Gamma/Hadron Separation using Machine Learning Methods with the IceAct Telescopes

Logan Molchany

UD Conference on Machine Learning

January 29, 2025

Outline:

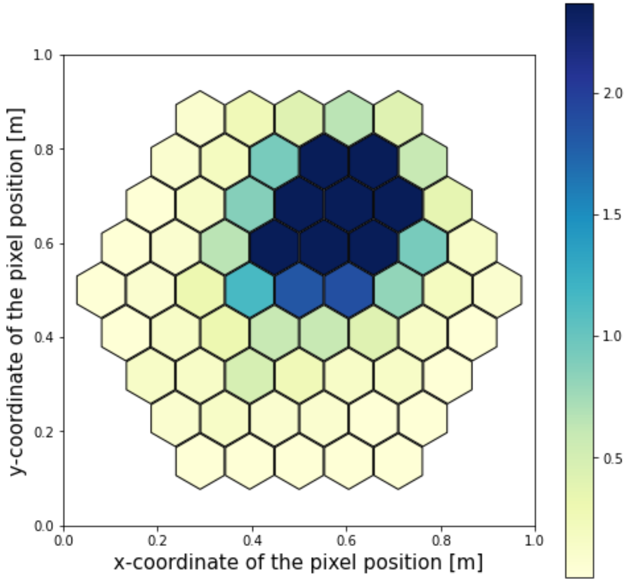
- Introduction to IceAct
- Simulations and Image Cleaning
- Hybrid Hillas Method Info
- Machine Learning Methods

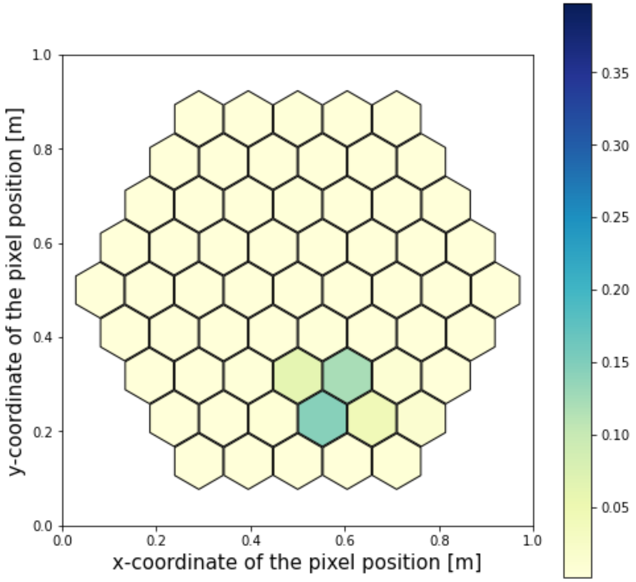
and Analysis

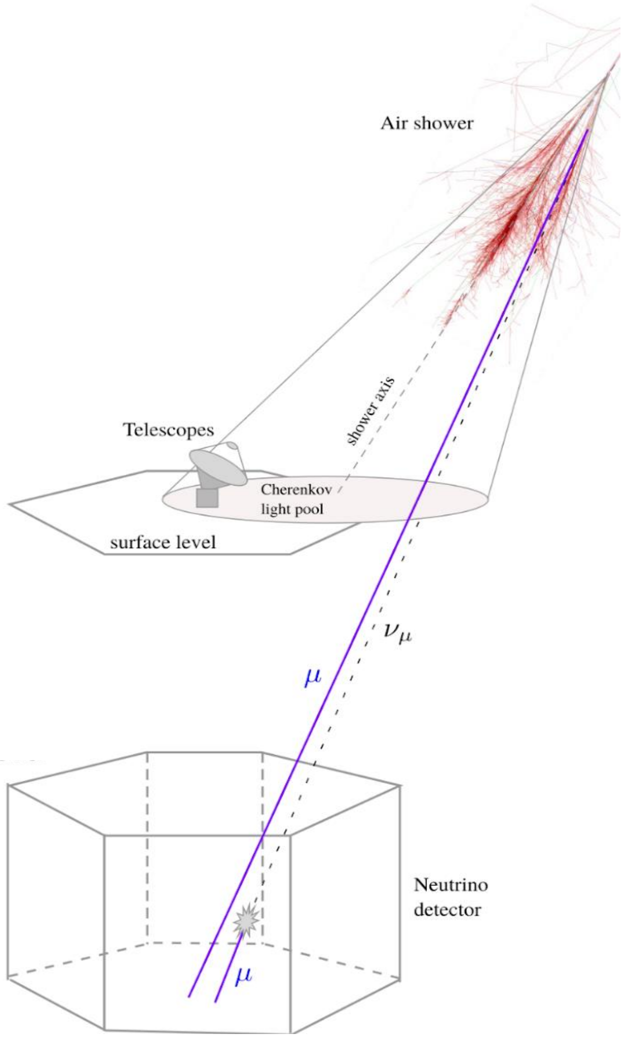
- Conclusion and Outlook

What IceAct Measures:

- Energy Range: 10 TeV – 100 PeV
  - Overlap in direct detection and radio detection
- Duty Cycle: 20% - 30%
- IACT Measurements:
  - Imaged Cherenkov light from EM component of atmospheric CR showers
  - Energy and shower maximum visible
  - Complimentary to:
    - Surface array ground particle footprint
    - In-Ice muonic component

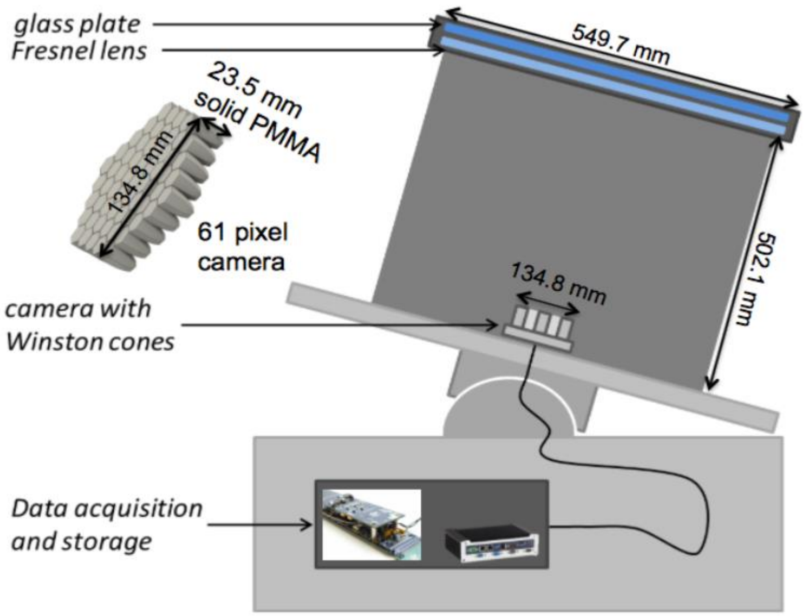




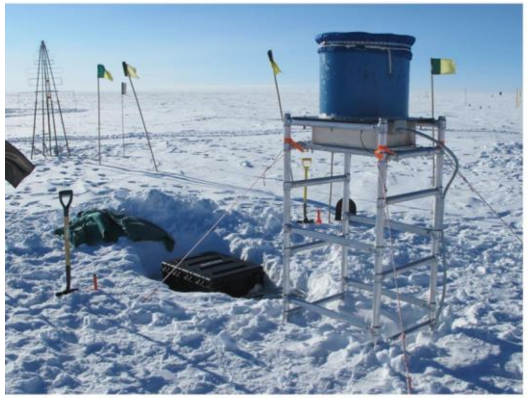


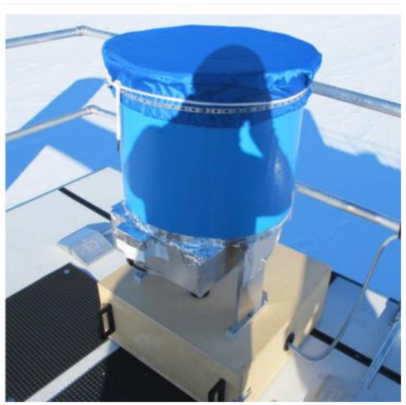
The IceAct Telescope:

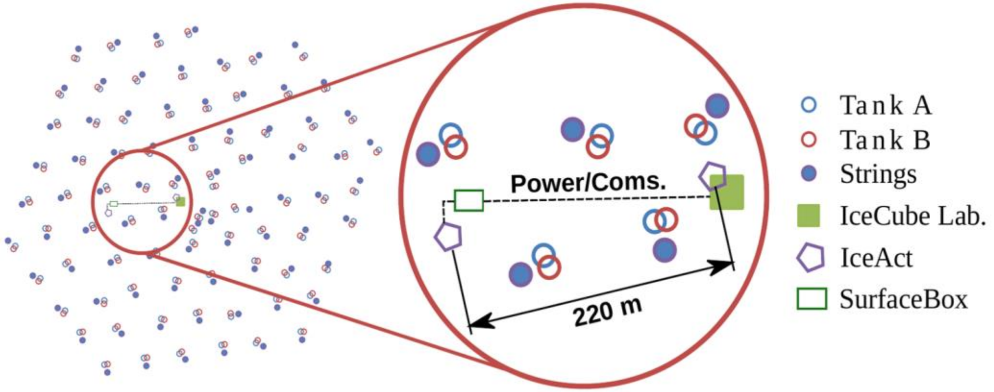
- Lightweight and compact
  - Weight: ~ 25kg, Diameter: ~ .5m
- 61 Silicon Photomultiplier (SiPM) pixel cameras
- Pixel Field of View (FOV): ~ 1.5° / pixel
- Singular telescope FOV: ~ 12°



- Three telescopes deployed:
  - All running at same rate with stable operations







Simulation Overview:

- Dataset (Upgraded V3)
  - CORSIKA (FLUKA + SIBYLL2.3c) Simulations:
  - Proton, Helium, Nitrogen, Aluminum, Iron, Neon, Photons from 3 TeV to 6 PeV (E^-1 spectrum) with atmospheric interactions

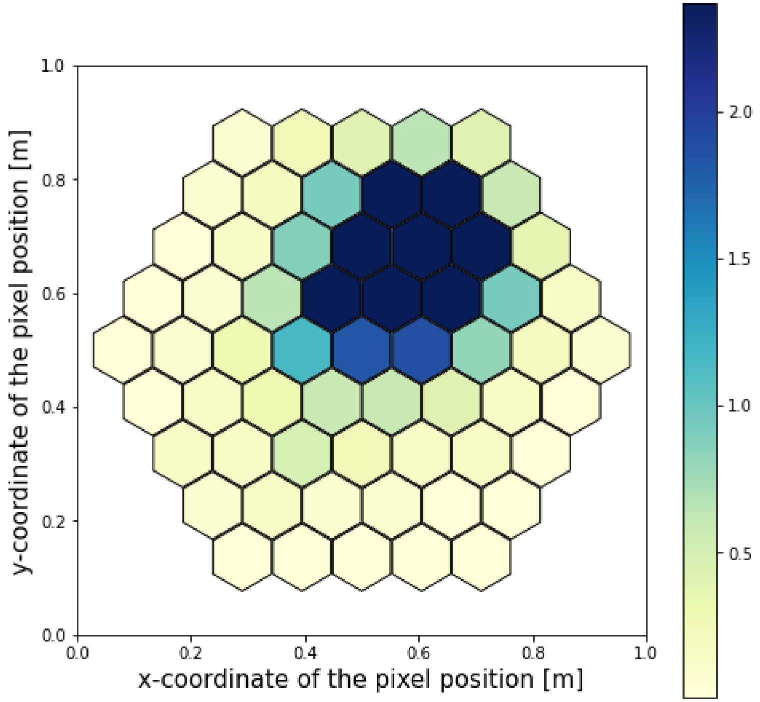
  - Detector simulations are done using IceAct optics, SiPM’s, and a full electronic simulation. The IceCube/IceTop detector response is simulated using the standard IceCube simulation framework.

