

Estimación de Energía de Rayos Cósmicos para el Observatorio HAWC aplicando algoritmos de Inteligencia Artificial



Universidad
Industrial de
Santander



Jorge Alfredo Jaimes Teherán – jjalfredo68@gmail.com

Director: Christian Sarmiento Cano

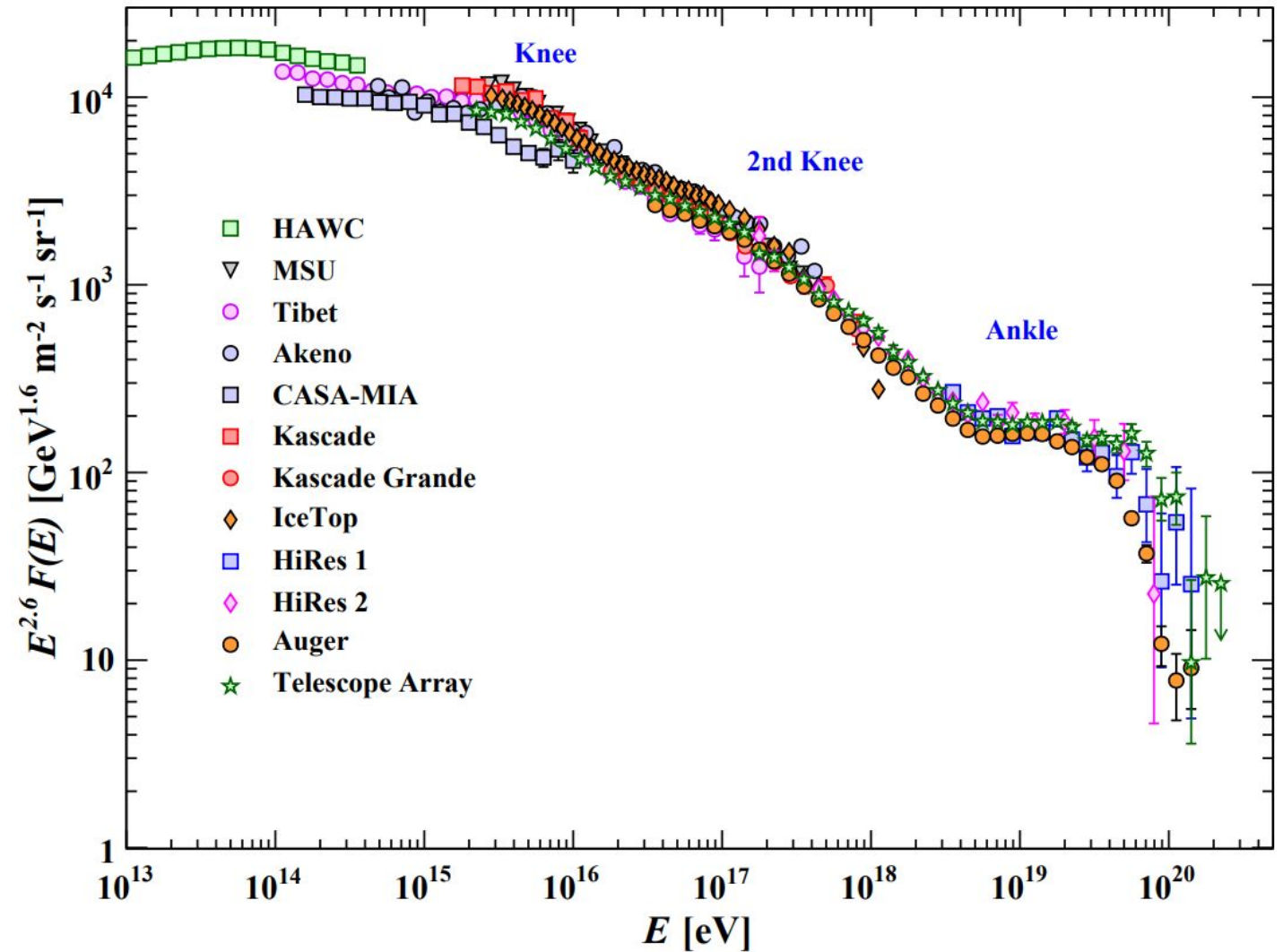
Codirectores: Ibrahim Torres, Tomás Capistrán, Luis Nuñez

Universidad Industrial de Santander
Instituto Nacional de Astrofísica, Óptica y Electrónica
Università degli Studi di Torino



