Improving the gamma-hadron separation for air showers at the IceCube Neutrino Observatory

Machine Learning workshop @ Delaware University

Karlsruher Institut für Technologie (KIT)
Institut für Astroteilchenphysik (IAP)

February 03, 2022

Bontempo, Federico

PhD student
Institut für Astroteilchenphysik (IAP)
Campus Nord, Geb. 425
Karlsruher Institut für Technologie (KIT)
Postfach 3640
D-76021 Karlsruhe
Telefon: +49-721-608-22426
Email: federico.bontempo@kit.edu

European Research Council
Established by the European Commission

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 802729).

ICECUBE
SOUTH POLE NEUTRINO OBSERVATORY

Previous works

2 PhD students from Delaware University (USA):

- SEARCH FOR PEV GAMMA RAYS WITH THE ICECUBE OBSERVATORY
  Zachary Dean Griffith
  ➔ Focused on ML techniques for Gamma Hadron discrimination

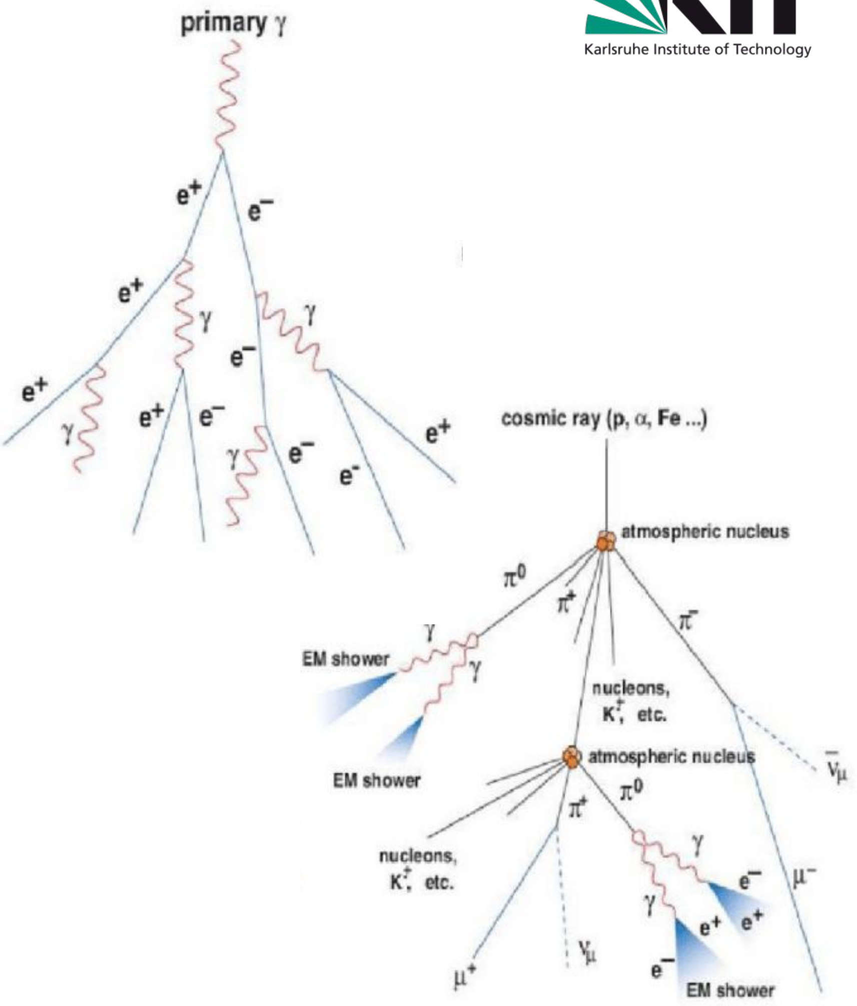
- SEARCH FOR PEV GAMMA RAYS AND ASTROPHYSICAL NEUTRINOS WITH ICETOP AND ICECUBE
  Hershal Pandya
  ➔ Developed Log Likelihood Ratio parameter

Gamma hadron separation

- Production of air showers:
  - gamma-ray primary
  - hadronic primary

- Gamma-ray induced air showers:
  - fewer muons
  - less shower fluctuations
  - narrower lateral spread

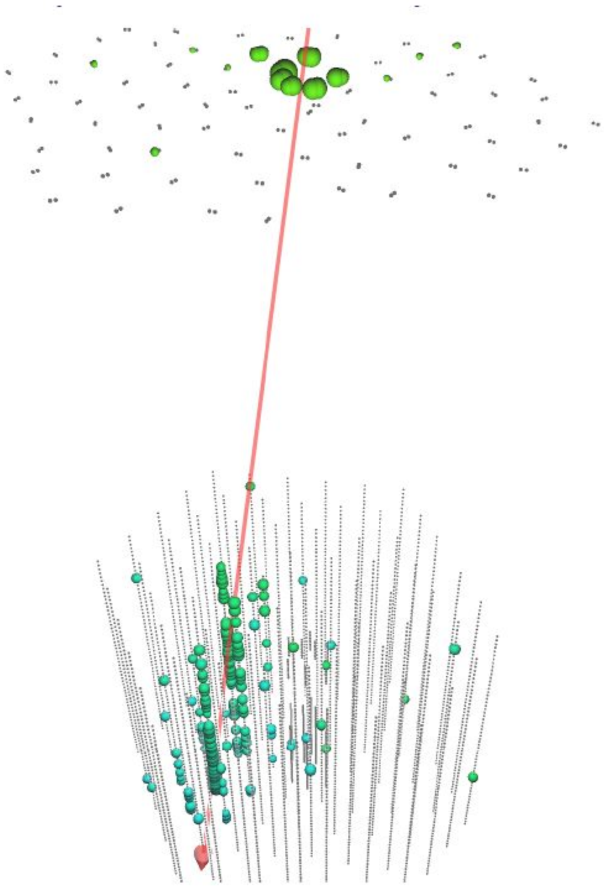
- Hadronic air showers:
  - richer in muon content
  - more shower fluctuations
  - wider lateral spread



Images source: Multi-year Campaign of the Gamma-Ray Binary LS I +61◦ 303 and Search for VHE Emission from Gamma-Ray Binary Candidates with the MAGIC Telescopes
doi: 10.13140/RG.2.1.4140.4969

IceCube Neutrino Observatory

- Ice-Top:
  - 1 km2
  - 162 Cherenkov tanks
  - measurement of the electromagnetic component of the shower
- In-ice:
  - 1 km3
  - 86 strings with in-ice with 60 DOMs each
  - measurement of the high energy muons



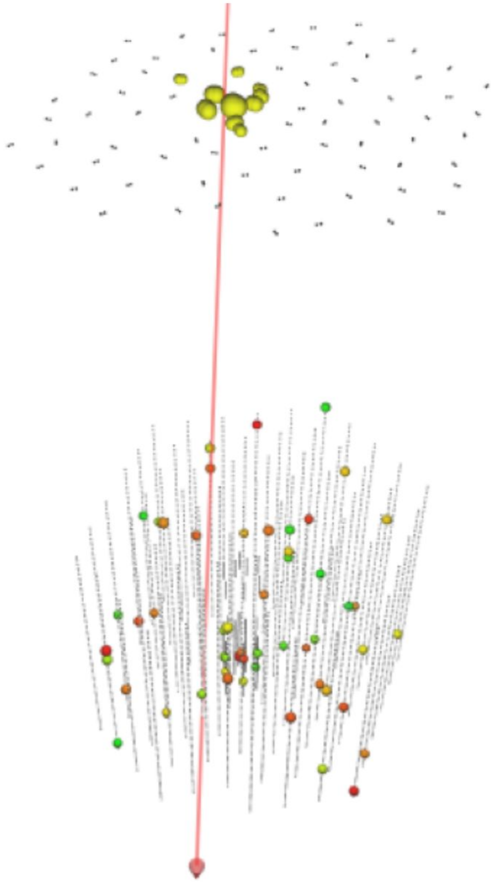
A cosmic-ray air shower event with a shower axis that passes through both components of the detector. (Z. Griffith)

Source: SEARCH FOR PEV GAMMA RAYS WITH THE ICECUBE OBSERVATORY, Zachary Dean Griffith

IceCube Neutrino Observatory

- Ice-Top:
  - 1 km2
  - 162 Cherenkov tanks
  - measurement of the electromagnetic component of the shower

- In-ice:
  - 1 km3
  - 86 strings with in-ice with 60 DOMs each
  - measurement of the high energy muons



An event from the 2012 point source sample dataset with one of the ten highest Random Forest scores. (Z. Griffith)

Source: SEARCH FOR PEV GAMMA RAYS WITH THE ICECUBE OBSERVATORY, Zachary Dean Griffith

Find Gamma ray sources

1. Data (burning sample) and Sim (Sibyll 2.1) .i3 -> .hdf5
2. Cleaning & selection
3. Train Random Forest
4. Data selection with the Random Forest -> Gamma selection
5. Background creation
6. All sky search
7. Test statistic
8. P-value
9. Final plot

Completed for year 2012, 2013, 2014 and 2015
burning sample year 2011 is now available

Simulations:
https://wiki.icecube.wisc.edu/index.php/Cosmic-ray_IC86_Datasets

Data and Sim Cleaning & selection

- Quality cuts:
  - Number of triggered Station >= 5
  - Fraction Containment < 1.0 (contained in the in the surface array)
  - Laputop zenith < arccos(0.8)
  - log10(S125) > -0.25
  - Laputop Beta > 1.4 & Laputop Beta < 9.5
  - log10(Energy / GeV) < 8.0

- Features for Random Forest:

1. charges: total in-ice charge
2. Laputop in-ice FractionContainment
3. Laputop log10(s125)
4. sin(dec)
5. LLH_Ratio: log-likelihood parameter (Hershal PhD thesis)

Log likelihood ratio

The log-likelihood ratio is constructed for each event via 3 PDFs

e.g. The log-likelihood ratio for charge Q and lateral distance R:

$$\Lambda_{QR} = \log_{10}\left(\frac{L_{QR}(\text{event}|H_\gamma)}{L_{QR}(\text{event}|H_{CR})}\right), \qquad L_{QR}(\text{event}|H) = \prod_{\mathrm{i}=1}^{162} P\left(Q_\mathrm{i},\ R_\mathrm{i}|\,H\right),$$

with P (Qi, Ri | H) being the probability of observing a tank with measured charge Qi and at lateral distance Ri, for the hypothesis H.

