Cosmic Rays Primary Energy estimation using ML and combined reconstruction

Diana Leon Silverio
South Dakota School of Mines and Technology

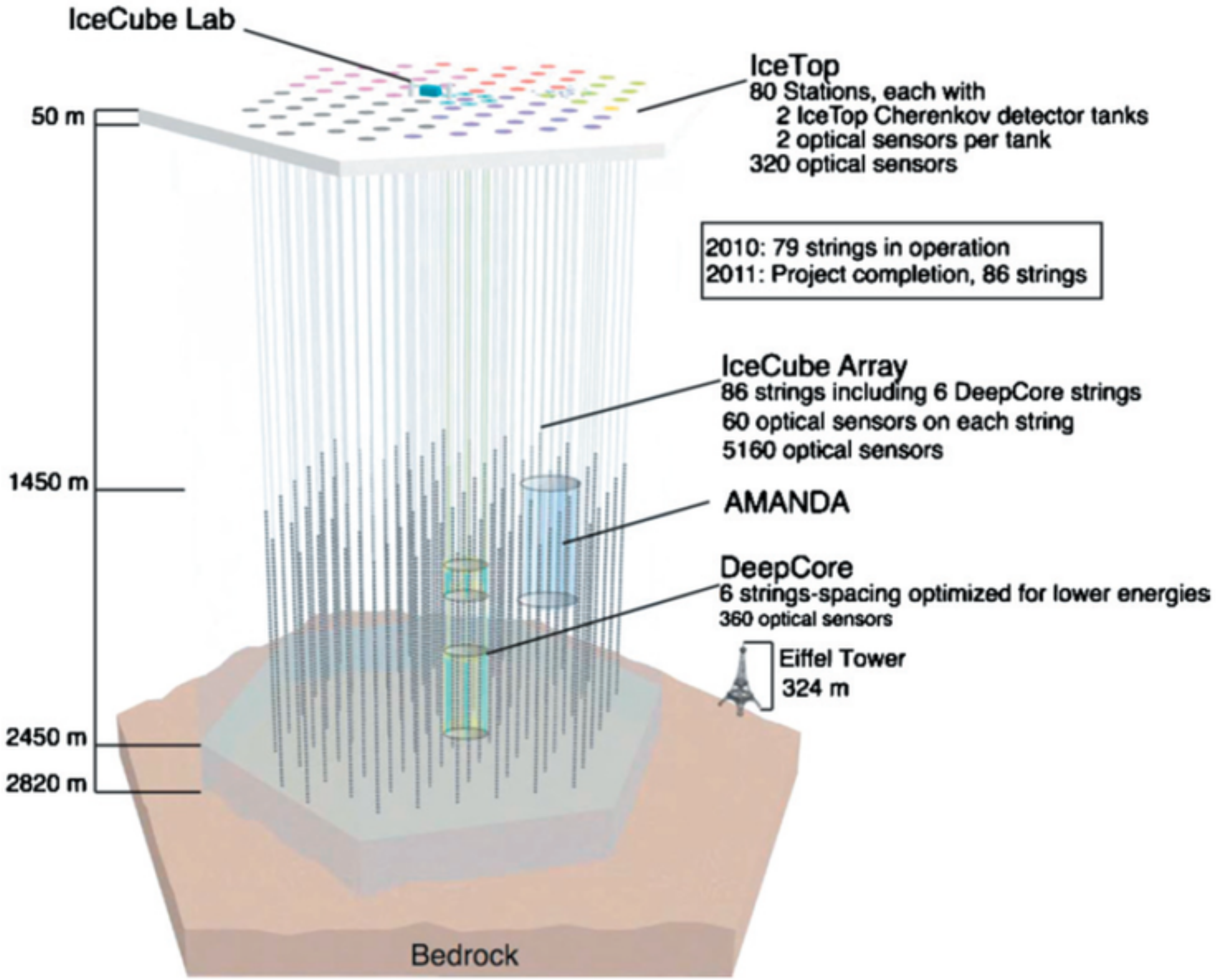
Co-authors: IceCube Collaboration, Xinhua Bai, Matthias Plum

1/Feb/2022

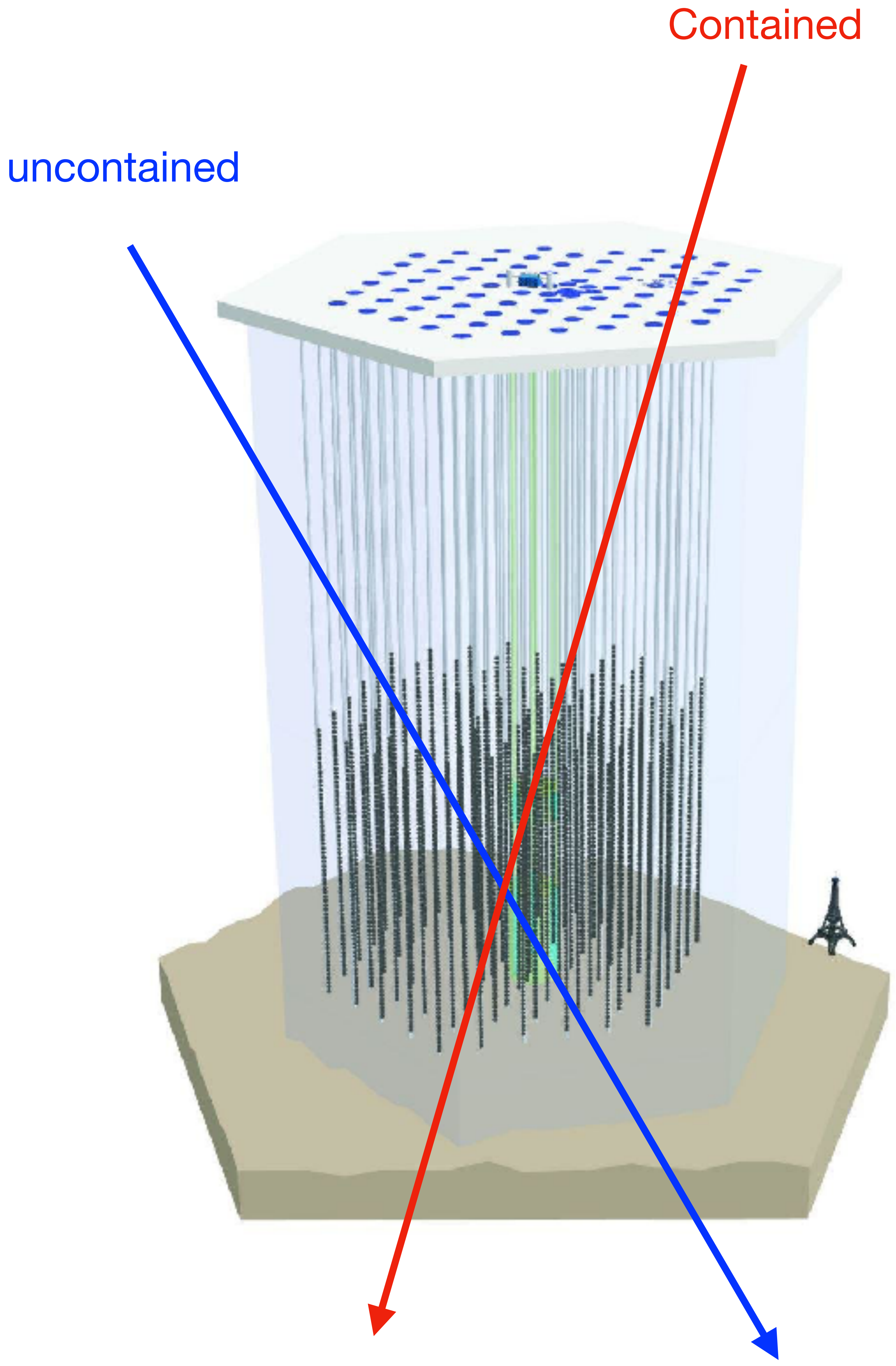
Overview

- Introduction and Motivation
- General description of the reconstruction
- Energy estimation method using Linear fit
- Energy estimation/prediction using Random Forest Regression
- Comparison between Liner Fit and Random Forest Regression performance
- Summary

Introduction



- A lot of techniques and parameters are available and can be combined to determine the primary energy and mass from air showers.



Motivation

- There is no conventional energy estimation function (each analysis is unique; simulation, snowmodel, cuts, etc.).
- Using Random Forest Regression and a combined reconstruction to estimate cosmic rays's primary energy
- Using Monte Carlo events for proton, iron, helium and oxygen under two containment conditions (contained and uncontained).

Goal

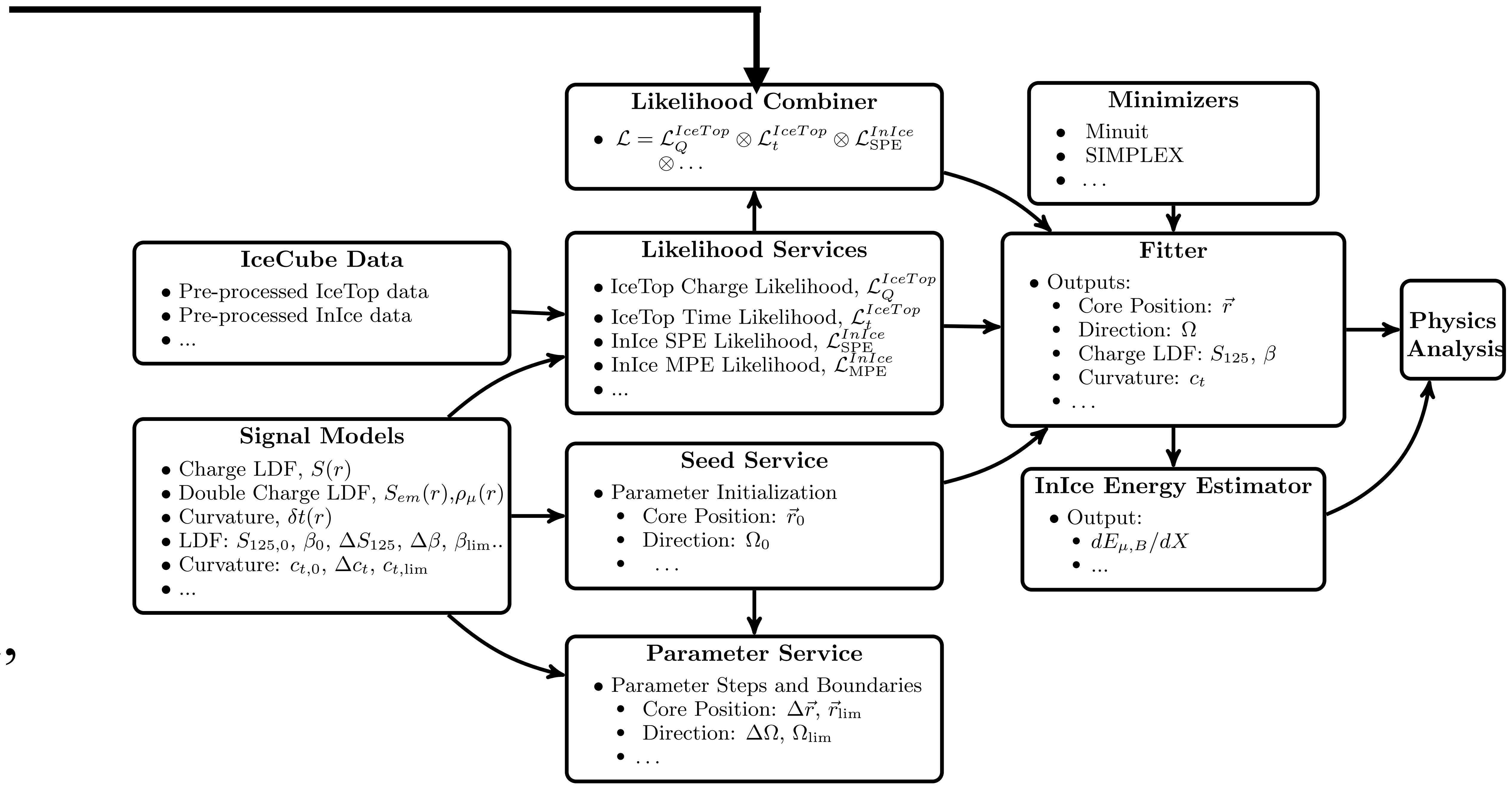
- Investigate a possible improvement for cosmic rays' primary energy estimation

Reconstruction

Events are reconstructed using a <u>combined reconstruction</u> (3D Reconstruction)

<u>Combined Reconstruction</u>

- Combines the likelihoods of IceTop and InIce together for event reconstruction
- Timing Likelihood: Implementation of a flexible curvature and new timing fluctuation

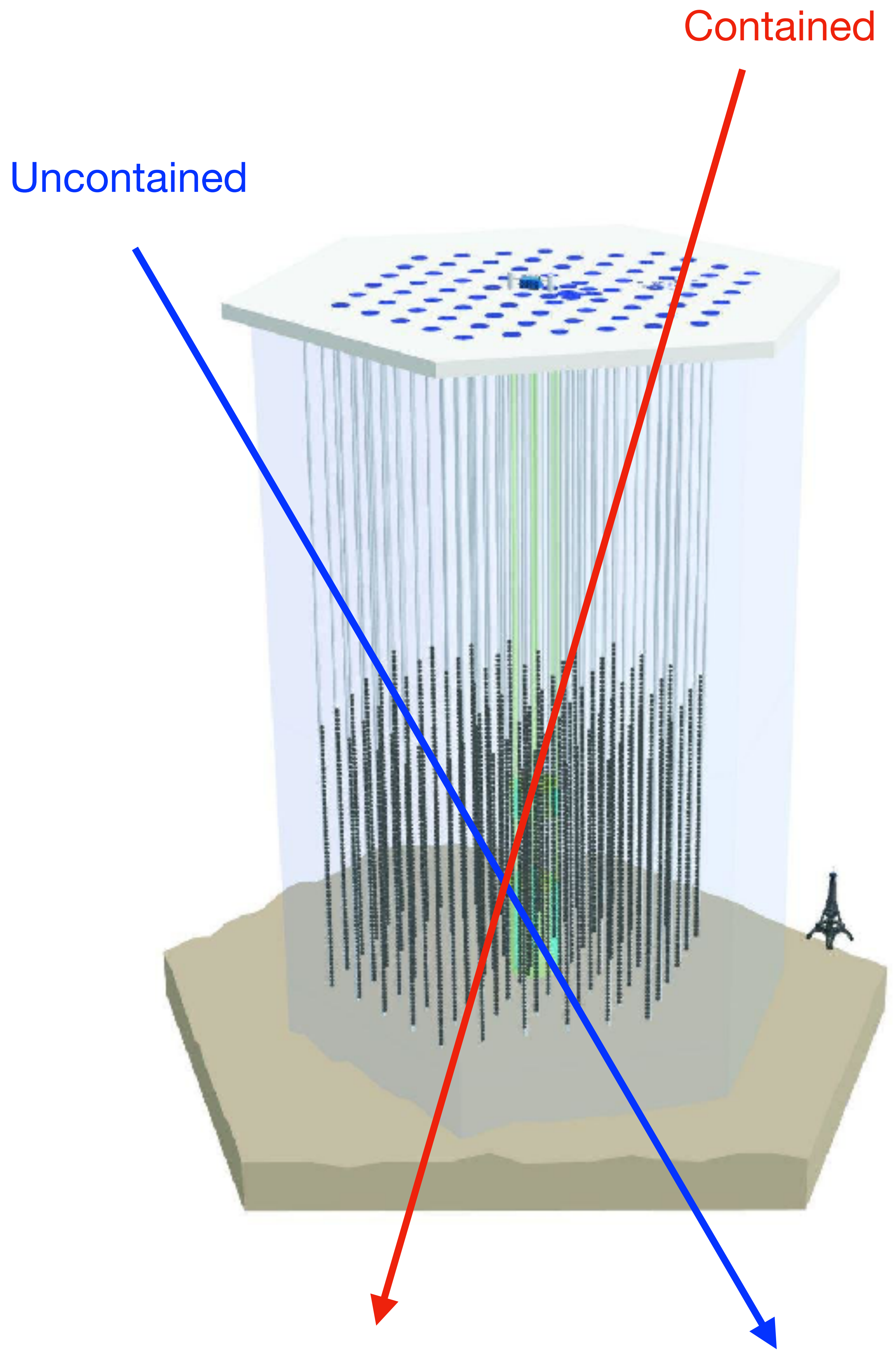


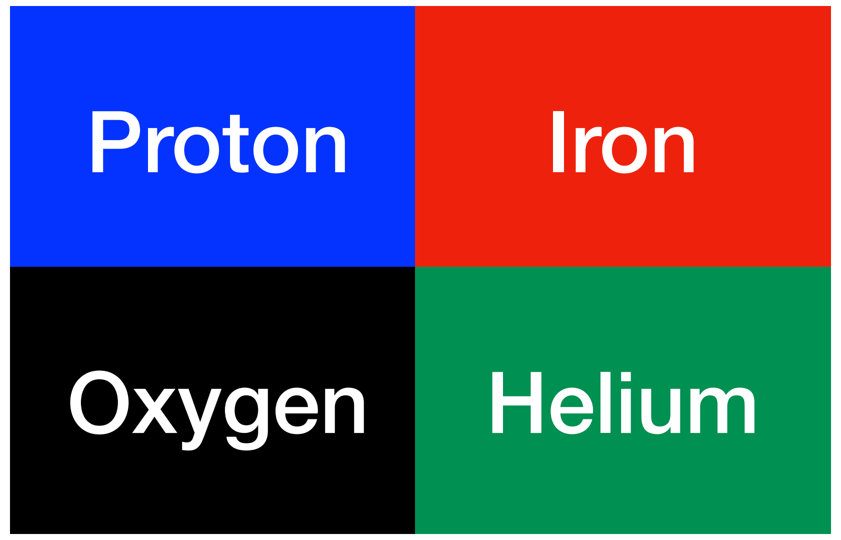
https://wiki.icecube.wisc.edu/index.php/RockBottom

<u>Events</u>

- Monte Carlo events for proton, helium, oxygen, and iron primaries (IC86-2012).