Espectro de rayos cósmicos

- La rodilla
- Segunda rodilla
- Tobillo
- Supresión

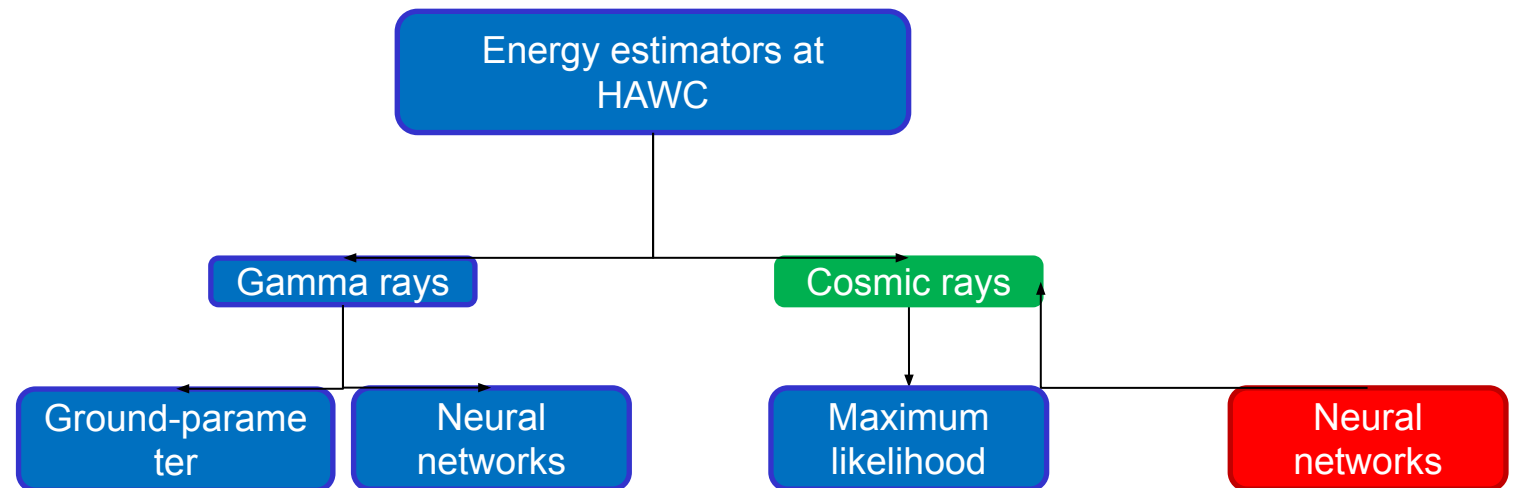
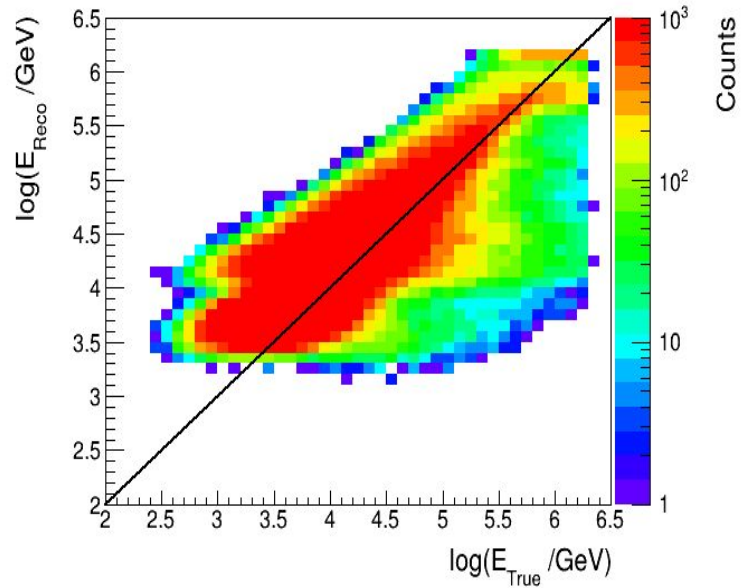


HAWC

- Situado en el volcán Sierra Negra
- 100 GeV hasta ~ 1 PeV (RC)
- 300 Detectores Cherenkov en agua
- 4 foto multiplicadores por tanque



Importancia de la estimación de energía de los rayos cósmicos

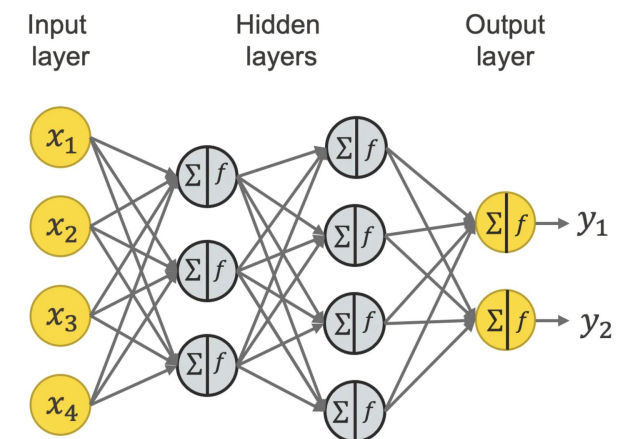
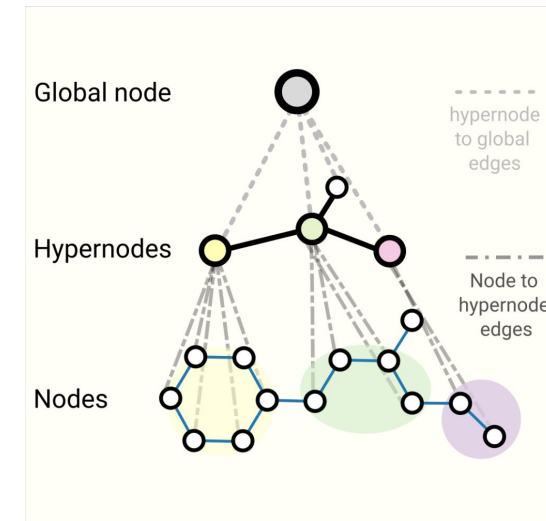


Machine learning

- Técnicas comunes: **Árboles de decisión**, **Máquinas de soporte vectorial**, **máquinas de gradiente de boosting**
- Son efectivos incluso con conjunto de datos pequeños o medianos

Deep Learning

- Suele requerir **grandes cantidades de datos** para su entrenamiento.
- Se utilizan **redes neuronales profundas (DNNs)**, **Redes Convolucionales (CNNs)** y **Redes Neuronales basadas en grafos (GNNs)**.

