COSMIC RAY MASS COMPOSITION STUDY USING A RANDOM FOREST APPLIED TO DATA FROM THE ICEACT TELESCOPES

Larissa Paul[a], Matthias Plum[e], Karen Andeen[a]

on behalf of the IceCube Collaboration,

with associated members: Thomas Bretz[b], John W. Hewitt[c], Adrian Zink[d]

Workshop on Machine Learning for Cosmic-Ray Air Showers 2022

[a]Marquette University, Milwaukee,WI USA
[b]RWTH Aachen University Aachen, Germany
[c]University of North Florida, Jacksonville,FL USA
[d]Erlangen Centre for Astroparticle Physics, Erlangen, Germany
[e]South Dakota School of Mines & Technology, Rapid City, SD USA

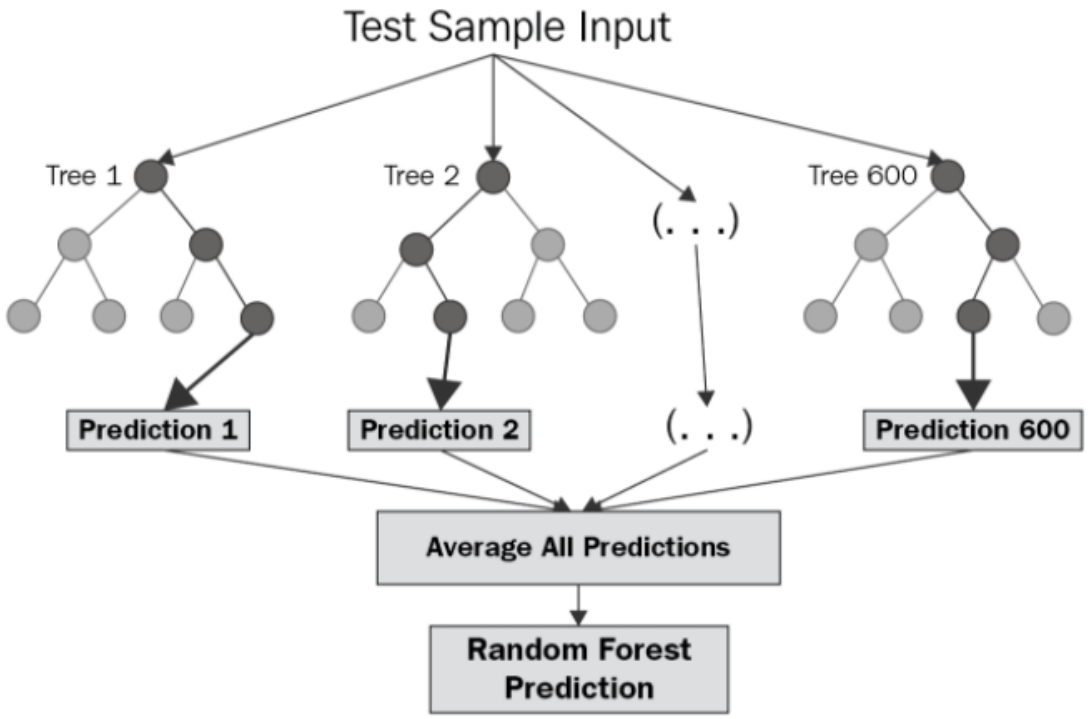
Outline:

- Introduction into Random forest trees
- Cosmic ray composition analysis
- Summary and outlook

Introduction into random forest regression

- Simple Random Forest Regressor Analysis
- Using python package :Sklearn
- Random forest consist of a sample of decision trees
- Each decision tree divides its data until a parameter can describe the remaining sample
- There is no connection between the trees
- The random forest result is been produced by the average of the single decision trees

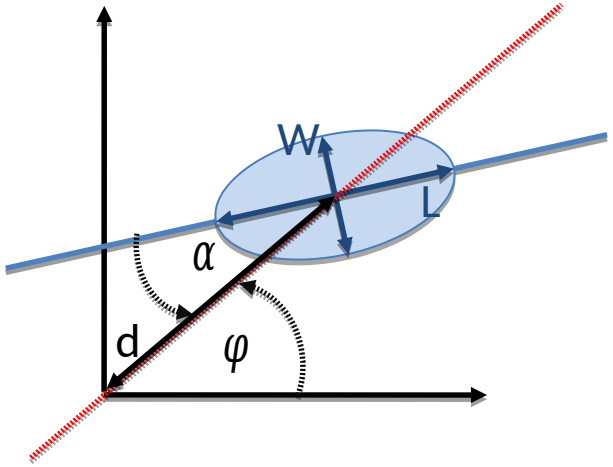
  => Advantage: errors of a single tree are negligible