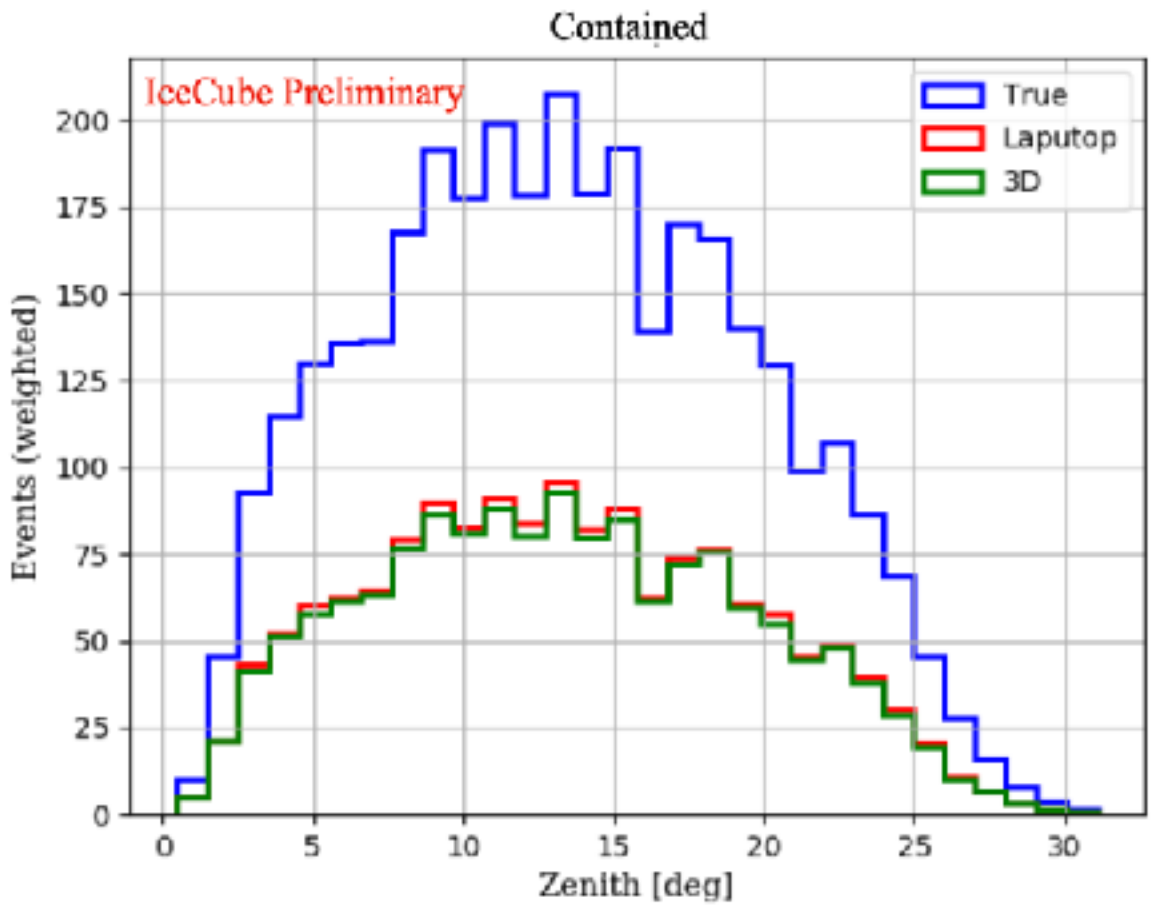
Energy estimation method using Linear fit

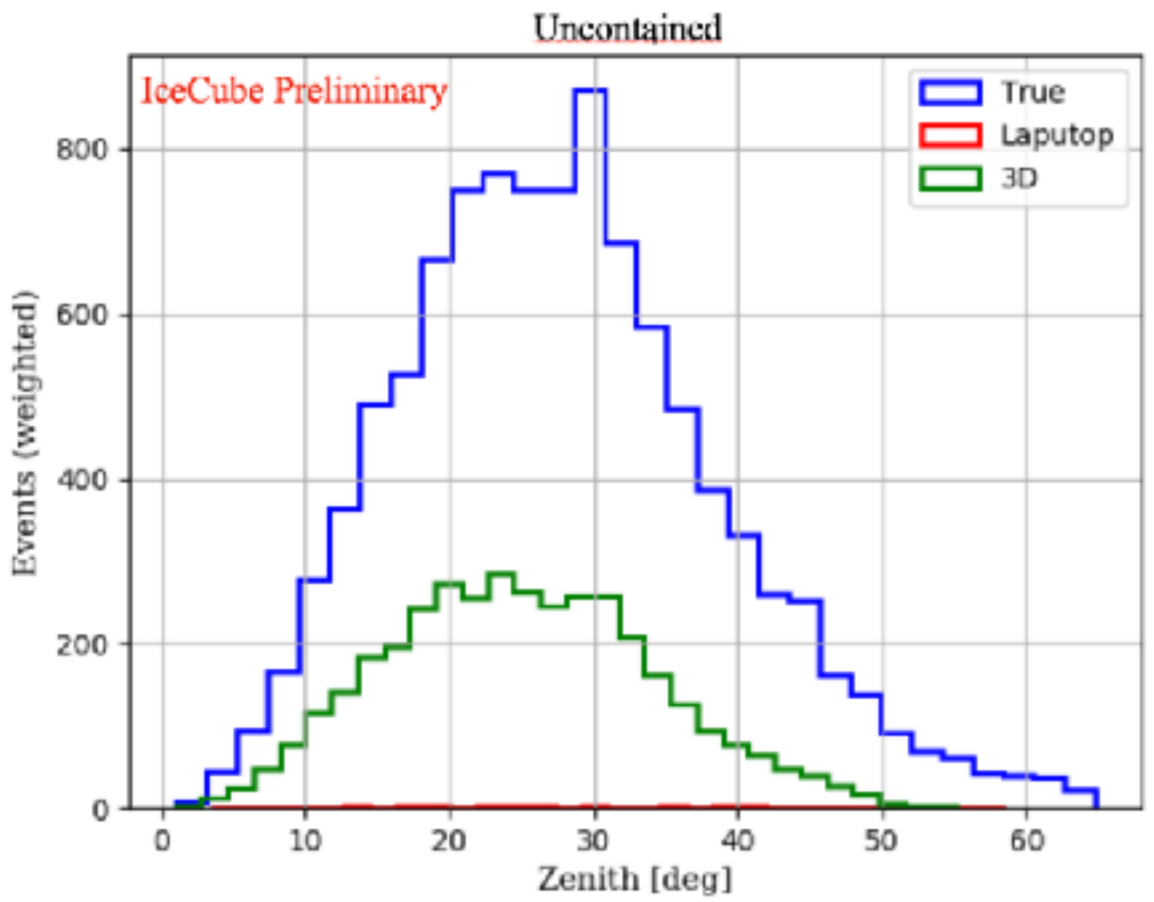




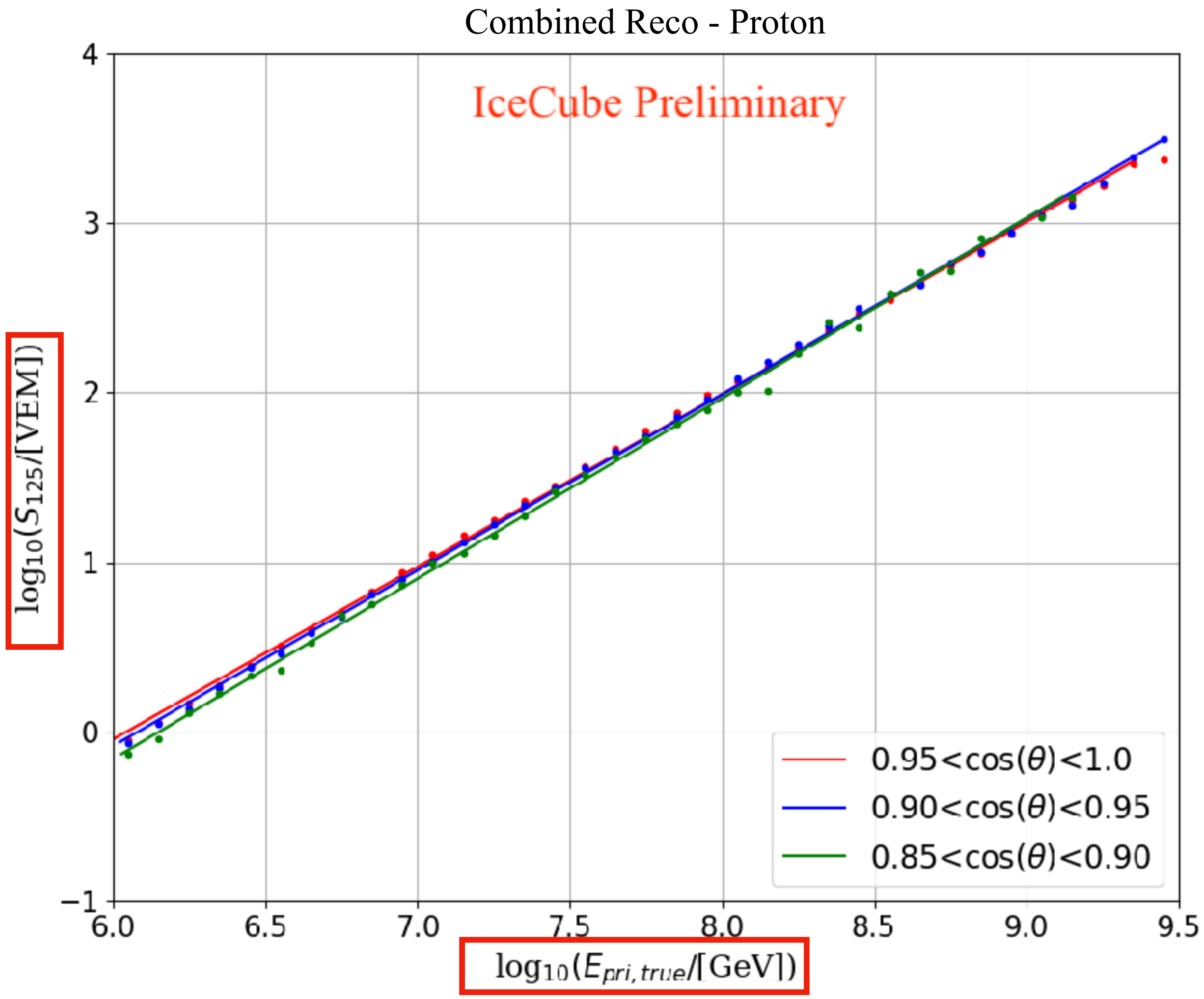
Linear Fit method:

- Analyze proton, helium, oxygen, and iron primaries separately.
- For each primaries split events in contained and uncontained events
- Considering different zenith ranges





Energy estimation method using Linear fit

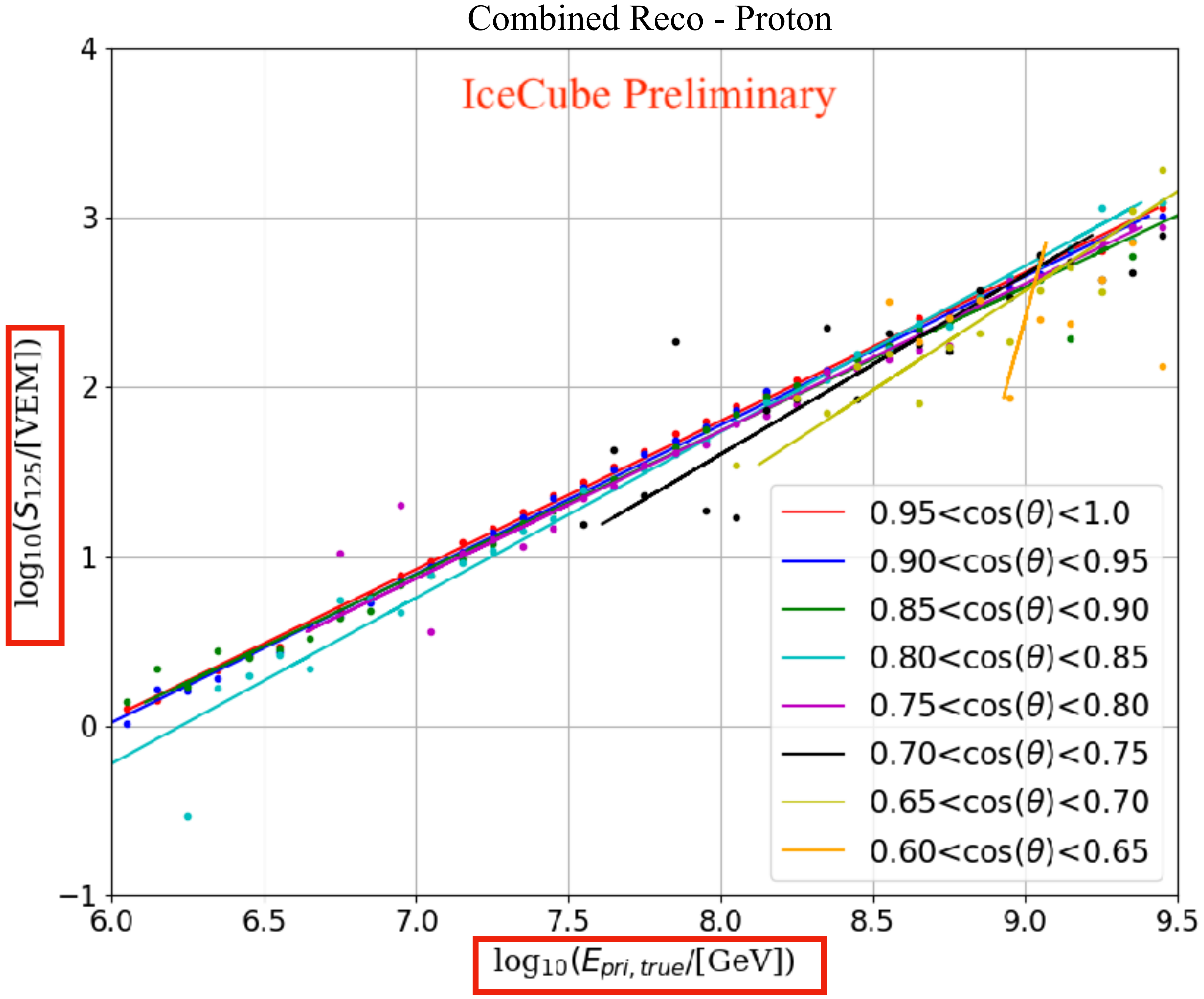


$$log_{10}(E[GeV]) = p_0 + p_1 log_{10}(S_{125}[VEM])$$

Combined Reco - Contained Events		
Zenith	p_0	p_1
$0.95< cos(\theta) <1.0$	6.0446	0.9824
$0.90< cos(\theta) <0.95$	6.0822	0.9639
$0.85< cos(\theta) <0.90$	6.1517	0.9399

Note: Calculation for Helium, Oxygen and Iron at backup

Energy estimation method using Linear fit



$$log_{10}(E[GeV]) = p_0 + p_1 log_{10}(S_{125}[VEM])$$

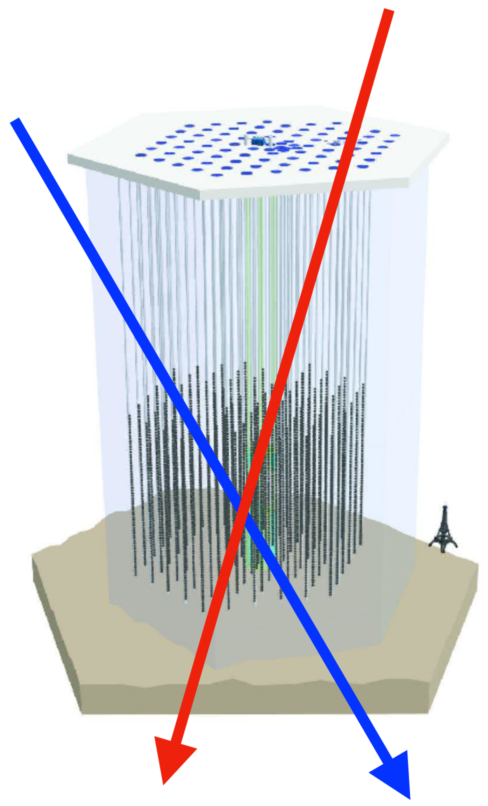
Combiend Reco - Uncontained Events		
Zenith	p_0	p_1
$0.95< cos(\theta) <1.0$	5.9447	1.1413
$0.90< cos(\theta) <0.95$	5.9767	1.1390
$0.85< cos(\theta) <0.90$	5.9451	1.1796
$0.80< cos(\theta) <0.85$	6.2266	1.0214
$0.75< cos(\theta) <0.80$	5.9997	1.1487
$0.70< cos(\theta) <0.75$	6.4873	0.9432
$0.65< cos(\theta) <0.70$	6.8099	0.8525
$0.60< cos(\theta) <0.65$	8.6340	0.1522

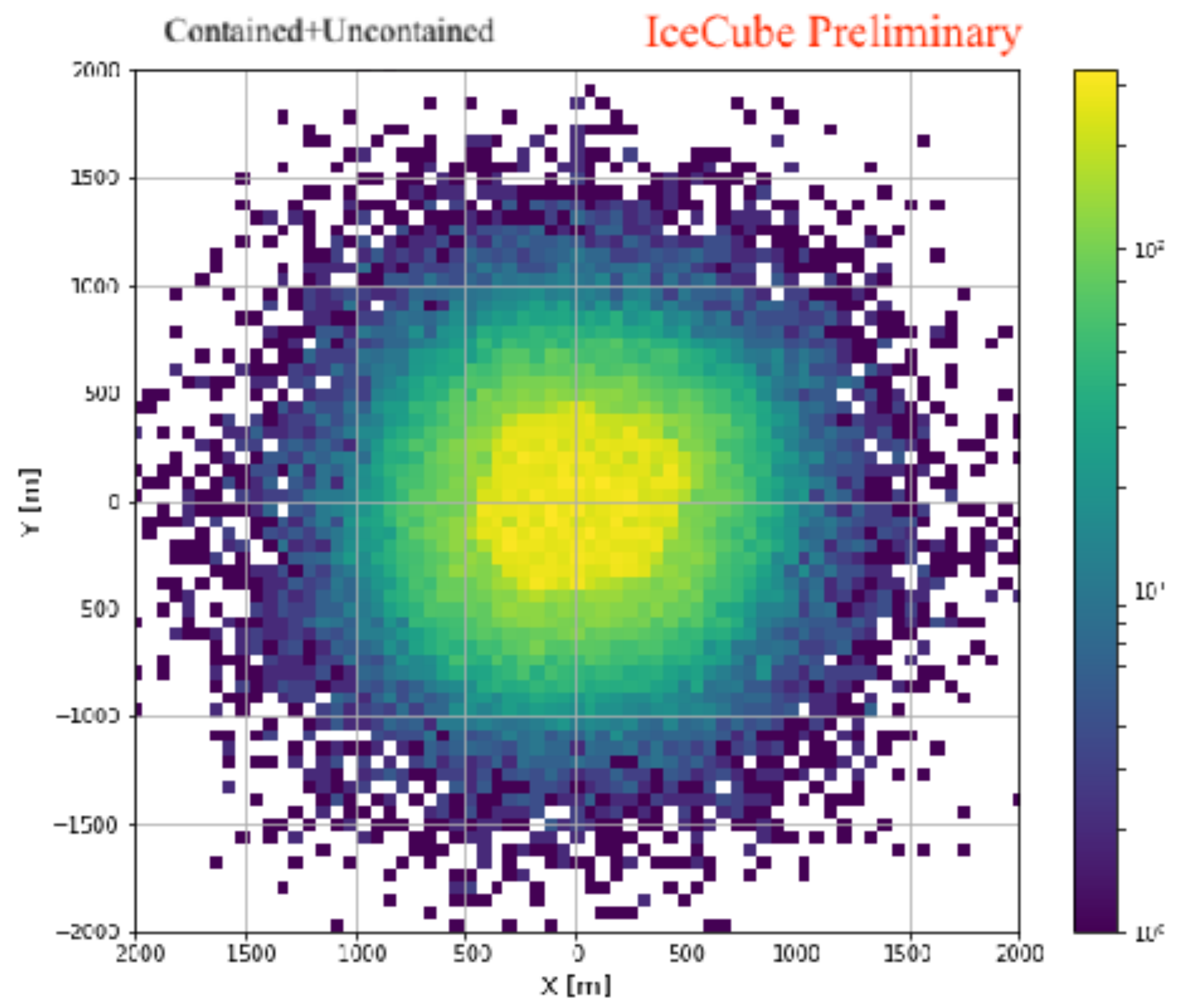
Note: Calculation for Helium, Oxygen and Iron at backup

SOUTH DAKOTA MINES

Energy estimation/prediction using Random Forest Regression

Proton	Iron
Oxygen	Helium





Random Forest Regression prediction:

- Consider proton, helium, oxygen and iron primaries together!!.
- Consider contained and uncontained events together!!.
- Zenith dependence as one feature parameter for training.
- Give the possibility to consider InIce feature parameters.

