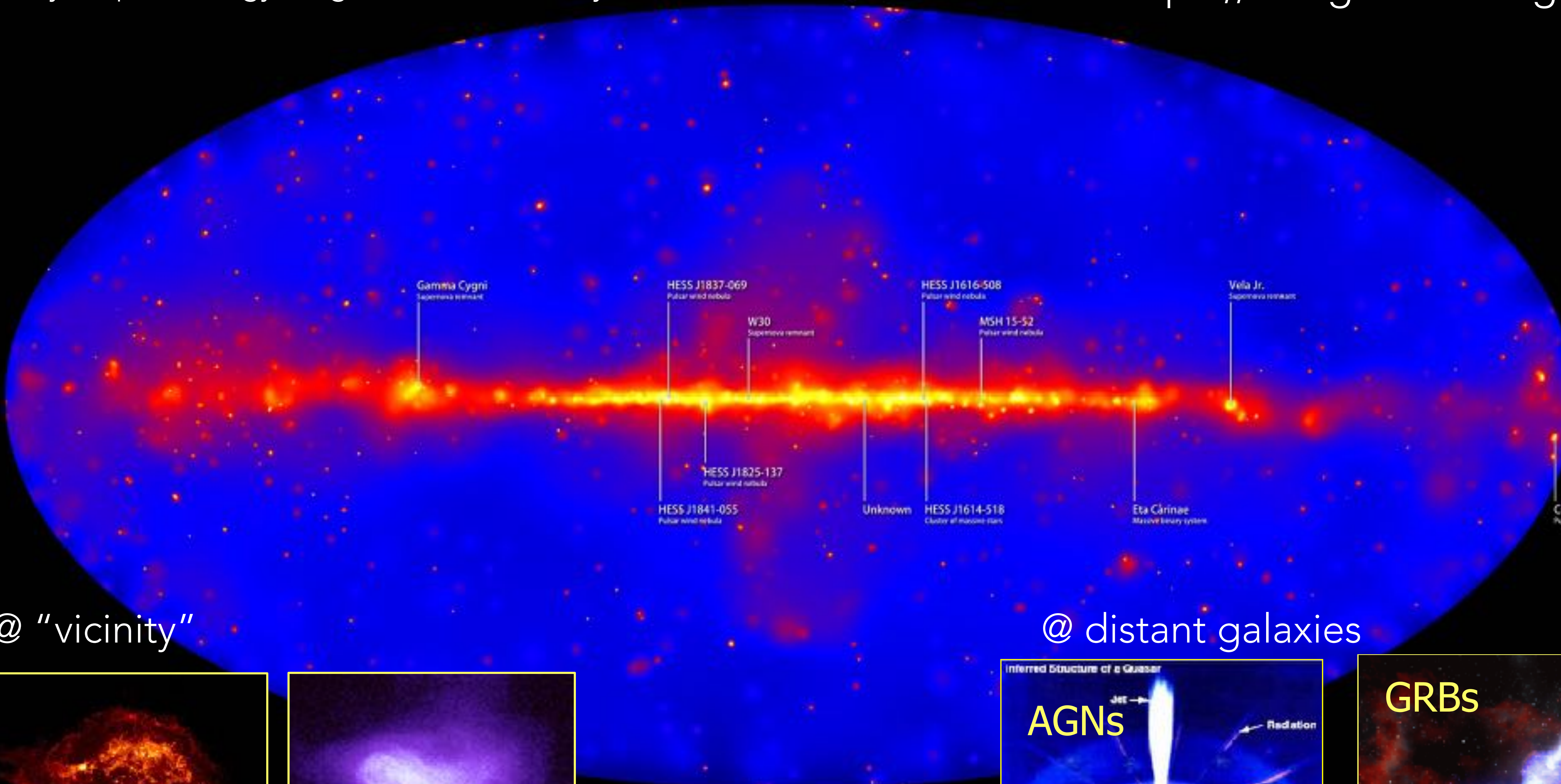
"Improvement of energy reconstruction by using machine learning algorithms in MAGIC"

Kazuma Ishio, David Paneque
Max-Planck-Institut für Physik
Galina Maneva, Petar Temnikov
Institute for Nuclear Research and Nuclear Energy, Sofia, Bulgaria
Abelardo Moralejo
Institut de Fisica d'Altes Energies (IFAE),
The Barcelona Institute of Science and Technology, Bellaterra (Barcelona), Spain
Julian Sitarek
Division of Astrophysics, University of Lodz, Lodz, Poland

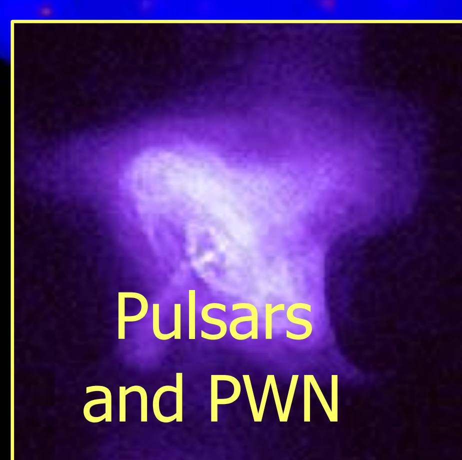
Universe is bright in Gamma rays

Sky map in energy range 50GeV - 2TeV by Fermi satellite

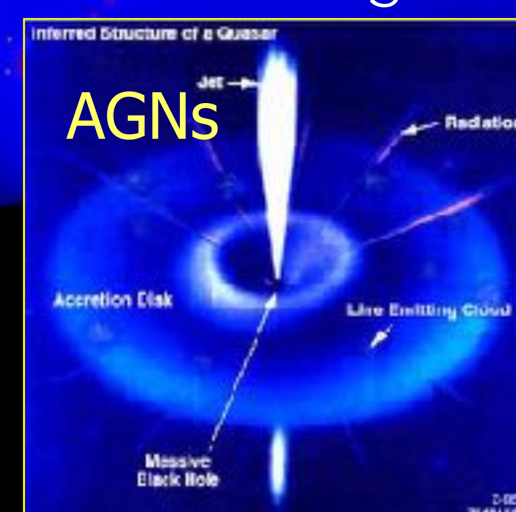
https://svs.gsfc.nasa.gov/



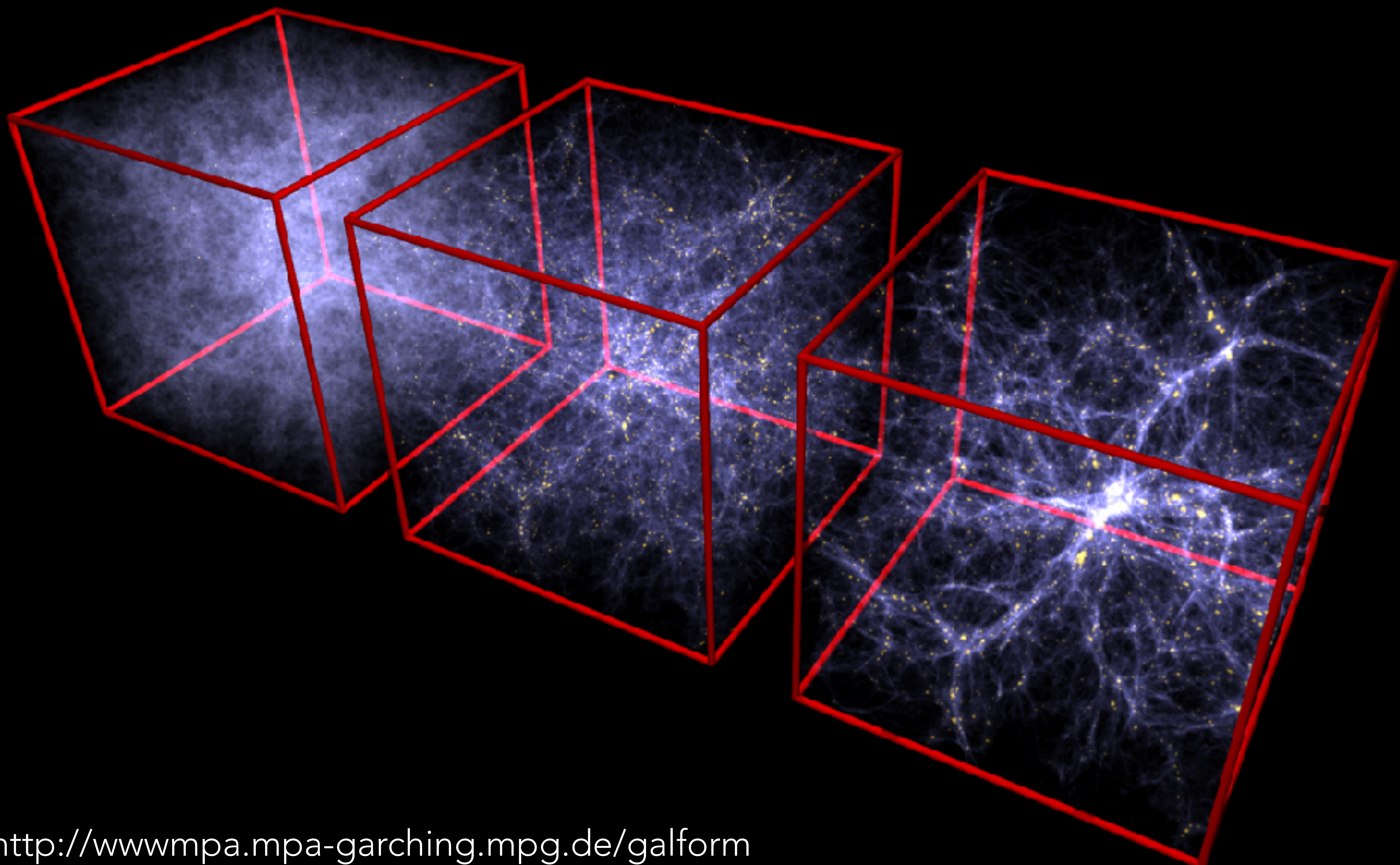
@ "vicinity"



@ distant galaxies

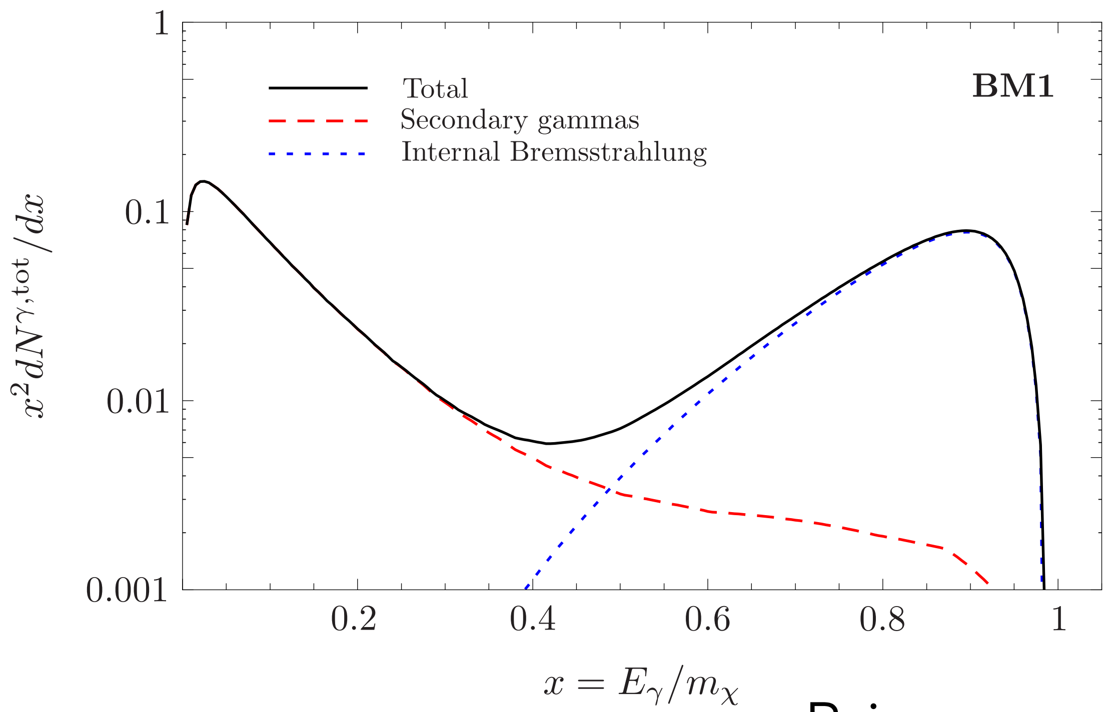


Another possible source — "Dark Matter"



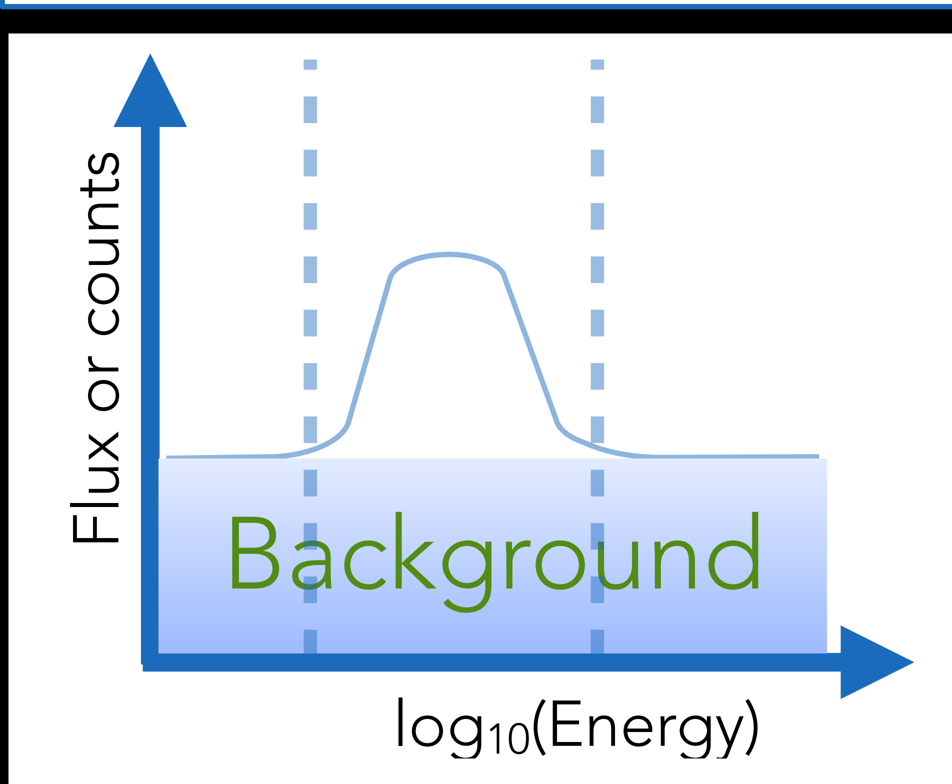
http://wwwmpa.mpa-garching.mpg.de/galform