[1] https://levelup.gitconnected.com/random-forest-regression-209c0f354c84

Composition sensitivity of Hillas parameter

- Image size ~ number of PE
- Width & length ellipse parameter

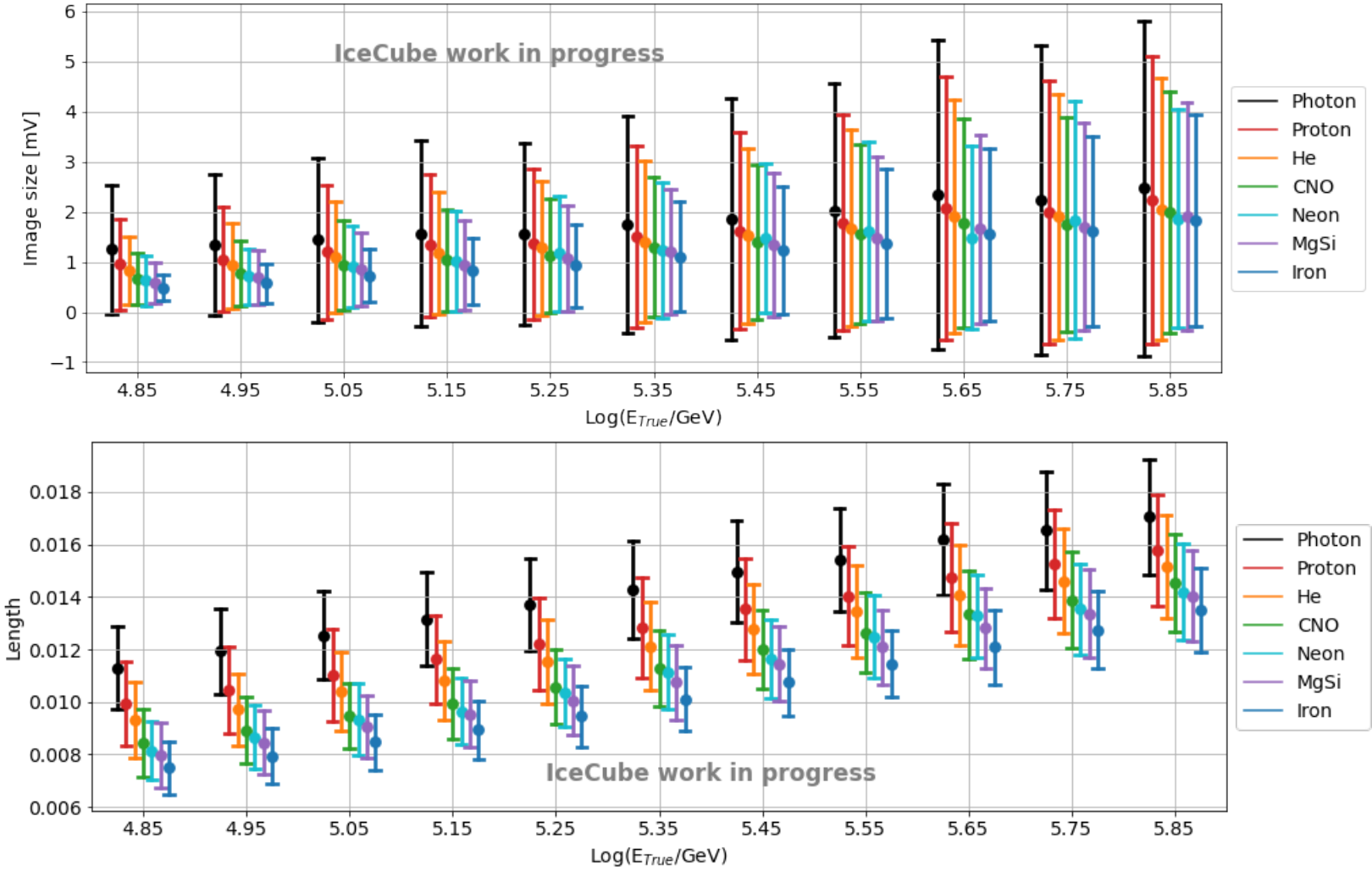


Other composition sensitive parameters:

- distance between the shower core and telescope

  =>MC truth smeared by 50m
- cosine(zenith)

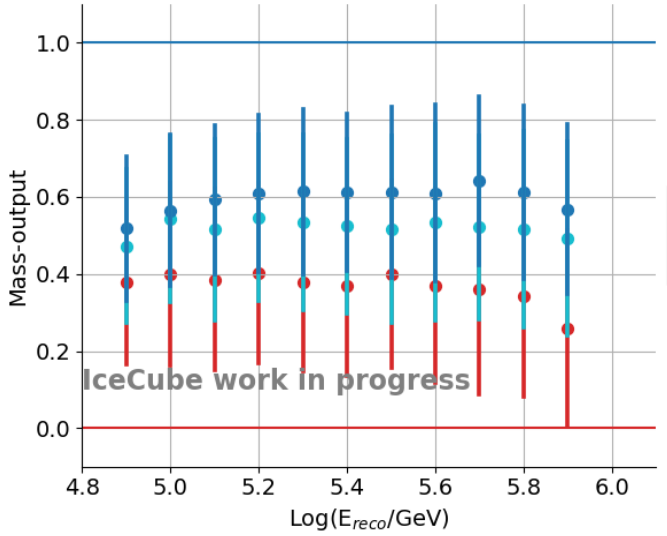
  => MC truth smeared 1degree

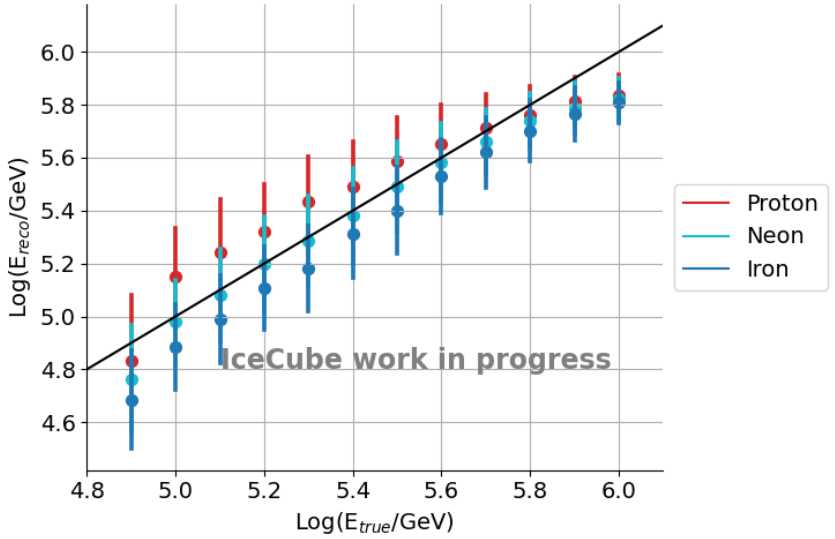


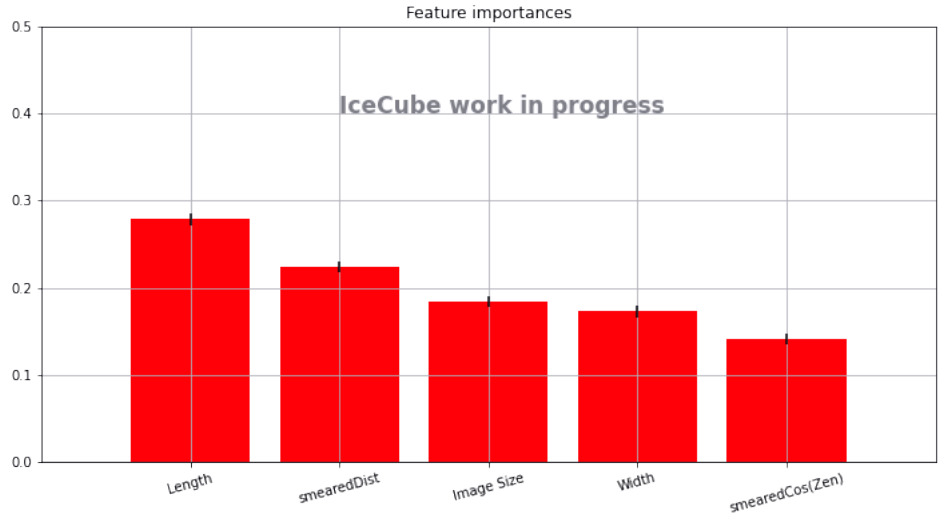
Composition analysis: IceAct parameters

- Simple Random Forest Regressor Analysis => reconstruct energy and mass output
- Events induced by proton and iron nuclei are used for the training and validation of the Random Forest Tree. The result is tested with events induced by proton, neon and iron nuclei.
- The using just the Hillas parameters and simple geometric reconstructions as input variables enables the separation of different nuclei.

