



# Estimador de energía para eventos detectados por HAWC's Eye empleando técnicas de Machine Learning

F. González<sup>a</sup>, R. Alfaro<sup>b</sup>, T. Capistrán<sup>a</sup>, M. M. González<sup>a</sup>, Y. F. Pérez<sup>a</sup>, J. Serna<sup>b</sup> e I. Torres<sup>c</sup>.

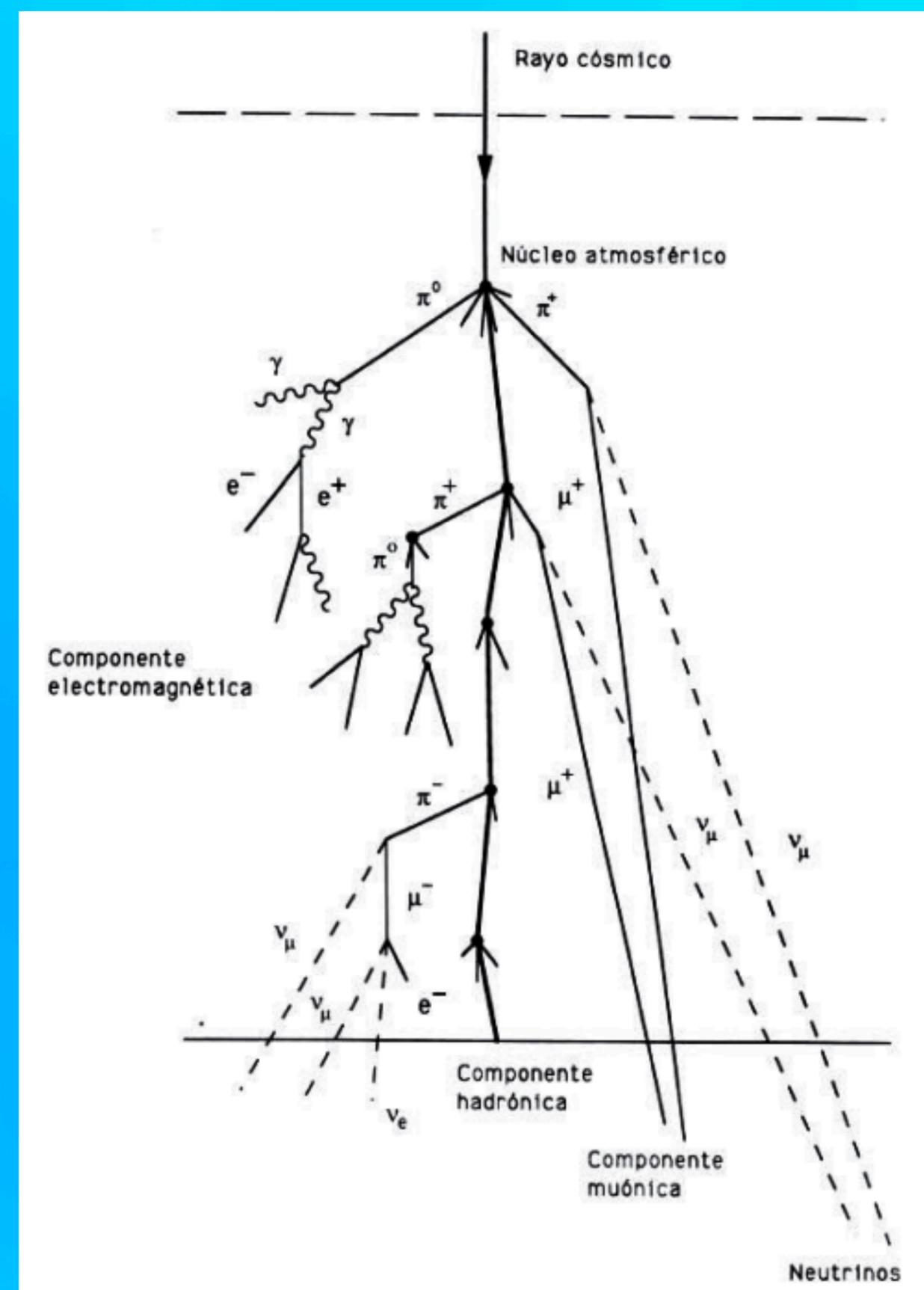
<sup>a</sup> Instituto de Astronomía, UNAM.

<sup>b</sup> Instituto de Física, UNAM.

<sup>c</sup> Instituto Nacional de Astrofísica, Óptica y Electrónica.

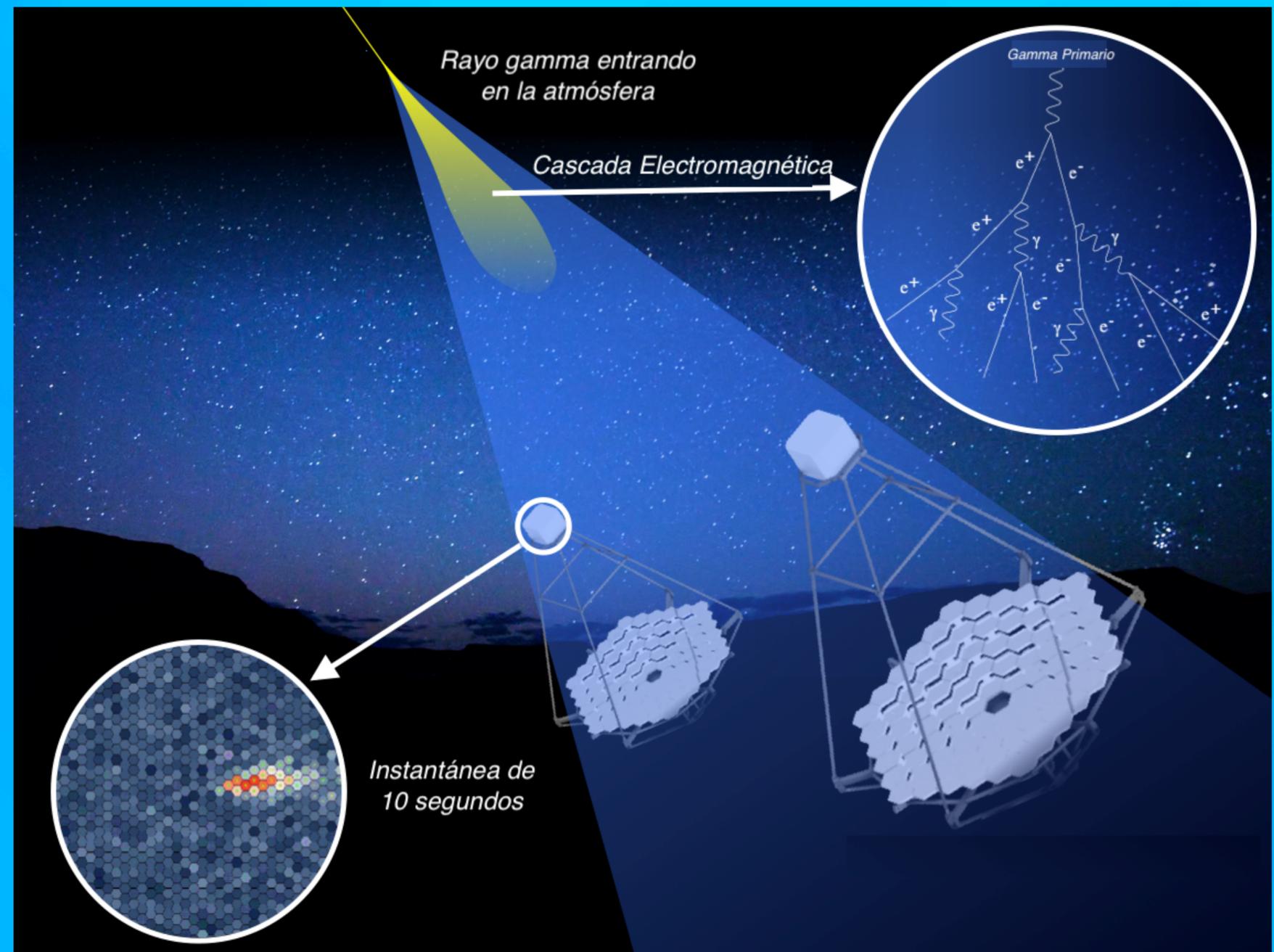
# Cascadas Atmosféricas Extensas

- Cuando un rayo gamma o rayo cósmico interacciona con los átomos de la atmósfera se produce una Cascada Atmosférica Extensa.
- La radiación Cherenkov se genera cuando una partícula cargada viaja a una velocidad superior que la velocidad de la luz en un medio.



# Telescopio Cherenkov Atmosférico

- Se colecta la radiación Cherenkov generada en la atmósfera.
- Compuesto por un área reflectante, una cámara y un sistema de adquisición de datos (DAQ).
- El campo de visión es pequeño.



R. White (MPIK) / K. Bernlohr (MPIK) / DESY

# Arreglos Extendidos

- Cubren una gran área para detección.
- Están compuestos por una gran cantidad de detectores, un ejemplo son los detectores de Cherenkov.
- Los detectores Cherenkov tienen en su interior un medio distinto del aire (lo más común es el agua), para que se genere la radiación Cherenkov y colectarla.



# Comparación entre técnicas

## Telescopios Cherenkov Atmosféricos

- Diámetro del telescopio: 5 - 25 metros.
- Ciclo de trabajo: 5% - 10%.
- Condiciones ambientales específicas.
- Resolución angular:  $\sim 0.01^\circ$ .
- Detección de fuentes puntuales.
- Rango de energía:  $\sim 20$  GeV - 30 TeV.
- Excelente discriminador gamma/hadrón.
- Alto costo de producción.