Another possible source — "Dark Matter"



Bringmann,2008

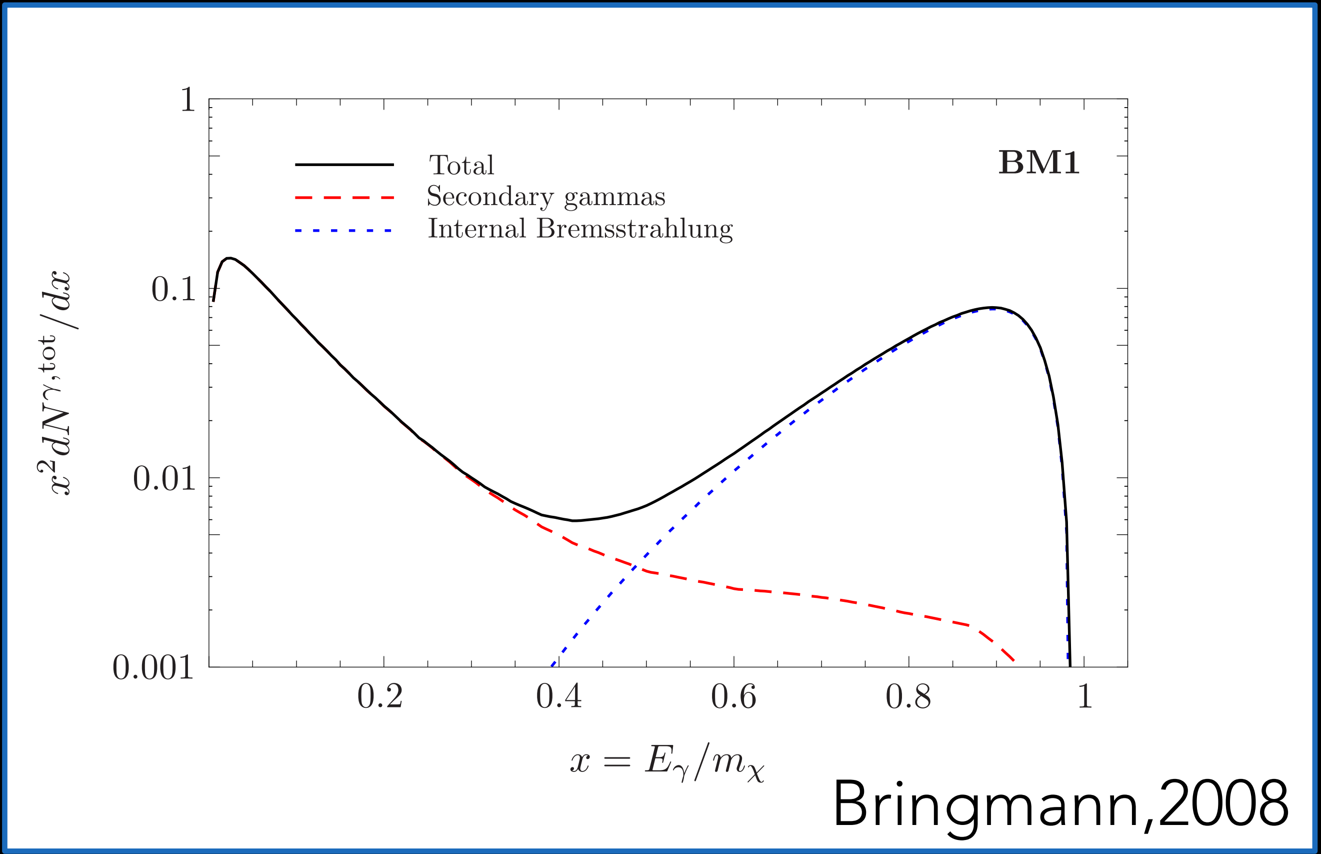
Internal bremsstrahlung from produced charged particles in the annihilations could yield a detectable "bump".



Additional feature in a spectrum to be searched

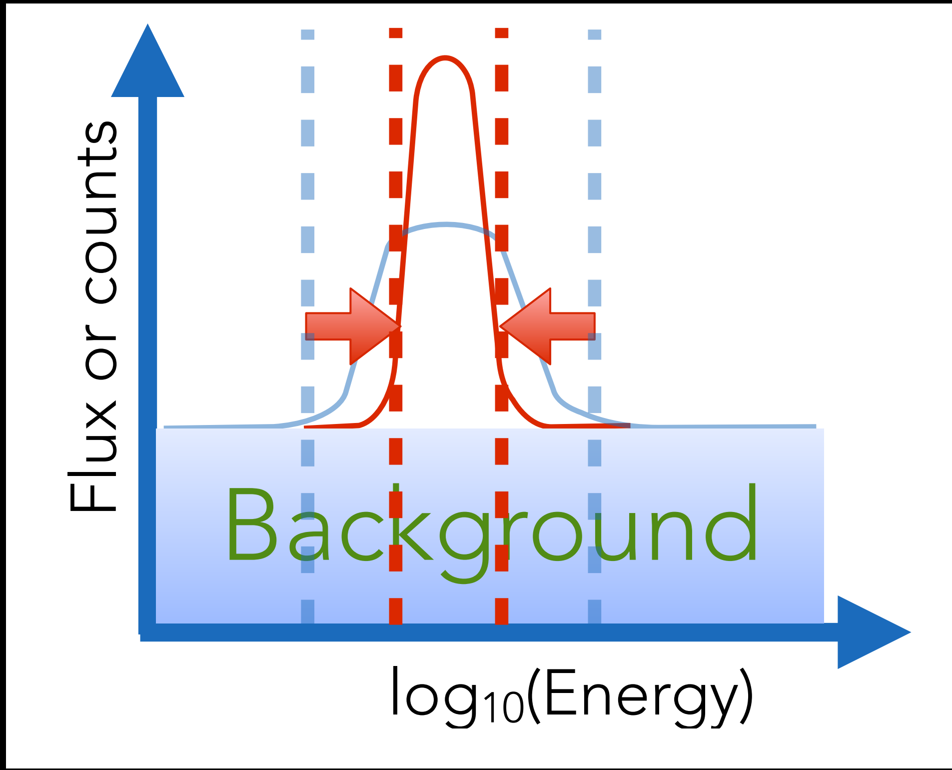
$$S = \frac{N_s}{\sqrt{N_b}}$$

For DM search, energy resolution "matters"



Bringmann,2008

Internal bremsstrahlung from produced charged particles in the annihilations could yield a detectable "bump".



If energy resolution becomes 4 times better, significance would be double!

$$2\times\ S = \frac{N_s}{\sqrt{N_b \times 1/4}}$$

TeV gamma ray with MAGIC telescope

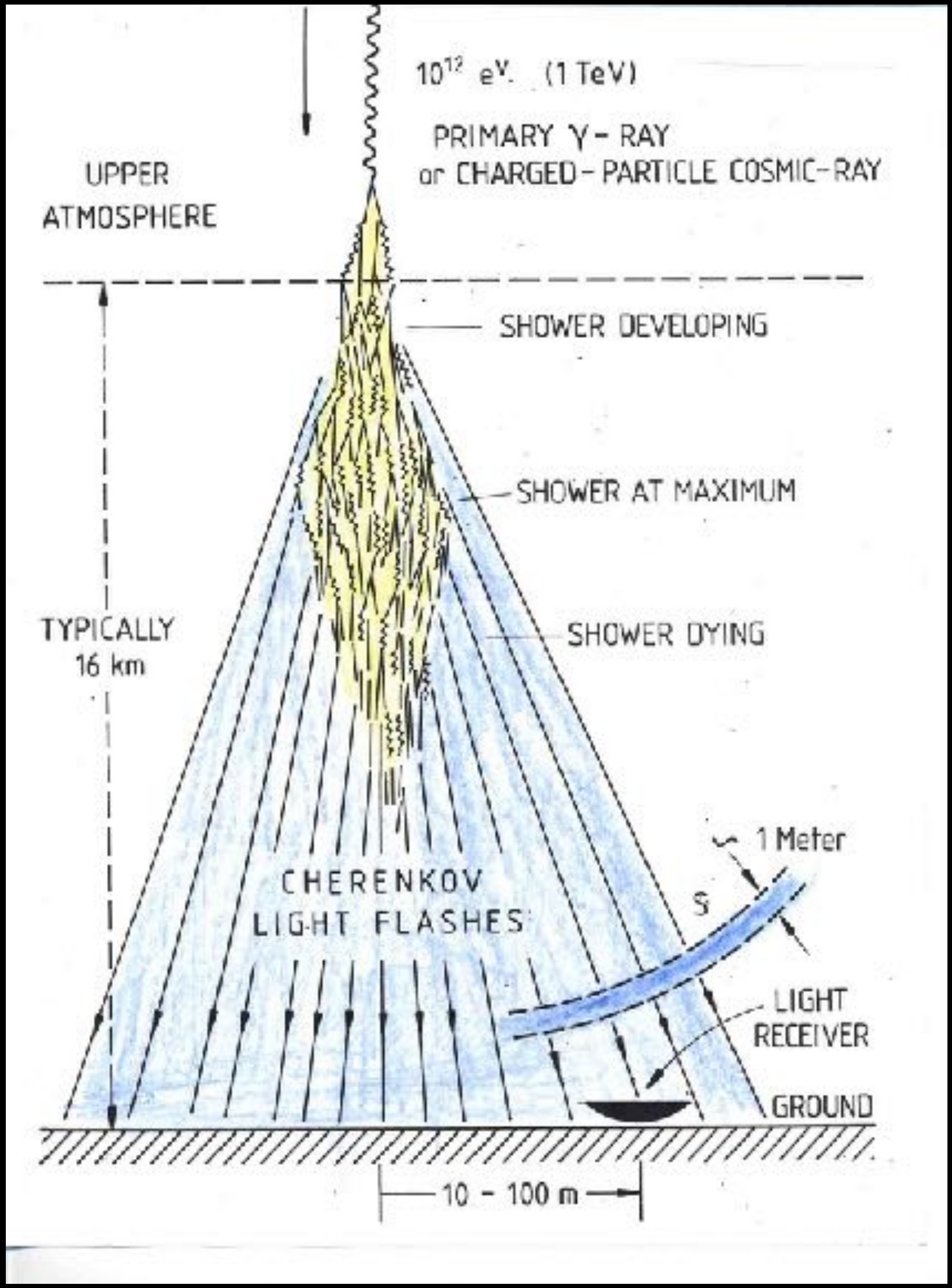
La Palma(29°N, 18°W), asl. 2200m

Imaging Atmospheric Cherenkov Telescope (IACT)

2 telescopes with

- **Dish diameter : 17m**
- **Camera FoV : 3.5deg**
- **Trigger Threshold of gamma ray : ~50 GeV**
- **Sensitivity : ~0.7% Crab flux 0.2TeV**

How energy is estimated?



A high energy particle interacts with atmosphere, which initiates "air shower", consists of so many secondary particles traveling faster than speed of light in the air.