IceAct + geometry input parameters used:

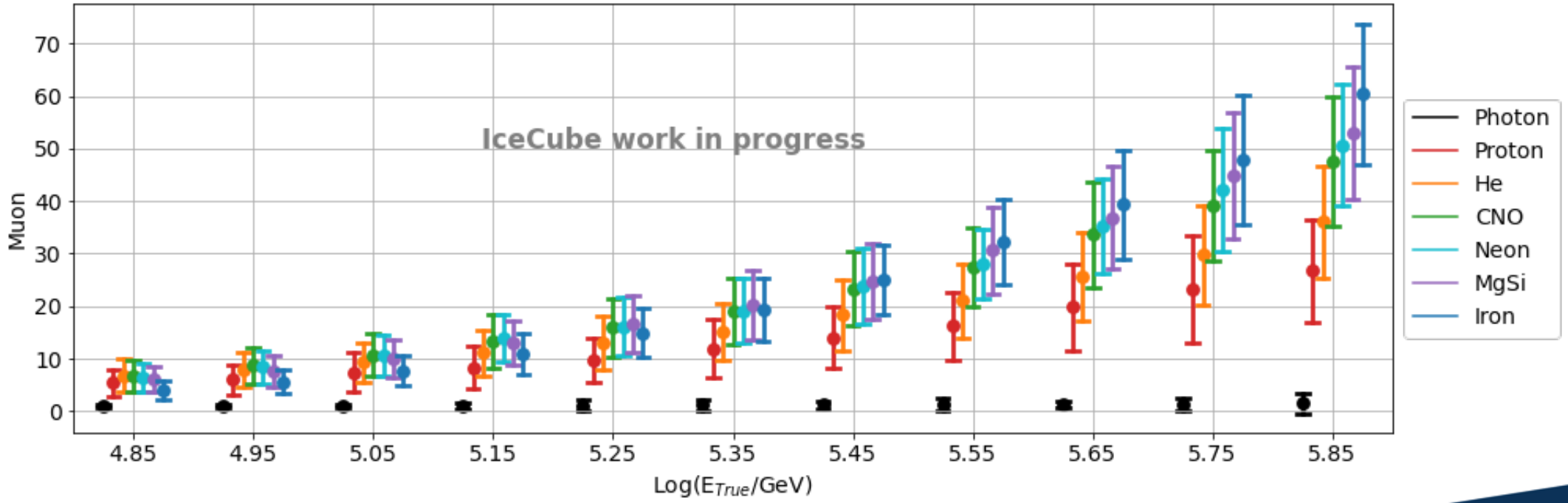






Composition sensitive IceCube parameter:

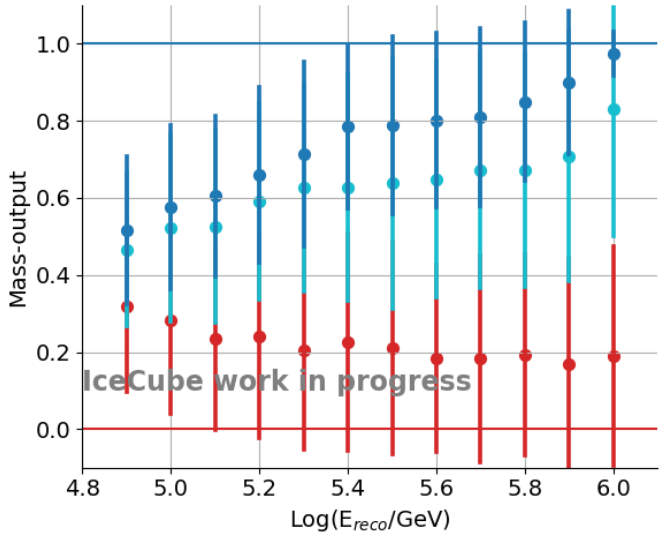
- Number of muons in the inice detector derived from the MC truth of CORSIKA showers.

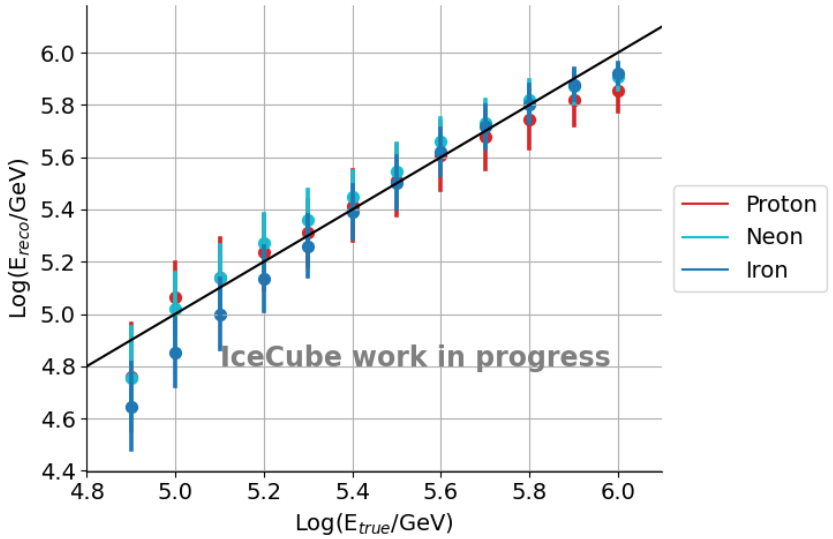


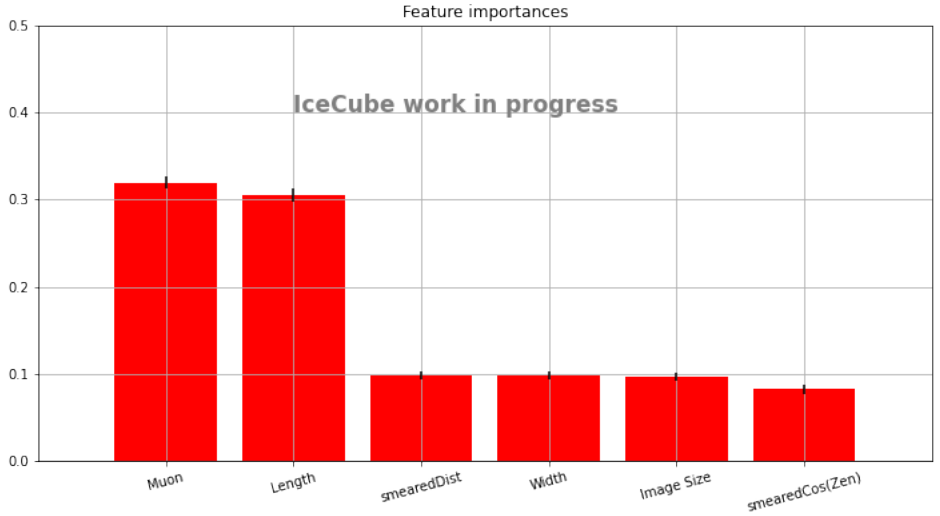
Composition analysis: IceAct and IceCube

- Simple Random Forest Regressor analysis => reconstruct energy and mass output
- The number of inice muons and the length of the Hillas ellipse are by the most important parameters in the Random Forest analysis.

