Updates on Inelasticity and PID Reconstruction with 2D CNNs

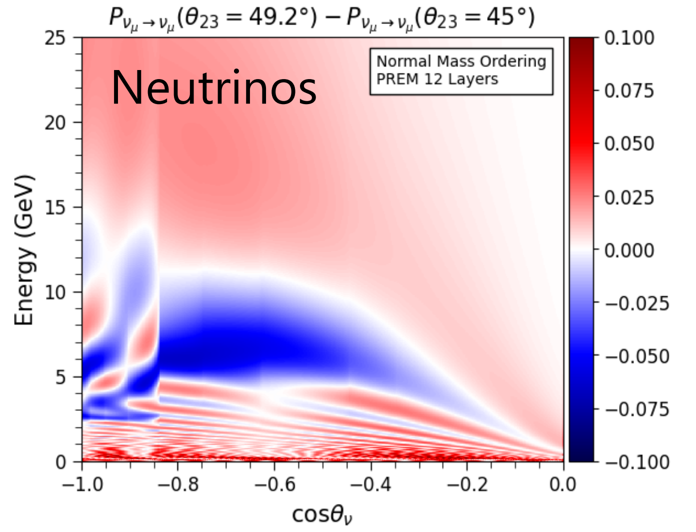
WIPAC
Wisconsin IceCube Particle Astrophysics Center

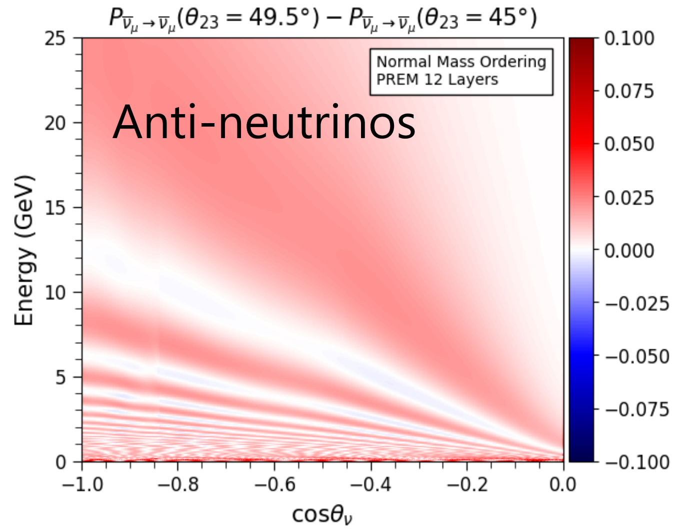
Wisconsin
University of Wisconsin–Madison

Josh Peterson
2023 IceCube Collaboration Meeting
Oscillations Parallel Session

Motivation

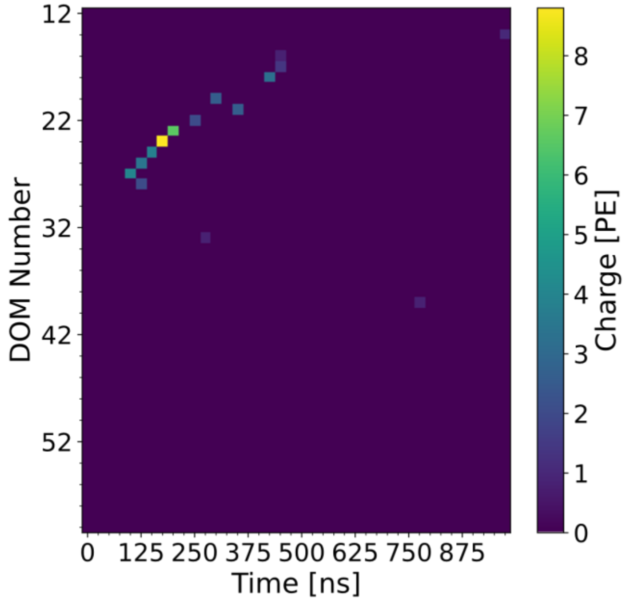
- **Maria and I developed a deep convolutional neural network for PID, I then used the same model for inelasticity**
    - Links to presentations on these networks: Link, Link
- **PID classification is one of the main limiting factors in current oscillation studies**
- **Inelasticity could be used to statistically separate neutrinos from antineutrinos**
    - MSW resonance only occurs for neutrinos or antineutrinos depending on the NMO
    - This could help with studies about NMO, tomography, breaking the octant degeneracy of θ_{23}, BSM studies at higher energies (arXiv:1303.0758, arXiv:1406.3689)

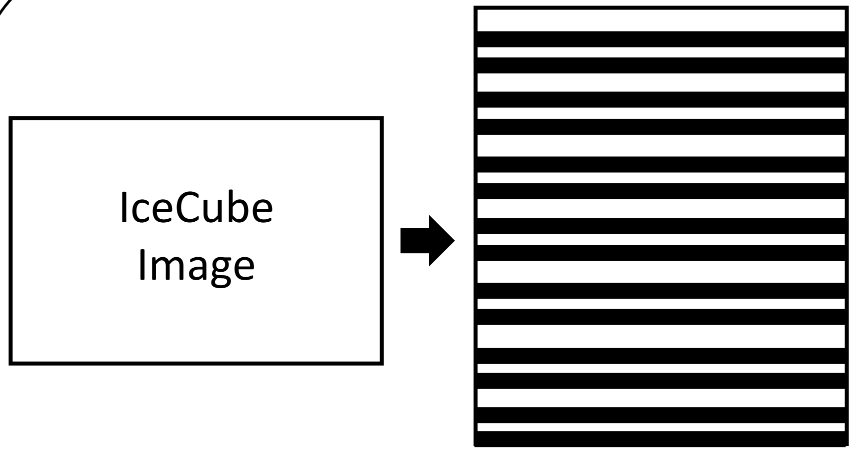




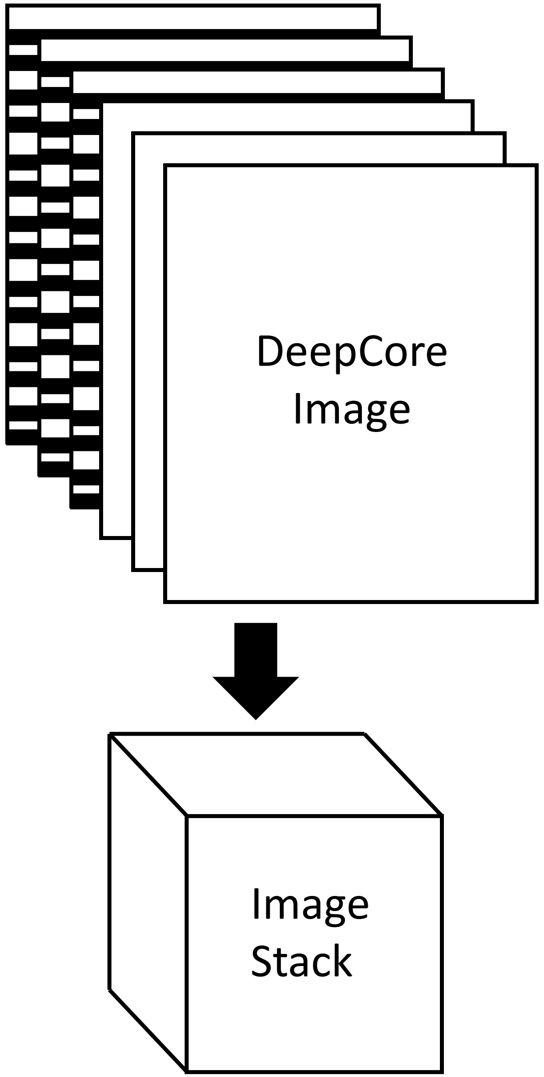
Data Input

- Each row represents a DOM on the string
    - Only use DOMs below the dust layer
- Each column is a slice in time
- The value of each pixel represents the total charge detected