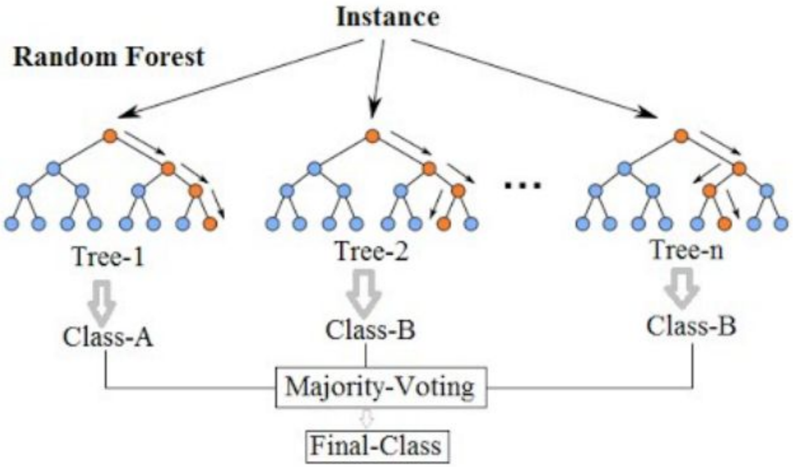
Total log-likelihood ratio is defined by the sum of all three log-likelihood ratios

$$\Lambda = \Lambda_{QR} + \Lambda_{Q\Delta T} + \Lambda_{\Delta TR}$$

Source: SEARCH FOR PEV GAMMA RAYS AND ASTROPHYSICAL NEUTRINOS WITH ICETOP AND ICECUBE, Hershal Pandya

Random Forest

The Random Forest: open-source
Python package Scikit-Learn



A Random Forest classifier consists of
a combination of many simple decision trees

Each tree acquires its n events from
the training sample, such that
each tree trains on a different set of events every time

The splitting condition: minimization of the Gini impurity in the child nodes

$$I_G = 1 - \frac{w_S^2 + w_B^2}{(w_S + w_B)^2}$$

wS and wB are the total weights in a node
for the signal and background classes

“probability of misclassification”:
all of the weight in the node is in one class IG = 0
an even split in the node results in IG = 0.5.

Source: https://scikit-learn.org/stable/

Random Forest

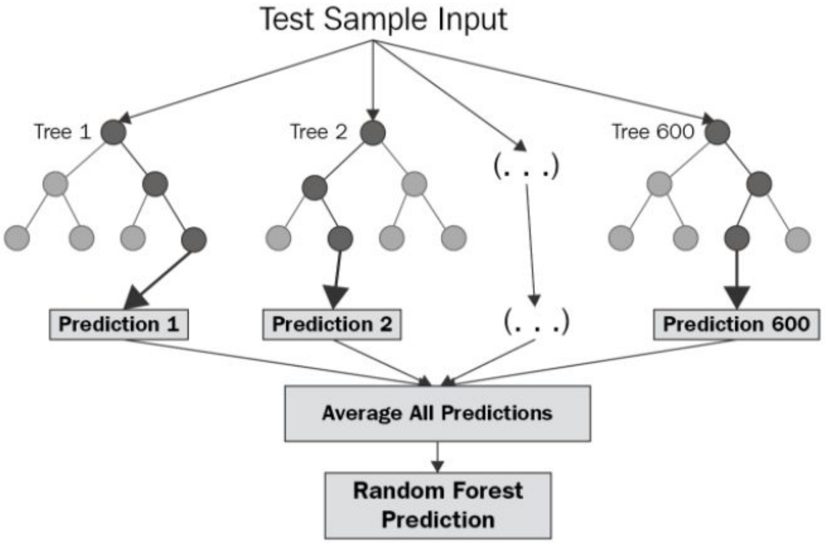
The Random Forest: open-source Python package Scikit-Learn

A Random Forest regressor consists of a combination of many simple decision trees

The splitting condition: minimization of the Mean Square Error (MSE)

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

MSE	= mean squared error
n	= number of data points
Y_i	= observed values
$\hat{Y}_i$	= predicted values



Source: https://scikit-learn.org/stable/

Training & Testing

- Training sample:
  - about 80.000 Gamma MC
  - about 80.000 Data as background
- Test sample:
  - about 20.000 Gamma MC
  - the complete burn sample about 4 Millions event

- Note: a different forest was trained for each year (snow accumulation)

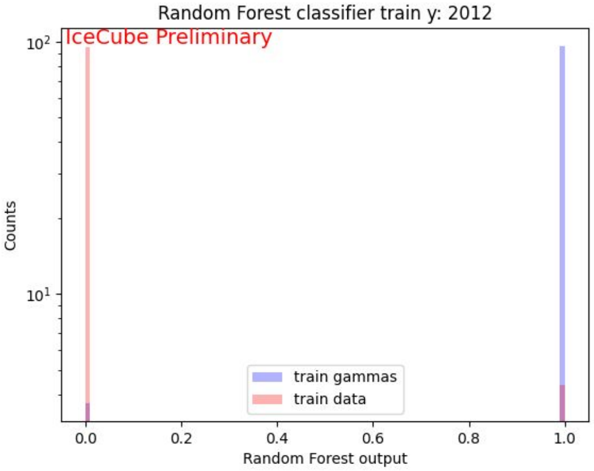
```
RandomForestClassifier(n_estimators=100,
                       random_state=0,
                        n_jobs=5, verbose=0,
                        max_depth=6)
```

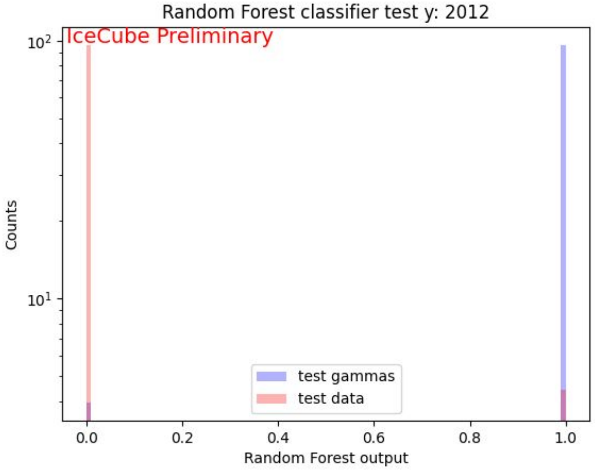
cut value=0.7

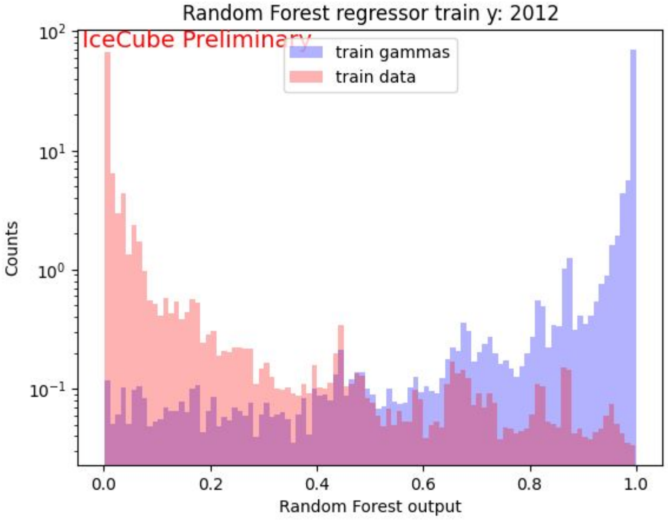
```
RandomForestRegressor(n_estimators=100,
                      random_state=0,
                      n_jobs=5, verbose=0,
                      max_depth=8)
```