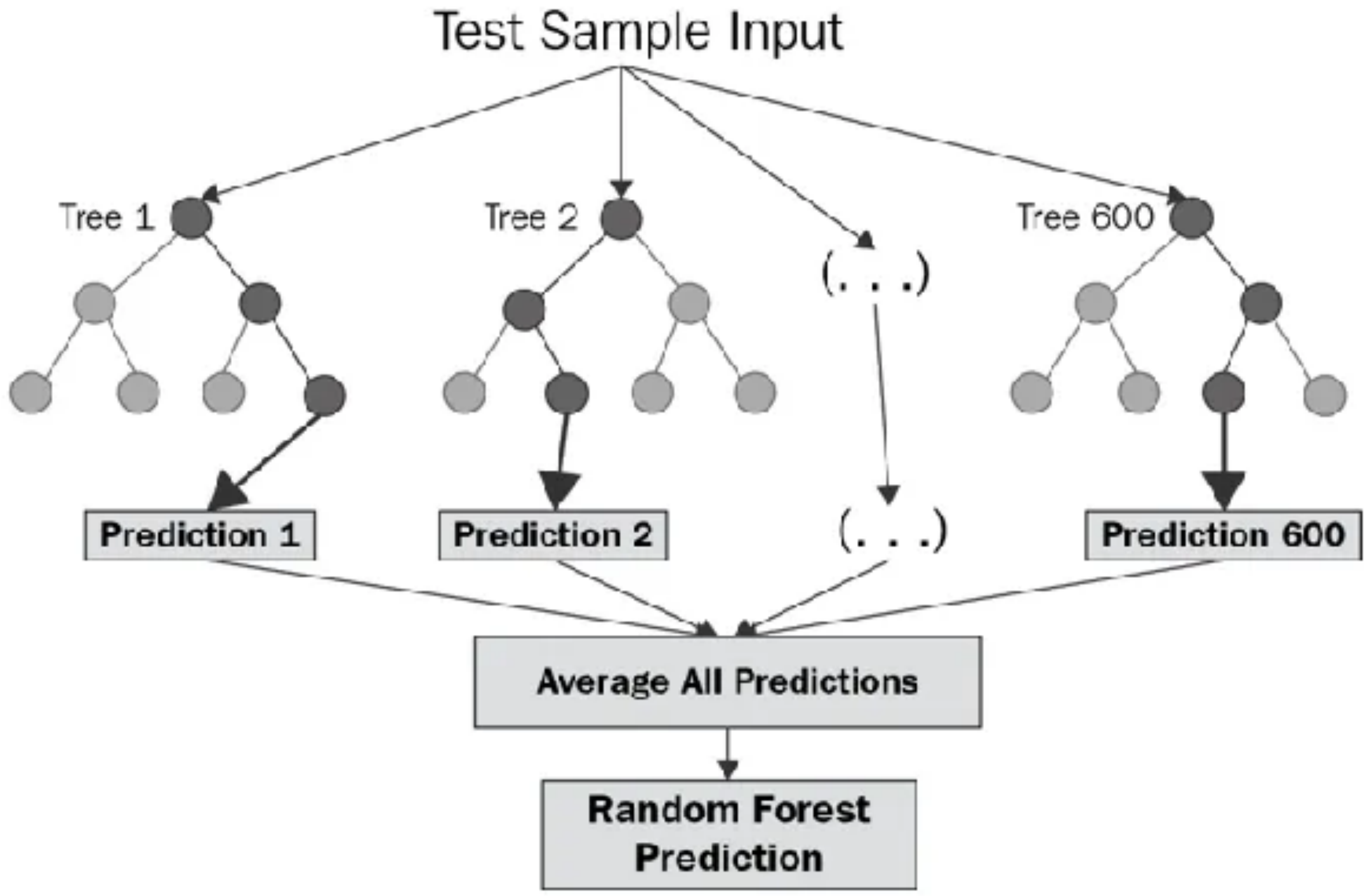
Energy estimation/prediction using Random Forest Regression

Events Selection

- Monte Carlo events for proton, iron, helium and oxygen (IC86-2012)
- Selection Criteria:
  - Contained: Core location inside of IceTop Array and InIce muon track
  - Uncontained: Core location outside IceTop Array and InIce muon track
- Requiring a successful reconstruction

Machine Learning

- Random Forest Regression (open-source Python package Scikit-Learn).



- R^2 regression score function.

Energy estimation/prediction using Random Forest Regression

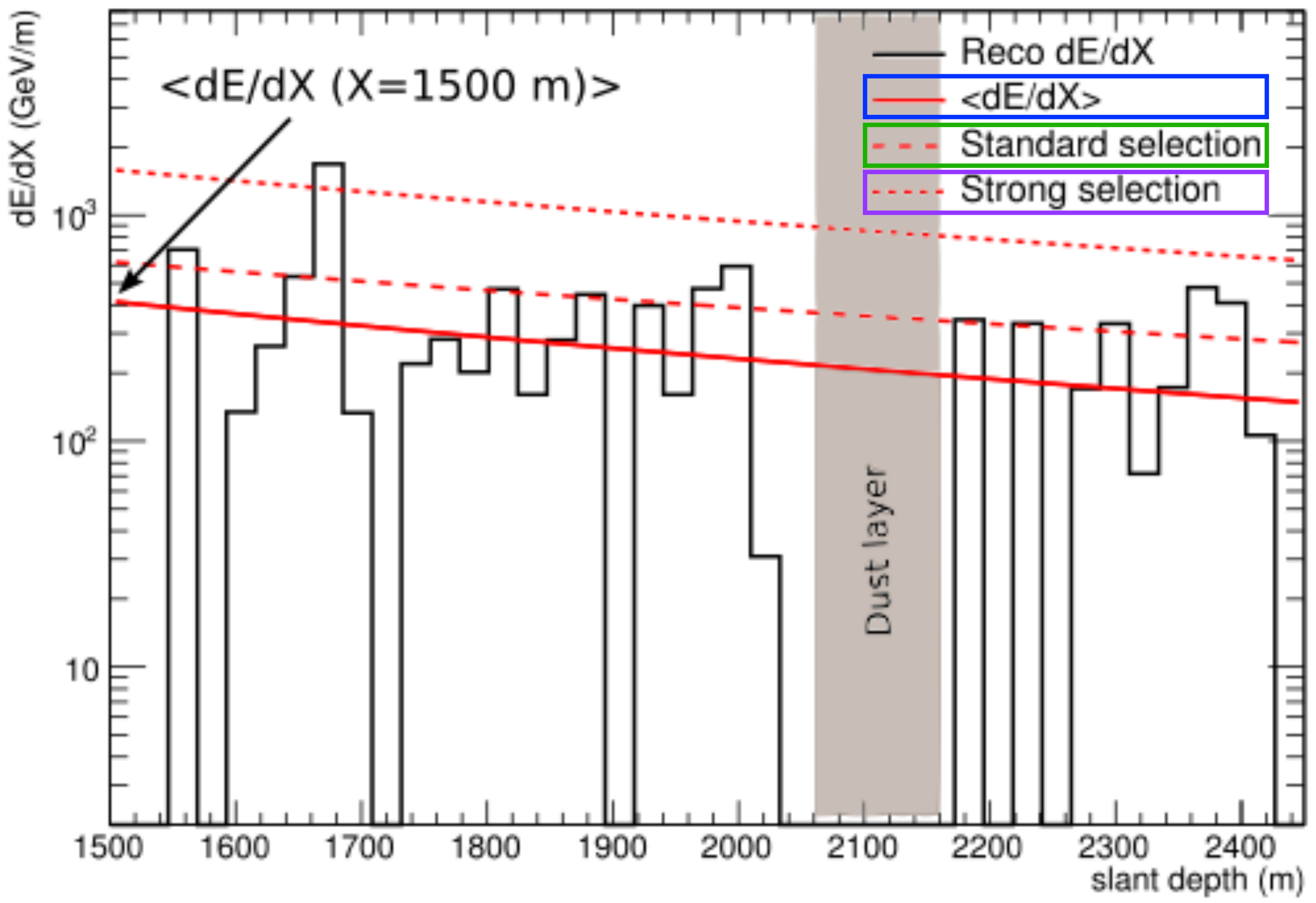
LDF

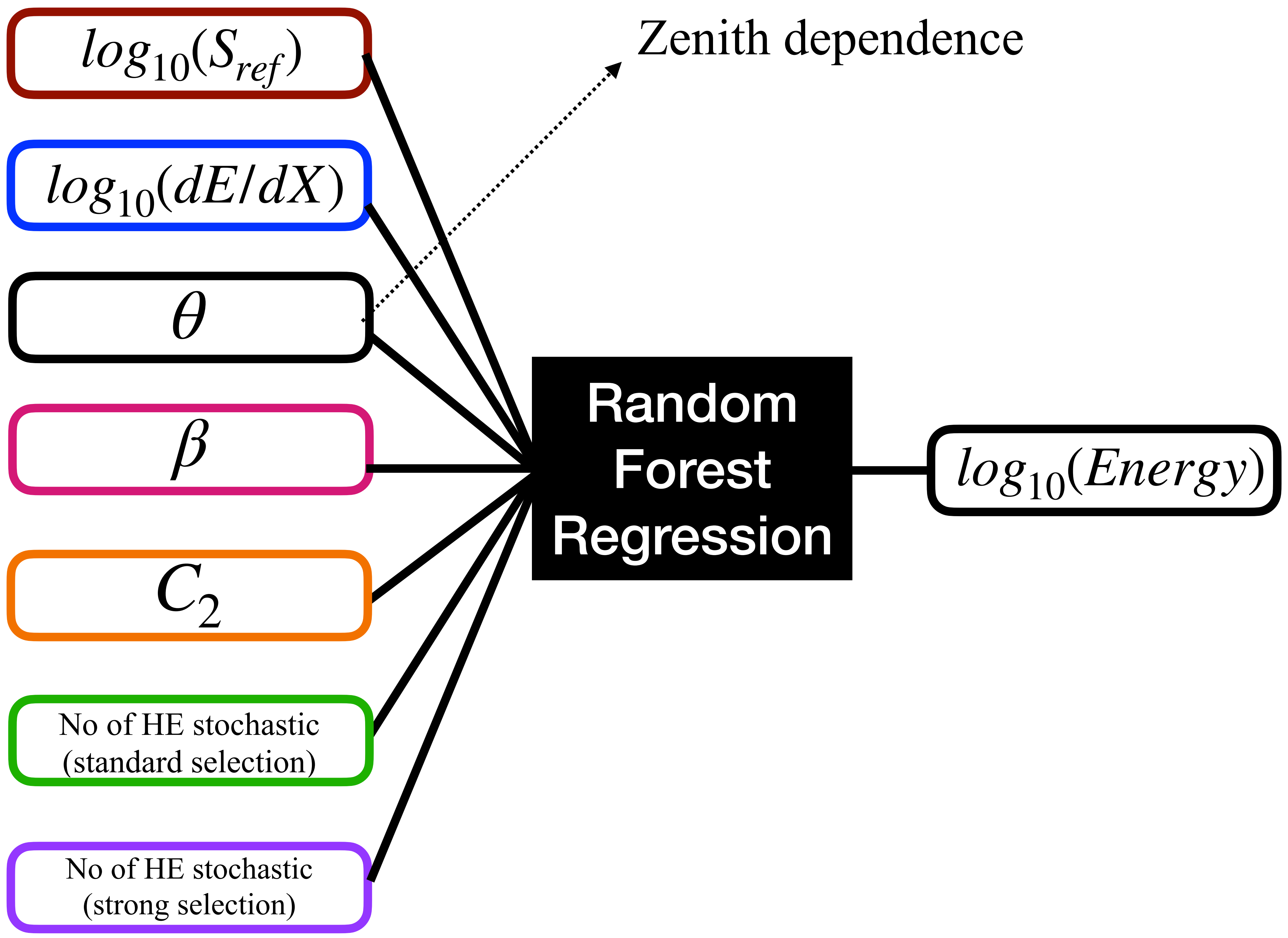
$$S(r) = \boxed{S_{ref}}(\frac{R}{R_{ref}})^{\boxed{-\beta}-\kappa log_{10}(R/R_{ref})}$$

Time Residuals

$$\Delta t(R) = \boxed{C_2}R^2 + C_1(1 - exp(-\frac{R^2}{2\sigma^2}))$$

https://arxiv.org/pdf/1906.04317.pdf

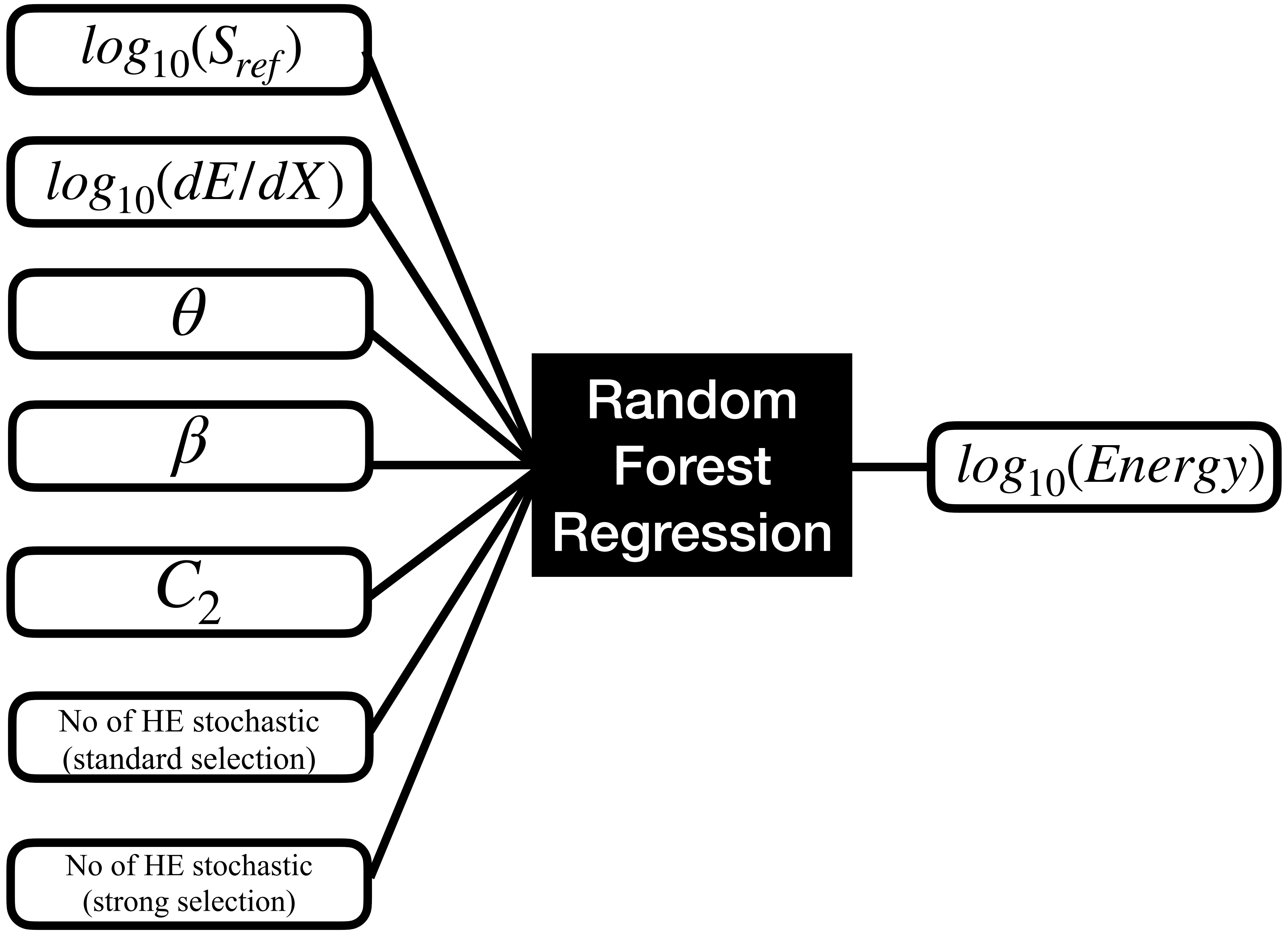




Energy estimation/prediction using Random Forest Regression

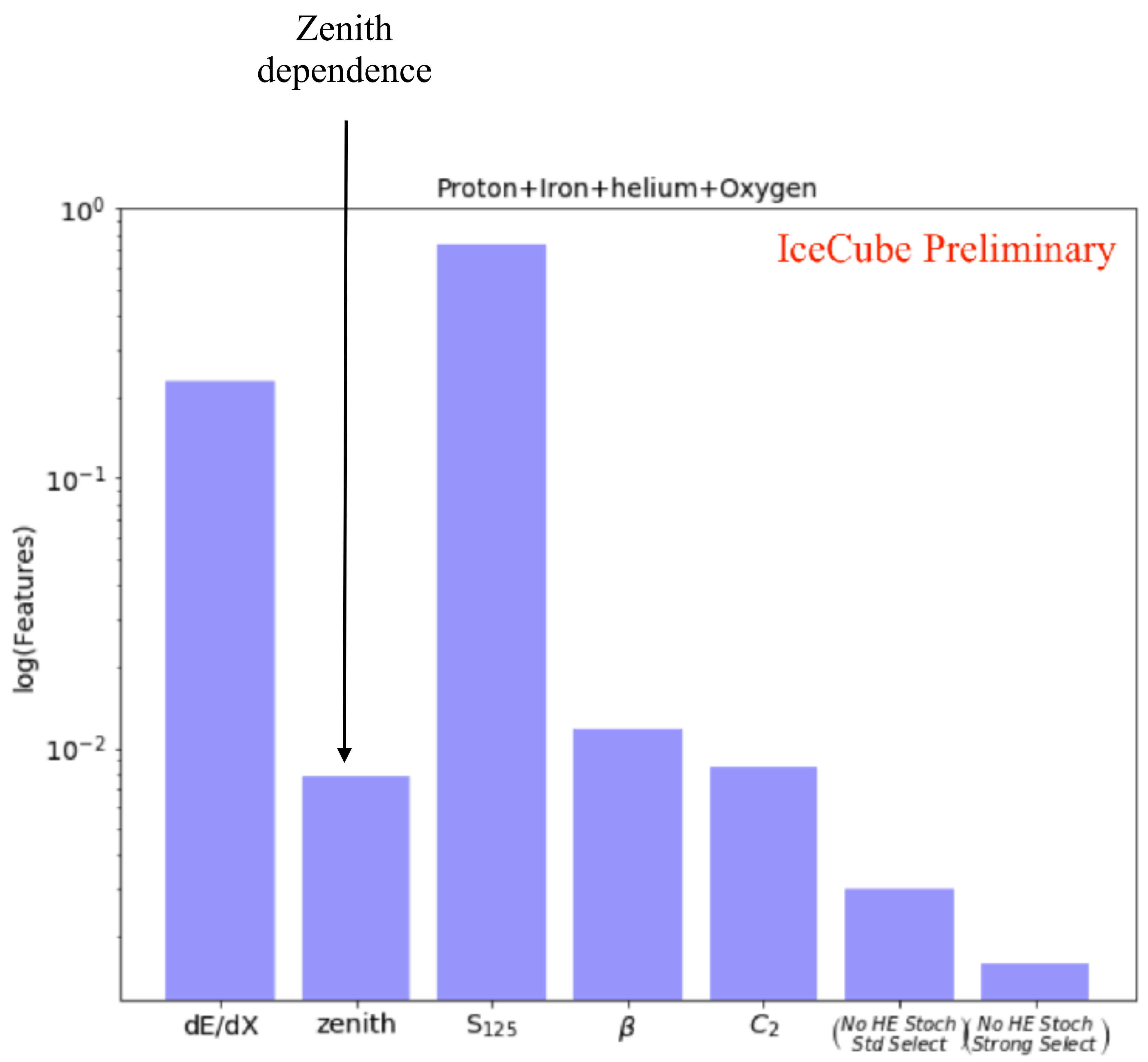
The training and testing dataset has a total 176070 events (proton, iron, helium and oxygen).