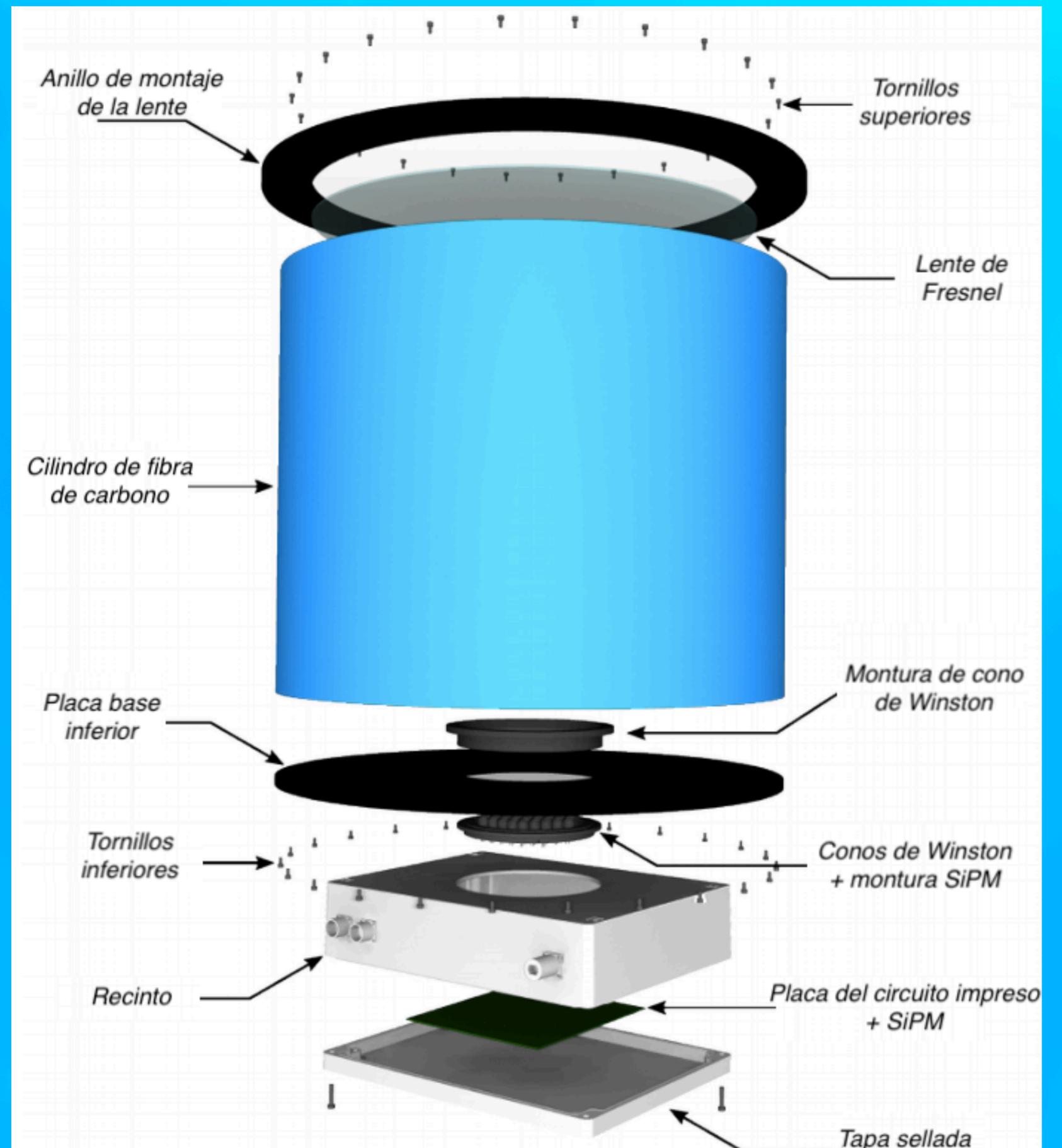
## Arreglos Extendidos

- Tamaño del arreglo de cientos de  $m^2$ .
- Ciclo de trabajo: 95%.
- Sin condiciones ambientales específicas.
- Resolución angular:  $\sim 0.1^\circ$ .
- Detección de fuentes extendidas.
- Rango de energía:  $\sim 100$  GeV - 100TeV.
- Buen discriminador gamma/hadrón.
- Alto costo de producción.

# HAWC's Eye

Prototipo de Telescopio Cherenkov Atmosférico refractivo, compacto y de bajo costo

- Diámetro de 0.5 metros.
- Lente de Fresnel.
- La radiación Cherenkov se refracta a diferencia de otros telescopios Cherenkov.
- Cámara compuesta por 64 pixeles SiPM.
- Campo de visión de  $12^\circ$ .
- DAQ es una adaptación al utilizado en el telescopio FACT.
- Control de manera remota.
- Costo aproximado de 10,000 euros.



# Simulaciones

Se simuló la detección de las cascadas atmosféricas extensas con HAWC's Eye.

CORSIKA V7.69:

- Simulación de las cascadas atmosféricas extensas producidas por rayos gamma y protones.
- Rango de energía: 1 TeV - 100 TeV.
- Índice espectral de -1.5.

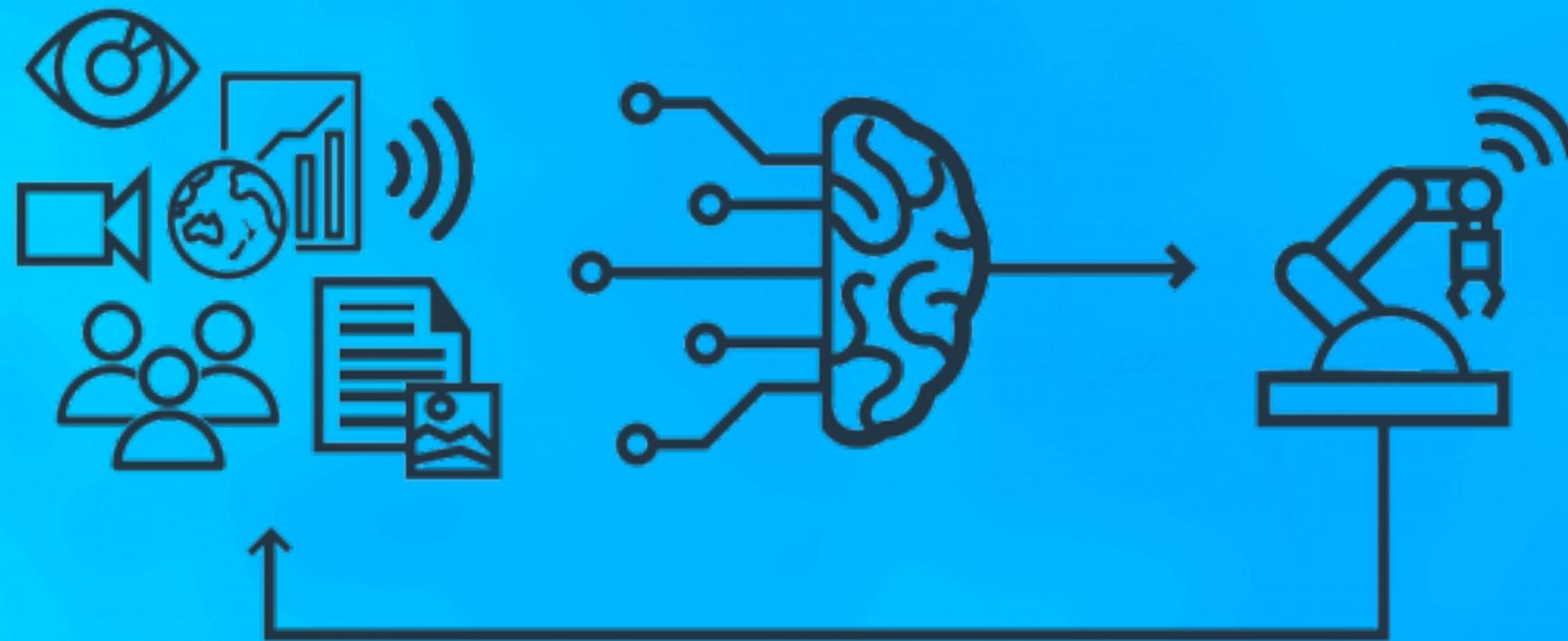
MARS:

- Simulación de la respuesta del detector.
- Calibración de los datos.
- Reconstrucción de los parámetros de las cascadas.

**Parámetros de la  
partícula primaria:  
Energía  
Dirección**

# Machine Learning

- Machine Learning es una rama de la Inteligencia Artificial.
- Los modelos se obtienen a partir de patrones en los datos.
- Existen distintos tipos de aprendizaje.