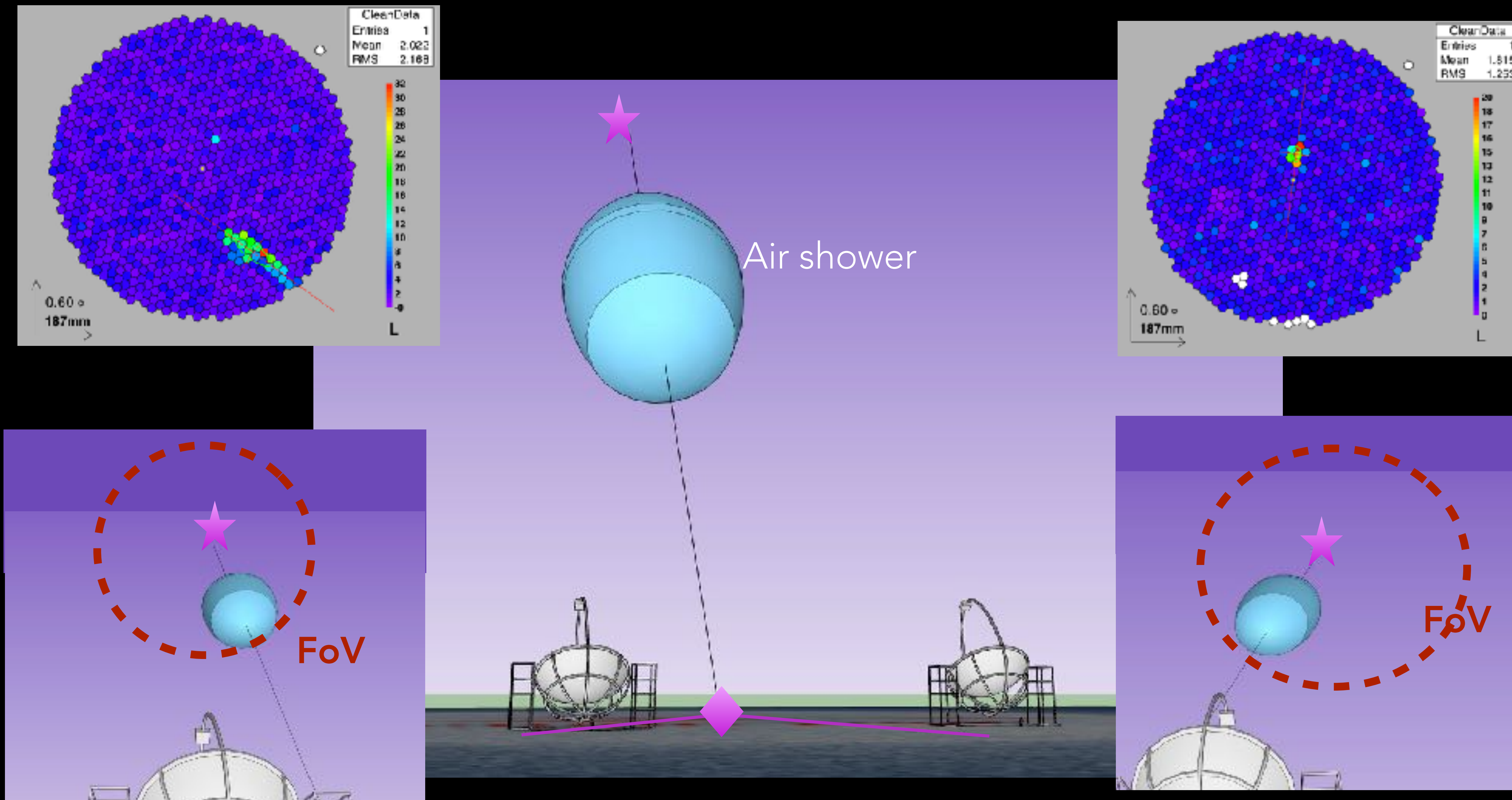
→ Cherenkov radiation

- 10^4 times higher sensitivity than satellites !

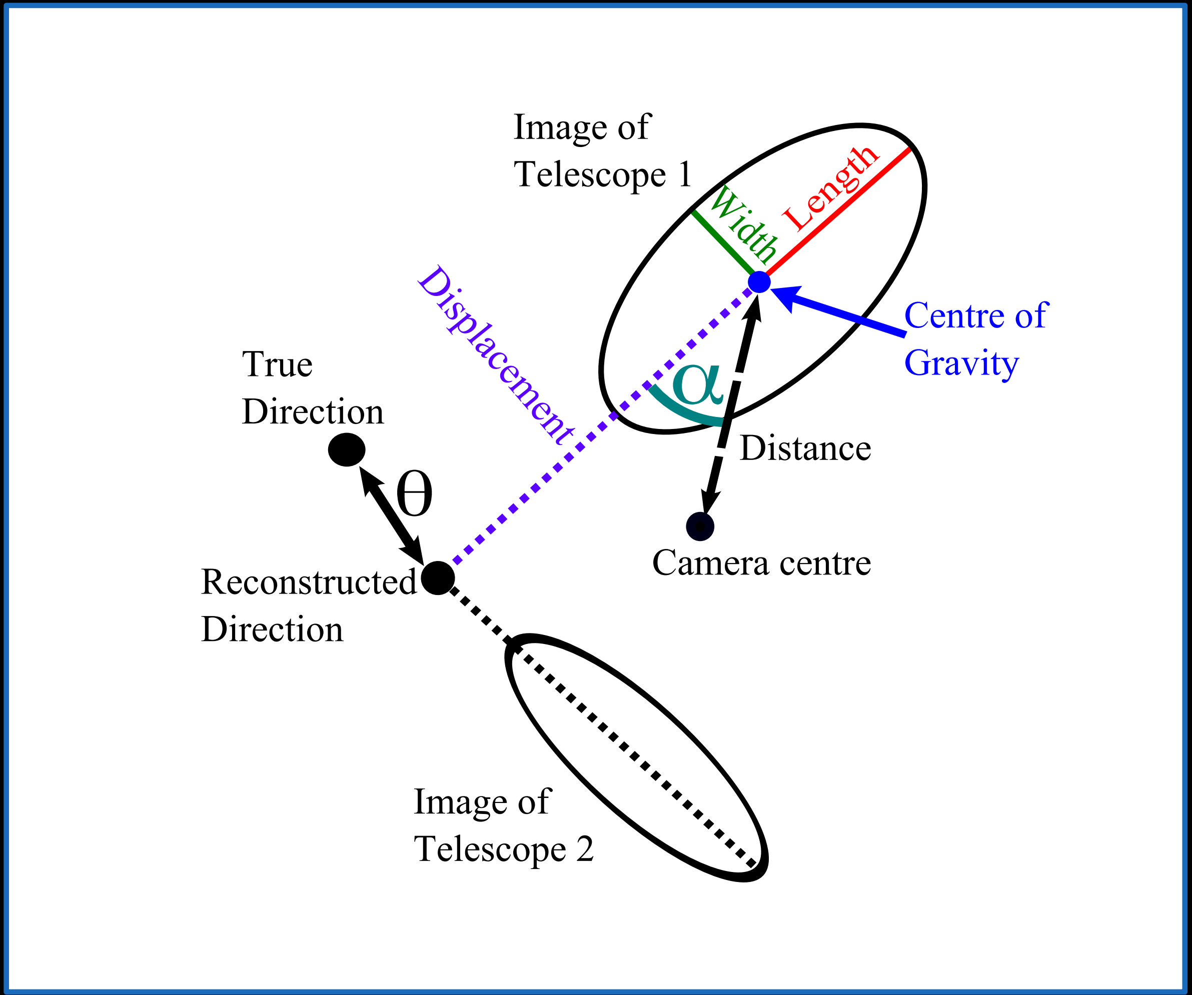
-The higher the gamma ray's energy, the more the secondary particles, and hence the brighter the image of the shower (cherenkov light)

The higher the gammaray's energy, the brighter the shower image. But…location matters!

Darker when more distant. —> correction with geometrical information is needed



Parametrisation



For each event,
a vector value is stored
with many components.

- Brightness (light content)
directly indicates initial energy.
It needs to be corrected by the location parameters.

- Shape
useful for background rejection.

- Orientation and location
important for correction.

Energy can be estimated from light content
corrected by location parameter etc.
=> **15 components are used in the Look Up Table method**

Performance should improve by adopting machine learning

Specifications of the ANN & RF

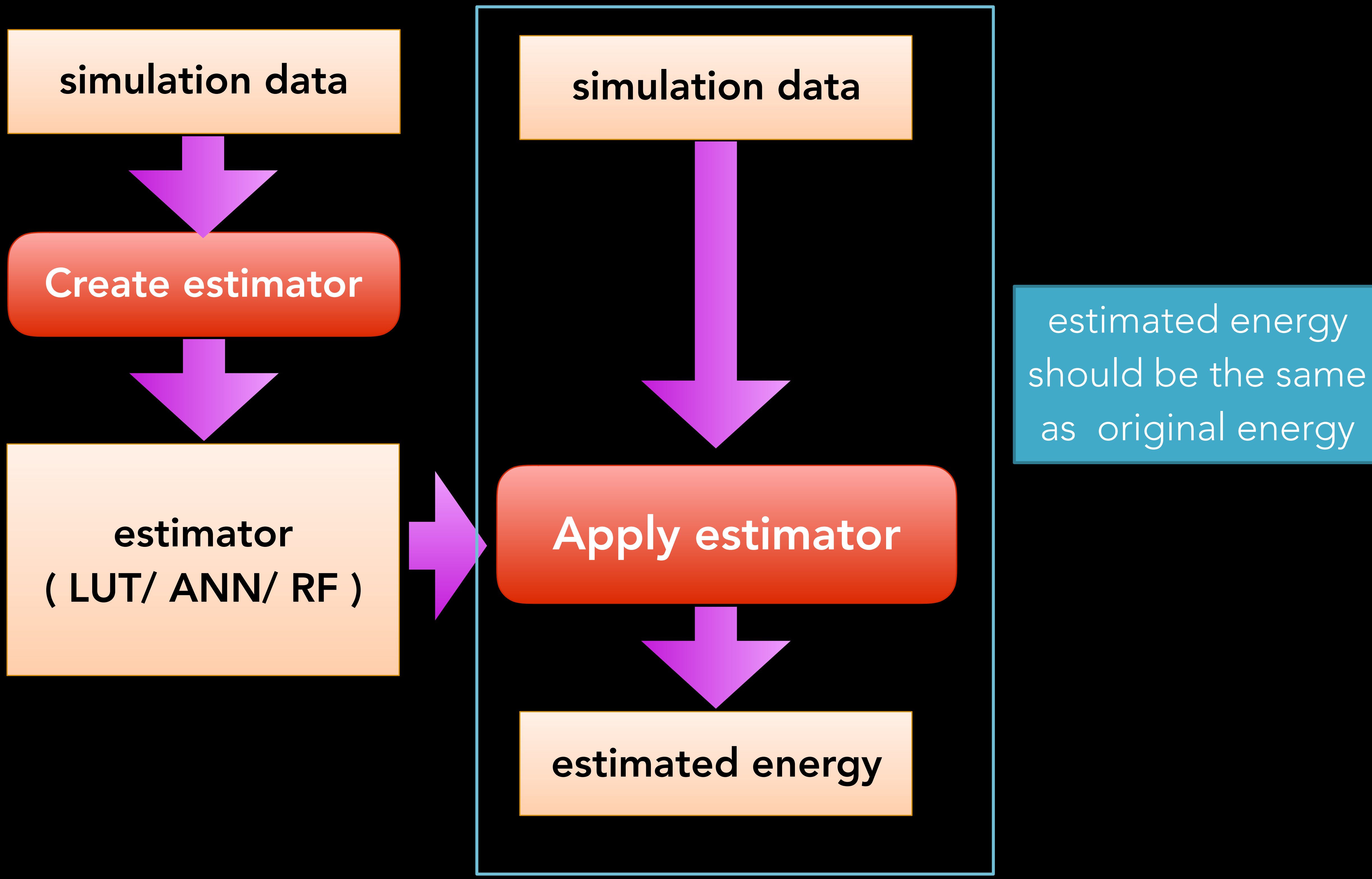
Artificial Neural Network

- JETNET package
- node structure = 15-12-09-05-01
- Better performance when proper cuts on the simulation events for training are performed.