IceAct + IceCube input parameters used:



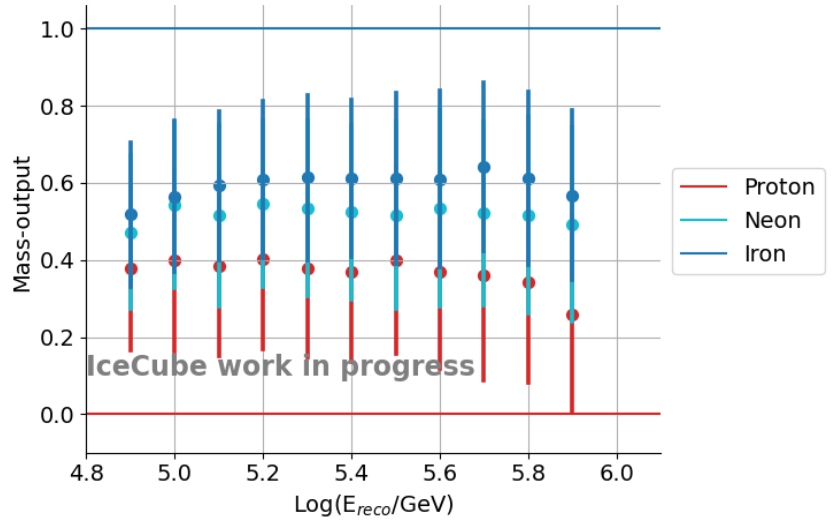


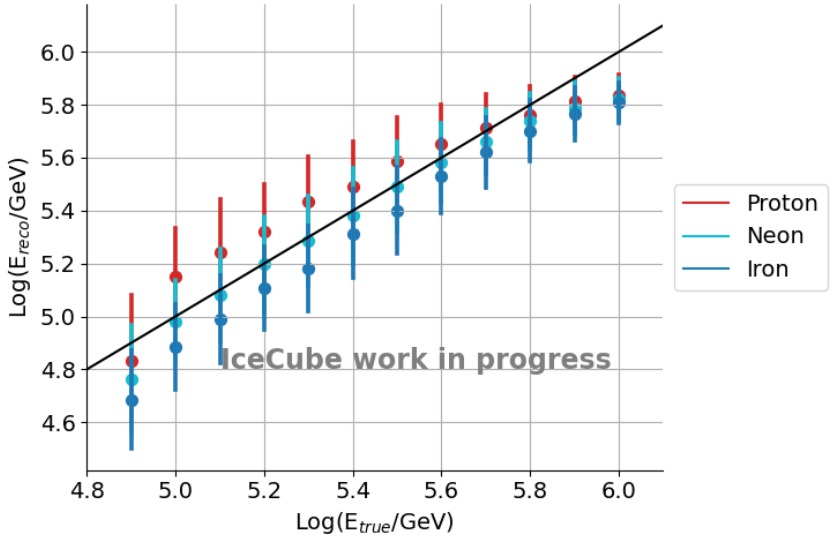


Comparison of the two different input parameter sets

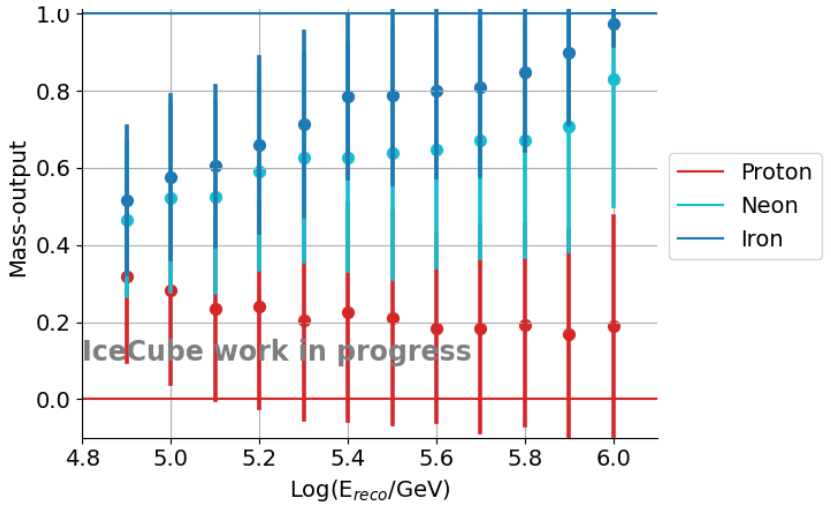
- Adding the number of inice muons separates the mass output value clearer.

