FACULTAD DE CIENCIAS FÍSICAS

MÁSTER EN ASTROFÍSICA



Búsqueda de subhalos de materia oscura en la banda de rayos gamma de alta energía mediante aprendizaje automático

Por:

Juan Bernete Medrano

Tutor:

Daniel Nieto Castaño

TRABAJO FIN DE MÁSTER

Curso académico 2020-21

Madrid, 17 de junio de 2021

Búsqueda de subhalos de materia oscura en la banda de rayos gamma de alta energía mediante aprendizaje automático

Resumen:

Bajo la hipótesis de que la materia oscura está formada por partículas masivas con interacción débil (WIMPs) con masa en el entorno del GeV - TeV capaces de autoaniquilarse, un subproducto de esta aniquilación serían rayos gamma de alta energía. Las sobredensidades locales de estas partículas dentro de nuestra galaxia predichas por simulaciones de N-cuerpos (subhalos) podrían ser detectadas por el instrumento *Fermi*-LAT. El último catálogo de *Fermi* (4FGL-DR2) incluye un gran número de fuentes no asociadas, entre las que podrían encontrarse estas emisiones. Para clasificar estas fuentes no asociadas, una estrategia es recurrir al aprendizaje automático y redes neuronales.

En este trabajo se crea el código necesario para generar grandes catálogos de fuentes simuladas de materia oscura, útiles para entrenar a las redes neuronales de clasificación. Se trata de un código que utiliza las herramientas de *Fermi* (*fermitools*) con este propósito. El código automatiza la producción de observaciones simuladas de estas fuentes en 216 posiciones en el cielo y su análisis para ajustar el espectro observado a una distribución Log-Parábola. Todo el proceso se puede ejecutar para diferentes masas de las WIMPs, canales de aniquilación y flujos.

Con el objeto de verificar el correcto funcionamiento del código, esta técnica se prueba para una masa de WIMP de 500 GeV, canal de aniquilación $b\bar{b}$ y flujo de $10^{-4}m^{-2}s^{-1}MeV^{-1}$, obteniendo buenos resultados, pero con margen de mejora.

Abstract:

Under the hypothesis of dark matter formed by weakly interacting massive particles (WIMPs) with a mass of order GeV - TeV capable of self-annihilation, high-energy gamma rays would be a subproduct of such annihilation. Local overdensities of these particles inside our galaxy predicted by N-body simulations (subhalos) could be detected by the *Fermi*-LAT instrument. The last *Fermi* catalogue (4FGL-DR2) includes a large number of unassociated sources, among which these emissions could be found. To classify these unassociated sources, one strategy is to use machine learning and neural networks.

In this work, the necessary code to generate large catalogs of simulated dark matter sources, useful for training classification neural networks, is created. It is a code that uses the *Fermi* tools (*fermitools*) for this purpose. This code automates the production of simulated observations of these sources in 216 positions in the sky and its analysis to fit the observed spectrum to a Log-Parabola distribution. All of this process can be executed for different WIMPs masses, annihilation channels and fluxes.

This technique is tested for a WIMP mass of 500 GeV, annihilation channel $b\bar{b}$ and flux of $10^{-4}m^{-2}s^{-1}MeV^{-1}$, obtaining good results, but with room for improvement.

Índice

Página

1. Introducción y objetivo. **2**

2. Telescopio *Fermi*-LAT y herramientas de trabajo. **4**

3. Simulación de observaciones de DM. Automatización de las simulaciones. **6**

4. Análisis de las simulaciones para la creación de catálogos sintéticos. **10**

5. Conclusiones y posibles continuaciones del estudio. **16**

Referencias **19**

---

1. Introducción y objetivo.

La materia oscura es uno de los principales componentes del universo. Constituye aproximadamente un 26 % de la densidad de materia-energía del universo, tal y como se obtiene de los datos del satélite Planck [1]. Su existencia es esencial para la comprensión de la formación, evolución y dinámica de las galaxias y de la estructura del universo a gran escala debido a sus grandes efectos gravitatorios, asumiendo el modelo cosmológico ΛCDM. Numerosas evidencias como las curvas de rotación de las galaxias o las lentes gravitacionales desvelan la presencia de una masa mucho mayor que la de los objetos visibles en una galaxia. La materia oscura sería este exceso de masa invisible [2].

En cuanto a su composición, el modelo estándar de la física de partículas no presenta ningún candidato para este tipo de materia. Se desconoce prácticamente todo sobre su naturaleza, más allá de sus efectos gravitatorios y el hecho de que no emite ni absorbe radiación electromagnética, y es por esto por lo que hay una gran variedad de hipótesis que están siendo exploradas para dar una explicación a este exceso de masa invisible en las galaxias y clústeres de galaxias. La búsqueda de candidatos para la materia oscura es uno de los mayores problemas actuales en la física fundamental. Algunos ejemplos de estos candidatos son: agujeros negros primordiales, modificaciones en la dinámica newtoniana (MONDs, Modified Newton's Dynamics) que tratan de evitar que la materia oscura sea necesaria o partículas aún no detectadas como neutrinos estériles, axiones o WIMPs (Weakly Interactive Massive Particles, partículas masivas con interacción débil) naturales o no naturales [3]. Las WIMPs naturales son partículas que derivan directamente de modelos supersimétricos de física de partículas y tienen características concretas. Estas últimas son los llamados neutralinos supersimétricos y durante mucho tiempo han parecido el mejor candidato. Sin embargo, la falta de evidencias en estas teorías de supersimetrías lleva a pensar que esta no es más probable que el resto de las hipótesis. Las WIMPs no naturales son una clase mucho más general de partículas al no derivar de supersimetrías y su masa puede tomar un amplio rango de valores desde 1 MeV hasta 100 TeV [4]. En cualquiera de los casos, el descubrimiento de una partícula o fenómeno responsable de la materia oscura deberá implicar modificaciones o ampliaciones en el modelo estándar actual.

En cuanto a su distribución, los efectos gravitatorios de la materia oscura proporcionan indicios de que esta se agrupa en grandes halos esféricos que rodean a las galaxias, con el máximo de concentración en el centro, y subhalos de mayor concentración dentro del halo principal de las galaxias [5]. Esta distribución jerárquica de la materia oscura, conocida gracias a las simulaciones de N-cuerpos, se representa en la Figura 1, donde se observa que a las escalas más pequeñas de la simulación hay subestructuras con sobredensidades de materia oscura, a las que llamamos subhalos y que pueden no contener materia bariónica.

En este trabajo se parte de la hipótesis de que las partículas de materia oscura son WIMPs de masas muy elevadas, del orden desde los GeV hasta los TeV. Además, se supone que estas partículas pueden aniquilarse por pares, produciendo productos del modelo estándar diferentes a fotones, que a su vez pueden generar fotones (por la desintegración de piones, en el caso de que se aniquilen a quarks, o por curvatura o Bremsstrahlung, en el caso de leptones). Véase Figura 2. De esta forma, la mayor parte de los fotones generados en estos procesos serían rayos gamma de alta energía, detectables desde la Tierra usando telescopios Cherenkov o el telescopio espacial *Fermi*-LAT (Large Area Telescope) de la NASA, apuntando a subhalos de materia oscura dentro de la Vía Láctea [6], al centro y halo galáctico, donde la concentración de materia oscura es muy elevada, a clústeres de galaxias o a galaxias esferoidales enanas, que son ricas en materia oscura y apenas tienen otros procesos generadores de rayos gamma. Esto es lo que se llamaría una detección indirecta de materia oscura. Hasta ahora no ha habido ninguna detección directa (en aceleradores) ni indirecta de estas

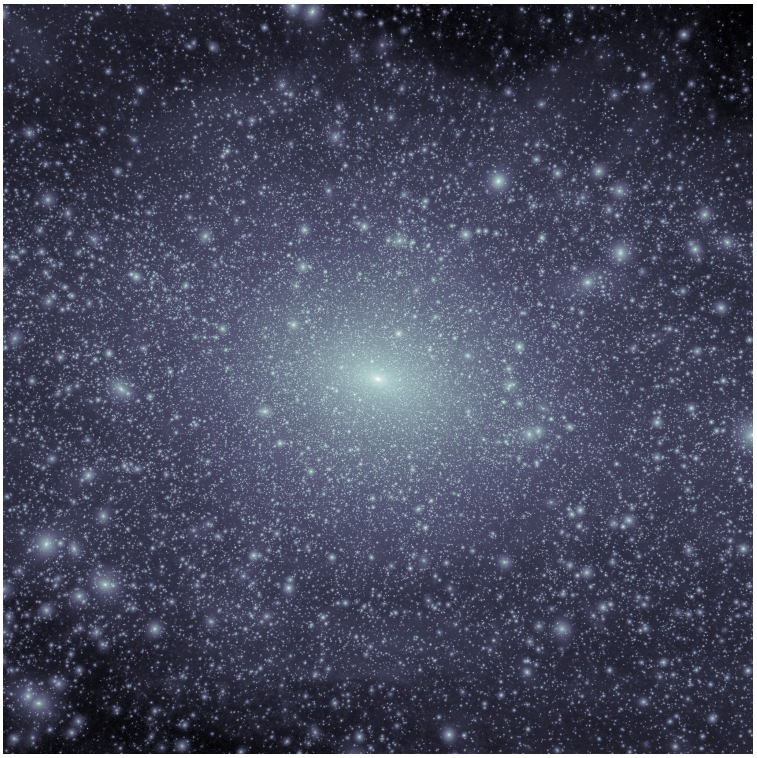


Figura 1: Representación de la distribución local de subhalos de materia oscura por la simulación "Via Lactea II" (cubo de 800kpc). [5].

partículas.