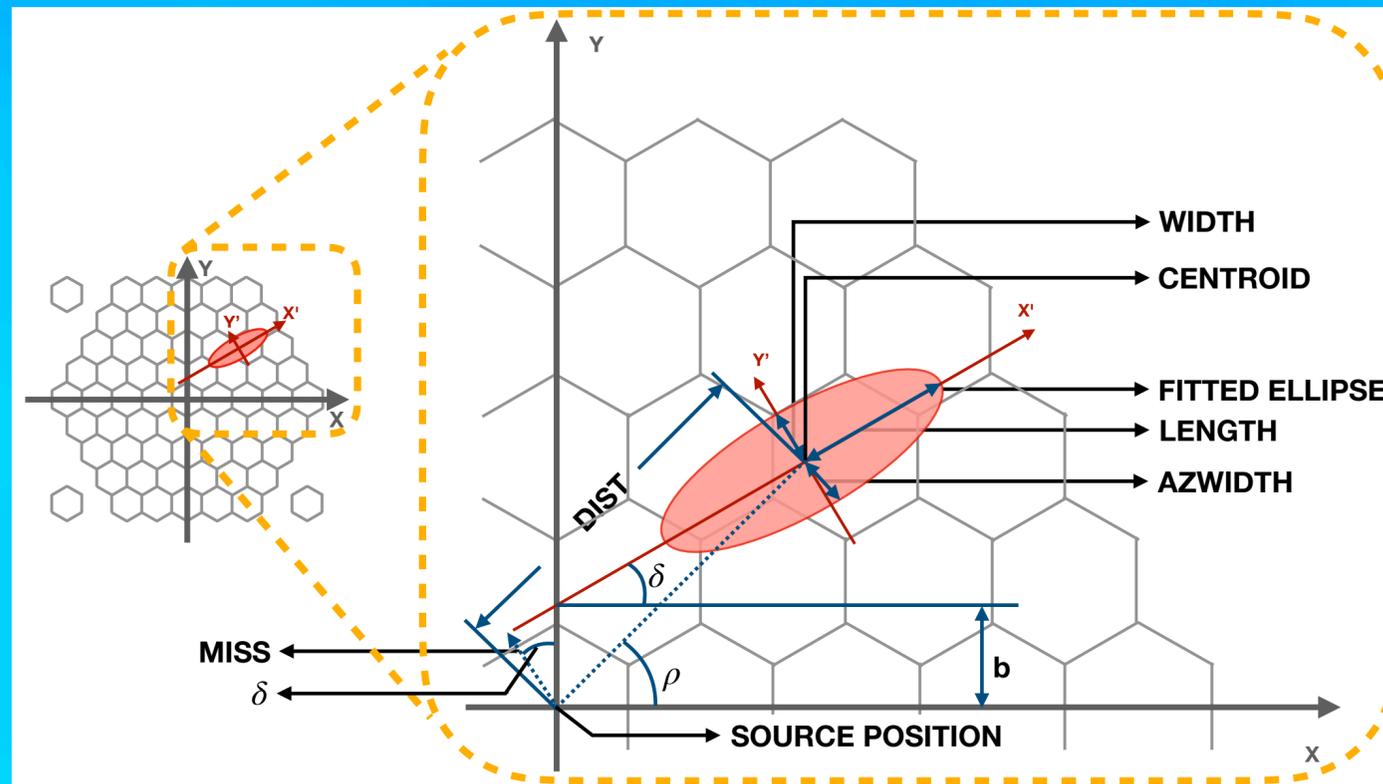


# Selección de Variables de entrenamiento

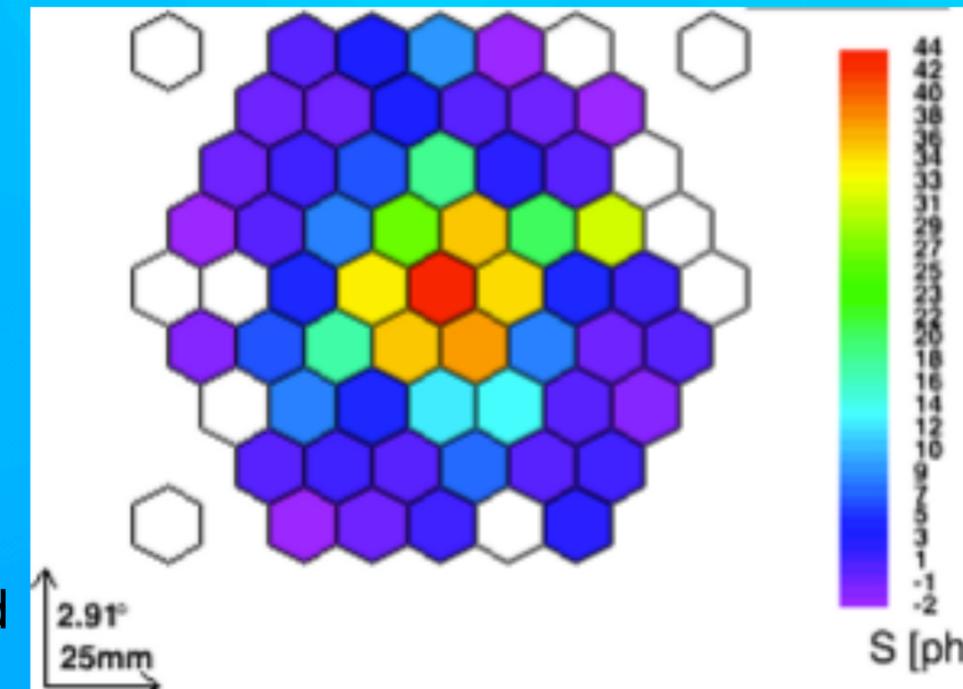
## Parámetros de Hillas

## Variables de tiempo y carga

- Alpha
- Delta
- Dist
- Length
- MeanX
- MeanY
- Width



- Leakage1
- NumUsedPixels
- Size
- SizeMainIsland
- TimeSpread
- TimeSpreadWeighted



# Modelos

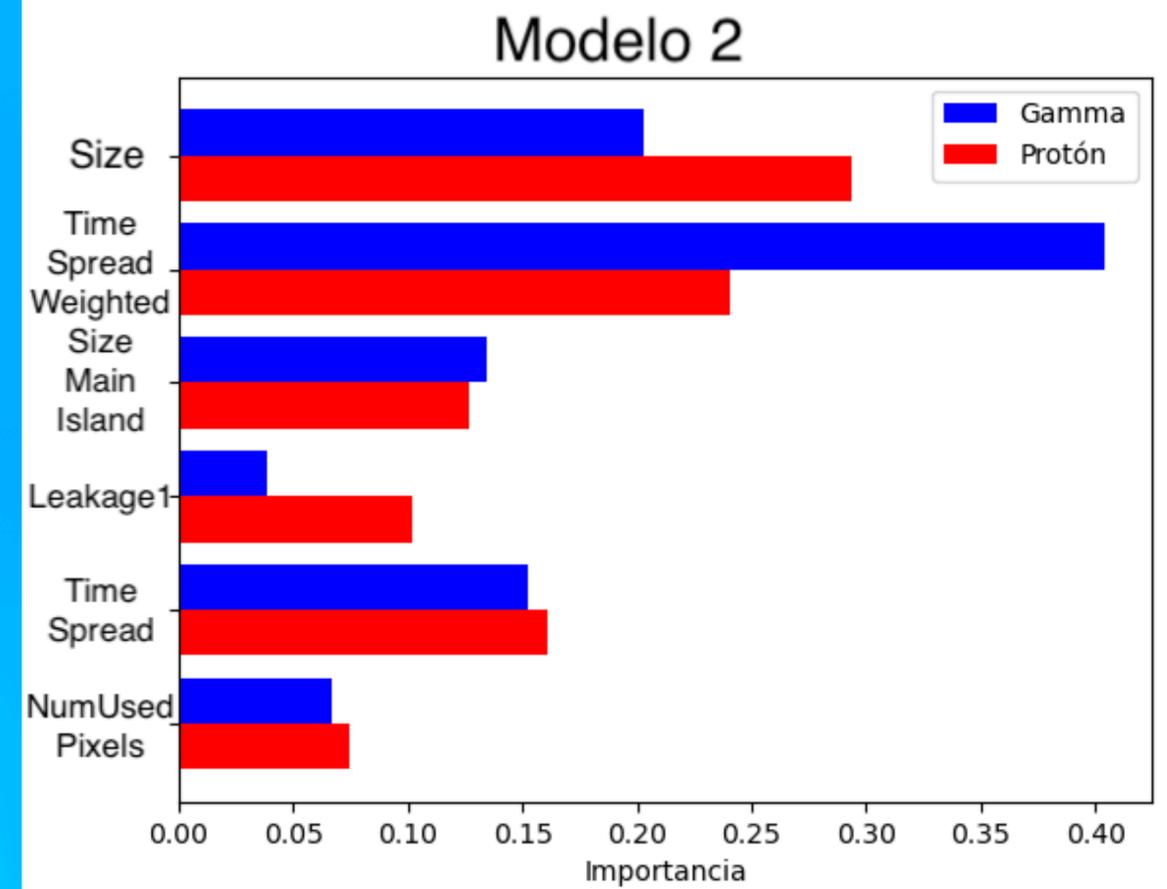
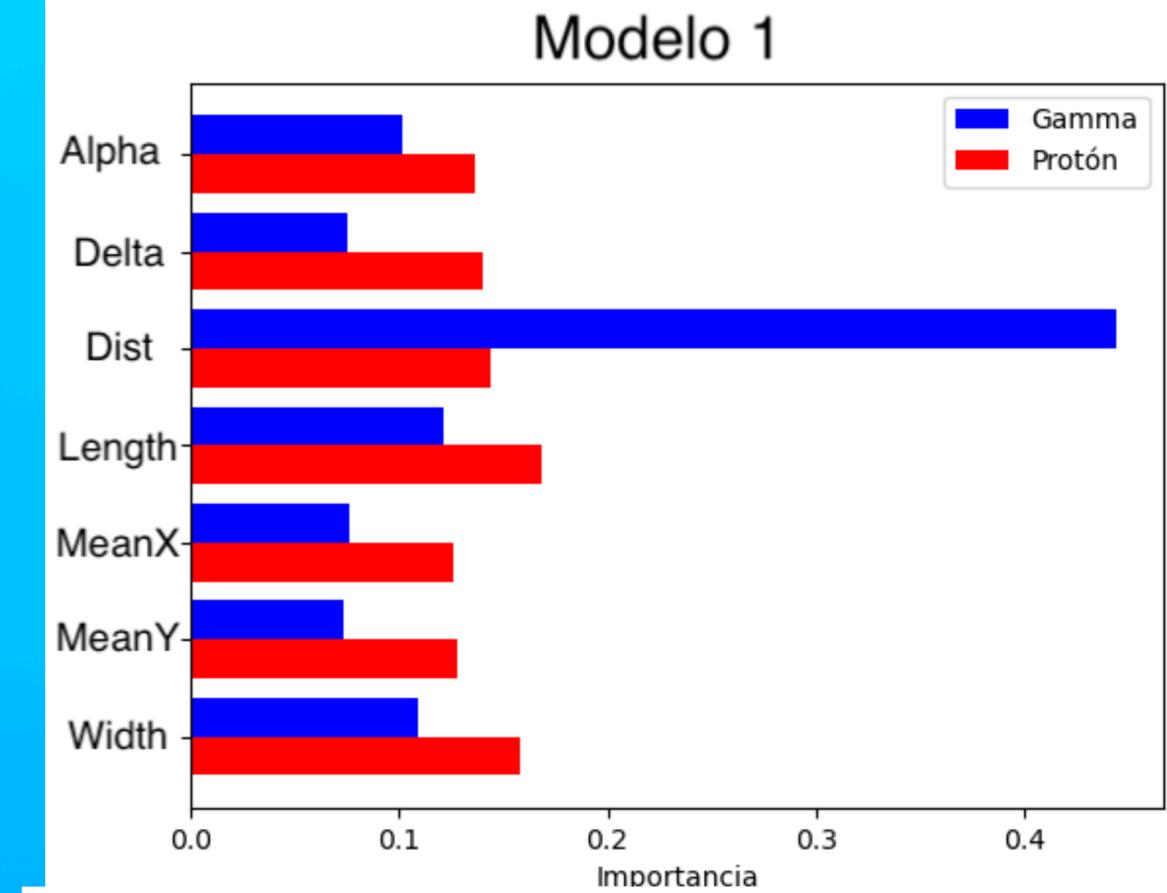
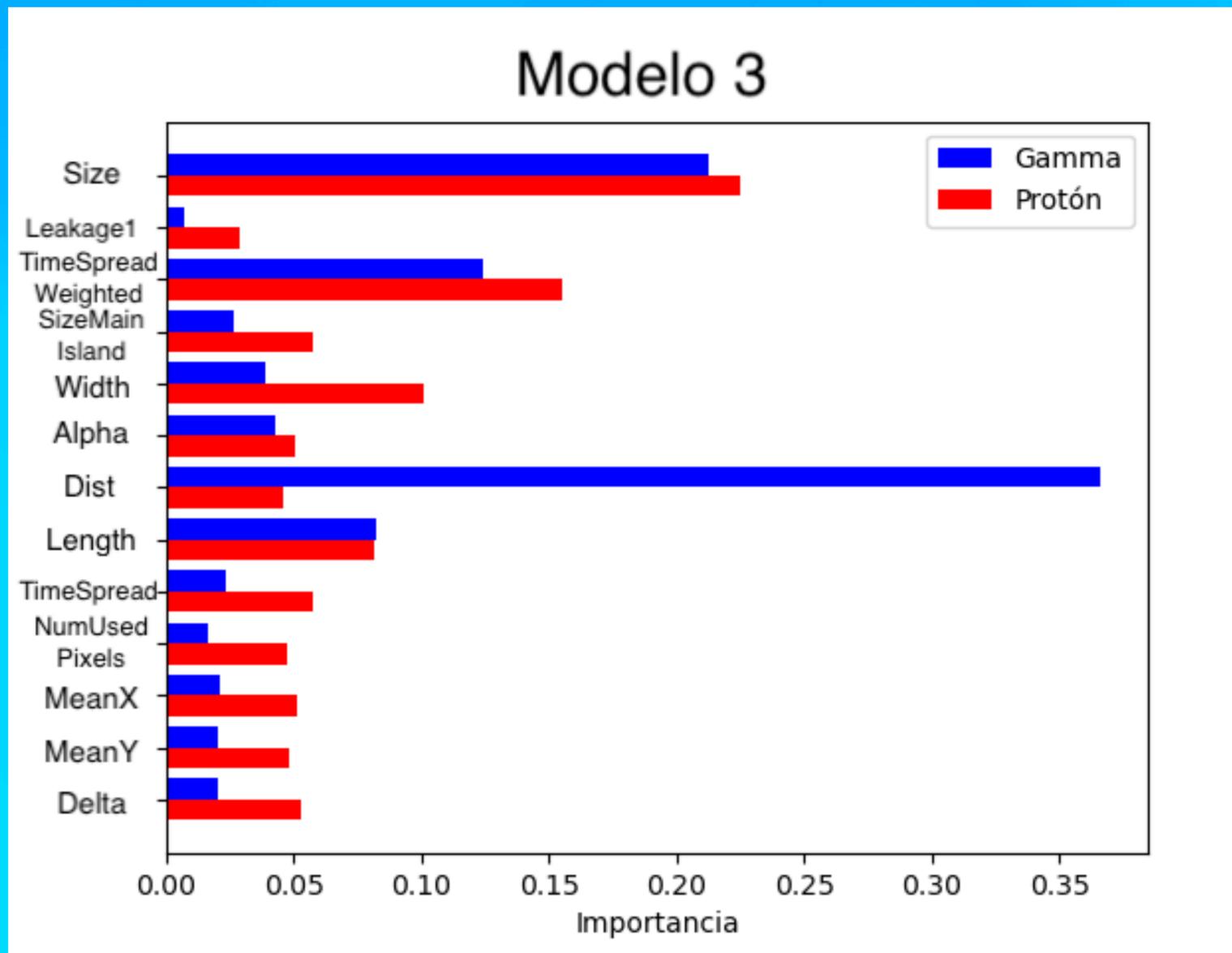
De todos los eventos simulados se utilizó el 70% para entrenar el modelo y el 30% para verificar el rendimiento del modelo.