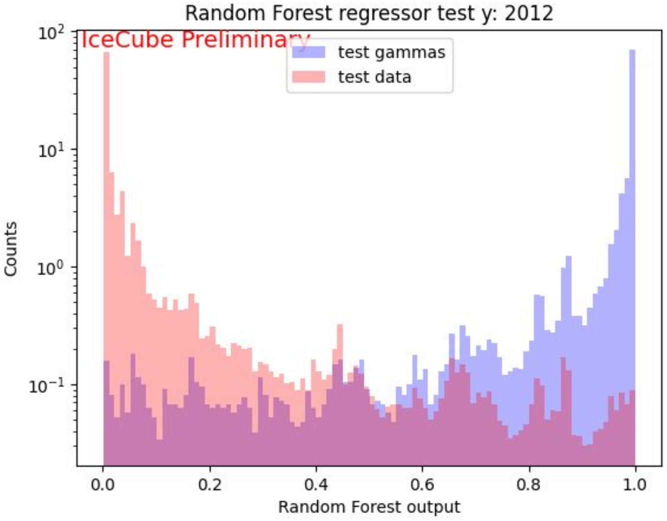
cut value=0.99

Training & Testing year 2012



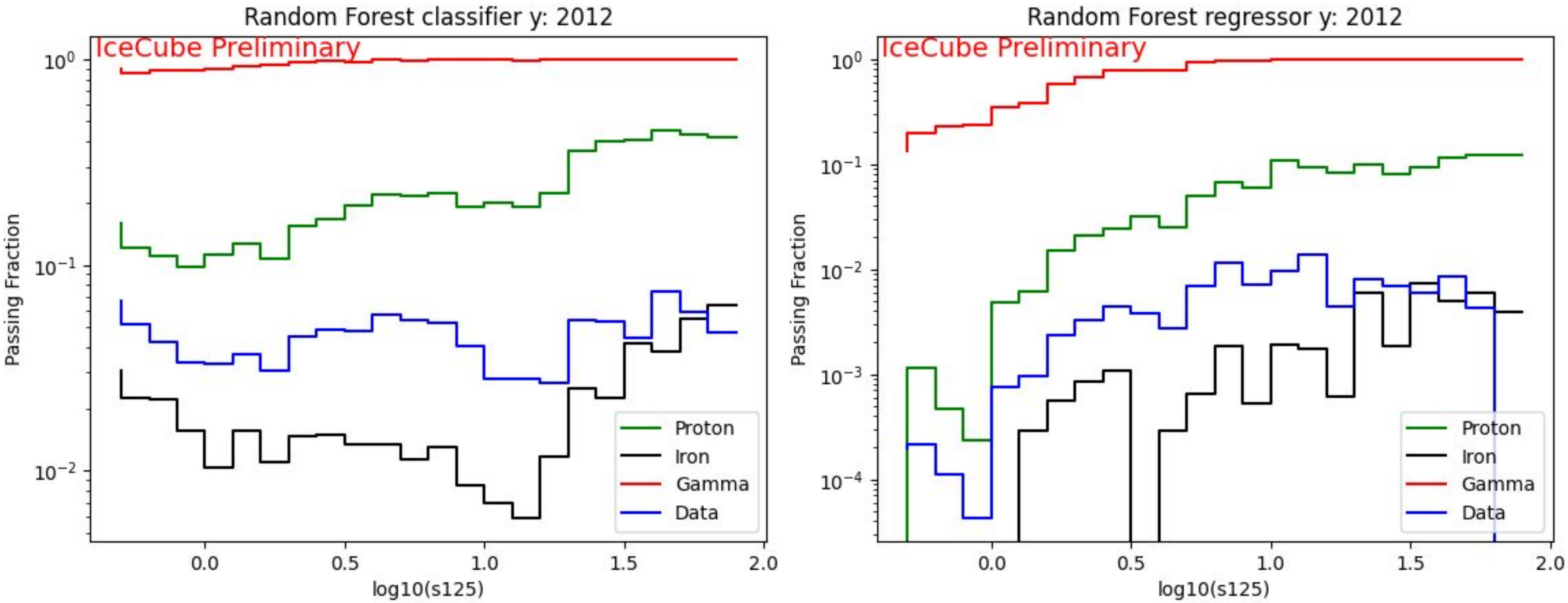






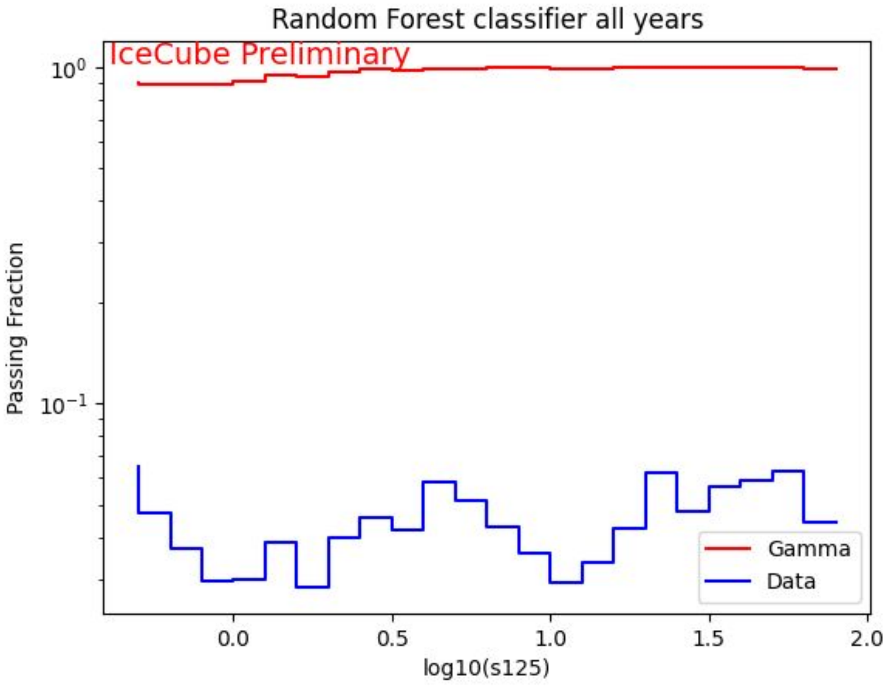
Passing fraction year 2012

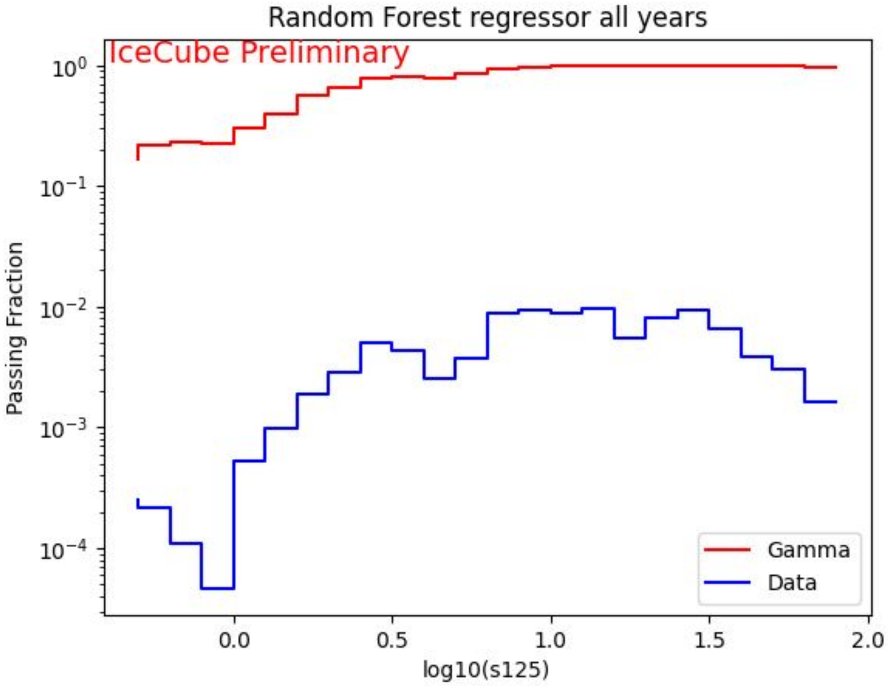
- event that passed the prediction cut / Total event
- plotted as function of energy



Passing fraction all years

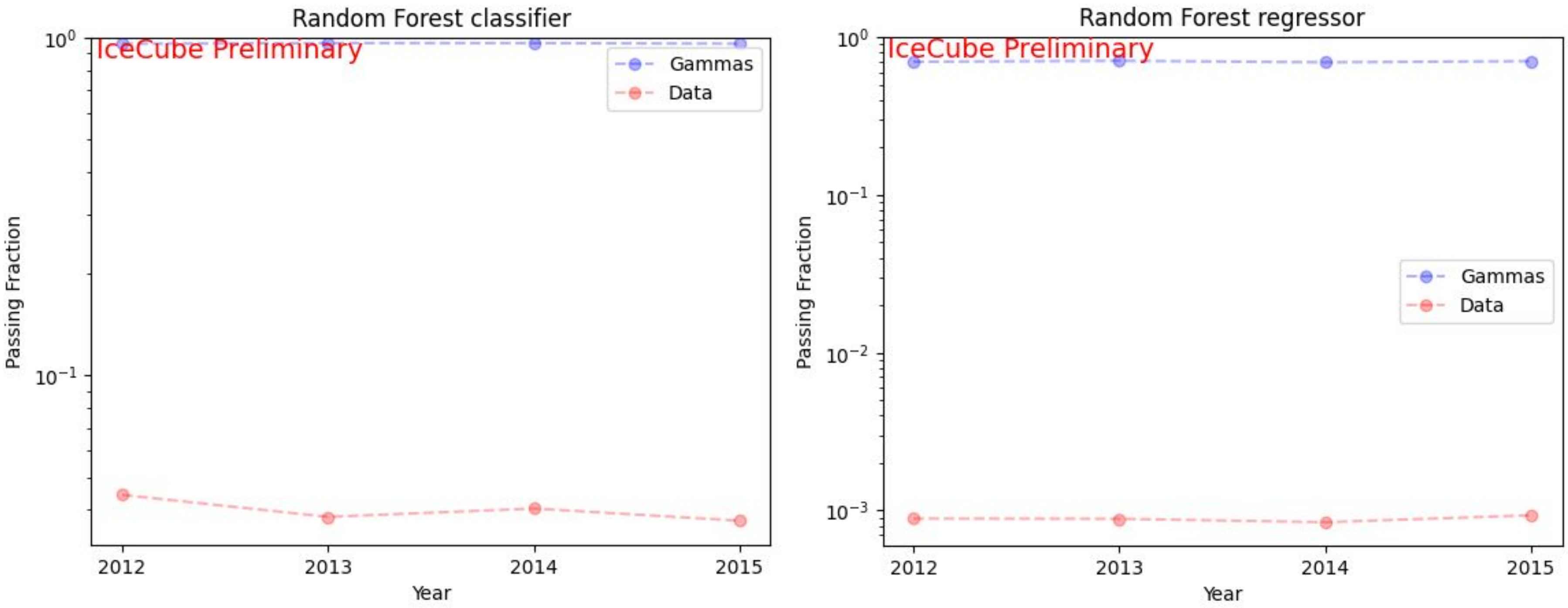
- event that passed the prediction cut / Total event
- plotted as function of energy