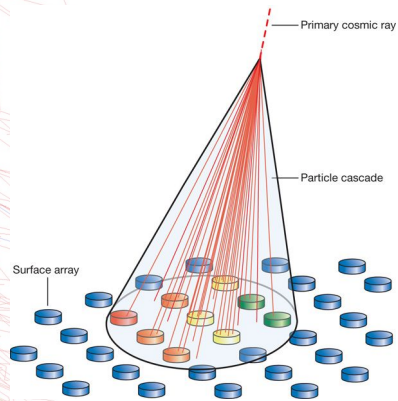


Construcción del modelo

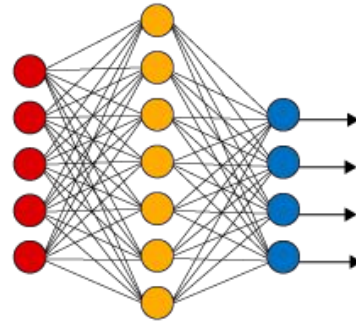
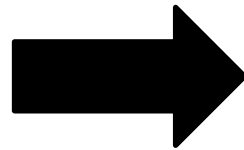
Entrenamiento del modelo en eventos simulados de **Monte Carlo**

Entrenamos en **simulaciones**, pero usamos en **datos reales**

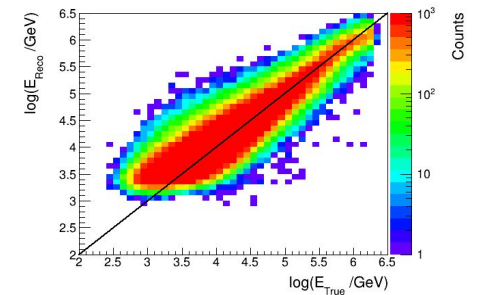
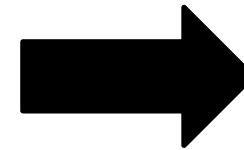
Las estimaciones mediante inteligencia artificial mejoran la estimación de la energía frente a métodos tradicionales.



Datos simulados



Red neuronal

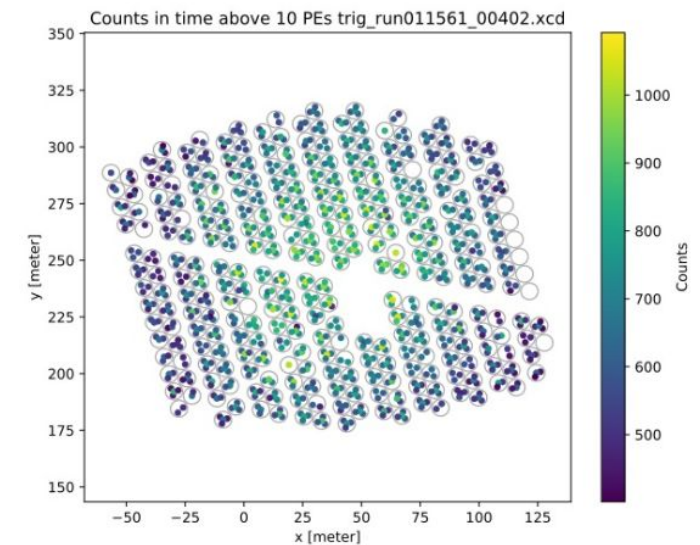
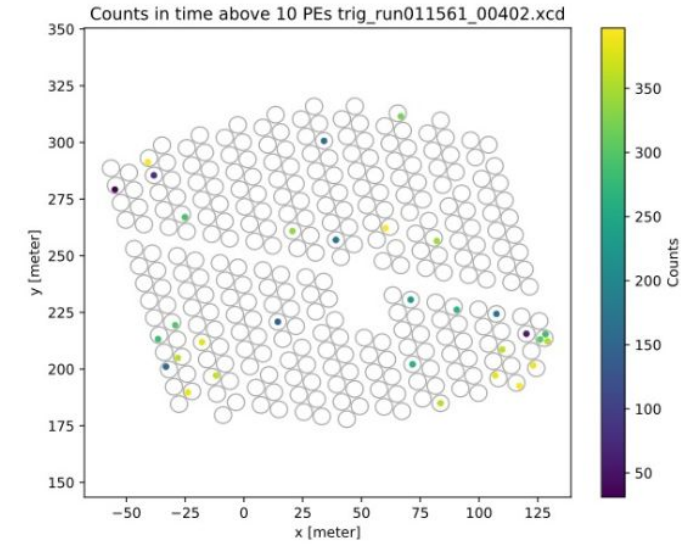
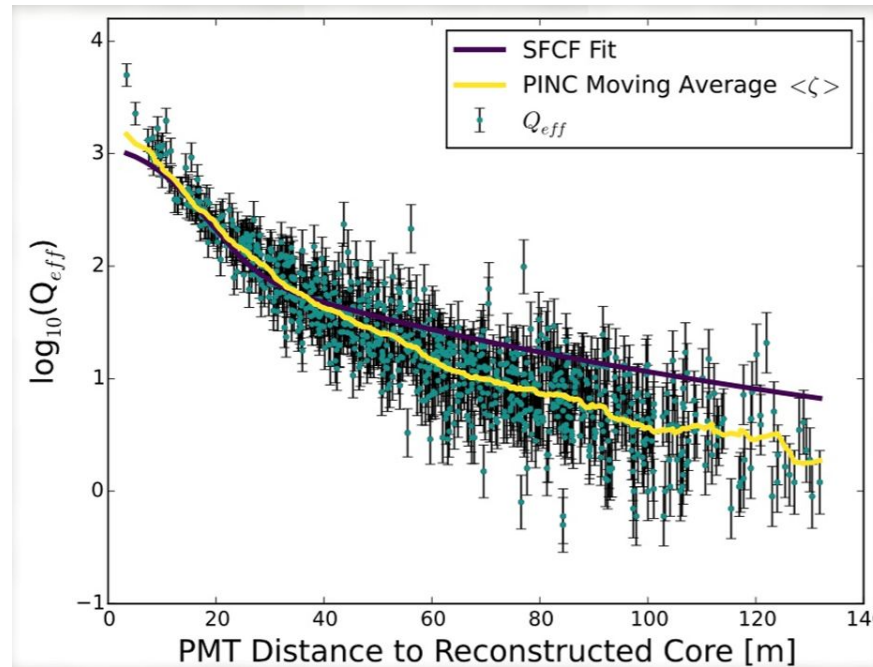