Se utiliza Random Forest para estimar la Energía de los eventos detectados.

Para este trabajo se utilizaron 3 modelos diferentes para las variables de entrenamiento:

- Modelo 1: Parámetros de Hillas.
- Modelo 2: Variables de tiempo y carga.
- Modelo 3: Combinación entre los modelos 1 y 2.

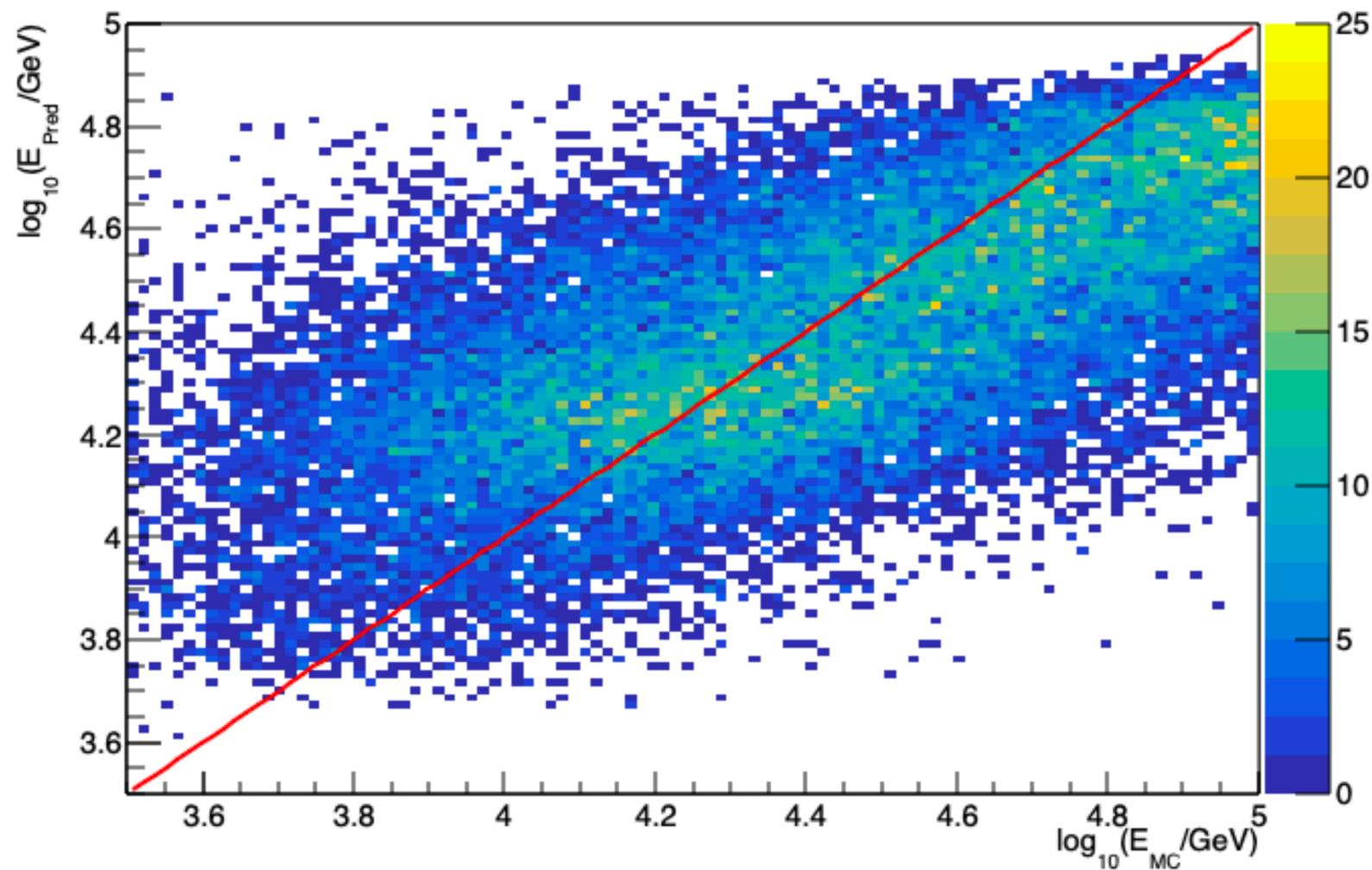
# Importancia de las variables de entrenamiento



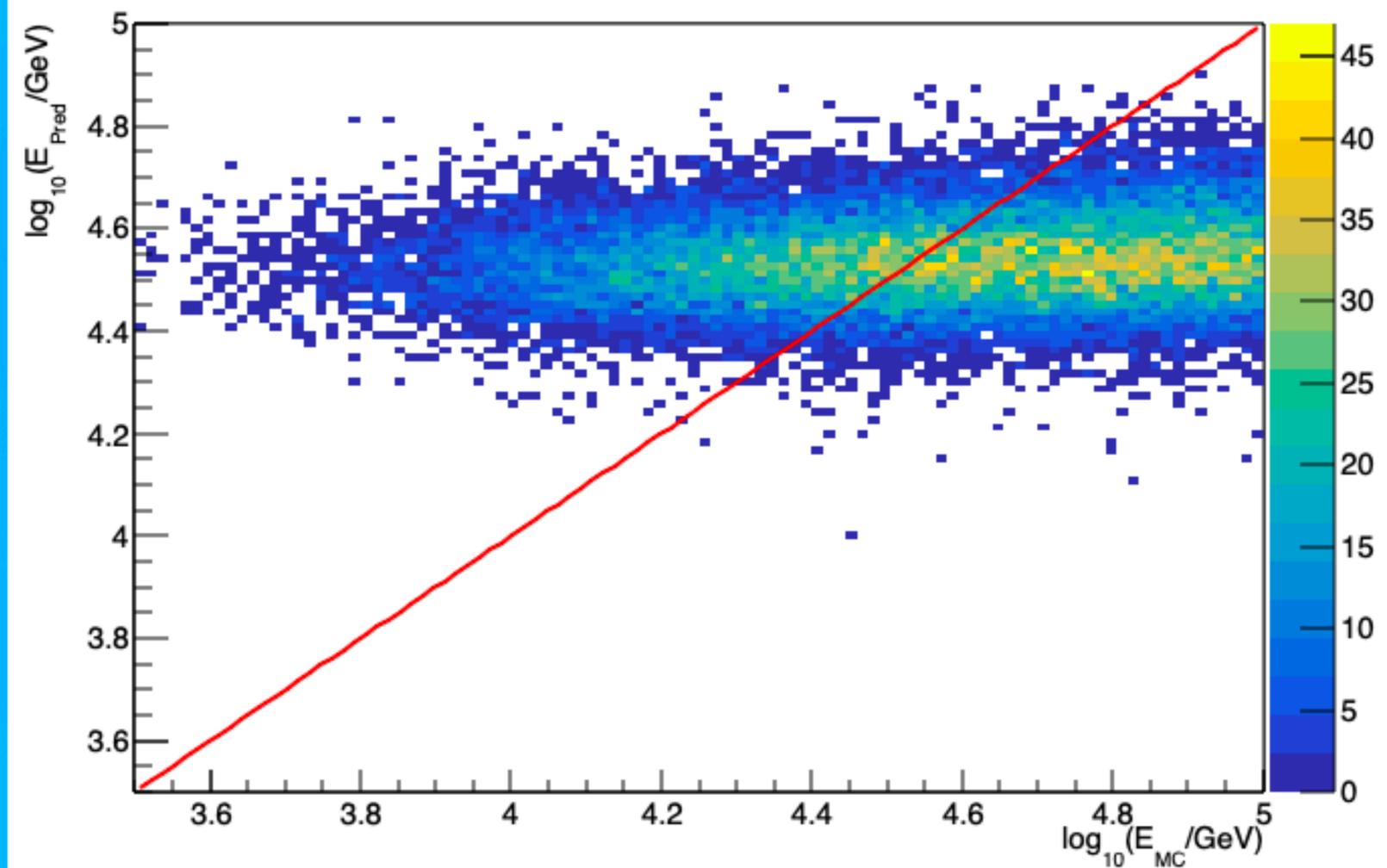
# Modelo 1 (Parámetros de Hillas)

