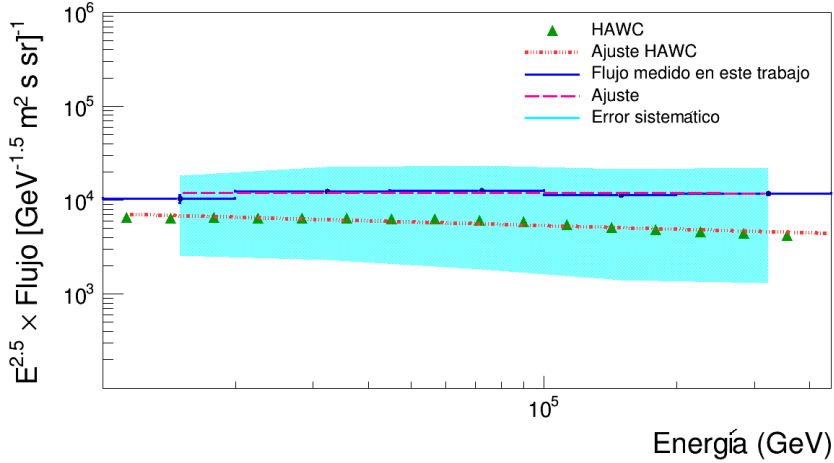


Figura 4.15: Flujo diferencial de energía escalado por $E^{2.5}$.

4.7.1. Conclusiones sobre el caso de *unfolding* sencillo

La importancia de realizar este ejercicio es tener claro el procedimiento a llevar a cabo, desde obtener datos experimentales hasta presentar los resultados a un nivel que se pueda comparar con otros experimentos que tengan la misma finalidad, que en este caso es reconstruir el espectro de energía de los rayos cósmicos. Además de hacer énfasis en lo indispensable que es realizar un proceso de *unfolding* para tomar en cuenta, mediante simulaciones, los efectos que un detector tiene sobre los resultados de una medición.

En el caso de este ejercicio el proceso de *unfolding* fue representado por una aproximación muy sencilla, sin embargo el procedimiento en esencia es el mismo para cualquier otra forma de llevar a cabo el *unfolding*.

Los resultados principales se obtuvieron de las siguientes gráficas: de la Figura 4.14 se obtiene la forma de la estructura del espectro de energía, cuya importancia ya se ha mencionado en la Sección 1.2. Dicho espectro de energía se ha caracterizado por un ajuste exponencial que, en general,

puede variar en el índice espectral por secciones y mostrar *rompimientos* en la estructura, como los mencionados en la Sección 1.2, conocidos como *knee*, *second knee* y *ankle*. Para este caso se ha ajustado con un solo índice espectral en el rango de energía estudiado (10 *TeV* a 447 *TeV*) ya que no se ha observado un indicio que indique que haya un cambio en el índice dentro de dicho rango. El valor que se ha encontrado para el índice espectral es $\gamma = -2.50 \pm 0.02$.

De la Figura 4.15 se obtiene una comparación entre el espectro obtenido con este proceso sencillo y el presentado por HAWC. El resultado ideal esperado tendría que ser la reproducción del espectro de energía obtenido por HAWC, o al menos una transposición de mis resultados dentro de sus barras de error.

Si bien únicamente se ha obtenido una sobreposición dentro del error sistemático (realizado a partir del uso de diferentes valores para el área efectiva), incluso aunque se hubiera obtenido un buen resultado dentro de las barras de error del espectro de energía esto no significaría que el proceso es totalmente confiable ya que debemos recordar que se ha llevado a cabo mediante una aproximación muy sencilla del proceso de *unfolding*.

La utilización de una herramienta más confiable para llevar a cabo el *unfolding* es debido a que, en general, asumir que el factor por el que se corrigen los efectos de un detector es simplemente el factor multiplicativo entre la energía verdadera de datos simulados y su energía reconstruida correspondiente, no es correcto. El código utilizado de *TUnfold* por sí mismo representa una mejor opción ya que ha sido diseñado específicamente para realizar el proceso de *unfolding*, implementando lo repasado sobre álgebra lineal además de otros métodos de optimización de su algoritmo para dar como salida el resultado más confiable. En este trabajo no me he centrado en explicar el desarrollo y funcionamiento del algoritmo de *TUnfold* y de los métodos que implementa, sino en el uso práctico de esta herramienta como se ha visto en la Sección 4.3.3. Para consultar todo los fundamentos matemáticos usados para el desarrollo de *TUnfold* se sugiere revisar [61].

4.8. Análisis con *TUnfold*

Los histogramas obtenidos utilizando el código *TUnfold* se presentan de la siguiente forma: matriz de respuesta, histograma de datos reales e histograma de la distribución unfolded, por ultimo se muestra la forma en que se realiza la normalización. A continuación se muestra el proceso del análisis usando como ejemplo el caso más significativo para cada estimador: *NN* θ_2 (caso angular a 45.0°) para el estimador *Neural Network* y *GP* θ_0 (caso angular a 16.8°) para el estimador *Ground Parameter* (basado en el error porcentual en el índice espectral obtenido en cada caso angular para cada estimador de energía).

Los histogramas de datos reales y las matrices de respuesta están definidos desde 10^3 *GeV* hasta 10^6 *GeV*. Tomando en cuenta las consideraciones que se sugieren en el uso de *TUnfold* descritas en la Sección 4.3.1 y el esquema2 de *bines* de energía de la Sección 4.4 (el cual consta de 9 *bines*), se deben tener 18 *bines* para el histograma de datos reales así como para la energía reconstruida en la matriz de respuesta. Para obtener el doble de *bines* a partir del esquema2, cada *bin* ha sido dividido por la mitad. Por ejemplo, el *bin* 1 va de 10^3 *GeV* a $10^{3.3}$ *GeV* en espacio $\log E$. Luego se definen dos nuevos *bines* de la siguiente forma; *bin* 1 de 10^3 *GeV* a $10^{3.15}$ *GeV* y el *bin* 2 de $10^{3.15}$ *GeV* a $10^{3.3}$ *GeV* en espacio $\log E$.

1. Matriz de respuesta.

En la Figura 4.16 se muestra la matriz de respuesta para el estimador *Neural Network* en el caso angular con corte cenital a 45° . Este histograma ha sido llenado con 3.35×10^7 eventos que han pasado los cortes generales descritos en la Sección 3.5.4. En la Figura 4.17 se muestra la correspondiente matriz para el estimador *Ground Parameter* en el caso angular con corte cenital a 16.8° . Para este caso se tienen 6.77×10^5 eventos que han pasado los cortes generales antes mencionados, además del corte extra utilizado en el estimador *Ground Parameter*.

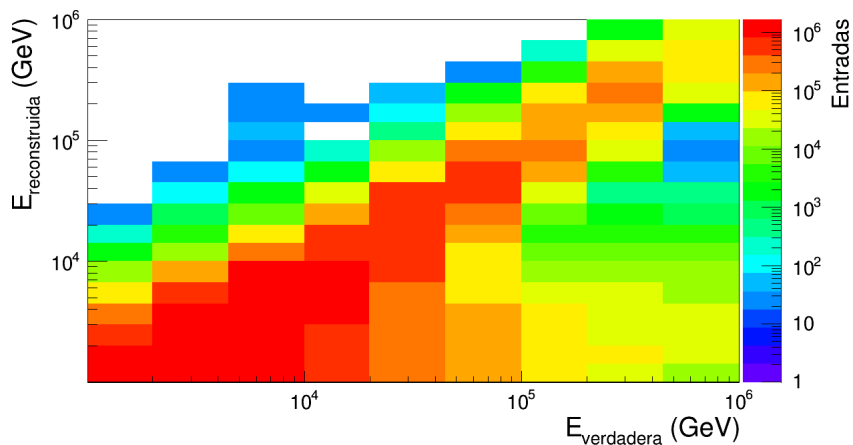


Figura 4.16: Matriz de respuesta para el caso $NN \theta_2$.

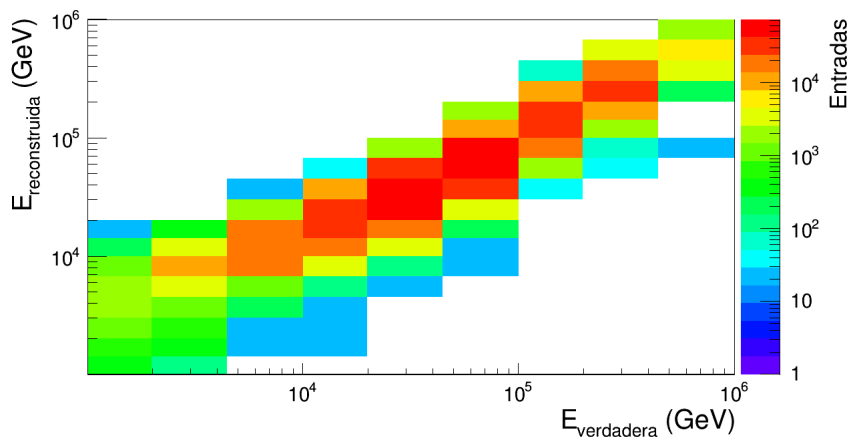


Figura 4.17: Matriz de respuesta para el caso $GP \theta_0$.

Cada evento simulado tiene una energía verdadera predefinida, luego la primer diferencia entre las energías reconstruidas es debida al estimador utilizado. Sin embargo la principal diferencia, que se refleja en la estadística

y apariencia de las dos matrices presentadas anteriormente, es el corte extra utilizado para el estimador *Ground Parameter*. En la Sección 3.5.5 se puede observar la forma en que este corte descarta un gran número de eventos. Por otra parte, para ejemplificar el proceso del análisis se ha tomado un caso angular diferente para cada estimador, lo cual también repercute en la estadística de cada matriz. En el *Apéndice C* se muestra una comparación entre las matrices de los dos estimadores usando el mismo caso angular, además de comparar el efecto del corte extra sobre la matriz obtenida con el estimador *Ground Parameter*.

2. Histograma de datos reales y distribución *unfolded* con *TUnfold*.

Ya que se tiene generada la matriz de respuesta se procede a cargarla en el código de *TUnfold*. El siguiente paso es cargar el histograma de datos reales y ejecutar el código para obtener una distribución *unfolded* como salida. Las Figuras 4.18 y 4.19 muestran, respectivamente, el resultado de este proceso para el estimador *Neural network* en el caso con corte angular a 45° ($NN \theta_2$) y para *Ground parameter* en el caso con corte angular a 16.8° ($GP \theta_o$).

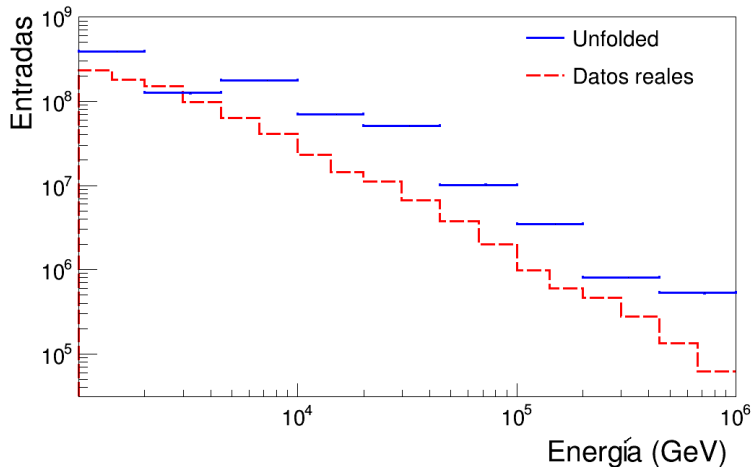


Figura 4.18: Histograma de datos reales y distribución *unfolded* para el caso $NN \theta_2$.