Datasets:

- Pulse Series:
  - No filter max 40mV (in volts)
  - Calibrated pixel charges (converts volts to PE using 0.00144 PE /1V)
  - Images cleaned (using cta's two-tailed threshold cut)
  - Atmospheric absorption simulated

- Two tailed cut:
  - Picture threshold: keeps all pixels above this value (9.5 PE)
  - Boundary Threshold: keeps all pixels above the boundary threshold (4.75 PE) and that have a neighbor in the image, with a minimum of 2 neighbors

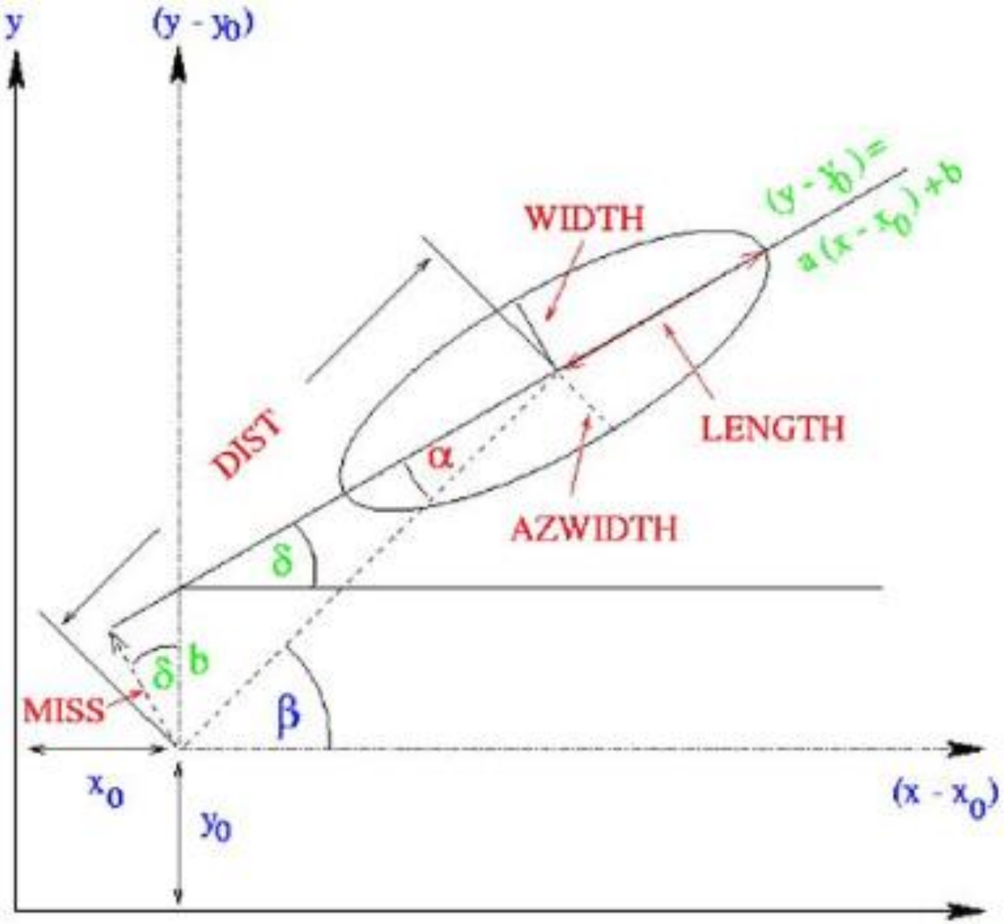
Ex. of good image to pass cleaning



Hillas Analysis Parameters:

$$A = \begin{bmatrix} \mathrm{var}(\Delta x) & \mathrm{cov}(\Delta x, \Delta y) \\ \mathrm{cov}(\Delta x, \Delta y) & \mathrm{var}(\Delta y) \end{bmatrix}$$

- First used by M. Hillas in 1985:
  - First 9 sigma detection with an IACT of the Crab Nebula
- Parametrizes the image using width, length, and orientation on the camera pixels
- Works for coarse pixel sizes, but we use a filter length of greater than 3 pixels to prevent binning too finely
- Length and width are calculated geometrically or using the square eigenvalues of a covariance matrix (CTA method)

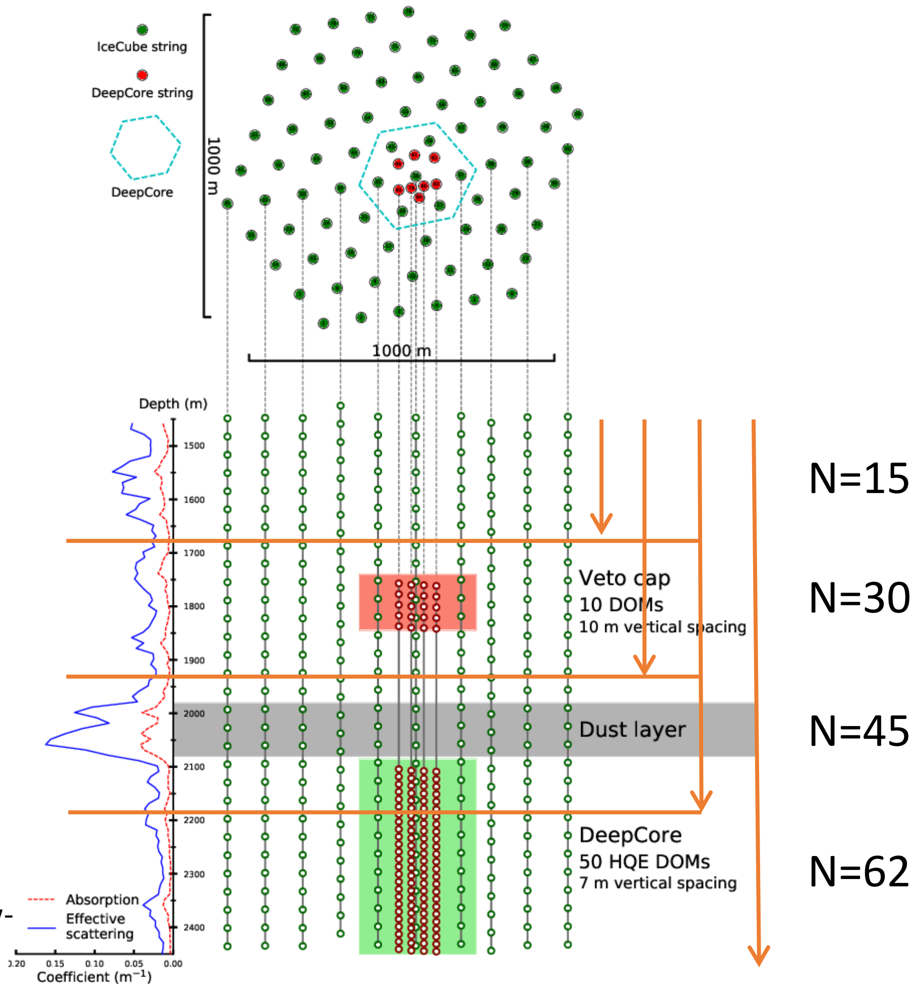


Hampton, Muriel. (2018). *Analysis of the MAGIC data.* MAGIC Collaboration. https://slideplayer.com/slide/13433695/

- Cos(zenith) and distance from shower core come MC simulations

N-Layer Charge Total Parameters:

- InIce Parameter that sums the DST Dom Charges in n-layers of DOMs
- Starting from the first 15 DOMs, then the first 30, 45 and the entire detector at 62 DOMs
- Needs more exploration into where the greatest mass separation is, and how many layers of DOMs that corresponds too
- For now, only the first 30, and 45 DOM charge totals are used



https://www.researchgate.net/publication/365147756_Graph_Neural_Networks_for_low-energy_event_classification_reconstruction_in_IceCube

Machine Learning Features and Models:

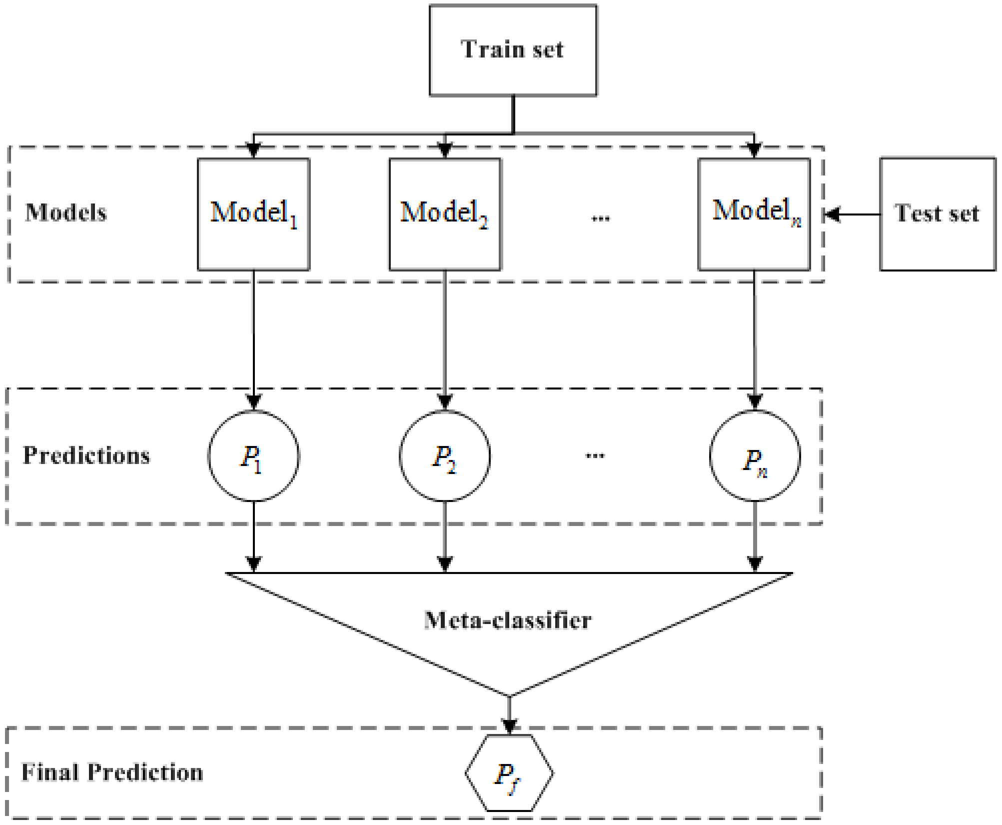
- "Stacking Method" – both classification and regression:
  - 33% test, 66% training - Base models
  - CORSIKA MC Sims (Fluka+Sibyll2.3c)
  - Reconstruction Output (E and type):
    - 0.0 = Photon, 1.0 = Hadrons
    - Nuclei consist of: Proton, Helium, Nitrogen, Aluminum, Iron
- Features:
  - Shower Size
  - Ellipse Width
  - Ellipse Length
  - Distance (from shower core)
  - Cos(zenith)
  - DOM Charge Total (In n-layers of DOMs)
  - Muon Proxy
  - Skewness
  - Kurtosis

Hillas Image Params

Primary Geometry

In-Ice Param

Statistical Params



https://www.researchgate.net/figure/The-architecture-of-the-stacking-ensemble-learning-In-the-base-classifiers-the-training_fig1_335156833

Machine Learning Methods: Stacking Method

- Meta Models:
  - Classifier: LogisticRegression()
  - Regressor: LinearRegression()
  - 10-Fold Validation Set