Comparación de la Energía MC vs la Energía Predicha por el modelo

Energía MC vs Energía Predicha Modelo1 - Gamma



Energía MC vs Energía Predicha Modelo1 - Proton



# Modelo 1 (Parámetros de Hillas)

**El Bias es una forma de medir el rendimiento del modelo.**

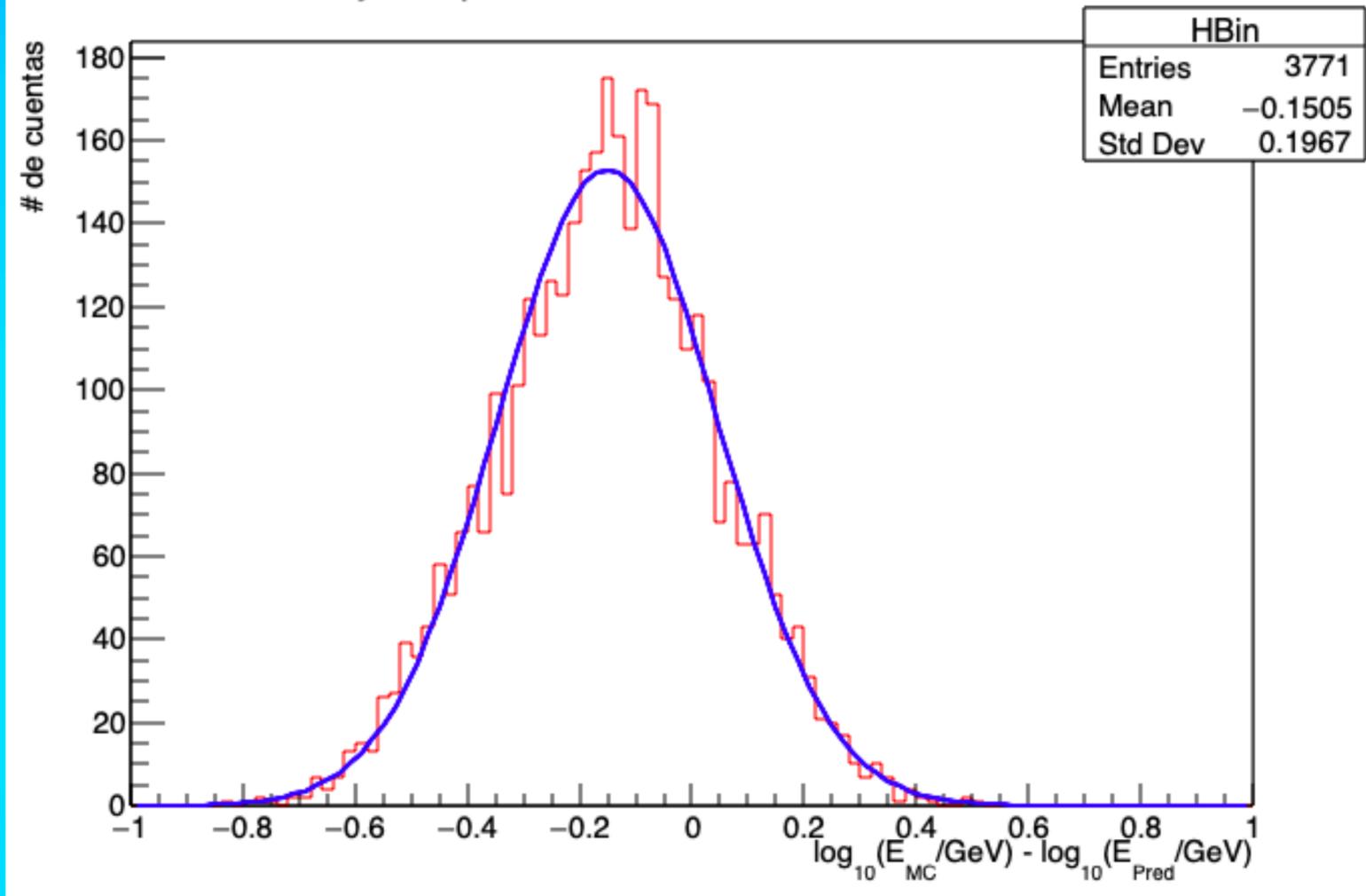
$$Bias = \log_{10}(E_{True}/GeV) - \log_{10}(E_{Pred}/GeV)$$

Bin	Rango de Energía log(E/GeV)
0	3.5-4
1	4-4.2
2	4.2-4.4
3	4.4-4.6
4	4.6-4.8
5	4.8-5

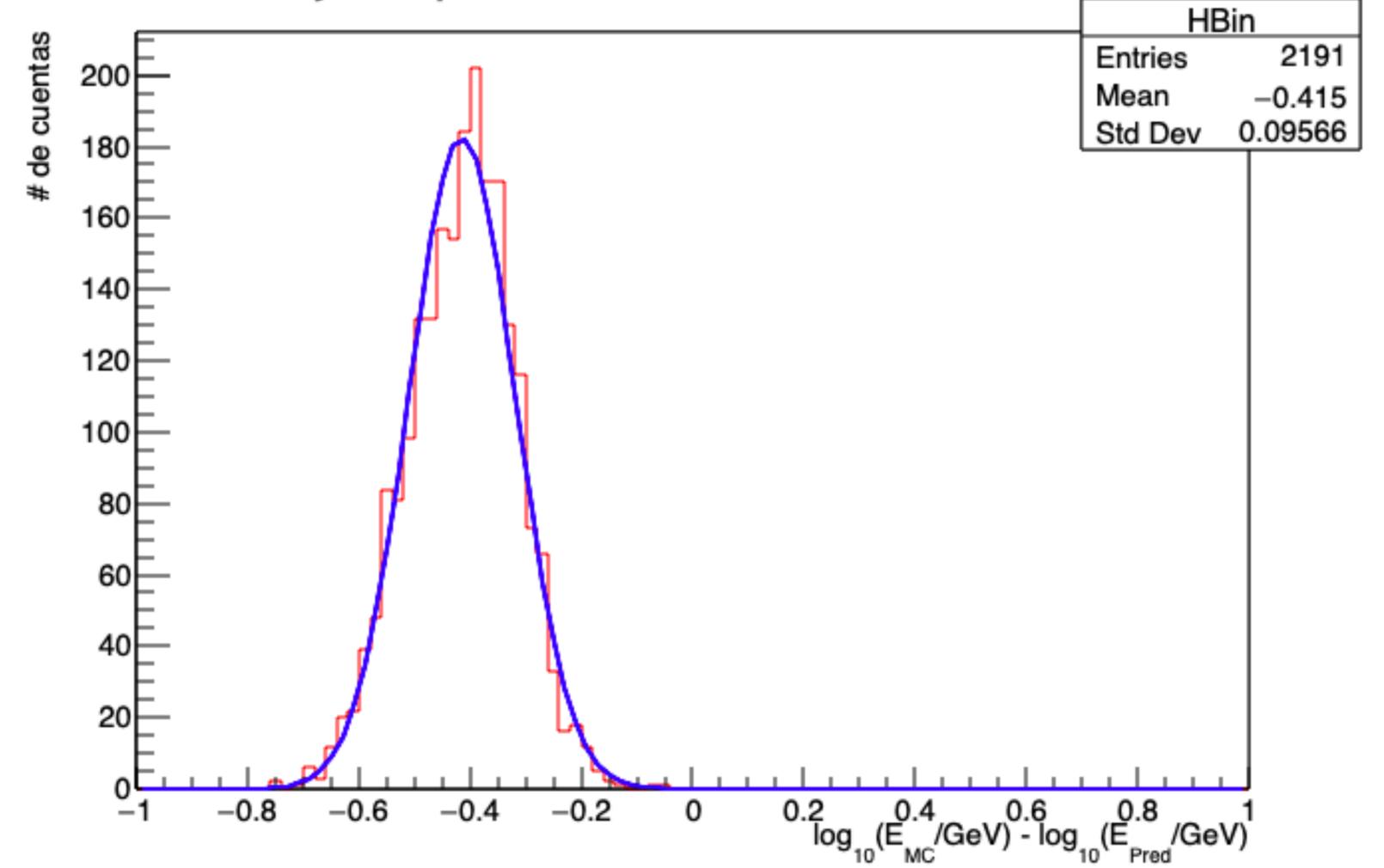
# Modelo 1 (Parámetros de Hillas)

Para cada bin de energía se ajustó una distribución Gaussiana, obteniendo la media y desviación estándar