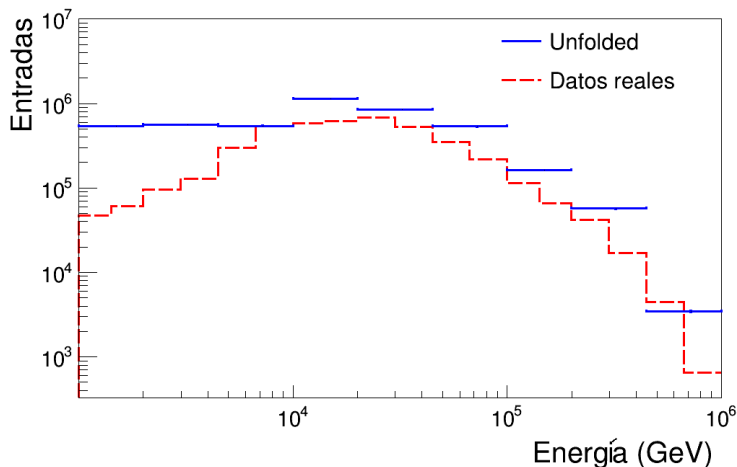


Figura 4.19: Histograma de datos reales y distribución *unfolded* para el caso $GP \theta_0$.

Para el histograma de datos reales del caso $NN \theta_2$ (Figura 4.18) se tienen 8.28×10^8 eventos que pasan los cortes generales, mientras que para el caso $GP \theta_0$ se tienen 4.39×10^6 eventos que han pasado los cortes generales y el corte extra aplicado para el estimador *Ground Parameter*. De nuevo es este último corte extra el que genera un gran diferencia entre las estadísticas de las distribuciones de datos reales (aproximadamente 188 veces más eventos para el estimador *Neural Network* que para *Ground Parameter*), así como una apariencia muy diferente entre ambas distribuciones (ver Figuras 4.18 y 4.19).

Es importante mencionar que, a partir de este punto, en lo siguiente se decidió llevar a cabo el análisis únicamente en el rango de energía que cubren los *bines* del 4 al 8, es decir, de 10 TeV a 447 TeV . Esto debido a que HAWC tiene una mejor precisión en dicho rango (ver Figura 1.4) por encima de los 10 TeV [7]. Lo anterior se ve confirmado por las matrices de respuesta (ver Figuras 4.16 y 4.17), si bien las matrices obtenidas para el estimador NN no ayudan por completo a definir este rango, las obtenidas para GP sí lo hacen pues los *bines* tienen una clara tendencia a la identidad

además de presentar un mayor número de eventos en los *bines* del 4 al 8 (10 *TeV* a 447 *TeV*).

3. Normalización.

La normalización es el último paso para obtener el espectro de energía a partir de la distribución *unfolded*. Como se vio en la Sección 4.5 el flujo diferencial de energía de los rayos cósmicos está dado por la Ecuación 4.4 la cual ha de ser aplicada *bin* por *bin*.

El tiempo de observación es un valor fijo para todos los *bines*, el ángulo sólido solo depende de cada caso angular y en cada caso se mantiene fijo también para todos los *bines* mientras que el valor del ancho de *bin* depende del esquema de *bines* utilizado. Estos tres valores no presentan incertidumbre asociada en este trabajo.

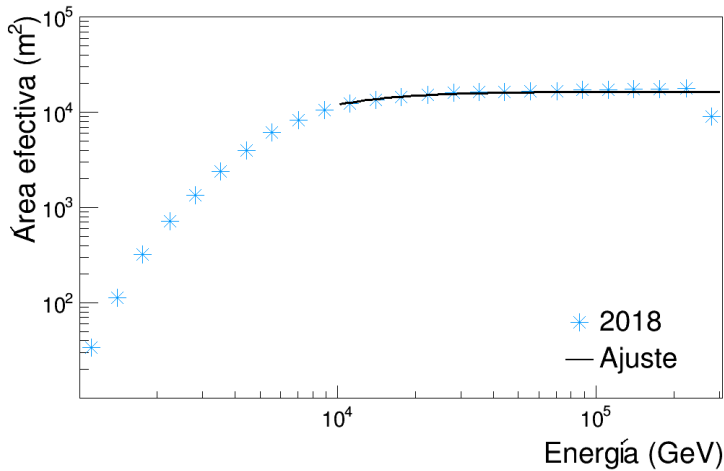


Figura 4.20: Ajuste a curva de área efectiva de 2018.

El área efectiva es un poco más compleja de calcular ya que, como se habrá podido observar a partir de la Figura 4.5, el tomar gráficas de áreas efectivas ya calculadas no permite poder dar un valor en el centro de cada

bin así como en algunos casos no cubre el rango de energía utilizado. Para poder extraer información utilizable a partir de las gráficas de área efectiva se hizo un ajuste a cada curva de la Figura 4.5. La Figura 4.20 muestra un ejemplo para la curva de área efectiva de 2018.

Después de varias pruebas se ha encontrado la siguiente función que ajusta de buena forma a la curva de área efectiva.

$$A_{eff}(E) = \exp [p_0 - p_1 E^{-c}] \quad (4.8)$$

en donde E es la energía, $c = 1.71$, $p_0 = 9.71 \pm 0.04$ y $p_1 = 2.22 \times 10^6 \pm 1.19 \times 10^6$ son los parámetros del ajuste.

En la Tabla 4.4 se muestran los valores de c , p_0 y p_1 así como sus errores asociados, para todas las curvas de área efectiva.

Curva A_{eff}	c	$p_0 \pm \Delta p_0$	$p_1 \pm \Delta p_1$
16° - 2011	0.97	11.6 ± 0.06	1051 ± 1054
35° - 2011	0.61	11.9 ± 0.26	207 ± 146
45° - 2011	0.53	12.1 ± 0.18	197 ± 48.6
2018	1.71	9.71 ± 0.04	$2.22 \times 10^6 \pm 1.19 \times 10^6$

Tabla 4.4: Resultados del ajuste a las curvas de área efectiva. Los valores se obtuvieron pidiendo que se realizara el ajuste en el rango de energía cubiertos desde el *bin* 4 al *bin* 8 (energía de 10 *TeV* a 447 *TeV*).

Una vez teniendo una función que describe el área efectiva lo siguiente es evaluarla en el centro de cada *bin*. Con esto ya se tienen todas las cantidades necesarias para calcular el flujo *bin* por *bin* sin olvidar que se debe considerar la propagación de errores debidos al ajuste del área efectiva. En la Sección D.1 del *Apéndice D* se explica la propagación de errores debida al uso de las curvas a de área efectiva.

Con el área efectiva calculada, y las cantidades mencionadas en la Sección 4.5, es decir, tiempo de 830, 400 s, ángulo sólido dado por la Tabla 4.3,

y la energía correspondiente al ancho de cada *bin*, se puede hacer uso directamente de la Ecuación 4.4 para obtener, *bin* por *bin*, el flujo diferencial de rayos cósmicos.

Como se puede observar, en la Ecuación 4.4 de flujo diferencial de rayos cósmicos, se tiene una división en la que la distribución de energía $N(E)$ (dividendo) está dada por el número de entradas en cada *bin* del histograma de la distribución *unfolded*. Esta cantidad $N(E)$ tiene un error asociado que proviene del error asociado a los eventos registrados en cada *bin* del histograma de datos reales y del proceso de *unfolding* realizado por el código de *TUnfold*. Por otra parte, el área efectiva dada por A_{eff} (en el divisor) tiene un error asociado que ha sido propagado manualmente. Por lo tanto se tiene un error asociado al flujo diferencial de rayos cósmicos, el cual ha sido calculado en la Sección D.2 del *Apéndice D*.

Por último se presenta el resultado de la normalización, es decir, el flujo diferencial o espectro diferencial de energía de rayos cósmicos. A continuación se muestran los histogramas obtenidos para los dos casos que se eligieron para ejemplificar el proceso de análisis con *TUnfold*.

La Figura 4.21 muestra el resultado para el estimador *Neural Network* en el caso angular θ_2 (45.0°), mientras que la Figura 4.22 muestra el resultados para el estimador *Ground Parameter* en el caso angular θ_0 (16.8°).

De acuerdo al ajuste exponencial $\propto E^\gamma$ en el rango de 10 *TeV* a 447 *TeV* discutido en la Sección 1.2, para el caso *NN* θ_2 se obtuvo un índice espectral con valor de $\gamma = -2.63$ mientras que para el caso *GP* θ_0 se obtuvo $\gamma = -2.14$

El proceso anterior fue realizado para todos los casos angulares en cada estimador de energía. Los resultados obtenidos de acuerdo al juste realizado y sus respectivos errores asociados propagados son reportados en la siguiente sección .

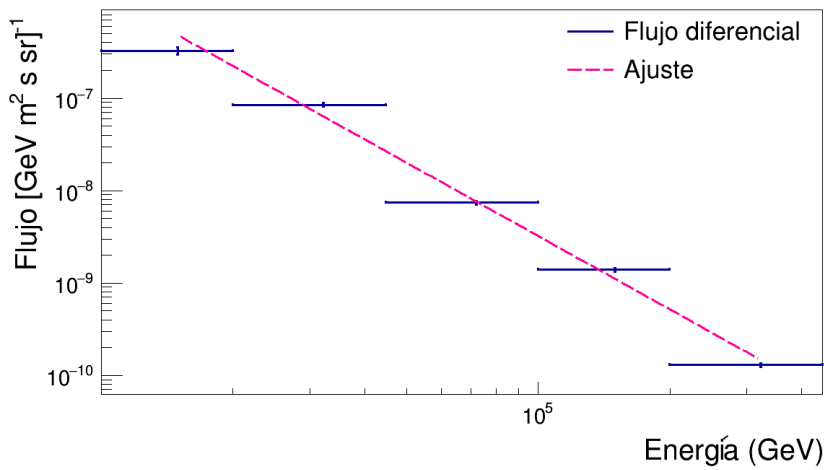


Figura 4.21: Espectro diferencial de energía de rayos cósmicos ajustado por una ley de potencias simple para el caso $NN \theta_2$.

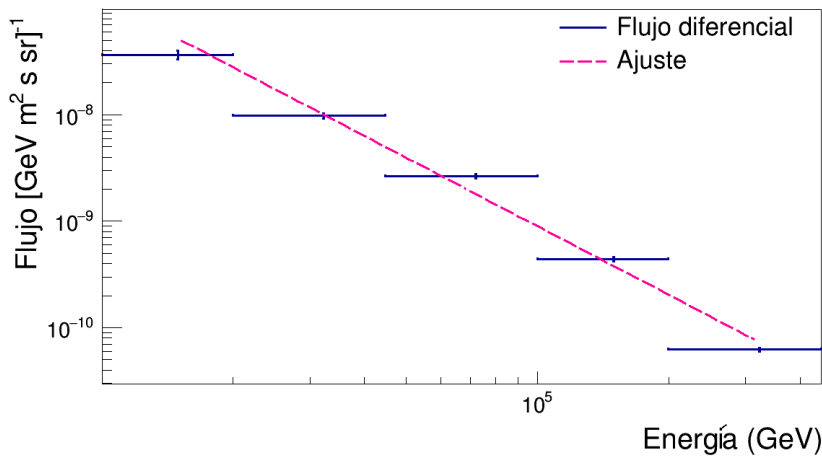


Figura 4.22: Espectro diferencial de energía de rayos cósmicos ajustado por una ley de potencias simple para el caso $GP \theta_0$.