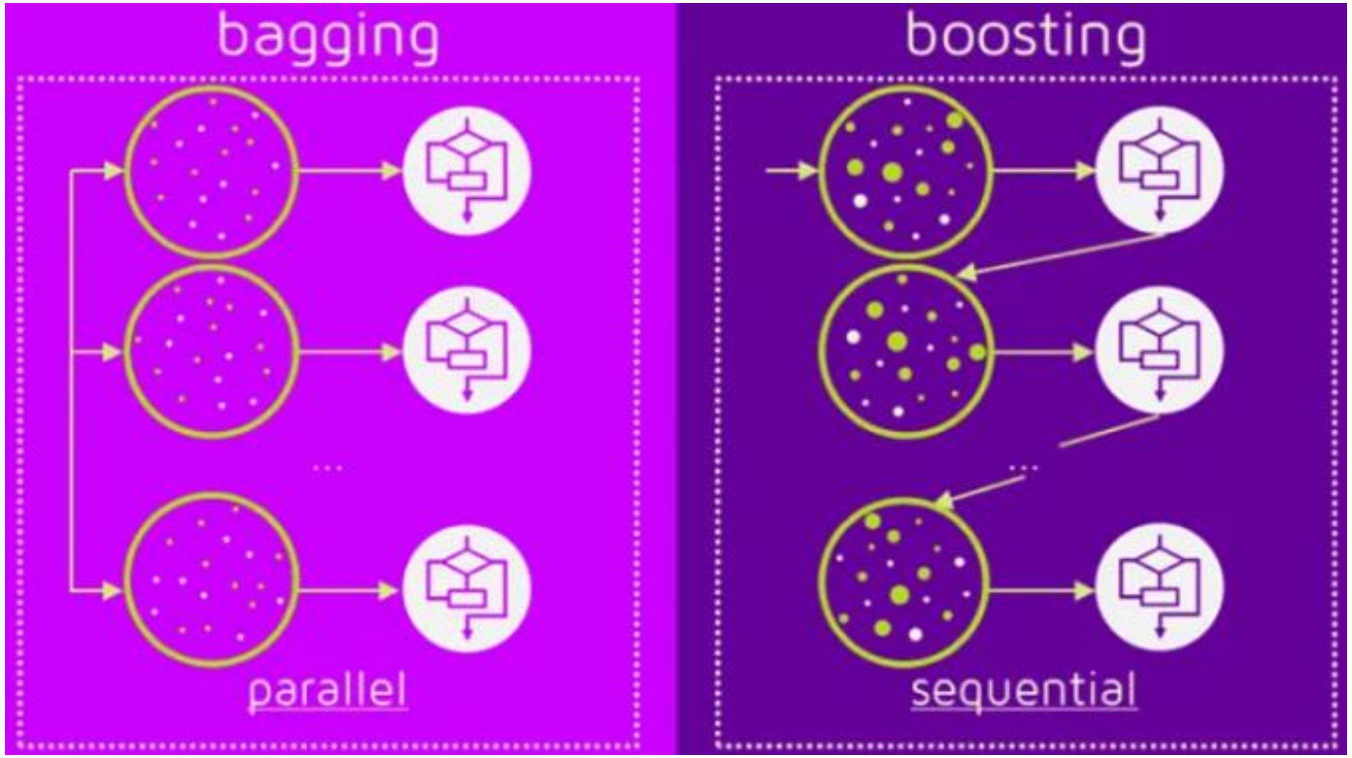
  - Want to test other meta models
    - Gradient Boosting, SVM, KNN

CLASSIFIERS	REGRESSORS
CatBoostClassifier()	CatBoostRegressor()
GradientBoostingClassifier()	GradientBoostingRegressor()
RandomForestClassifier()	RandomForestRegressor()

  - Dimension reduction needs tested (PCA / LDA)

Why Stacking Method?

- Uses the principle of ensemble learning
  - Multiple weaker learners can combine to make stronger models
  - Uses the strengths of the “base models” in different energy bins
- Allows us to use multiple diverse methods
  - Bagging, gradient boosting, voting/ tagging, etc
  - Can be used for both classification and regression

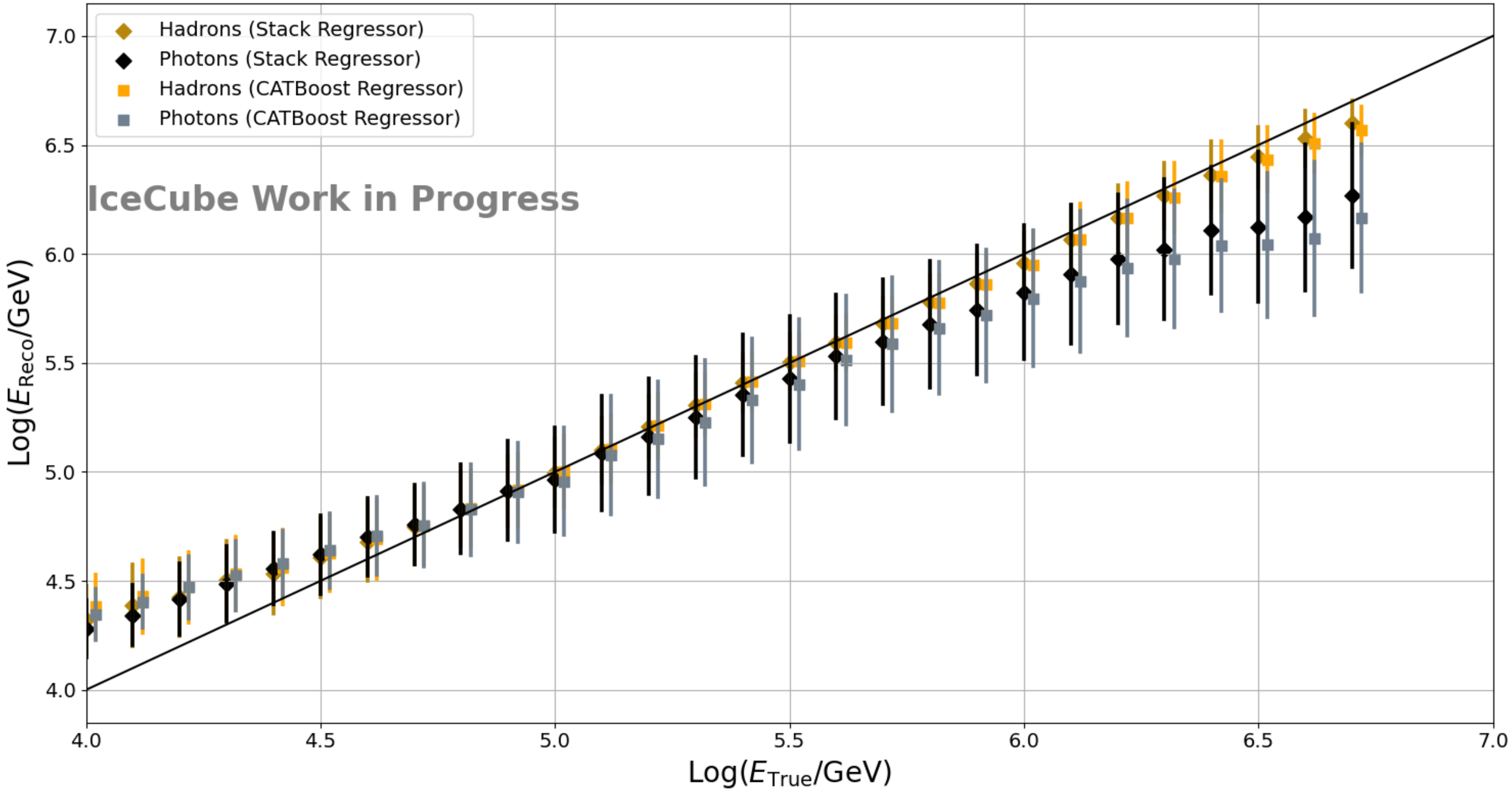


https://medium.com/@mygreatlearning/everything-you-need-to-know-about-ensemble-learning-eed79a704a45

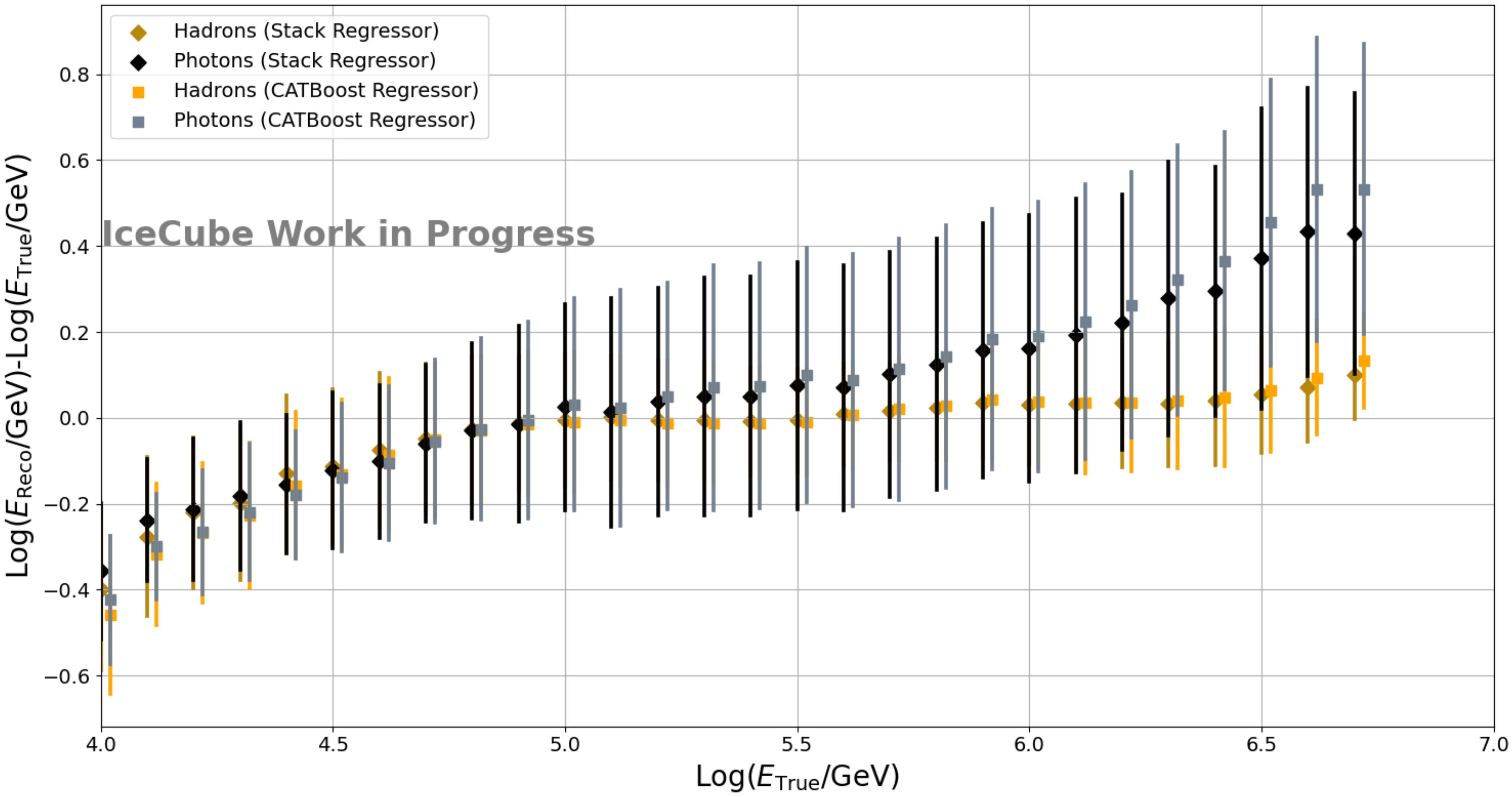
Regressor Analysis:

Energy Reconstruction:

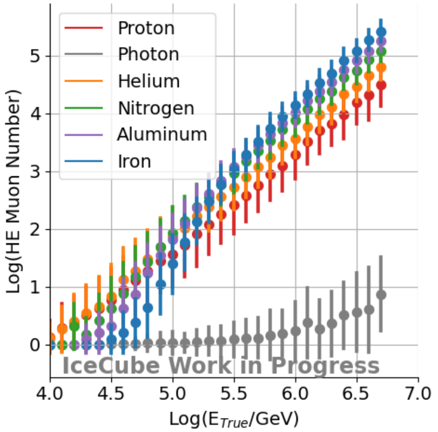
- Bias increases with energy
  - Move toward separating the photon and hadron energy reconstructions
  - Stacking Method increases performance over single base model

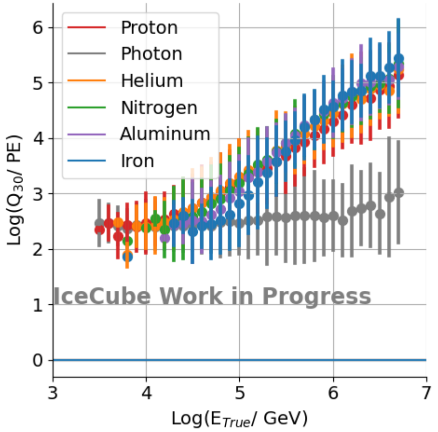


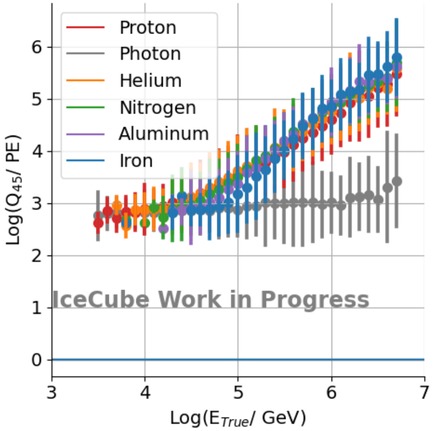
Energy Reconstruction Residuals:

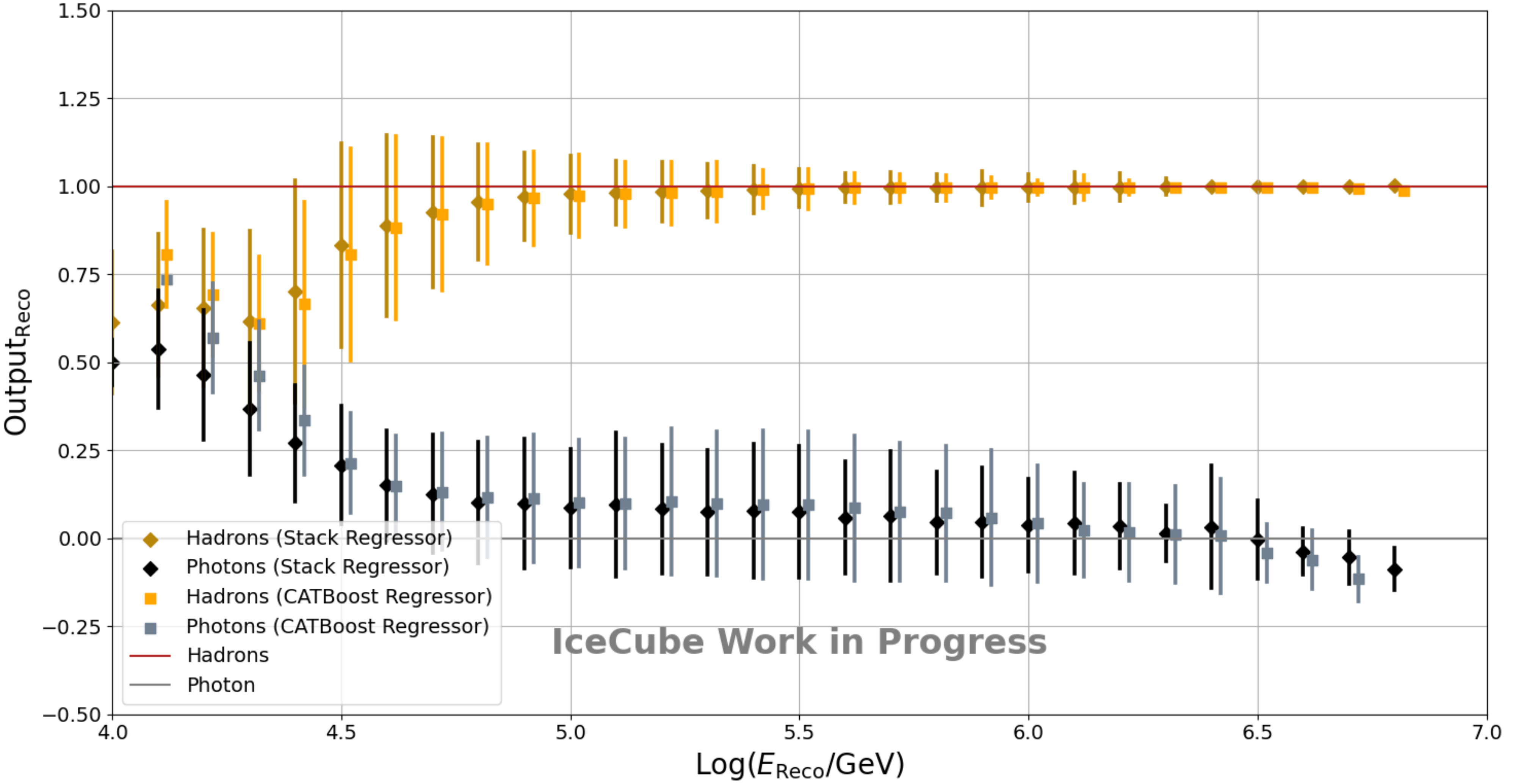


Reconstructed Output:

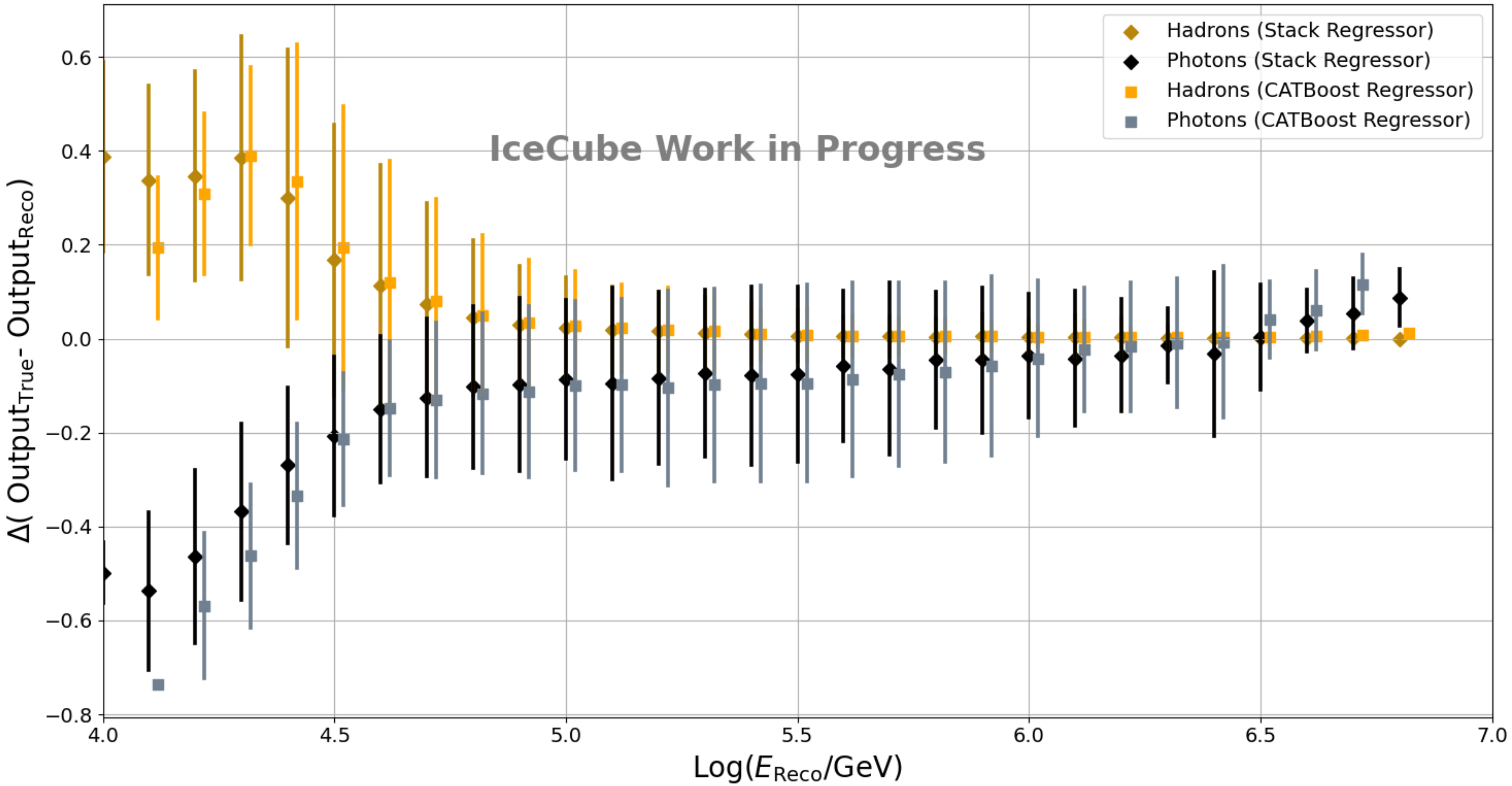








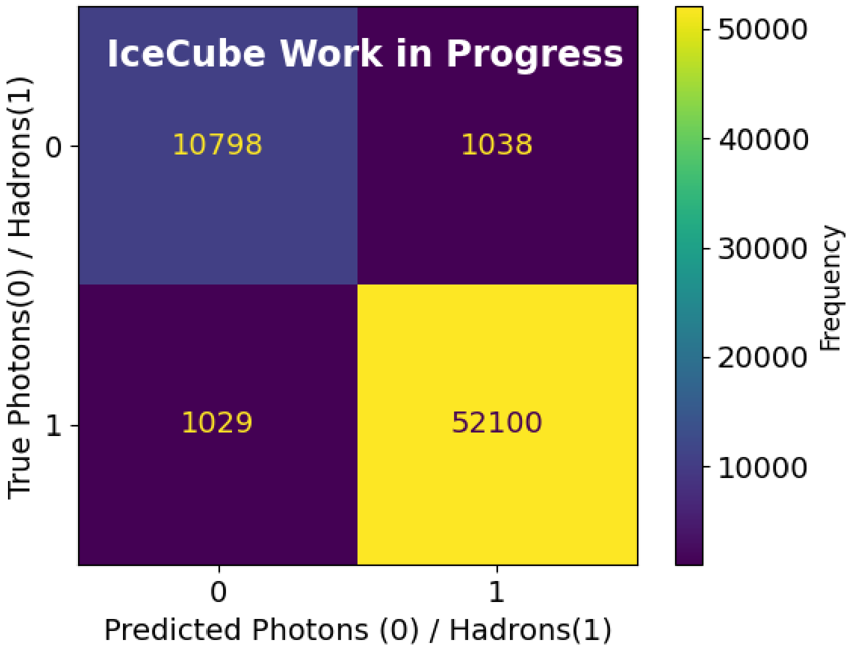
Reconstructed Output Residuals:



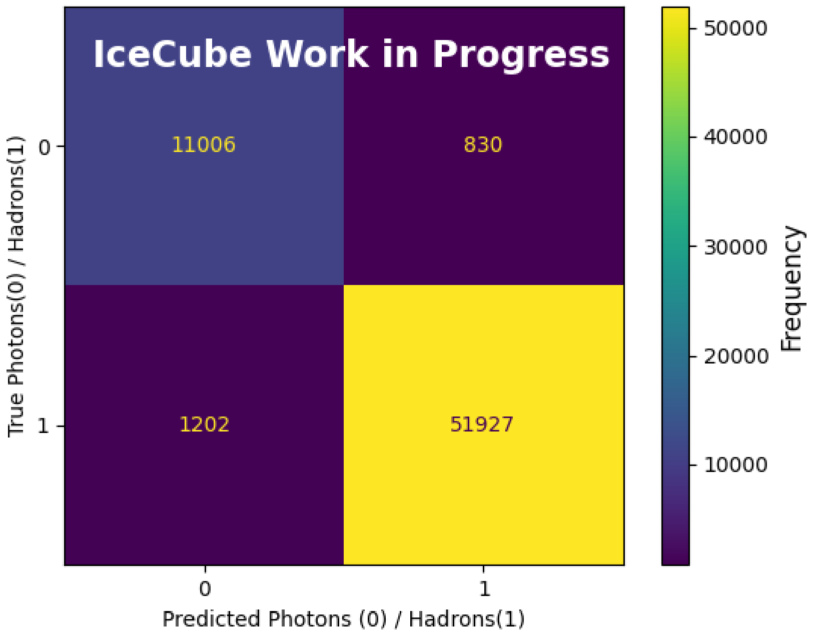
Classifier Analysis:

Confusion Matrices:

Stack

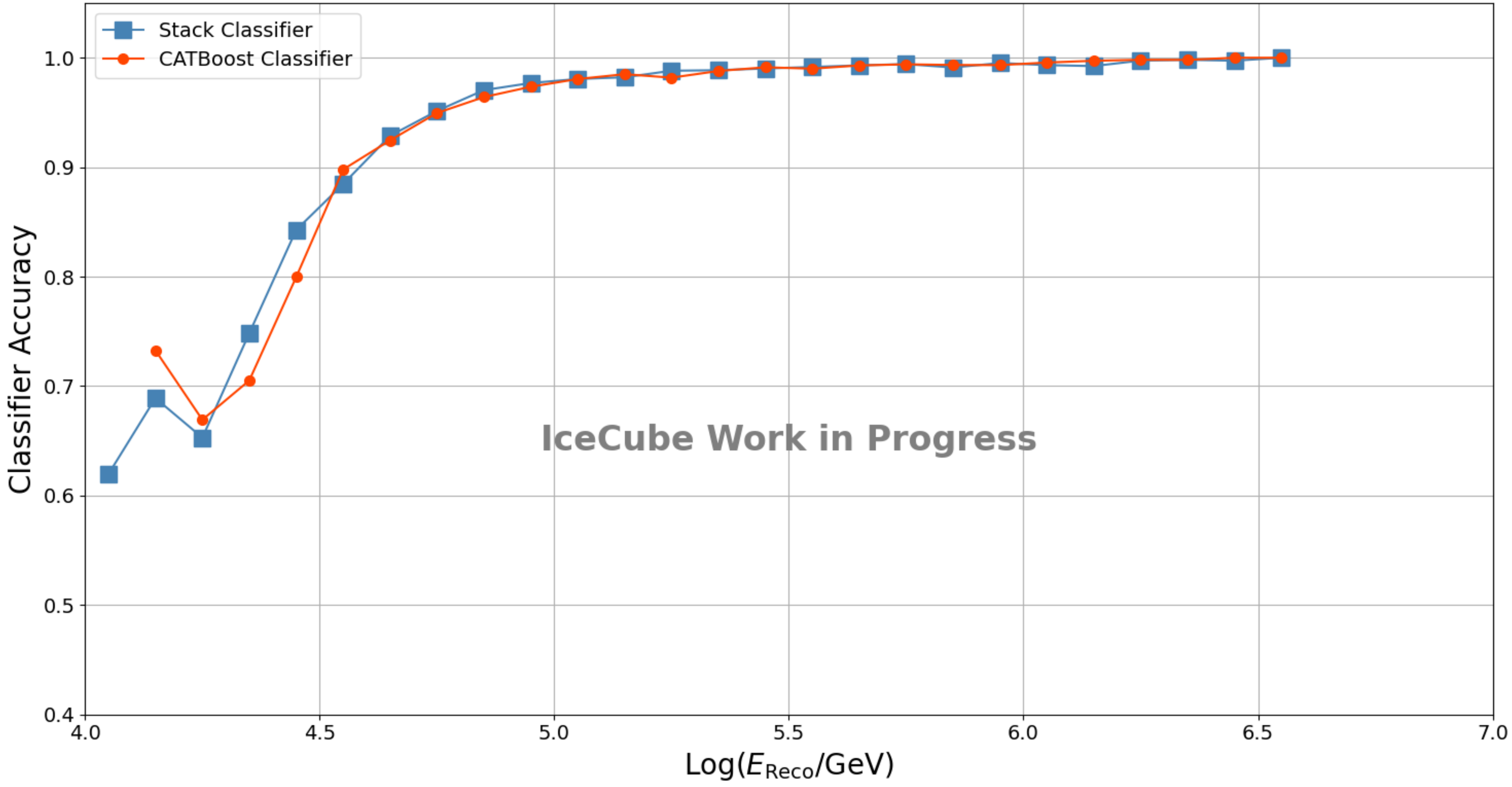


CATBoost



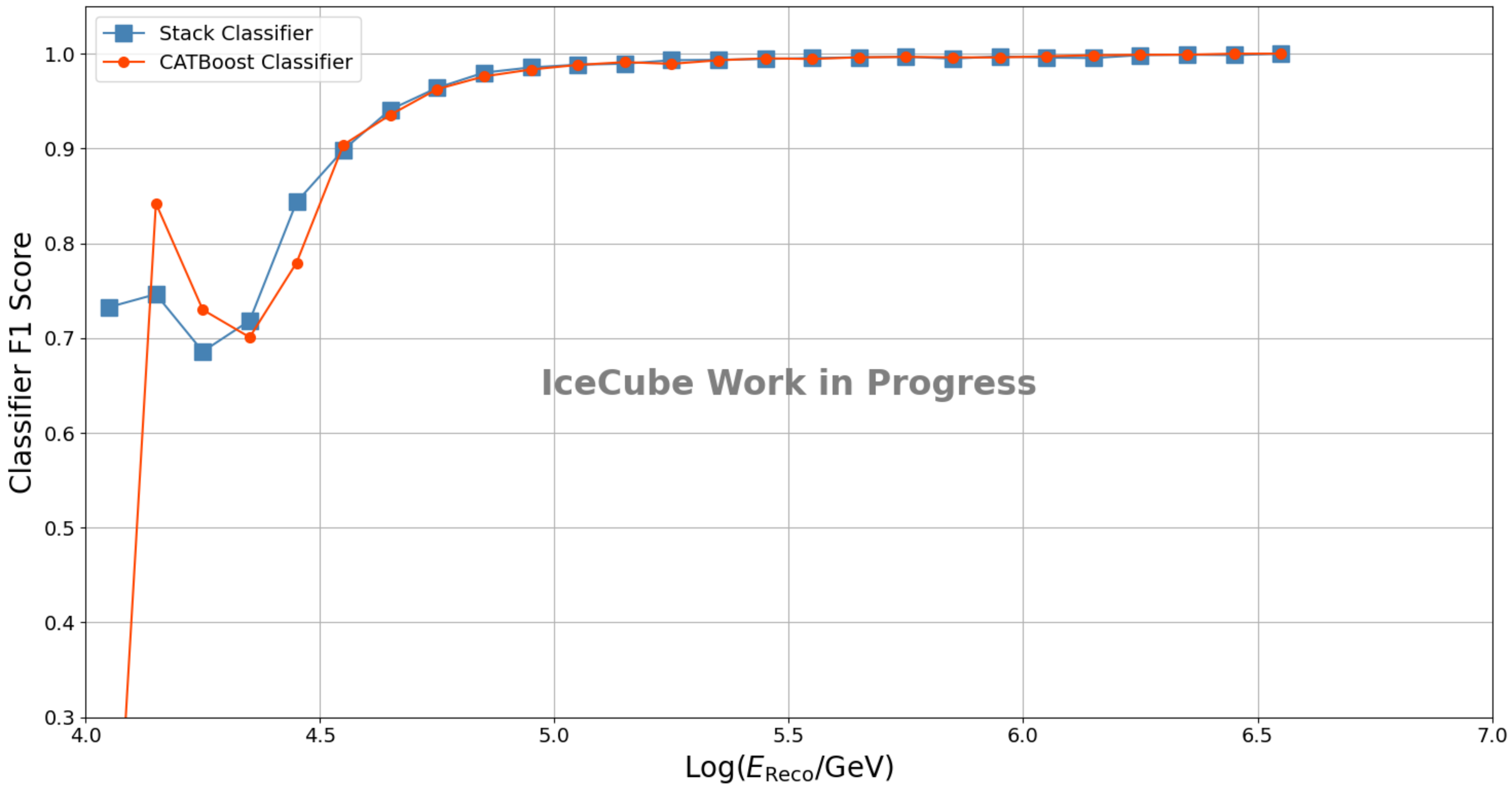
Classifier Accuracy:

- Accuracy score: $\frac{TP+FN}{TP+TN+FP+FN}$



Classifier F1:

- F1 score: $2 * \frac{Precision * Recall}{Precision + Recall}$



Conclusion and Outlook:

- Immediate work:
- Study containment cuts and new low level observables (InIce or IceTop)
- Study other Gamma/Hadron separation methods
- Use GCNN for energy and primary type reconstructions
- Retune all models and meta model (compare to various meta models)

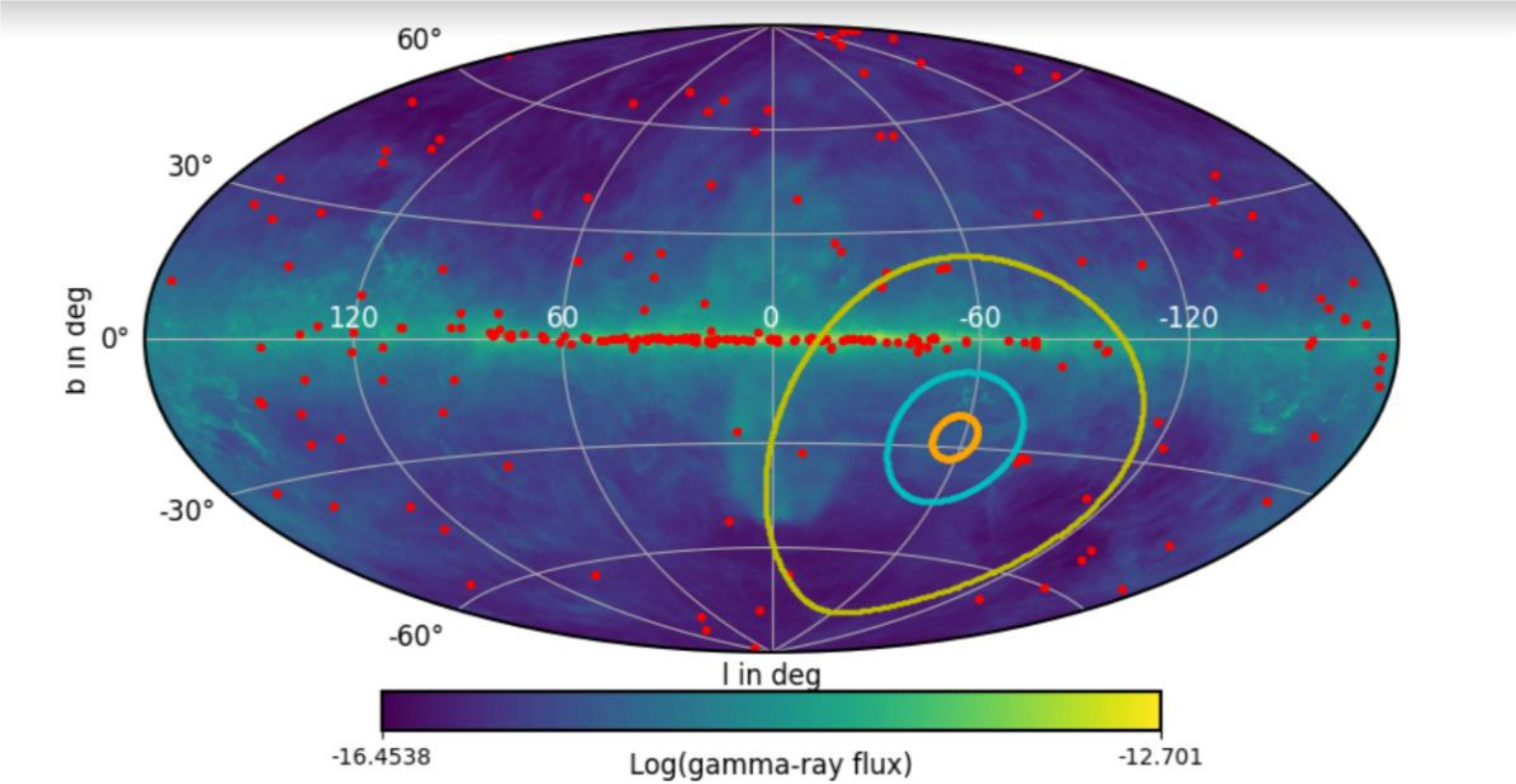
- Future work:
- Determine IceAct sensitivity / background rates / effective areas
- Determine diffuse flux of photons if possible

- Further Future Work:
- Use full IceAct 7 telescope array in the future
- Use real data for analysis