El espectro de radiación gamma emitido por la autoaniquilación de WIMPs en una región del cielo viene dado por la ecuación (1).

$$\frac{d\phi}{dE} = \int_{los} \rho^2(r) dr \frac{\langle \sigma v \rangle}{2\pi m_\chi^2} \Sigma_i Br_i \frac{dNi}{dE} \tag{1}$$

Donde hay un primer factor astrofísico $J = \int_{los} \rho^2(r) dr$ que integra la densidad de partículas de materia oscura al cuadrado $\rho^2(r)$ dentro del cono de observación (l.o.s., siglas de "line of sight"). Este factor se puede ver como una normalización del flujo. Y el otro factor es de física de partículas, donde hay una dependencia en el cuadrado inverso de la masa de la partícula m_χ y una suma de "branching ratios" $Br_i = \langle \sigma v \rangle_i / \langle \sigma v \rangle$ para cada canal de aniquilación i, y donde $\langle \sigma v \rangle$ es la sección eficaz promediada térmicamente y $\frac{dN_i}{dE}$ el rendimiento fotónico. De esta forma, el espectro de una fuente de estas características vendrá determinado por la combinación de estos tres parámetros: masa, canal de aniquilación y factor de normalización del flujo.

El objetivo de este trabajo es tomar parte en la siguiente estrategia para la búsqueda de materia oscura por detección indirecta:

- Predecir el espectro de emisión gamma de la aniquilación de partículas de materia oscura.
- Con dicho espectro simular observaciones de subhalos de materia oscura, utilizando las herramientas de *Fermi* y generar un catálogo de observaciones simuladas suficientemente grande.

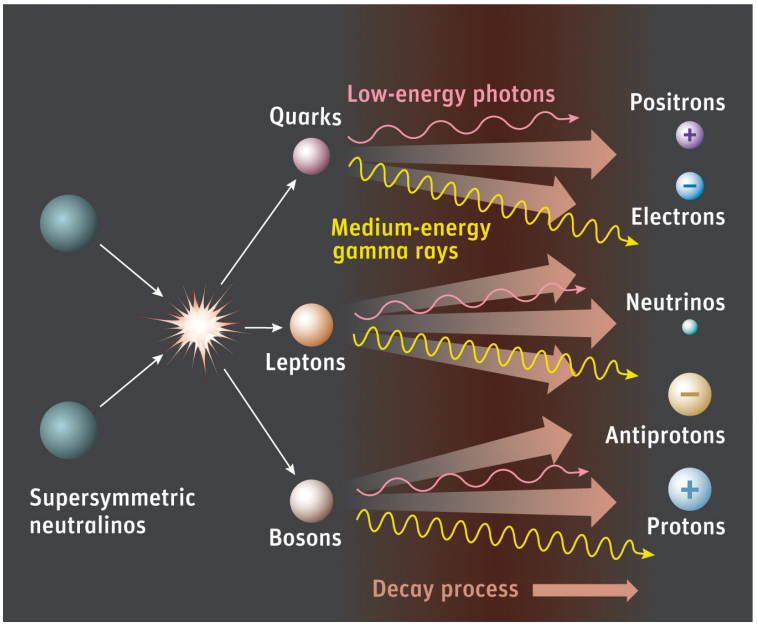


Figura 2: Producción de elementos del Modelo Estándar a partir de la aniquilación de partículas de materia oscura.

- Generar también un catálogo de observaciones simuladas del resto de tipos de fuentes conocidas presentes en el catálogo, como AGNs, púlsares, remanentes de supernovas, etc.
- Entrenar a una red neuronal de Deep Learning con los catálogos generados para poder diferenciar unos procesos de otros.
- Introducir los datos reales del último catálogo de *Fermi* a la red neuronal, una vez haya terminado su aprendizaje, para que categorice las posibles fuentes de materia oscura.
- Estudiar, si las hubiera, las fuentes del catálogo de *Fermi* que la red neuronal señalara como probables fuentes de materia oscura, utilizando telescopios Cherenkov y, si es posible, confirmar si se trata o no de una observación indirecta de materia oscura.

Ya se cuenta con una herramienta para predecir los espectros en función de la masa de la partícula de materia oscura y el canal de aniquilación; de modo que este trabajo se centra en el segundo punto, es decir, generar un catálogo de observaciones simuladas de materia oscura, como paso previo necesario para avanzar en el proceso de búsqueda.

Para crear este catálogo, se simula una gran cantidad de fuentes a partir de su espectro utilizando las herramientas del telescopio *Fermi*-LAT y posteriormente se analizan estas simulaciones como si se tratara de datos reales. De esta forma, se puede entender cómo responde el telescopio a este tipo de señales y reducir esta información a un ajuste de parámetros espectrales, que es precisamente la información que aporta el último catálogo del telescopio respecto a sus fuentes no asociadas y con lo que, en un trabajo futuro, la red neuronal trabajaría para diferenciar unas fuentes de otras.

2. Telescopio *Fermi*-LAT y herramientas de trabajo.

El *Fermi* Large Area Telescope [7] es el principal instrumento a bordo del telescopio espacial de rayos gamma *Fermi* [8], lanzado en 2008. Este instrumento es un detector de rayos gamma por

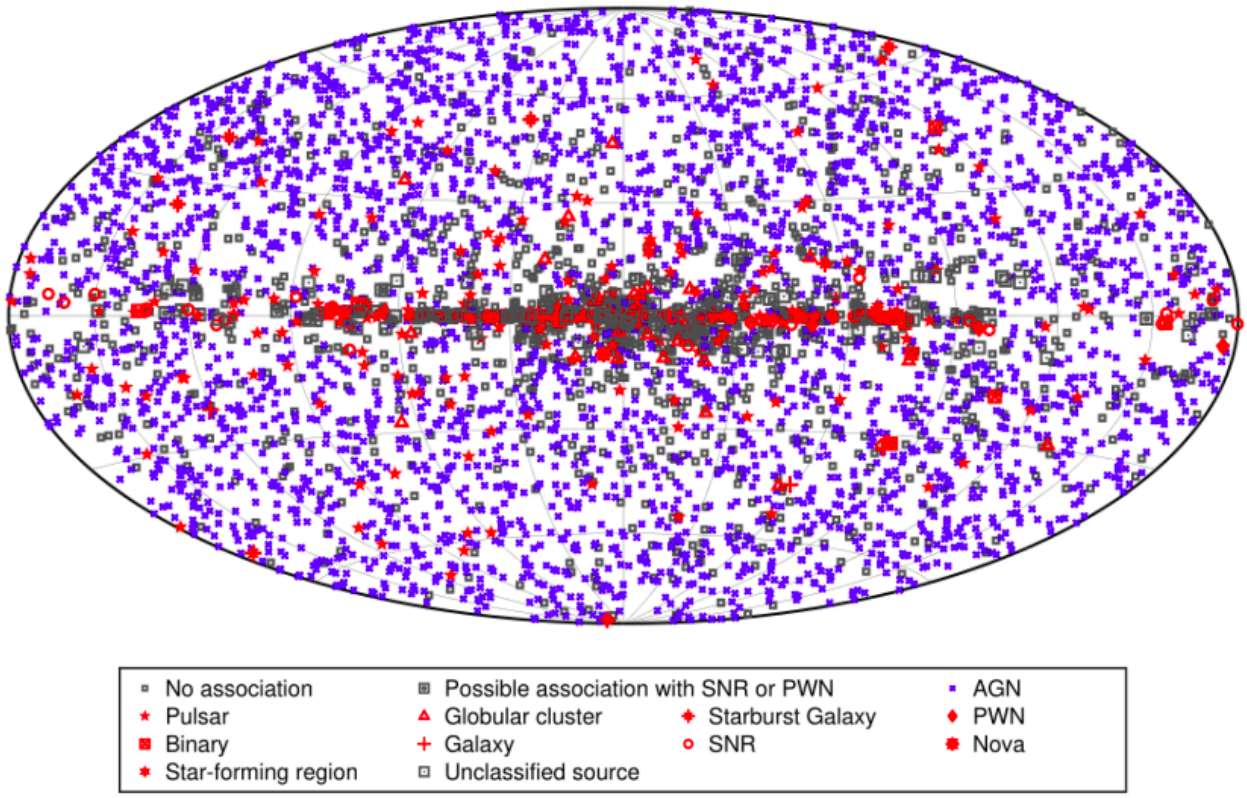


Figura 3: Fuentes del catálogo de *Fermi*-LAT 4FGL. Basado en 8 años de observación. 1336 de las 5064 fuentes de este catálogo no están asociadas a procesos conocidos.