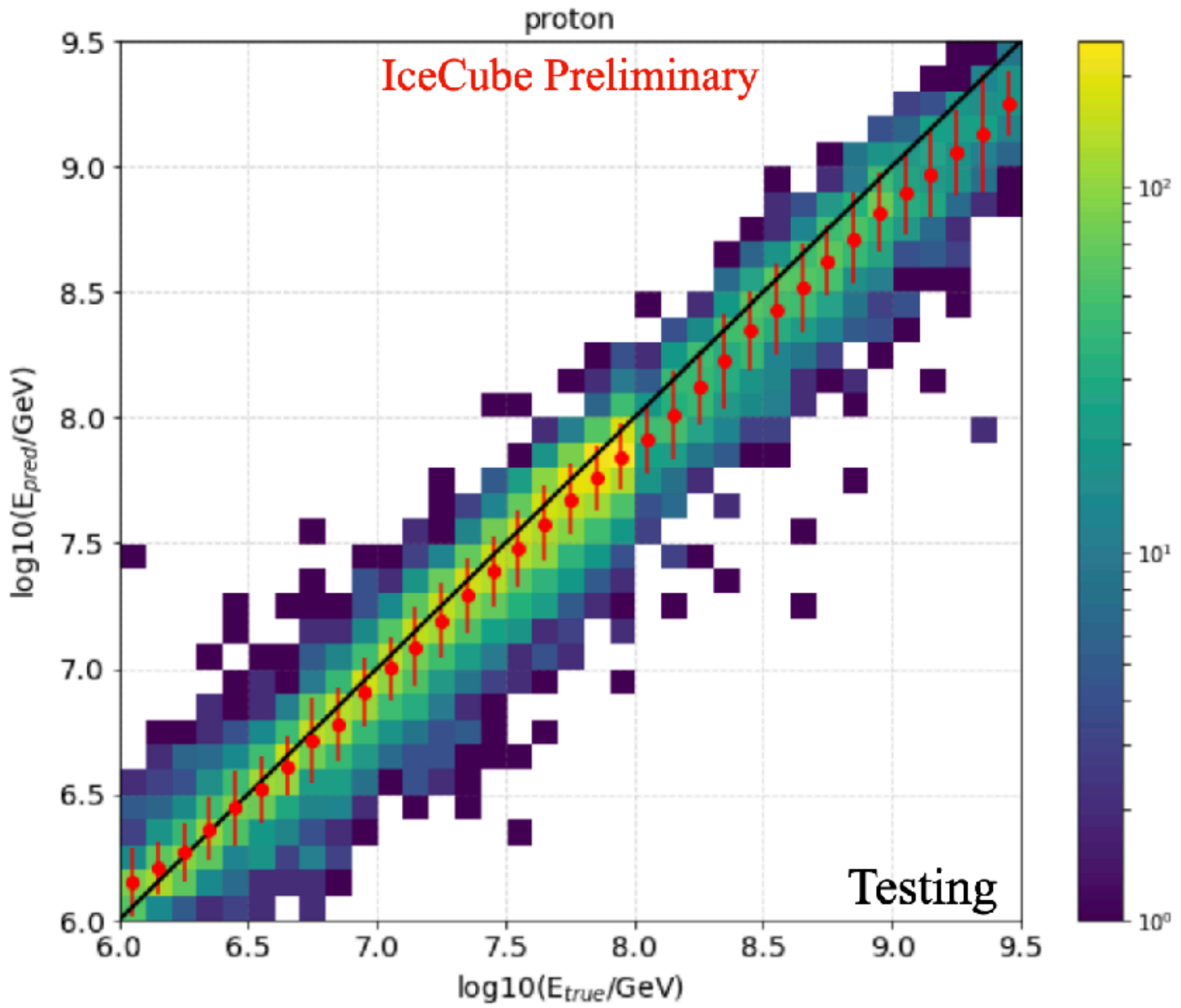
- 75% for Training
- 25% for Testing



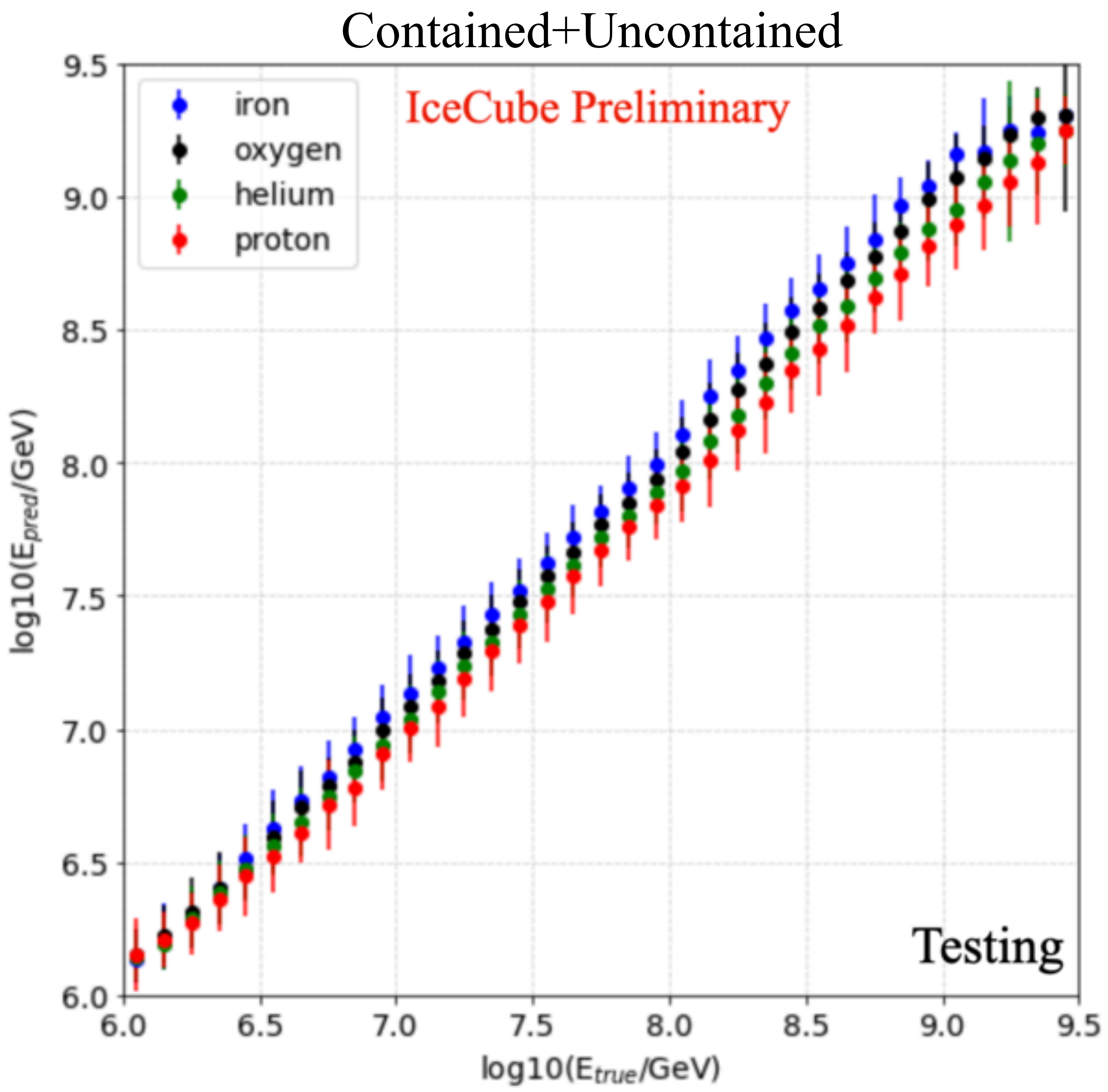
Regression Score	
Training	99.48%
Testing	96.37%

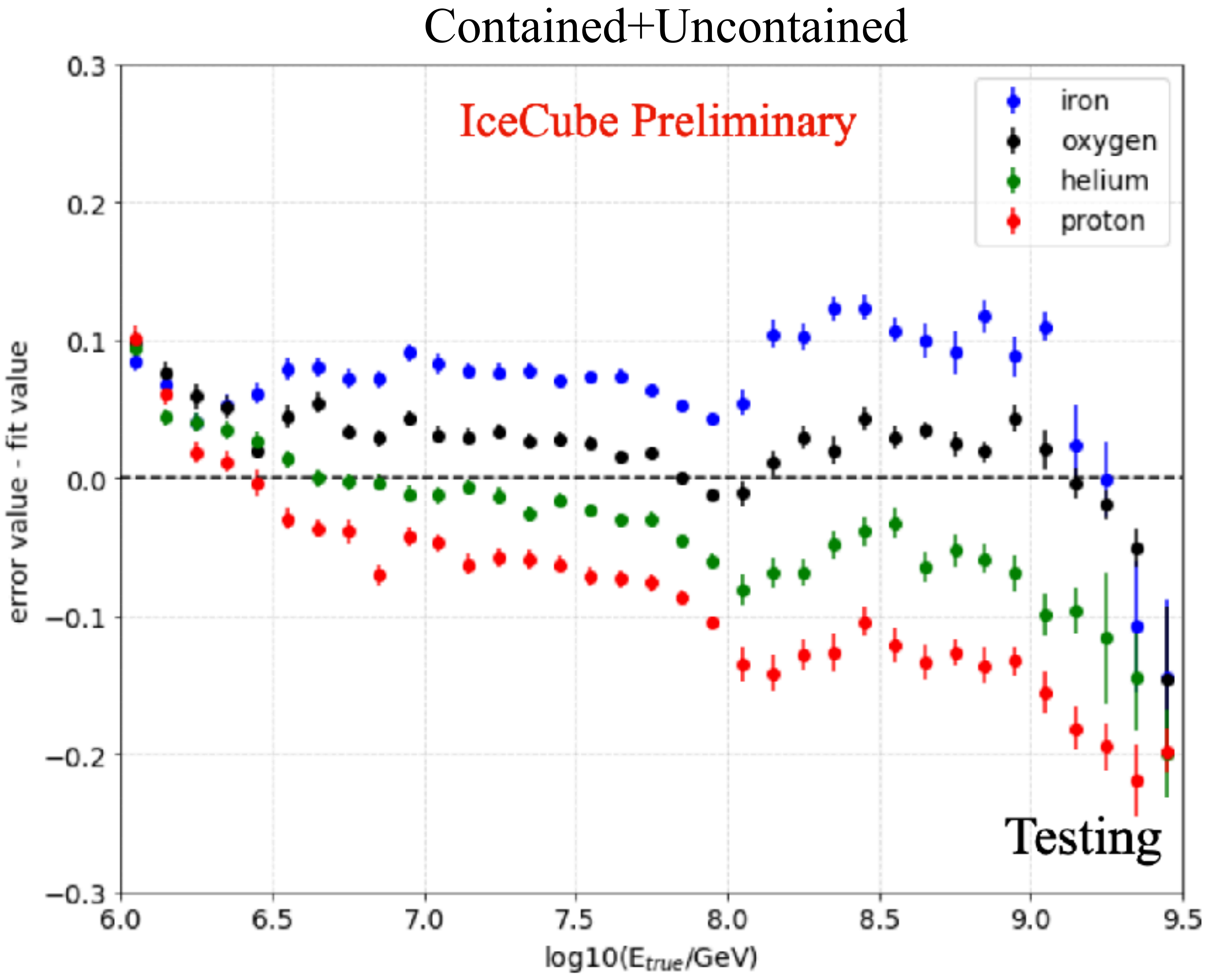
Energy estimation/prediction using Random Forest Regression



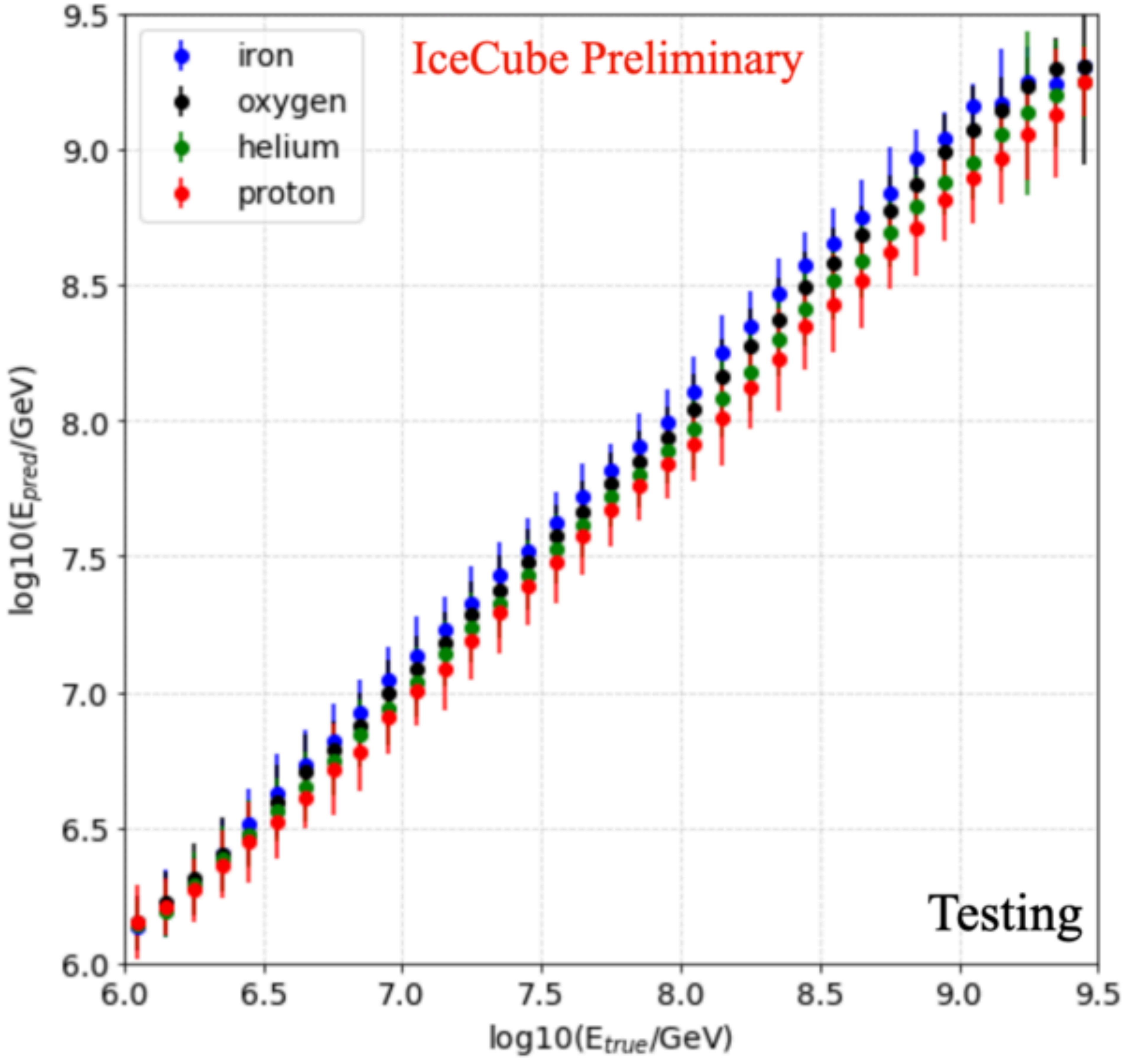


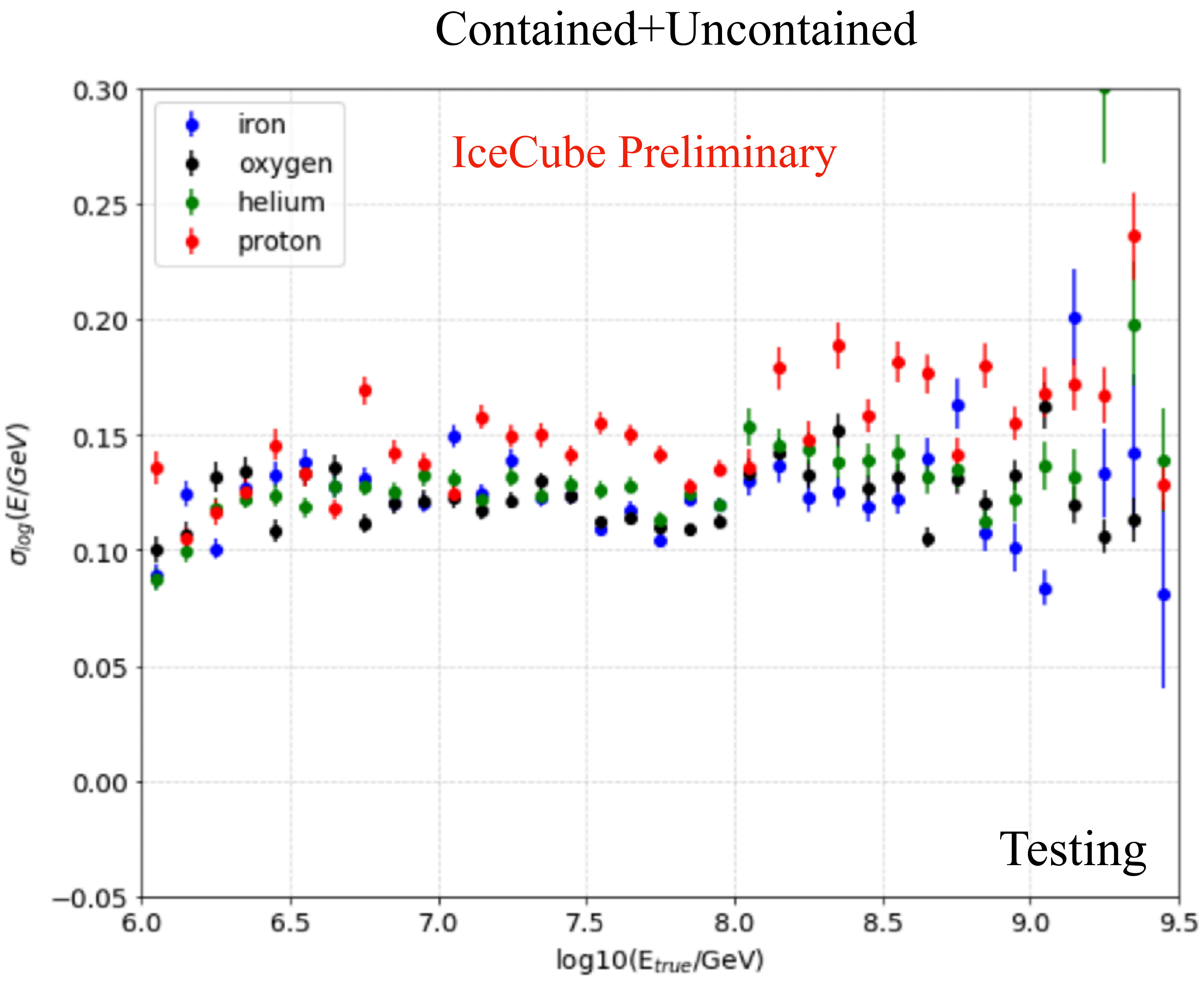
Energy estimation/prediction using Random Forest Regression