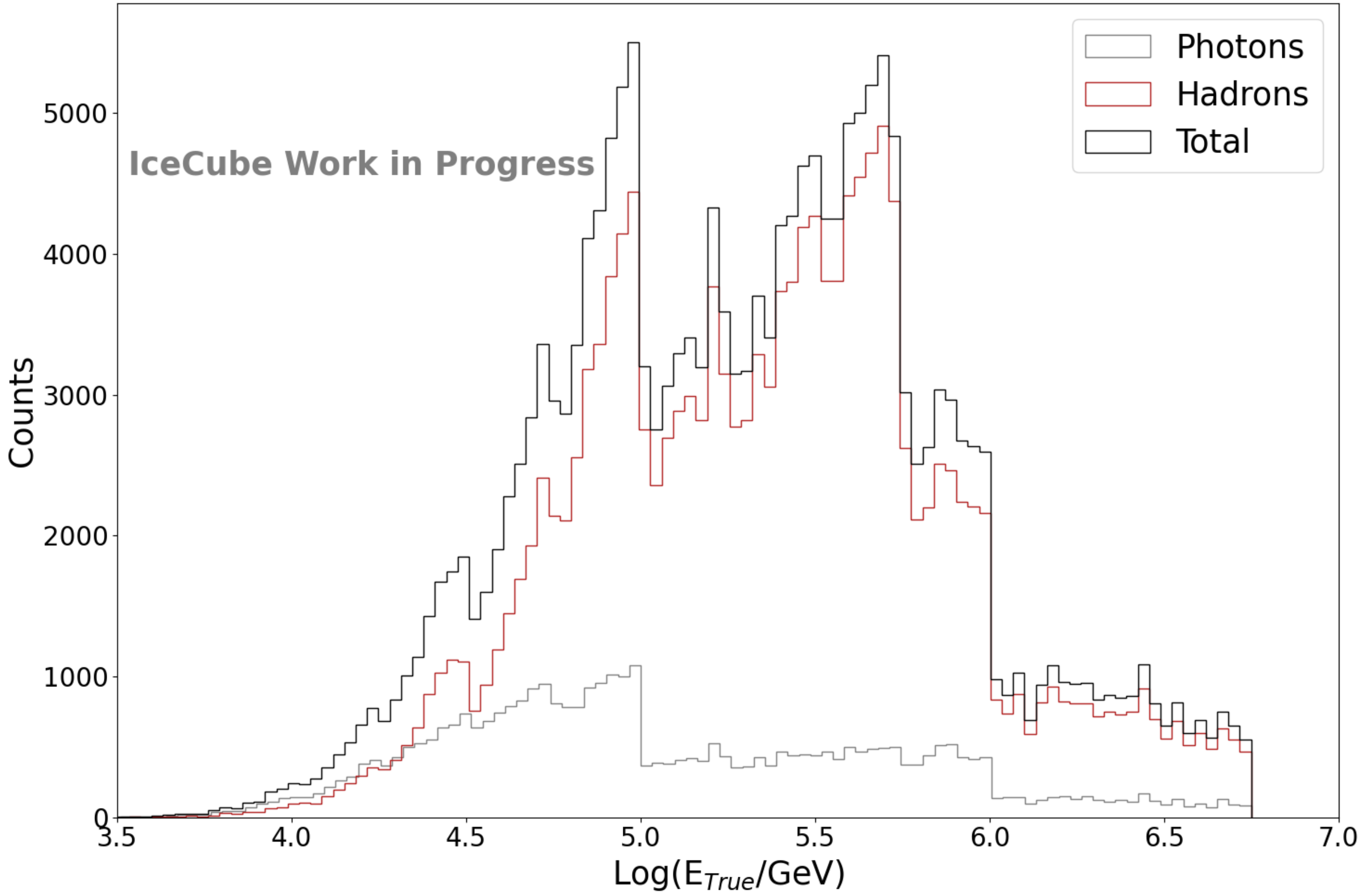
Backup

All-Sky Background Coverage:

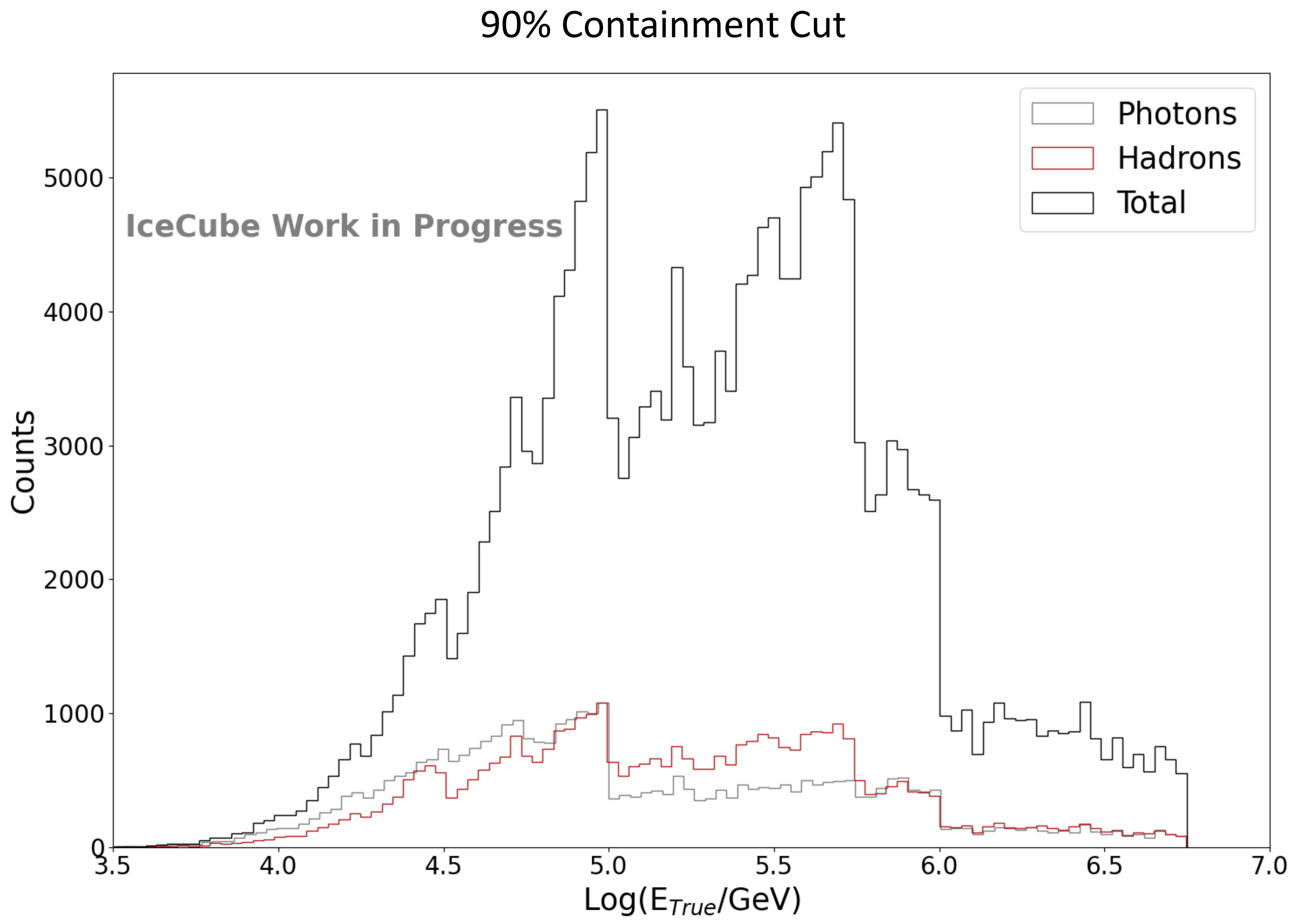
- Red dots are TeV gamma-ray sources
- Field of View:
    - Orange: current IceAct and IceCube
        - Opening Angle: 12°
    - Cyan: IceAct 7 Telescope station and IceCube
        - Opening Angle: 36°
    - Yellow: IceCube and IceTop
        - Opening Angle: 50°



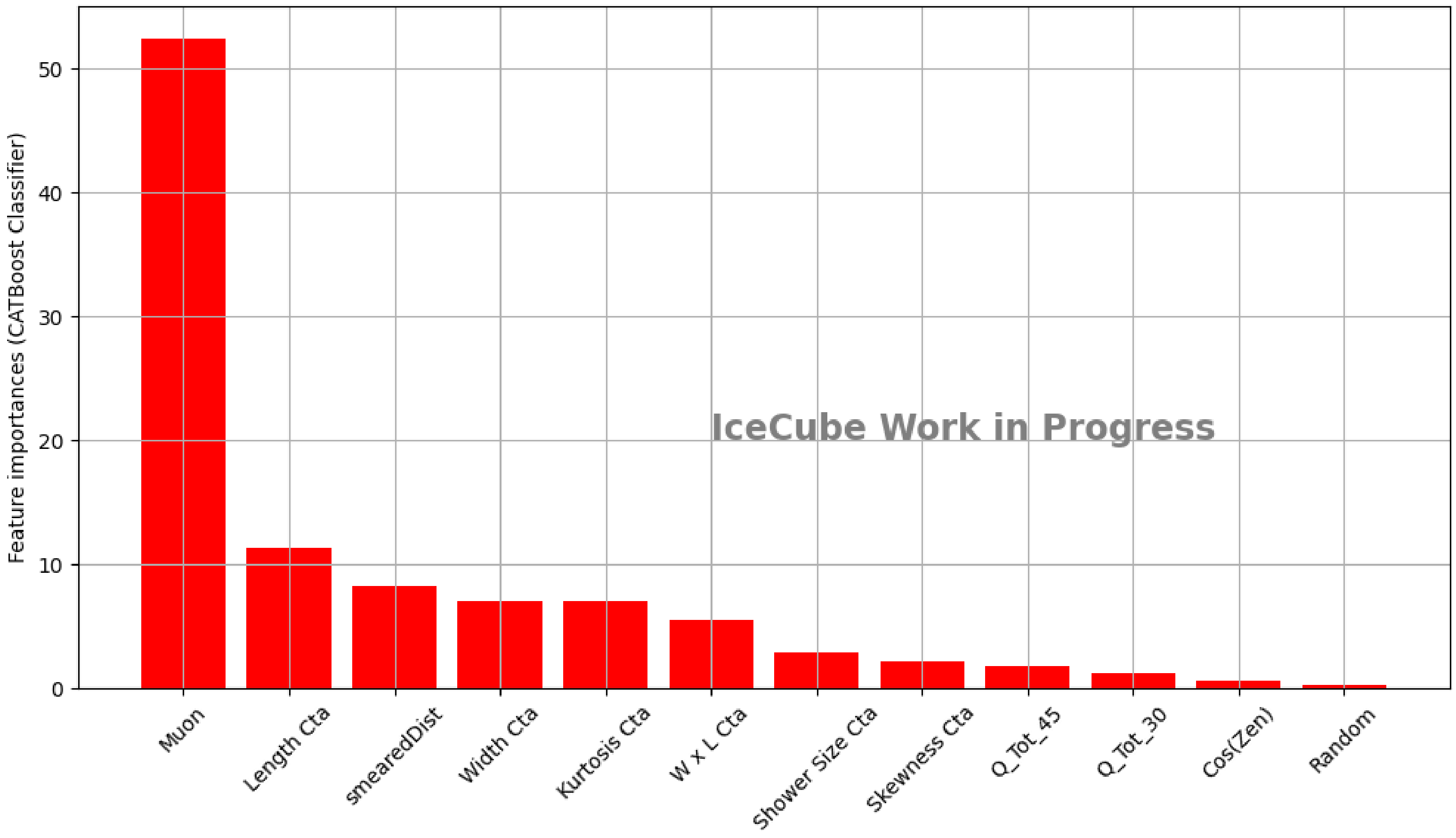
Data Distribution (No Cuts):



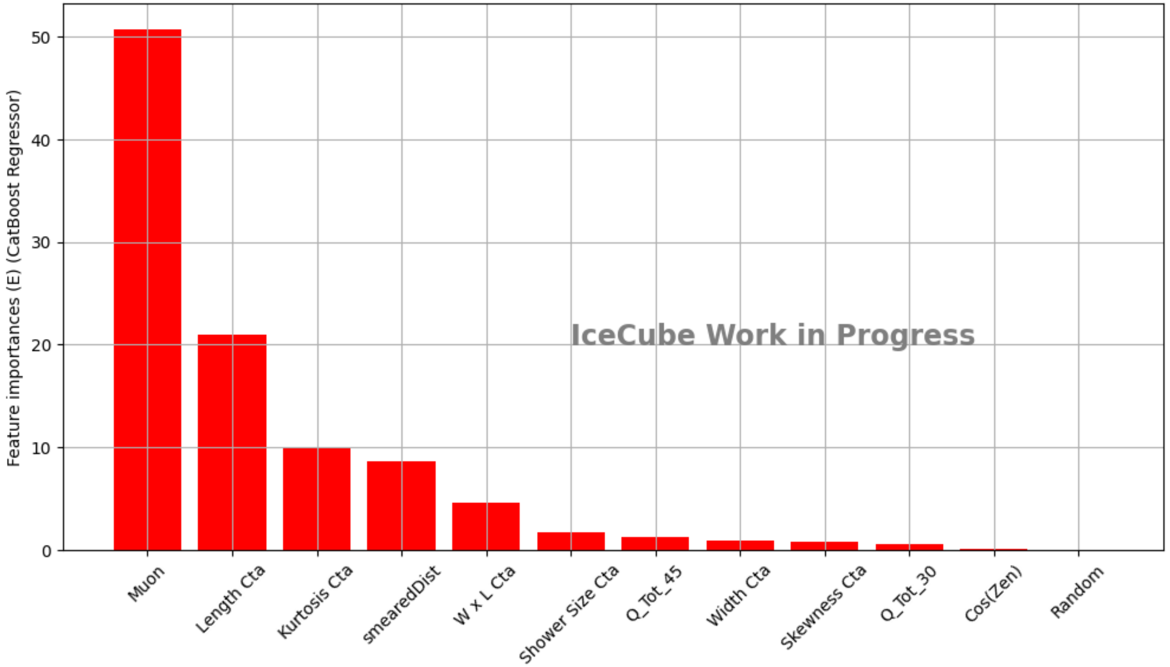
Data Distribution (Containment Cut):