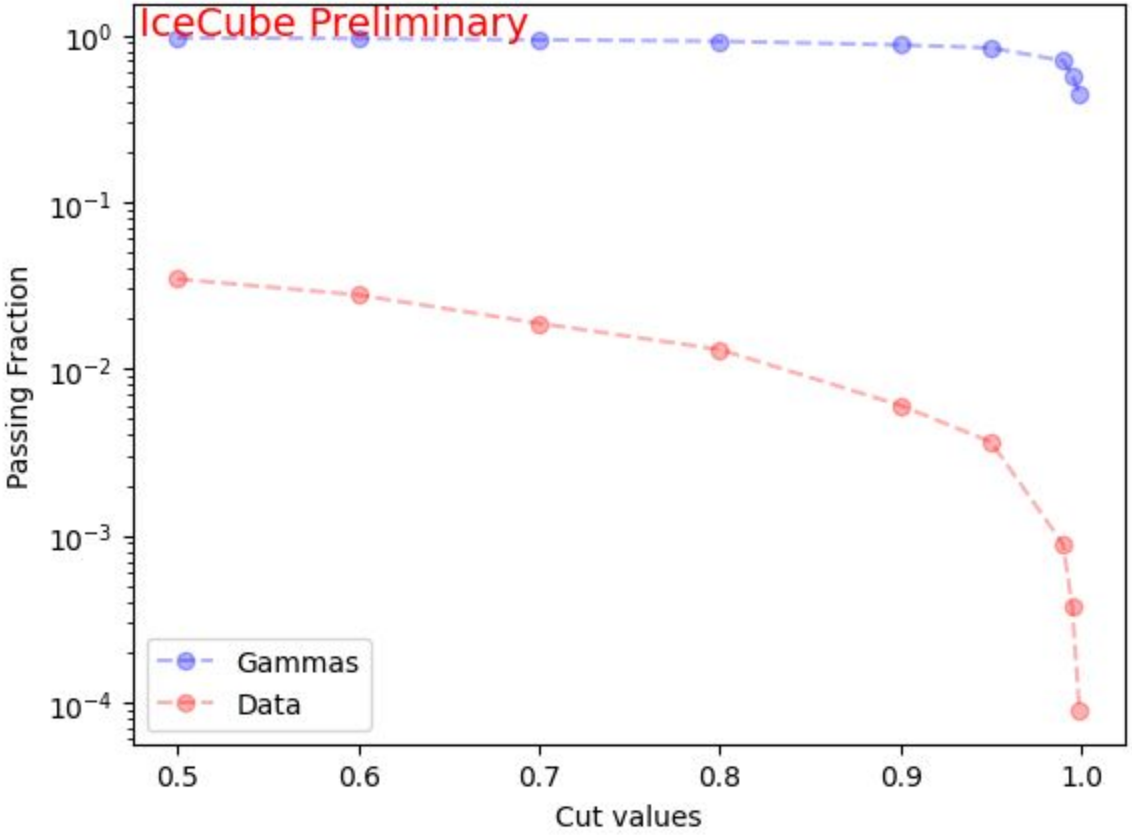
Passing fraction all years

- event that passed the prediction cut / Total event

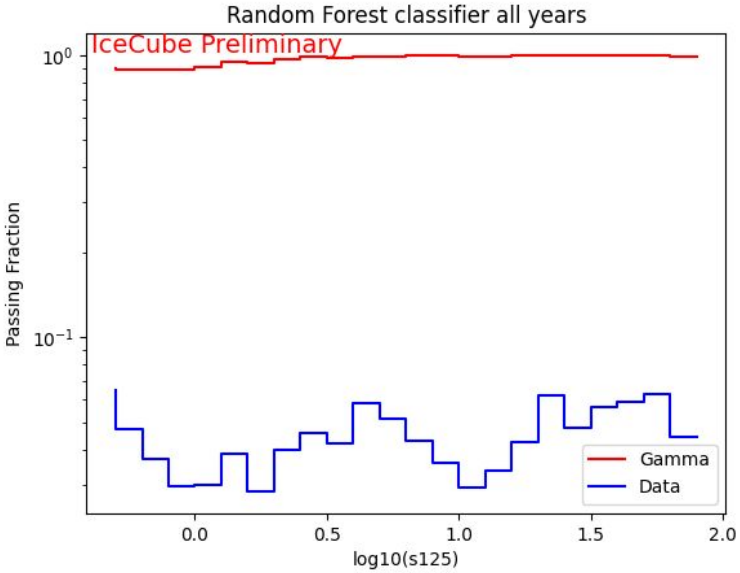


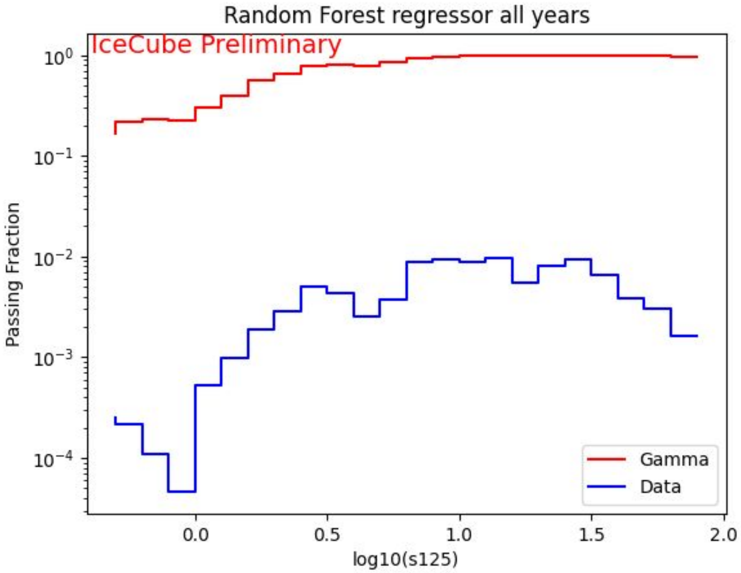
Passing fraction regressor all years

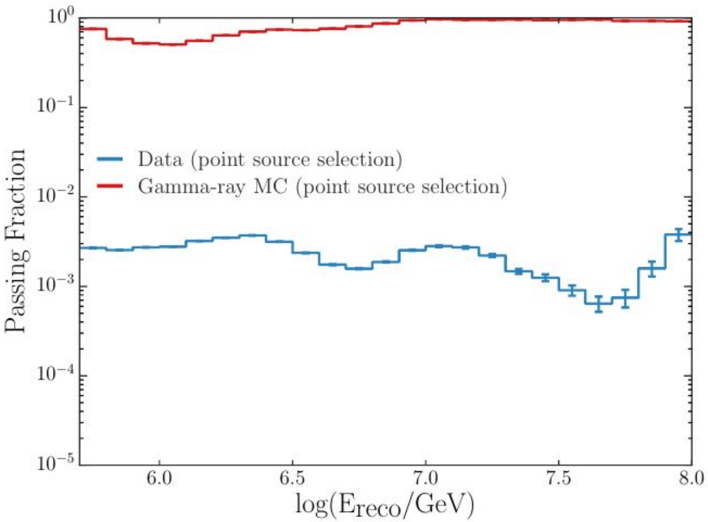
- Passing fraction for different cut values: [0.5,0.6,0.7,0.8,0.9,0.95,0.99,0.995,0.999]



Comparison with Zach’s results



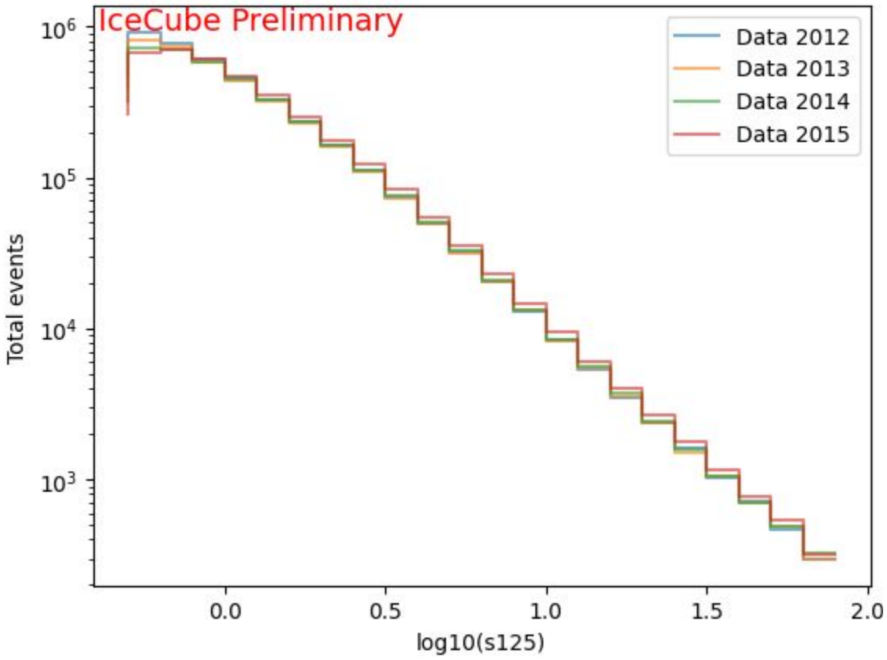


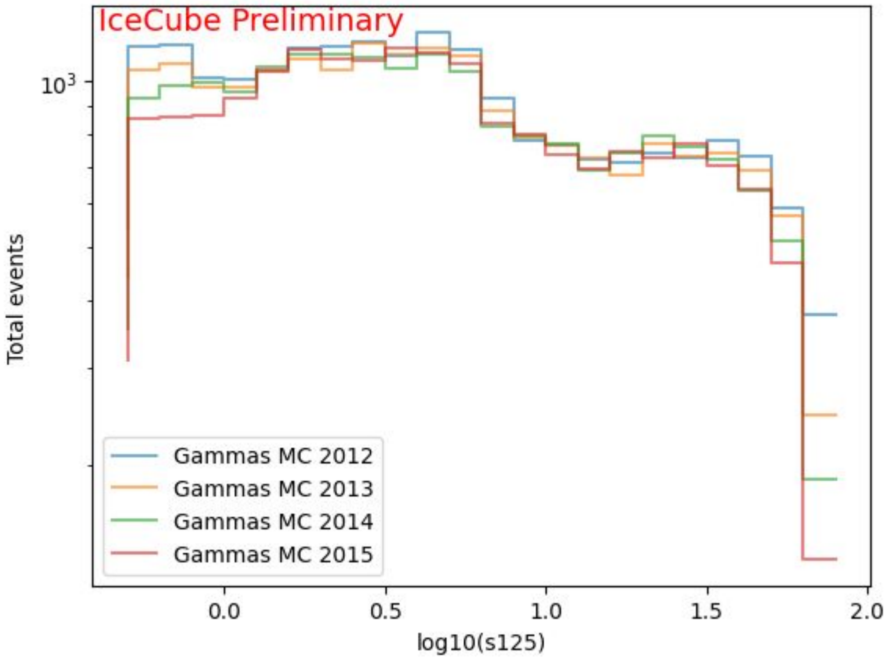


Source: SEARCH FOR PEV GAMMA RAYS WITH THE ICECUBE OBSERVATORY, Zachary Dean Griffith