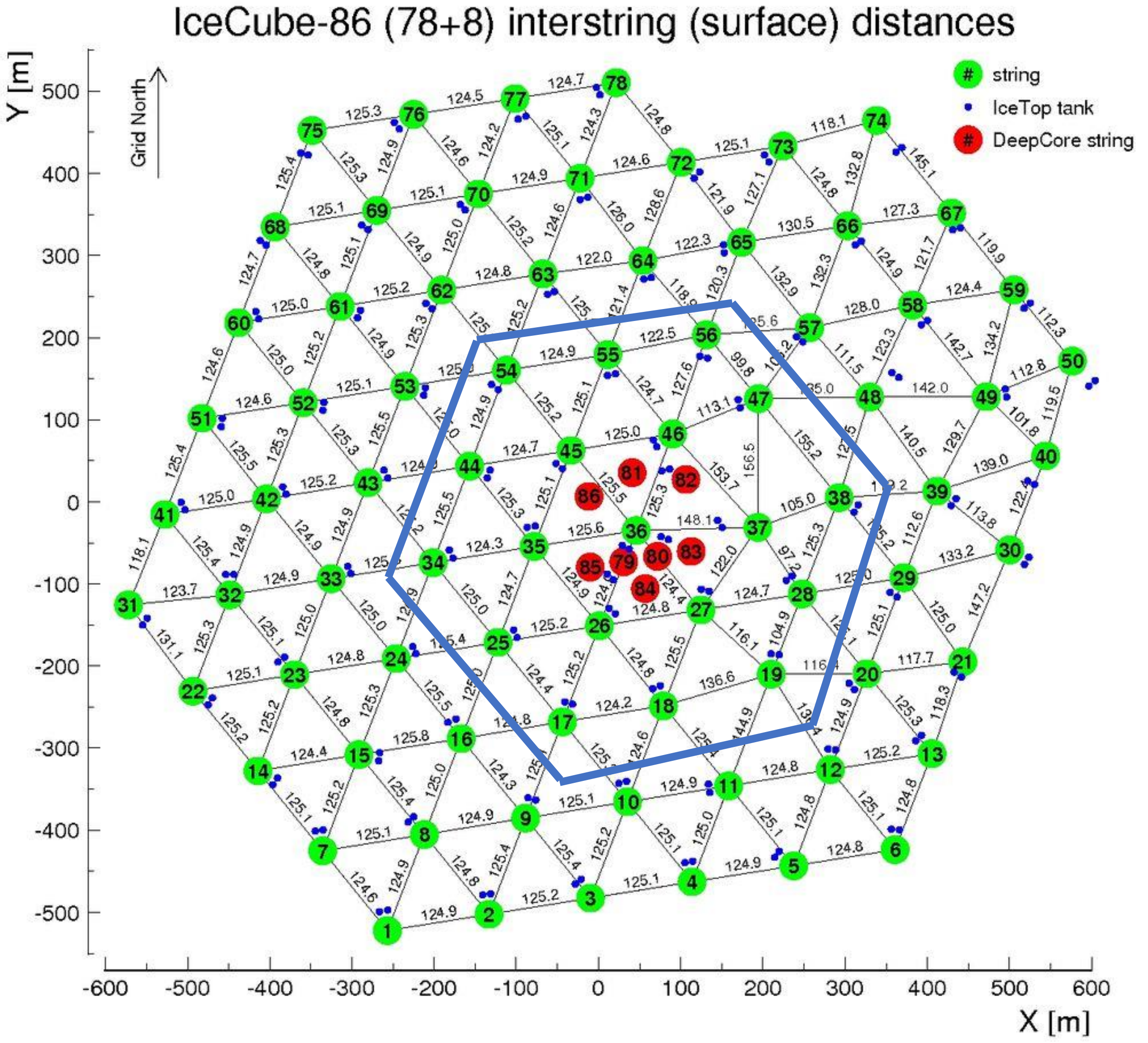
IceCube strings have less DOMs in the DeepCore region than DeepCore strings, so add filler rows



Stack IceCube and DeepCore string images

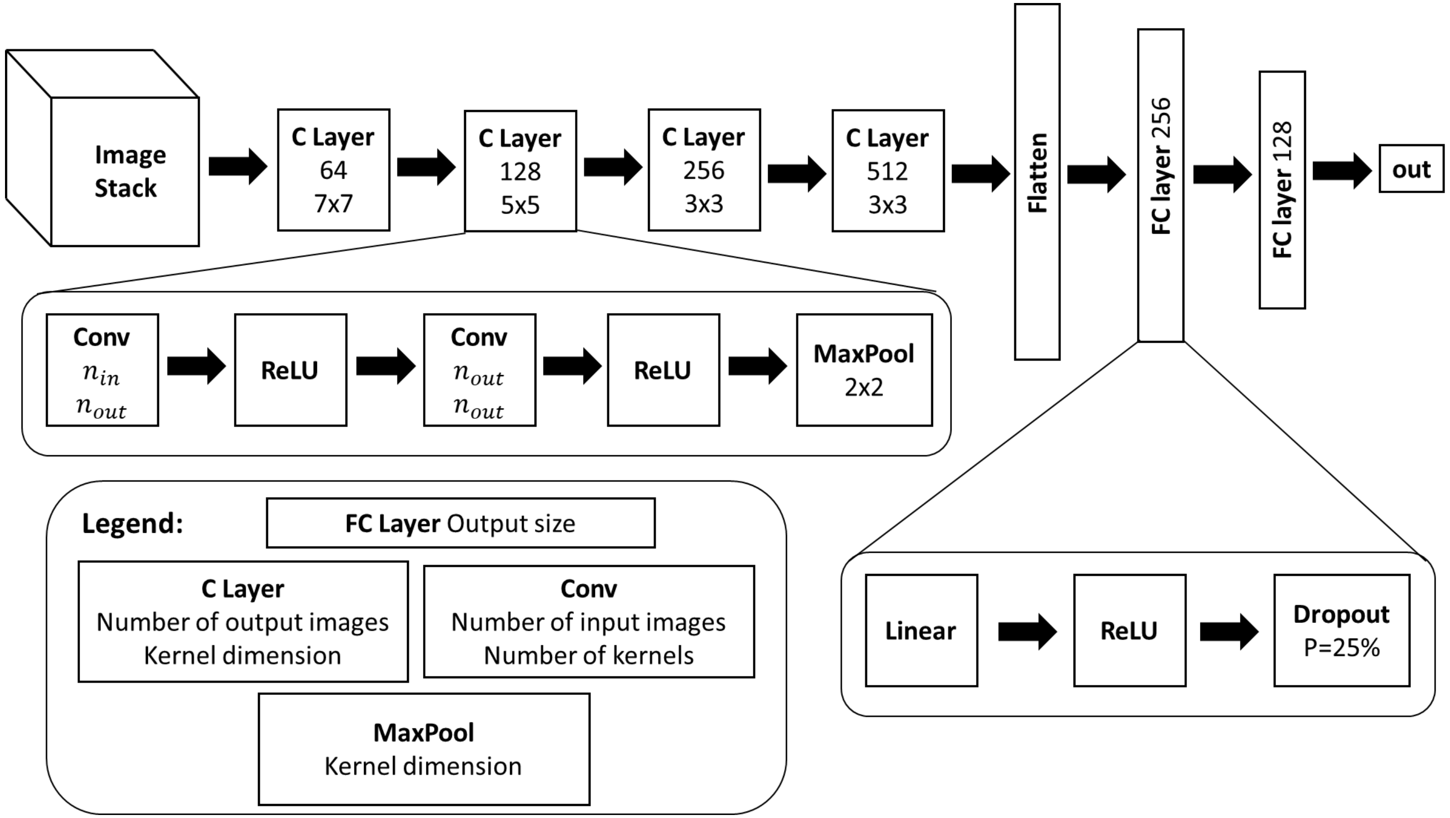
New Geometry

- **I use DeepCore and the IceCube strings that surround DeepCore (shown in the blue hexagon)**
- **Only DOMs below the dust layer are used**
- **In later slides, I will refer to this portion of the detector as “DeepCore Plus”**

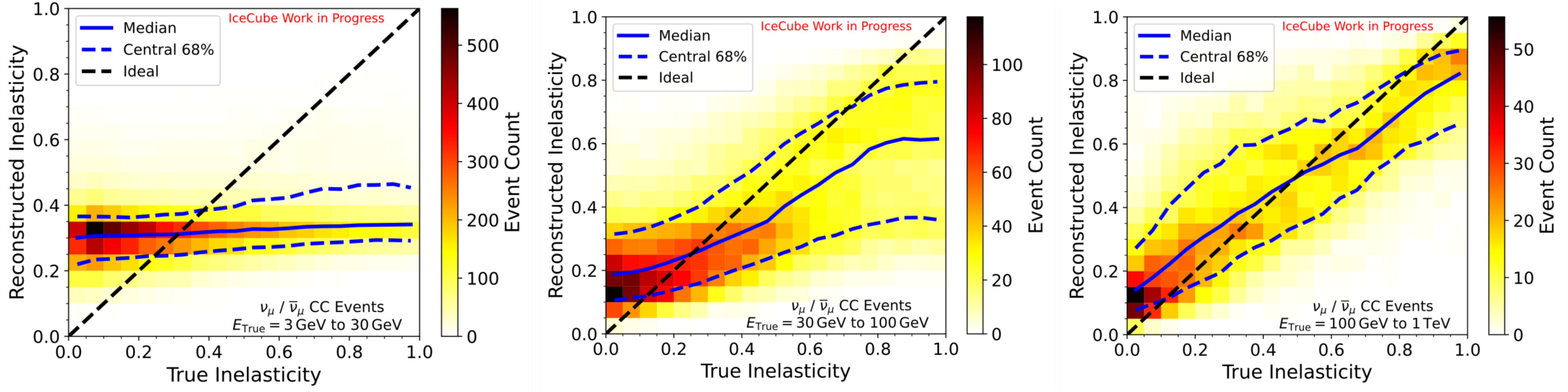


CNN Information

- **10 layer convolutional neural network**
  - 8 convolution layers
  - 2 fully connected layers
- **Trained on level 3 data with some simple charge and containment cuts**
  - Low level data chosen for larger training set
  - Can apply network to higher levels
- **For inelasticity:**
  - Train with ν_μ events
  - Utilize the L1 loss
- **For PID:**
  - Train with ν_μ and ν_e events
  - Utilize a modified binary cross entropy loss