Reconstrucción de energía

Variables físicas



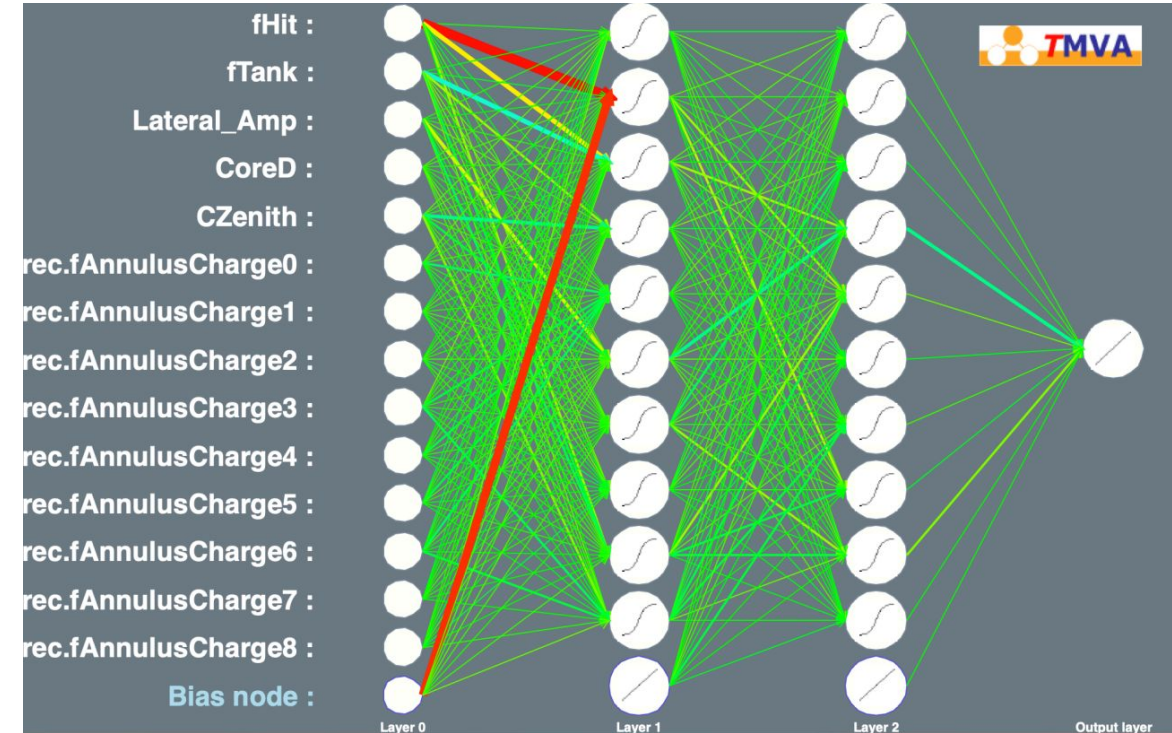
- fHit
- fTank
- Lateral_Amp
- CoreD
- CZenith
- fAnnulusCharge (9 variables)
- 22 variables +



- Actualizar estimadores de energía para rayos cósmicos en el HAWC
- Mejora continua
- Reducción de sesgos