Data and Sim Cleaning & selection

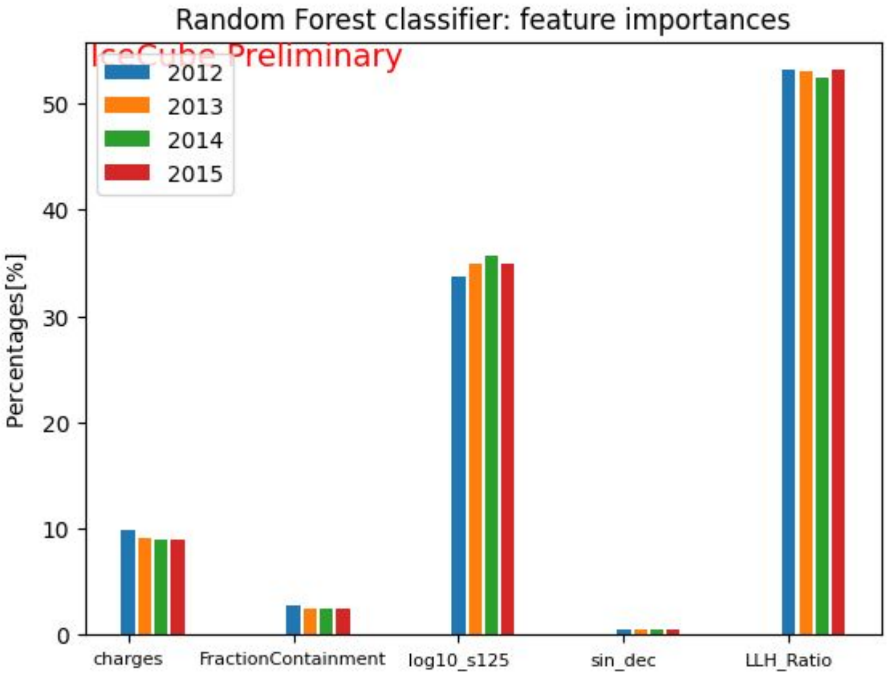
Sim and Data do not have the same steepness

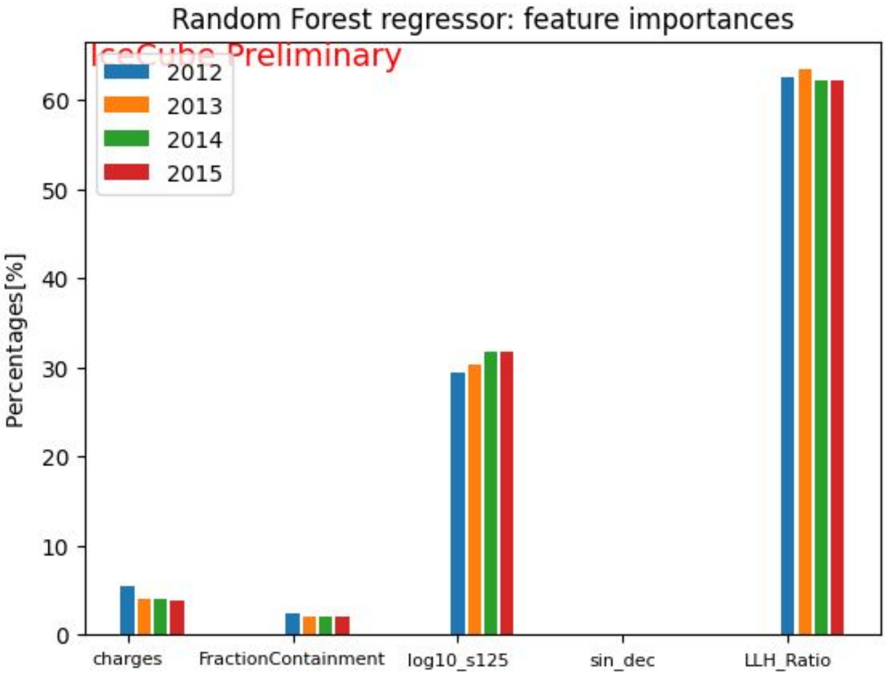




Feature importance all years

Each of the input feature has a different importance.
For each year, here they are plotted in percentages and sum up to 1

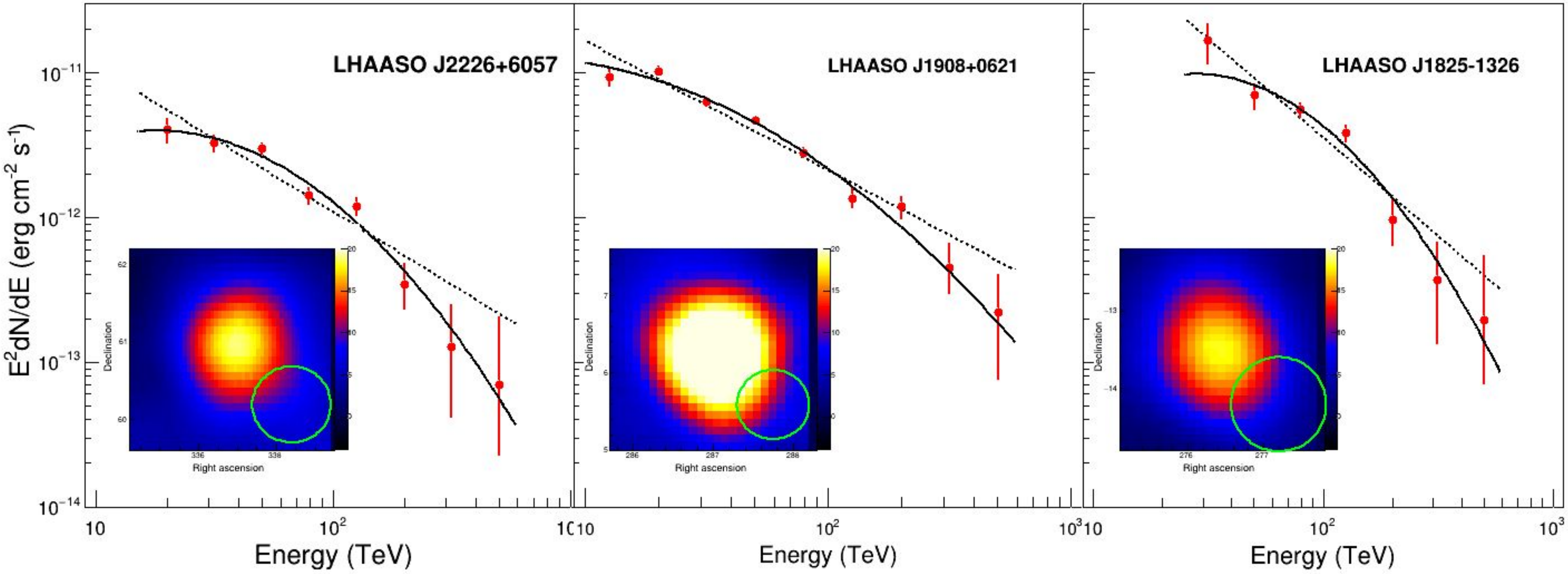




LHAASO sources

- Ultra High-energy photons up to 1.4 petaelectronvolts from 12 γ-ray Galactic sources

- Is IceCube able to detect a LHAASO like UHE photon source?

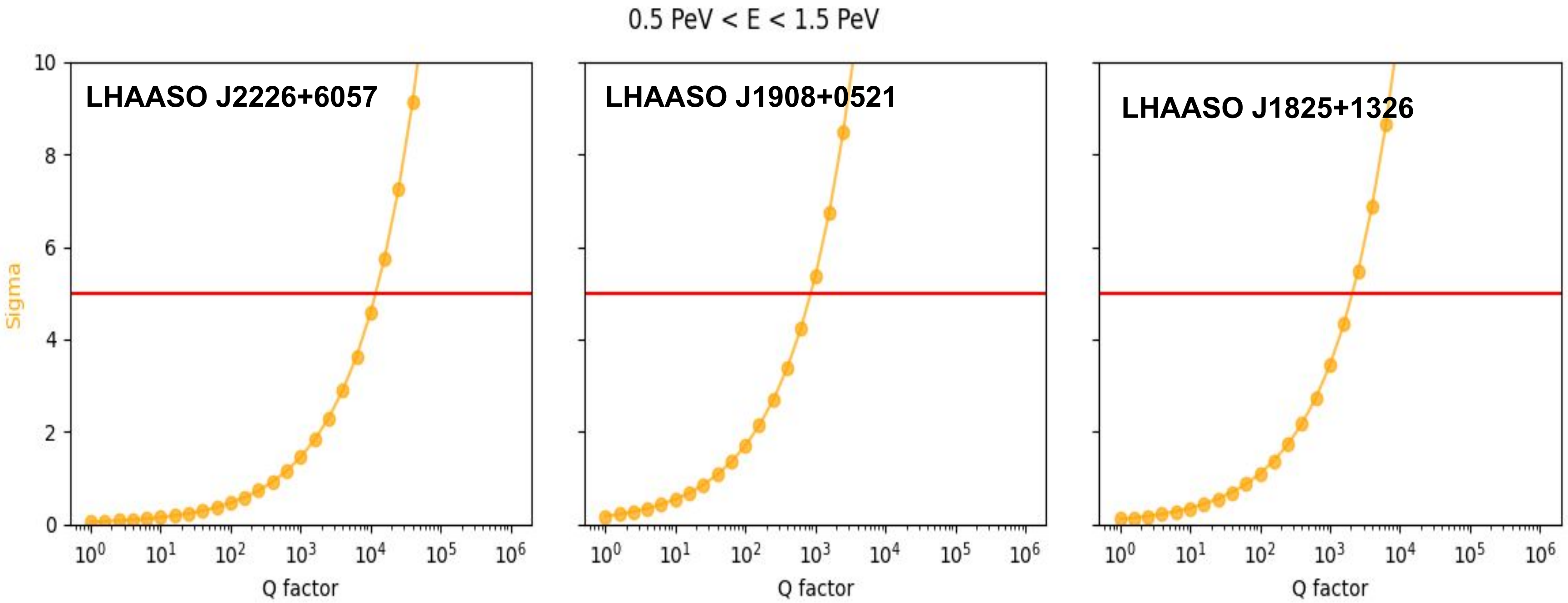


Source: https://www.nature.com/articles/s41586-021-03498-z

Sigma and Q-factor

- Q-factor indicates how good the CR background can be rejected
- A Q-factor between $10^3 \div 10^4$ is required for signal bg separation for a 5 sigma observation