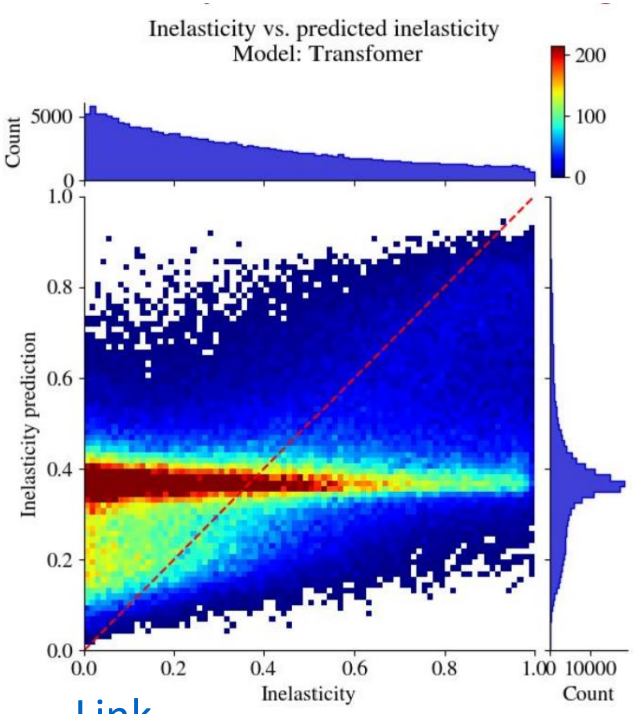
y Results at Different Energies



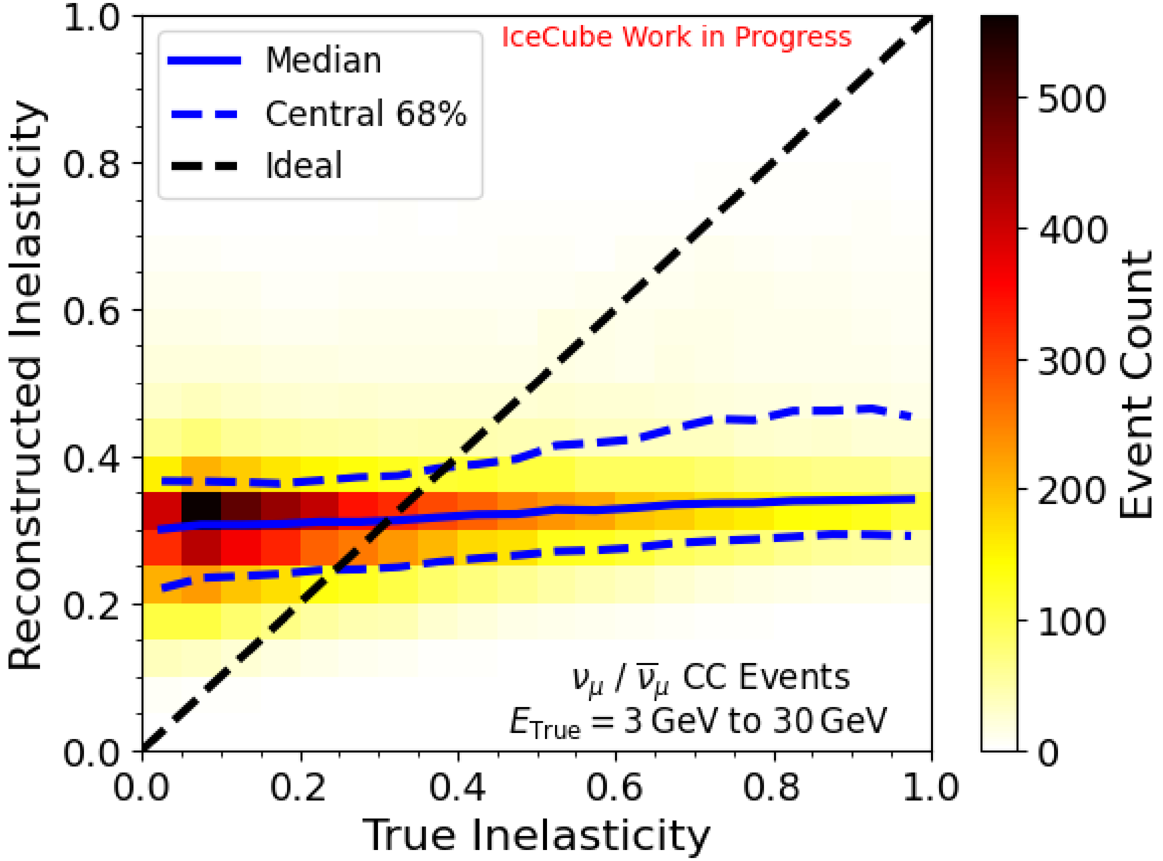
- **Failure to reconstruct y for events below 30 GeV**
  - Matter effects with standard three flavor oscillations are below 20 GeV

- **Reconstruction performance improves with larger energies (expected)**

Why a Flat Distribution?



[Link]

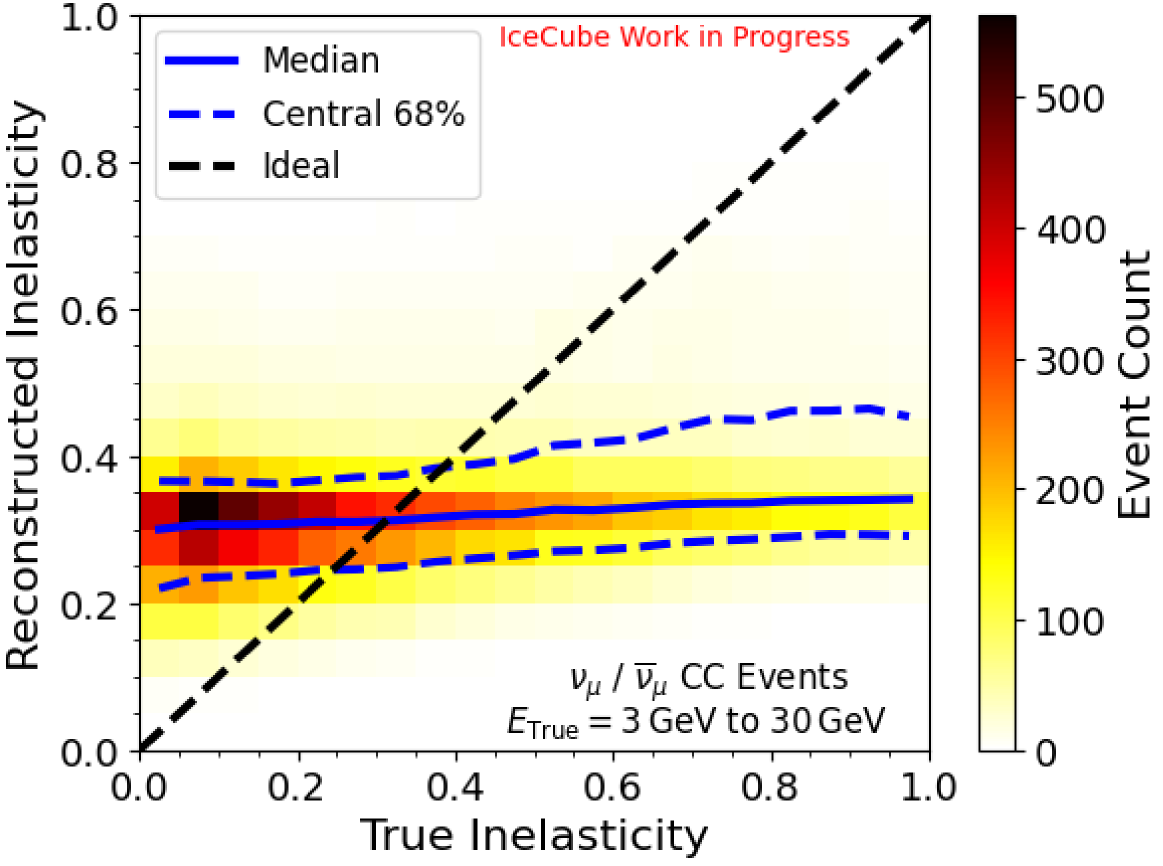


- **Below 30 GeV I get a mostly flat distribution**
  - Seeing similar behavior with transformers
- **Average inelasticity of the training set: 0.36**
- **Likely, there just isn't enough information for effective reconstruction and so the optimal thing to do is guess the mean**

ν_μ / $\overline{\nu_\mu}$ Separation

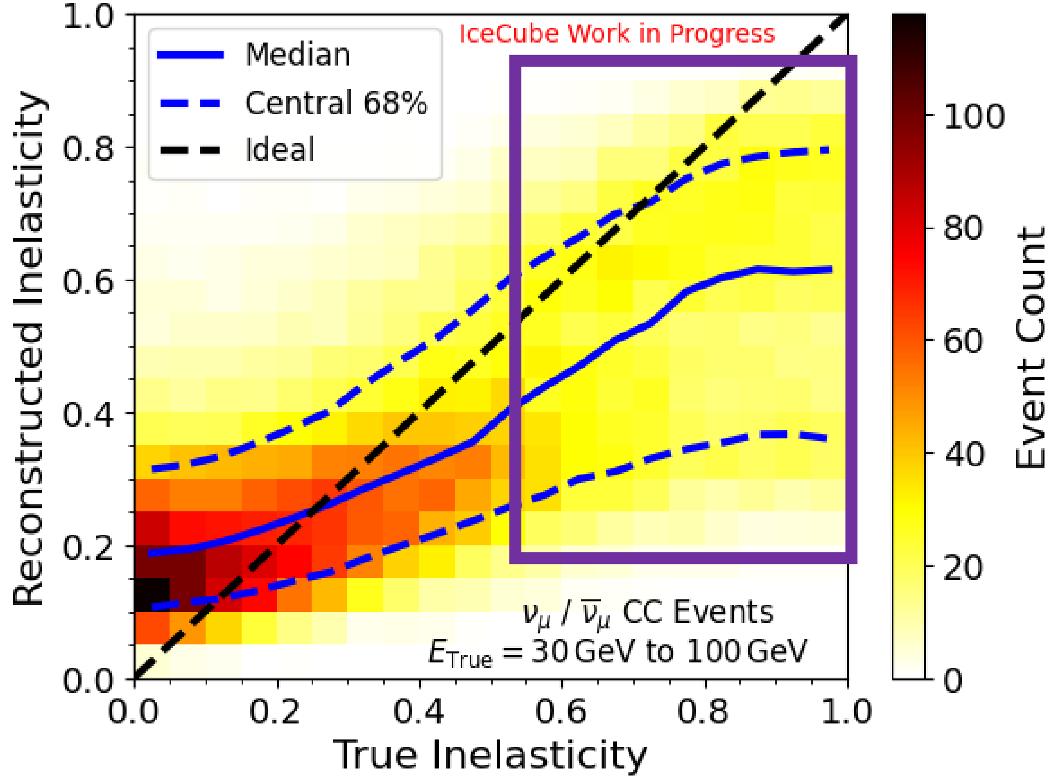
- **The reconstructed inelasticity distribution isn't completely flat**
  - May have some neutrino/antineutrino separation power with just two reconstructed inelasticity bins
- **Bin the track events with true energies from 3 GeV to 30 GeV into two bins, above and below the mean reconstructed inelasticity**