conversión de pares con energías de entre 20 MeV y >300 GeV con alta resolución temporal, angular y de energías y bajo tiempo muerto. Tiene un amplio campo de visión, lo que le permite a lo largo de su órbita obtener imágenes de todas las direcciones del cielo cada día. El último catálogo de fuentes de radiación gamma detectadas por este telescopio, 4FGL-DR2 [9] se basa en 10 años de observación del cielo entero en un rango de energías entre 50 MeV y 1 TeV y tiene un total de 5787 fuentes, de las cuales entorno a un tercio no se han asociado con un fenómeno conocido. Entre estas fuentes no asociadas podría haber subhalos de materia oscura detectados, pero no identificados. Véase Figura 3.

Todos los datos obtenidos por el telescopio son de libre acceso a través de su página web, así como las herramientas para trabajar con ellos, llamadas *fermitools* [10]. La página web de *Fermi* dispone de explicaciones y tutoriales para el tratamiento de todos los archivos, la realización de simulaciones de observaciones, análisis estadísticos, etc. [11]. Todos los datos extraídos del telescopio deben pasar una serie de procesos que se realizan con estas herramientas para poder ser analizados.

Fermi-LAT mide fotones gamma, que con ayuda de sus herramientas, se convierten en flujos diferenciales de fotones procedentes de un determinado ángulo sólido del cielo en un intervalo de energías y para un intervalo de tiempo. Estos datos se almacenan en formato de tablas FITS (Flexible Image Transport System). Se guardan y se extraen por separado la información de los sucesos detectados y la información sobre el satélite y su trayectoria. El tratamiento de estos archivos se realiza mediante las *fermitools*, tal como se explica en posteriores apartados.

Las *fermitools* funcionan en terminales de Linux o Mac instaladas en un entorno de Conda [12]. Estas herramientas sufrieron una actualización en septiembre de 2020 por el paso a la tercera

versión de Python, su instalación es muy complicada y además, su funcionamiento con los datos que se tratan en este trabajo requiere una gran capacidad de computación. Por estos motivos se han utilizado de forma remota para todo el proceso de este trabajo a través de Secure Shell [13] en una máquina del Grupo de Altas Energías de la UCM.

Dado que es necesario hacer procesos repetitivos estos se han automatizado con programas de Python compartidos a través de un repositorio en GitHub creado específicamente para este trabajo [14] para facilitar su edición, el control de versiones y el traspaso de información a la máquina en remoto.

3. Simulación de observaciones de DM. Automatización de las simulaciones.

Para simular las observaciones de subhalos de materia oscura se han utilizado las herramientas de *Fermi*. La herramienta principal para este proceso es *gtobssim*, esta herramienta requiere dos archivos de entrada que deben prepararse previamente: un archivo XML con toda la información de las fuentes que se quieren simular y un archivo de texto con los nombres de las fuentes a simular para seleccionarlas en el XML. El tutorial Observation Simulation Tutorial de la web de *Fermi* explica todo el proceso de forma general, pero con algunos errores.

Al hacer una simulación de una observación hay que tener en cuenta que el telescopio no observa una fuente individual separada de cualquier otro suceso, sino que la medición se ve afectada por otros objetos presentes en el campo de visión y por el ruido de fondo. Este ruido de fondo se divide en dos componentes, una es extragaláctica, isotrópica, y otra es galáctica, que sigue una distribución espacial y espectral concreta, de forma que el ruido es máximo en dirección al centro galáctico y mínimo en una dirección perpendicular al plano de la Vía Láctea. Ambos fondos se definen en el archivo XML de entrada como fuentes extensas, mientras que las fuentes de materia oscura serán siempre puntuales en esta simulación.

La escritura del archivo XML de entrada de la simulación es uno de los pasos más complicados de todo el proceso, ya que requiere una sintaxis muy específica para que el programa pueda entender toda la información de la simulación. En el XML se indican en forma de etiquetas el nombre de la fuente, el tipo de función espectral, todos los parámetros que requiere dicha función, posición o distribución espacial, etc. En casos de necesitar utilizar funciones espectrales y/o distribuciones espaciales específicas sin aproximarlas a las funciones que el programa entiende en esta sintaxis, como es el caso para esta simulación de observación de materia oscura, se redirige desde el XML a los archivos (FITS, TXT o DAT) que contienen dichas distribuciones.

Para el caso de los fondos galáctico y extragaláctico, la página web de *Fermi* aporta los archivos necesarios para sus definiciones, sin embargo, el archivo que aporta para el fondo galáctico en formato FITS carece de determinados "headers" necesarios para funcionar con *gtobssim*. Para solucionar este problema, se emplea un código de Jaryd Ulbricht [15] que incorpora manualmente los "headers" que faltan con Astropy [16].

En el caso de los espectros de las fuentes de materia oscura, se generan a partir de las tablas numéricas publicadas por Cirelli et al. (2011) [17]. Estas tablas incluyen la información para generar espectros de 8 canales diferentes de aniquilación de WIMPs y cada tabla se corresponde al flujo de un producto final concreto. Para este caso, ya que solo se detectan rayos gamma, solo es necesaria la tabla correspondiente al flujo de rayos gamma. Un programa de Python descargado junto a las tablas permite extraer de la tabla los datos pertinentes, según los tres parámetros de entrada para

este programa: la masa de las partículas iniciales de materia oscura, el canal de aniquilación y el factor de normalización del flujo. Previo a su uso, el programa es modificado para que el archivo de salida tenga el formato adecuado para poder ser utilizado en la simulación con *gtobssim*. Se encuentra con la modificación en el repositorio de GitHub asociado a este trabajo [14].

Se pretende generar un catálogo de observaciones de materia oscura simuladas. Para cada observación se fijan los parámetros de masa de la partícula de materia oscura, canal de aniquilación, factor de normalización del flujo y posición en el cielo. Para generar un catálogo completo y útil para el propósito de entrenar una red neuronal sin sesgos, habría que variar todos estos parámetros y hacer múltiples simulaciones con ellos. En este trabajo, por limitaciones de tiempo, se demuestra el funcionamiento del código para un caso concreto: se escoje una masa para la partícula de materia oscura de 500 GeV, un canal de aniquilación $b\bar{b}$ y un factor de normalización de $10^{-4}m^{-2}s^{-1}MeV^{-1}$ y variando únicamente la posición en el cielo, de tal forma que se generan un total de 216 muestras.

En esta simulación las posiciones elegidas son tales que cada fuente está separada de otra en 15^{o} en ascensión recta, desde 0^{o} hasta 345^{o} y 15^{o} en declinación, desde -60^{o} a 60^{o}, generando una malla de 216 fuentes por todo el cielo. Con esta separación se espera que los fotones provenientes de una fuente no influyan demasiado en el análisis de otras. Se debe tener en cuenta, aún así, que por la geometría de este sistema de coordenadas esta separación no es equivalente en el ecuador y en las declinaciones más extremas, donde las fuentes estarán más cerca. Por este motivo no se toman declinaciones más altas.

La simulación realizada cubre un total de 10 años de observación, el mismo tiempo que abarca el último catálogo de *Fermi*, 4FGL-DR2. Sin embargo, debido a un fallo de la herramienta *gtobssim* que el equipo de *Fermi* aún no ha podido arreglar, no es posible realizar una simulación para un intervalo temporal tan largo, ya que el programa da errores aleatoriamente cuando se simula el fondo galáctico para tiempos largos. Simulaciones de un mes de observación o más reproducen este error de forma casi segura. Tras numerosas pruebas, se estima que un buen intervalo de tiempo para realizar estas simulaciones es de dos semanas; de esta forma, entorno a un 60 % de las veces la simulación se realiza correctamente sin ningún error. Por lo tanto, la estrategia en este punto es concatenar simulaciones de dos semanas, repitiendo aquellas que sufran el error hasta que salgan bien. Y así llegar a cumplir un total de 10 años simulados de observación.