- # sigma = signal / bg**0.5



Bontempo, Federico - *federico.bontempo@kit.edu*

Summary & Outlook

- The classifier and the regressor are two different approaches
- Better gamma hadron separation for lower energies probably due to lower statistics for data at higher energies
- The declination is almost irrelevant for the final selection
- Regressor and Classifier depend on the cut value selected

Coming next:

- Correct the energy bias between simulation and data
- Search for gamma sources in the sky
- Use Deep Learning neural networks for the classification

Karlsruher Institut für Technologie

Improving the gamma-hadron separation for air showers at the IceCube Neutrino Observatory

Machine Learning workshop @ Delaware University

Karlsruher Institut für Technologie (KIT)
Institut für Astroteilchenphysik (IAP)

February 03, 2022

Bontempo, Federico

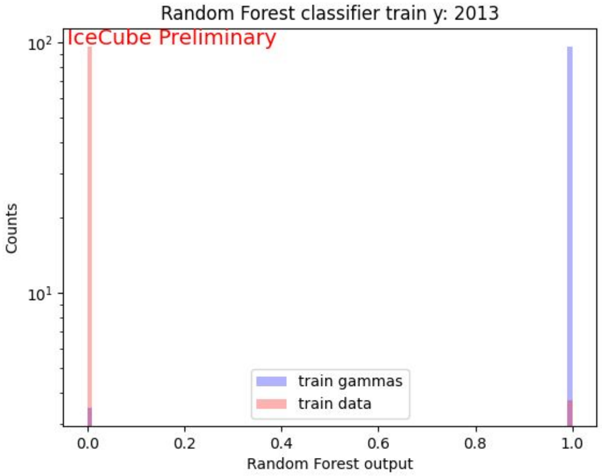
PhD student
Institut für Astroteilchenphysik (IAP)
Campus Nord, Geb. 425
Karlsruher Institut für Technologie (KIT)
Postfach 3640
D-76021 Karlsruhe
Telefon: +49-721-608-22426
Email: federico.bontempo@kit.edu

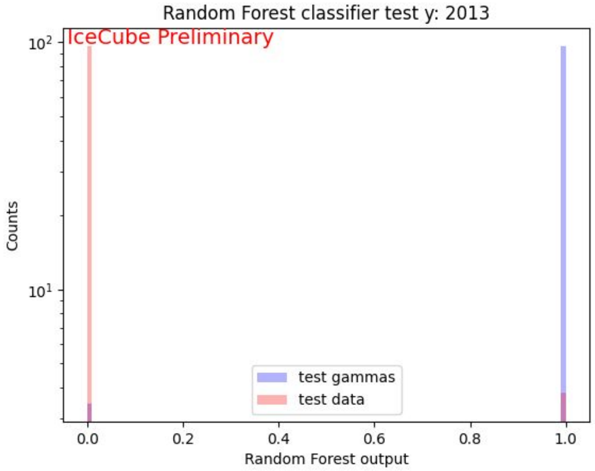
European Research Council
Established by the European Commission

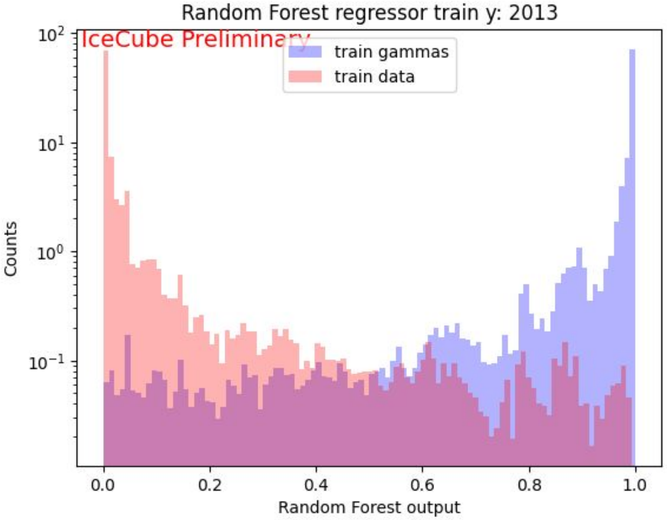
This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 802729).