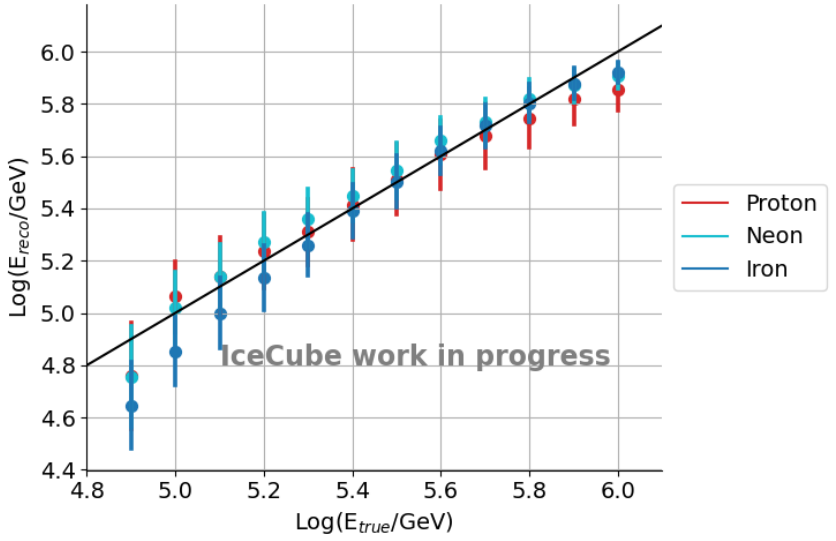
IceAct + Geometry





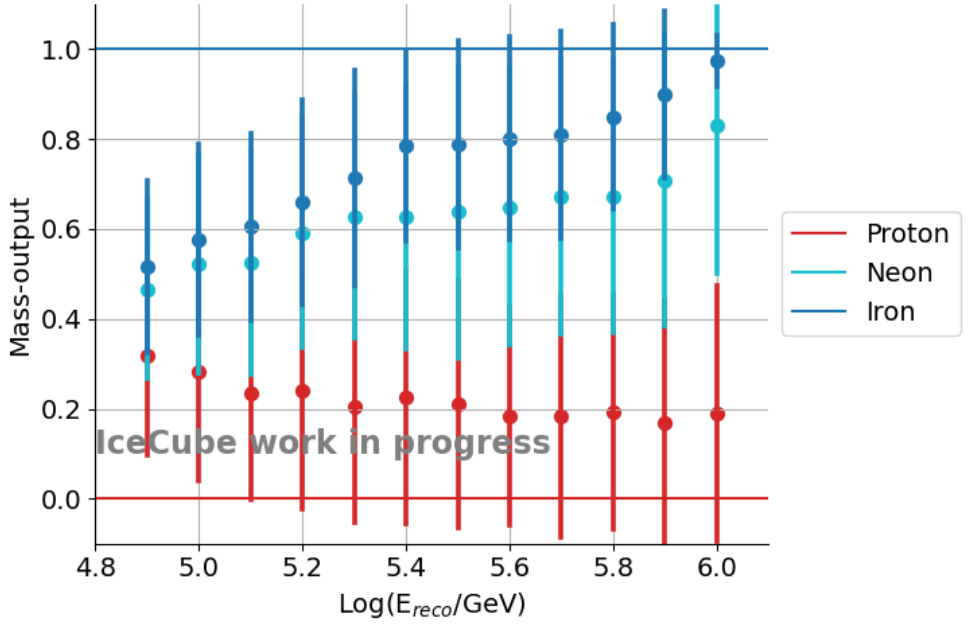
IceAct + IceCube

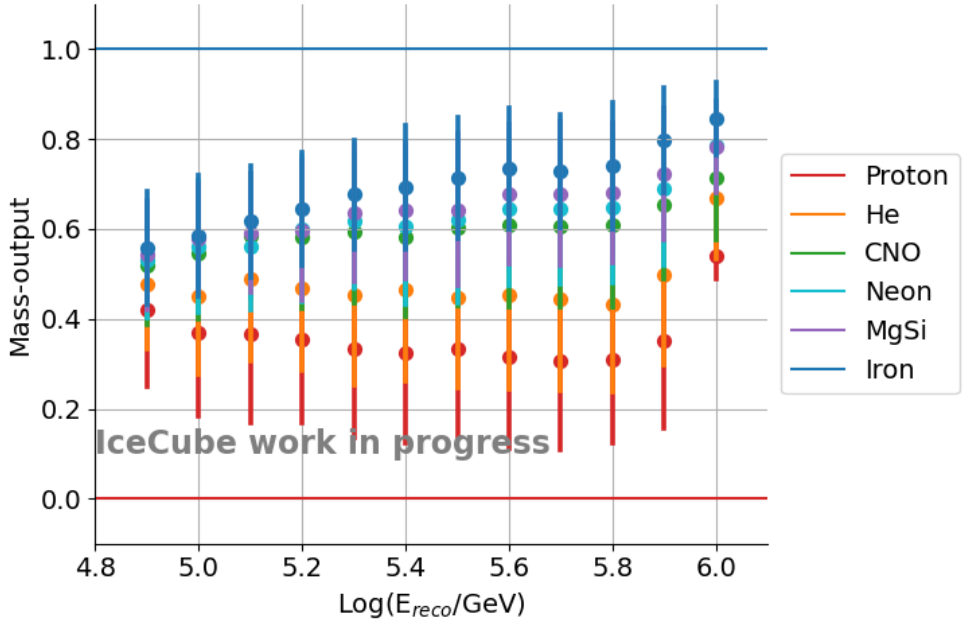




Comparison of between different training data sets

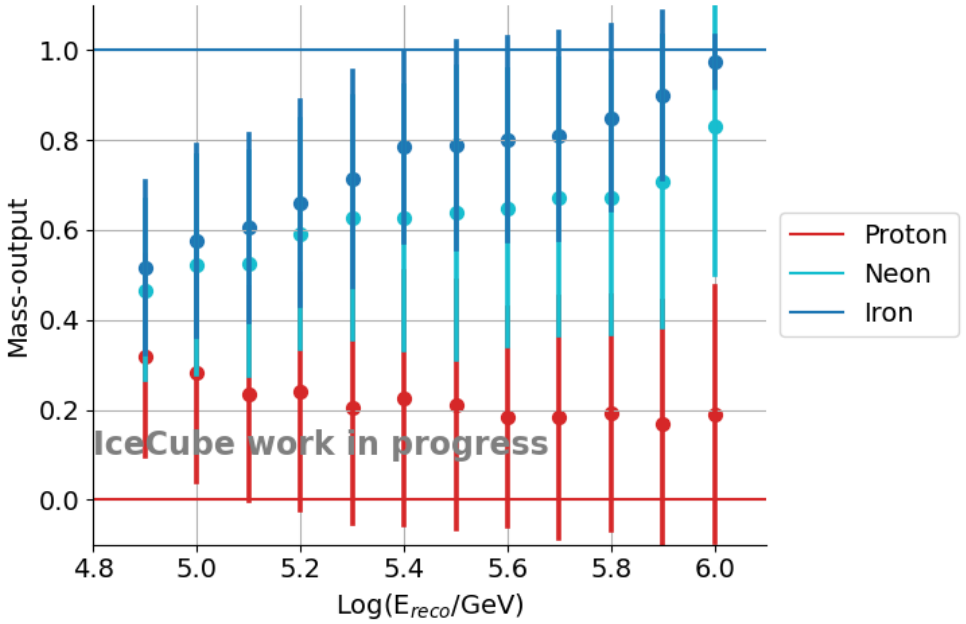
- Exchanging the simple proton and iron induced event data set with a more realistic data set using proton, helium, oxygen, aluminum and iron induced events for training and validation, and adding neon induced events for testing results in a slightly worse mass output separation, but with smaller standard deviations.
- The highest energy bin is not very trustworthy because of missing higher energy MC.
- Room for improvement by adding input parameters and tuning modeling parameters.

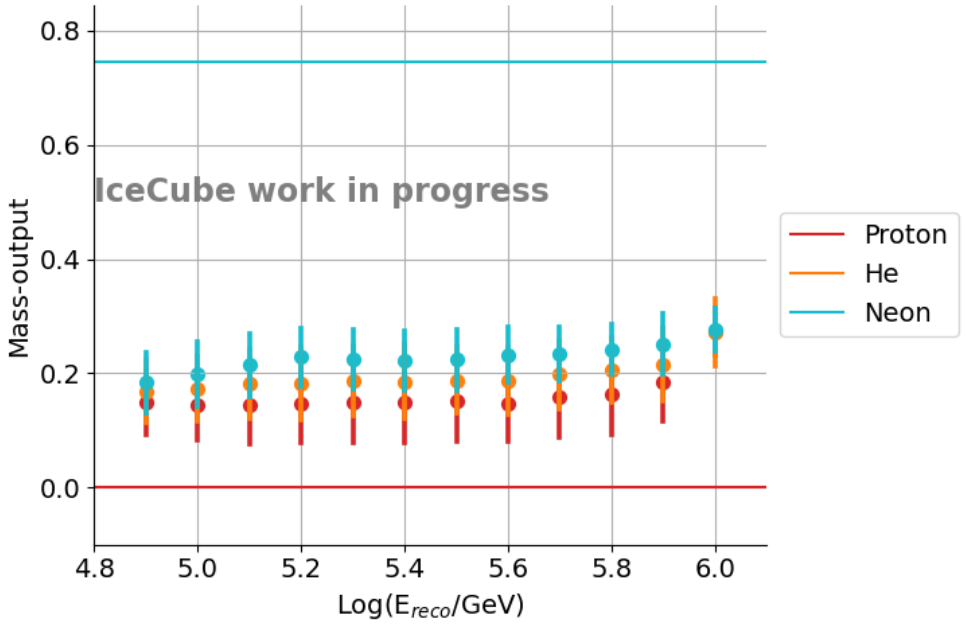




Comparison of between different training data sets

- Exchanging the iron nuclei induced events with helium induced events, does not show a very good separation, but testing the results with neon induced events shows that on average the neon events are reconstructed heavier than the helium events.
- The highest energy bin is not very trustworthy because of missing higher energy MC.
- Room for improvement by adding input parameters and tuning modeling parameters.