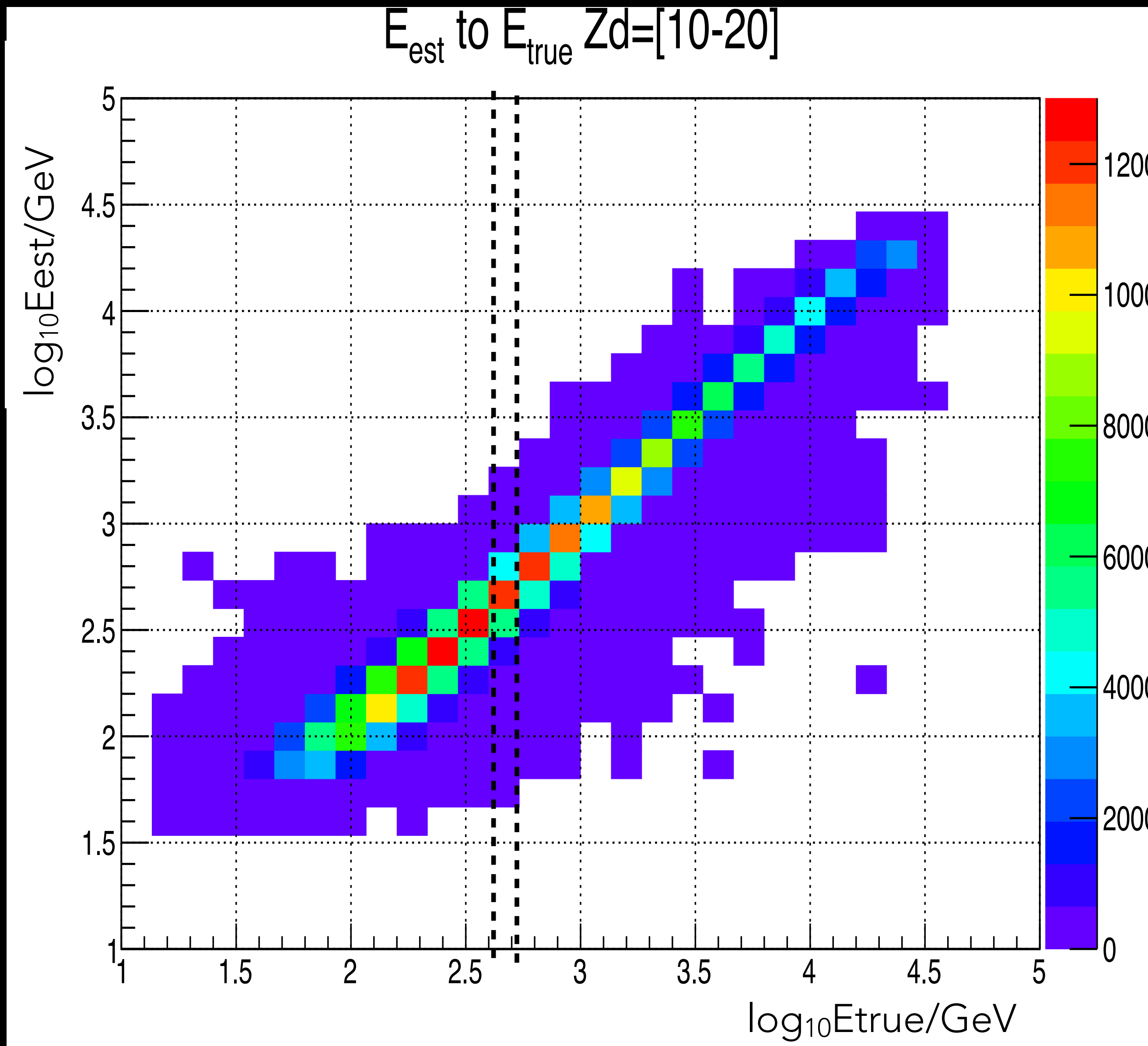
Random Forest

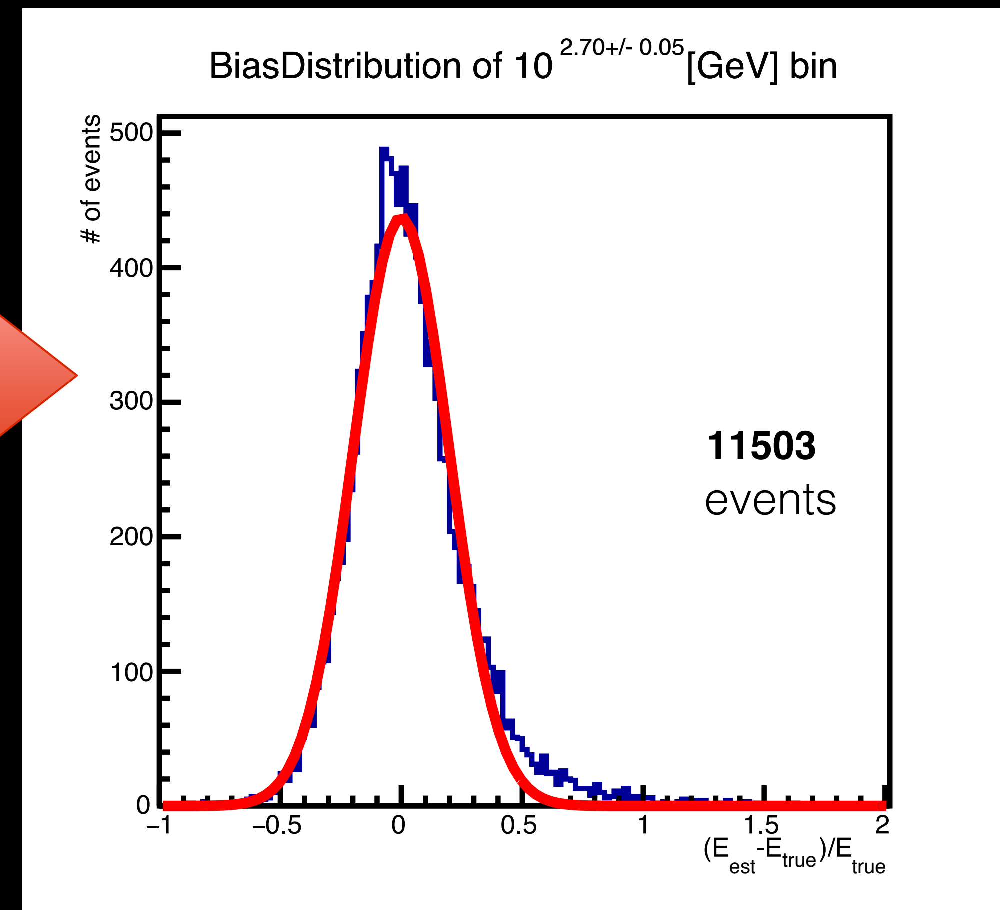
- coded from scratch.
- bootstrap bagging of events for training
- number of trees = 200
- minimum node size = 5
- number of trials =3
  (to choose the most effective parameter to separate)
- Better performance when all the simulation events are used for training

Performance evaluation



Performance evaluataion



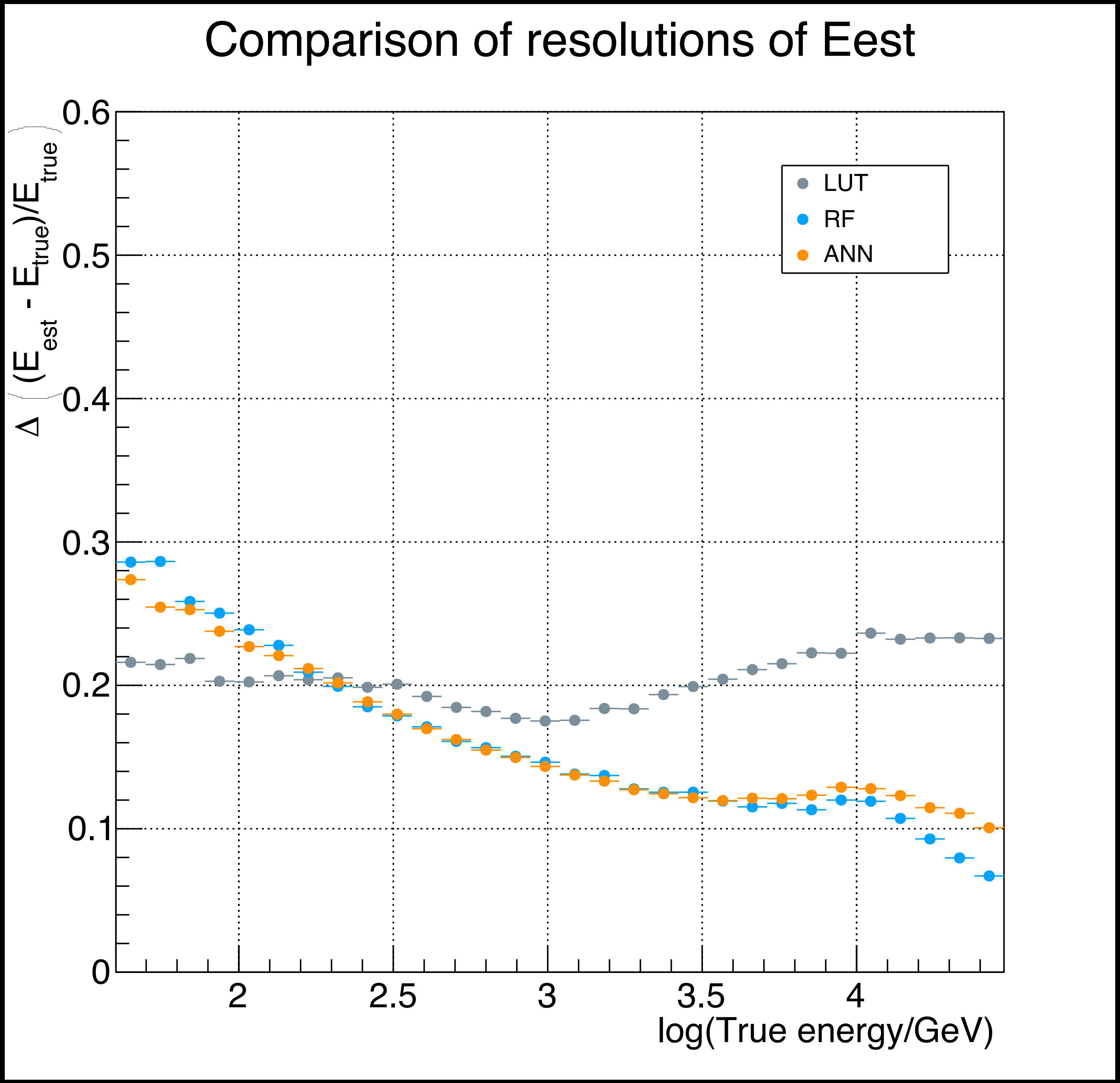


$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Gaussian fit to each distribution of
E_{est} - E_{true} in energy ranges

Bias := μ
Resolution := σ

Improvement by machine learning



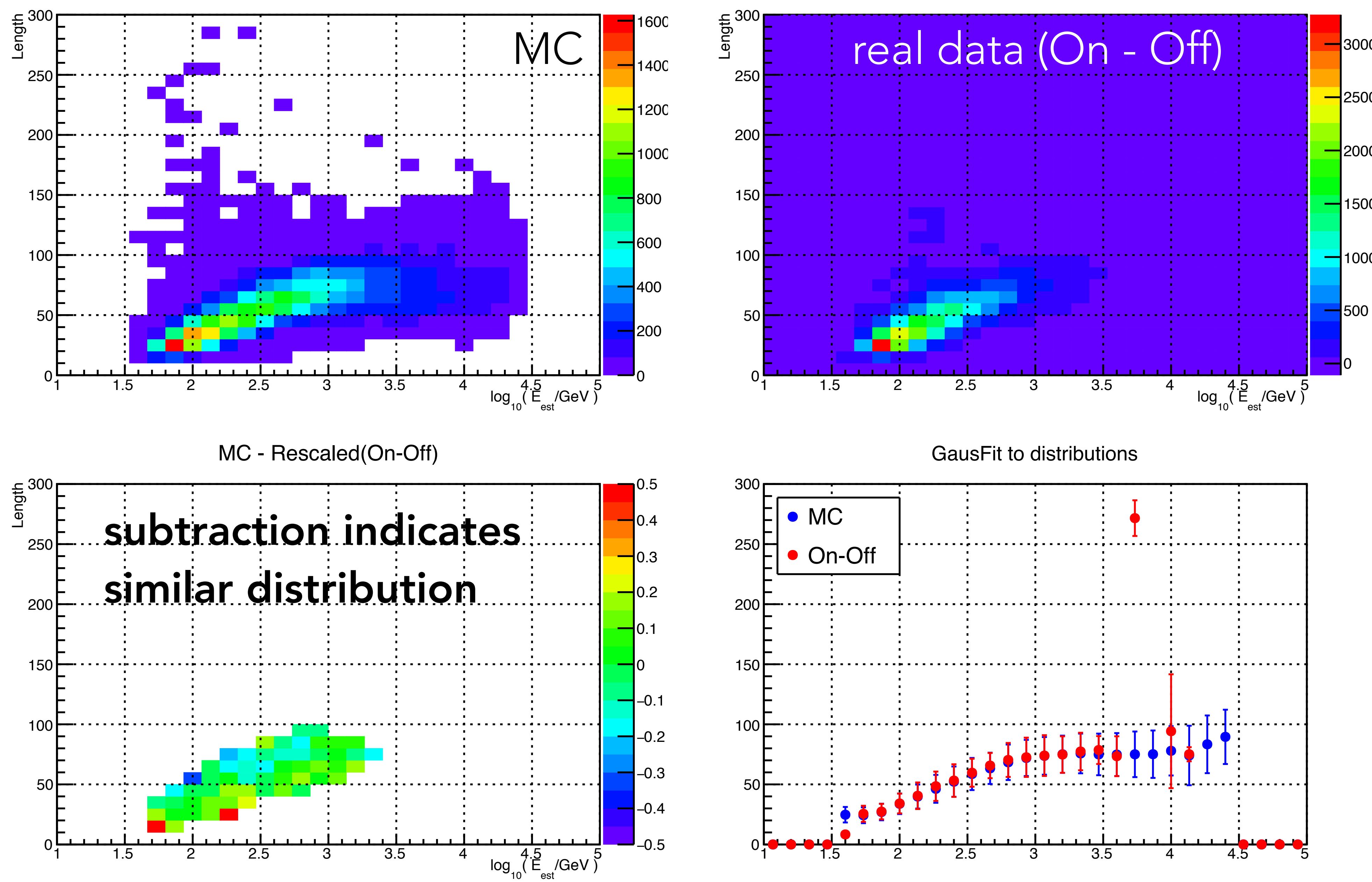
Both machine learning techniques perform very similarly

Better resolution above ~200GeV than the LUT (Current standard in MAGIC)

Sanity check (in RF strategy)

All the parameter should distribute similarly
under the same estimated energy and incoming direction.

An example of the comparison：MaxHeight (Zd =[10,20]deg)



Summary and Conclusions

Gamma-ray astronomy is a novel discipline that addresses many scientific topics.
A good energy resolution can play important roles in many scientific studies (e.g. identification of bumps).

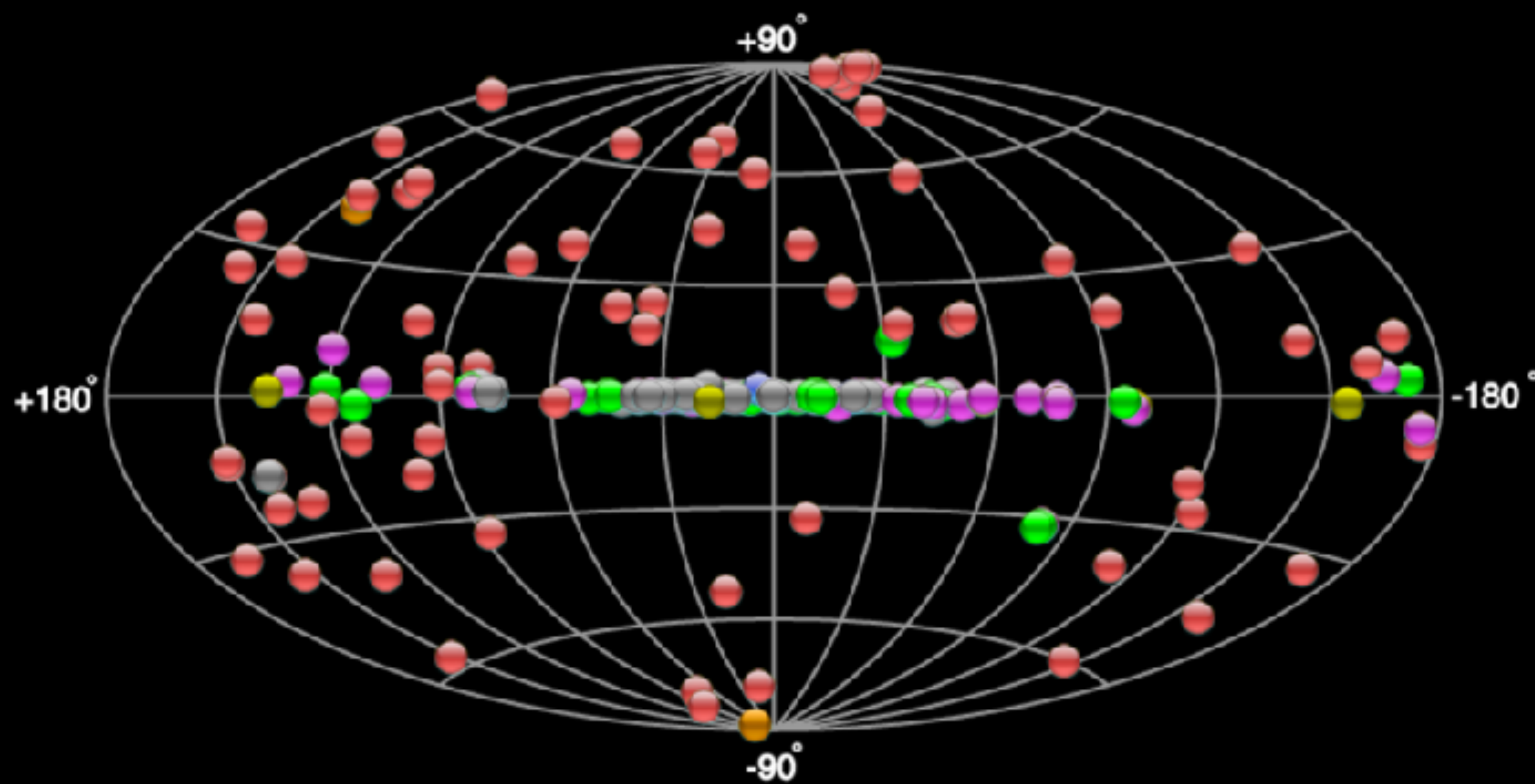
In IACT technique, the gamma-ray energy is derived from many image parameters. It is an excellent case for a room to be improved by machine learning techniques.