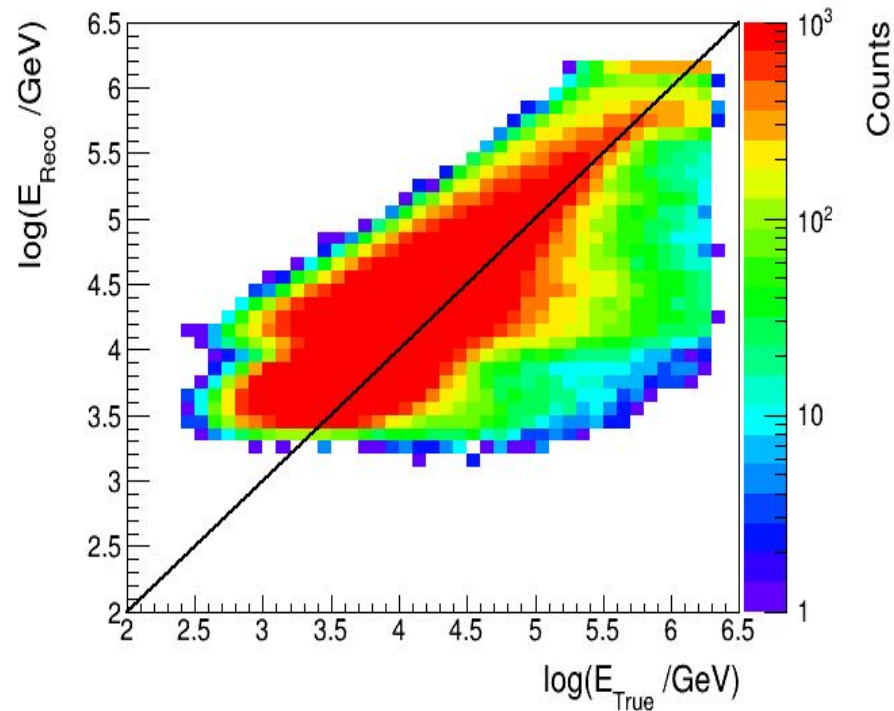
Arquitectura

- Entrenamiento con eventos simulados de 5,181,079 protones
- 2 capas ocultas de 10 neuronas

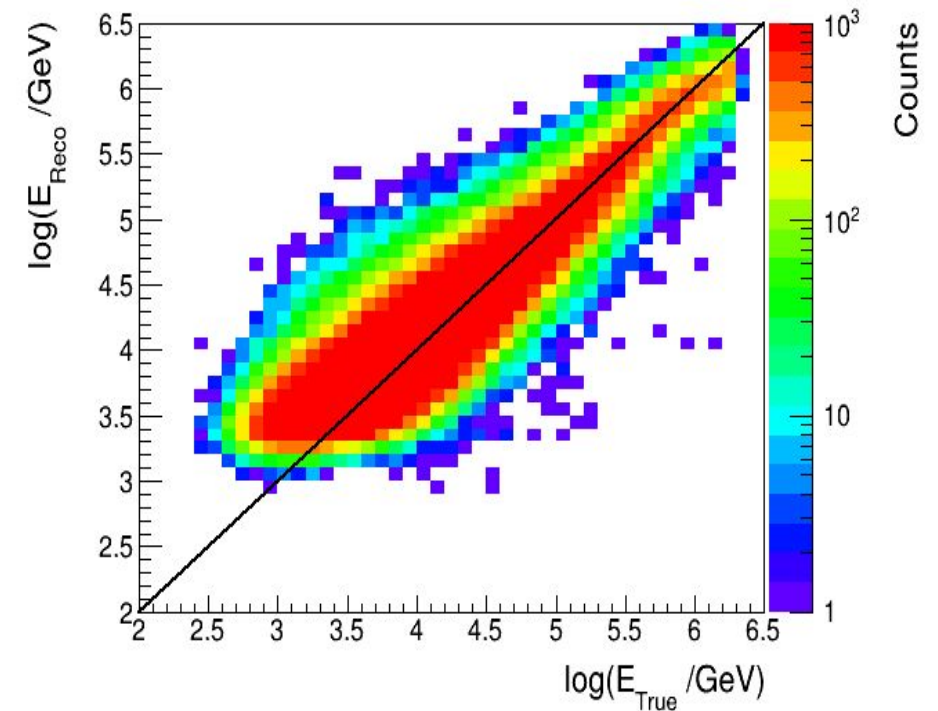


Método clásico VS Red neuronal

Método clásico

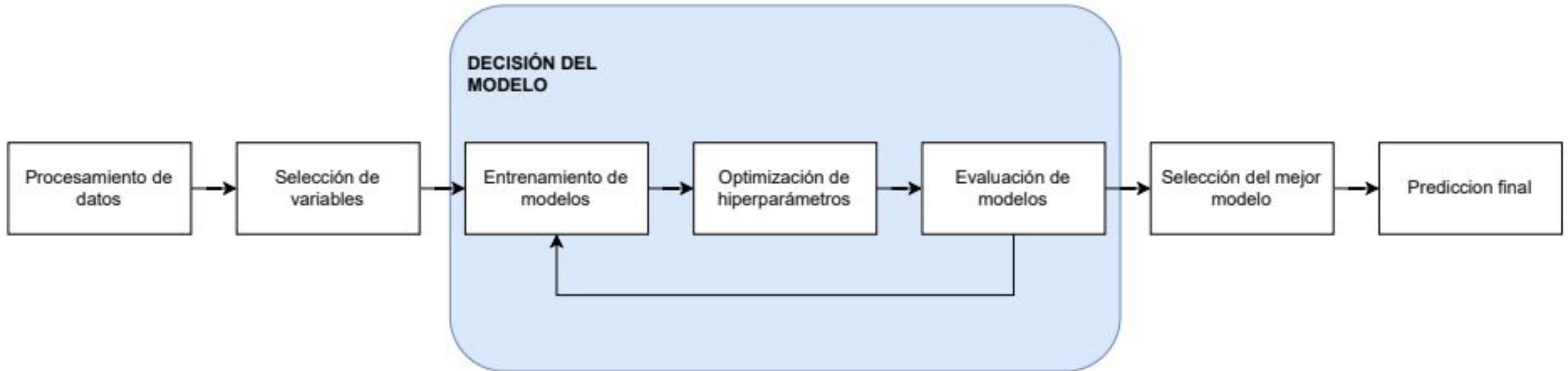


Red neuronal



Ambos casos muestran la estimación de energía para una red entrenada para protones