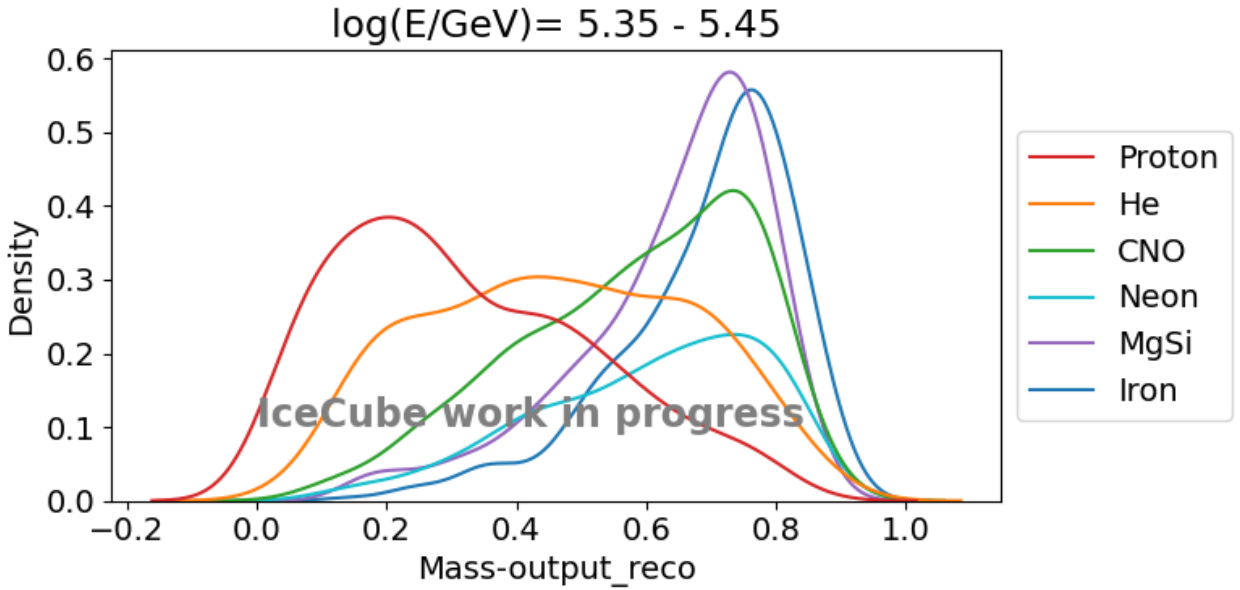




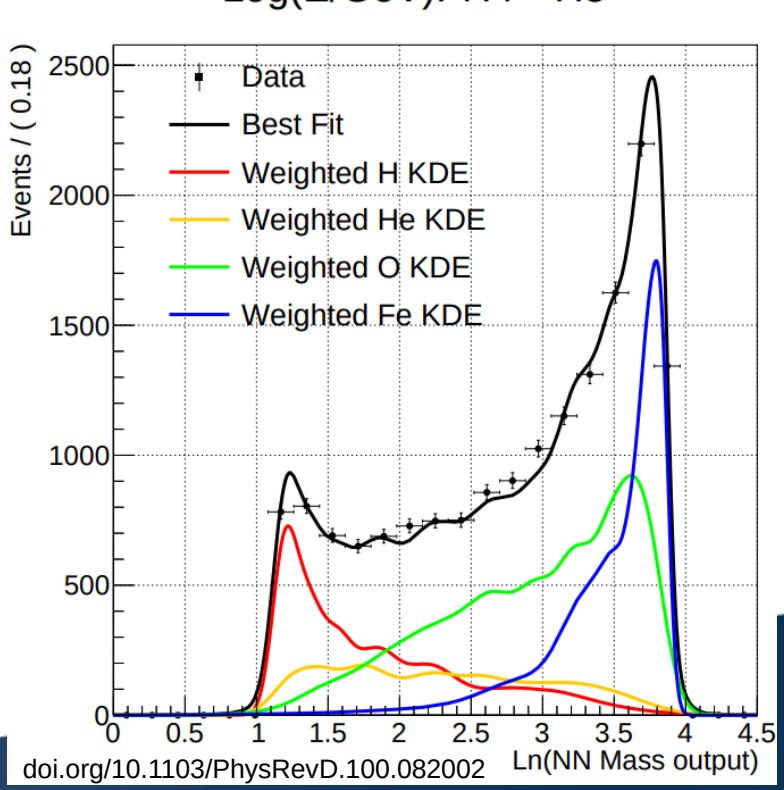
Template analysis

- Use the resulting distributions for the mass output in each energy bin and weighted them to reproduce the reconstructed data.

Single bin for the mass output of the IceAct and IceCube analysis

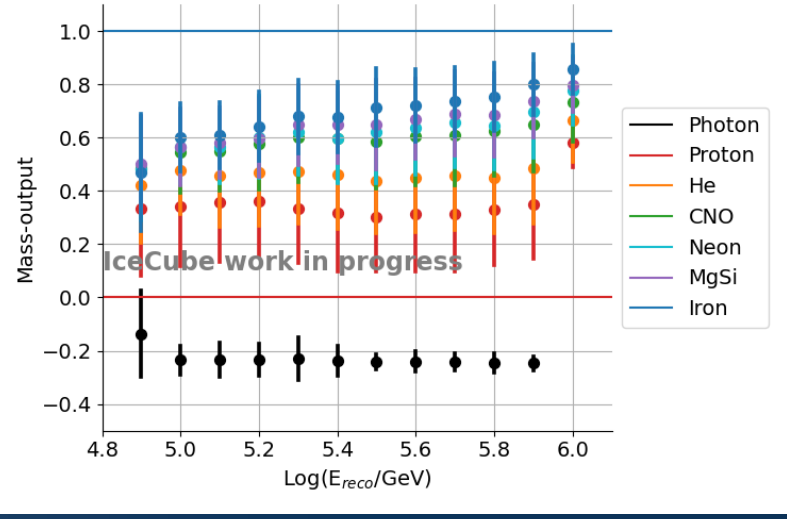


Single bin for the mass output of the 3year IceCube and IceTop composition analysis