Energy estimation/prediction using Random Forest Regression





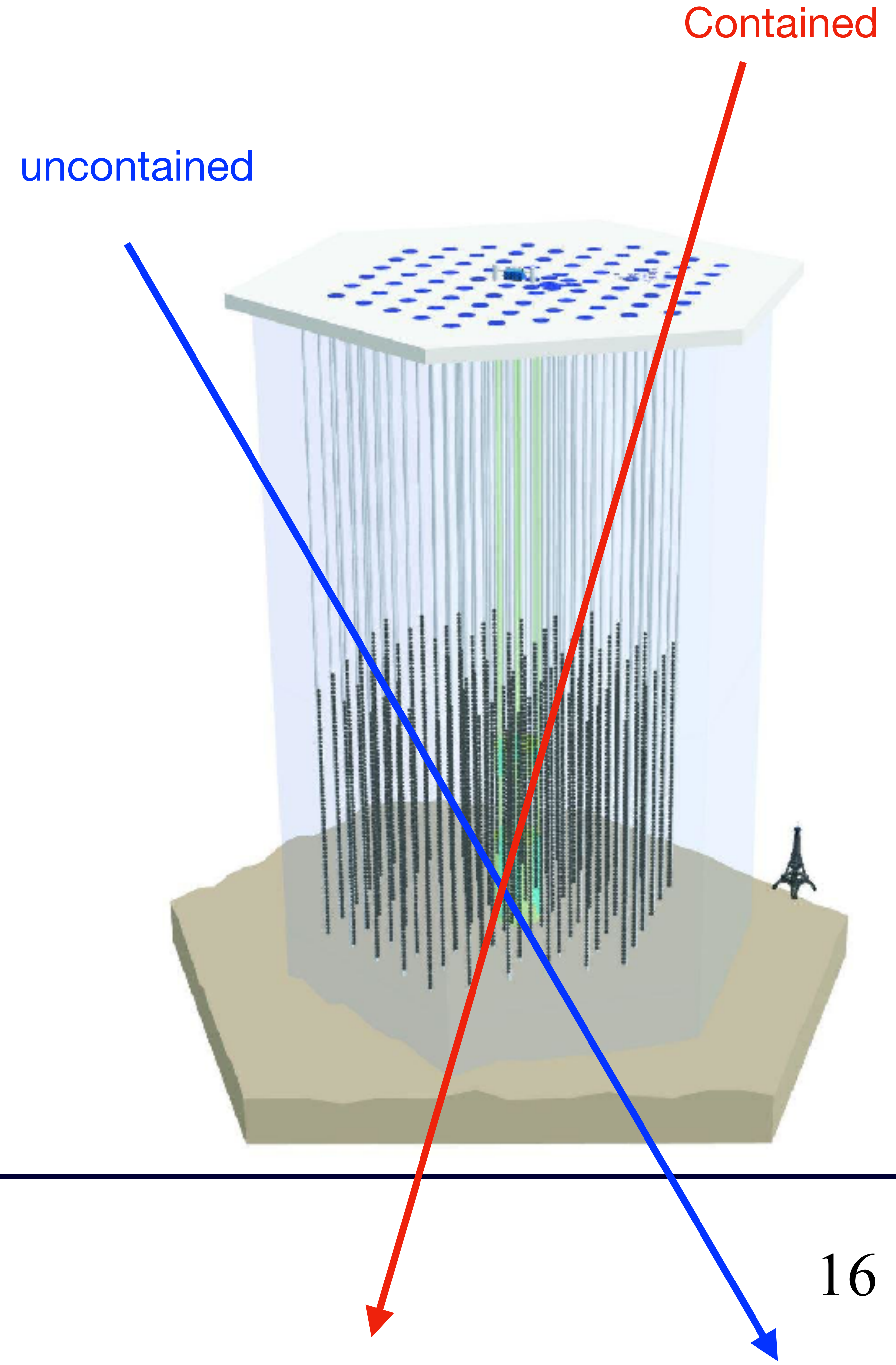
SOUTH DAKOTA MINES

Comparison between Linear Fit and Random Forest Regression performance

- Preliminary comparison considering proton and iron events (contained and uncontained)
- Specific energy ranges:

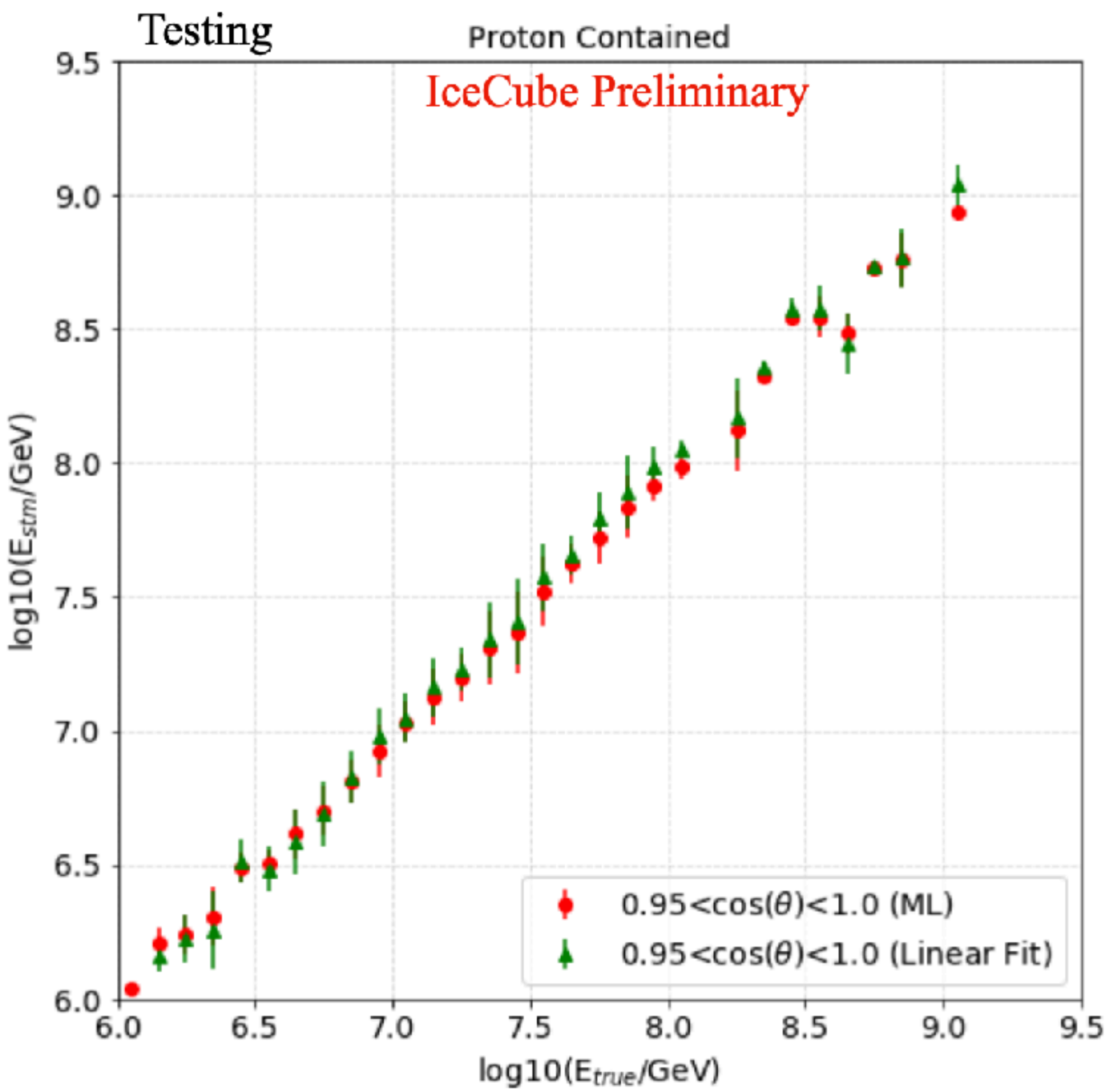
  -Contained: $0.95 < cos(\theta) < 1.0$

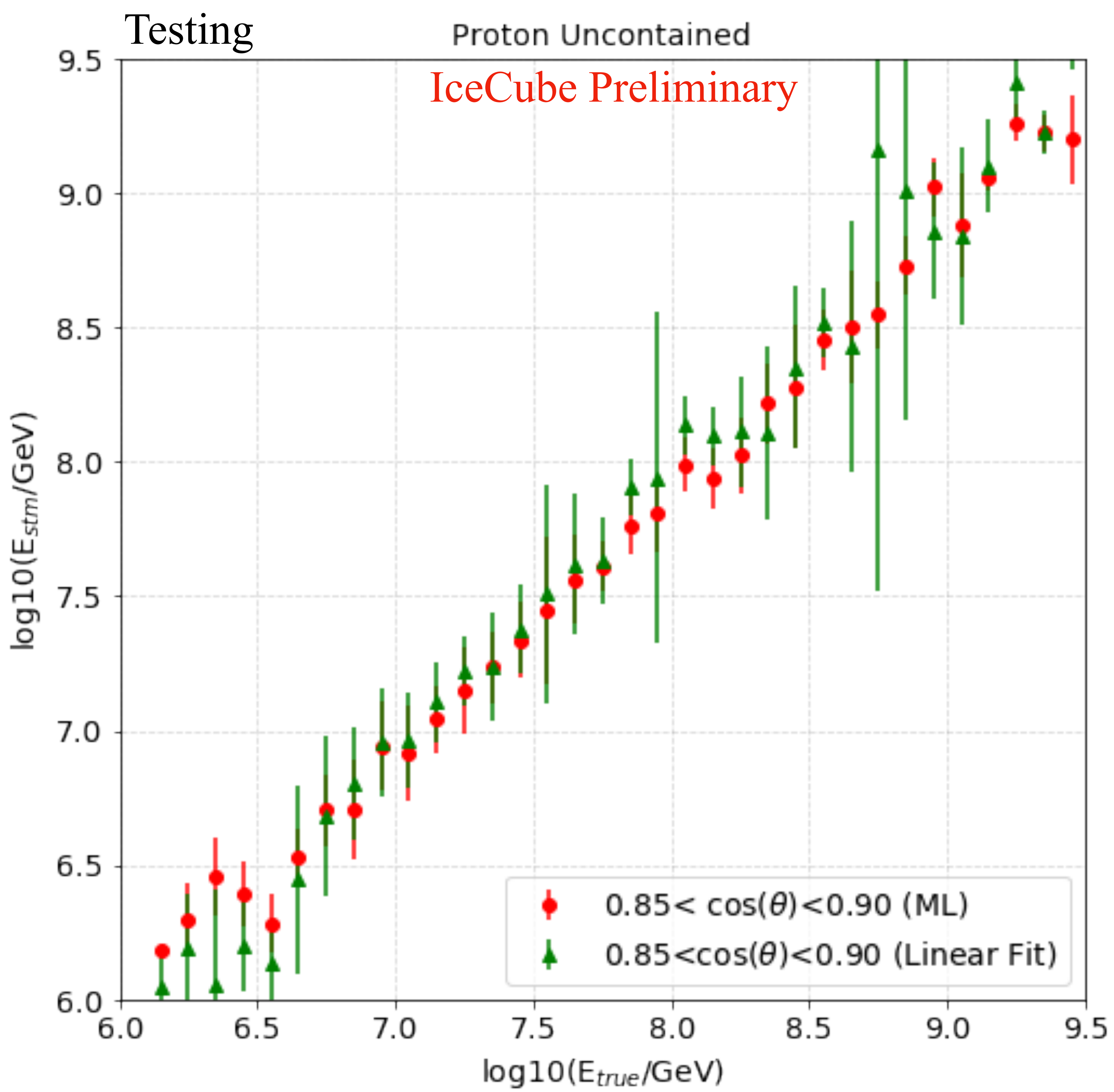
  -Uncontained: $0.85 < cos(\theta) < 0.90$



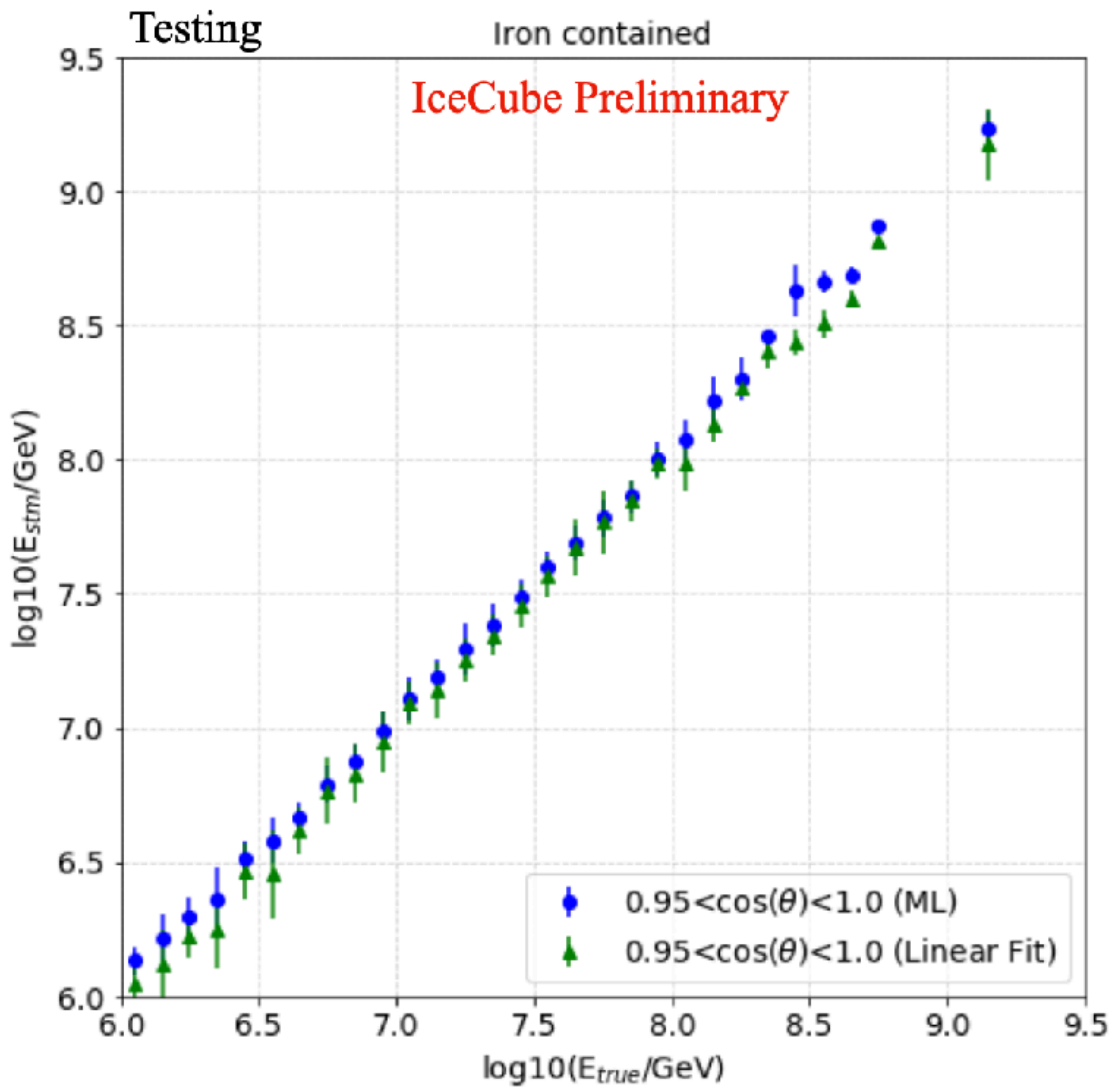
SOUTH DAKOTA MINES

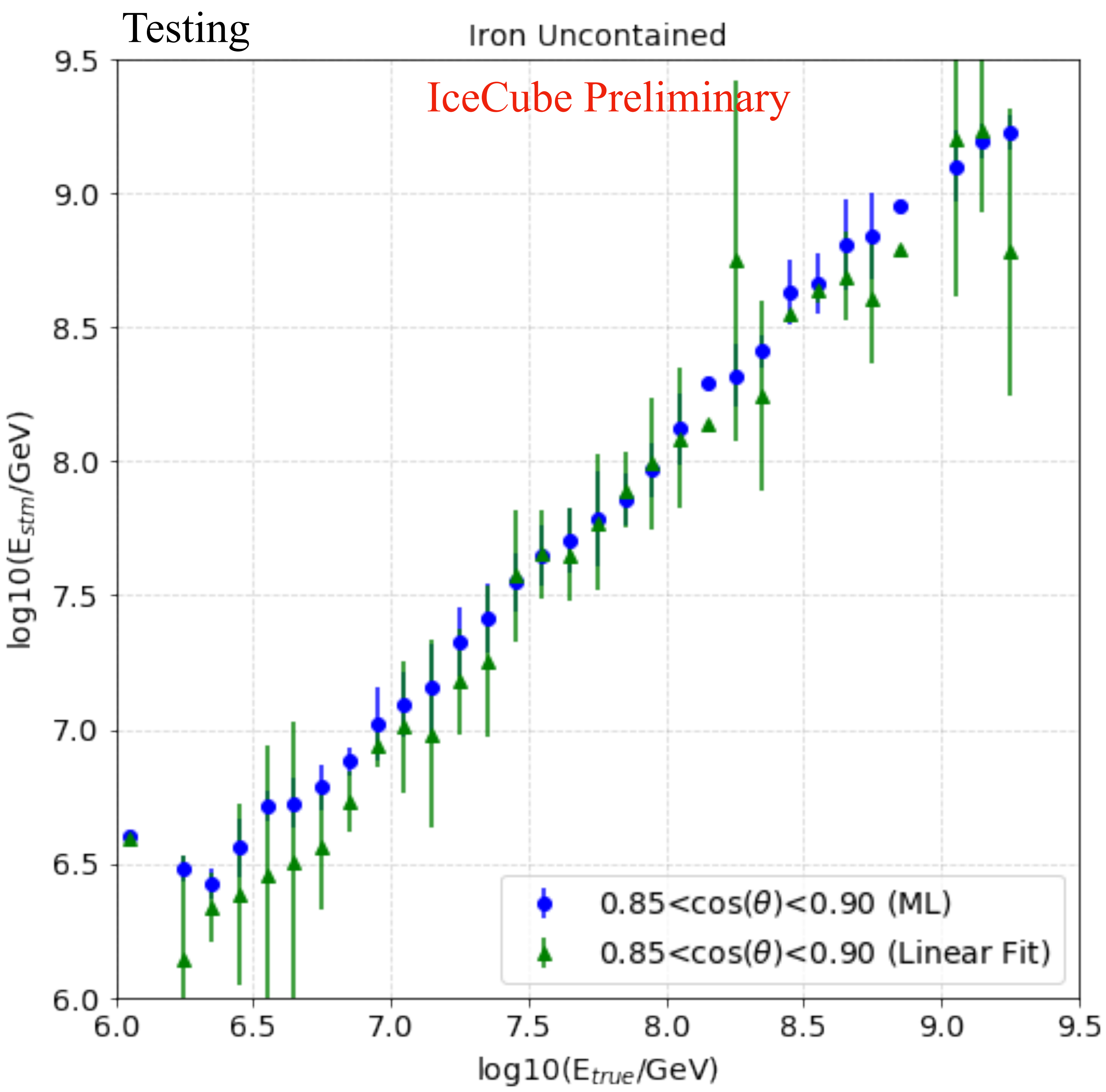
Comparison between Linear Fit and Random Forest Regression performance





Comparison between Linear Fit and Random Forest Regression performance





SOUTH DAKOTA MINES

Summary

- In this work we implemented a combined reconstruction considering contained and uncontained events for proton, iron, helium and oxygen for IceTop and IceCube.