We have developed strategies and tools for the application of machine learning techniques (ANN and RF) for the reconstruction of the gamma-ray energy in MAGIC data.

When compared with the LUTs (standard method used in MAGIC), both ANN and RF show a performance improvement above 0.2 TeV, with a factor ~2 improvement at multi-TeV energies.
In case of bump-like feature search, up to 40% higher significance can be expexted.

BACKUP

~200 emit even higher energy!



http://tevcat.uchicago.edu/

Source types

Extragalactic sources

- Active Galactic Nuclei
- Starburst Galaxy

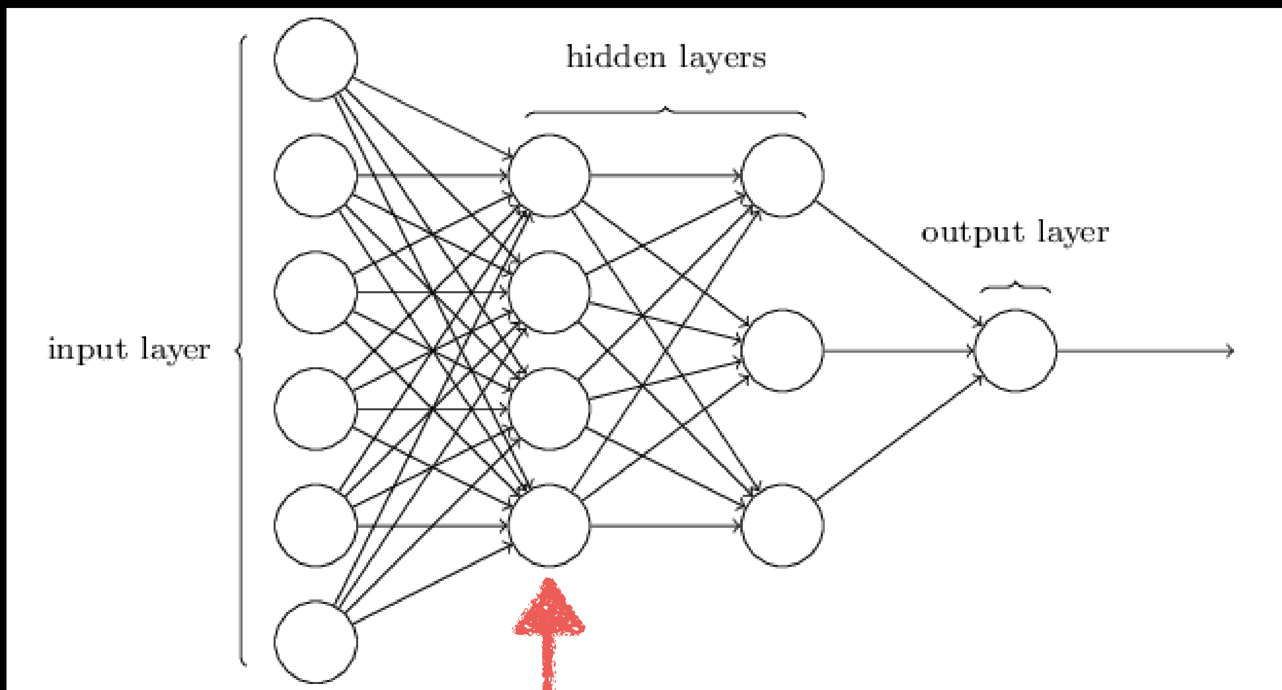
Galactic sources

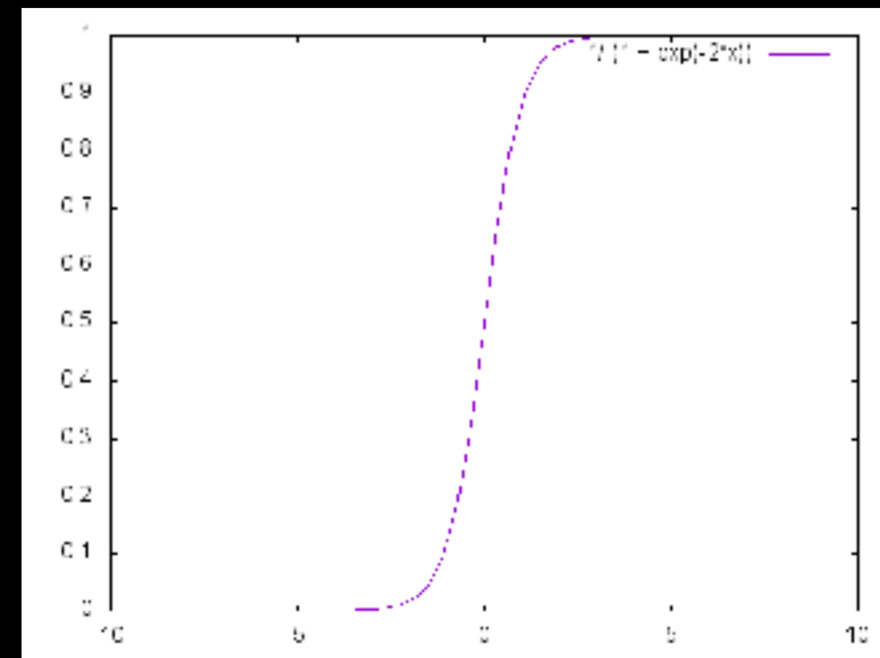
- Pulsar Wind Nebula
- Super Nova Remnant
- Compact object (Pulsars,binaries etc.)
- Star forming region Globular cluster

- Unidentified

The sources which are detected by
IACT : “Imaging Atmospheric Cherenkov Telescopes”
which are x10^4 more sensitive than satellites!

Artificial Neural Network (ANN)





The output of j th node in l th layer is the activation function σ

$$a_j^l = \sigma\left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\right)$$

σ : "Activation function" such as Sigmoid function

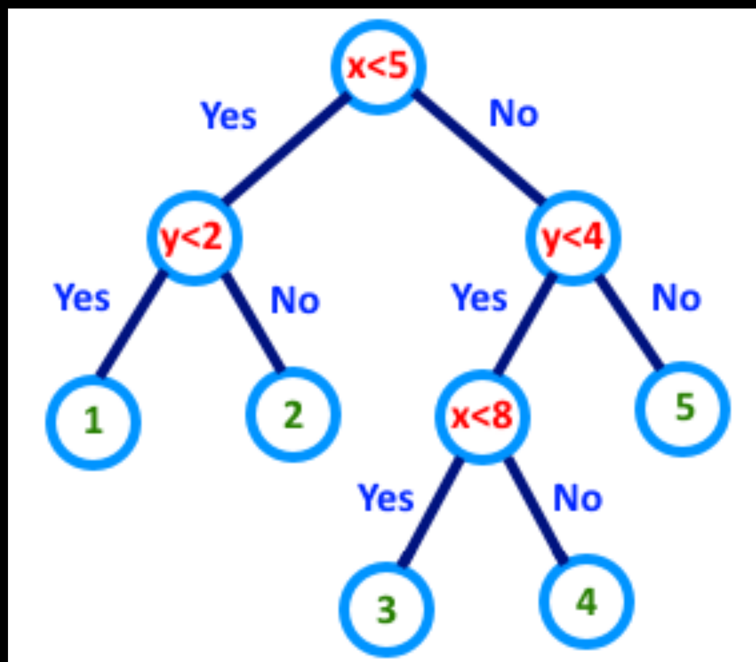
Weight w_{jk}
(Strength of connection)

Input a_k^{l-1}
(output of k th node in l-1 th layer)

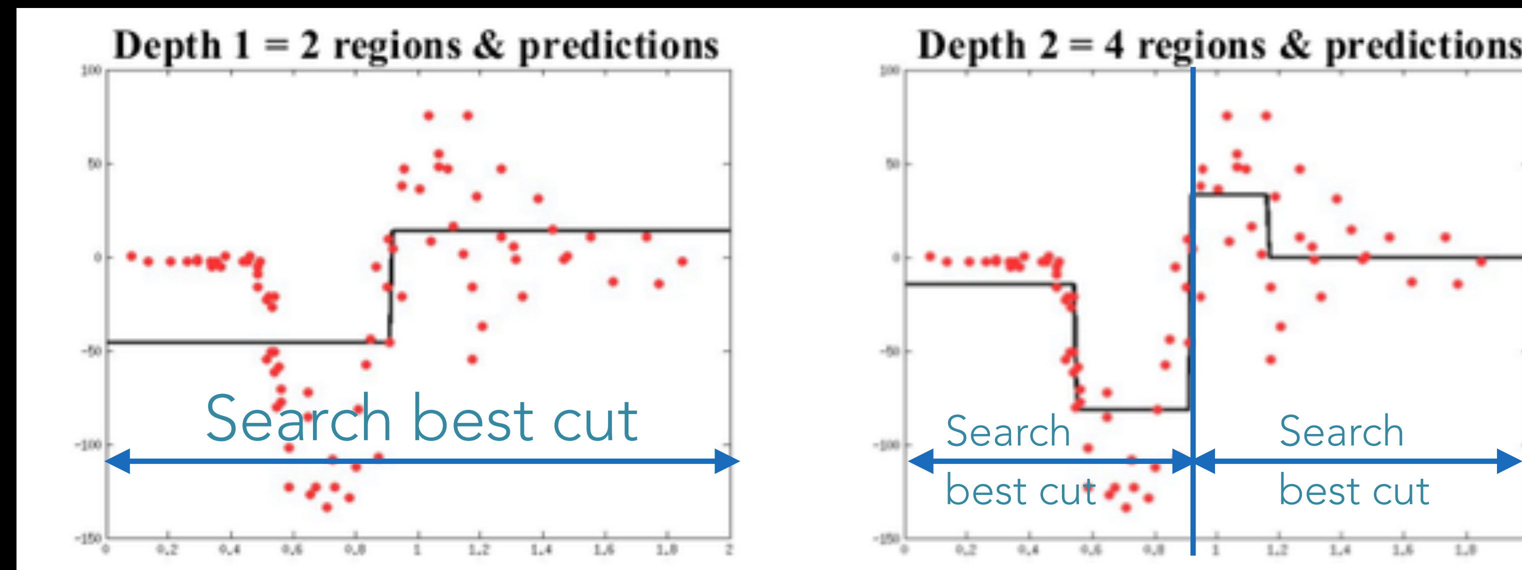
bias b_j

The network can become almost any kind of nonlinear function

Random Forest (RF)



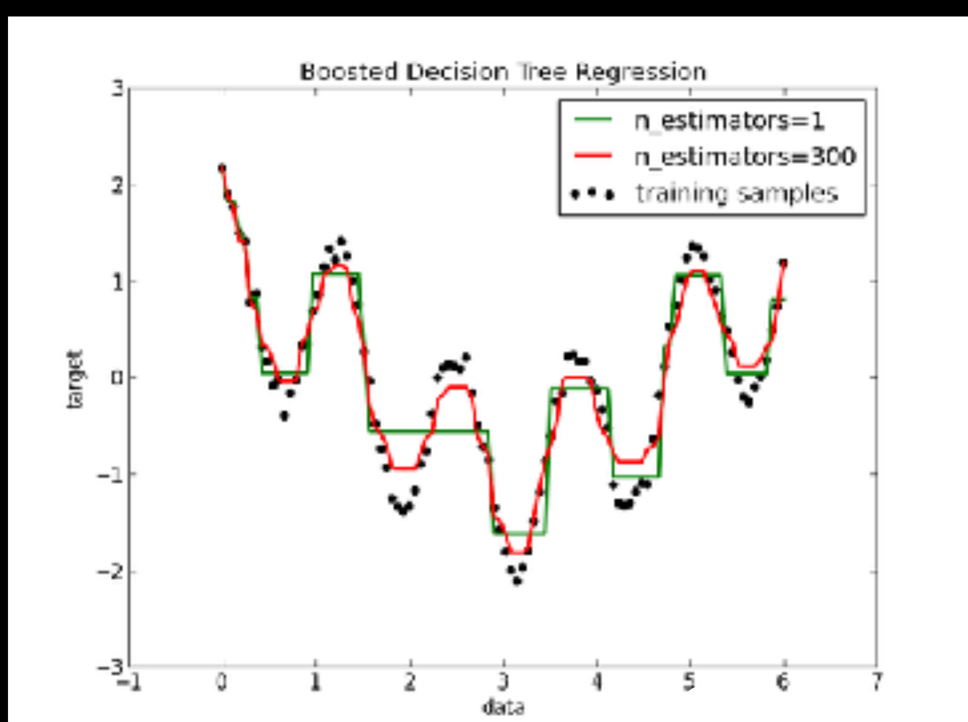
A decision Tree classifies events by energy classes.