Bin	$\overline{\nu_\mu}/(\overline{\nu_\mu}+\nu_\mu)$
$> y_{reco} = 0.33$	0.30
$< y_{reco} = 0.33$	0.33

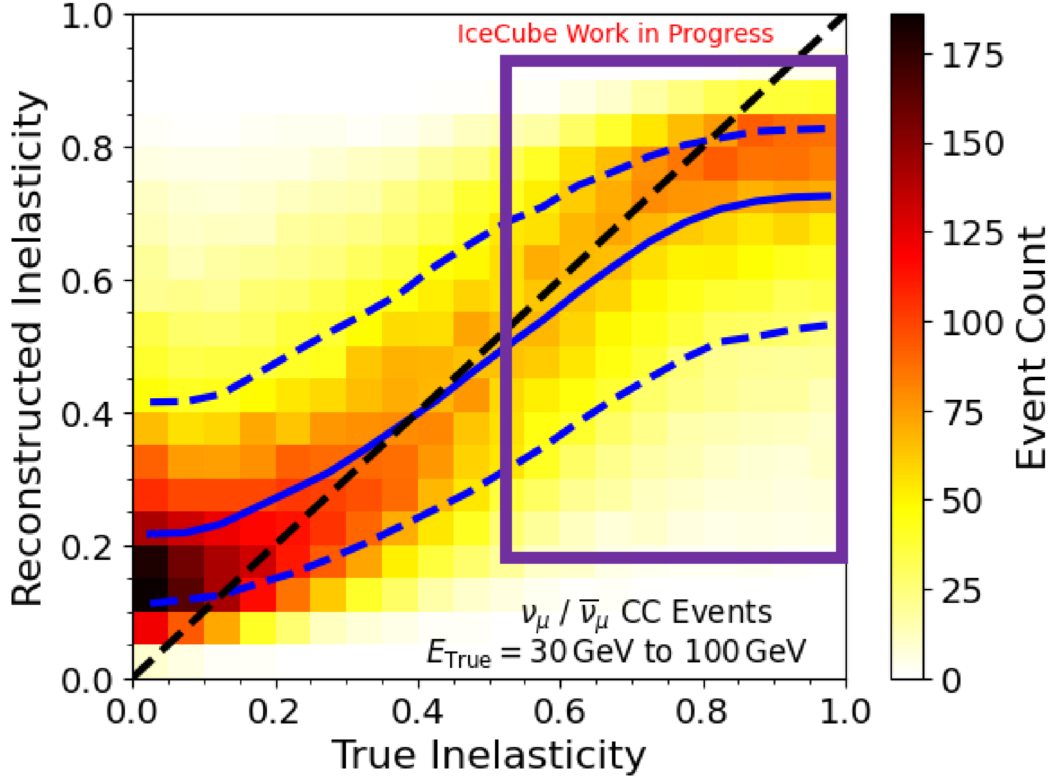


Improving Higher Energy Reconstruction

Training set energy: 3 GeV to 1 TeV

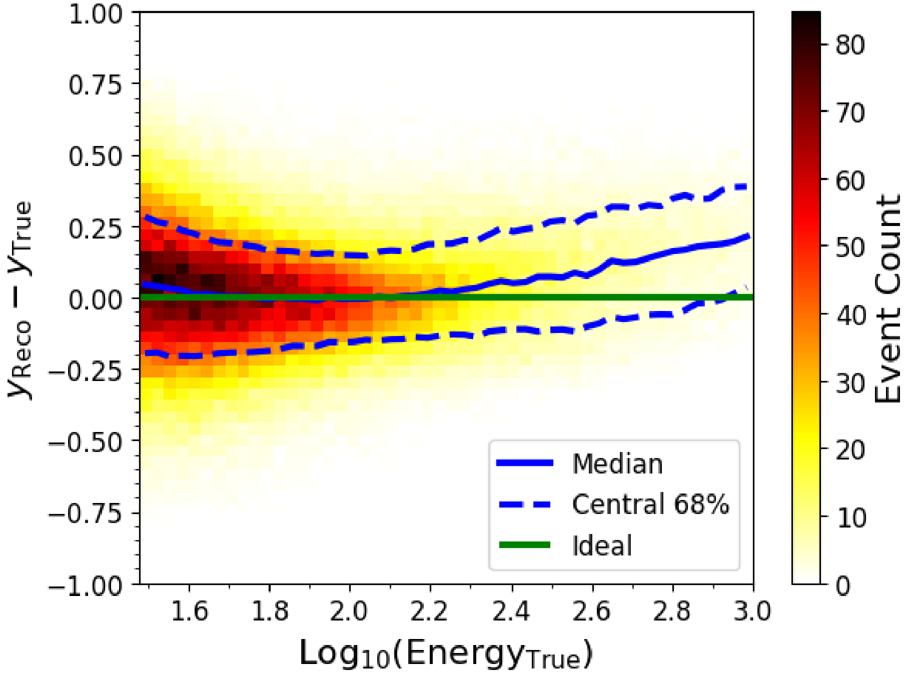


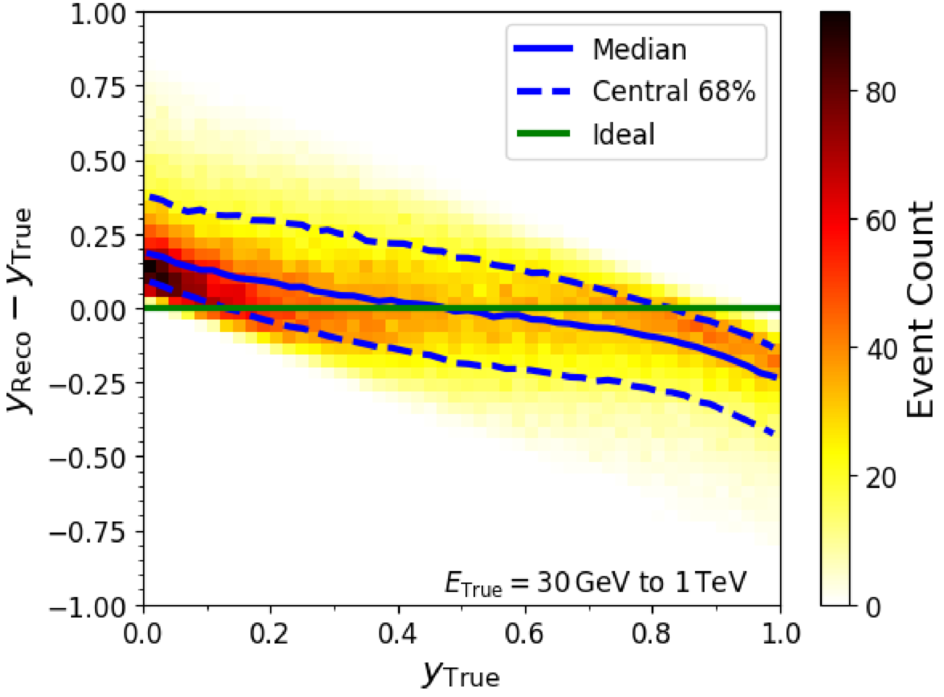
Training set energy: 30 GeV to 1 TeV



- **Definite improvement in reconstruction of larger inelasticities when we remove events with energies below 30 GeV from the training set**

Checking Difference (Above 30 GeV)