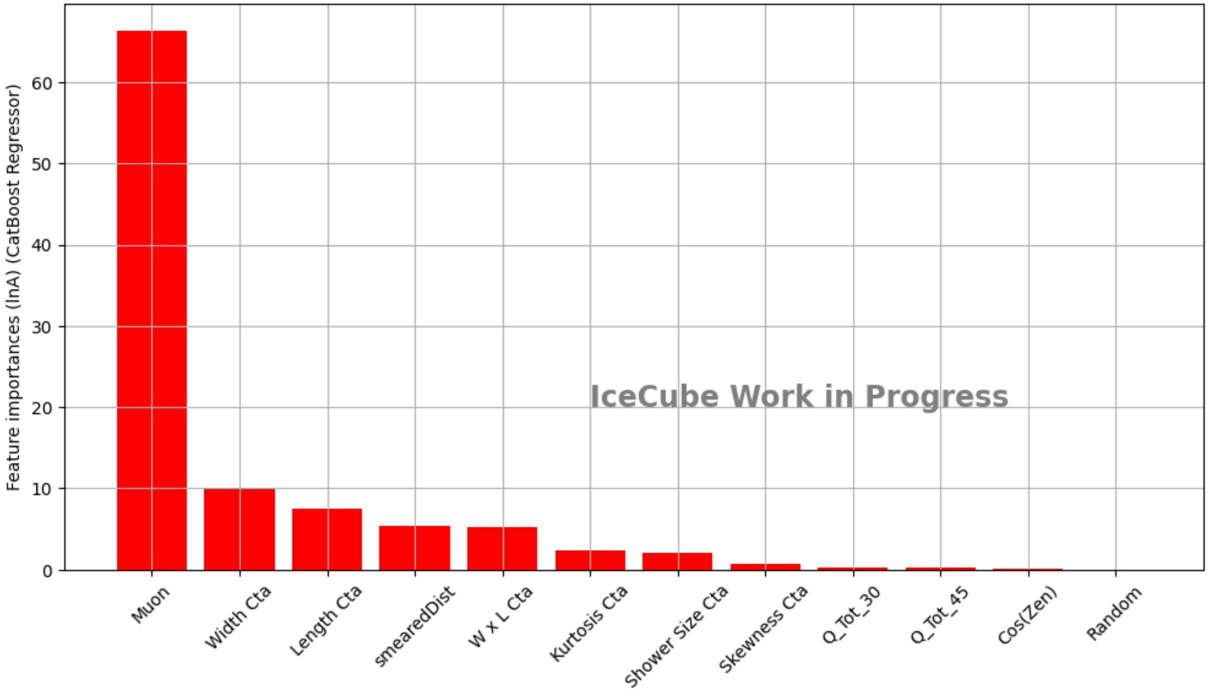


Classifier Feature Importances:



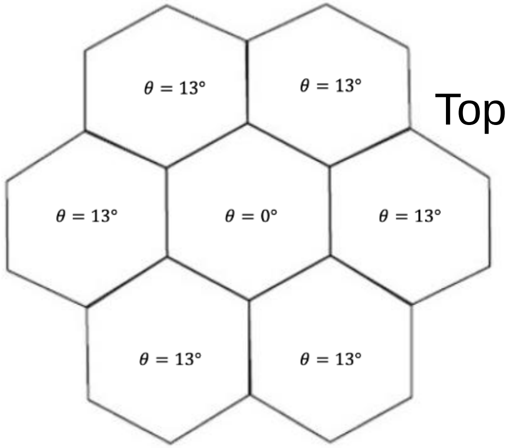
Regressor Feature Importances:





IceAct Station:

- Station = 7 telescopes in fly's eye configuration + surface box
- Central telescope points to zenith
  - Neighboring 6 point 13° off zenith
- Full station FOV: ~36°
- Gen2 proposal:
  - 4 stations
  - Moveable and can be expanded



Top view

Hillas Methods Comparison:

- Both use same shower size, position deltas, and position COG

$$s = \sum_{i=1}^{61} q_{\rho_{ix,i}} \qquad x_{COG} = \sum_{i=1}^{61} \frac{q_{\rho_i x,i^*} x_{p_i x,i}}{s} \qquad \Delta x = x_{P_i x,i} - x_{cOG}$$

- Our method:

$$M_{xx} = \sum_{i=1}^{61} q_{p_i x,i^*} \Delta x_i^2 \qquad M_{yy} = \sum_{i=1}^{61} q_{p_i x,i^*} \Delta y_i^2$$

$$k = \sqrt{\left(M_{yy} - M_{xx}\right)^2 + (2Mxy)^2 + \left(M_{yy} - M_{xx}\right)}$$

$$\Delta \tan = \frac{k}{2Mxy}$$

$$w = \sqrt{\frac{\left(M_{xx} * \Delta \tan^2 - k + M_{yy}\right)}{(\Delta \tan^2 + 1)s}} \qquad l = \sqrt{\frac{\left(M_{yy} * \Delta \tan^2 + k + M_{xx}\right)}{(\Delta \tan^2 + 1)s}}$$

- CTA method:
- l and w are squared eigenvalues of a covariance matrix

$$A = \begin{bmatrix} \operatorname{var}(\Delta x) & \operatorname{cov}(\Delta x, \Delta y) \\ \operatorname{cov}(\Delta x, \Delta y) & \operatorname{var}(\Delta y) \end{bmatrix}$$

Optimization Algorithms:

- Grid Search:
  - Uninformed search
  - Tests every unique combination in parameter space
- Randomized Search:
  - Uninformed search
  - Randomly picks parameters per iteration
- Bayesian Optimization:
  - Informed search
  - Tunes parameter space by previous iterations scores
- Successive Halving:
  - Sample parameter space according to algorithm
  - Evaluate the performance
  - Throw out bottom half of worst scoring parameters
  - Repeat until most optimal parameters remain

Pros and Cons of Hyperparamter Tuning Methods:

- Grid Search:
- Pros:
  - Exhaustive search
- Cons:
  - Computation time exponentially increases with parameter space size
  - Can be prone to overfitting
- Randomized Search:
- Pros:
  - Reduced computation time
  - Throws out parameters causing overfitting quicker than grid search
  - Highly scalable
- Cons:
  - Takes more iterations to reach the same performance as grid search
  - High variance between runs

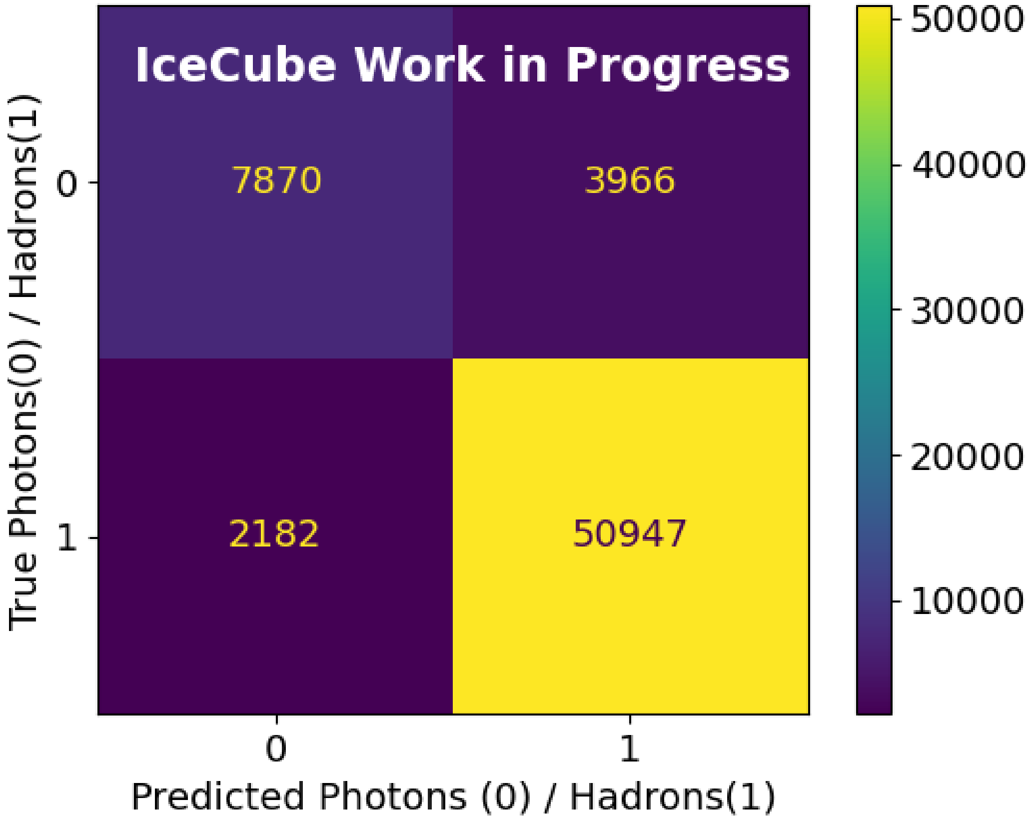
Pros and Cons Continued:

- Bayesian Optimization:
- Pros:
  - Better at finding optimal hyperparameters for larger and more complex models
  - Takes 20x less computing time than other tuning algorithms
- Cons:
  - More complex to implement
  - Requires more computational resources
  - More sensitive to parameters

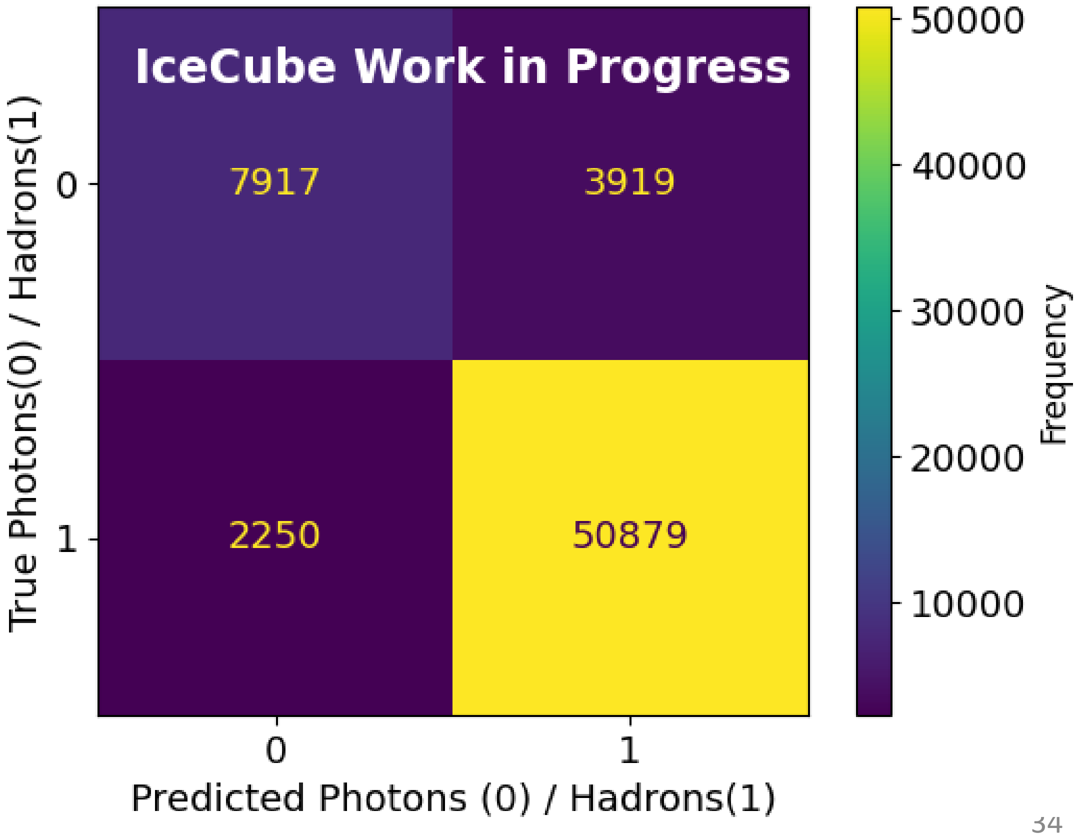
No Muon Proxy Outlook Preliminary Plots

Stacking Classifier CM's:

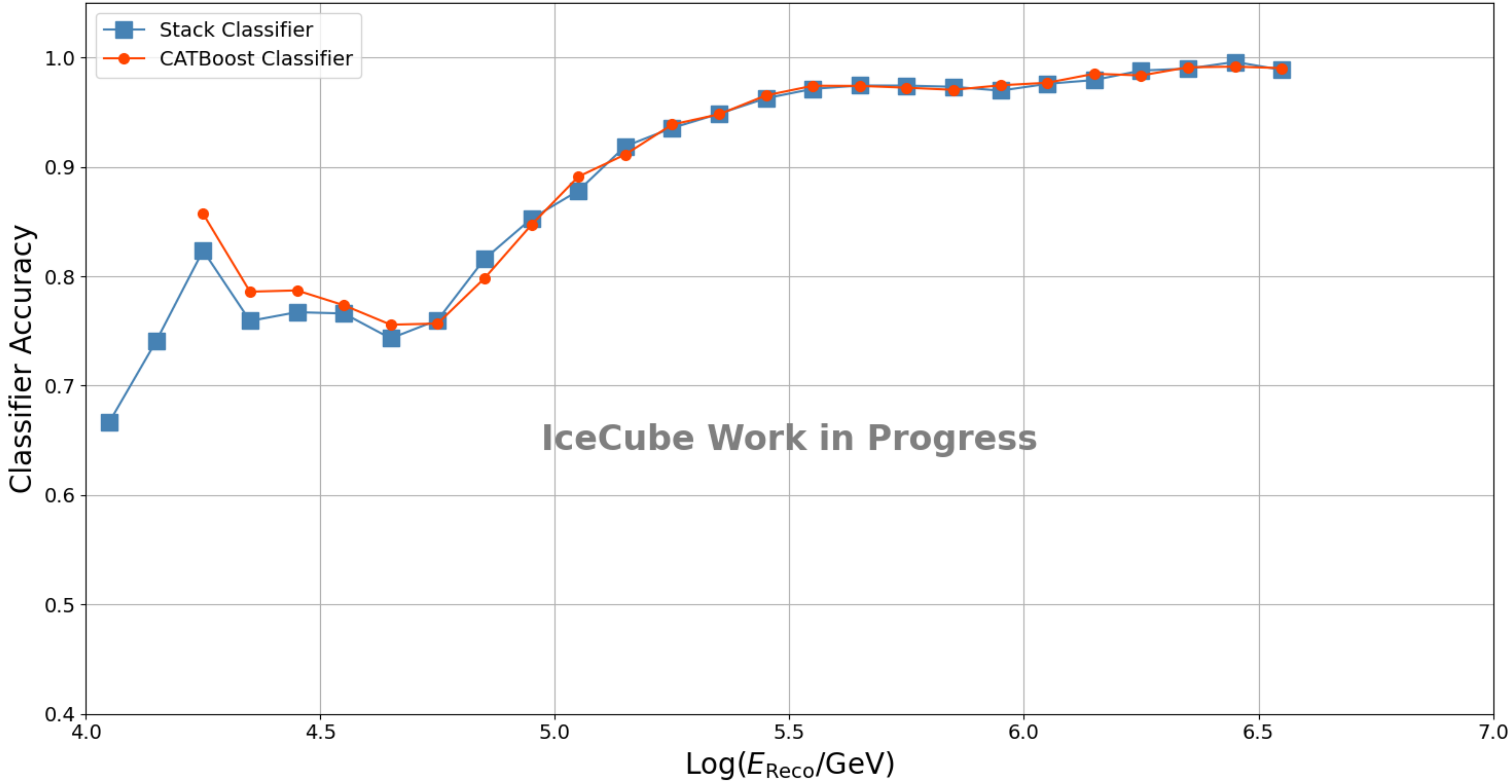
Stack



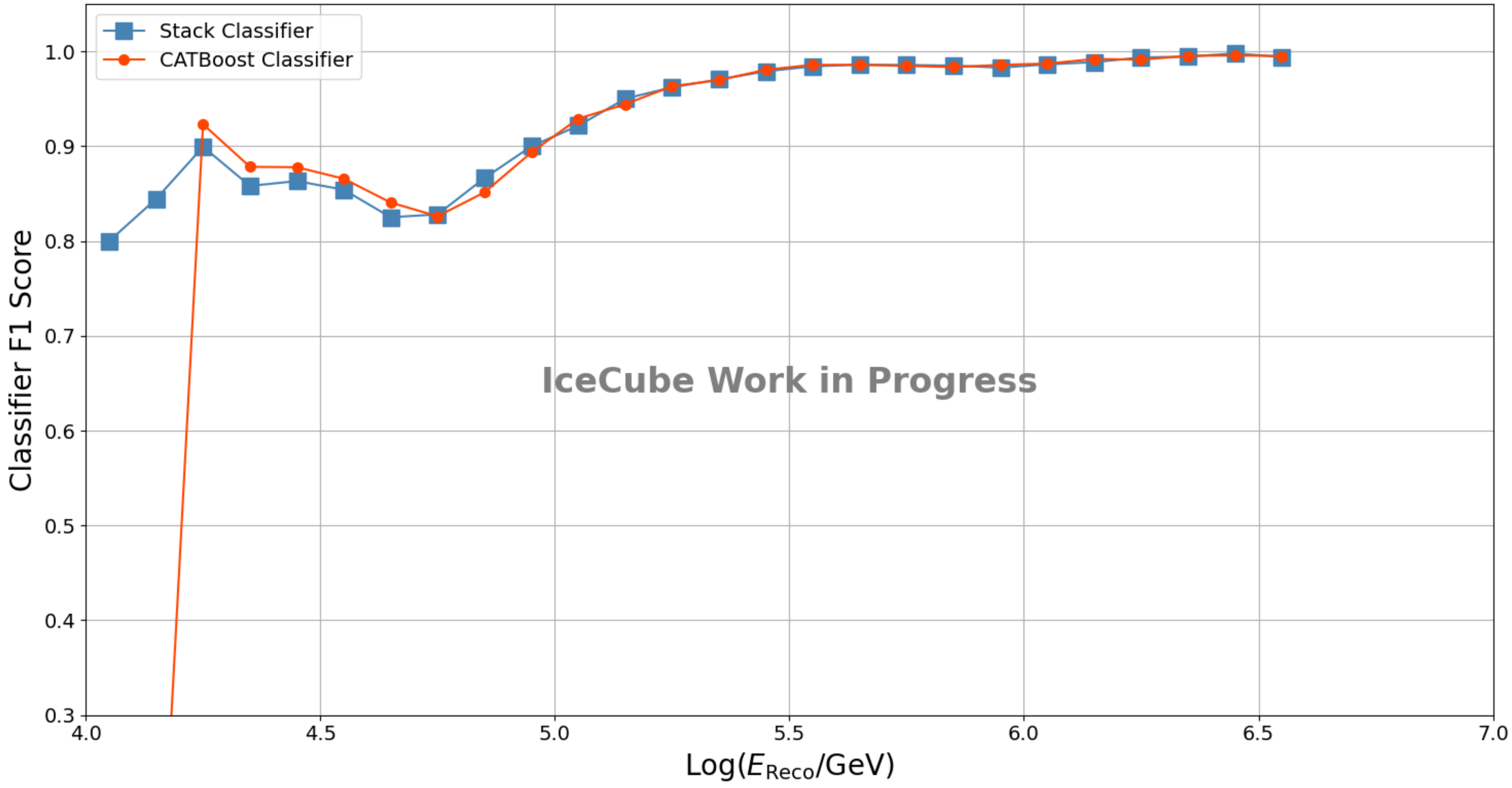
CATBoost



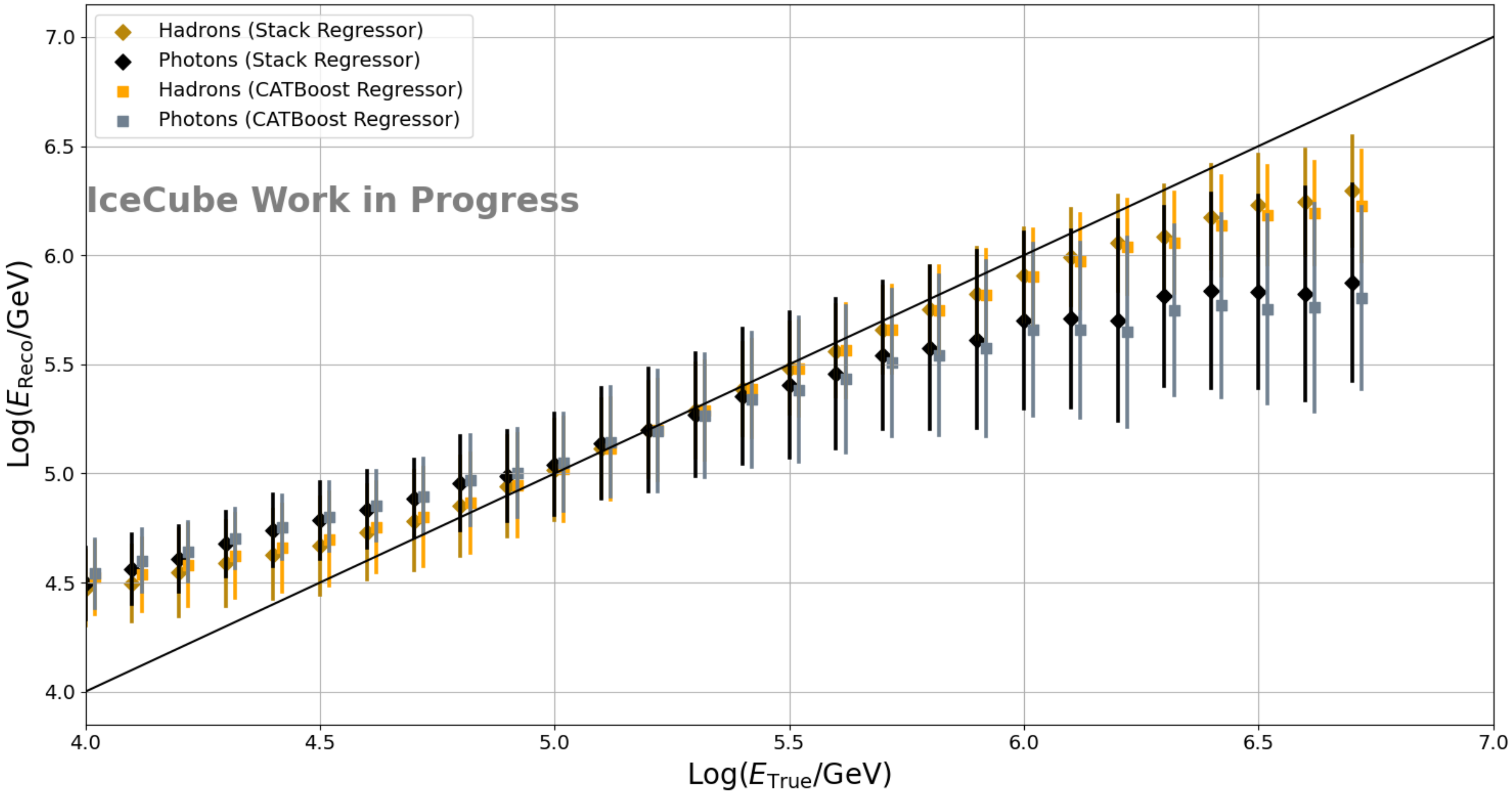
Reco Classifier Accuracies (90% Cut):



Reco Classifier F1 Scores (90% Cut):



Energy Reco (90% Cut):



LnA Reco (90% Cut):