Capítulo 5

Resultados

5.1. Índices espectrales obtenidos

Del análisis anterior se obtiene el flujo diferencial de energía de los rayos cósmicos que, como se vio en la Sección 1.2, puede ser ajustado por una ley de potencias simple de donde nos interesa obtener el valor del índice (por las razones también mencionadas en la Sección 1.2). En la publicación de HAWC [7], que es la fuente principal en la que se basa este trabajo, se reporta un valor del índice de -2.63 ± 0.01 . A continuación se muestran los resultados para todos los casos estudiados dentro del rango de energía elegido (10 TeV a 447 TeV).

Caso angular	Índice	Error $\times 10^{-2}$	Error porcentual	Chi ² /ndf
θ_0	-2.72	2.59	0.95	42.65/3
θ_1	-2.66	2.28	0.86	16.26/3
θ_2	-2.63	2.02	0.77	83.29/3

Tabla 5.1: Resultados para 10 *runs* con el estimador *Neural Network*

En las Tablas 5.1 y 5.2 se observa que los mejores resultados son el caso *NN* θ_2 y *GP* θ_0 ya que tienen los errores asociados y errores porcentuales más pequeños que los otros casos de sus respectivos estimadores.

Caso angular	Índice	Error $\times 10^{-2}$	Error porcentual	Chi2/ndf
θ_0	-2.14	1.90	0.89	101.84/3
θ_1	-2.03	1.92	0.94	83.49/3
θ_2	-1.96	1.95	0.99	79.48/3

Tabla 5.2: Resultados para 10 *runs* con el estimador *Ground Parameter*

5.2. Cambio en los resultados respecto a la cantidad de datos analizados

Se realizó también un análisis sobre la forma en que cambian los resultados obtenidos conforme se aumenta la cantidad de datos analizados. El número de eventos consta de aquellos que pasaron los cortes aplicados en el rango de energía del análisis (10 *TeV* a 447 *TeV*, del *bin* 4 al *bin* 8). Tanto en los valores del índice como de su error asociado se muestran tres cifras después del punto decimal ya que los valores presentan cambios notables hasta esas cifras. En las Tablas 5.3 y 5.4 se muestra, respectivamente, lo obtenido para los dos caso más relevantes: *NN* θ_2 y *GP* θ_0 .

- Para *NN* θ_2 :

Runs	Eventos $\times 10^8$	Índice	Error $\times 10^{-2}$
1	0.84	-2.632	2.049
2	1.68	-2.633	2.030
3	2.51	-2.635	2.035
4	3.35	-2.635	2.030
5	4.17	-2.635	2.030
6	5	-2.635	2.028
7	5.82	-2.635	2.026
8	6.64	-2.636	2.027
9	7.46	-2.636	2.026
10	8.28	-2.636	2.026

Tabla 5.3: Resultados para el caso *NN* θ_2 respecto al aumento de datos.

Para poder observar mejor la dependencia entre el índice espectral y la cantidad de datos (*runs*) se realizaron las gráficas mostradas en las Figuras 5.1 y 5.2.

Se hizo un ajuste a los valores del índice espectral y de su error asociado mediante una función del tipo Ax^B en donde x representa la cantidad de datos analizados, A y B son los parámetros determinados por el mejor ajuste. Para el ajuste a los índices espectrales (Figura 5.1) se obtuvo $A = -2.63 \pm 0.28 \times 10^{-3}$ y $B = 0.62 \times 10^{-3} \pm 0.06 \times 10^{-3}$.

Realizando un ajuste del mismo tipo para los valores de los errores asociados (Figura 5.2) se obtuvo $A = 2.04 \pm 2.55 \times 10^{-3}$ y $B = -4.04 \times 10^{-3} \pm 0.75 \times 10^{-3}$.

Para el resultado obtenido en [7] se utilizaron 234 días de datos. Evaluando la función encontrada en este valor se puede dar un resultado aproximado de lo que se obtendría con 234 *runs* en este análisis. El valor del índice con esta aproximación converge rápidamente a $-2.64 \pm 2.00 \times 10^{-2}$.

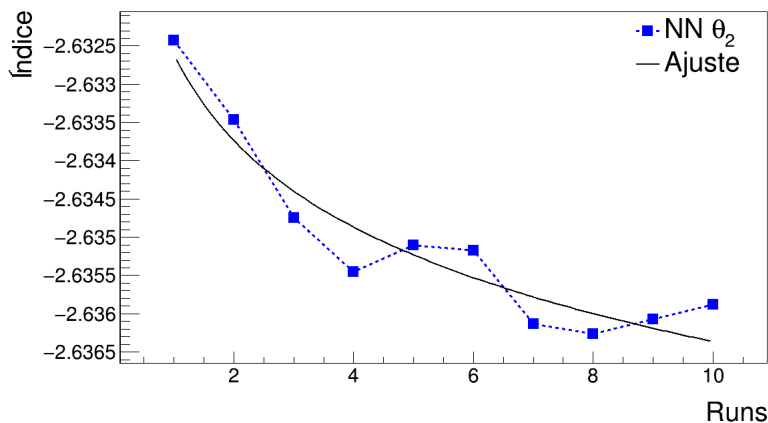


Figura 5.1: Valores del índice espectral respecto al número de *runs*.

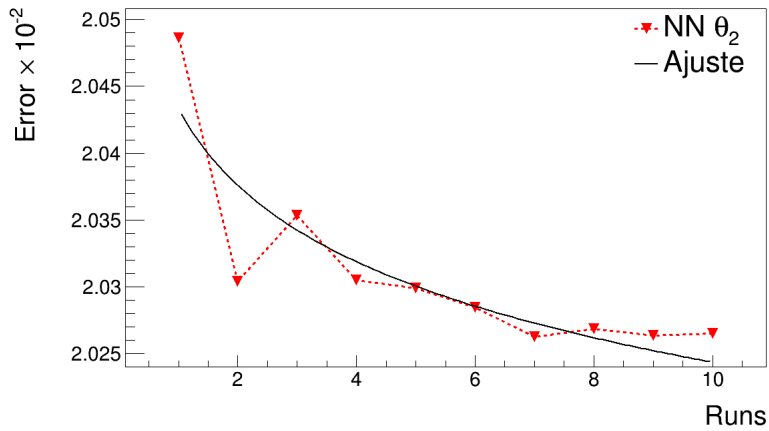


Figura 5.2: Valores del error asociado del índice espectral respecto al número de *runs*.

- Para $GP \theta_0$:

Para el caso del estimador *Ground Parameter* los resultados de este análisis se muestran en la Tabla 5.4 y las Figuras 5.3 y 5.4

Runs	Eventos $\times 10^6$	Índice	Error $\times 10^{-2}$
1	0.44	-2.124	1.957
2	1.88	-2.133	1.925
3	1.32	-2.133	1.916
4	1.76	-2.135	1.909
5	2.2	-2.135	1.906
6	2.64	-2.135	1.901
7	3.07	-2.137	1.899
8	3.51	-2.136	1.899
9	3.95	-2.137	1.897
10	4.39	-2.136	1.897

Tabla 5.4: Resultados para el caso $GP \theta_0$ respecto al aumento de datos.

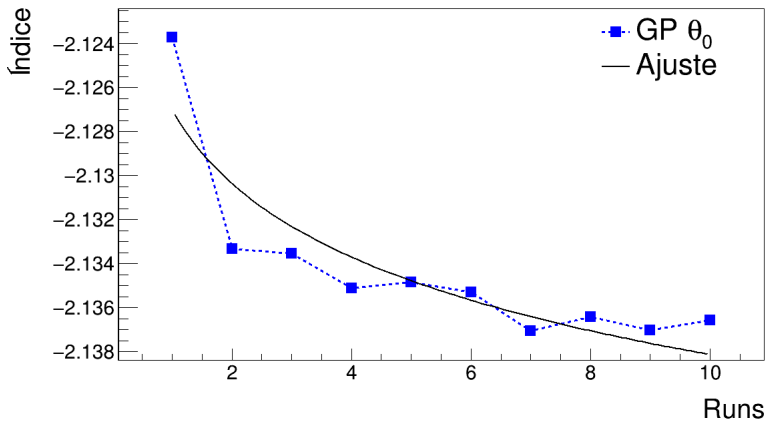


Figura 5.3: Valores del índice espectral respecto al número de *runs*.

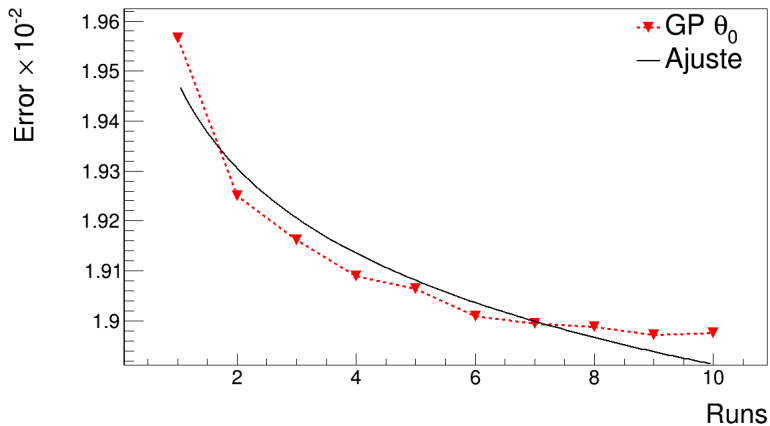


Figura 5.4: Valores del error asociado del índice espectral respecto al número de *runs*.

Del ajuste al valor de los índices (Figura 5.3) se obtuvo $A = -2.13 \pm 1.38 \times 10^{-3}$ y $B = 2.26 \times 10^{-3} \pm 0.39 \times 10^{-3}$. Para el valor del error asociado

(Figura 5.4) se obtuvo $A = 1.95 \pm 3.96 \times 10^{-3}$ y $B = -12.77 \times 10^{-3} \pm 1.23 \times 10^{-3}$.

Evaluando en 234 días la función descrita con los parámetros anteriores el valor del índice converge a $-2.15 \pm 1.82 \times 10^{-2}$.

5.3. Comparación con los resultados de HAWC

Con la finalidad de poder comparar los resultados obtenidos con el publicado por HAWC en R. Alfaro et al [7] se multiplica el flujo de rayos cósmicos por un factor de $E^{2.5}$. Este escalamiento por el valor de la energía elevada a un exponente positivo similar al índice exponencial del espectro obtenido se suele hacer con la finalidad de percibir con mayor facilidad algún posible cambio en la estructura del espectro.

De esta forma, la Figura 5.5 representa el resultado, para el caso del estimador *Neural Network* en el caso angular con corte a 45.0° ($NN \theta_2$), mostrado en la Figura 4.21 una vez que se ha multiplicado por el factor $E^{2.5}$. La Figura 5.6 muestra este escalamiento aplicado al resultado para el caso del estimador *Ground Parameter* en el caso angular con corte a 16.8° ($GP \theta_0$) mostrado en la Figura 4.22.