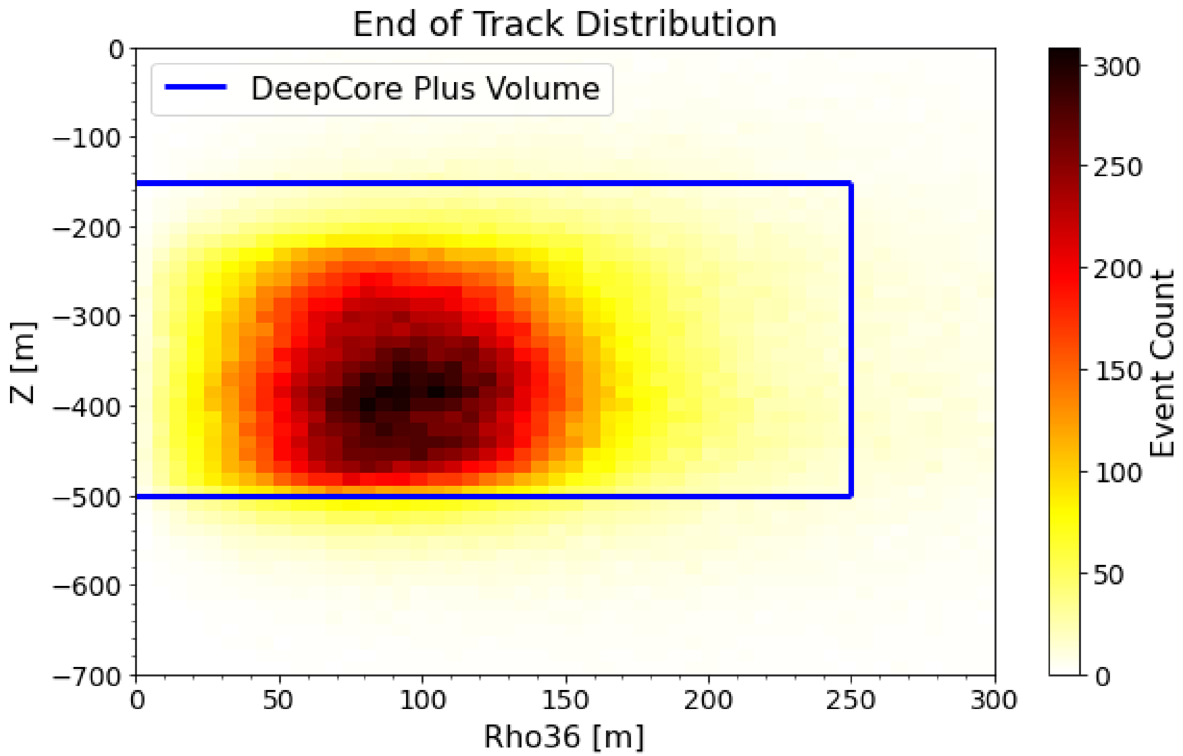


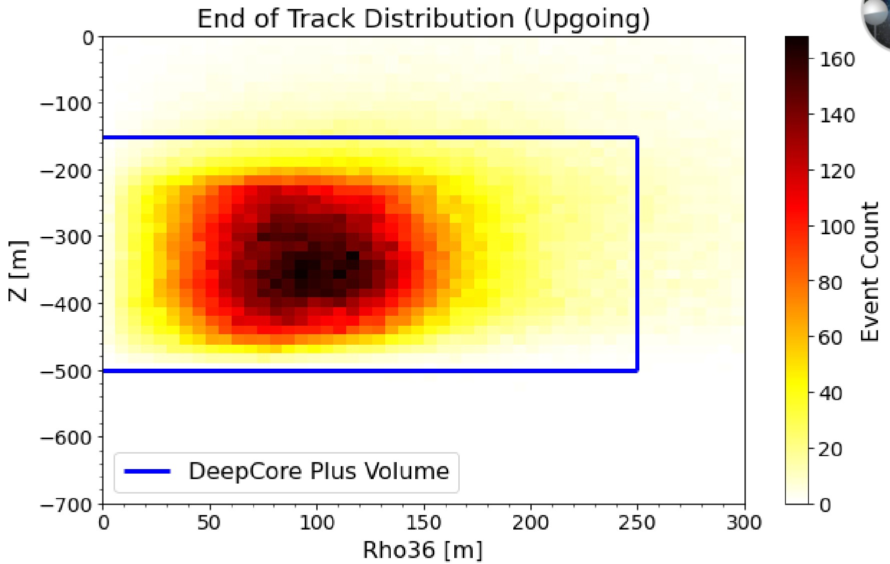


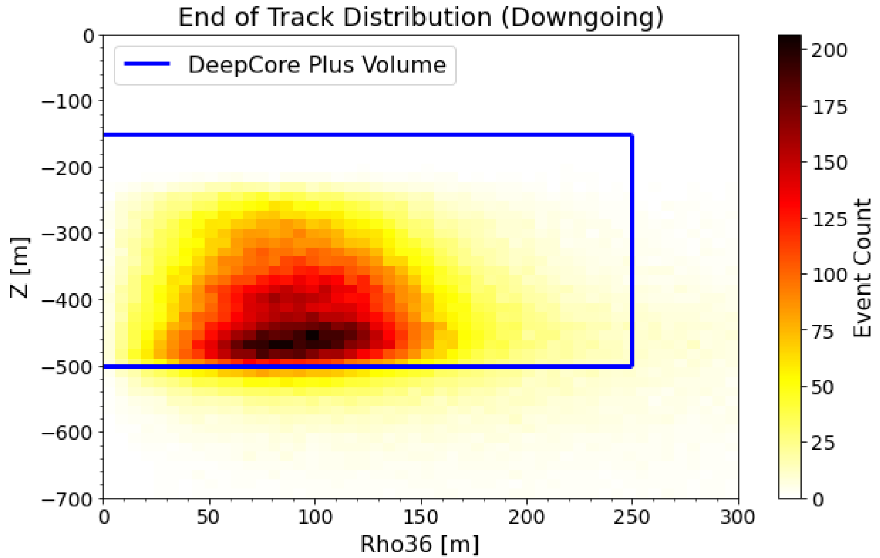
- **The difference vs. energy plot (left) shows a skewing of reconstruction above 300 GeV**
  - Skewing is towards higher inelasticity values, suggesting that the outgoing muon is escaping, which means we see less of the track
  - This gives an upper-end energy limit on what we can do with the “DeepCore Plus” volume

- **The difference vs. true inelasticity plot shows a “squeezing” of the CNN output**

Escaping Muons

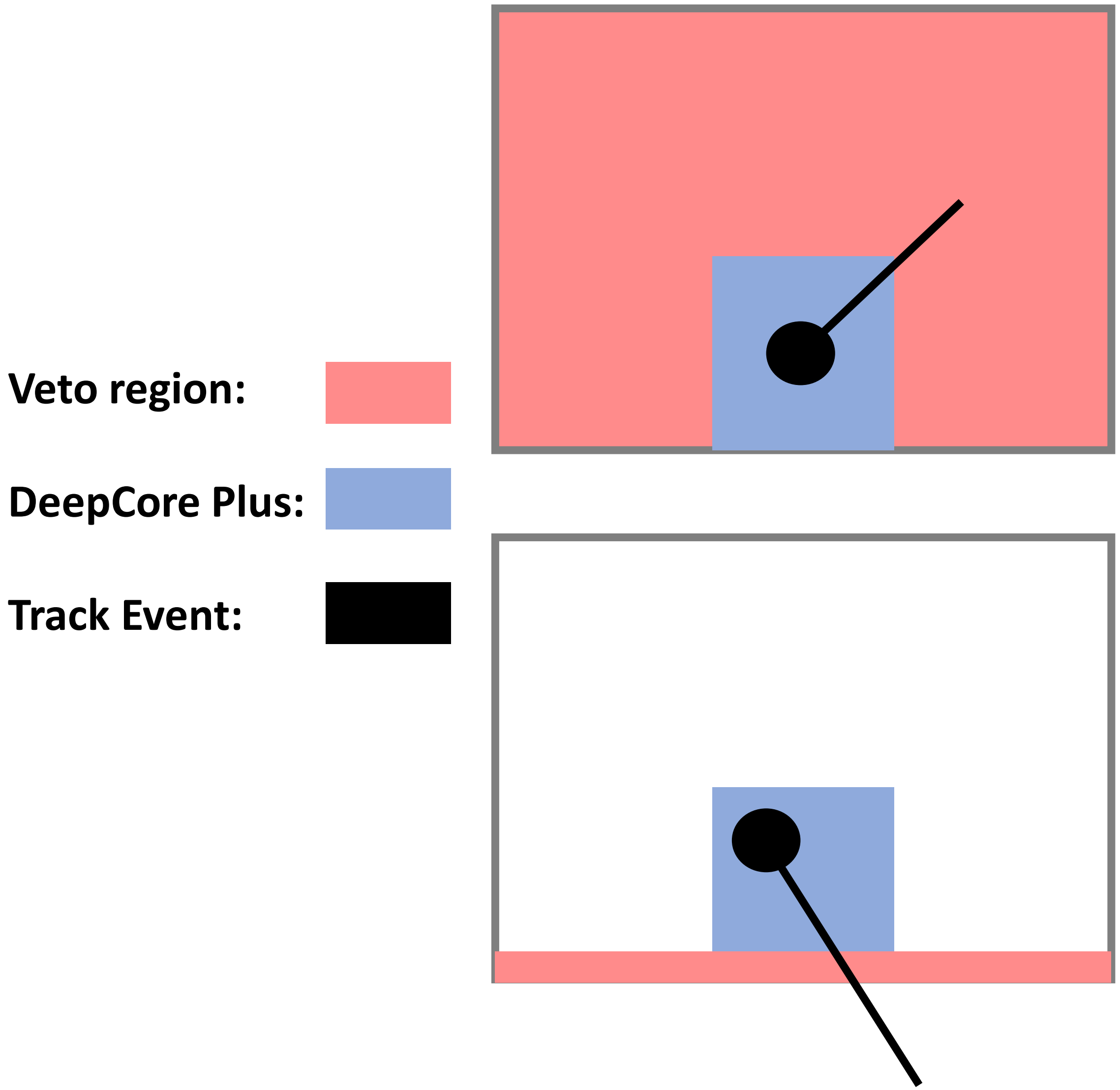






- **When we plot the end of the muon tracks, we see tracks escaping the detector, especially down going events**