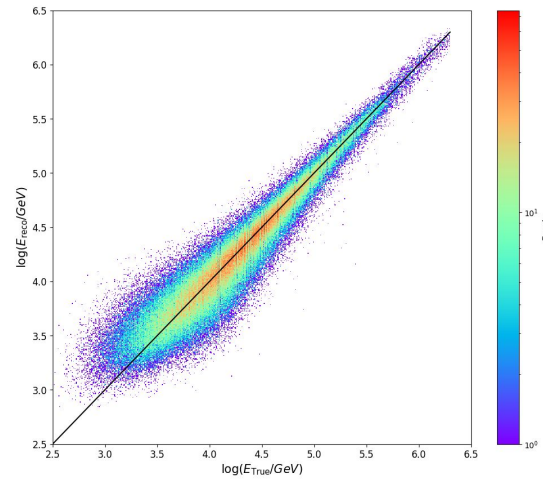
Pipeline de entrenamiento



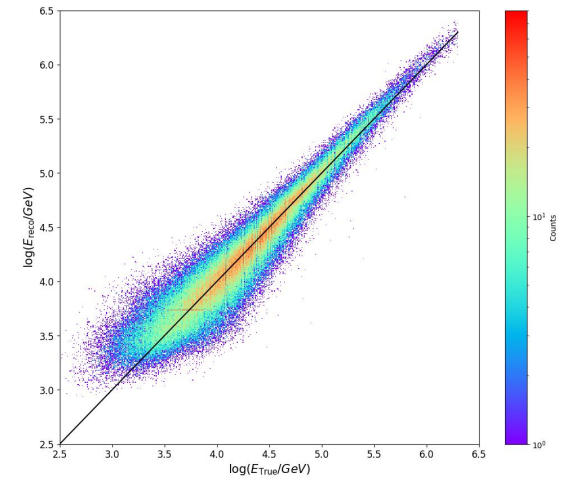


Modelo	Conjunto	MSE	MAE	R^2
MLP	Entrenamiento	0.0276	0.1212	0.9182
	Prueba	0.0279	0.1218	0.9109
XGBoost	Entrenamiento	0.0250	0.1165	0.9201
	Prueba	0.0281	0.1225	0.9101
LightGBM	Entrenamiento	0.0262	0.1192	0.9163
	Prueba	0.0282	0.1228	0.9099
CNN	Entrenamiento	0.0365	0.1518	0.8832
	Prueba	0.0367	0.1522	0.8825

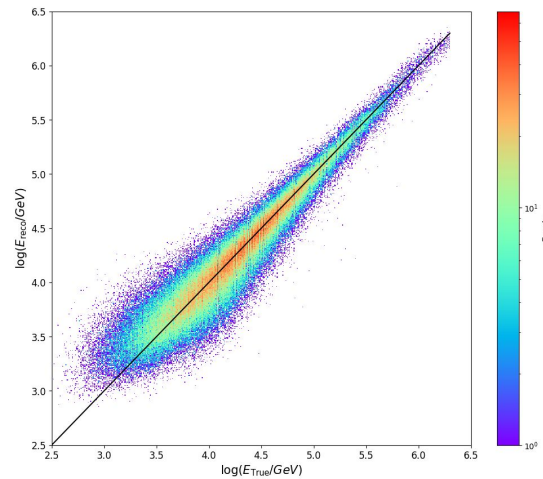
MLP



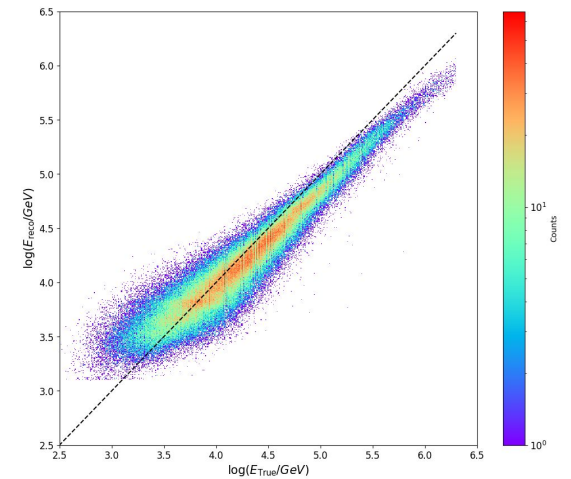
XGBOOST



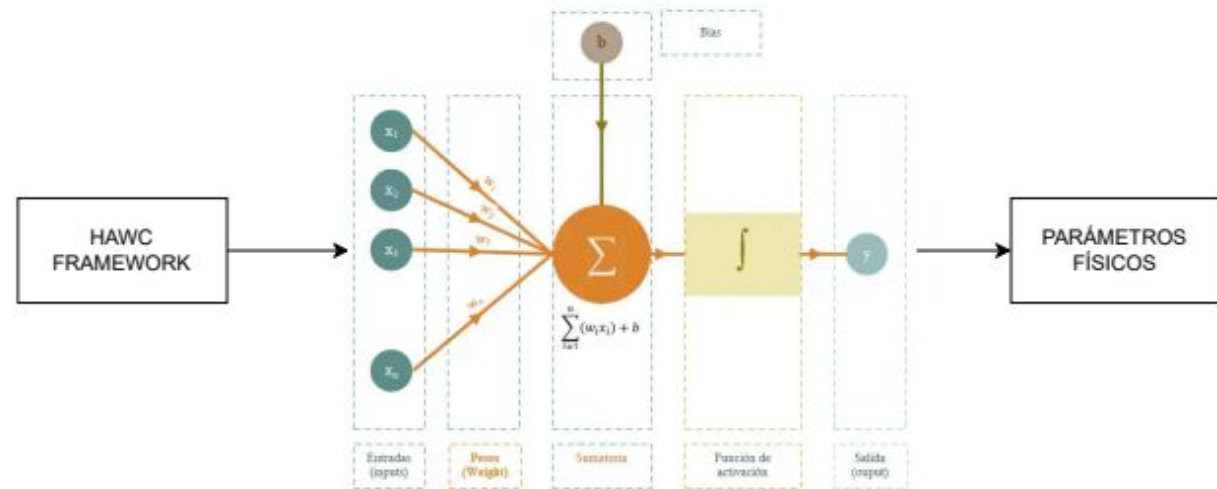
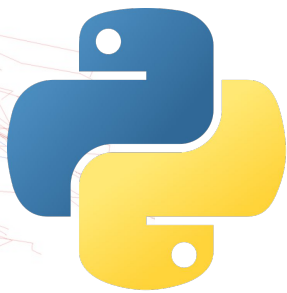
LGBM