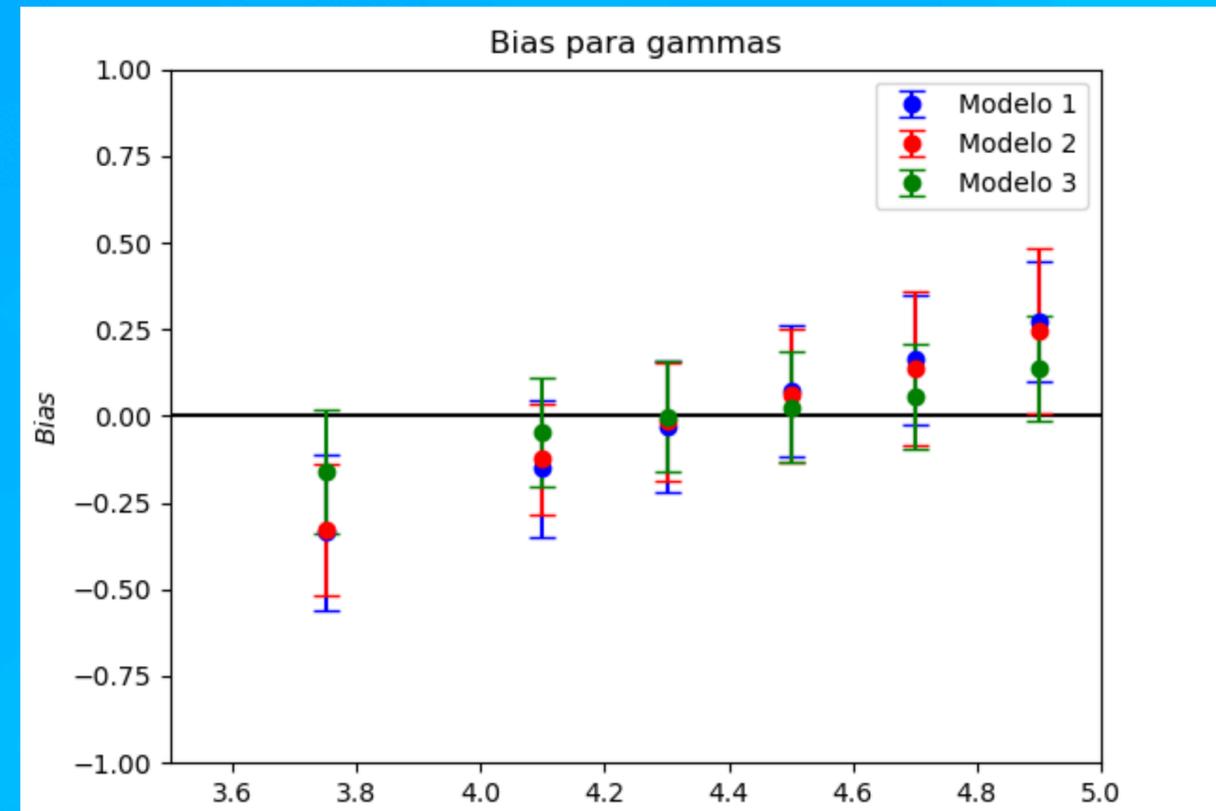
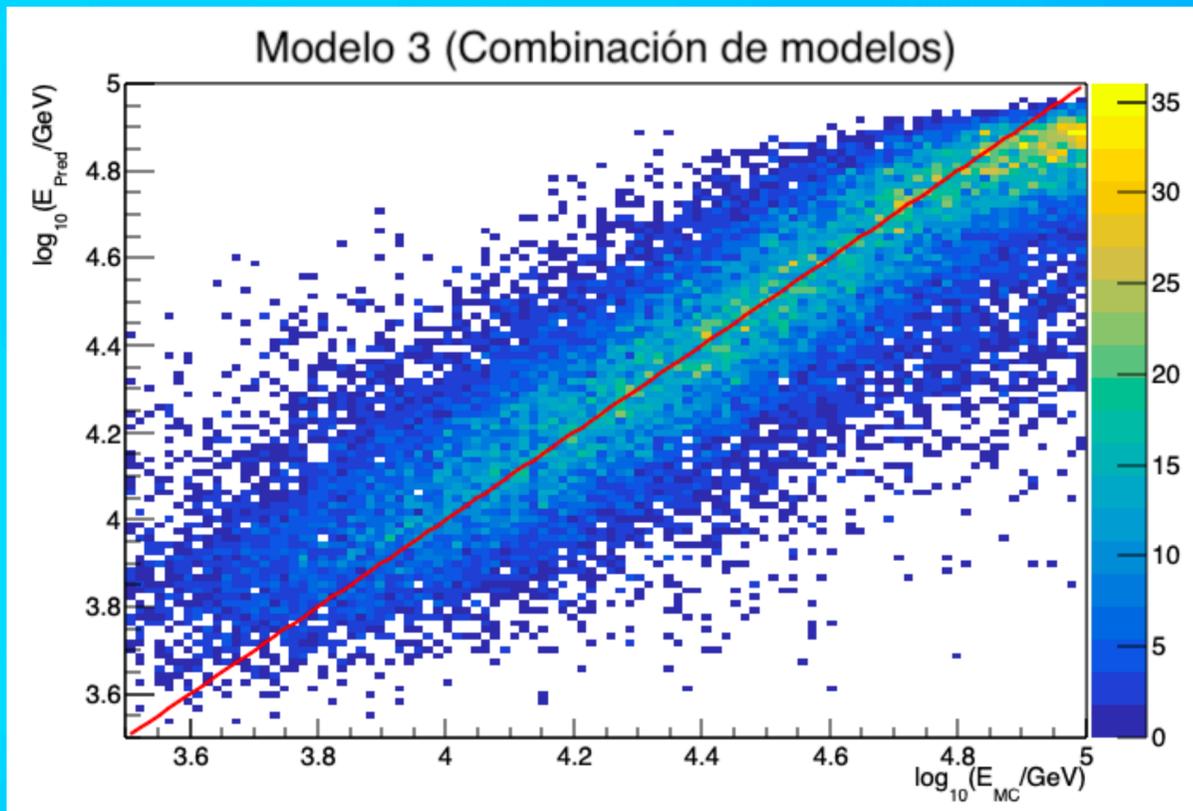
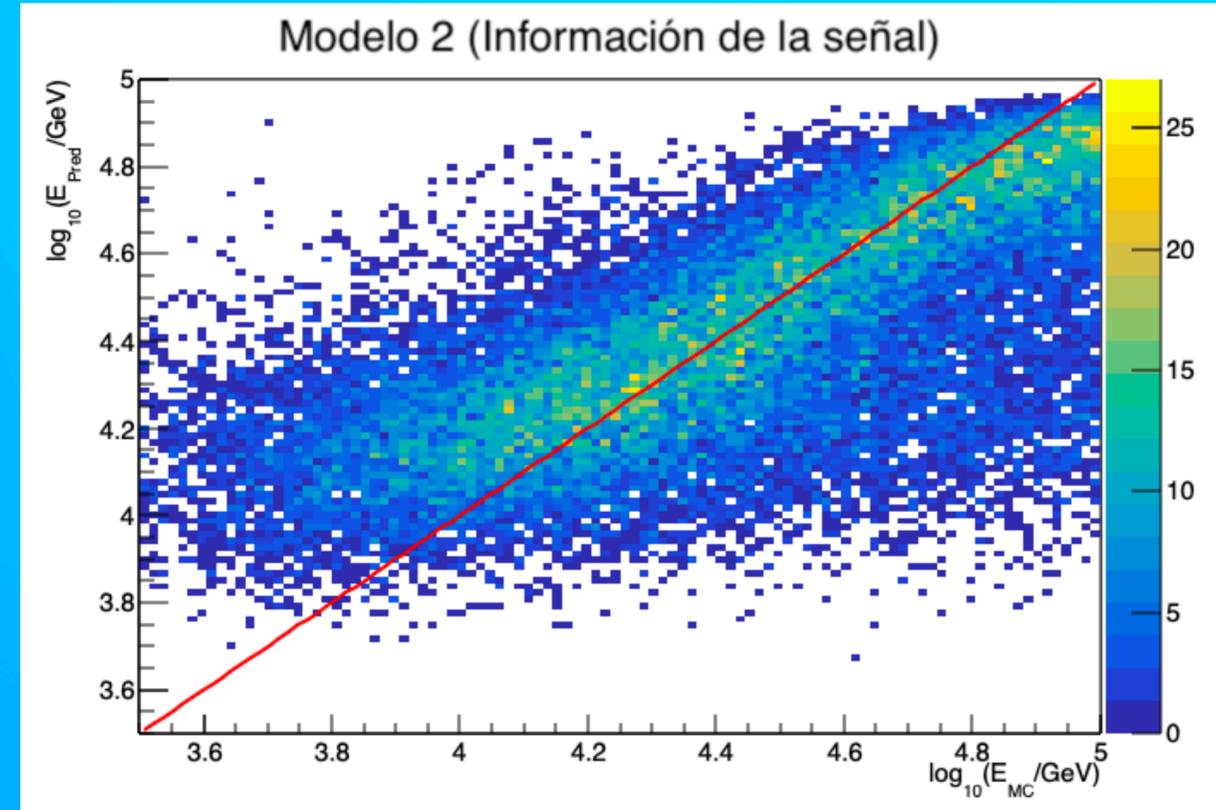
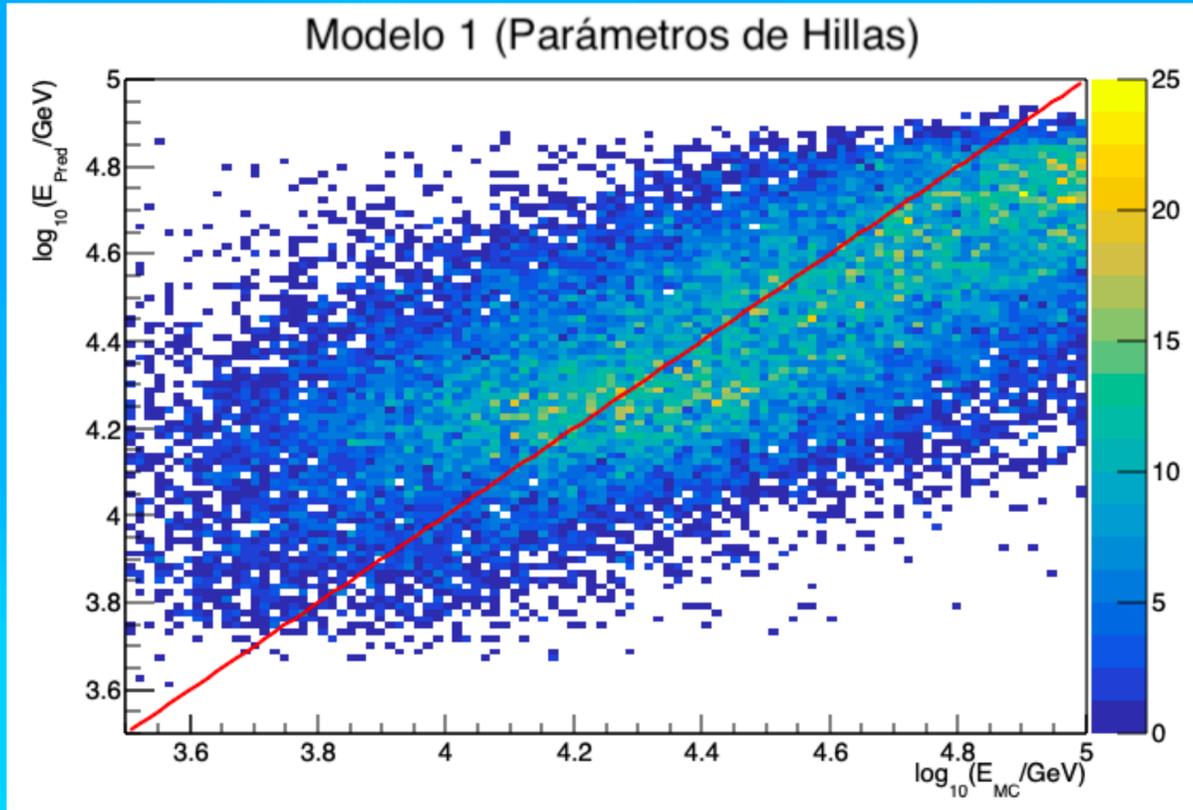
Ajuste para el bin 1 Gamma - Modelo1