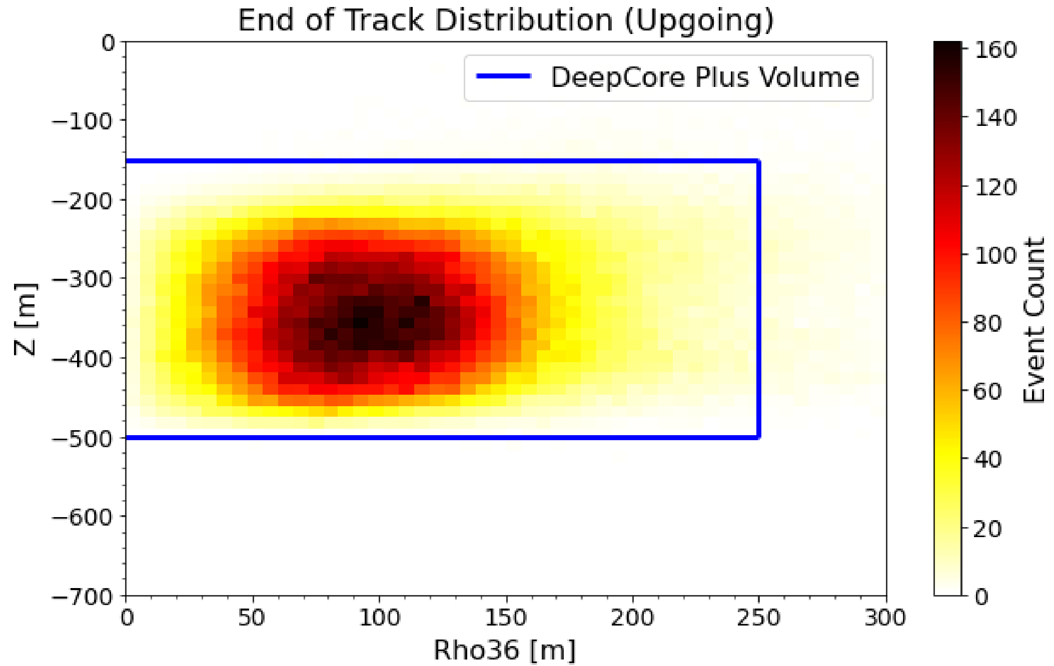
Containment Cuts

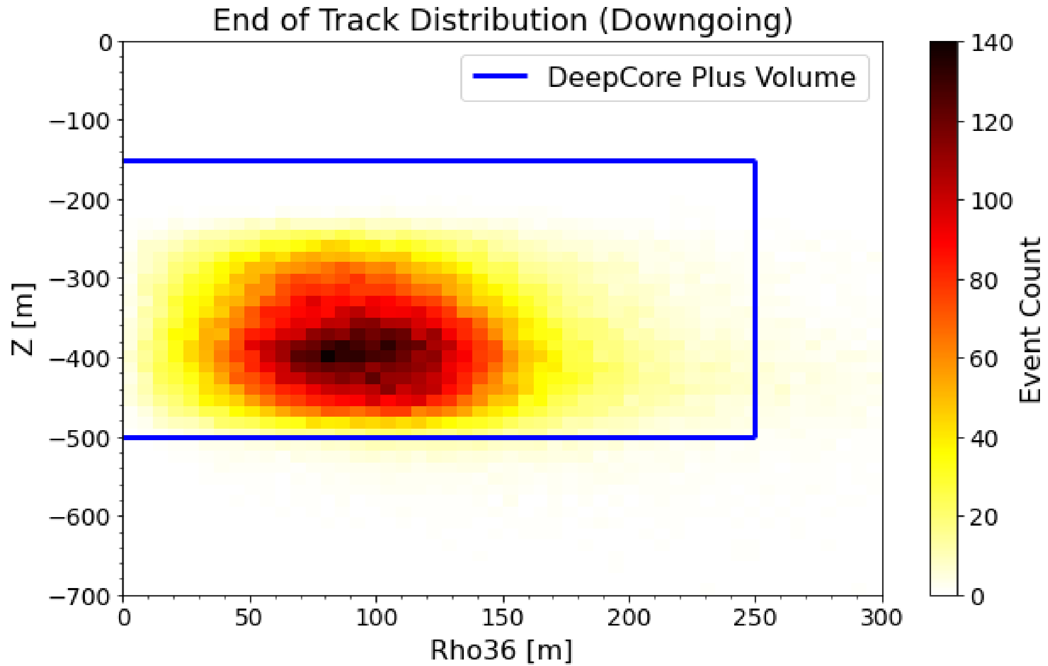


- **Since we are only using part of the detector, we can use the rest of the detector as a veto region**
  - Calculate the ratio of charge in DeepCore Plus to total charge, cut harshly on that ratio (I used 0.99)

- **Use the bottom layer of DOMs as a veto region for down going events. If any charge is seen on the bottom layer, we cut the down going event**
  - Cannot make charge cuts on muons that escape out the bottom of the detector, since there aren't any DOMs to pick up charge

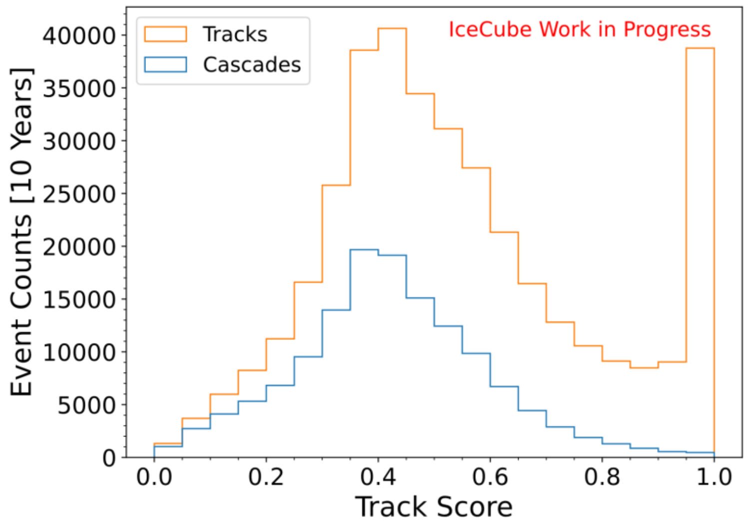
Cut Results

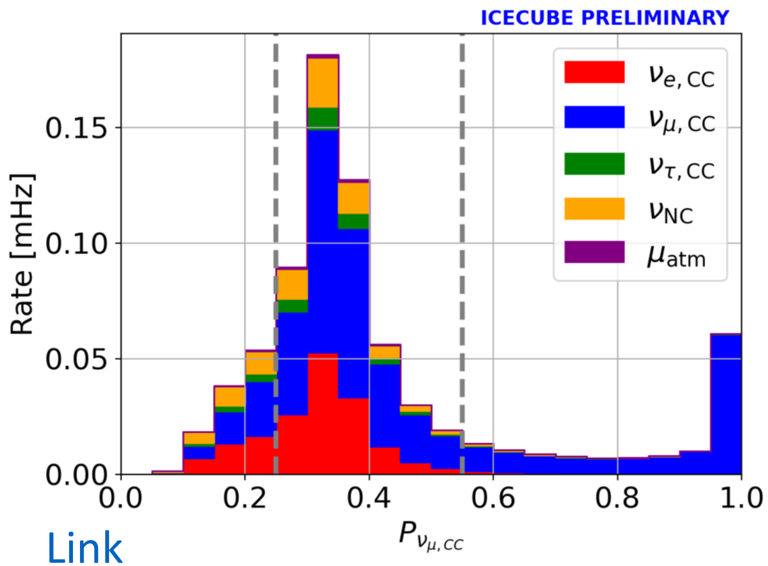




- **With these new cuts we see significantly less events escaping the DeepCore Plus volume**
- **Currently training CNNs for PID and y with FLERCNN Level 6 & 8 MC sets using these new cuts**
  - Also limiting the energy to 300 GeV