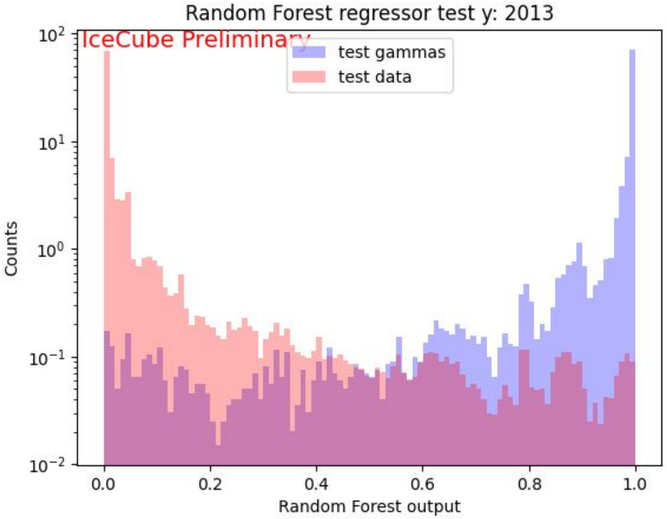
ICECUBE
SOUTH POLE NEUTRINO OBSERVATORY

Training & Testing year 2013



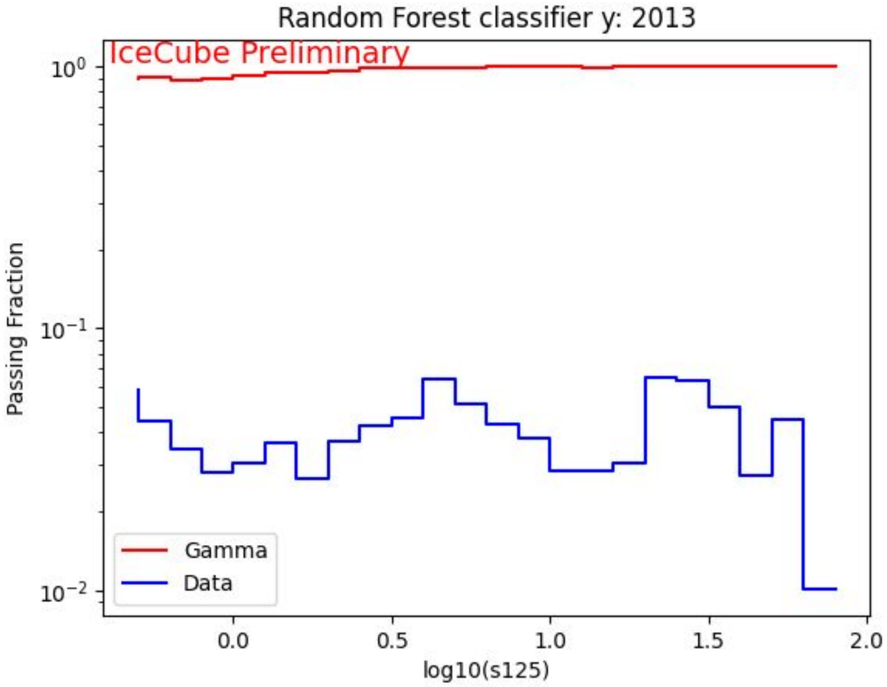


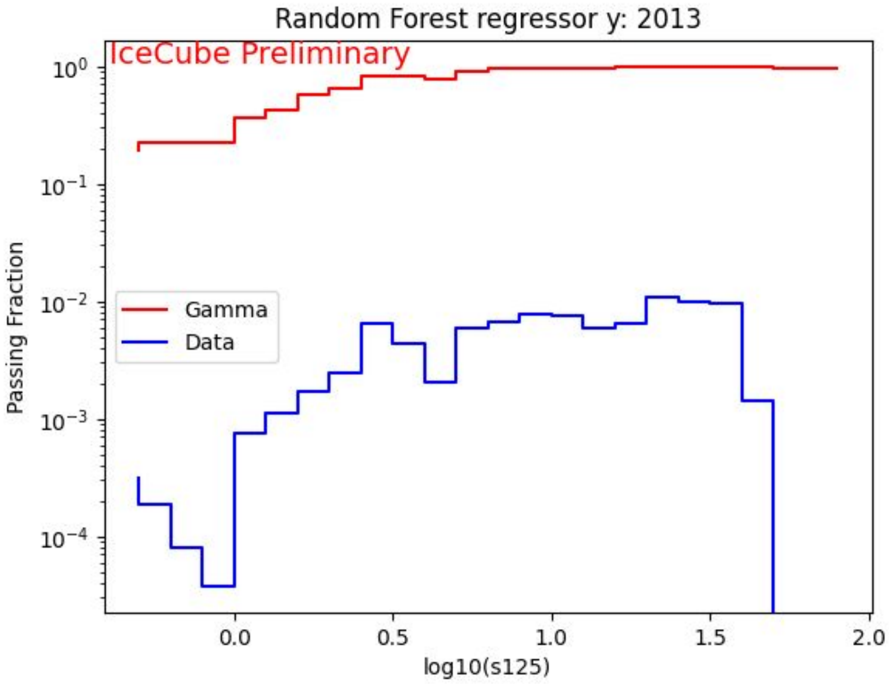




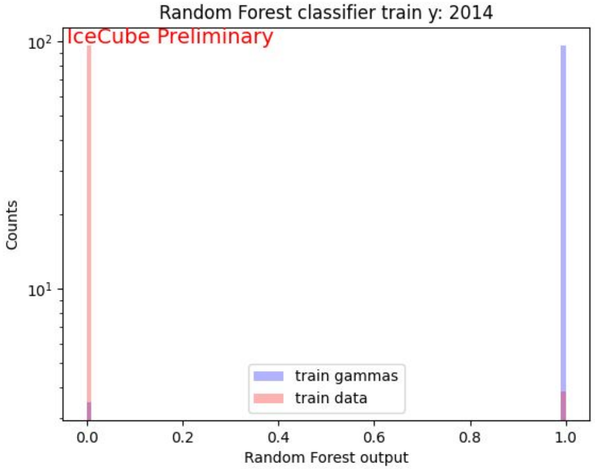
Passing fraction year 2013

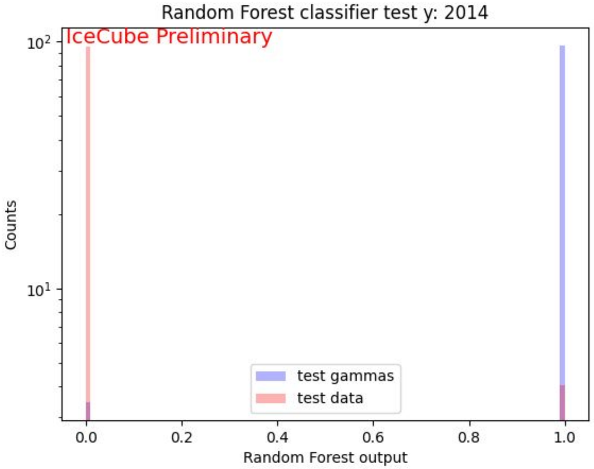
- event that passed the prediction cut / Total event
- plotted as function of energy

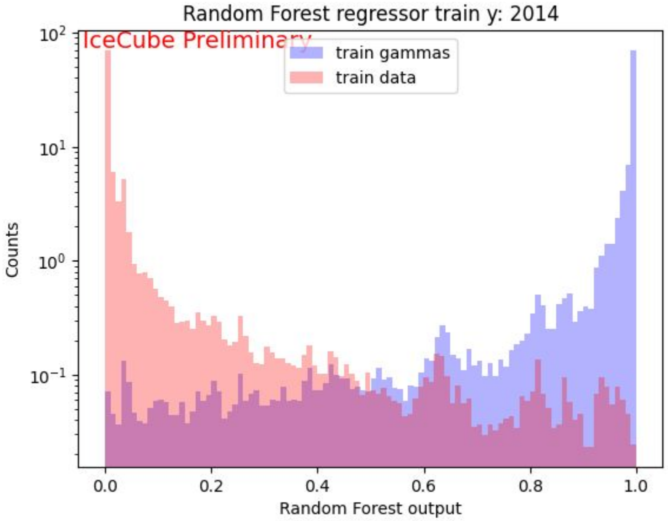


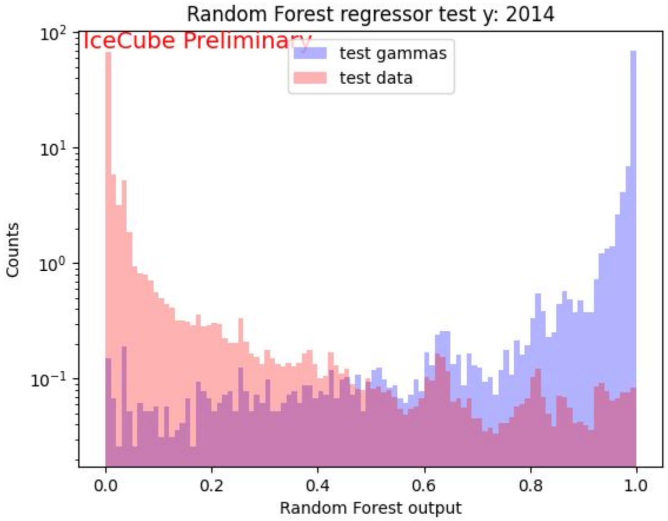


Training & Testing year 2014


Random Forest classifier train y: 2014

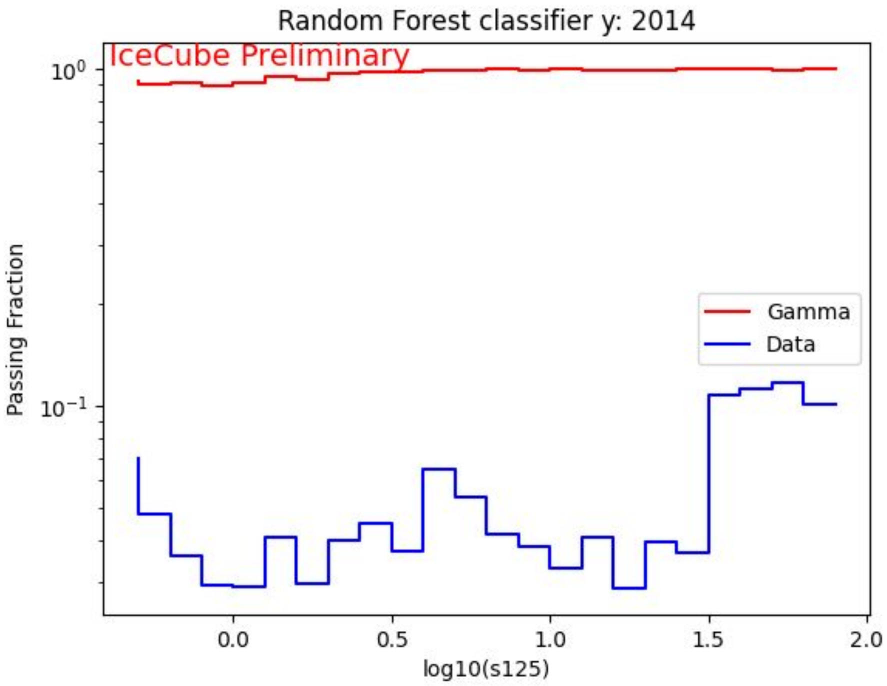

Random Forest classifier test y: 2014

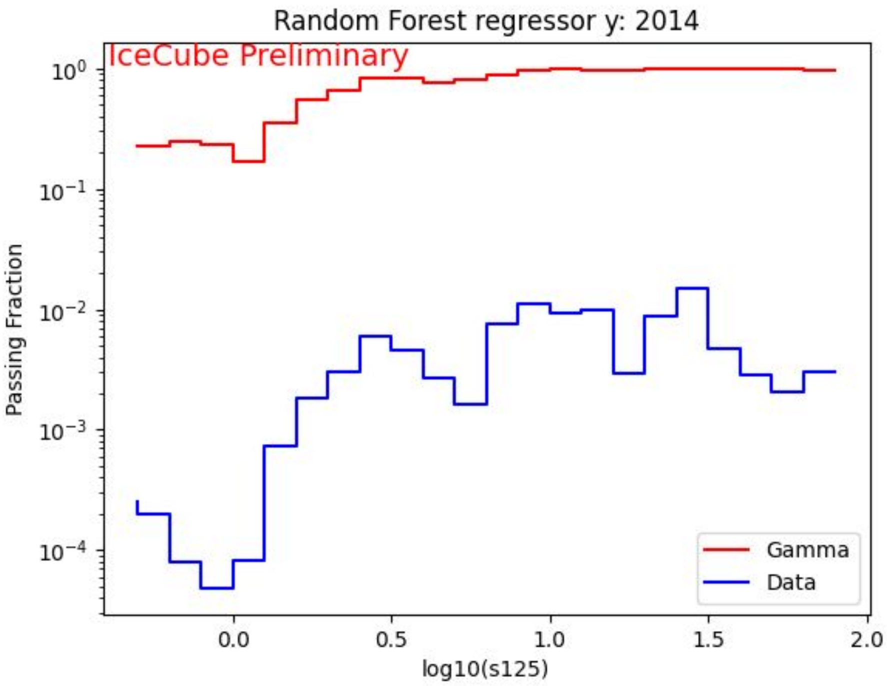

Random Forest regressor train y: 2014


Random Forest regressor test y: 2014

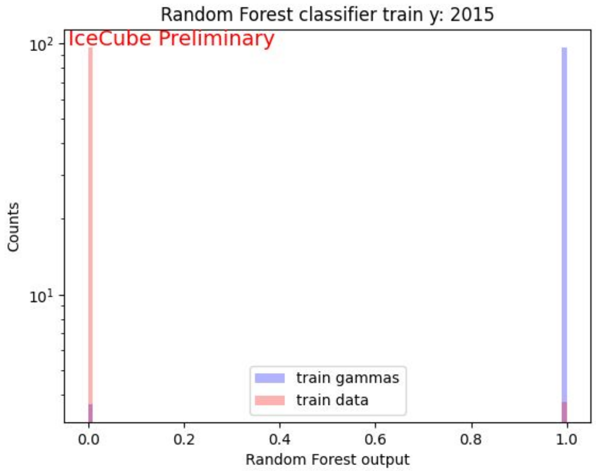
Passing fraction year 2014

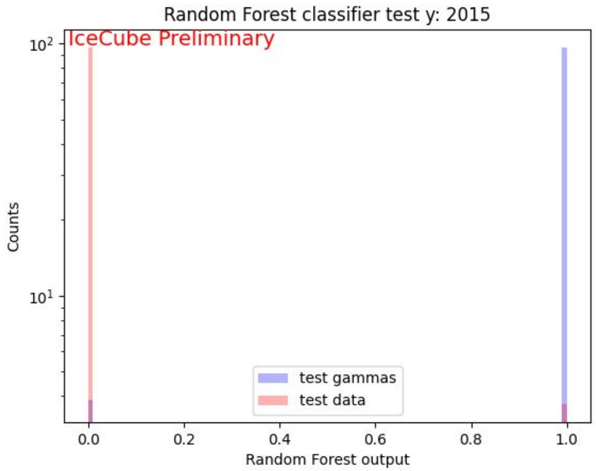
- event that passed the prediction cut / Total event
- plotted as function of energy