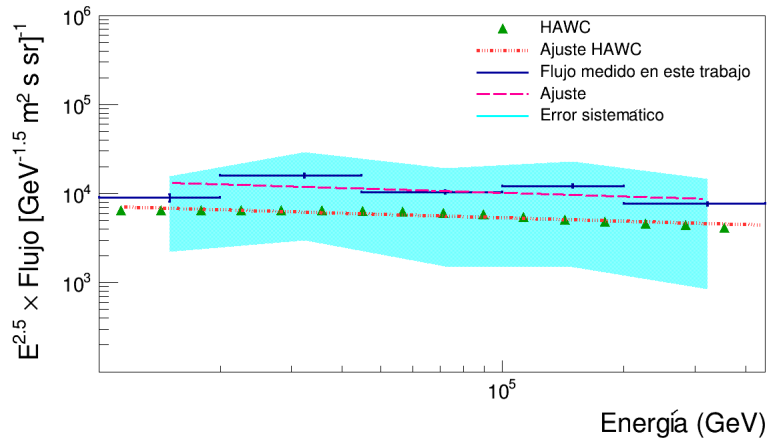


Figura 5.5: Espectro de energía obtenido para el caso $NN \theta_2$.

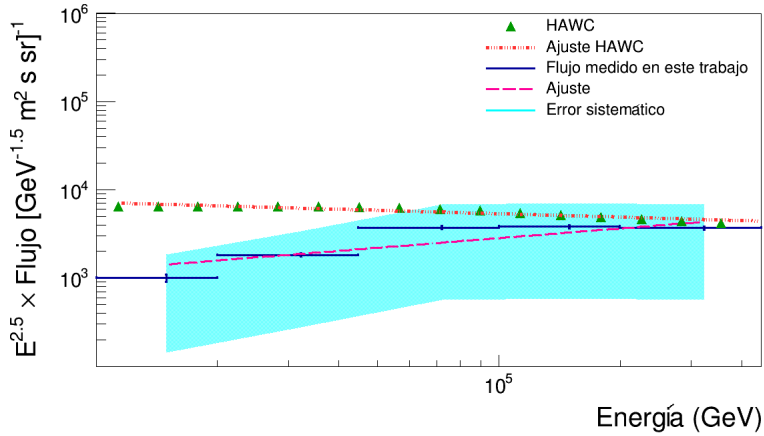


Figura 5.6: Espectro de energía obtenido para el caso $GP \theta_0$.

El resultado publicado por HAWC se ajusta con una ley de potencias $p_0 E^{p_1}$ en donde $p_0 = 10.06 \pm 0.15$ y $p_1 = -12.79 \times 10^{-2} \pm 1.38 \times 10^{-2}$.

Las Tablas 5.5 y 5.6 muestran una comparación entre el error estadístico y el error sistemático (el cual ha sido calculado a partir de las áreas efectivas utilizadas) para el caso más relevante de cada estimador de energía.

Bin	Contenido $\times 10^3$	Error sistemático $\times 10^3$	Error estadístico $\times 10^3$
4	8.96	6.74	0.86
5	15.97	12.99	0.78
6	10.42	8.9	0.45
7	12.12	10.61	0.52
8	7.70	6.85	0.33

Tabla 5.5: Errores sistemáticos y estadísticos para el caso $NN \theta_2$.

Bin	Contenido $\times 10^3$	Error sistemático $\times 10^3$	Error estadístico $\times 10^3$
4	1.00	0.86	0.10
5	1.83	1.55	0.09
6	3.72	3.15	0.16
7	3.84	3.26	0.17
8	3.70	3.14	0.16

Tabla 5.6: Errores sistemáticos y estadísticos para el caso $GP \theta_0$.

El error sistemático en ambos casos es aproximadamente un orden de magnitud mayor que el error estadístico. Para el caso $NN \theta_2$ el error sistemático cubre los resultados de HAWC (Figura 5.5) mientras que para el caso $GP \theta_0$ el error sistemático solo cubre a los resultados de HAWC en los *bines* 6, 7 y 8 (Figura 5.6), los cuales que cubren la energía de 44.7 TeV a 447 TeV . En ambos casos el error estadístico no cubre los resultados de HAWC. A continuación se muestran los resultados al realizar un ajuste de tipo $p_0 E^{p_1}$ al flujo escalado por $E^{2.5}$ (ver Figuras 5.5 y 5.6).

Caso	p_0	Error p_0	$p_1 \times 10^{-2}$	Error $p_1 \times 10^{-2}$
HAWC	10.06	0.15	-12.79	1.38
$NN \theta_2$	10.79	0.23	-13.59	2.03
$GP \theta_0$	1.2	0.23	54.13	1.96

Tabla 5.7: Comparación entre el resultado de HAWC y los resultados de los mejores casos de cada estimador. El resultado mas consistente con el publicado por HAWC es el caso $NN \theta_2$

Capítulo 6

Discusión y conclusiones

En este trabajo se llevó a cabo la reconstrucción del espectro de energía de los rayos cósmicos en un rango de 10 TeV a 447 TeV haciendo uso de dos algoritmos de estimación de energía: *Neural Network* y *Ground Parameter*.

Se estudiaron tres diferentes casos angulares para cada estimador con la finalidad de comparar los resultados en cada caso, ya que para cortes angulares pequeños (16.8°) el detector muestra una mejor resolución obteniendo señales con menos ruido. Por otro lado, con un corte angular mayor (45.0°) se obtiene una mayor estadística. Se realizó también un corte angular intermedio (35.2°) para tener un tercer caso con qué comparar los otros dos caso extremos.

Después se prosiguió a llenar histogramas con datos reales y simulados para poder llevar a cabo el análisis. En este punto se tomó la elección de un cierto esquema de *bines* de energía para poder realizar las distribuciones de datos reales y las matrices de respuesta con eventos simulados.

Se llevó a cabo un análisis de distribución de *bias* con eventos simulados, el cual muestra la diferencia entre la energía verdadera y aquella reconstruida por un estimador. De esta forma se obtuvo el sesgo que hay entre ambas cantidades y la resolución de energía para cada *bin*. El sesgo obtenido a partir de este análisis es cercano a cero, lo que indica una adecuada reconstrucción de la energía. De esta forma se dedujo que el uso de los estimadores resulta confiable para reconstruir la energía de los rayos cósmicos. Por otra parte, la resolución de los *bines* dan un indicio sobre el intervalo de energía que debería cubrir cada *bin*, sin embargo debido a los

errores discutidos en la Sección 4.4 y mostrados en el *Apéndice B*, se optó por seguir el análisis con el esquema de bins elegido.

Se estudiaron los aspectos generales necesarios para la comprensión del proceso de *unfolding* el cual nos permite, a través de simulaciones, obtener una medición más precisa. En breves palabras, este proceso lo podemos describir como la obtención de una relación entre el valor verdadero de una variable y la medición que se obtiene a través de un experimento. Puesto que en la vida real no podemos saber el valor verdadero de dicha variable (en tal caso no habría necesidad de hacer otra cosa más que medir directamente), en el proceso de *unfolding* siempre se van a utilizar simulaciones. Una vez obtenida la relación entre el valor verdadero y el reconstruido, se procedió a resolver el problema inverso para, a partir de una variable reconstruida, encontrar el valor verdadero, o al menos el más preciso posible.

Resolver dicho problema inverso resulta no ser una tarea trivial. Para tener más claro el proceso descrito anteriormente sobre el problema inverso y entender de manera más intuitiva de que trata el *unfolding*, se realizó un ejercicio de *unfolding sencillo* en el que la relación entre la energía verdadera y reconstruida era el factor multiplicativo entre ambas distribuciones. De esta forma el problema inverso fue simplemente aplicar el inverso multiplicativo del factor encontrada sobre la distribución de energía medida. Posteriormente se realizó el proceso de *unfolding* haciendo uso de un algoritmo más especializado llamado *TUnfold* en el que la relación entre la energía verdadera y la reconstruida es un histograma en dos dimensiones llamado matriz de respuesta.

Después de obtener las distribuciones de energía con *TUnfold*, estas debían ser normalizadas para poder compararlas con los resultados de HAWC (y en general con los resultados de otros experimentos sobre el espectro de rayos cósmicos, ya que la normalización quita las características del detector). Para llevar a cabo la normalización se requirió del área efectiva del detector. Calcular el área efectiva supondría todo un análisis y estudio aparte por lo que se optó por hacer uso de áreas efectivas ya obtenidas por la colaboración de HAWC. Se tomó una área efectiva calculada en 2018 como principal ya que esta contaba con algunos cortes similares a los utilizados para este trabajo. También se realizó el análisis de manera alternativa usando otra tres áreas efectivas, calculadas en 2011, cada una con un corte en el ángulo cenital muy similar a los utilizados para cada caso angular de

este trabajo. Estas áreas efectivas no se tomaron como principales ya que en 2011 no existían algunos cortes que sí se utilizaron para la del 2018. El resultado obtenido a partir de estas áreas efectivas alternativas fue utilizado para establecer un error sistemático.

Al final se obtuvo como resultado, para cada caso angular y cada estimador, el espectro de energía de los rayos cósmicos (también llamado flujo diferencial de energía) el cual fue ajustado por una ley de potencias simple. El índice espectral obtenido de dicho ajuste caracteriza cada resultado obtenido y el error asociado fue propagado tomando en cuenta el error asociado a la distribución de datos reales (error calculado por el software *Root* al momento de crear los histogramas) y el error asociado al uso de áreas efectivas ya calculadas.

Basado en el error porcentual (como se compara el valor del error asociado con el valor del índice espectral), se llegó a la conclusión que el mejor resultado obtenido para el estimador *Neural Network* fue en el caso angular con corte a 45.0° mientras que para el estimador *Ground Parameter* fue el caso angular con corte a 16.8° .

Por último se realizó una comparación entre los mejores resultados de cada estimador y el resultado publicado por HAWC. Debido a que el proceso implementado por la colaboración es más especializado y detallado, el índice que ellos obtuvieron es más preciso por lo que, en un caso ideal, los resultados de este trabajo deberían coincidir con los de HAWC dentro de las barras de error, lo cual no se logró con ninguno de los estimadores. Dentro de la comparación, haciendo uso del error sistemático debido a la utilización de dos diferentes áreas efectivas (la principal del 2018 y la alternativa de cada caso angular), el resultado más cercano al obtenido por HAWC fue el caso *NN* θ_2 .

Los resultados de este trabajo han sido muy buenos tomando en cuenta que se logró utilizar estimadores de energía diseñados para rayos gammas para poder medir el espectro de energía de los rayos cósmicos. Para el caso del estimador *Neural Network* con corte angular a 45° , el índice espectral obtenido coincide con el publicado por HAWC dentro de sus errores asociados.

Sin embargo, para obtener mejores resultados, el proceso completo llevado a cabo en esta tesis puede ser mejorado. Dentro de los aspectos más técnicos se debería considerar un análisis más detallado en la elección de

un esquema de *bines* y por lo tanto dar un peso más relevante al análisis de la *distribución de bias*, así como la comprensión a fondo de todas las herramientas que ofrece el algoritmo de *TUnfold* para la optimización de resultados.

Dentro de los aspectos teóricos, buscar la optimización de los resultados haciendo una análisis sobre los cortes que deberían ser aplicados para cada caso. Por ejemplo, el corte extra utilizado en el estimador *Ground Parameter* tuvo como resultado una matriz de respuesta que presentaba menos sesgo, sin embargo los valores para este corte extra sobre el parámetro de edad fueron elegidos tomando en cuenta las referencias encontradas en la literatura. Un estudio específico sobre como implementar este corte para las características de HAWC podría ser aplicado.