Many Reconstruction Algorithms



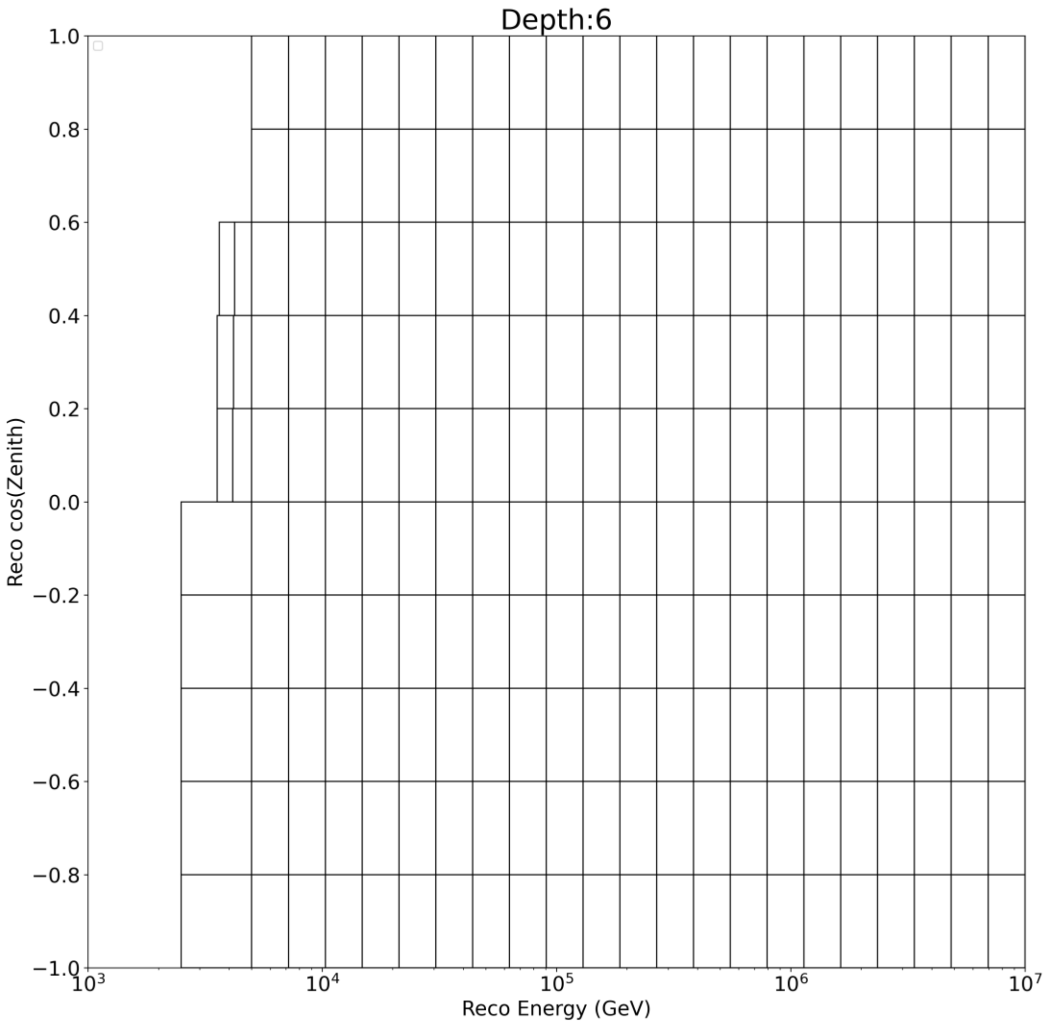


Link

- **At this point in time, we have multiple reconstruction algorithms for PID, Energy, etc.**
- **Rather than sticking with one, it might be worth using multiple**
  - Stacked Generalization – Use outputs of many classifiers / regressors as input for another classifier / regressor
  - Voting – select classification based on majority vote
  - Averaging
- **Going to investigate combining FLERCNN and my 2D CNNs for improved PID classification**

Binning w/ Decision Trees

- **Can use a decision tree to generate binning schemes**
- **Bin sizes correspond to the amount of information available in that region of parameter space**
    - Can enforce minimum bin counts, good for hypersurfaces!
- **Vedant Basu and I are working on figuring out how to encode physics information into the generation of the decision trees**
- **Maybe applicable to oscillation analyses as well?**



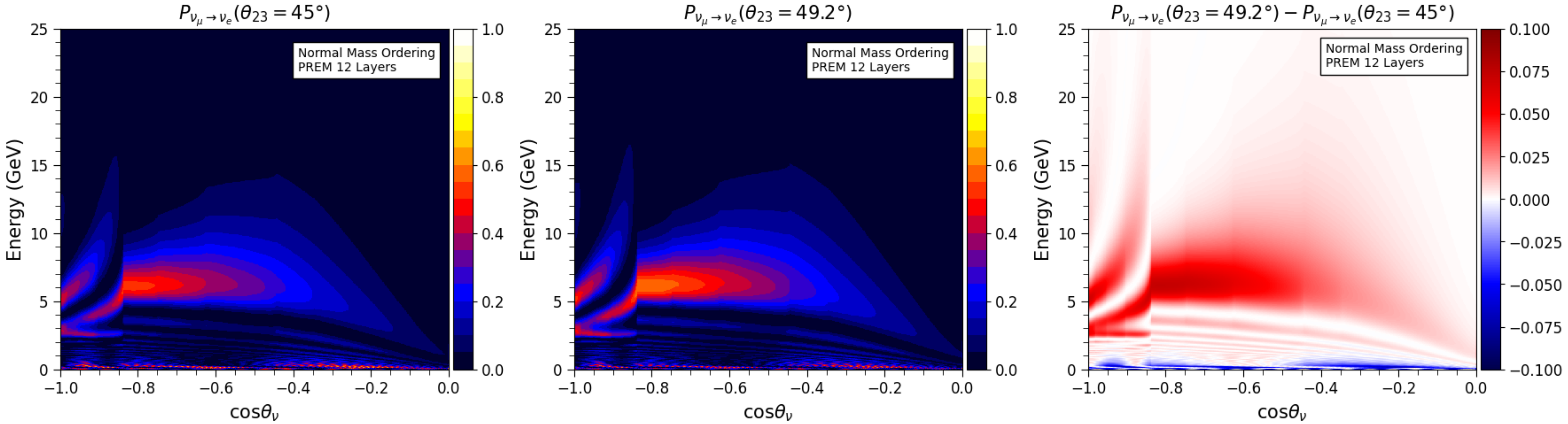
Summary

- **Have trained 2D CNNs for inelasticity and PID reconstruction**
  - Can reconstruct inelasticity above 30 GeV
  - Minimal neutrino / antineutrino separation below 30 GeV

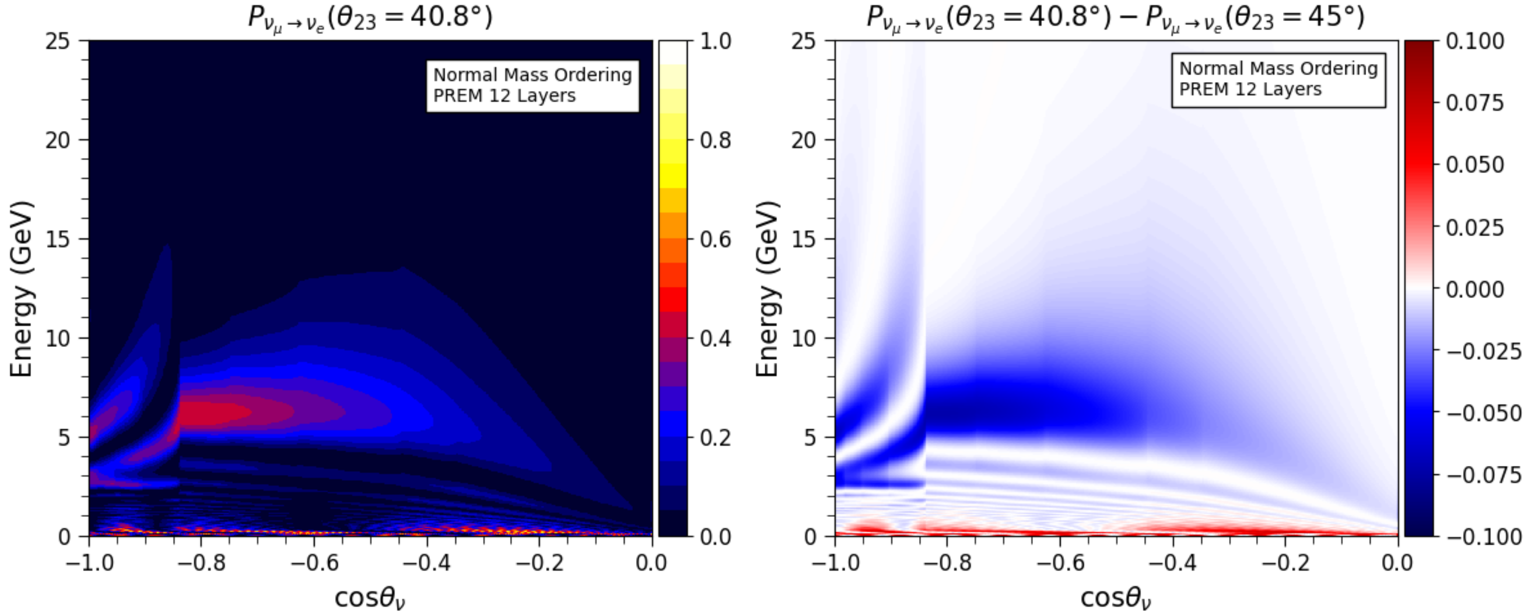
- **Determined better energy cuts and developed new containment cuts for training**

- **Coming in the near future:**
  - CNNs trained with new energy and containment cuts
  - Combining FLERCNN and my 2D CNNs for PID
  - Binning with decision trees?