The distributions are separated at minimum of the covariance σ^2.

$$\sigma^2(E) = \frac{1}{N_L + N_R}(N_L\sigma_L^2(E) + N_R\sigma_R^2(E))$$

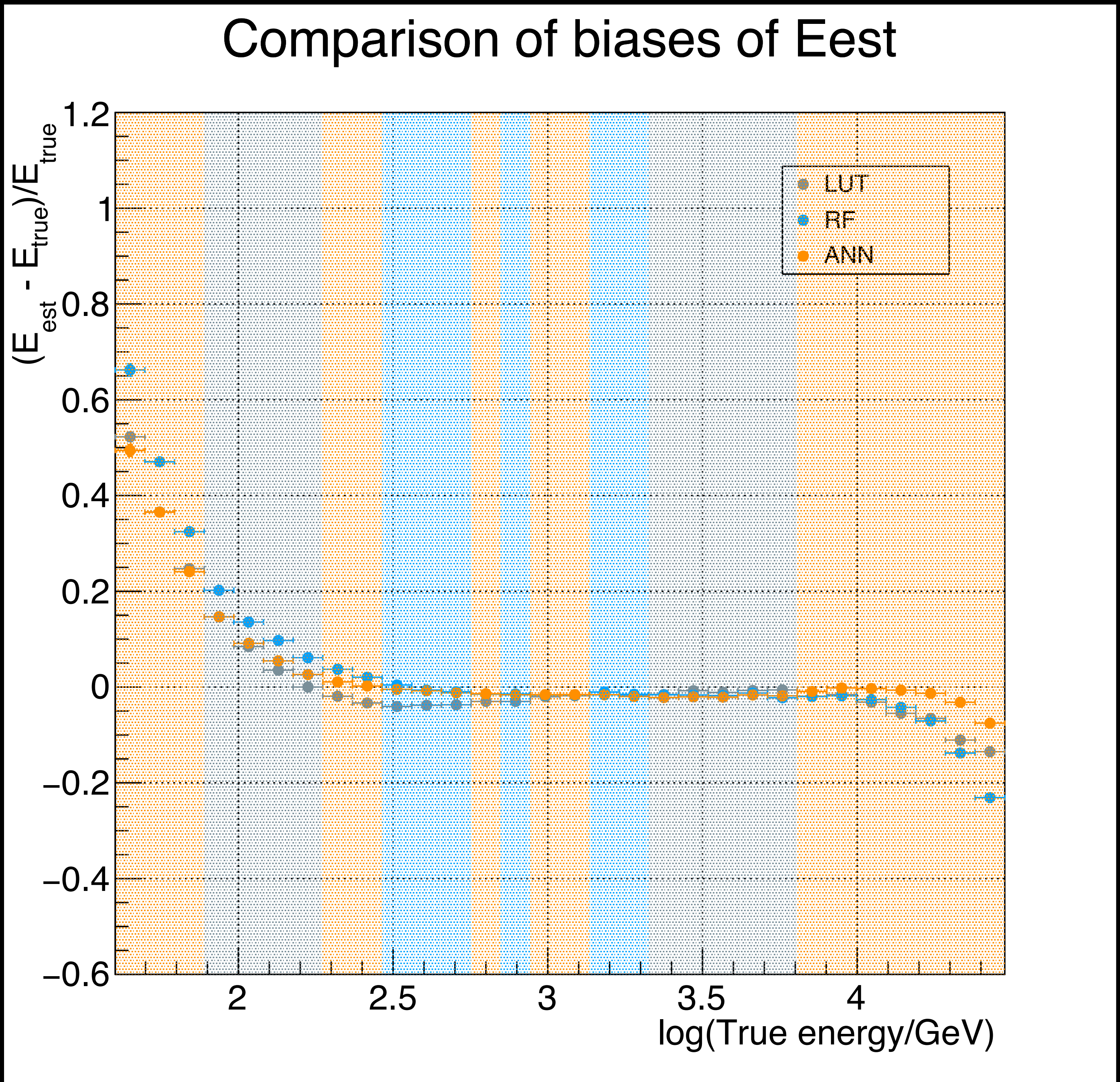
Ei (the energy in class i) is determined as average of Ni events in final nodes



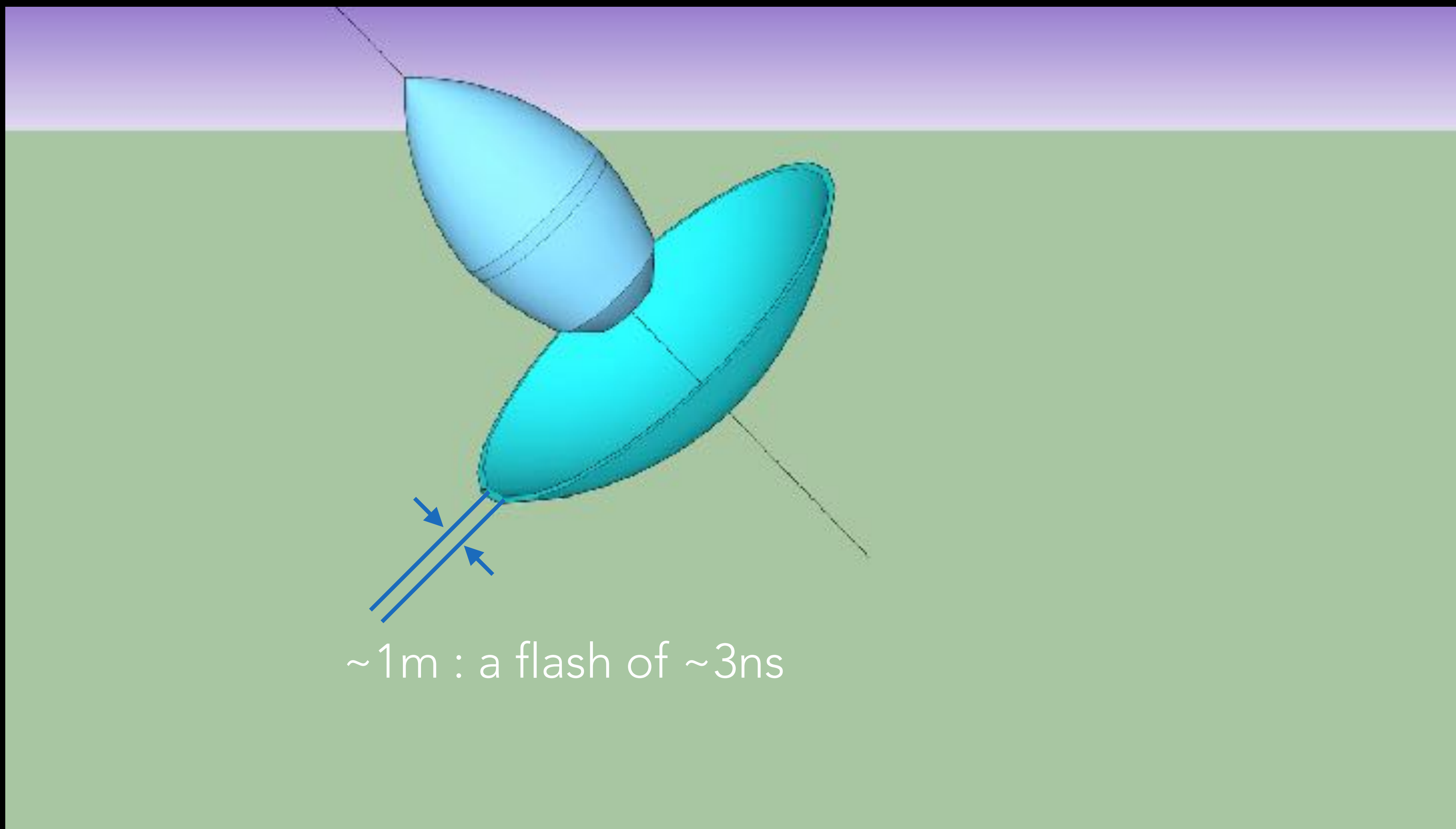
A forest is created by growing different trees,
-> Average of estimators follows true value well!

$$E_{est} = \frac{\sum_{i=0}^{n-1} E_i \cdot N_i}{\sum_{i=0}^{n-1} N_i}$$

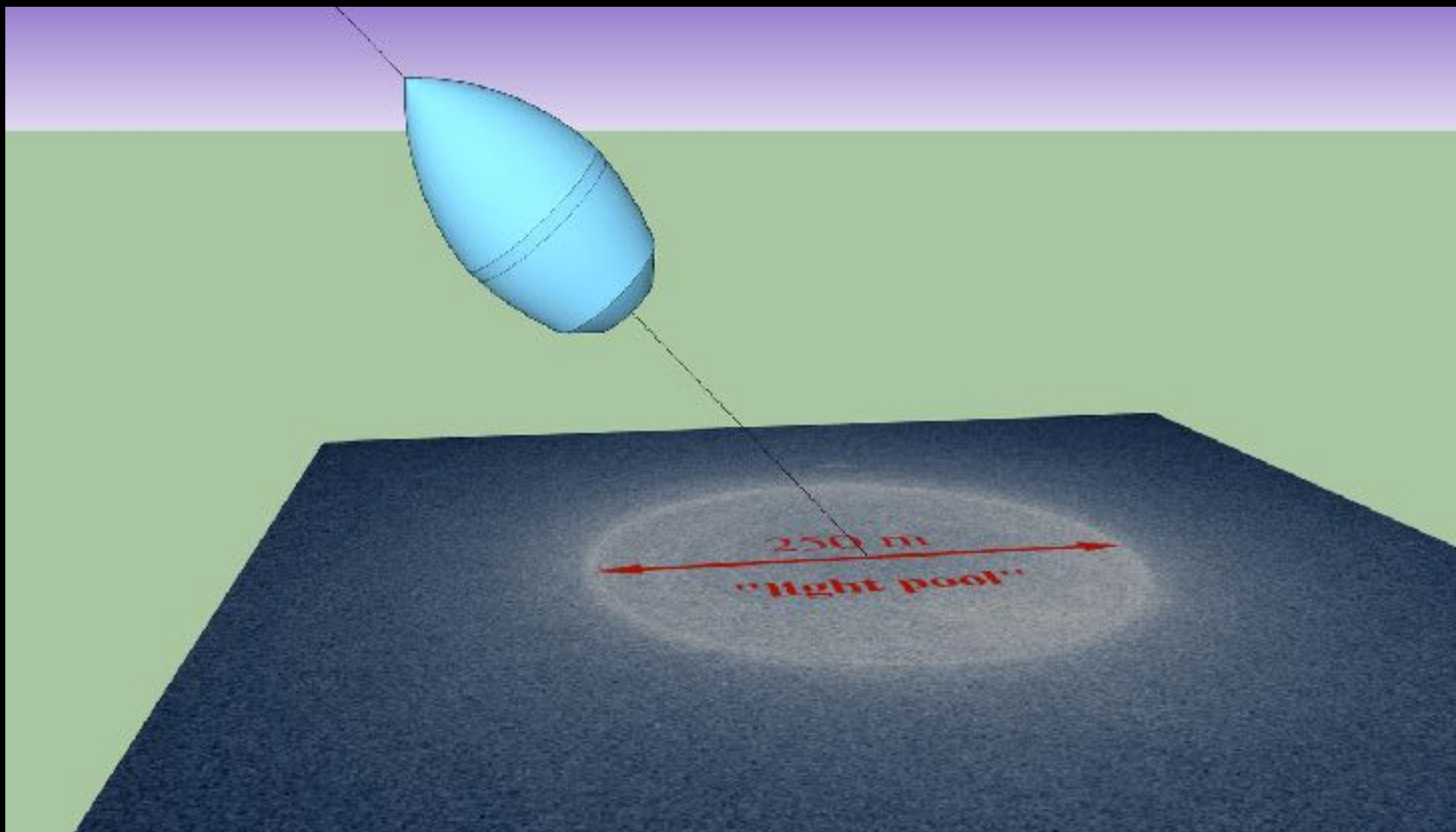
https://www.quora.com/How-does-random-forest-work-for-regression-1


Comparison of biases of Eest
$(E_{est} - E_{true})/E_{true}$
1.2
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
2
2.5
3
3.5
4
log(True energy/GeV)
LUT
RF
ANN