CNN



Integración en HAWC

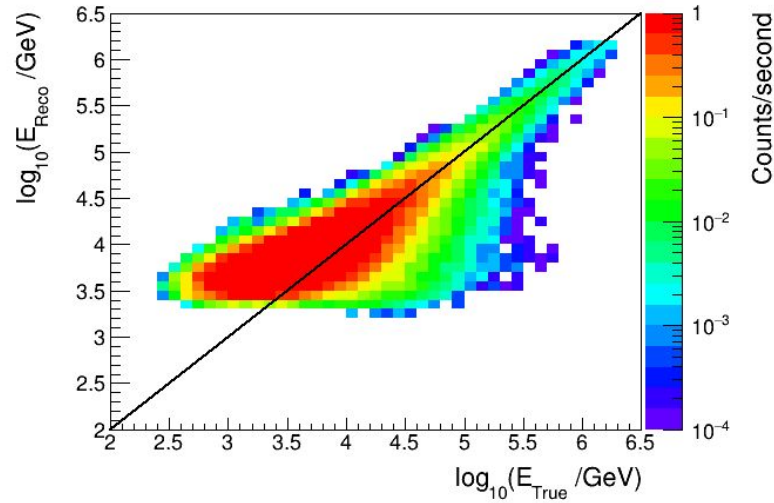


Método clásico VS Red neuronal

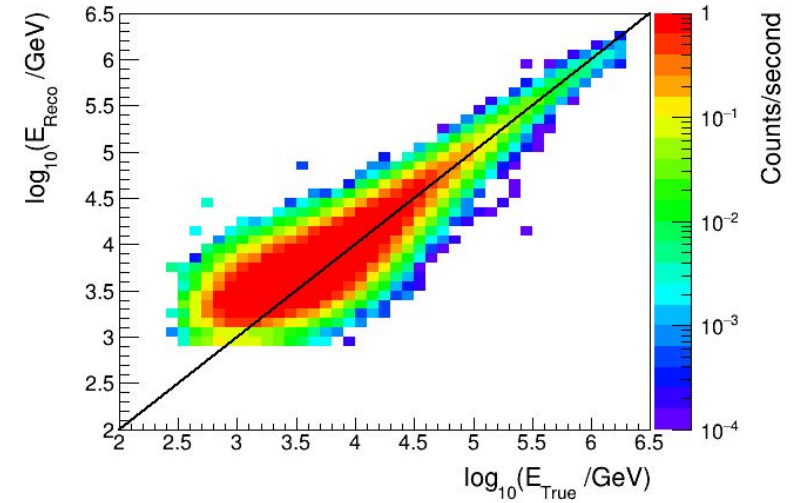


Protones

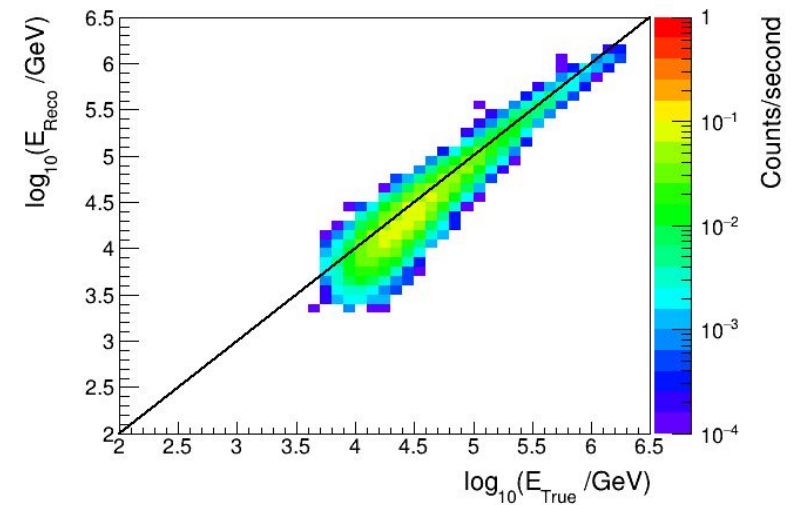
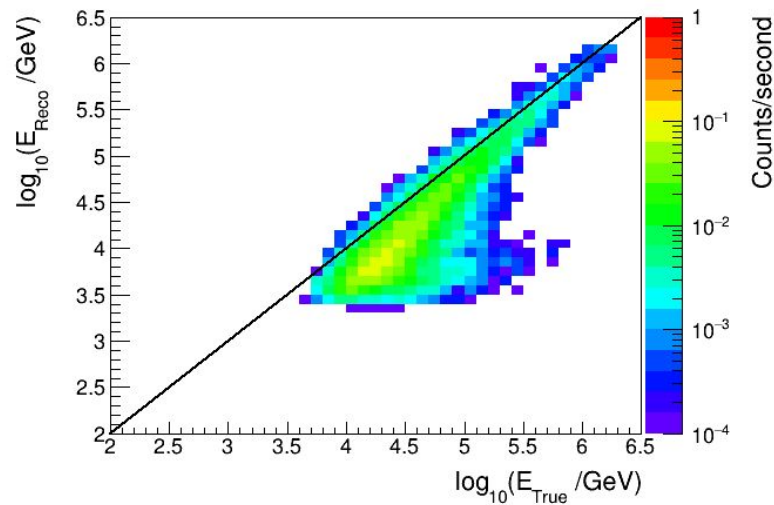
Método clásico