Otro gran aspecto para mejorar los resultado (y quizá el más relevante para este trabajo) es la utilización de una área efectiva calculada específicamente para los casos analizados, es decir, con exactamente los mismo cortes utilizados.

Dentro de la colaboración de HAWC se siguen llevando a cabo investigaciones con el propósito de construir algoritmos de estimación de energía específicos para rayos cósmicos por lo que el presente trabajo presentado puede servir de apoyo. También se pretende que los temas estudiados puedan servir a aquellos estudiantes que se introducen en temas de física experimental (y más en específico para aquellos que pretendan colaborar en líneas de investigación llevadas a cabo con HAWC) a comprender algunos de los aspectos más relevantes involucrados en este campo de la ciencia: desde la obtención de datos, el software con el que se procesan dichos datos, así como la utilización de simulaciones para obtener mediciones más precisas mediante el proceso de *unfolding* y por último llegar a presentar resultados que pueden ser puestos en comparación directamente con los de otros experimentos que tengan la misma finalidad.

Apéndice A

Mínimos cuadrados

A.1. Ejemplo introductorio

Supongamos que se observa el movimiento de un objeto sobre un plano. Desde nuestro punto de vista se determina que el movimiento del objeto es sobre una línea recta; $y = mx + b$ en donde m es la pendiente de la recta y b la ordenada al origen. Debido a la naturaleza experimental se toman tres datos en diferentes posiciones; (x_1, y_1) , (x_2, y_2) y (x_3, y_3) .

Se obtiene el siguiente sistema de ecuaciones:

$$\begin{aligned}b + mx_1 &= y_1. \\b + mx_2 &= y_2. \\b + mx_3 &= y_3.\end{aligned}\tag{A.1}$$

En el caso ideal en que se tienen mediciones exactas, las ecuaciones serían compatibles entre sí y la pendiente y ordenada al origen podrían ser determinadas usando solo dos de las ecuaciones. Pero en el mundo real experimental los errores surgen por todas partes por lo que el sistema será inconsistente (la Figura A.1 muestra un ejemplo). Sin embargo se puede obtener una solución aproximada en el sentido de los mínimos cuadrados.

La idea es encontrar m y b tal que se minimiza la suma del cuadrado de los errores.

$$\Sigma = (b + mx_1 - y_1)^2 + (b + mx_2 - y_2)^2 + (b + mx_3 - y_3)^2. \quad (\text{A.2})$$

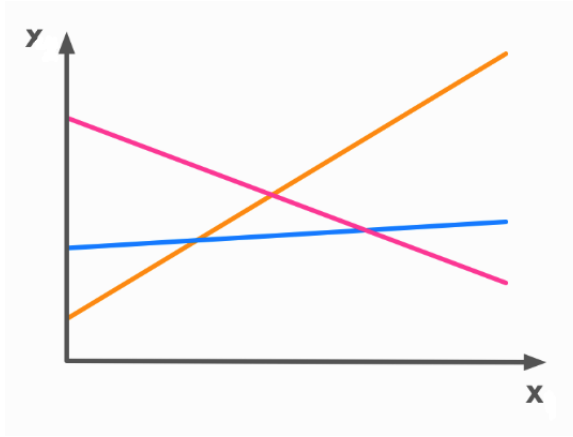


Figura A.1: Representación de un sistema inconsistente. Se observa que no existe un punto en el que se interseccionen las tres rectas por lo que el sistema no tiene una solución exacta [70].

A.2. Caso general

Cuando se tiene una matriz $m \times n$ sobre-determinada de un sistema $Ax = b$ Gauss y Legendre encontraron una solución que minimiza:

$$\| Ax - b \|^2. \quad (\text{A.3})$$

Dicha solución está dada por la matriz *pseudo-inversa* A^\dagger .

$$A^\dagger b = x^* \approx x, \quad (\text{A.4})$$

en donde x^* es la solución que más se aproxima a x en el sentido de los mínimos cuadrados. A continuación se hace un repaso de los conceptos necesarios para la deducción del resultado de la Ecuación A.4 también llamada *pseudo-inversa de Moore-Penrose*.

A.3. Herramientas de álgebra lineal

- Espacio columna de una matriz. Sea $A \in \mathbb{M}_{m \times n}$ una matriz de m renglones por n columnas; $A = [\vec{a}_1 \ \vec{a}_2 \ \cdots \ \vec{a}_n]$ en donde $\vec{a}_i \in \mathbb{R}^m$ con $i = 1, 2, \dots, n$. El espacio columna de dicha matriz $C(A) \subseteq \mathbb{R}^m$ se define como el espacio vectorial generado por las columnas (vectores columna) linealmente independientes de A . Del mismo modo se puede definir el espacio generado por los renglones de la matriz llamado *espacio fila*. Se puede comprobar que el espacio fila de una matriz es igual al espacio columna de la matriz transpuesta.
- Vector proyección (Figura A.2).

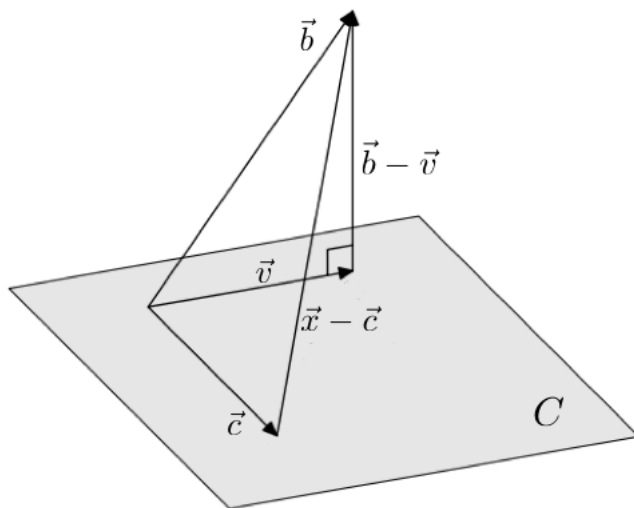


Figura A.2: Diagrama del Teorema 1 [71].

Teorema 1. Sea $C \subseteq \mathbb{R}^m$ un subespacio de \mathbb{R}^m , para todo $\vec{b} \in \mathbb{R}^m$ existe un único vector $\vec{v} \in C$ tal que;

- (1) $(\vec{b} - \vec{v}) \perp \vec{c} \quad \forall \vec{c} \in C$.
- (2) $\|\vec{b} - \vec{v}\| \leq \|\vec{b} - \vec{c}\| \quad \forall \vec{c} \in C$.

Al vector \vec{v} se le conoce como la proyección ortogonal de \vec{b} sobre C ; $\vec{v} = \text{Proy}_C \vec{b}$. Ver C. Arce, W. Castillo, y J. González [71, Pág. 252] para conocer la demostración.

- Espacio nulo de una matriz. Sea $A \in \mathbb{M}_{m \times n}$ una matriz de m renglones por n columnas, el espacio nulo de A se define como:

$$N(A) = \{\vec{x} \in \mathbb{R}^n \mid A\vec{x} = \vec{0} \in \mathbb{R}^m\}. \quad (\text{A.5})$$

- Espacio complemento ortogonal. Sea $C \subseteq \mathbb{R}^m$ un subespacio de \mathbb{R}^m , el complemento ortogonal de C se define como:

$$C^\perp = \{\vec{x} \in \mathbb{R}^m \mid \vec{x} \cdot \vec{c} = 0 \ \forall \vec{c} \in C\}. \quad (\text{A.6})$$

Teorema 2. Sea $A \in \mathbb{M}_{m \times n}$ una matriz de m renglones por n columnas. El espacio nulo de A es igual al complemento ortogonal del espacio columna de A^\perp .

$$N(A) = C(A^T)^\perp.$$

Demostración. Para llevar a cabo la demostración se va a probar la doble contención.

Caso 1: $N(A) \subseteq C(A^T)^\perp$.

Supongamos que tomamos un vector \vec{x} que pertenece al espacio nulo de A , tenemos entonces que:

$$A\vec{x} = \vec{0},$$

y reescribimos la matriz con sus vectores renglón \vec{r}_i con $i = 1, 2, \dots, m$:

$$\begin{aligned} \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix} \vec{x} &= \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \\ \Rightarrow \quad \vec{r}_1 \cdot \vec{x} &= 0 \\ \vec{r}_2 \cdot \vec{x} &= 0 \\ &\vdots \\ \vec{r}_m \cdot \vec{x} &= 0 \end{aligned}$$

Por lo que para cualquier vector $\vec{x} \in N(A)$ se tiene que este es ortogonal a todas las filas de A ; $\vec{x} \cdot \vec{r}_i = 0$. Lo que implica que \vec{X} es ortogonal a toda combinación lineal de las filas de A , es decir, a cualquier elemento del espacio fila de A también denotado por $C(A^T)$.

Entonces, si $\vec{c} \in C(A^T)$ y $\vec{x} \in N(A)$:

$$\vec{x} \cdot \vec{c} = 0 \quad \forall \vec{x} \in N(A), \vec{c} \in C(A^T).$$

Que es una forma de expresar que todo elemento de $N(A)$ es ortogonal a cada elemento de $C(A^T)$. Lo anterior implica (usando la Ecuación A.6) que si $\vec{x} \in N(A)$ entonces $\vec{x} \in C(A^T)^\perp$. Es decir:

$$N(A) \subseteq C(A^T)^\perp. \quad (\text{A.7})$$

Caso 2: $C(A^T)^\perp \subseteq N(A)$.

Tomemos los siguientes vectores; $\vec{c} \in C(A^T)$ y $\vec{u} \in C(A^T)^\perp$.

$$\Rightarrow \vec{c} \cdot \vec{u} = 0.$$

En particular los renglones \vec{r}_i con $i = 1, 2, \dots, m$ de A están en $C(A^T)$.

$$\Rightarrow \vec{r}_i \cdot \vec{u} = 0.$$

$$\Rightarrow A\vec{u} = \vec{0}.$$

$$\Rightarrow \vec{u} \in N(A).$$

Por lo que:

$$C(A^T)^\perp \subseteq N(A). \quad (\text{A.8})$$

Y de las Ecuaciones (A.7) y (A.8) se obtiene:

$$N(A) = C(A^T)^\perp. \quad (\text{A.9})$$

que es lo que se quería demostrar ■

Si en el resultado anterior se sustituye la matriz A por su transpuesta se obtiene:

$$N(A^T) = C(A)^\perp. \quad (\text{A.10})$$

A.4. Deducción de la pseudo-inversa

Supongamos que se tiene una matriz de m renglones por n columnas; $A \in \mathbb{M}_{m \times n}$ y se tiene la ecuación:

$$A\vec{x} = \vec{b}, \quad (\text{A.11})$$

con $\vec{x} \in \mathbb{R}^n$ y $\vec{b} \in \mathbb{R}^m$. Supongamos también que el sistema no tiene solución (con la finalidad de encontrar una solución para estos casos). Ahora reescribimos la Ecuación A.11 de la siguiente forma:

$$[\vec{a}_1 \quad \vec{a}_2 \quad \cdots \quad \vec{a}_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \vec{b}, \quad (\text{A.12})$$

en donde las entradas \vec{a}_i son los vectores columna de la matriz A y x_i son las entradas del vector \vec{x} (con $i = 1, 2, \dots, n$). Desarrollando el lado izquierdo de la Ecuación A.12:

$$x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_n\vec{a}_n. \quad (\text{A.13})$$

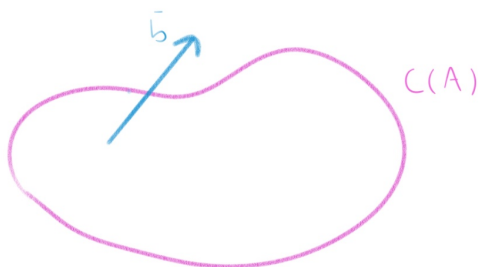
Observamos que al no haber solución, no existen coeficientes (entradas de \vec{x}) que hagan que \vec{b} se pueda escribir como combinación lineal de los vectores columna \vec{a}_i . Este resultado implica que el vector \vec{b} no se encuentra en $C(A)$ el espacio columna de la matriz (ver Figura A.3a).