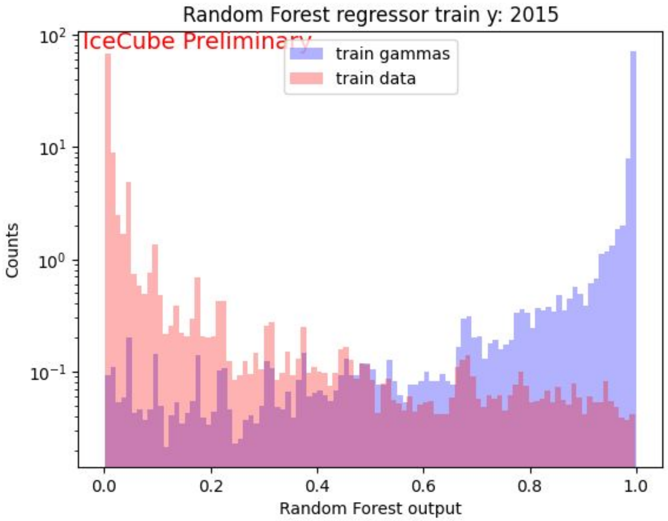


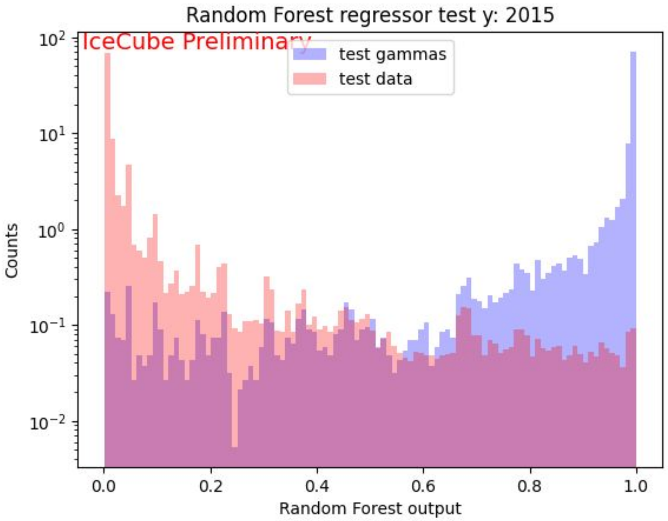


Training & Testing year 2015


Random Forest classifier train y: 2015

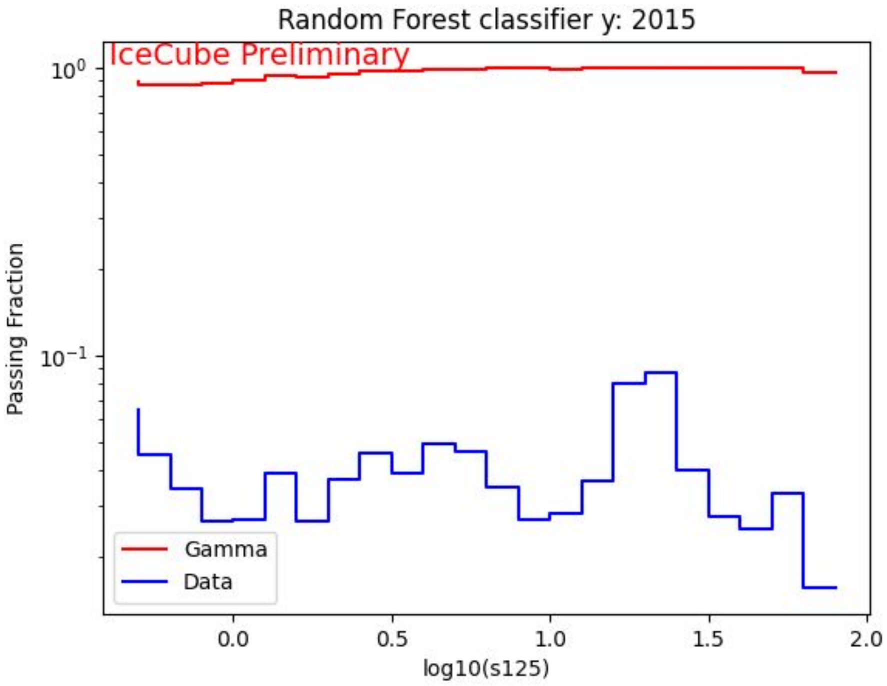

Random Forest classifier test y: 2015

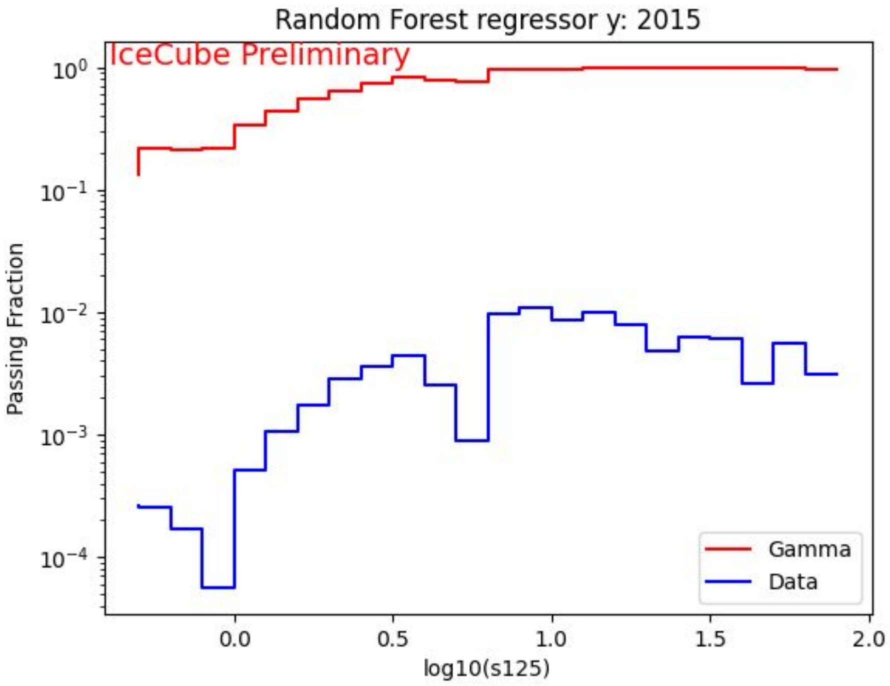

Random Forest regressor train y: 2015


Random Forest regressor test y: 2015

Passing fraction year 2015

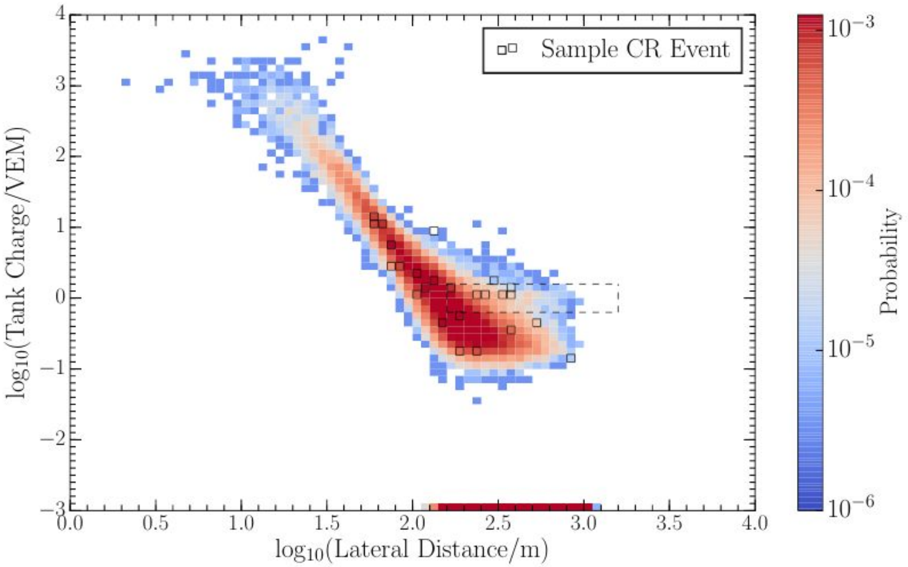
- event that passed the prediction cut / Total event
- plotted as function of energy



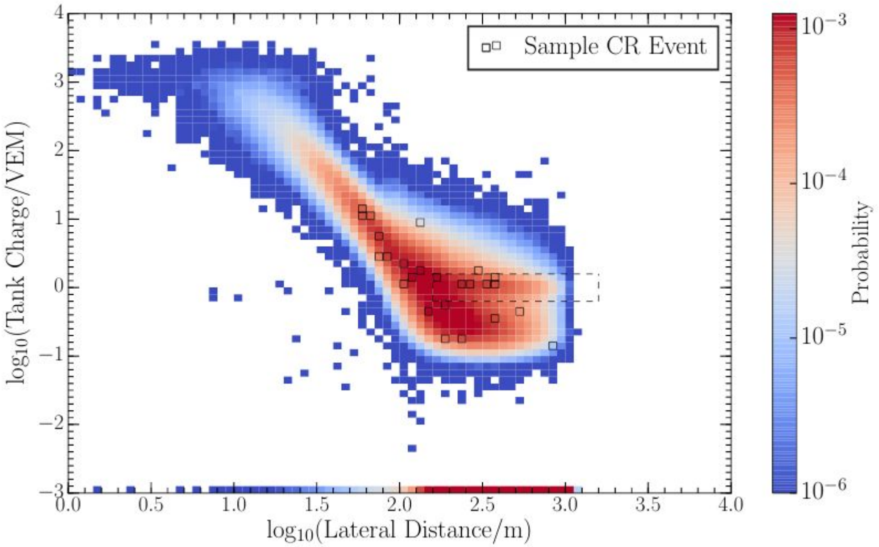


Log likelihood ratio

- example of a photon and CR background probability map



(a) Lateral PDF for gamma rays.



(b) Lateral PDF for cosmic rays.

Source: SEARCH FOR PEV GAMMA RAYS WITH THE ICECUBE OBSERVATORY, Zachary Dean Griffith