Para incluir fácilmente toda esta información en el archivo XML y generar la simulación con la concatenación temporal explicada, se ha creado un programa en Python, disponible en el repositorio de GitHub [14] y válido para cualquier otra elección de parámetros, que automatiza este proceso de la siguente forma:

- Toma de entrada los parámetros del canal de aniquilación, masa de la partícula de materia oscura (en GeV), y el factor de normalización. Incluye la posibilidad de introducir los valores mínimo y máximo de energías para la simulación (parámetros por defecto: 100 MeV y 500 GeV, respectivamente). También se puede variar el tiempo de corte de la simulaciones (el tiempo por defecto es de dos semanas).
- Se llama al programa modificado de Cirelli para que genere el espectro de emisión de la materia oscura dados los parámetros de canal de aniquilación, masa y normalización. Este, para los parámetros dados, es el representado en la Figura 4.
- Se crean el archivo XML de la simulación y el listado de nombres. Estos archivos se van rellenando con la información de todas las fuentes a simular: las 216 fuentes de materia oscura con su espectro asociado y los 2 fondos galáctico y extragaláctico.
- Se divide el tiempo total de la simulación en los cortes que se han hecho de dos semanas (o el

tiempo de corte que se haya elegido) y para cada corte se corre la simulación con *gtobssim* con los archivos creados y los parámetros dados. Se evalúa en cada caso si ha habido un error en la simulación o no y en caso de haberlo repite el mismo corte cambiando la semilla aleatoria de la simulación.

Durante las pruebas al crear este código, se fueron variando todos los parámetros con los que se ejecutaba. De esta forma, se segura que el código no está sesgado para funcionar solo en ciertos casos, sino que es válido para cualquier combinación de parámetros.

Es muy importante tener en cuenta en este punto que las unidades para el flujo que dan las tablas de Cirelli son $cm^{-2}s^{-1}MeV^{-1}$, mientras que la simulación toma de entrada los datos en $m^{-2}s^{-1}MeV^{-1}$. Es decir, que introduciendo en el programa de Cirelli un factor de normalización 10^{-4} y utilizando el archivo generado se consigue, efectivamente, que la simulación tenga un flujo de $10^{-4}m^{-2}s^{-1}MeV^{-1}$. Pero a la hora de realizar el análisis y la comparación con el espectro inicial, en el que las unidades vuelven a cambiar a $cm^{-2}s^{-1}MeV^{-1}$, el flujo obtenido será $10^{-8}cm^{-2}s^{-1}MeV^{-1}$.

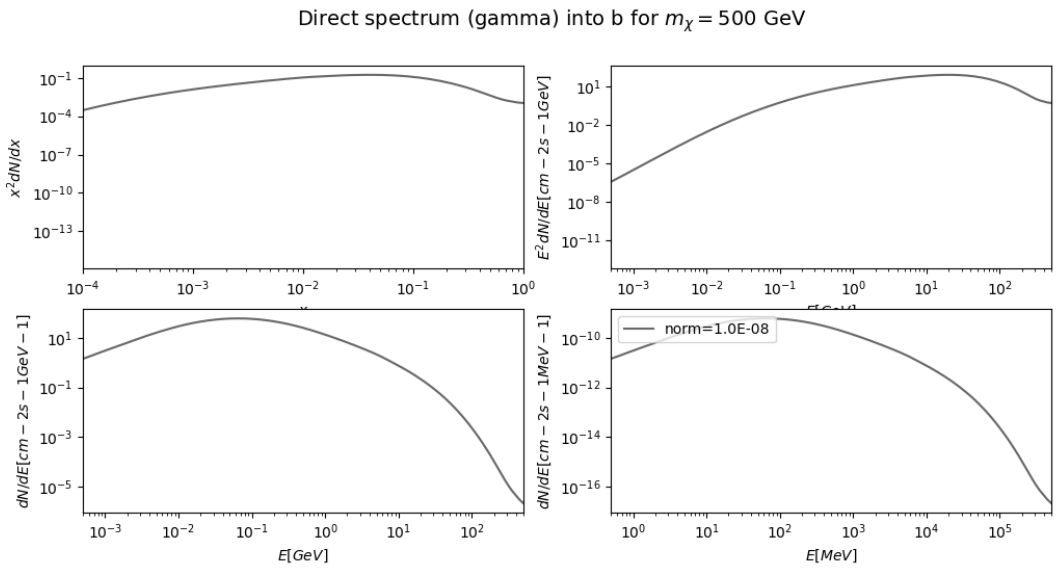


Figura 4: Espectro de emisión en gamma de materia oscura aniquilándose a través del canal $b\bar{b}$, con masa de 500GeV y factor de normalización de $10^{-8}cm^{-2}s^{-1}MeV^{-1}$, generado mediante el código de Cirelli [17]. La variable x representa energía dividida por la masa de la partícula de materia oscura.

Este proceso genera un total de 260 archivos de sucesos y sus respectivos archivos de información del satélite.

Se realiza un procesado rápido de estos archivos, utilizando la herramienta *gtselect* (herramienta para filtrar sucesos según sus características) para unir los archivos de sucesos en uno único y *gtbin* (herramienta para subdividir los sucesos en intervalos concretos de ángulo en el cielo y/o energías) y *DS9*(visualizador de imágenes en formato de tablas FITS) para generar una imagen (*cmap* o mapa de cuentas) de estos bajo determinadas proyecciones (AIT o CAR [18]). En la Figura 5 se representa esta imagen en coordenadas celestes y galácticas para la proyección AIT, que es la utilizada más frecuentemente y en la que se comprueba que las distancias entre fuentes en la realidad tal como se han elegido las posiciones, y para la proyección CAR en coordenadas celestes, que coloca las coordenadas de ascensión recta y declinación como cartesianas, mostrando efectivamente el grid que se ha escogido hacer.

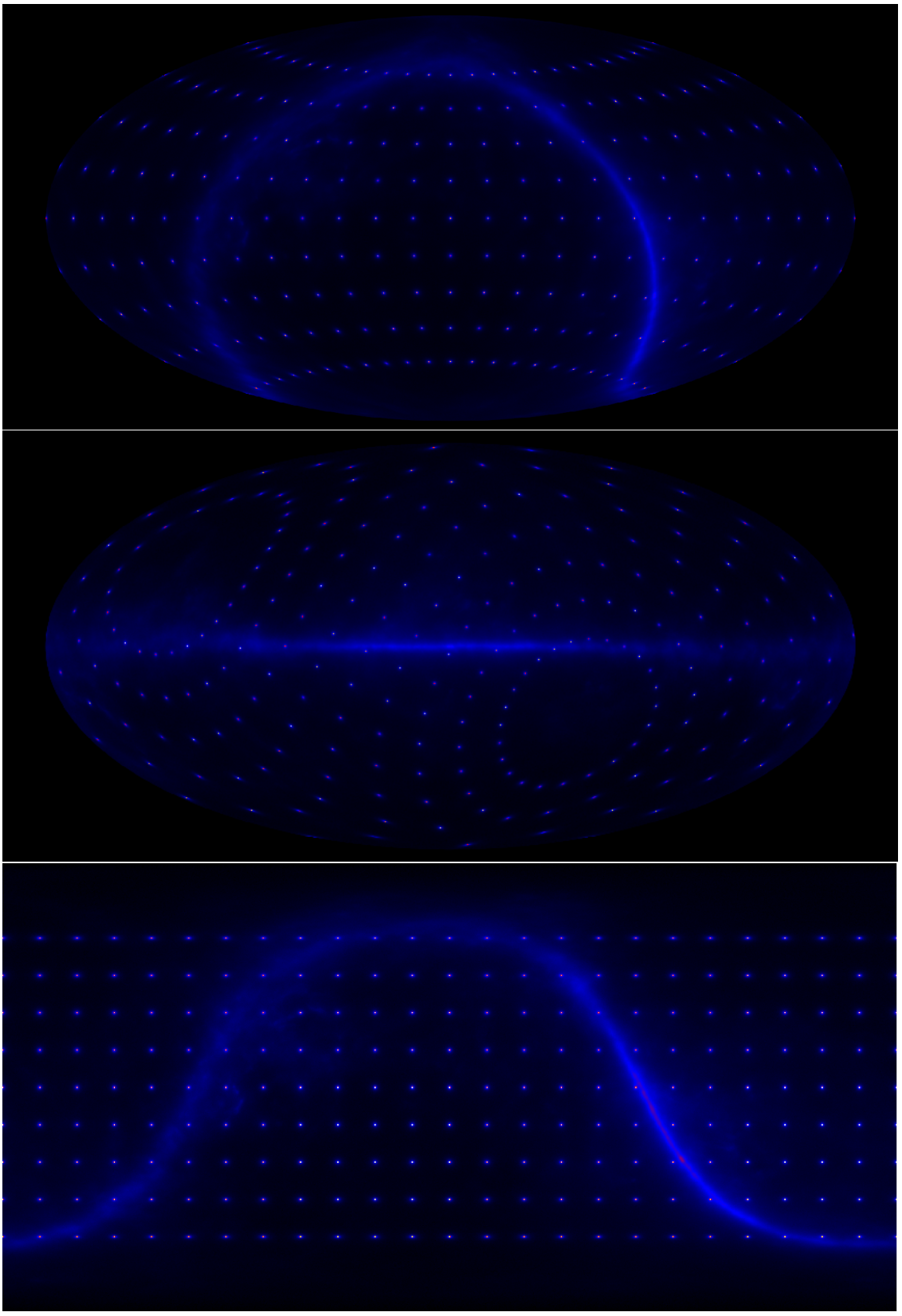


Figura 5: Mapa de la simulación de 216 fuentes de DM y los fondos galáctico y extragaláctico. Primera imagen en coordenadas celestes y proyección AIT, segunda imagen en coordenadas galácticas y proyección AIT y tercera imagen en coordenadas celestes y proyección CAR.