Summary and outlook

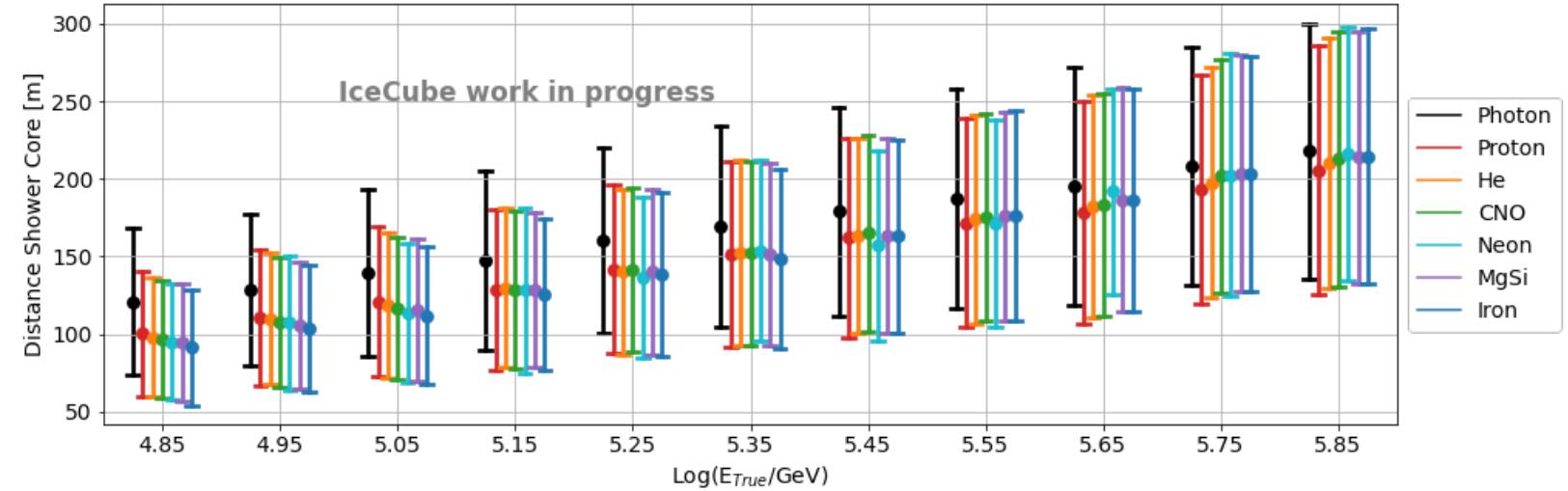
- A simple random forest tree analysis using Hillas parameter for reconstruction of the el.mag air shower component shows promising results.
- To Do:
  - Exchange from smeared MC truth input values to reconstructed values.
  - Test normalization of the input or/and output parameters
  - Find reconstructed IceCube and IceTop variables which improve the mass output and energy reconstruction.
  - Increase the MC statistic and the energy range of the MC.
- A first look at photon-nuclei separation is promising:



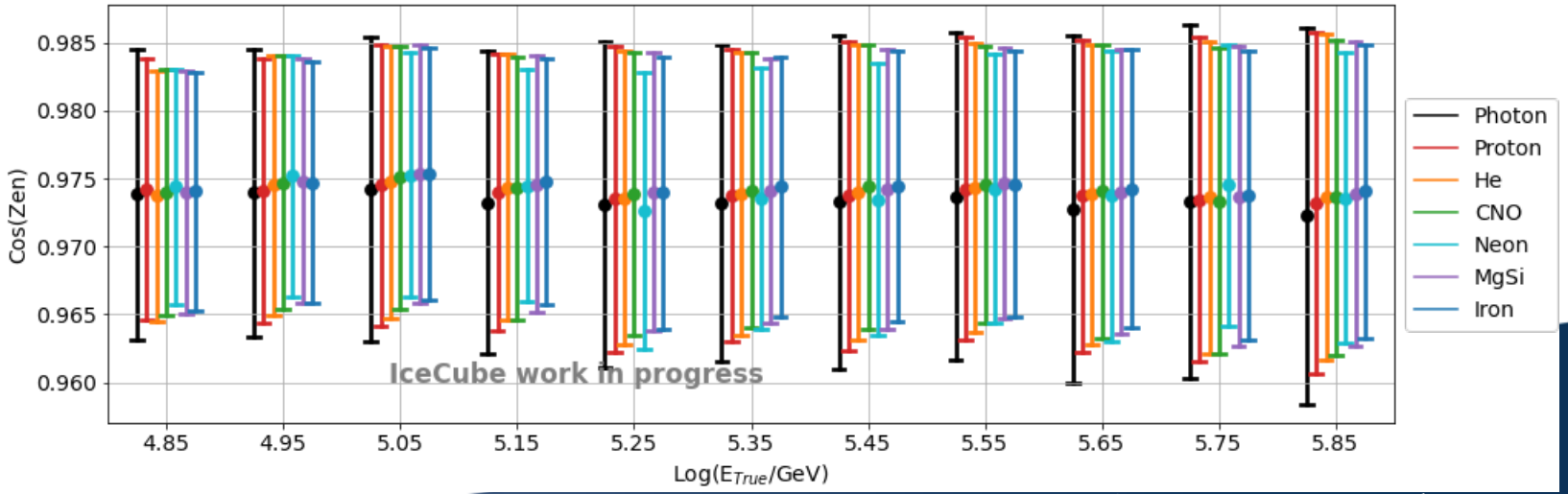
BACKUP

Other composition sensitivity of parameter

- distance between the shower core and telescope

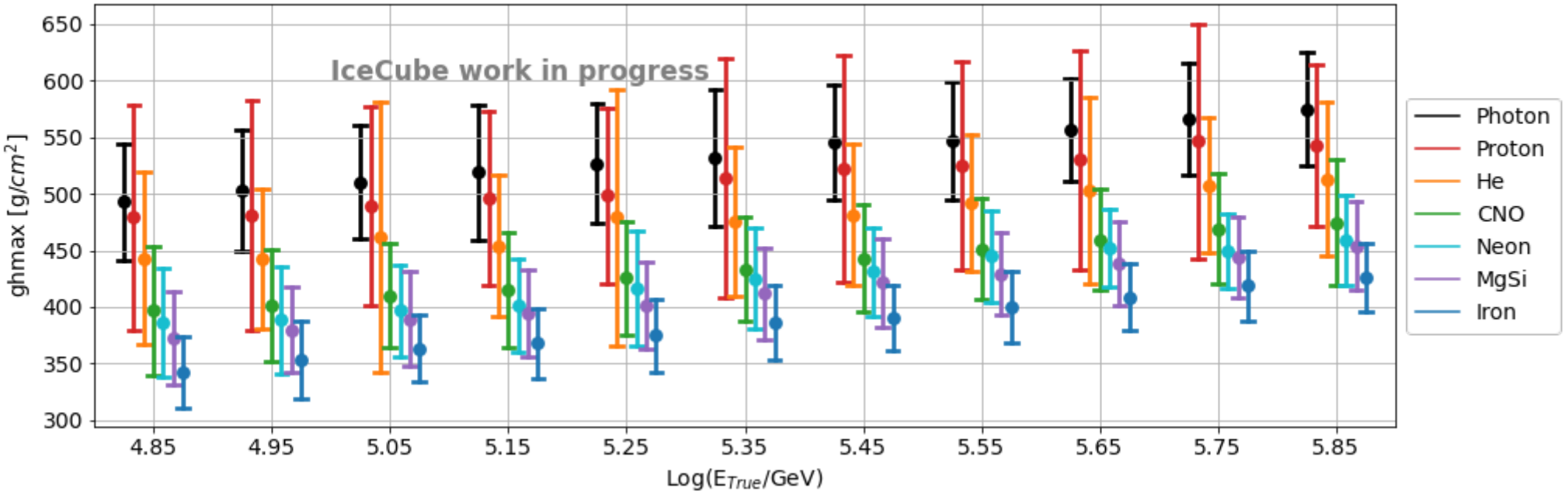


- cosinus(zenith)



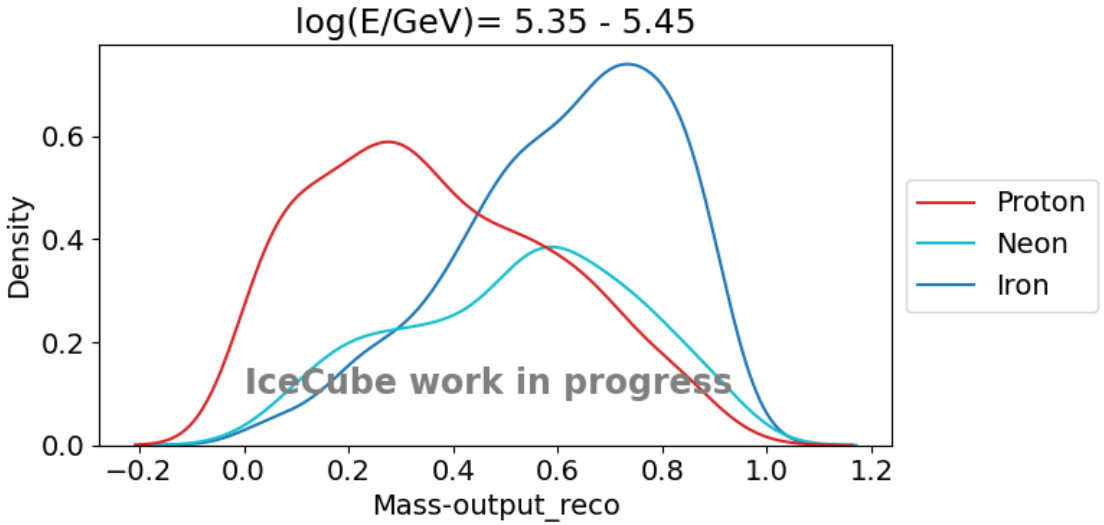
Composition Sensitive IceAct parameter:

Height of the shower maximum:

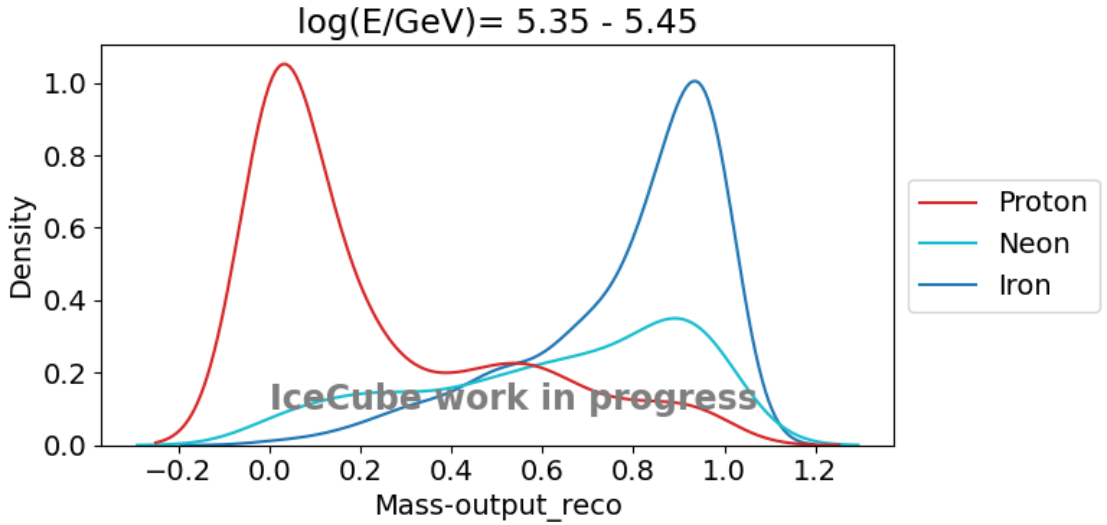


Composition analysis

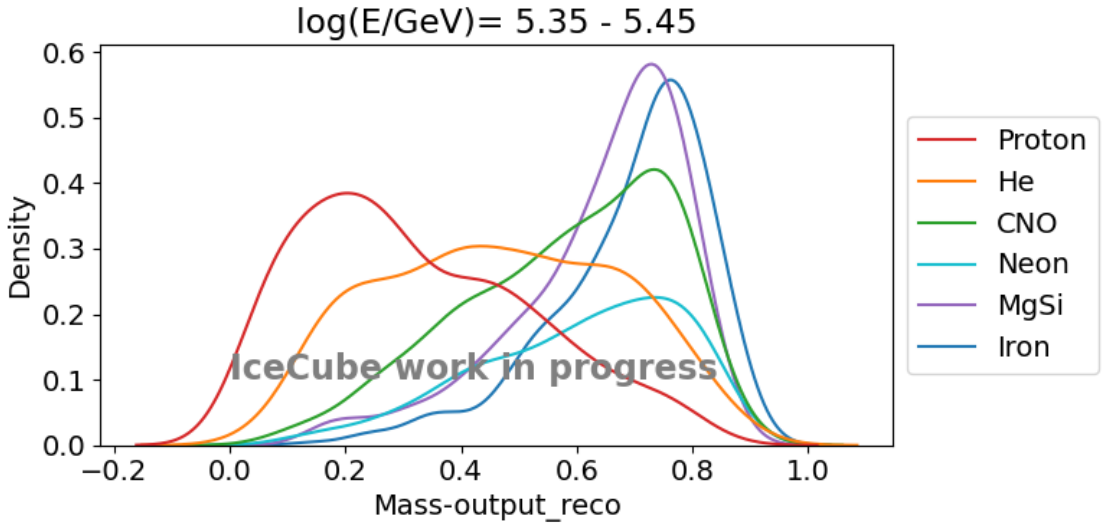
IceAct + Geometry



IceAct + IceCube



IceAct + IceCube



Composition analysis