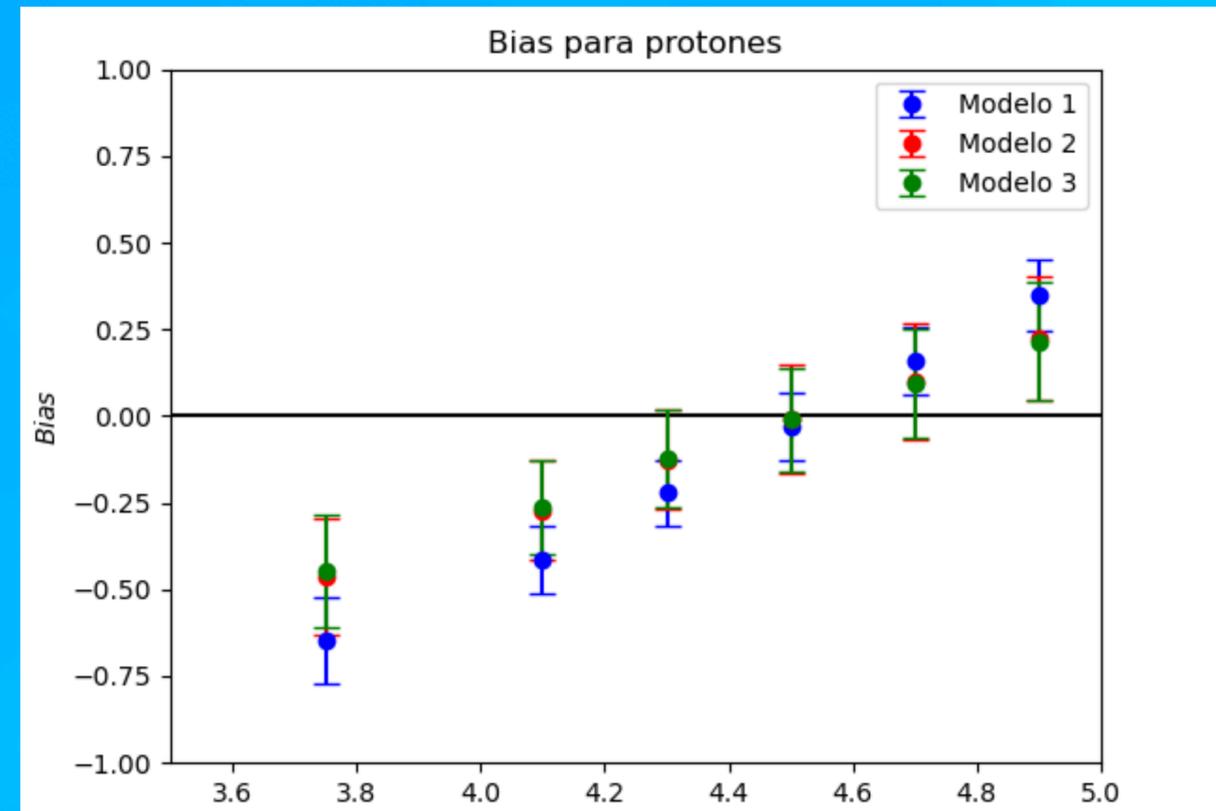
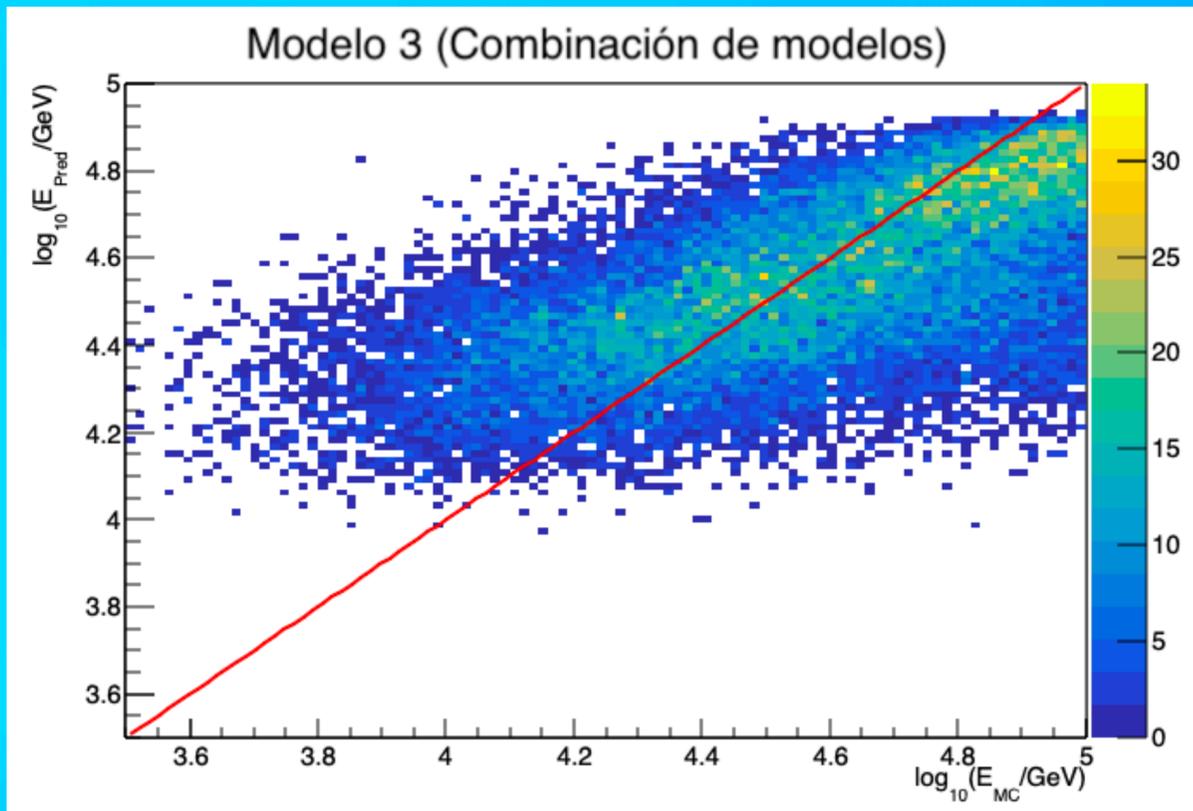
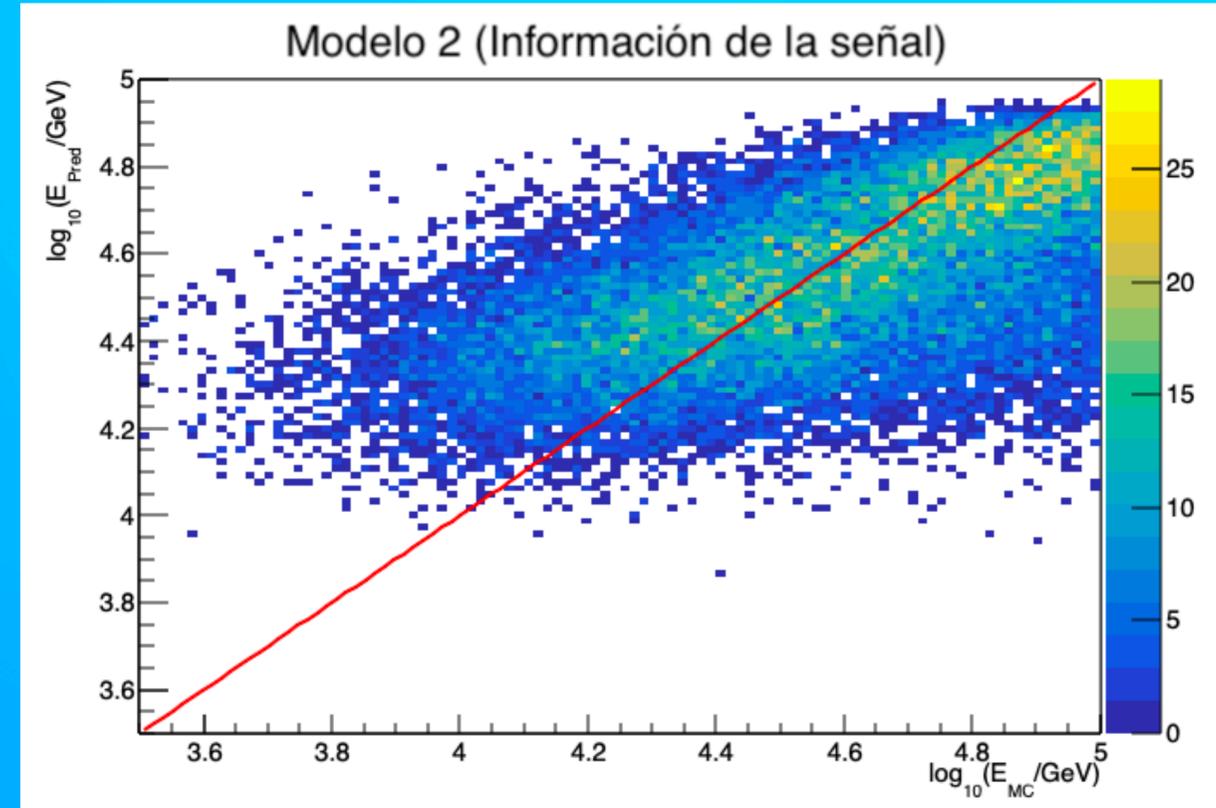
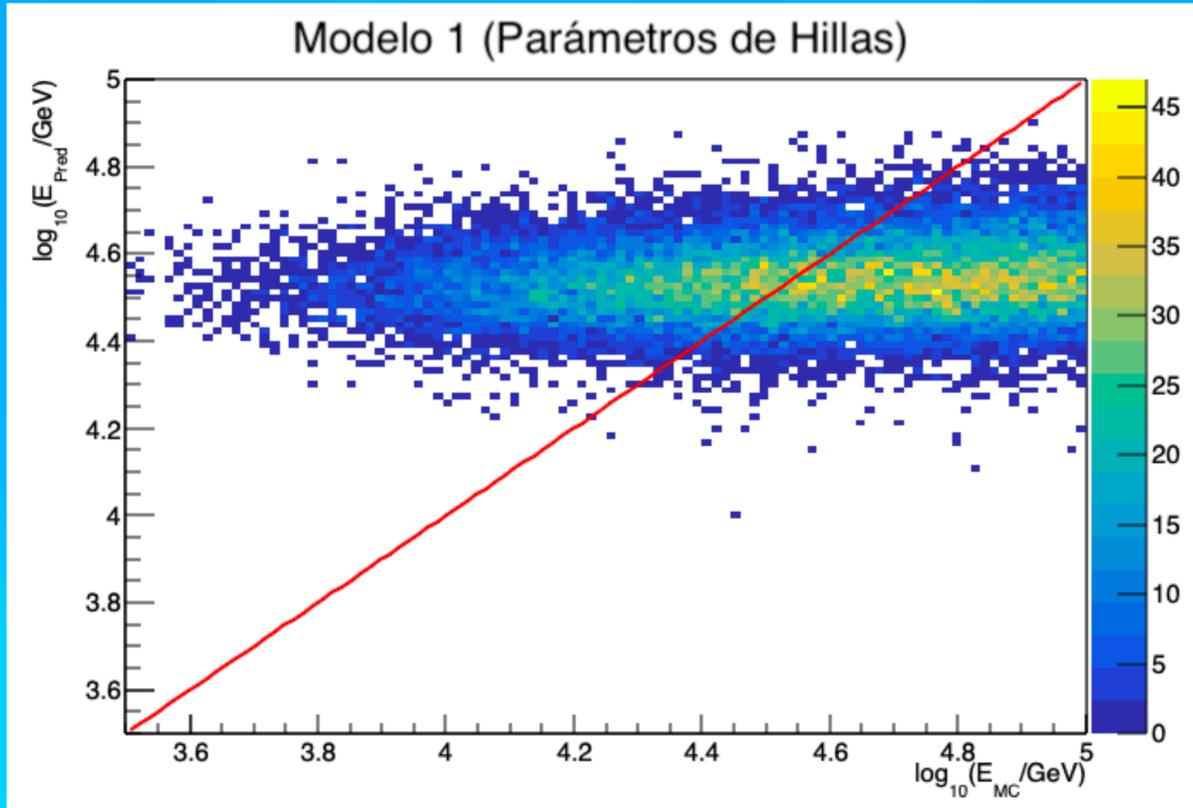
Ajuste para el bin 1 Protón - Modelo1



# Rendimiento para gammas



# Rendimiento para protones



# Resumen

- Utilizar Machine Learning es un campo prometedor para obtener un estimador de energía.
- Combinando las variables de tiempo y carga con los parámetros de Hillas se obtiene un mejor rendimiento para gammas que para protones.