Debido a que no se tiene solución exacta, lo más que se puede hacer es tratar de buscar una solución aproximada. El propósito es encontrar un vector \vec{x}^* tal que $A\vec{x}^* \approx \vec{b}$.

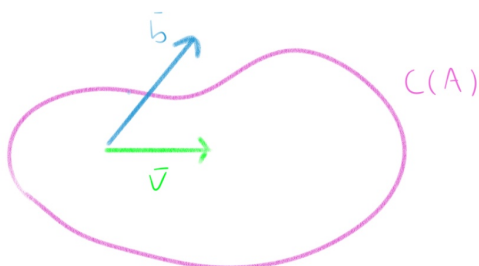
Sea:

$$\vec{v} = A\vec{x}^*. \quad (\text{A.14})$$

Y como debe ser una solución (es una combinación lineal de los vectores columna de A) se tiene que $\vec{v} \in C(A)$ (ver Figura A.3b).



(a) El vector \vec{b} no es combinación lineal de las columnas de A .



(b) El vector \vec{v} es resultado de la solución aproximada y es combinación lineal de las columnas de A .

Figura A.3: Representación gráfica del espacio columna y los vectores de resultados para el sistema sin solución y para la solución aproximada.

Ahora buscamos minimizar la distancia entre los vectores \vec{b} y \vec{v} con el fin de asegurar que la solución aproximada sea la mejor posible. Para hacer esto no queda otra opción más que tomar el vector proyección de \vec{b} sobre el espacio columna $C(A)$ (ver la hipótesis (2) del Teorema 1).

$$\vec{v} = \text{Proy}_{C(A)} \vec{b}. \quad (\text{A.15})$$

De esta forma se encuentra la solución que minimiza el cuadrado de la

norma del vector resta:

$$\left\| \begin{bmatrix} b_1 - v_1 \\ b_2 - v_2 \\ \vdots \\ b_m - v_m \end{bmatrix} \right\|^2 = (b_1 - v_1)^2 + (b_2 - v_2)^2 + \cdots + (b_m - v_m)^2, \quad (\text{A.16})$$

por lo que a \vec{x}^* se le conoce como solución por mínimos cuadrados.

De las Ecuaciones A.14 y A.15 se obtiene que:

$$A\vec{x}^* = \text{Proy}_{C(A)}\vec{b},$$

y restando \vec{b} de ambos lados de la ecuación:

$$\Rightarrow A\vec{x}^* - \vec{b} = \text{Proy}_{C(A)}\vec{b} - \vec{b}.$$

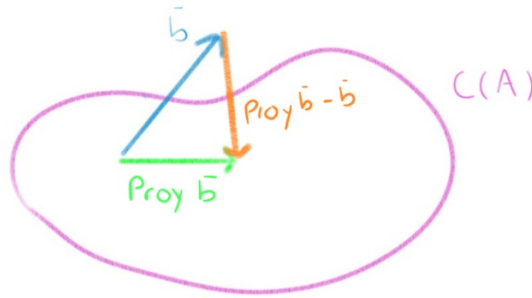


Figura A.4: El vector diferencia entre la proyección de \vec{b} y \vec{b} es ortogonal al espacio columna (ver la hipótesis (1) del Teorema 1).

Por construcción, el vector $\text{Proy}_{C(A)}\vec{b}$ es ortogonal al espacio columna $C(A)$ (ver Figura A.4), por lo que $A\vec{x}^* - \vec{b}$ se encuentra en el espacio complemento ortogonal $C(A)^\perp$ del espacio columna.

$$\Rightarrow A\vec{x}^* - \vec{b} \in C(A)^\perp.$$

Pero sabemos que $C(A)^\perp = N(A^T)$ (ver Ecuación A.10).

$$\Rightarrow A\vec{x}^* - \vec{b} \in N(A^T).$$

Algebraicamente, que un vector esté en el espacio nulo de una matriz (ver Ecuación A.5) tiene la siguiente implicación:

$$\begin{aligned} \Rightarrow A^T(A\vec{x}^* - \vec{b}) &= \vec{0}. \\ \Rightarrow A^T A\vec{x}^* - A^T \vec{b} &= \vec{0}. \\ \Rightarrow A^T A\vec{x}^* &= A^T \vec{b}. \end{aligned}$$

Y cuando las columnas de A son linealmente independientes ($A^T A$) es invertible:

$$\Rightarrow (A^T A)^{-1} A^T A\vec{x}^* = (A^T A)^{-1} A^T \vec{b}.$$

Por lo tanto se obtiene:

$$\therefore \vec{x}^* = (A^T A)^{-1} A^T \vec{b}, \tag{A.17}$$

de donde se define la matriz pseudo-inversa de A como:

$$A^\dagger = (A^T A)^{-1} A^T. \tag{A.18}$$

Apéndice B

Selección de bins

B.1. Errores en la selección de bins

Al momento de trabajar con las matrices e histogramas necesarios para ejecutar el código de *unfolding* se pudieron distinguir algunos errores que no son descritos por los problemas comunes para el código mencionados en la Sección 4.3.1.

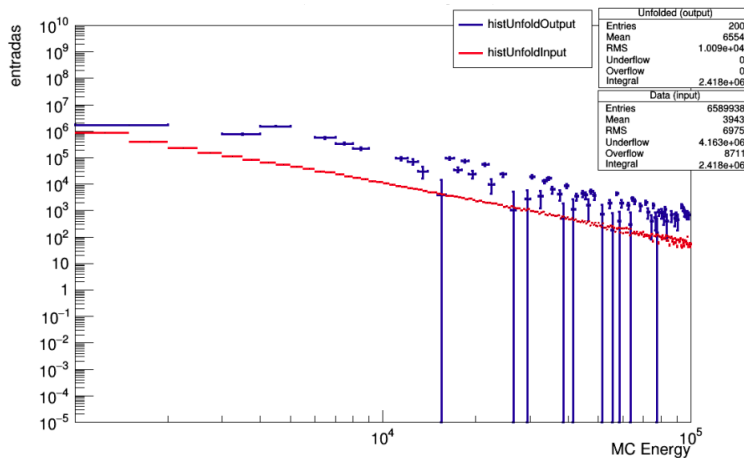


Figura B.1: Primeras pruebas de *unfolding*. En color rojo se muestra el histograma de datos reales y en azul el histograma resultado del *unfolding*.

La Figura B.1 muestra una de las primeras pruebas de *unfolding* que se hicieron con datos reales y matrices provenientes de simulaciones. Lo importante de resaltar con este histograma son dos aspectos que se pueden observar a simple vista: espacios vacíos en algunos *bines* en donde tendría que haber entradas para el histograma que resulta del *unfolding* y barras de error (barras verticales) demasiado grandes en algunos *bines*, comparables con el número de entradas en esos *bines*.

Al realizar un análisis rápido se obtuvo que, más que estar vacíos, aquellos *bines* que aparentemente no registran entradas en realidad presentan entradas negativas. Las barras de error para algunos *bines* son tan grandes que cubren valores negativos en las entradas, al presentarse el histograma en forma logarítmica estas barras parecen infinitas. Después de algunas pruebas se descartó que estos problemas sean debidos a la estadística en el número de eventos (cantidad de archivos utilizados) tanto de los datos reales como de simulaciones. Estos problemas parecen disminuir o corregirse cuando se ocupa otra selección de *bines*. De esta forma se llegó a la selección de *bines* mostrada en la Sección 4.4 con el nombre esquema1.

Apéndice C

Comparación de matrices

C.1. Cambio en la estadística de una matriz debido al estimador utilizado y al corte extra

Como se mencionó en la Sección 4.8 las principales diferencias que se pueden encontrar entre las matrices (para cada caso angular, en este ejemplo se utiliza el caso angular θ_0 , e decir, con corte a 16.8°) son debidas al estimador utilizado y al corte extra en el caso del estimador *Grpound Parameter*.

Inicialmente se tiene la misma cantidad de eventos con energía verdadera predefinida, luego, a pesar de que se apliquen los cortes generales sobre ambos estimadores (ver Sección 3.5.4), la energía reconstruida con cada uno es diferente y por lo tanto, *bin* por *bin*, la cantidad de eventos registrados cambia dando lugar a diferentes estadísticas para cada estimador.

Las figuras C.1 y C.2 muestran visualmente este cambio en estadística debido al estimador utilizado sin usar el corte extra para *Ground Parameter*. En ambas matrices se observa una tendencia simétrica respecto a la identidad, sin embargo se registran bastantes eventos fuera de la diagonal, lo que indica la presencia de un sesgo entre las energías verdaderas y reconstruidas.

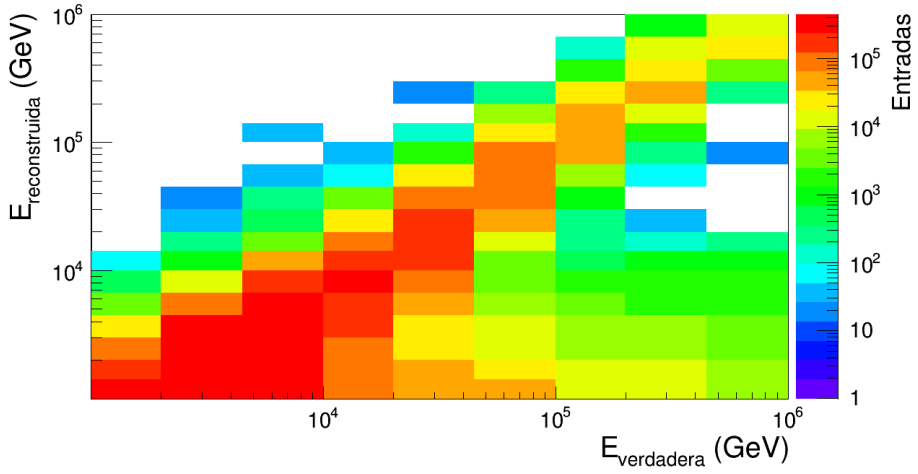


Figura C.1: Matriz de respuesta para el caso $NN \theta_0$.