Red neuronal



Hierros



¿Parámetros físicos?

Proximamente

**Este viernes
11:10 – 11:30 AM
Auditorio Jorge Zalamea**



**ASTRO
FEST²⁴**
BUCARAMANGA

NOVIEMBRE
13 AL 23

**3er Encuentro Internacional Julio Garavito
(13nov- 15nov).**
Alianzas estratégicas entre la Agencia Espacial Europea y Colombia.

**1er Encuentro de la Red Iberoamericana CyTED
LAGO INDICA (18nov-19nov).**
Consolidando la comunidad de Astropartículas a escala continental en torno al proyecto LAGO-INDICA.

**8vo Congreso Colombiano de Astronomía y Astrofísica,
CoCoA2024 (20nov-23nov).**
Construyendo puentes y proyectos con contenido e impacto social desde la Astronomía y Astrofísica en Colombia.

<https://jupyterhd.redclara.net/category/11/> 

Información y registro

 **Universidad Industrial de Santander
Bucaramanga, Santander**  **Halley@uis.edu.co**



¡Gracias!



GRUPO
HALLEY





rec.nChAvail: Número de canales disponibles para la detección de eventos.

rec.nHitTot: Número total de hits detectados en todo el observatorio.

rec.nHit: Número de hits detectados en el evento actual.

rec.nHitSP10: Número de hits dentro de una ventana temporal de 10 ns.

rec.nHitSP20: Número de hits dentro de una ventana temporal de 20 ns.

rec.nTankAvail: Número de tanques de agua disponibles para la detección.

rec.nTankHitTot: Número total de tanques que detectaron hits durante el evento.

rec.nTankHit: Número de tanques que detectaron hits en el evento actual.

rec.windowHits: Número de hits detectados en una ventana temporal definida.

rec.planeNDOF: Grados de libertad en el ajuste del plano de la onda del evento.

rec.mPFnHits: Número de hits asociados con el método de partición de planos (PlaneFit).

rec.mPFp0nAssign: Número de hits asignados al primer paso de reconstrucción del ajuste de plano.

rec.mPFp1nAssign: Número de hits asignados al segundo paso del ajuste de plano.

rec.coreFiduScale: Escala de fiducialidad para la posición del núcleo del evento.

rec.zenithAngle: Ángulo cenital del evento, indicando la dirección del rayo cósmico con respecto a la vertical.

rec.planeChi2: Valor de Chi-cuadrado para la calidad del ajuste del plano de la onda.

rec.fAnnulusCharge0: Carga detectada en el primer anillo concéntrico del detector.

rec.fAnnulusCharge1: Carga detectada en el segundo anillo concéntrico.

rec.fAnnulusCharge2: Carga detectada en el tercer anillo concéntrico.

rec.fAnnulusCharge3: Carga detectada en el cuarto anillo concéntrico.

rec.fAnnulusCharge4: Carga detectada en el quinto anillo concéntrico.

rec.fAnnulusCharge5: Carga detectada en el sexto anillo concéntrico.

rec.fAnnulusCharge6: Carga detectada en el séptimo anillo concéntrico.

rec.fAnnulusCharge7: Carga detectada en el octavo anillo concéntrico.

rec.fAnnulusCharge8: Carga detectada en el noveno anillo concéntrico.

rec.logMaxPE: Logaritmo del máximo número de electrones fotoemisores (PE) detectados en el evento.

rec.logNPE: Logaritmo del número total de electrones fotoemisores detectados.

rec.CxPE40: Número estimado de electrones fotoemisores en un área de 40 metros alrededor del núcleo.

rec.logGP: Logaritmo del ajuste de la función de probabilidad geométrica.

rec.mPFp0Weight: Peso asignado al primer paso en el ajuste de partición de planos.

rec.mPFp1Weight: Peso asignado al segundo paso en el ajuste de partición de planos.

rec.disMax: Distancia máxima entre el núcleo reconstruido y los hits detectados.

rec.TankLHR: Likelihood ratio asociado a los tanques que detectaron hits.

rec.LHLatDistFitXmax: Parámetro de ajuste para la distancia lateral máxima en el evento.

rec.LHLatDistFitEnergy: Ajuste de la energía basado en la distribución lateral del evento.

rec.LHLatDistFitGoF: Calidad del ajuste de la distribución lateral del evento.



$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

La función de verosimilitud para un conjunto de datos $X = \{x_1, x_2, \dots, x_n\}$ es:

$$L(\mu, \sigma^2 | X) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

En este caso, podemos derivar y resolver para μ y σ^2 que maximizan la verosimilitud. Esto nos llevará a los estimadores de máxima verosimilitud para una distribución normal:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i \quad (\text{media muestral})$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu})^2 \quad (\text{varianza muestral})$$