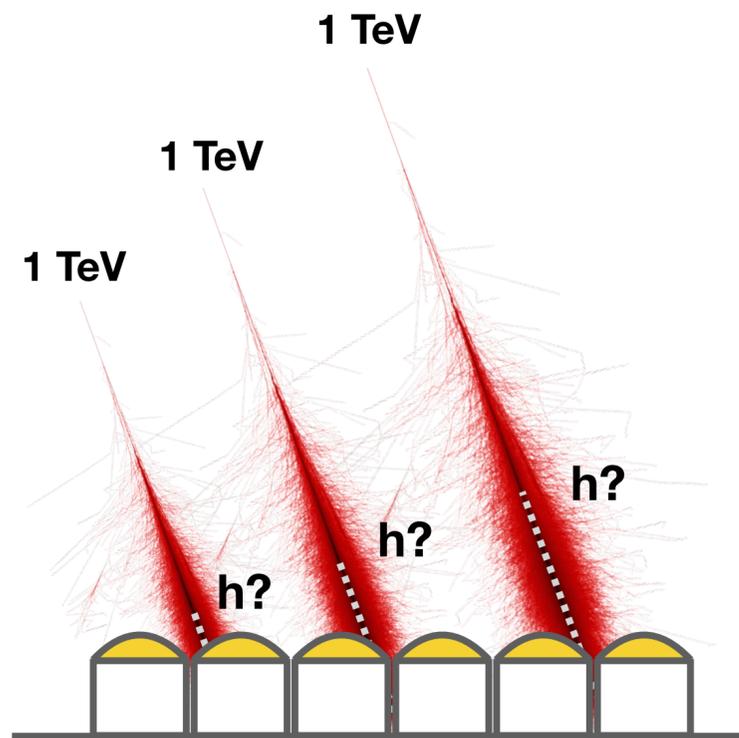
# Trabajo a futuro

- Se crearán nuevas variables de caracterización de eventos.
- Se buscará optimizar las predicciones con Random Forest o mediante el uso de otros modelos de Machine Learning.

# Respaldo

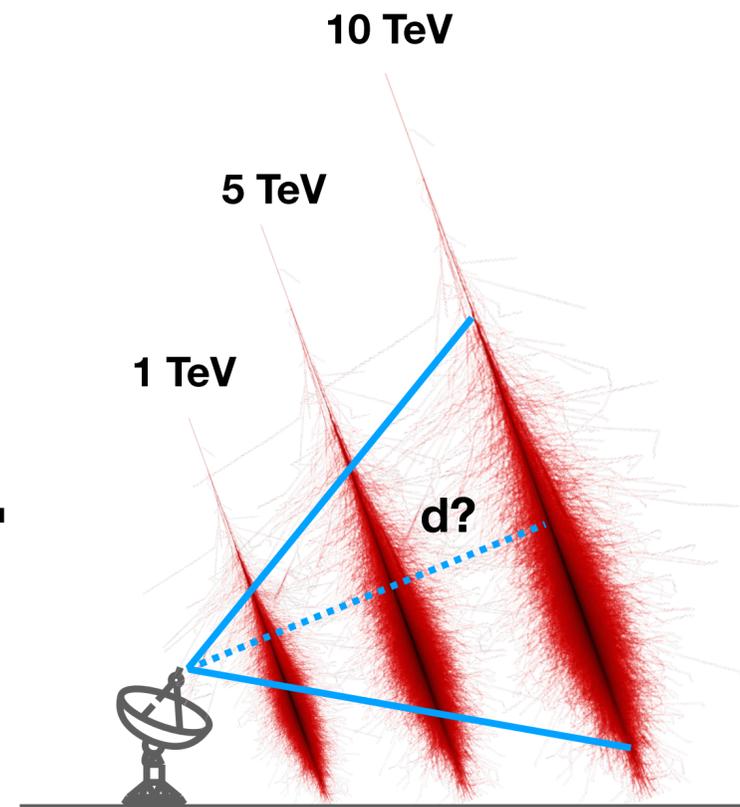
# Sistema híbrido

## DETECTOR DE ARREGLO EXTENDIDO



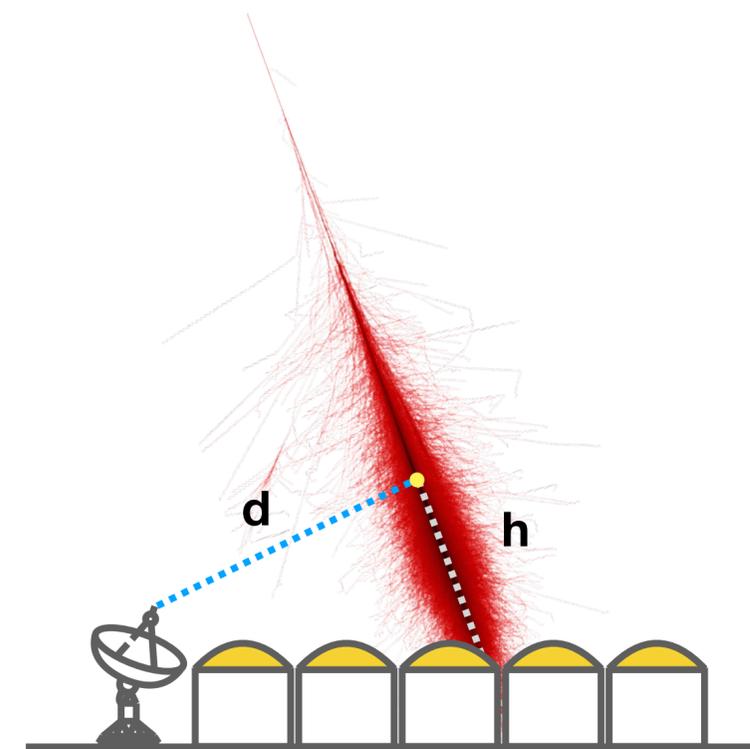
Cascadas idénticas a diferentes alturas ( $h$ )  
► Señales diferentes

## TELESCOPIO DE IMAGEN DE CHERENKOV ATMOSFÉRICO



Diferentes cascadas a diferentes distancias ( $d$ )  
► Señales similares

## DETECTOR HÍBRIDO



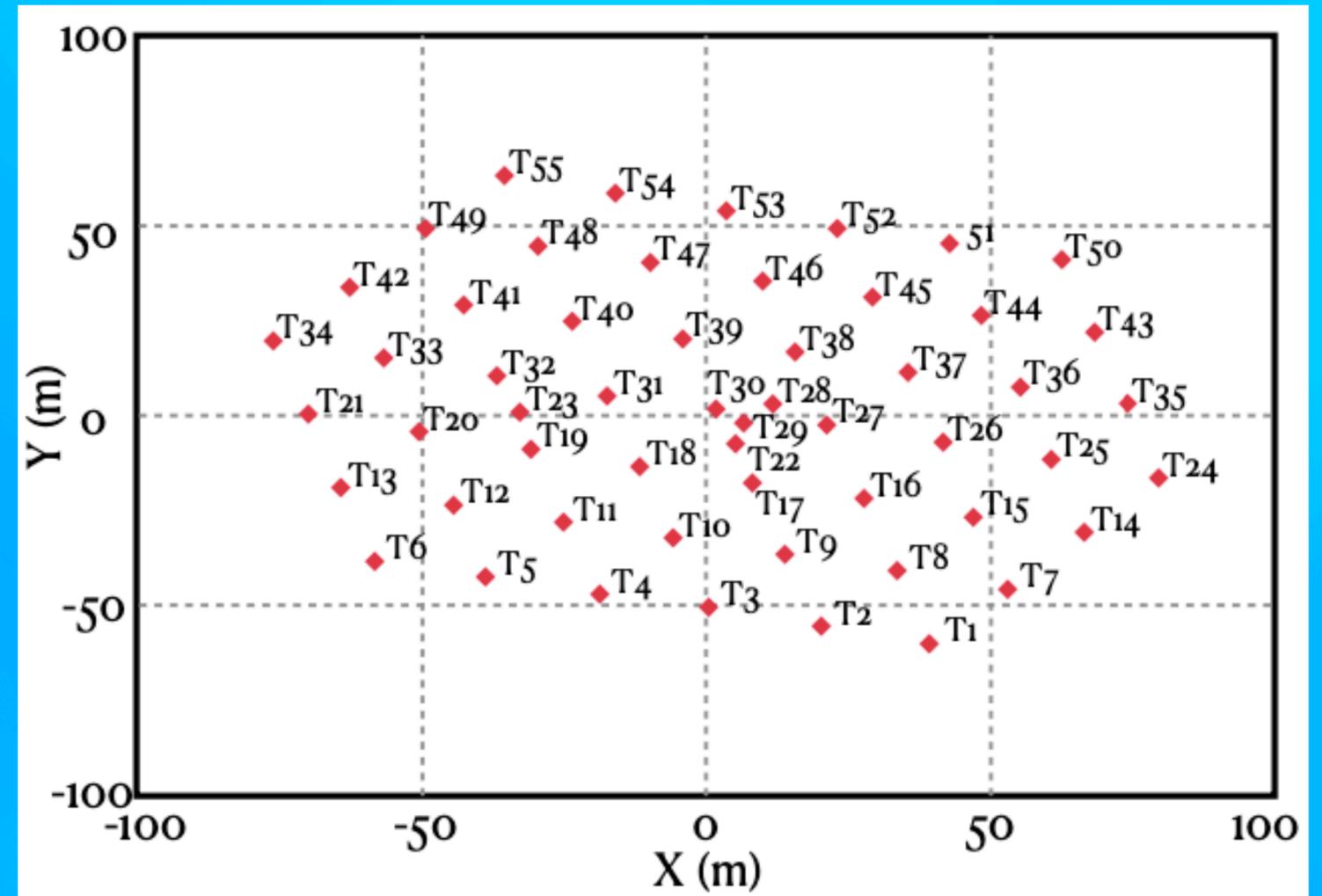
Ambigüedades resueltas  
► Mejor resolución de energía  
► Mejor resolución angular  
► Mejor identificación de partículas

# Configuración de la simulación

Simulación de las cascadas atmosféricas extensas con CORSIKA v7.96:

- Simulación de un arreglo de 55 telescopios HAWC's Eye.
- Partículas primarias:
  - Protones ~ 5 millones de eventos simulados.
  - Gammas ~ 2 millones de eventos simulados.
- Rango de energía: 1 TeV a 100 TeV.
- Altitud: 4100 m. s. n. m.
- Modelo de la atmósfera: Atmósfera estándar de US (7).
- Cono de visión:  $8^\circ$ .
- Área de dispersión de las cascadas simuladas:  $500 \times 500 \text{ m}^2$ .

Simulación de la respuesta del detector y reconstrucción de la imagen de la cámara con MARS.



# Configuración del Random Forest

- Número de estimadores: 200
- Profundidad máxima: Es expande hasta que las hojas sean puras.
- Mínimo de muestras para dividir: 2.
- Mínimo de muestras por hoja: 1.
- Máximo número de variables por árbol: Número de variables de entrenamiento.