La variabilidad en las coordenadas es positiva en cuanto a que permite representar toda la casuística de interacción entre las fuentes puntuales y los fondos a la hora de realizar los análisis.

4. Análisis de las simulaciones para la creación de catálogos sintéticos.

Una vez se han finalizado las simulaciones, estas se deben analizar como si se tratara de simulaciones reales, de forma que los resultados sean comparables a los de una observación real de una fuente del catálogo de *Fermi* no asociada.

La principal función de las herramientas de *Fermi* es precisamente analizar las observaciones del telescopio y ajustar los datos a un modelo preseleccionado, en el que se varían determinados parámetros hasta que el valor de la verosimilitud (probabilidad de obtener los datos dado el modelo) se maximiza. El esquema general para realizar un análisis de una fuente con las herramientas de *Fermi* es el siguiente:

1. Seleccionar los datos alrededor de la fuente de interés para poder tener en cuenta los fotones que se superpongan entre unas fuentes y otras.
2. Seleccionar el modelo. Se escribe un modelo que debe incluir todas las fuentes en esa región con sus parámetros espectrales y espaciales, en la que los parámetros espectrales de la fuente de interés se dejan libres para su ajuste.
3. Se calculan y se generan diferentes archivos que son necesarios para realizar el ajuste final.
4. Finalmente, se realiza el ajuste en el que algoritmos optimizadores van cambiando los parámetros libres hasta conseguir maximizar la verosimilitud del modelo para los datos que se han seleccionado.

En este trabajo, los espectros son ajustados a una distribución Log-Parábola (2), dejando libres los parámetros N_0, α, β y E_b.

$$\frac{dN}{dE} = N_0 \left(\frac{E}{E_b}\right)^{-(\alpha+\beta\log(E/E_b))} \tag{2}$$

Al contrario que con la simulación, no es posible realizar el análisis de las 216 fuentes de una sola vez debido a la cantidad de recursos computacionales que esto requeriría por el gran espacio de parámetros a ajustar y por limitaciones de las propias herramientas. Por este motivo, se repite el proceso del análisis para el ajuste de cada fuente individualmente.

Para realizar este proceso y poder almacenar los resultados de los ajustes de manera automática, se crea, del mismo modo que para la simulación, un programa en Python, disponible en el repositorio de GitHub [14] y que también es válido para otras elecciones de parámetros. Funciona del siguiente modo:

- Toma de entrada los mismos parámetros que el programa para la simulación, con la diferencia de que en este no es necesario el tiempo de los cortes de las simulaciones y el valor por defecto para la energía máxima es menor (100GeV), debido a que los eventos muy cercanos en energía al límite de la simulación dan problemas en el análisis, por un lado por la baja estadística de fotones que se obtiene a estas energías y, además, por errores de lectura que aparecen en *gtlike*.

- Crea las carpetas donde guardar los resultados y dos listas con los nombres de los 260 archivos de sucesos de la simulación y los de los archivos del satélite.
- A partir de los archivos del satélite, se utilizan las herramientas *gtltcube* y *gtexpcube2* para generar para todo el cielo unos archivos de "livetime cube" y "exposure cube", necesarios para el análisis de verosimilitud. Estos archivos guardan la información del tiempo de exposición que ha tenido cada región del cielo y bajo qué inclinaciones, imprescindible para el posterior cálculo de la verosimilitud con cada fuente. El cálculo del "livetime cube" es el proceso más largo, puede durar una gran cantidad de horas, pero solo es necesario hacerlo una vez, el mismo archivo de resultado se utiliza en todos los análisis. El "exposure cube", en cambio, se debe calcular para cada par de coordenadas.
- Para cada una de las fuentes a analizar, se crea un archivo XML con el modelo de su región de interés (ROI), que es de 15^{o}. En él se definen la fuente a analizar con todos sus parámetros libres para el ajuste como Log-Parábola, las ocho fuentes más cercanas a esta con el espectro de Cirelli y sin parámetros libres y los fondos galáctico y extragaláctico, dejando libre su factor de normalización. En este modelo, las unidades para los flujos son $cm^{-2}s^{-1}MeV^{-1}$.
- Para el análisis de cada una de las fuentes, se seleccionan los datos de los archivos de la simulación categorizados como fotones frontales y traseros (evclass=128 y evtype=3) que caen dentro de la ROI de 15^{o} alrededor de la fuente en cuestión, cortando en la nueva energía máxima y para los intervalos de tiempo buenos (GTIs). Para esto se utilizan las herramientas *gtselect* y *gtmktime*.
- Se genera para cada selección de datos un cubo de cuentas mediante la herramienta *gtbin*, dividiendo todos los sucesos en compartimentos espaciales y de energías. La división espacial se hace en píxeles cuadrados de 0.2^{o} de lado y la división en energías se hace en intervalos equivalentes en escala logaritmica, concretamente se dividen 10 intervalos por cada orden de magnitud, lo que con las energías utilizadas son 30 intervalos.
- Para el modelo en XML creado para cada fuente, se aplica la herramienta *gtsrcmaps* para crear un modelo de mapa de cuentas para utilizar en el análisis de verosimilitud. Toma los espectros del modelo en XML, los multiplica por la exposición de la posición de la fuente y convoluciona esta exposición con la PSF efectiva dada por la respuesta y resolución del detector.
- Finalmente, se utiliza la herramienta *gtlike* con el optimizador "New Minuit" [19] para encontrar los parámetros de la distribución Log-Parábola que maximiza la verosimilitud para el espectro de cada fuente y guardarlos como resultado.

Al definir los parámetros que se dejan libres para el ajuste a Log-Parábola, se escogen los mismos valores iniciales, de escala, máximo y mínimo de cada parámetro que aparecen en el ejemplo de esta definición en la web de *Fermi*. Esto es, los valores de la Tabla 1.

	min	**max**	**scale**	**value**
N_0	0.001	1000	1e-9	1
α	0	10	1	1
β	0	10	1	2
E_b	20	1e4	1	300

Tabla 1: Valores mínimo, máximo, de escala e inicial para cada parámetro del ajuste a Log-Parábola.

Como resultado se obtienen archivos en formato XML, uno por cada fuente, similares al que entra

como modelo, pero con todos los parámetros que se han dejado libres ajustados y cada uno con su estimación de error correspondiente. Para analizar los resultados, estos archivos se leen y se almacenan en una tabla de Pandas [20], mediante un código de Python incluido en el GitHub [14]. El código, además, genera varias figuras, que se muestran a continuación para comprobar la bondad de los ajustes.

Para empezar, se observa al tratar los resultados con este código de lectura que la convergencia no ha sido buena en todos los casos. De los 216 análisis de verosimilitud realizados, 7 no han generado ningún resultado y 31 han generado resultados incompletos, sin estimación de errores. Es decir, *gtlike* con el optimizador "NEWMINUIT" no ha alcanzado la convergencia en 38 casos a pesar de tratarse todos de fuentes que se han simulado con exactamente las mismas características, para las que solo varía la posición en el cielo. En este punto, se decide continuar el análisis con los 178 ajustes que sí han dado resultados y estimación de errores.

Representando en histogramas los valores de los parámetros del ajuste para las 178 muestras, se obtiene la Figura 6. Los valores no parecen seguir una tendencia clara y hay unos pocos ajustes cuyos valores se alejan significativamente de los demás, lo que hace que la media y la mediana de cada parámetro tengan grandes diferencias. Para comprobar cómo afectan estas desviaciones a la forma de los espectros, se representan todos ellos en la Figura 7. En esta última figura, la gran mayoría de los espectros aparecen juntos con los flujos más altos, mientras que hay 6 muestras cuyos flujos decaen a valores extremadamente pequeños y que no tienen sentido. Se decide, a partir de ahora, no contar tampoco con dichas 6 muestras que han convergido, pero de forma muy errónea. De esta forma, se cuenta finalmente con 172 ajustes "buenos", a los que se le pone la etiqueta "con discriminación".