Cherenkov flash

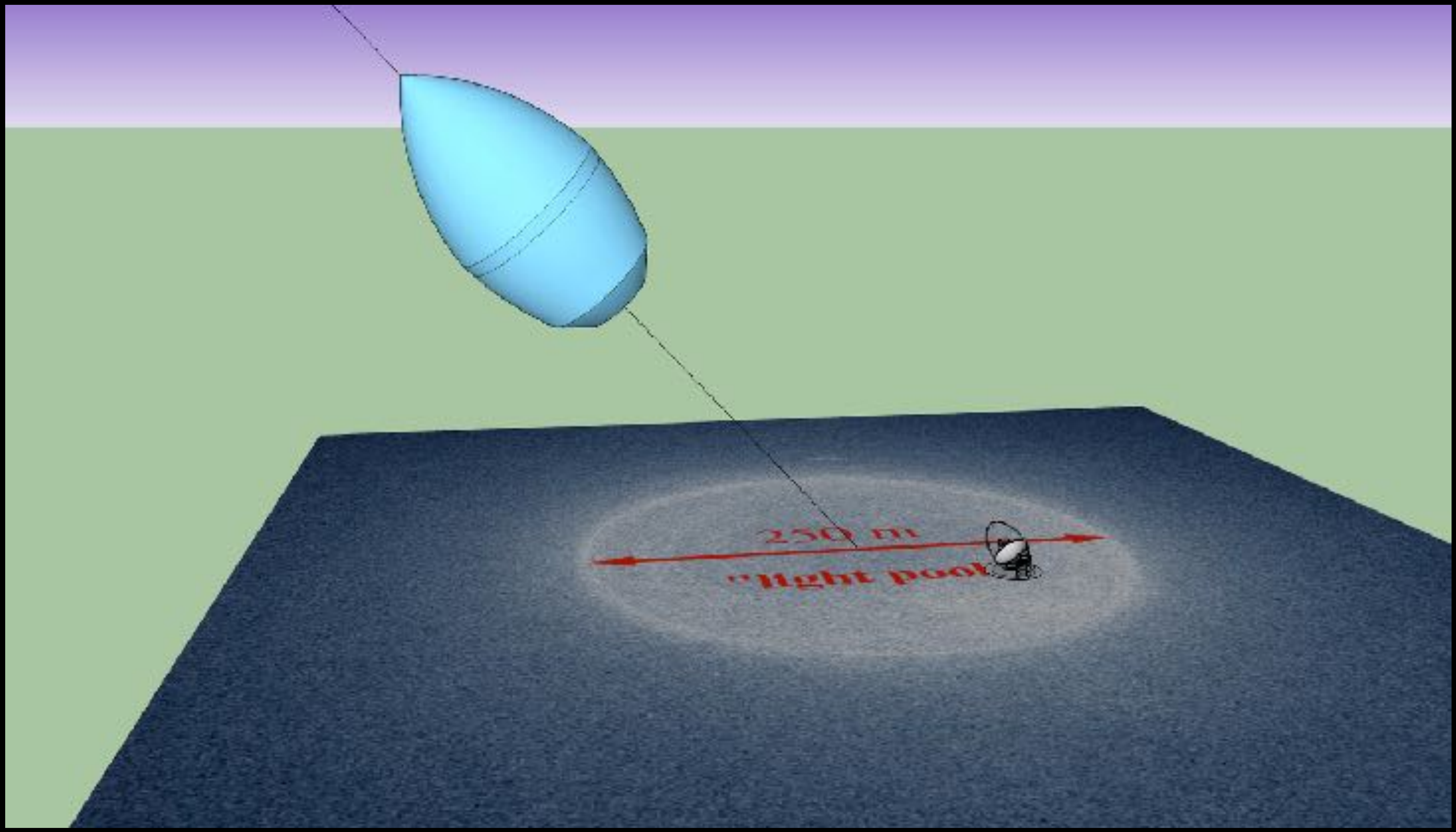


Light pool with diameter~250m



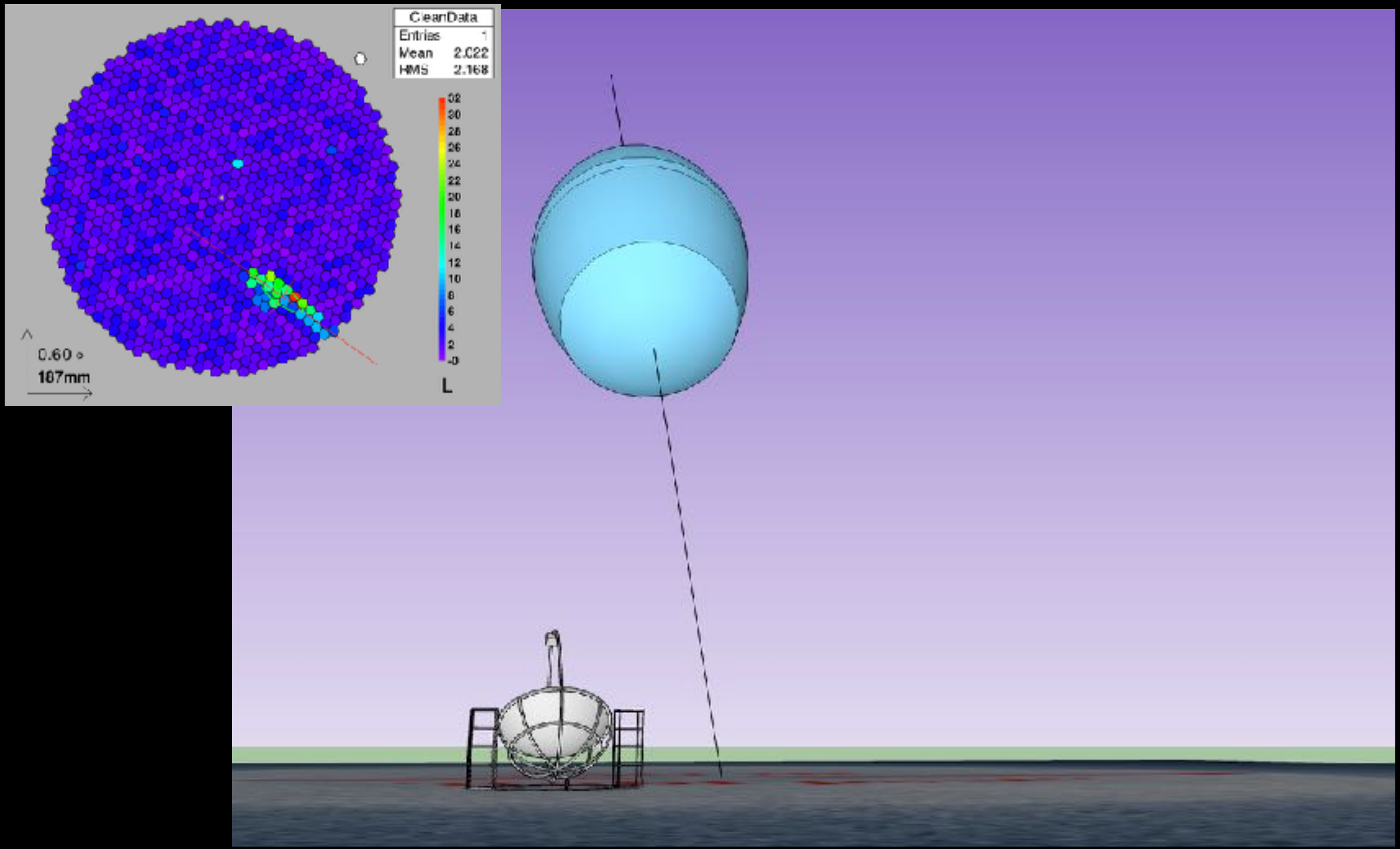
What is "Imaging Atmospheric Cherenkov Telescope (IACT)" ?

The shower can be seen if a telescope is within its lightpool.



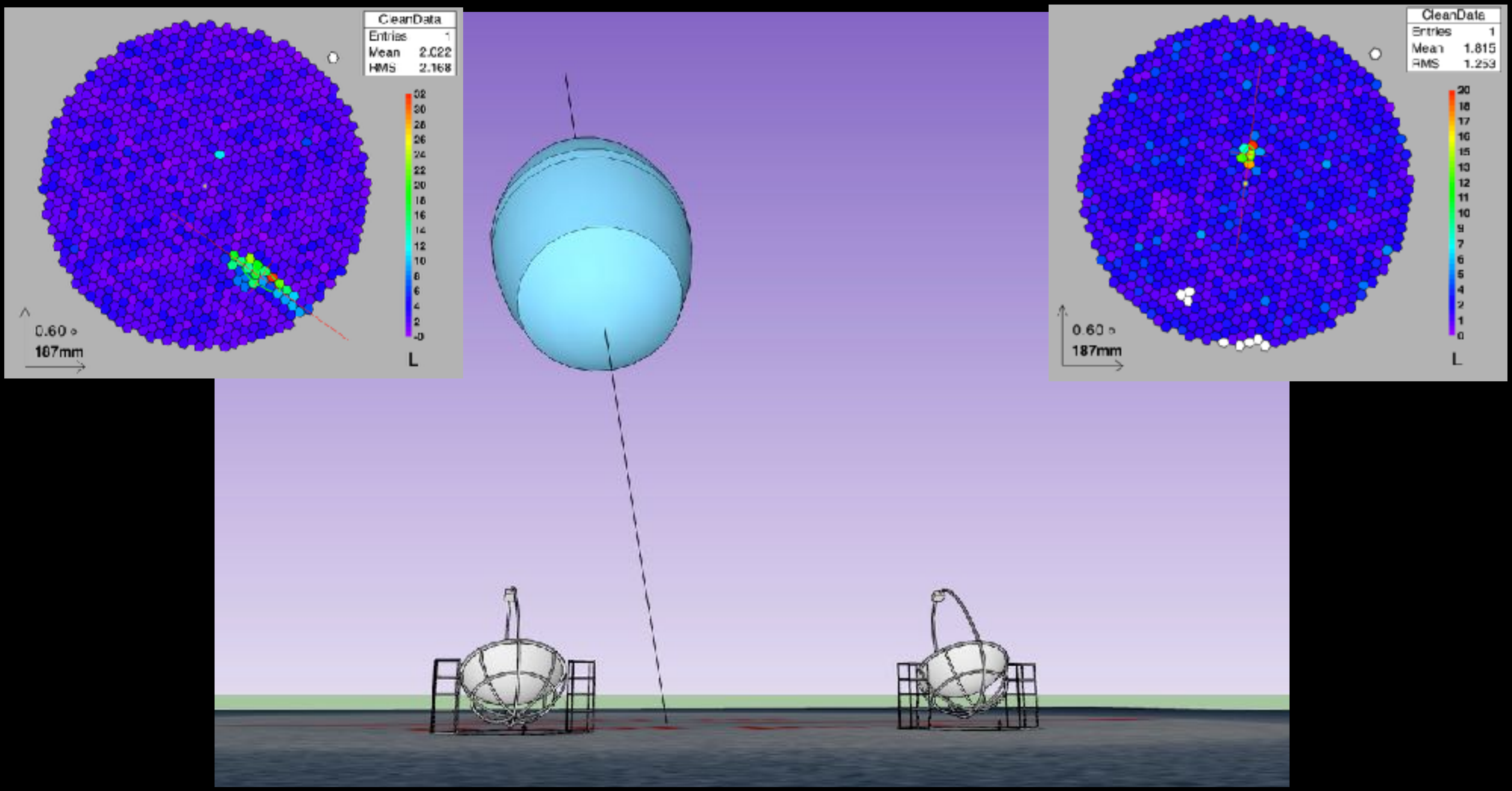
What is "Imaging Atmospheric Cherenkov Telescope (IACT)" ?

The shower shape can be seen as a elipse

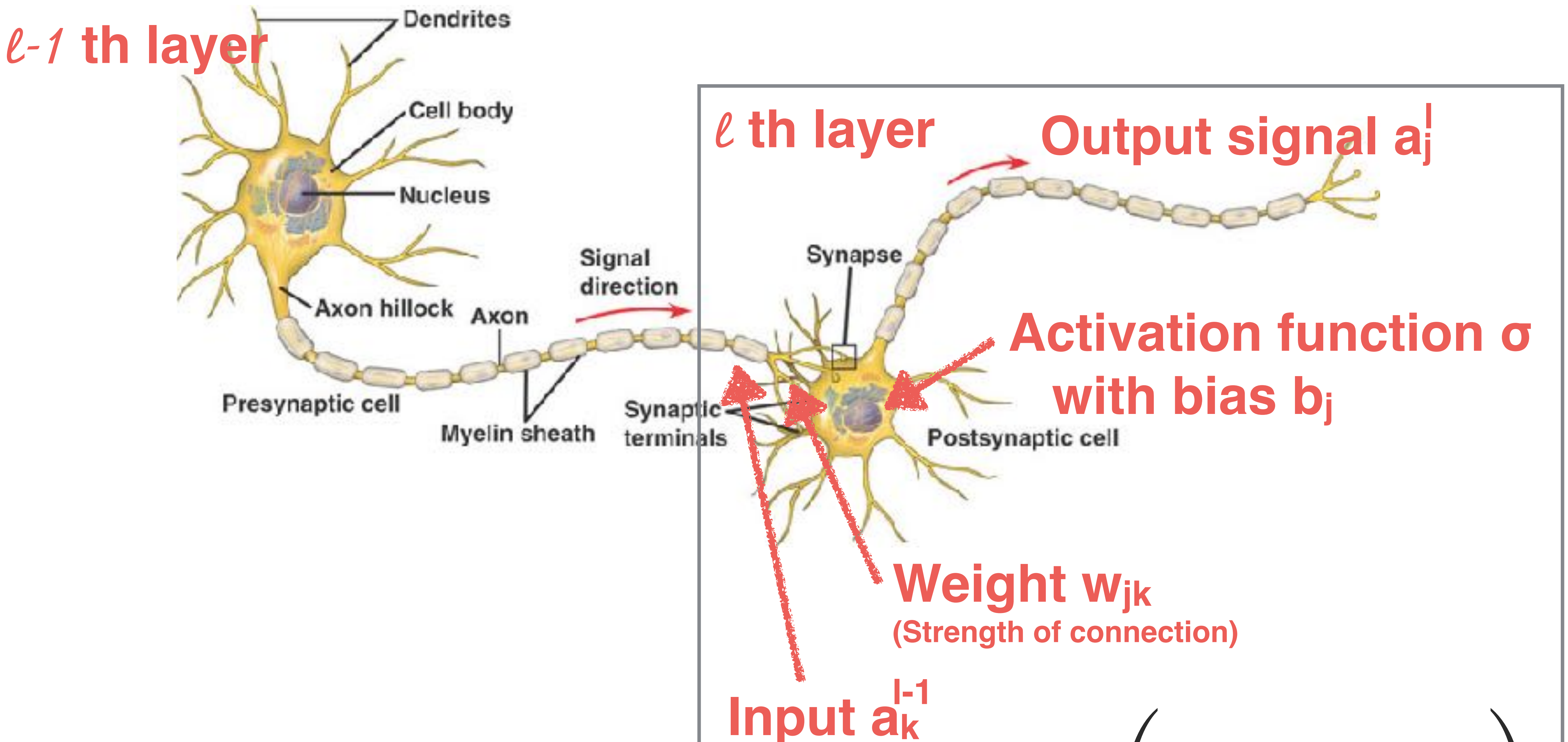


What is "Imaging Atmospheric Cherenkov Telescope (IACT)" ?