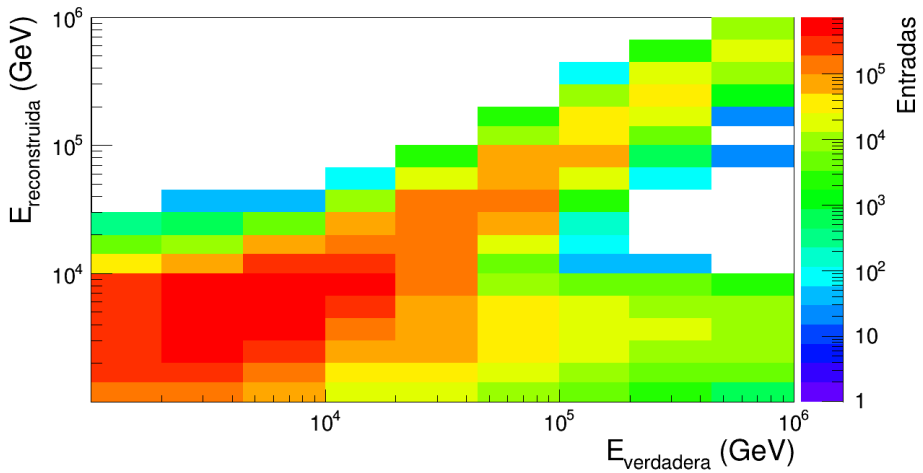


Figura C.2: Matriz de respuesta para el caso $GP \theta_0$. No se ha aplicado el corte extra.

Por otra parte, al usar el corte extra sobre el estimador *Ground Parameter* se tiene una reducción abrupta en el número de eventos registrados (ver la Sección 3.5.5), lo cual hace que la matriz cambie considerablemente. La Figura C.3 muestra dicho cambio.

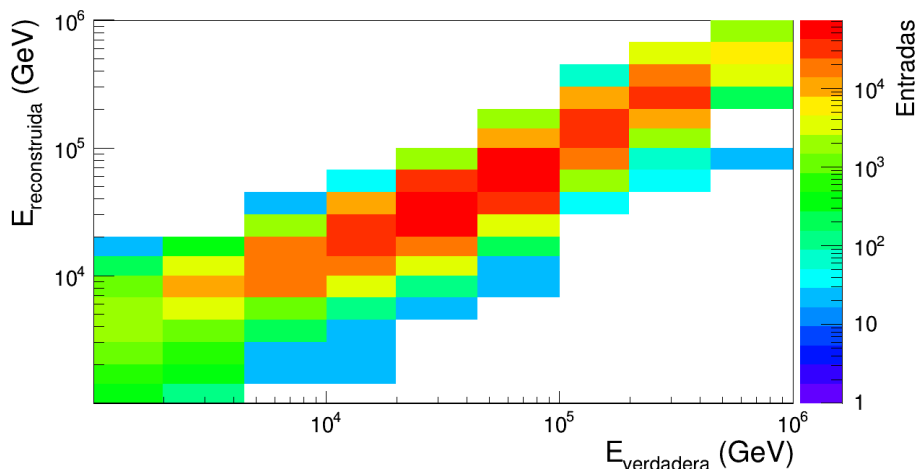


Figura C.3: Matriz de respuesta para el caso $GP \theta_0$. Se ha aplicado el corte extra.

La matriz obtenida para el caso *Ground Parameter* con el corte extra aplicado (Figura C.3) muestra una tendencia más clara sobre la identidad, es decir, presenta un menos sesgo entre la energía verdadera y la reconstruida.

En la Tabla C.1 se muestra la cantidad de eventos registrados en la matrices mostradas.

Estimador	Cortes	Entradas
$NN \theta_0$	Generales	7.20×10^6
$GP \theta_0$	Generales	1.04×10^7
$GP \theta_0$	Generales y extra	6.77×10^5

Tabla C.1: Número de eventos simulados utilizados en las matrices para el caso con corte cenital a 16.8° .

Se puede observar que la diferencia debida a los estimadores es, aproximadamente, 1.44 veces más la cantidad de eventos para *Ground Parameter* que para *Neural Network*. Después de aplicar el corte extra, el estimador *Ground Parameter* disminuye la cantidad de eventos aproximadamente 15.36 veces.

Apéndice D

Propagación de errores

D.1. Cálculo del error asociado a la curva de área efectiva

Las curvas de área efectiva han sido ajustadas por la función dada en la Ecuación 4.8:

$$A = \exp [p_0 - p_1 E^{-c}] . \quad (\text{D.1})$$

Por simplicidad se ha reescrito dicha ecuación de área efectiva llamándola A , en donde E es la energía y c una constante. Las variables que tienen un error asociado, y por lo tanto sobre las que se hará la propagación de errores, son p_0 y p_1 . Los valores obtenidos para estos parámetros se mostraron en la Tabla 4.4, de la cual resulta llamativo observar que el parámetro p_1 presenta errores asociados demasiado grandes, comparables con el valor mismo de p_1 . Es importante recordar que en este punto ya se ha calculado el mejor ajuste de las curvas de área efectiva y que dichos ajustes, dados por la Ecuación D.1, ya han sido evaluados en el centro de cada *bin* (con la finalidad de encontrar un valor de área efectiva para cada *bin*) y es precisamente de esta evaluación de la que se quiere calcular el error asociado. De lo anterior se quiere hacer notar que, mientras en la Ecuación 4.8 (al buscar el mejor ajuste) la energía E es la variable independiente definida en un rango de energía dado, en la Ecuación D.1 (una vez ya encontrado el mejor ajuste) la variable E representa la energía en el centro de cada *bin*.

Se utilizarán las siguientes ecuaciones de propagación de errores:

■ **Multipliación por una constante.**

Si se tiene una variable x con error asociado Δx y k una constante, y sea f una función definida como $f = kx$, entonces:

$$f = k(x \pm \Delta x) = kx \pm k\Delta x. \quad (\text{D.2})$$

Es decir, $\Delta f = k\Delta x$ es el error asociado a la función f .

■ **Resta de dos medidas**

Si se tienen las variables x y y con errores asociados Δx y Δy respectivamente, y sea f una función definida como $f = x - y$, entonces:

$$f = (x \pm \Delta x) - (y \pm \Delta y) = (x - y) \pm \sqrt{(\Delta x)^2 + (\Delta y)^2}. \quad (\text{D.3})$$

Es decir, $\Delta f = \sqrt{(\Delta x)^2 + (\Delta y)^2}$ es el error asociado a la función f .

■ **Exponencial**

Si se tiene una variable x con error asociado Δx , y sea f una función definida como $f = e^x$, entonces:

$$f = e^{(x \pm \Delta x)} = e^x \pm e^x \Delta x. \quad (\text{D.4})$$

Es decir, $\Delta f = f\Delta x$ es el error asociado a la función f .

De esta forma, se irá calculado el error asociado a cada componente de la Ecuación D.1 para poder calcular el error asociado total ΔA .

Primero, sea $f_1 = p_1 E^{-c}$ y utilizando la Ecuación D.2 para la multiplicación por una constante (en este caso E^{-c} es tratado como constante) se obtiene el siguiente resultado:

$$\Delta f_1 = \Delta p_1 E^{-c}.$$

Luego, sea $f_2 = p_0 - p_1 E^{-c} = p_0 - f_1$ y utilizando la Ecuación D.3 para la resta de dos medidas se obtiene que $\Delta f_2 = \sqrt{(\Delta p_0)^2 + (\Delta f_1)^2}$ y sustituyendo el valor encontrado para Δf_1 se llega a:

$$\Delta f_2 = \sqrt{(\Delta p_0)^2 + (\Delta p_1)^2 E^{-2c}}.$$

Por último, $A = \exp [p_0 - p_1 E^{-c}] = e^{f_2}$ y utilizando la Ecuación D.4 para la exponencial se obtiene que $\Delta A = A \Delta f_2$ y sustituyendo los valores de F y f_2 se llega a:

$$\Delta A = \exp [p_0 - p_1 E^{-c}] \sqrt{(\Delta p_0)^2 + (\Delta p_1)^2 E^{-2c}}.$$

Por la forma en que se implementó en un código del software *Root*, la he reescrito de la siguiente forma:

$$\Delta A = A \sqrt{(\Delta p_0)^2 + (\Delta p_1)^2 \left(\frac{1}{E^c}\right)^2}. \tag{D.5}$$

D.1.1. Evaluación del área efectiva y sus errores asociados en el centro de cada *bin* de energía

A continuación se muestra el resultado de evaluar la Ecuación D.1 de área efectiva en el centro de cada *bin* del esquema2 de *bines* (ver Tabla 4.2), así como el error asociado a cada *bin* usando la Ecuación D.5.

Bin	Área efectiva (m^2)	Error propagado (m^2)	Error porcentual
1	4.22058	18.6148	441.049
2	1788.54	2119.41	118.5
3	9414.2	2839.04	30.157
4	14022.6	1348.1	9.61374
5	15778.1	769.844	4.8792
6	16297.4	706.895	4.33747
7	16424.5	706.502	4.30152
8	16462.3	707.63	4.29848
9	16472.7	708.04	4.29826

Tabla D.1: Resultado de la evaluación de área efectiva (A) en el centro de cada *bin* y error propagado (ΔA). Se muestran los resultado para la curva de área efectiva del 2018.

En la Tabla D.1 se presentan los valores obtenidos para la curva de área efectiva del 2018 mostrada en la Figura 4.20, ya que es esta curva la utilizada

para todos los casos, tanto angulares como para ambos estimadores. Se puede observar que a pesar de que el ajuste por sí mismo tiene errores muy grandes asociados al parámetro del ajuste p_1 (ver Tabla 4.4), el error propagado no resulta demasiado grande en los *bines* del 4 al 8 (10 *TeV* a 447 *TeV*), en los que se lleva a cabo el análisis con *TUnfold*. Además se debe recordar que esta propagación de error solo es para la utilización de los valores de área efectiva dentro de la normalización de la distribución de energía. Aún falta realizar la propagación de error de dicha normalización.

D.2. Cálculo del error asociado al flujo diferencial de rayos cósmicos

Para el cálculo del flujo diferencial de rayos cósmico se utiliza, *bin* por *bin*, la Ecuación 4.4 la cual se reescribirá, por simplicidad, de la siguiente forma:

$$F = \frac{N}{T \Omega A E}. \quad (\text{D.6})$$

En dónde el tiempo T , el ángulo sólido Ω y la energía que cubre el ancho de *bin* E son valores que, en este trabajo, no tiene un error asociado, por lo que se tratarán como constantes. El número de entradas N en la distribución *unfolded* con error asociado ΔN y el área efectiva A con error asociado ΔA .

Como se vio en con la Ecuación D.2 para la multiplicación de una constante por una variable medida, el error asociado es la misma contante por el error asociado a dicha medición, por lo que por el momento solo se hará uso de la siguiente ecuación de propagación de error:

■ División de dos medidas

Si se tienen las variables x y y con errores asociados Δx y Δy respectivamente, y sea f una función definida como $f = \frac{x}{y}$, entonces:

$$f = \frac{(x \pm \Delta x)}{(y \pm \Delta y)} = \left(\frac{x}{y}\right) \pm \left(\frac{x}{y}\right) \sqrt{\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2}. \quad (\text{D.7})$$

Es decir, $\Delta f = \left(\frac{x}{y}\right) \sqrt{\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2}$ es el error asociado a la función f .