De nuevo, se realizan los histogramas de los parámetros y el cálculo de la media y la mediana una vez se ha hecho la discriminación de aquellos ajustes que no generaban un espectro realista, obteniendo la Figura 8. Sigue habiendo algunos parámetros con valores muy alejados de la mediana, pero, en general, la media y la mediana tienen menores diferencias que en los ajustes sin discriminación.

Tras la discriminación, la forma de los espectros se muestra en la Figura 9. En este caso, los espectros son mucho más parecidos entre sí, especialmente en las energías entre 1 GeV y 100 GeV, donde la estadística de fotones es mucho mejor (hay gran número de sucesos, poco ruido en comparación con energías más bajas y buena resolución espacial y en energías). Esto indica que la dispersión observada en los parámetros en la Figura 8 no es muy problemática a la hora de representar el espectro y probablemente es debida a una gran correlación entre los parámetros, algo esperable de un ajuste a una función multiparamétrica, pero que no se analiza en mayor profundidad en este trabajo debido a la limitación temporal.

Incluyendo en la Figura 9 el espectro generado por las tablas de Cirelli, el mismo que se introdujo de entrada para realizar las simulaciones, pero con las unidades correctas, se observa en la Figura 10 que los 172 ajustes a Log-Parábola respresentan muy bien el espectro en el rango de energías entre 1 GeV y 100 GeV. Si se alargan estos espectros hasta 500 GeV, tramo en el que se ha simulado, pero no se ha hecho análisis por los errores que produce, tal como se muestra en la Figura 11, el espectro deja de coincidir con los ajustes, ya que cerca de la masa de la partícula de materia oscura (500 GeV en este caso) se produce un cambio en la curvatura del espectro que una Log-Parábola no reproduce.

Finalmente, se pretende comprobar si el motivo de que algunas fuentes no hayan tenido convergencia o no hayan dado resultados satisfactorios está relacionado con su posición en el cielo, ya que esta es, en principio, su única diferencia posible respecto a las 172 fuentes que sí han generado buenos ajustes. Para ello, en la Figura 12 se representan las coordenadas de las fuentes que sí han dado

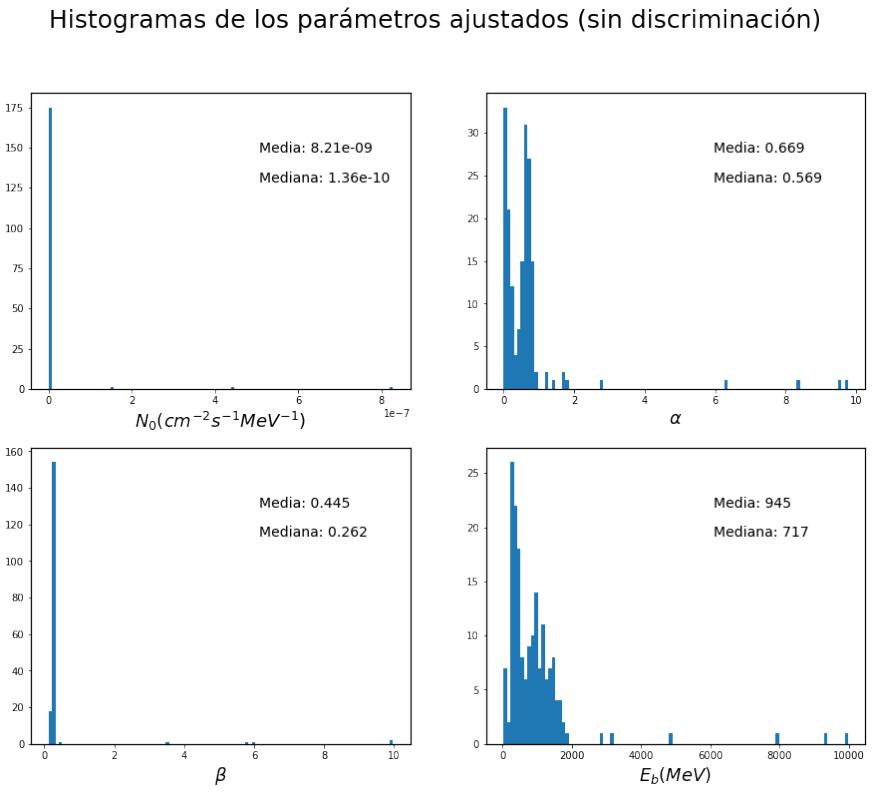


Figura 6: Histogramas de los cuatro parámetros del ajuste a Log-Parábola para las 178 muestras con convergencia en el análisis. Sin discriminación de si son buenos ajustes o no. Se incluye el valor de la media y la mediana para cada parámetro.

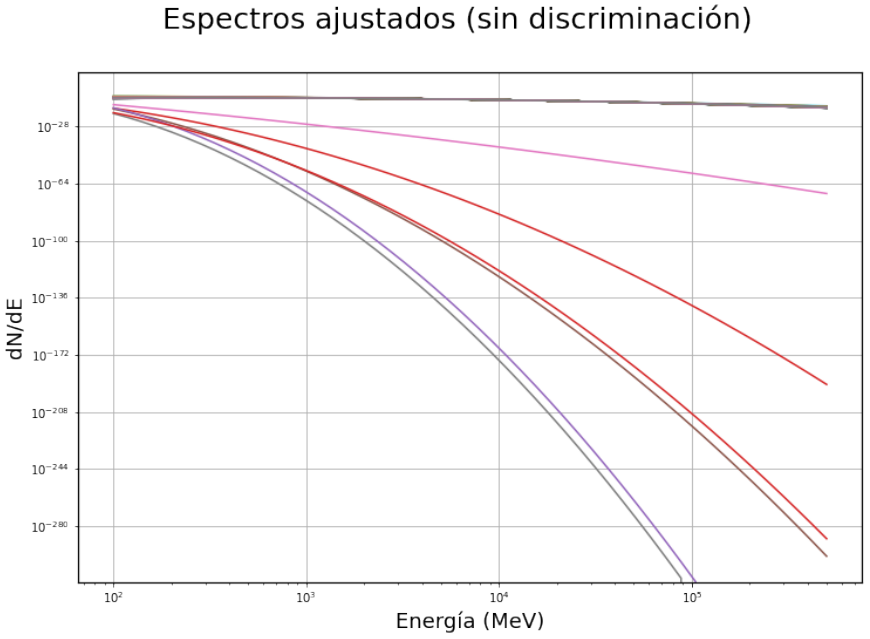


Figura 7: Espectros de los ajustes a Log-Parábola para las 178 muestras con convergencia en el análisis. Sin discriminación de si son buenos ajustes o no.

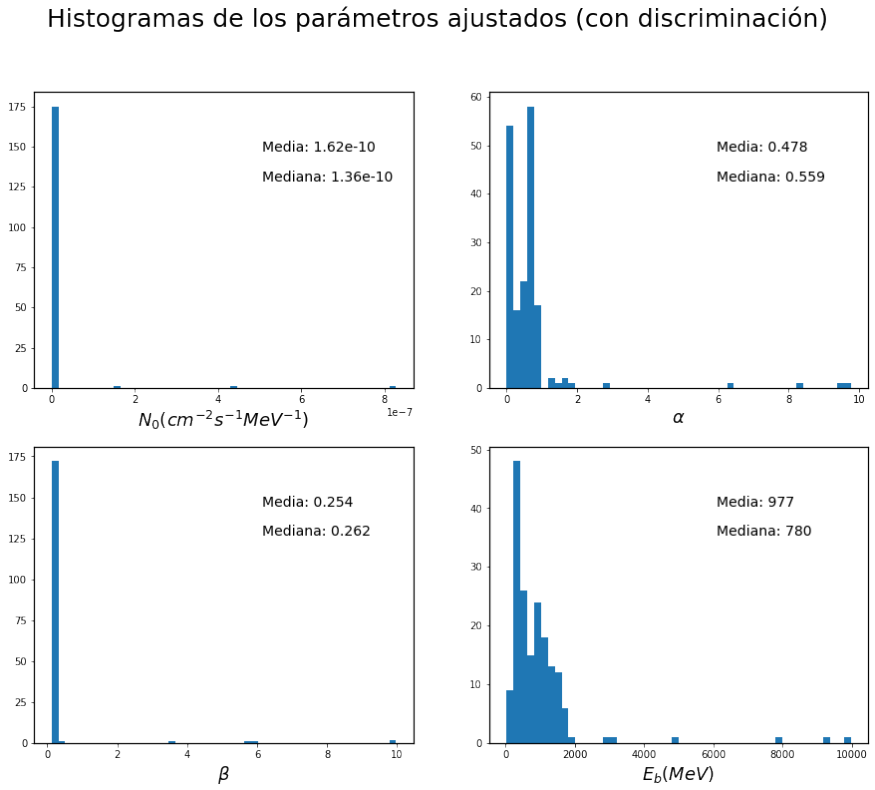


Figura 8: Histogramas de los cuatro parámetros del ajuste a Log-Parábola para las 172 muestras con convergencia en el análisis y con discriminación de si son buenos ajustes o no. Se incluye el valor de la media y la mediana para cada parámetro.