When a shower is seen from different positions

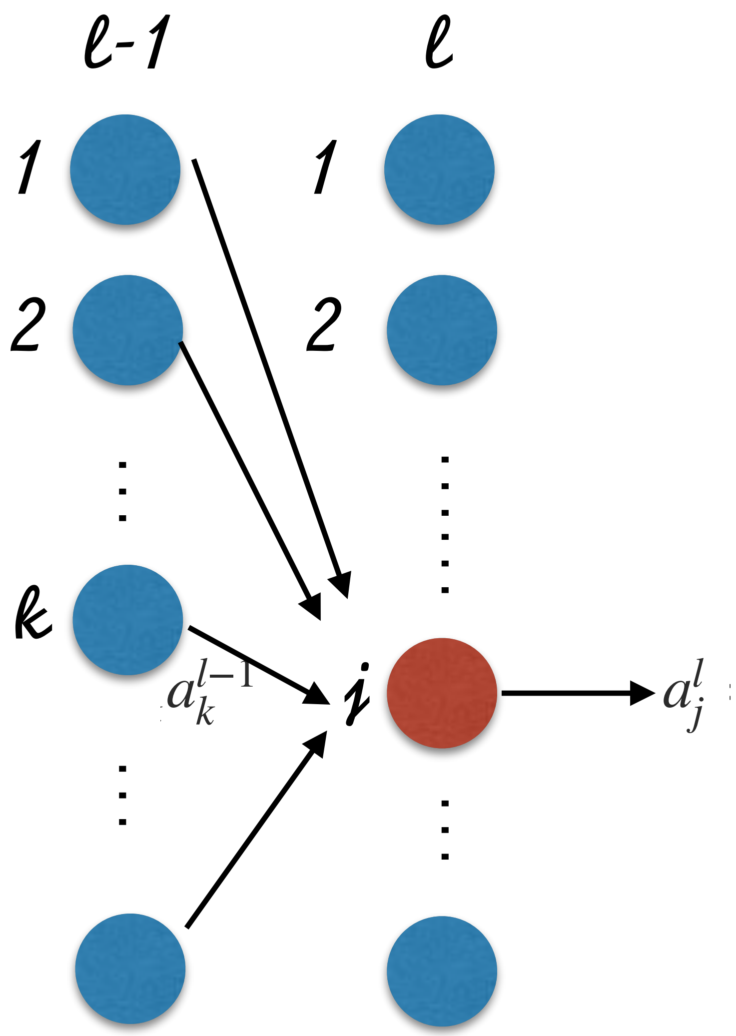


What's ANN?



$$a_j^l = \sigma\left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\right)$$

Back propagation(1)



At j-th node in l-th layer, output is

$$a_j^l = \sigma\left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\right) \qquad z_j^l = \sum_k w_{jk}^l a_k^{l-1} + b_j^l$$

Where a^l is output of activation function σ , w is weight to the input a^{l-1}, and b is bias.

Cost function C is

$$C = \frac{1}{2n}\sum_x \|y(x) - a^L(x)\|^2$$

Where y is true value, n is the number of train data x, and L is the number of layers.

$\partial C/\partial w_{jk}^l$ and $\partial C/\partial b_j^l$

are our interest, but let us define

$$\delta_j^l \equiv \frac{\partial C}{\partial z_j^l}.$$

Back propagation(2)

$$\delta_j^l = \frac{\partial C}{\partial z_j^l} = \sum_k \frac{\partial C}{\partial z_k^{l+1}} \frac{\partial z_k^{l+1}}{\partial z_j^l} = \sum_k \frac{\partial z_k^{l+1}}{\partial z_j^l} \delta_k^{l+1},$$

$$z_k^{l+1} = \sum_j w_{kj}^{l+1} a_j^l + b_k^{l+1} = \sum_j w_{kj}^{l+1} \sigma(z_j^l) + b_k^{l+1}$$

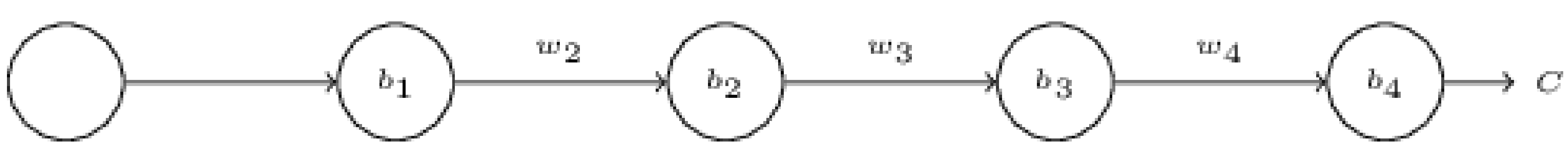
$$\frac{\partial z_k^{l+1}}{\partial z_j^l} = w_{kj}^{l+1} \sigma'(z_j^l).$$

$$\delta_j^l = \sum_k w_{kj}^{l+1} \delta_k^{l+1} \sigma'(z_j^l).$$

From the informations in l+1-th layer, we can obtain the error in l-th layer. (Back propagation)

In a very simple case(like composed of 4 layers, each has just one node)

$$\frac{\partial C}{\partial b_1} = \sigma'(z_1) \times w_2 \times \sigma'(z_2) \times w_3 \times \sigma'(z_3) \times w_4 \times \sigma'(z_4) \times \frac{\partial C}{\partial a_4}$$



1. **Input x:** Set the corresponding activation a^1 for the input layer.

2. **Feedforward:** For each $l = 2, 3, \ldots, L$ compute $z^l = w^l a^{l-1} + b^l$ and $a^l = \sigma(z^l)$.

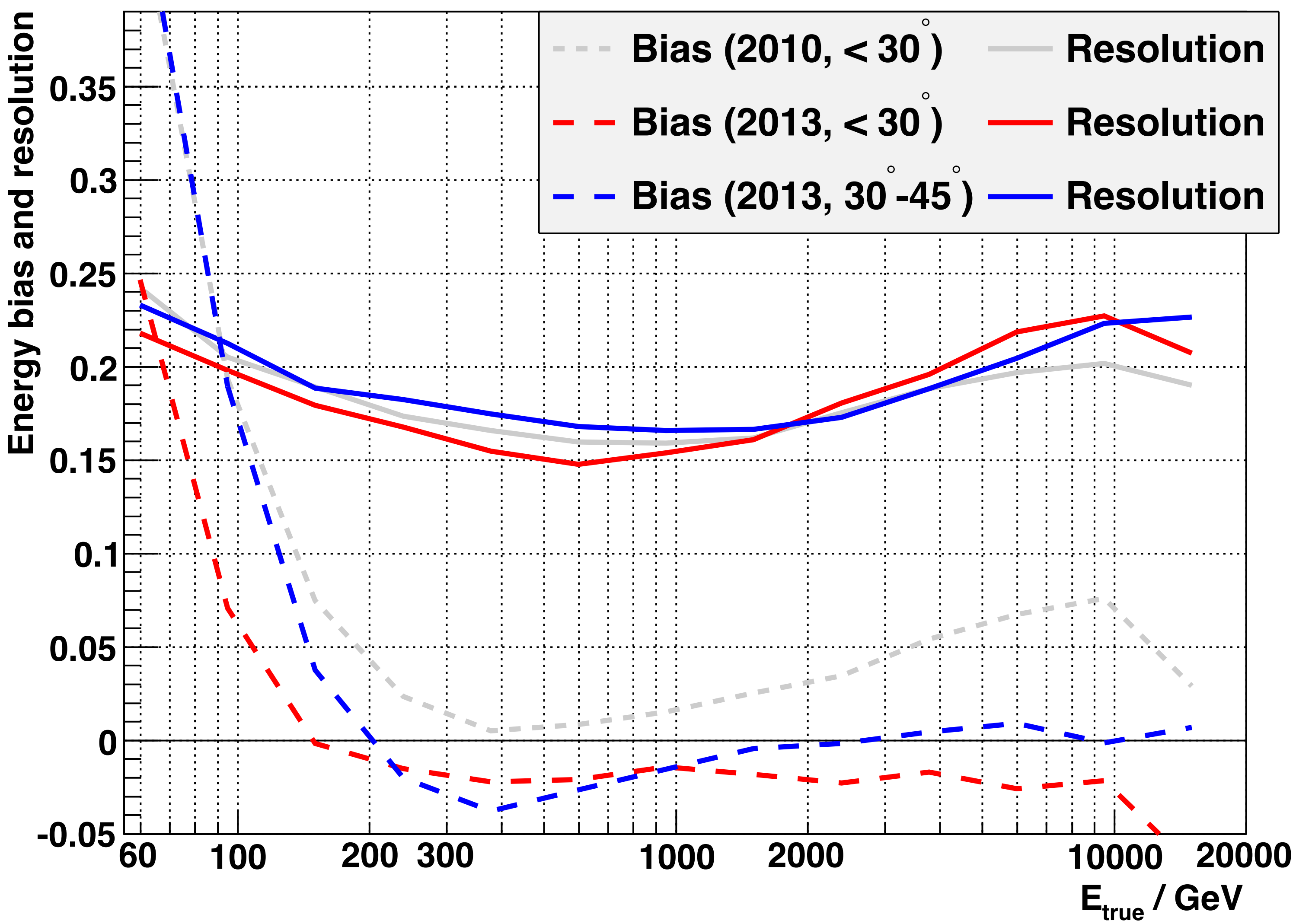
3. **Output error δ^L:** Compute the vector $\delta^L = \nabla_a C \odot \sigma'(z^L)$.

4. **Backpropagate the error:** For each $l = L-1, L-2, \ldots, 2$ compute $\delta^l = ((w^{l+1})^T \delta^{l+1}) \odot \sigma'(z^l)$.

5. **Output:** The gradient of the cost function is given by $\frac{\partial C}{\partial w^l_{jk}} = a^{l-1}_k \delta^l_j$ and $\frac{\partial C}{\partial b^l_j} = \delta^l_j$.

And move w and b in different direction to gradient

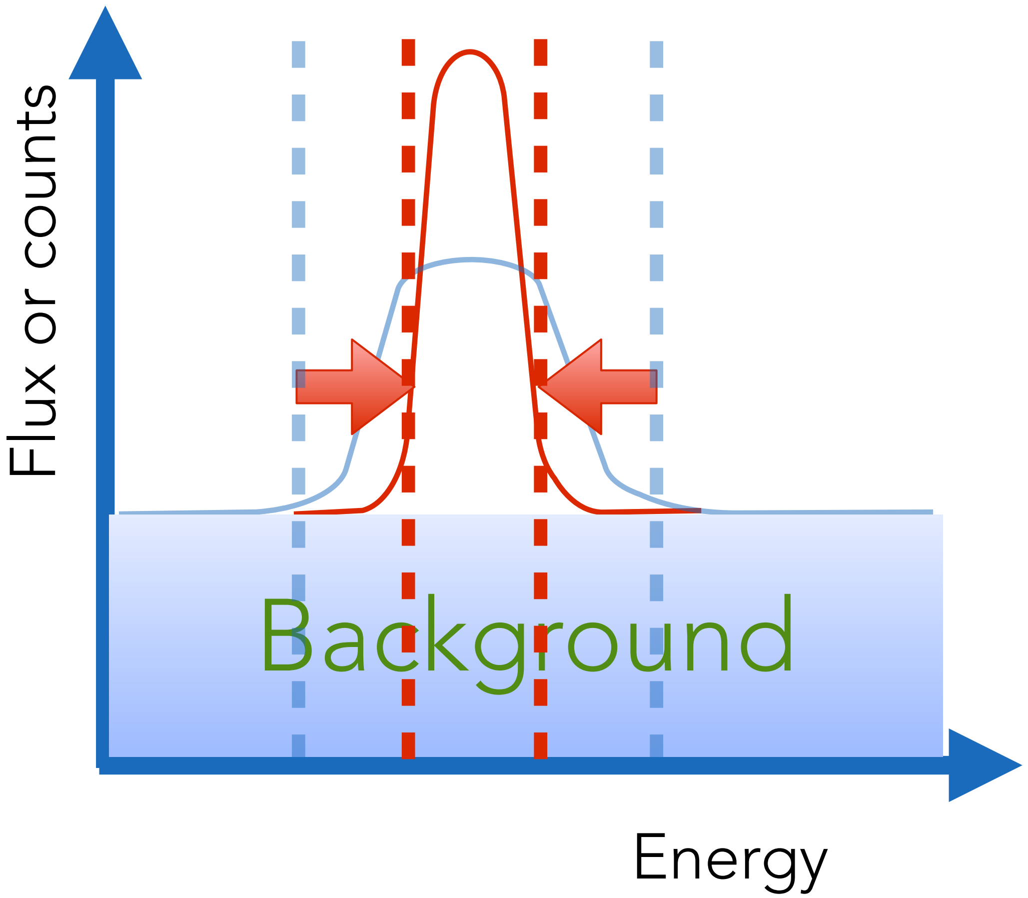
Official Performance



My study features
Zd = 05 - 50 deg

Figure 10: Energy resolution (solid lines) and bias (dashed lines) obtained from the MC simulations of $\gamma-$rays. Events are weighted in order to represent a spectrum with a slope of -2.6. Red: low zenith angle, blue: medium zenith angle. For comparison, pre-upgrade values from Aleksić et al. (2012a) are shown in gray lines.

Arxiv 1409.5594 The major upgrade of the MAGIC telescopes, Part II: A performance study using observations of the Crab Nebula


Flux or counts
Background
Energy