Utilizando lo anterior, el error ΔF asociado al flujo diferencial de rayos cósmicos es:

$$\Delta F = \frac{1}{T \Omega E} \frac{N}{A} \sqrt{\left(\frac{\Delta N}{N}\right)^2 + \left(\frac{\Delta A}{A}\right)^2}.$$

Por la forma en que se usa en el software *Root*, la ecuación anterior se expresará como:

$$\Delta F = F \sqrt{\left(\frac{\Delta N}{N}\right)^2 + \left(\frac{\Delta A}{A}\right)^2}. \quad (\text{D.8})$$

Bibliografía

- [1] J. Díaz. *Mensajeros cósmicos I: rayos cósmicos*. Recuperado en noviembre 2018. <https://conexioncasual.wordpress.com/2015/12/12/mensajeros-cosmicos-i-rayos-cosmicos/>.
- [2] M. Aguilar. Cien años de rayos cósmicos. *Rev.R.Acad.Cienc.Exact.Fís.Nat (Esp)*, 105(1):107–128, 2011.
- [3] Victor Hess. On the Observations of the Penetrating Radiation during Seven Balloon Flights. physics.hist-ph, 1808.02927, 2018.
- [4] Mario Bertolotti. *The Discovery: Victor F. Hess and the Balloon Ascents*, pages 33–44. Celestial Messengers. Astronomers Universe. Springer, Berlin, Heidelberg. 2013.
- [5] Jörg R. Hörandel. Early Cosmic-Ray Work Published in German. *AIP Conf. Proc.*, 1516(1):52–60, 2013.
- [6] Alessandra Abe Pacini. Cosmic rays: bringing messages from the sky to the Earth’s surface. *Revista Brasileira de Ensino de Física*, 39, 00 2017.
- [7] R. Alfaro et al. All-particle cosmic ray energy spectrum measured by the HAWC experiment from 10 to 500 TeV. *Phys. Rev.*, D96(12):122001, 2017.
- [8] IceCube MasterClass an authentic astrophysics research experience. *El Espectro de Energía de los Rayos Cósmicos*. https://masterclass.icecube.wisc.edu/es/analisis/espectro_rc.

- [9] A. D. Erlykin and A. W. Wolfendale. Cosmic ray ions, dust and the interstellar medium. *Physica Scripta*, 63:504, 03 2006.
- [10] Pasquale Blasi. The Origin of Galactic Cosmic Rays. *Astron. Astrophys. Rev.*, 21:70, 2013.
- [11] Alberto Carramiñana. *Detectors for γ - ray astronomy*. <https://indico.nucleares.unam.mx/event/1495/contribution/19/material/slides/0.pdf>.
- [12] Z. Hampel-Arias. *Cosmic Ray Observations at the TeV Scale with the HAWC Observatory*. PhD thesis, The University of Wisconsin - Madison, August 2017.
- [13] Jochem D. Haverhoek. Ultra high energy cosmic ray extensive air shower simulations using corsika. University of Leiden, April University of Leiden, 2016.
- [14] Karl-Heinz Kampert and Alan A. Watson. Extensive air showers and ultra high-energy cosmic rays: a historical review. *The European Physical Journal H*, 37(3):359–412, Aug 2012.
- [15] V.A Kudryavtsev. Muon pair production by photons in atmosphere: is any room left for high-energy muon astronomy? *Astroparticle Physics*, 18(1):97 – 105, 2002.
- [16] W. Heitler. *The quantum theory of radiation*, volume 5 of *International Series of Monographs on Physics*. Oxford University Press, Oxford, 1936.
- [17] Sonali Bhatnagar. In *Extensive Air Shower High Energy Cosmic Rays (II)*, Dayalbagh Educational Institute, Agra, 2009.
- [18] M. Tanabashi et al. (Particle Data Group). Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018.
- [19] J. C. Moreno. *Estudio de rayos cósmicos en el Observatorio Pierre Auger: nuevo modelo de profundidad atmosférica y su influencia en observables*. PhD thesis, Universidad Nacional de La Plata, Abril 2012.

- [20] Diego Garcia Pinto. Measurements of the longitudinal development of air showers with the pierre auger observatory. *Proceedings of the 32nd International Cosmic Ray Conference, ICRC 2011*, 2:87–, 01 2011.
- [21] H. Ayala. Core fitter analysis. 2014. Documento interno de HAWC.
- [22] K. Malone. Documentation for aerie module "latdist". 2015. Documento interno de HAWC.
- [23] Juan Carlos Arteaga-Velázquez, Z. Hampel-Arias, and J. D. Álvarez. Estimate of the energy spectrum of the light component of cosmic rays in HAWC using the shower age and the fraction of hit PMT's. *PoS, ICRC2017:501*, 2018.
- [24] J. A. Morales Soto y J. C. Arteaga Velázquez. *Estudio de distribución lateral de cascadas de partículas inducidas por rayos cósmicos usando HAWC*. <https://indico.nucleares.unam.mx/event/1495/contribution/6/material/slides/0.pdf>.
- [25] A. U. Abeysekara et al. Measurement of the Crab Nebula at the Highest Energies with HAWC. 2019.
- [26] A. U. Abeysekara et al. Observation of the crab nebula with the HAWC gamma-ray observatory. *The Astrophysical Journal*, 843(1):39, jun 2017.
- [27] K. Malone. Energy estimation via ground parameter. 2017. Documento interno de HAWC.
- [28] Koichi Kamata and Jun Nishimura. The Lateral and the Angular Structure Functions of Electron Showers. *Progress of Theoretical Physics Supplement*, 6:93–155, 02 1958.
- [29] Jun Nishimura and Koichi Kamata. On the Theory of Cascade Showers, I. *Progress of Theoretical Physics*, 7(2):185–192, 02 1952.
- [30] Bruno Rossi and Kenneth Greisen. Cosmic-ray theory. *Rev. Mod. Phys.*, 13:240–309, Oct 1941.

- [31] H. Bethe and W. Heitler. On the Stopping of fast particles and on the creation of positive electrons. *Proc. Roy. Soc. Lond.*, A146:83–112, 1934.
- [32] G. Puppi, H.S. Bridge, and K. Greisen. *Progress in Cosmic Ray Physics. Vol. 3. Edited by J.G. Wilson, ... Contributors: K. Greisen, H.S. Bridge, R.W. Thompson, G. Puppi.* North-Holland Publishing C°, 1956.
- [33] Peter Grieder. *Extensive air showers. High energy phenomena and astrophysical aspects. A tutorial, reference manual and data book. Vol. I and II*, volume 1. 01 2010.
- [34] David Newton, J. Knapp, and A. A. Watson. The Optimum Distance at which to Determine the Size of a Giant Air Shower. *Astropart. Phys.*, 26:414–419, 2007.
- [35] Todor Stanev. *High Energy Cosmic Rays.* Springer Praxis Books. ISBN 978-3-540-85147-9. Springer-Verlag Berlin Heidelberg, 2010.
- [36] S. Marinelli. Neural-net energy reconstruction. 2016. Documento interno de HAWC.
- [37] Thomas K. Gaisser, Ralph Engel, and Elisa Resconi. *Electromagnetic cascades*, page 302–312. Cambridge University Press, 2 edition, 2016.
- [38] Ahron S. Barber, David B. Kieda, and R. Wayne Springer. Simulation of Near Horizontal Muons and Muon Bundles for the HAWC Observatory with CORSIKA. *PoS, ICRC2017:511*, 2018. [35,511(2017)].
- [39] A. J. Smith and HAWC Collaboration. HAWC: Design, Operation, Reconstruction and Analysis. In *34th International Cosmic Ray Conference (ICRC2015)*, volume 34 of *International Cosmic Ray Conference*, page 966, Jul 2015.
- [40] O. Heaviside. *Electrical Papers Vol. II.* AMS Chelsea Publishing Series. American Mathematical Society, 2003.
- [41] A. G. Andrades. *Aplicaciones de la Radiación Cherenkov*, Trabajo Fin de Grado Inédita, Universidad de Sevilla, 2017.

- [42] J. V. Jelley. Cerenkov radiation and its applications. *British Journal of Applied Physics*, 6(7):227–232, jul 1955.
- [43] . https://en.wikipedia.org/wiki/Cherenkov_radiation#/media/File:Cherenkov.svg.
- [44] *Large Photocathode Area Photomultiplier Tubes*. https://www.hamamatsu.com/resources/pdf/etd/LARGE_AREA_PMT_TPMH1376E.pdf.
- [45] A. U. Abeysekara et al. On the sensitivity of the HAWC observatory to gamma-ray bursts. *Astroparticle Physics*, 35(10):641–650, May 2012.
- [46] A.U. Abeysekara et al. Data acquisition architecture and online processing system for the hawc gamma-ray observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 888:138 – 146, 2018.
- [47] *AERIE Offline/Online Software*. <http://hawclava.umd.edu/docs/aerie-docs/aerie-trunk-docs/>.
- [48] B. Baughman and A. Smith. Requirements and baseline design wbs 5.4 - detector simulation. 2010. Documento interno de HAWC.
- [49] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw. CORSIKA: A Monte Carlo code to simulate extensive air showers. Forschungszentrum und Universität Karlsruhe, 1998.
- [50] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003.
- [51] J. Pretz. Daqsim - code and parameter tuning. 2018. Documento interno de HAWC.
- [52] J. Braun and A. Smith. Requirements and baseline design wbs 5.2 - reconstruction. 2010. Documento interno de HAWC.
- [53] J. Pretz, S. Benzvi, and Z. Ren. Simulation weighting. 2016. Documento interno de HAWC.

- [54] B T Allen, G B Yodh, and the Milagro Collaboration. Measurement of the crab spectrum with milagro. *Journal of Physics: Conference Series*, 60:321–323, mar 2007.
- [55] J. Pretz. High energy paper pretz optimization, details of john optimization of cuts. 2017. Documento interno de HAWC.
- [56] D. Heck and T. Pierog. Extensive Air Shower Simulation with CORSIKA: A User’s Guide. 2019.
- [57] E. Lorenz and Wagner. Very-high energy gamma-ray astronomy. *R. EPJ H*, 37:459–513, 2012.
- [58] Stefan Schmitt. Data Unfolding Methods in High Energy Physics. *EPJ Web Conf.*, 137:11008, 2017.
- [59] Volker Blobel. Unfolding Methods in Particle Physics. pages 240–251. 12 p, Jan 2011. <https://cds.cern.ch/record/2203257>.
- [60] *Application of the Method of Least Squares to the Elements of the Planet Pallas*. <https://www.york.ac.uk/depts/maths/histstat/gausspallas.pdf>.
- [61] Stefan Schmitt. TUnfold: an algorithm for correcting migration effects in high energy physics. *JINST*, 7:T10003, 2012.
- [62] *TUnfoldV17 Class Reference*. http://www.desy.de/~sschmitt/TUnfold/doxygen_17.6/html/classTUnfoldV17.html#_details.
- [63] Schmitt, Stefan. The tunfold package: user manual. 2019. http://www.desy.de/~sschmitt/TUnfold/tunfold_manual_v17.9.pdf.
- [64] *class TUnfold: public TObject*. <https://root.cern.ch/root/html530/TUnfold.html>.
- [65] *Common problems when using TUnfold*. <http://www.desy.de/~sschmitt/tunfoldv16docu.html>.
- [66] *TUnfold tutorials*. https://root.cern.ch/doc/v606/group__tutorial__unfold.html.

- [67] *testUnfold3.C File Reference*. https://root.cern.ch/doc/v606/testUnfold3_8C.html.
- [68] Kathryn Sparks. HAWC effective area update. 2011. Documento interno de HAWC.
- [69] J.C. Arteaga-Velázquez and J. D. Álvarez. Update on the spectrum of the light component of cosmic rays measured with HAWC. 2018. Documento interno de HAWC.
- [70] *Deep Learning Book Series 2.9, The Moore Penrose Pseudoinverse*. <https://hadrienj.github.io/posts/Deep-Learning-Book-Series-2.9-The-Moore-Penrose-Pseudoinverse/>.
- [71] C. Arce, W. Castillo, and J. González. *Álgebra lineal*. Universidad de Costa Rica, Escuela de Matemática, Tercera Edición, 2012.