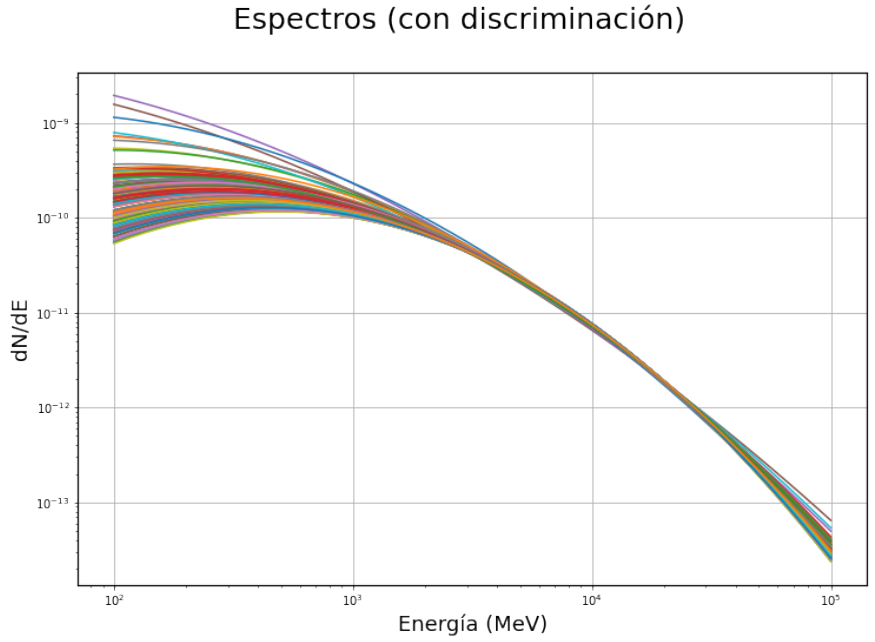


Figura 9: Espectros de los ajustes a Log-Parábola para las 172 muestras con convergencia en el análisis y con discriminación de si son buenos ajustes o no.

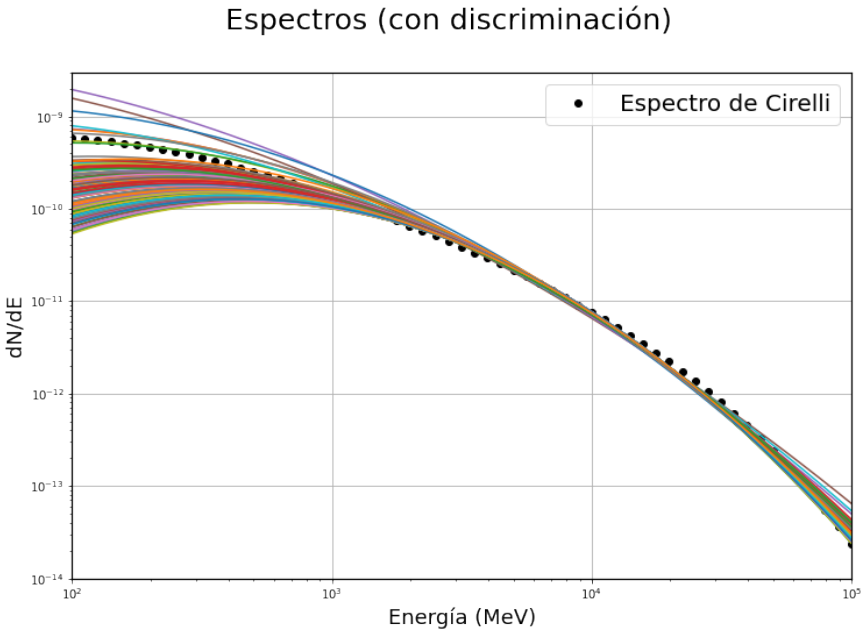


Figura 10: Espectros de los 172 ajustes a Log-Parábola en comparación con el espectro generado por las tablas de Cirelli. Rango desde 100 MeV hasta 100 GeV, rango en el que se hace el análisis de verosimilitud.

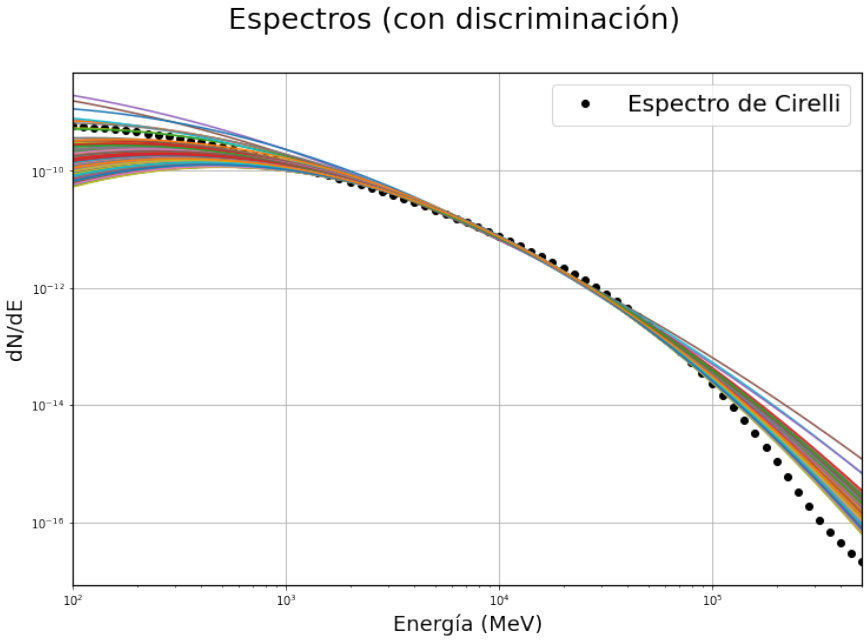


Figura 11: Espectros de los 172 ajustes a Log-Parábola en comparación con el espectro generado por las tablas de Cirelli. Rango desde 100 MeV hasta 500 GeV. 500 GeV es la masa de la partícula de materia oscura simulada.

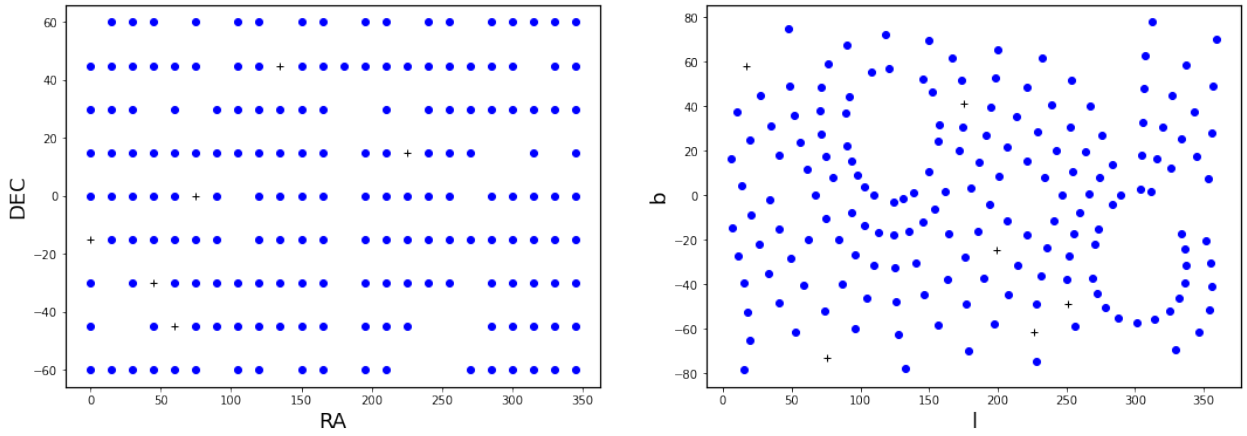


Figura 12: Coordenadas celestes a la izquierda y galácticas a la derecha de las fuentes que han tenido convergencia (cruces) y han dado buenos ajustes (puntos azules). Proyección CAR.

convergencia y que han dado buenos ajustes para sus espectros. No aparenta haber ninguna relación entre la convergencia o el buen ajuste con las coordenadas, de modo que se concluye por ahora que se trata de errores por números aleatorios, tal como ocurre con *gtobssim*, por lo que puede que su solución (no elaborada en este trabajo) deba ser similar a la solución que se da en las simulaciones, o bien, que esto requiera un análisis más profundo.