- Two methods for primary energy estimation were implemented. ML performs a comparable results related with contained events. While for uncontained events ML approach offers a better performance.

Backup

Flexible Curvature

Conventional

$$\Delta t(R) = \boxed{C_2} R^2 + C_1(1 - exp(-\frac{R^2}{2\sigma^2}))$$

Fix parameter

New Flexible Curvature

$$\Delta t(R) = \boxed{C_2} R^2 + C_1(1 - exp(-\frac{R^2}{2\sigma^2}))$$

$$(10^{-6}, 10^{-2}) \quad [ns \cdot m^{-2}]$$

New Timing Fluctuation

Conventional

$$\sigma_t(R) = a + bR^2$$

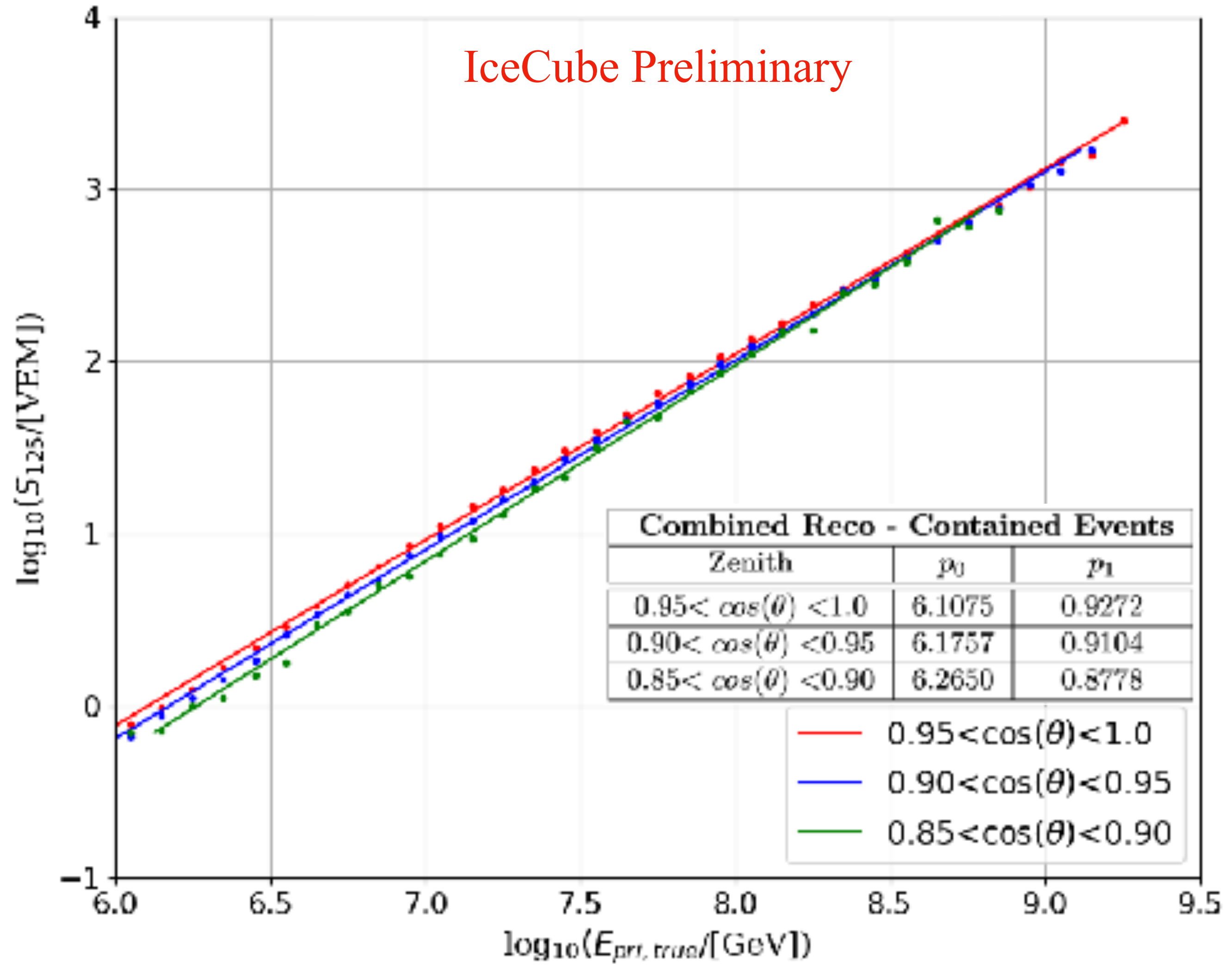
New Timing fluctuation

$$\sigma_{ti} = C\frac{\sqrt{\Sigma_{j=1}^{2}(t_{ij} - (\frac{t_{i1}+t_{i2}}{2}))^2}}{(\Sigma_{j=1}^{2} Q_{ij})^a} + b$$

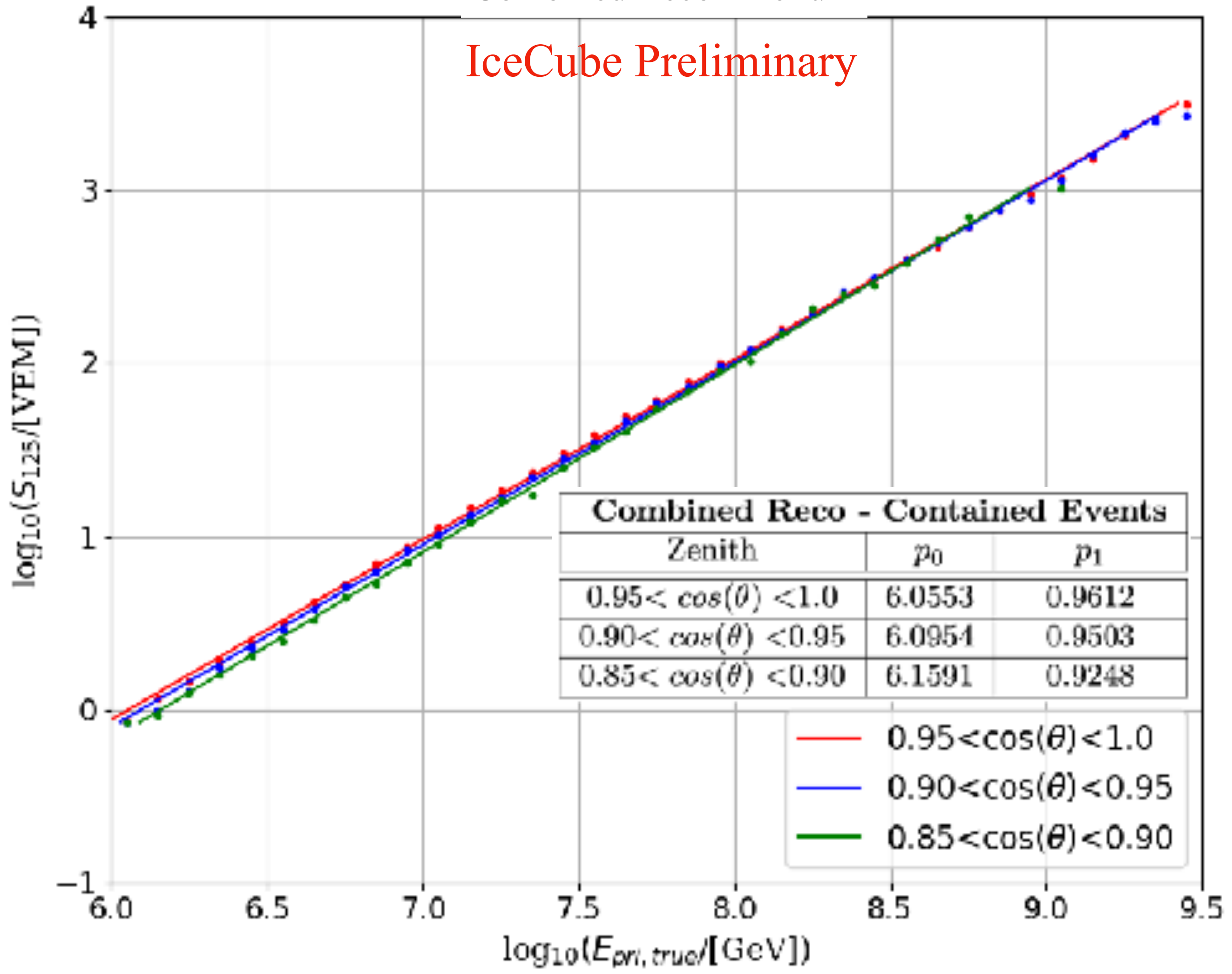
Not.	Value	Units
a	1.00	
b	1.22	ns
C	4	VEM