5. Conclusiones y posibles continuaciones del estudio.

Este trabajo cubre tres objetivos principales:

El primero de ellos es crear un código capaz de automatizar la simulación de la observación por parte de *Fermi*-LAT, utilizando su herramienta *gtobssim*, de numerosas fuentes de rayos gamma provenientes de la autoaniquilación de partículas de materia oscura (WIMPs). Estas simulaciones utilizan los espectros generados a partir de las tablas de Cirelli [17] y se pueden realizar para cualquier combinación de parámetros de canal de aniquilación, masa de la WIMP y factor de normalización, por lo que puede utilizarse masivamente para la generación de un catálogo de fuentes simuladas con diferentes combinaciones de estos parámetros, generando estadística suficiente para poder entrenar a una red neuronal para su reconocimiento.

El segundo objetivo alcanzado es automatizar también el tratamiento de datos y el análisis de verosimilitud de las simulaciones, tratadas del mismo modo que si fueran observaciones reales pasando por todas las herramientas de *Fermi* necesarias en el proceso. El resultado de estos análisis es un ajuste del espectro a una distribución Log-Parábola 2, con sus cuatro parámetros y sus respectivos errores para cada una de las fuentes simuladas.

Como tercer objetivo, se consigue llevar a cabo ambos procesos simulando y analizando 216 fuentes de materia oscura para un canal de aniquilación $b\bar{b}$, una masa de la partícula de 500 GeV y un factor de normalización de 10^{-4}. Esto permite conocer la respuesta de *Fermi* y sus herramientas a una emisión con estas características. La simulación de diez años de observación se genera correctamente en un proceso de aproximadamente dos días de duración, mientras que el análisis, de aproximadamente un día de duración, se realiza correctamente para 172 de las 216 fuentes simuladas.

Se concluye que este método es efectivo para crear en poco tiempo grandes catálogos de fuentes simuladas que podrían ser útiles para entrenar a redes neuronales de aprendizaje profundo, junto con otros catálogos de simulaciones de diferentes fuentes, para poder diferenciar automáticamente unas de otras. Sin embargo, el método presenta varios aspectos susceptibles de mejora:

- El código necesita optimización, tanto para el tiempo de ejecución como para espacio en disco duro que ocupan los archivos generados en el análisis.
- No todos los espectros se ajustan bien en el análisis. Parece tratarse de un error aleatorio que se debe detectar.
- Un ajuste a la distribución Log-Parábola no reproduce el cambio de curvatura y la caída que se produce en el espectro cerca del valor de la masa de la partícula de materia oscura, que es probablemente la característica de estos espectros que más puede ayudar a diferenciarlos de otros.
- Falta dar a los datos obtenidos un formato de catálogo real de *Fermi*, en formato FITS con sus headers correspondientes.

Los primeros pasos para continuar este estudio de búsquedas de materia oscura, antes de seguir generando catálogos, deben consistir en solucionar estos puntos. Para ello, una herramienta que podría ser muy útil es el paquete de Python "Fermipy-[21], una versión de las *fermitools* en Python, que facilita la optimización y permite analizar la distribución espectral de energías (SED). El estudio de la SED probablemente sí reflejaría mejor la forma del espectro a energías cercanas a la masa de la partícula, por lo que generaría información útil para que la red neuronal diferencie unas fuentes de otras.

Posteriormente, será necesario también tener la capacidad de generar catálogos falsos para otras fuentes diferentes a la materia oscura, para lo cual se deberá realizar un análisis estadístico de los espectros de estas otras fuentes y sus parámetros, según las ha detectado *Fermi*-LAT.

El siguiente paso sería generar finalmente todos estos catálogos. Se generarán catálogos para observaciones de materia oscura utilizando el método descrito en este trabajo con las mejoras pertinentes y variando los parámetros de masa de la WIMP, canal de aniquilación y factor de normalización del flujo, y catálogos para otros tipos de fuentes conocidas.

Finalmente, y una vez se tuvieran todos los catálogos, se utilizarían para entrenar a una red neuronal para que aprenda a diferenciar los tipos de fuentes, tal como otros autores han hecho anteriormente para clasificar otras fuentes en los catálogos de *Fermi* [22].

Referencias

[1] Planck Collaboration. Planck 2018 results. *Astronomy & Astrophysics Supplement Series*, 641(A6):67, 2020. doi: 10.1051/0004-6361/201833910.

[2] Gianfranco Bertone and Dan Hooper. History of dark matter. *Reviews of Modern Physics*, 90(045002), 2018. doi: 10.1103/revmodphys.90.045002.

[3] Gianfranco Bertone, Dan Hooper, and Joseph Silk. Particle dark matter: evidence, candidates and constraints. *Physics Reports*, 405(5-6):279–390, 2005. doi: 10.1016/j.physrep.2004.08.031.

[4] G. Bertone and T.M.P. Tait. A new era in the search for dark matter. *Nature*, 562:51–56, 2018. doi: 10.1038/s41586-018-0542-z.

[5] J. Diemand, M. Kuhlen, P. Madau, M. Zemp, B. Moore, D. Potter, and J. Stadel. Clumps and streams in the local dark matter distribution. *Nature*, 454:735–738, 2008. doi: 10.1038/nature07153.

[6] D. Nieto. Hunting for dark matter subhalos among the fermi-lat sources with veritas. *34th International Cosmic Ray Conference*, The Hague, The Netherlands., 2015.

[7] NASA. The fermi large area telescope — fermi lat. 2021. https://glast.sites.stanford.edu/.

[8] NASA. The fermi gamma-ray space telescope. 2021. https://fermi.gsfc.nasa.gov/.

[9] S. Abdollahi, F. Acero, M. Ackermann, M. Ajello, W. Atwood, and M. Axelsson et al. Fermi large area telescope fourth source catalog. *The Astrophysical Journal Supplement Series*, 247(1):33, 2020. doi: 10.3847/1538-4365/ab6bcb.

[10] Fermi-LAT. Fermi gamma-ray space telescope. 2021. https://github.com/fermi-lat.

[11] NASA. Fssc: Fermi data » data analysis » online documentation » fermitools: Cicerone. 2021. https://fermi.gsfc.nasa.gov/ssc/data/analysis/documentation/Cicerone/.

[12] Anaconda. Conda. 2021. https://anaconda.org.

[13] Tatu Ylonen. Secure shell (ssh). 2021. https://www.ssh.com/academy/ssh.

[14] Juan Bernete. Repositorio de github de este tfm. 2021. https://github.com/JBernete/TFM-Juan-Bernete.

[15] Jaryd Ulbricht. Fermimodel. *GitHub*, 2019. https://github.com/jaulbric/fermimodel/blob/master/READM

[16] The Astropy Project. Astropy. 2021. https://www.astropy.org/.

[17] M. Cirelli, G. Corcella, A. Hektor, G. Hütsi, M. Kadastik, and P. Panci et al. Pppc 4 dm id: a poor particle physicist cookbook for dark matter indirect detection. *Journal Of Cosmology And Astroparticle Physics*, 2011(3):051–051, 2011. doi: 10.1088/1475-7516/2011/03/051.

[18] Calabretta and Greisen. Spherical map projections. *A&A*, 395(1077), 2002. https://www.aanda.org/articles/aa/full/2002/45/aah3860/node5.html.

[19] James F. Function minimization and error analysis. *CERN Program Library*, D506, 1998. http://cdssls.cern.ch/record/2296388/files/minuit.pdf.

[20] Pandas. https://pandas.pydata.org/.

[21] M. Wood, R. Caputo, E. Charles, M. Di Mauro, J. Magill, J. S. Perkins, and Fermi-LAT Collaboration. Fermipy: An open-source python package for analysis of fermi-lat data. *35th International Cosmic Ray Conference, 10-20 July, 2017. Bexco, Busan, Korea, Proceedings of Science*, 301. https://pos.sissa.it/cgi-bin/reader/conf.cgi?confid=301 https://fermipy.readthedocs.io/en/latest/.

[22] S. Germani, G. Tosti, Lubrano, S. Cutini, I. Mereu, and A. Berretta. Artificial neural network classification of 4fgl sources. page 10, 2021. https://arxiv.org/abs/2106.08222.

Este trabajo de fin de máster me he visto obligado a escribirlo en muy poco tiempo, en una época de gran presión por el resto de asignaturas del máster y por la entrada al doctorado. A ello se ha unido una lesión en la mano y el brazo derechos, que ha afectado mucho a mi productividad. Es por ello que durante estas semanas el apoyo por parte de la gente que me rodea ha sido crucial y quiero agradecer este apoyo a mis amigos, mi familia, mis profesores y tutores y mis compañeros de clase; en especial a Víctor, a Leire y a mis padres por su ayuda.