Energy estimation method using Linear fit

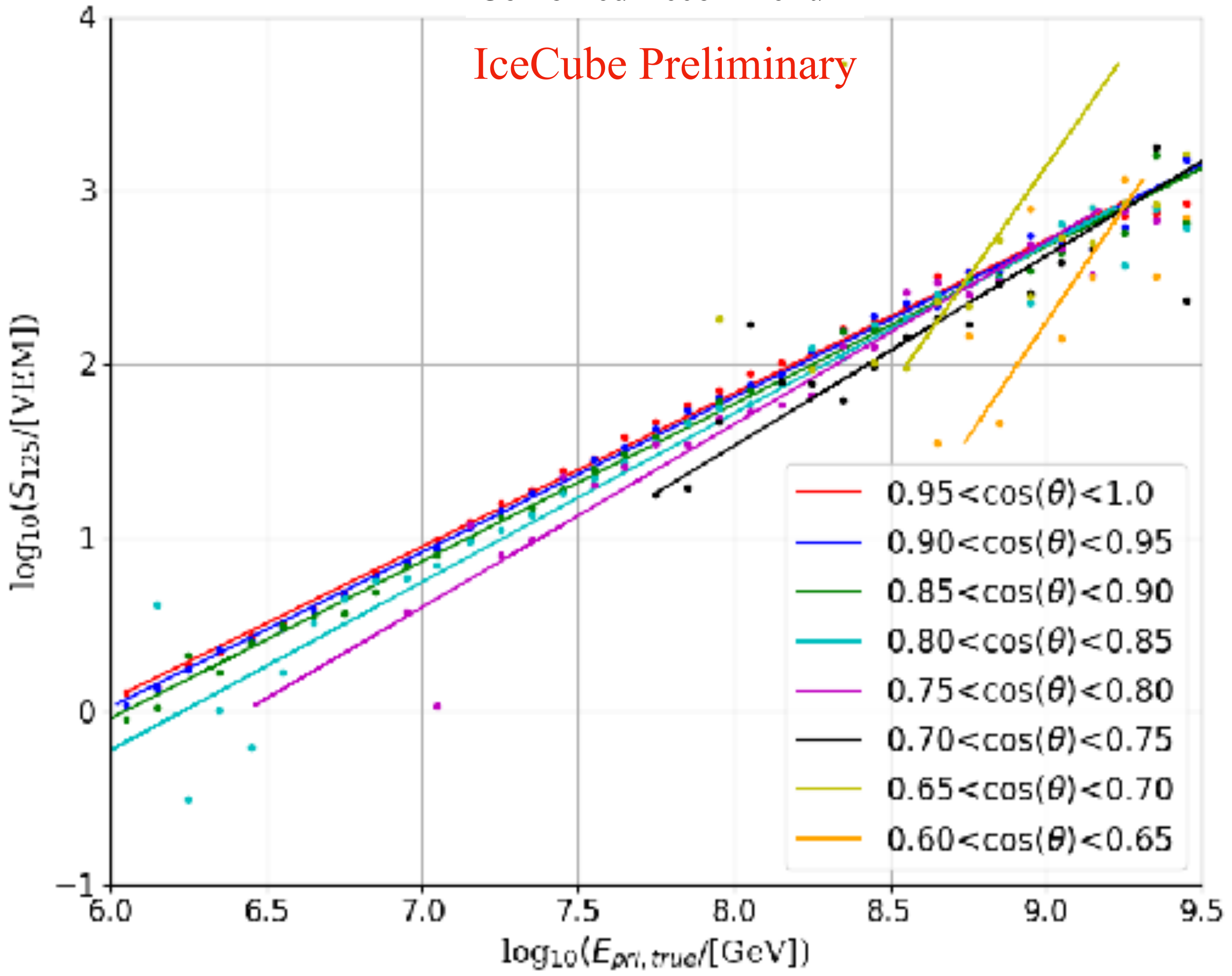
Combined Reco - Iron



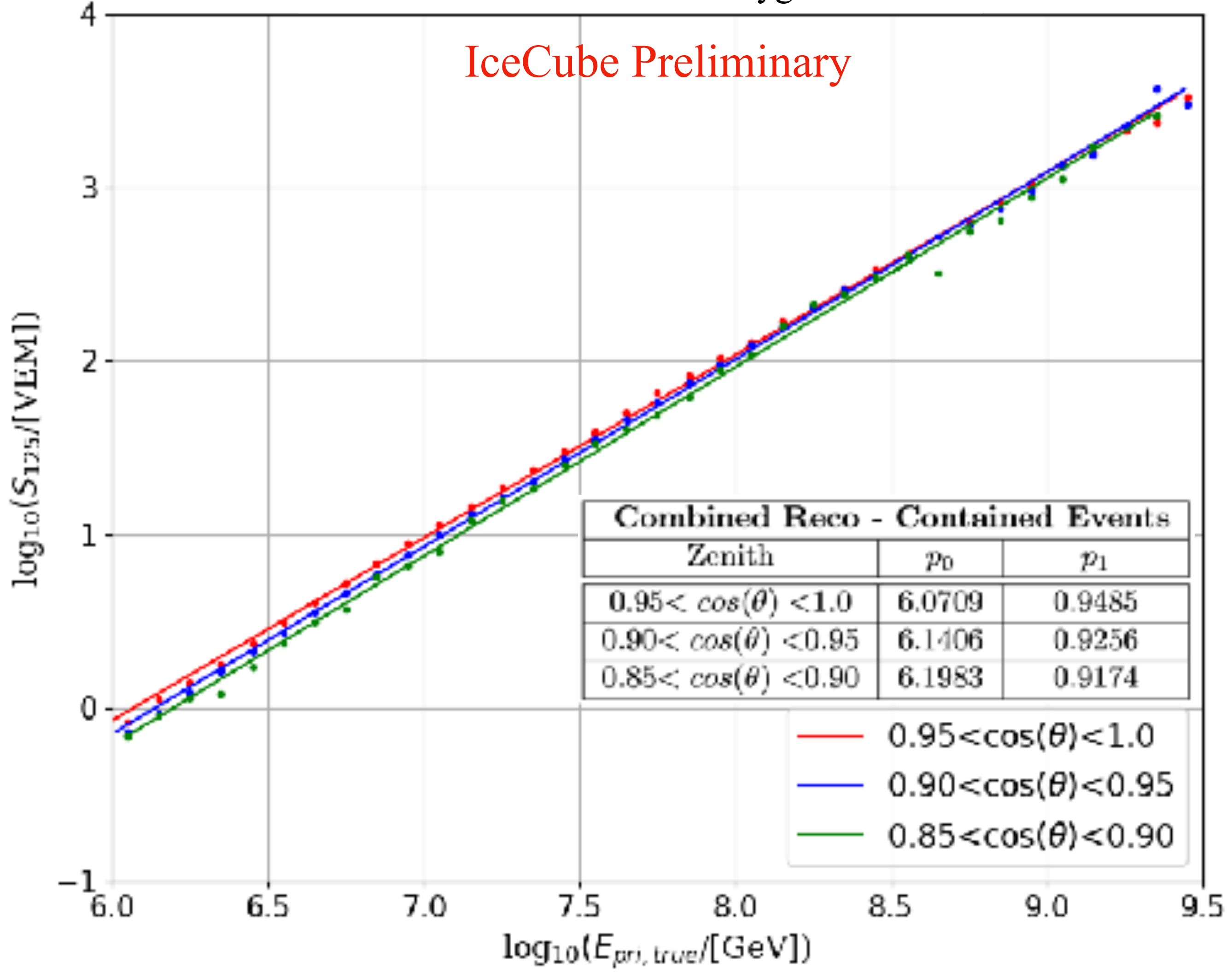
Combined Reco - Helium



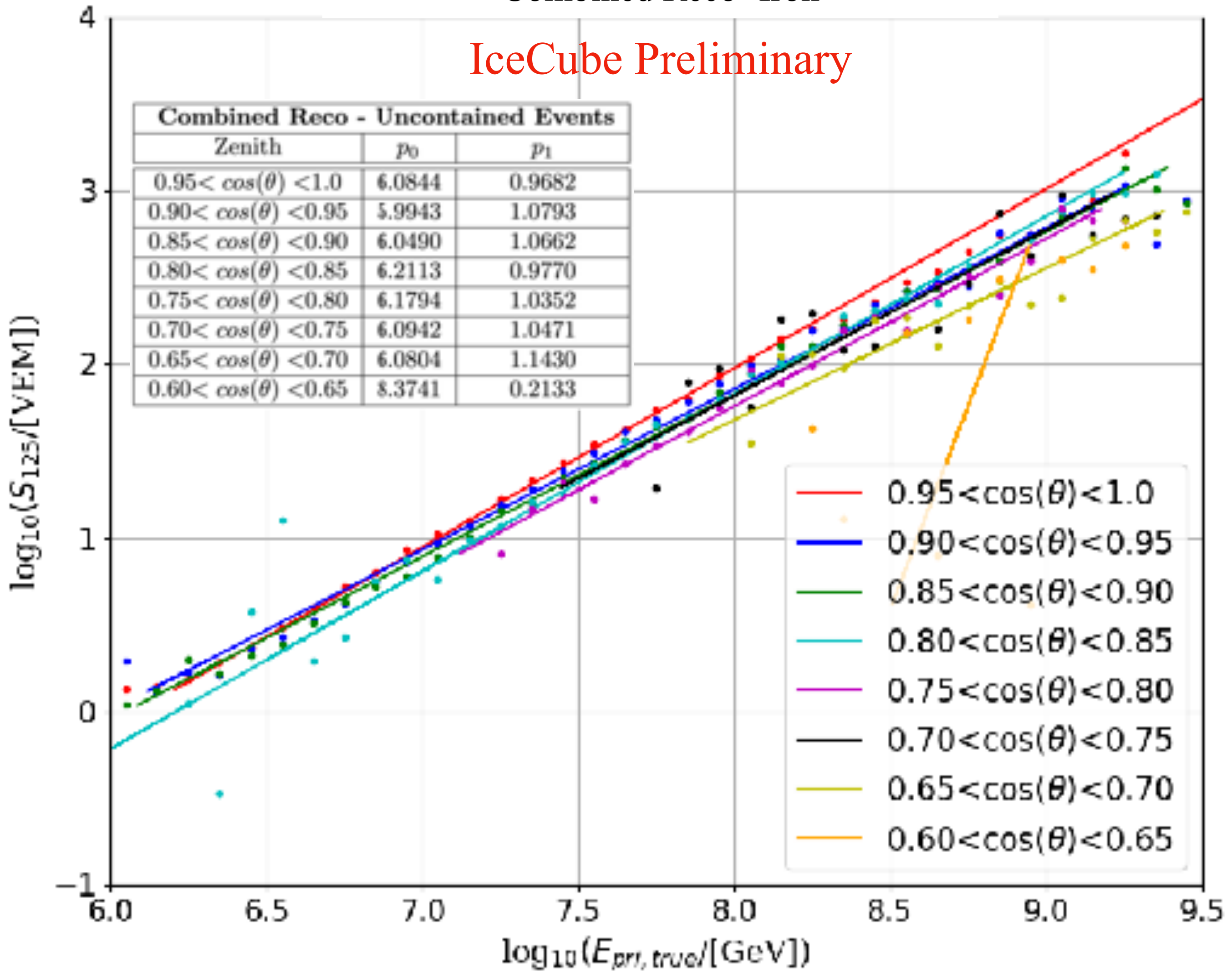
Combined Reco - Helium



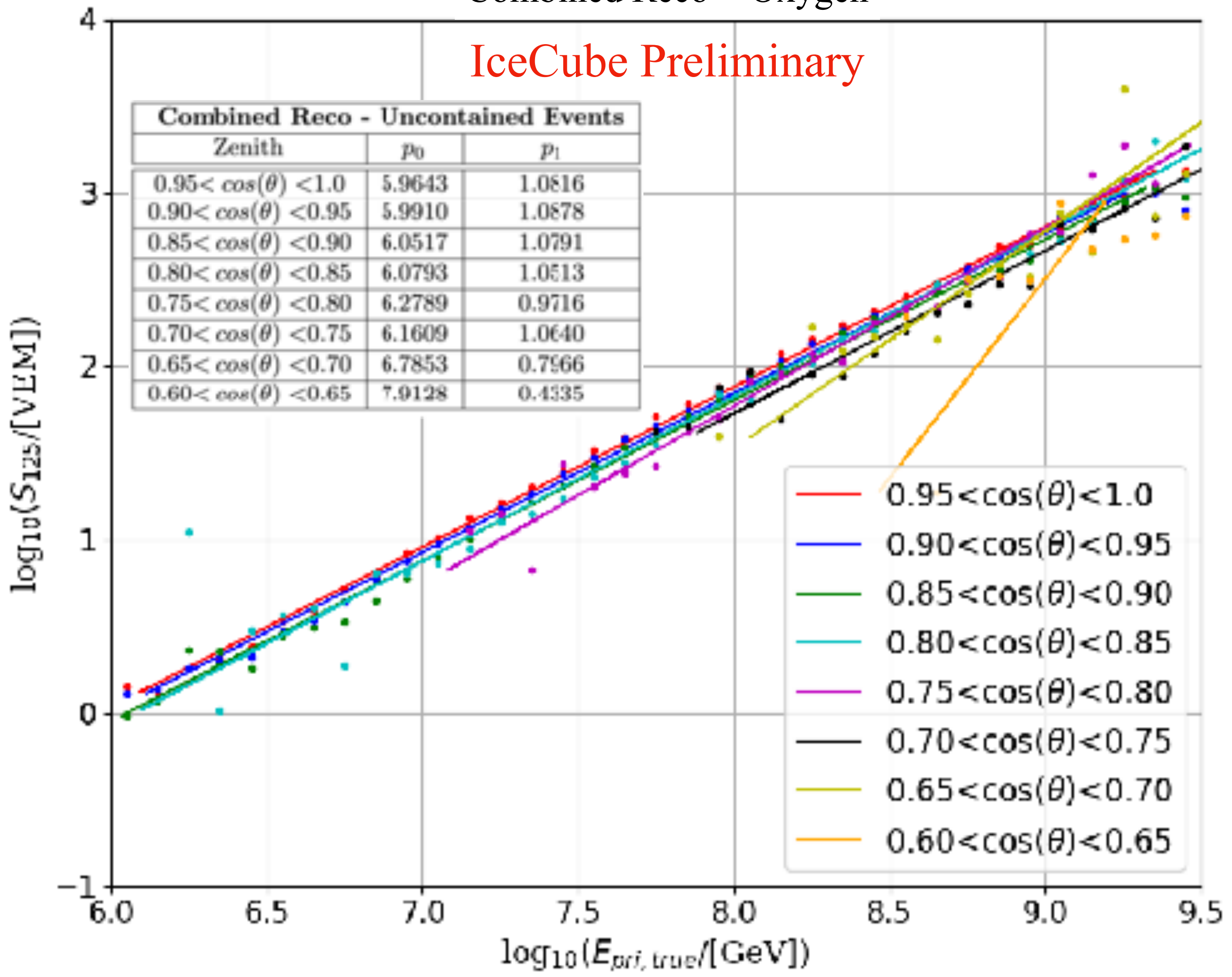
Combined Reco -Oxygen



Combined Reco -Iron

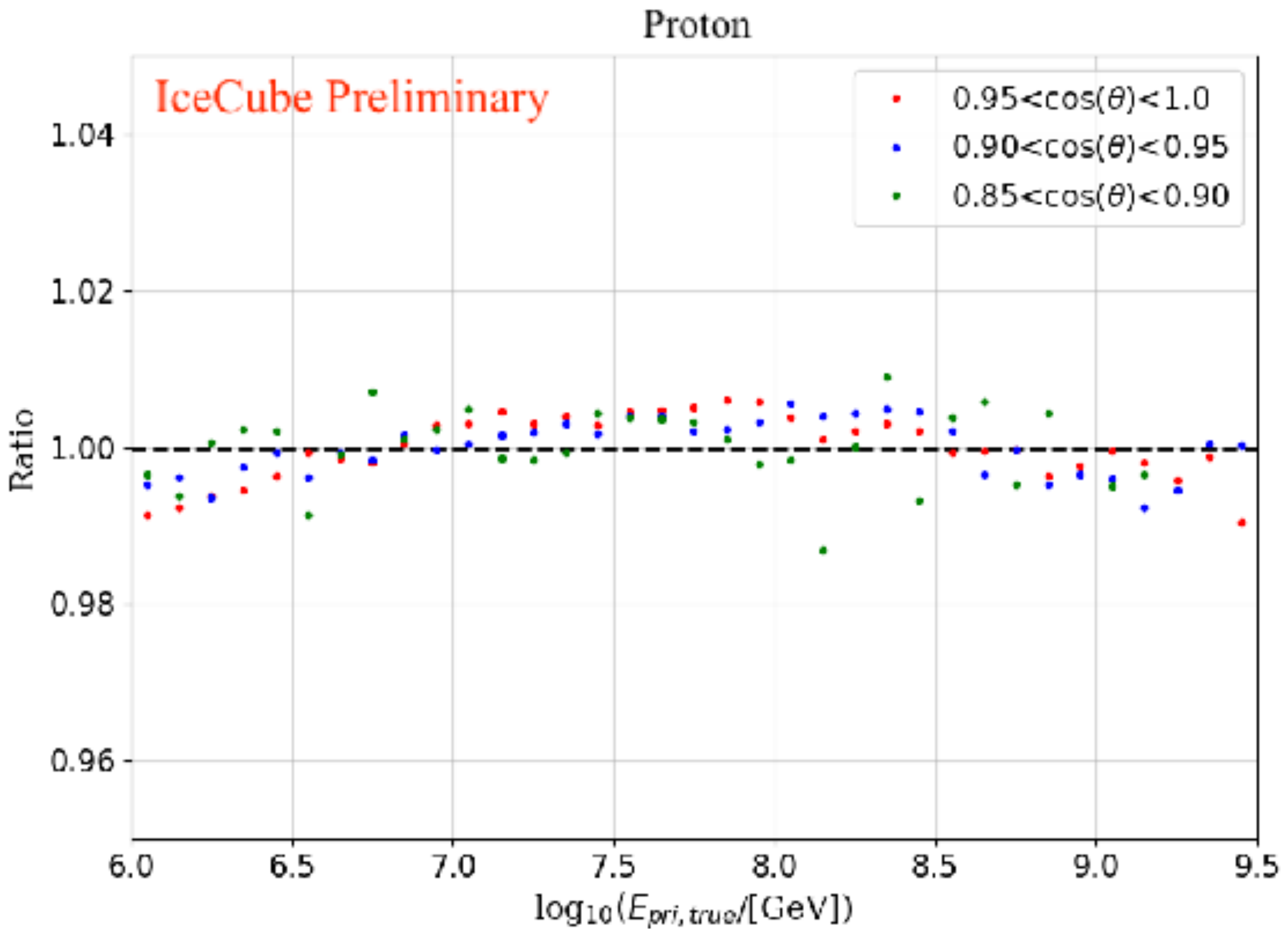


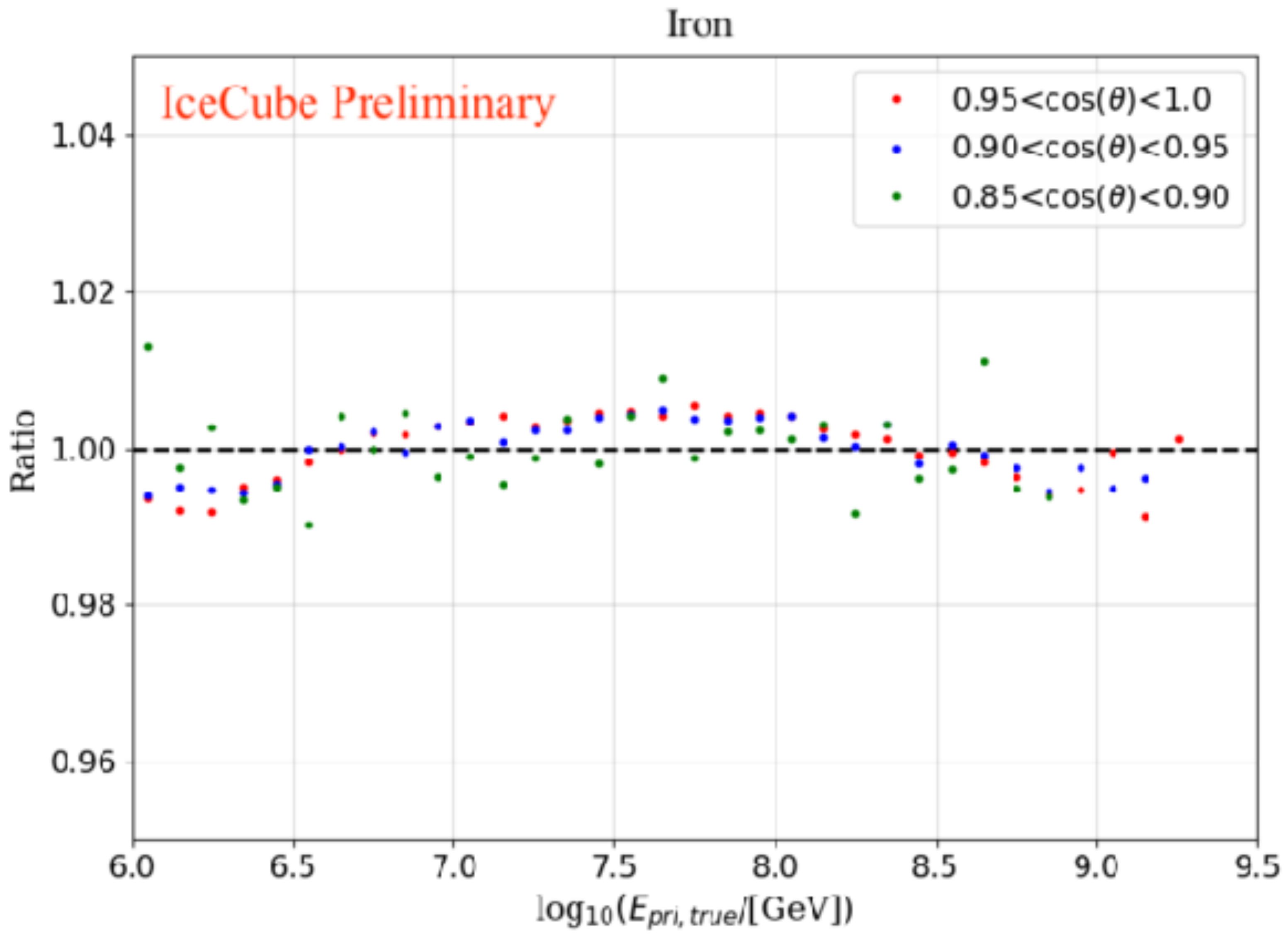
Combined Reco - Oxygen

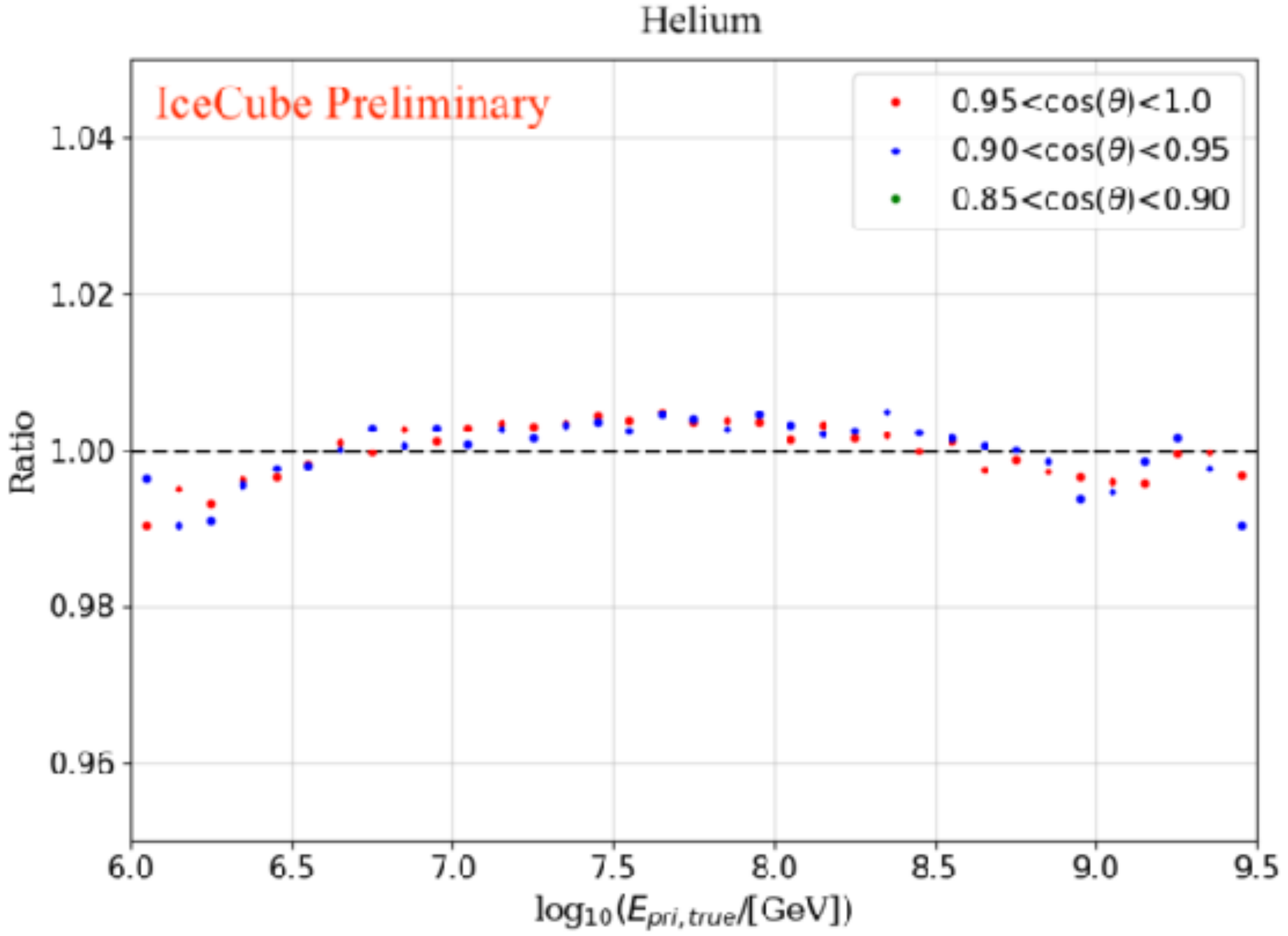


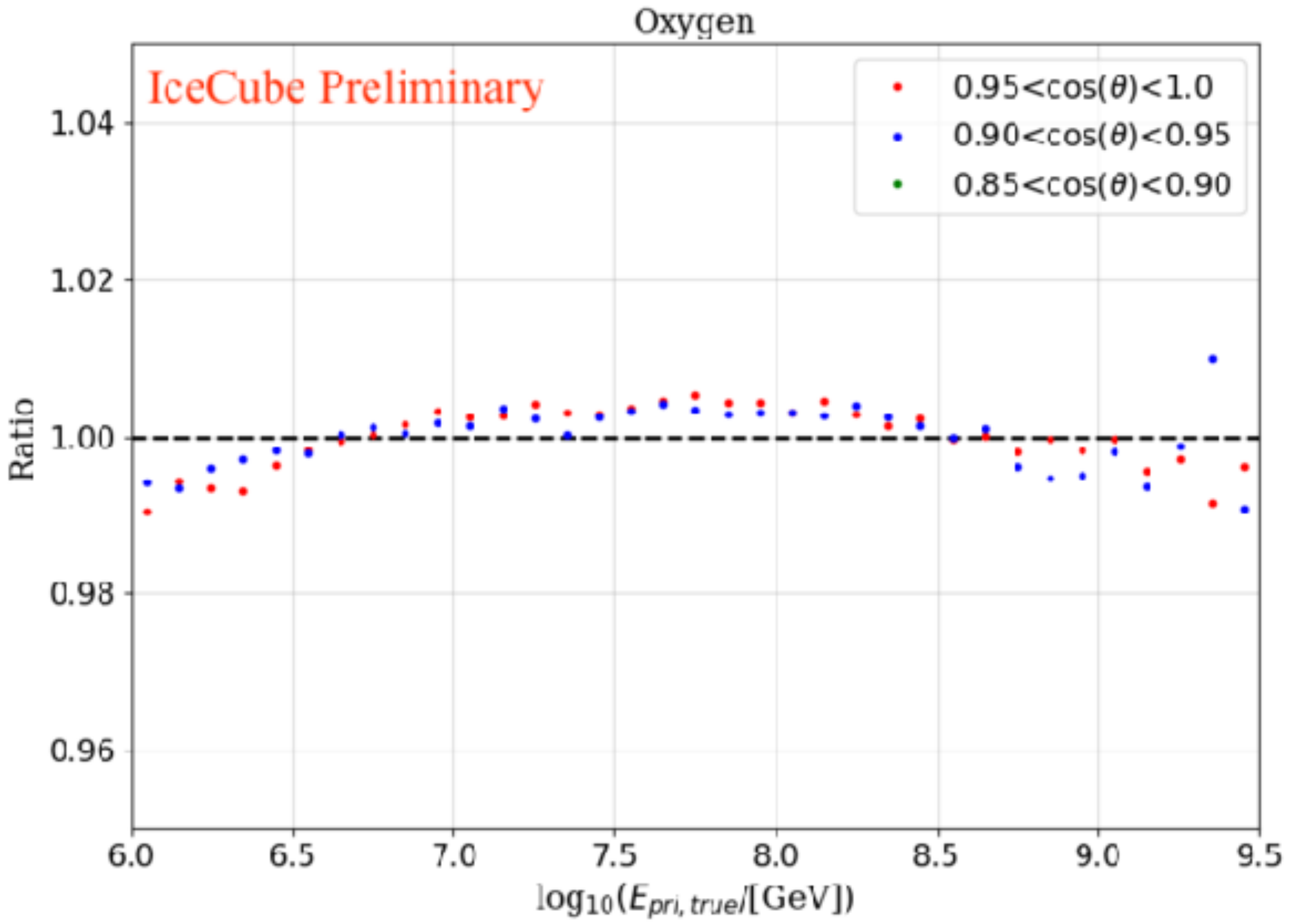
Energy estimation method using Linear fit

Contained Events





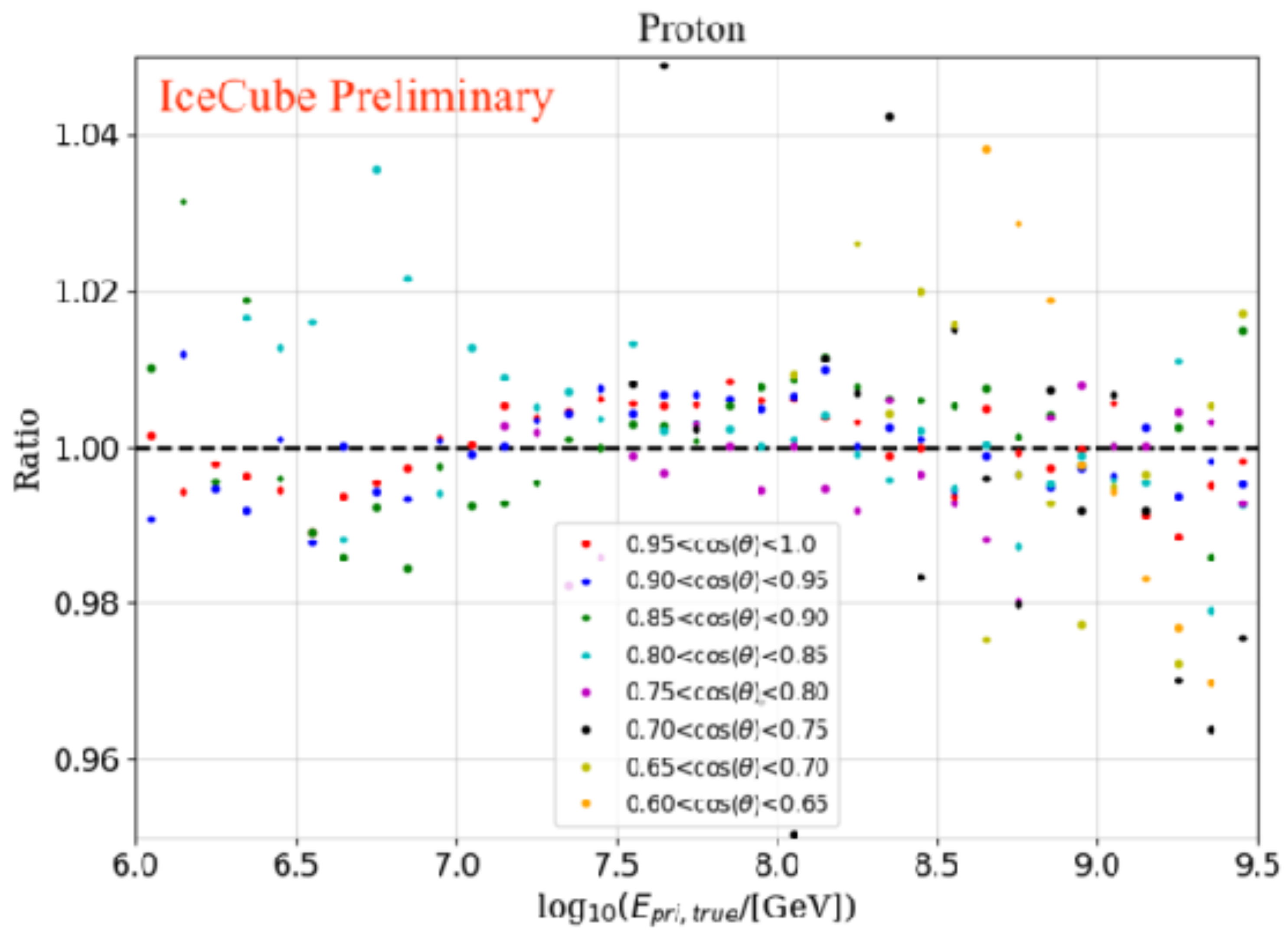


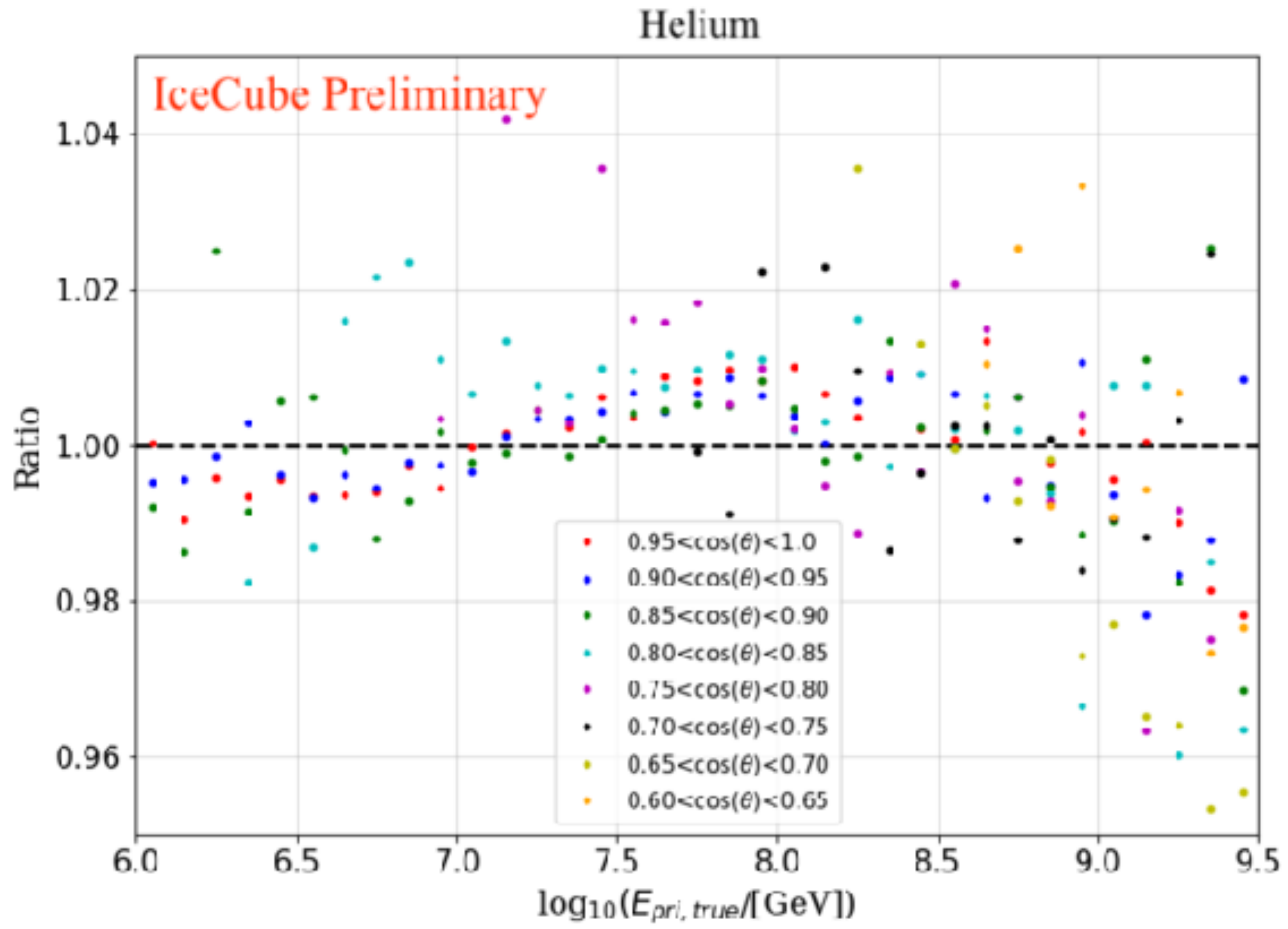


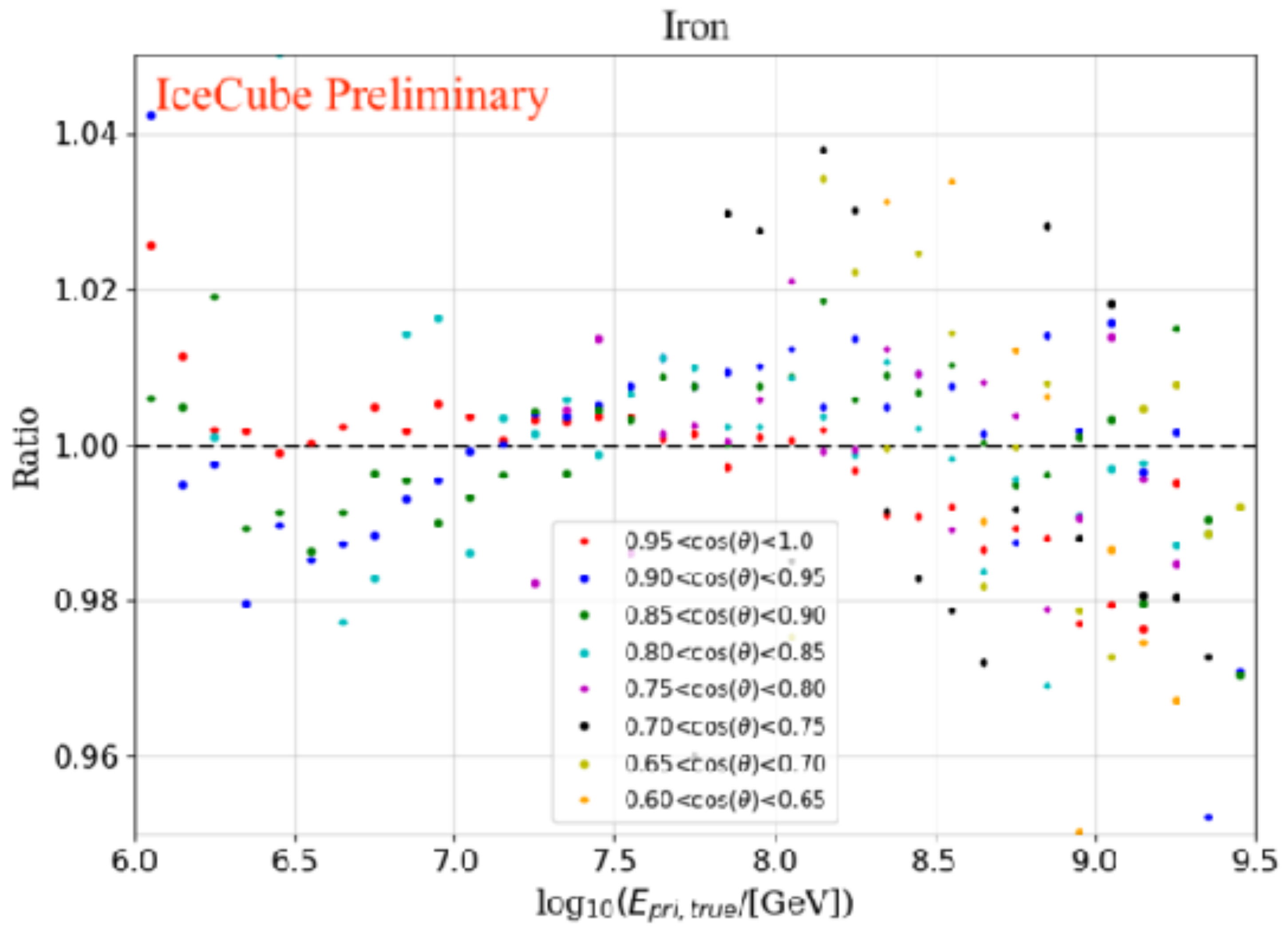
SOUTH DAKOTA MINES

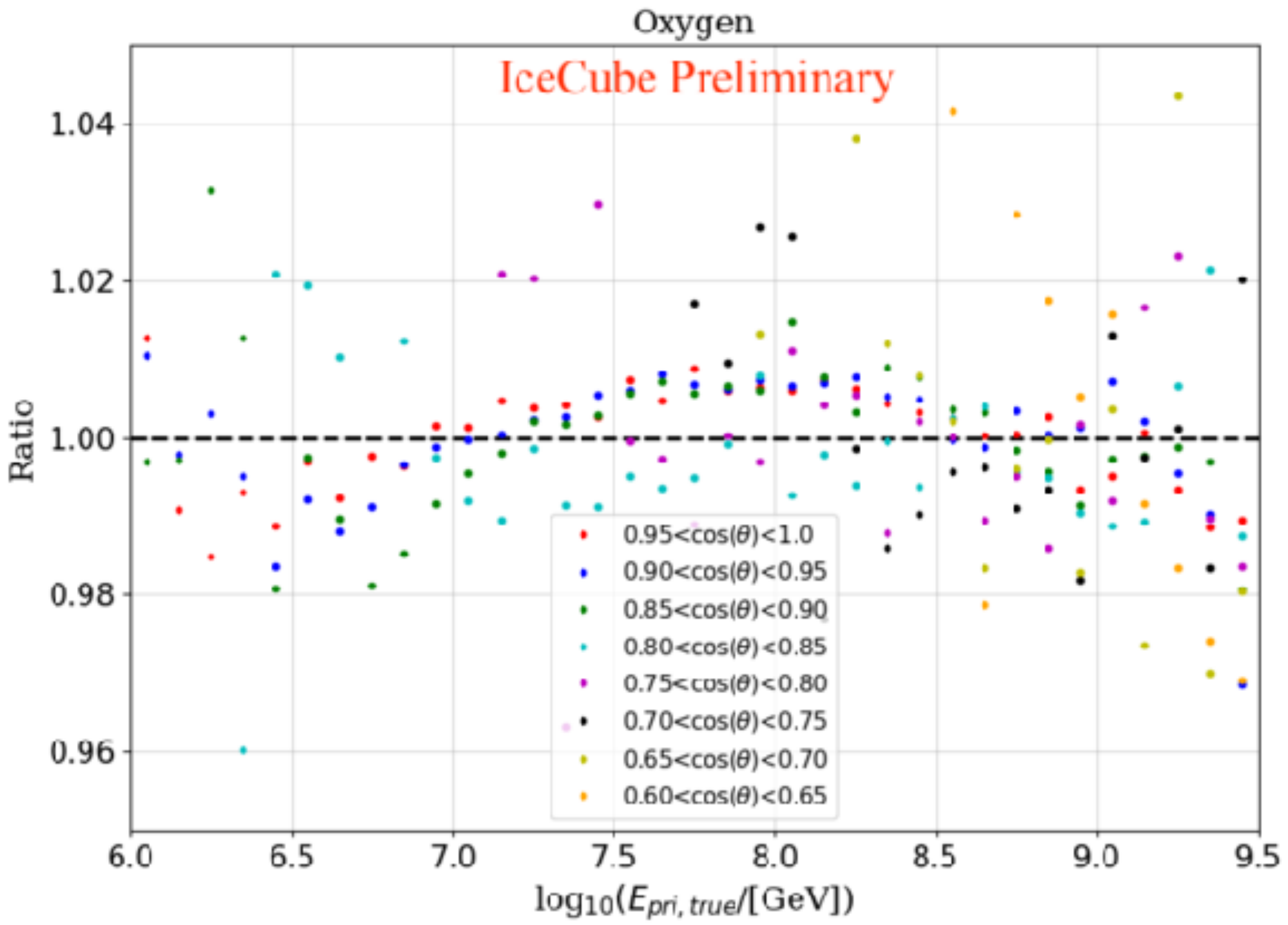
Energy estimation method using Linear fit

Uncontained Events









SOUTH DAKOTA MINES

Energy estimation/prediction using Random Forest Regression