Backup Slides



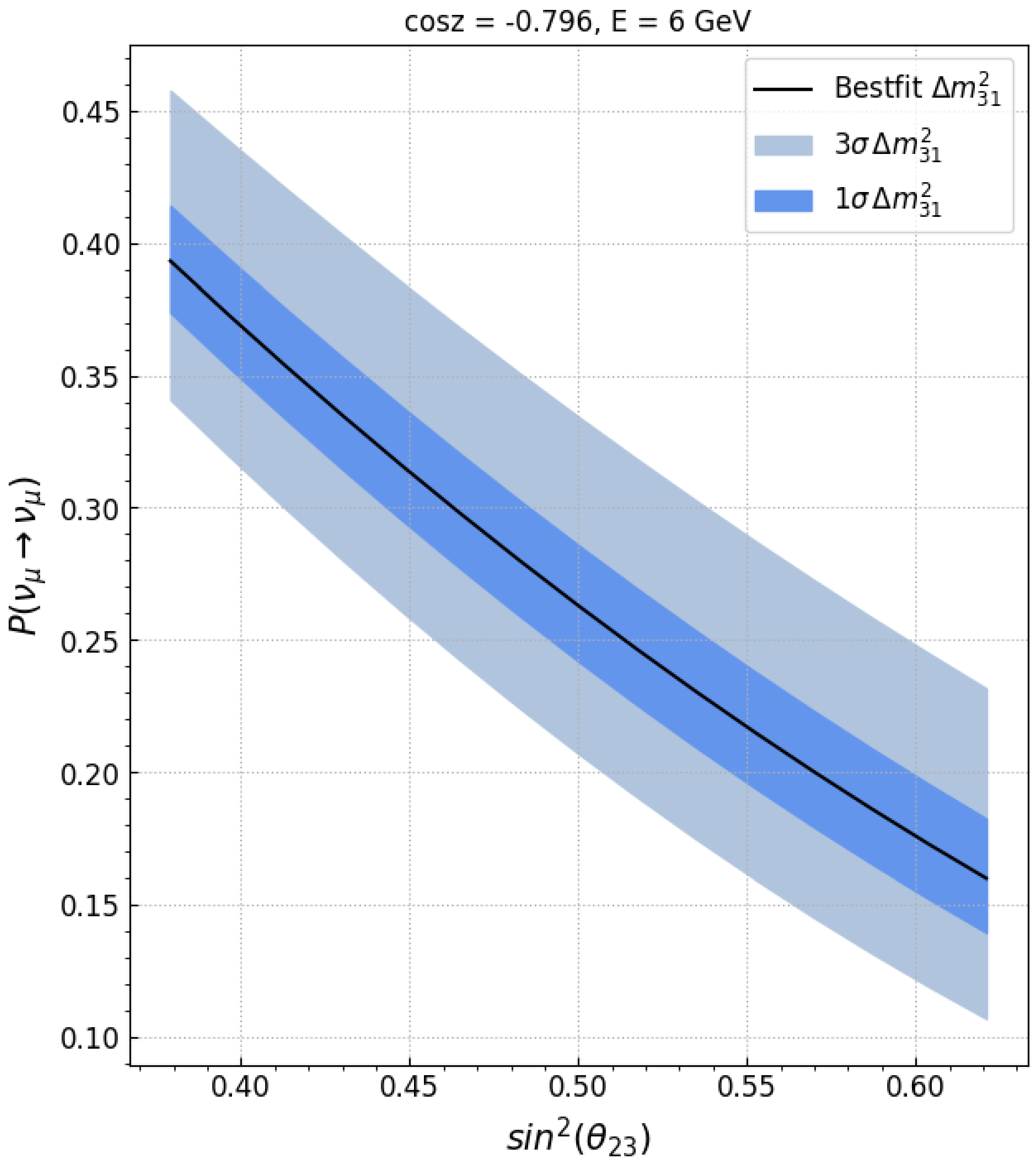
- $\nu_\mu \to \nu_e$ probability increases when $\theta_{23} > 45^\circ$ and decreases when $\theta_{23} < 45^\circ$
    - $\Rightarrow$ suppression of track events for $\theta_{23} > 45^\circ$
    - $\Rightarrow$ enhancement of track events for $\theta_{23} < 45^\circ$
- $P_{\mu e} = 4 s_{13}^2 {\color{red} s_{23}^2} \frac{sin^2(A-1)\Delta}{(A-1)^2} + H.O.T.$
- Leading probability term is *asymmetric* about $\theta_{23} = 45°$



$$A = \frac{2E}{\Delta m_{31}^2}(7.56 \times 10^{-14} eV)\left(\frac{\rho(x)}{g/cm^3}\right) Y_e(x)$$

Breaking the Octant Degeneracy

- Referring to the oscillogram difference plots on slide 2, I wanted to check the baseline corresponding to the maximum difference from $\theta_{23} = 45^{\circ}$
- Δm_{31}^2 seems to have a significant impact on the survival probability, and introduces degenerate solutions
- The next plots will show the degeneracy gets worse with increasing energy

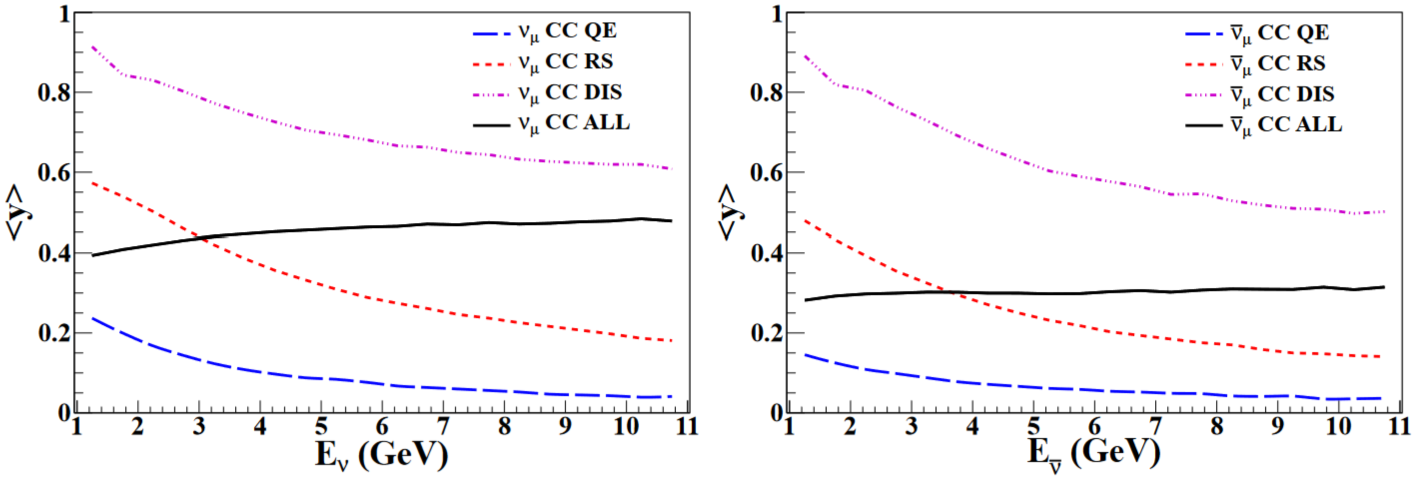


Inelasticity

- Bjorken y $:= \frac{p_2 \cdot (p_1 - p_3)}{p_2 \cdot p_1}$

- In the high energy limit:

$$y = \frac{E_\nu - E_\mu}{E_\nu}$$



arXiv:1406.3689

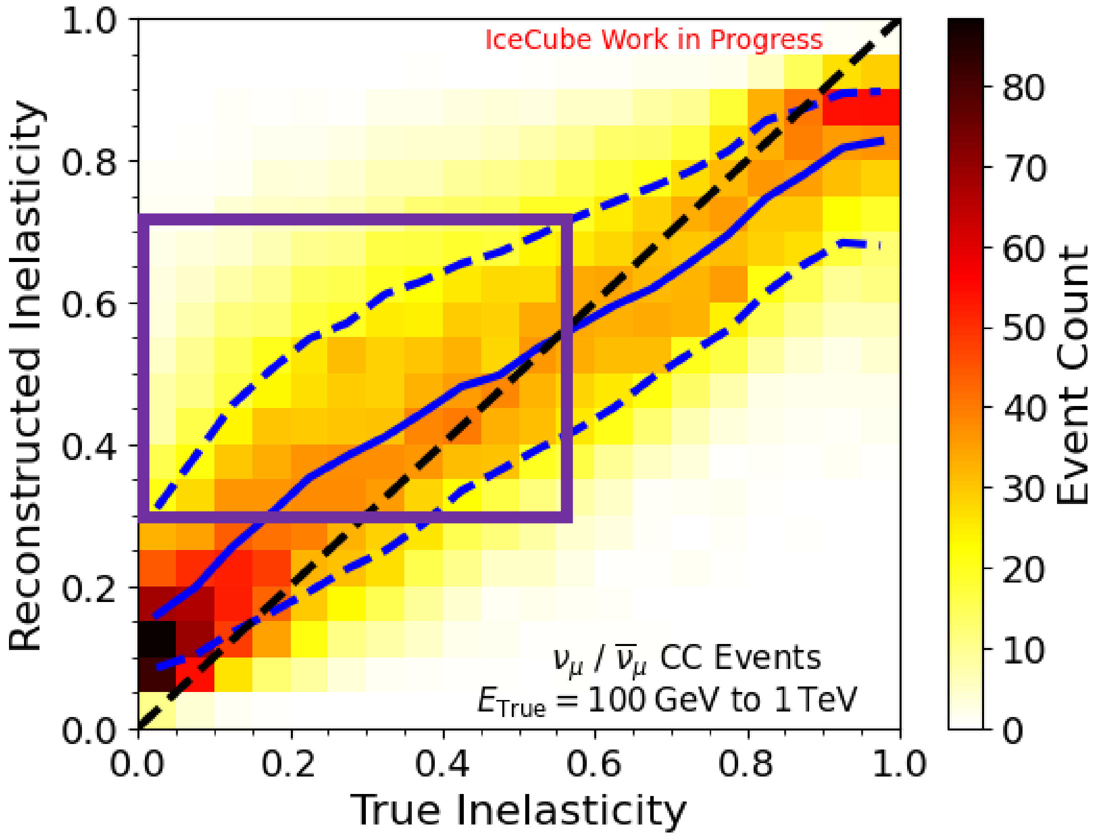
- Neutrinos have a higher average inelasticity than antineutrinos

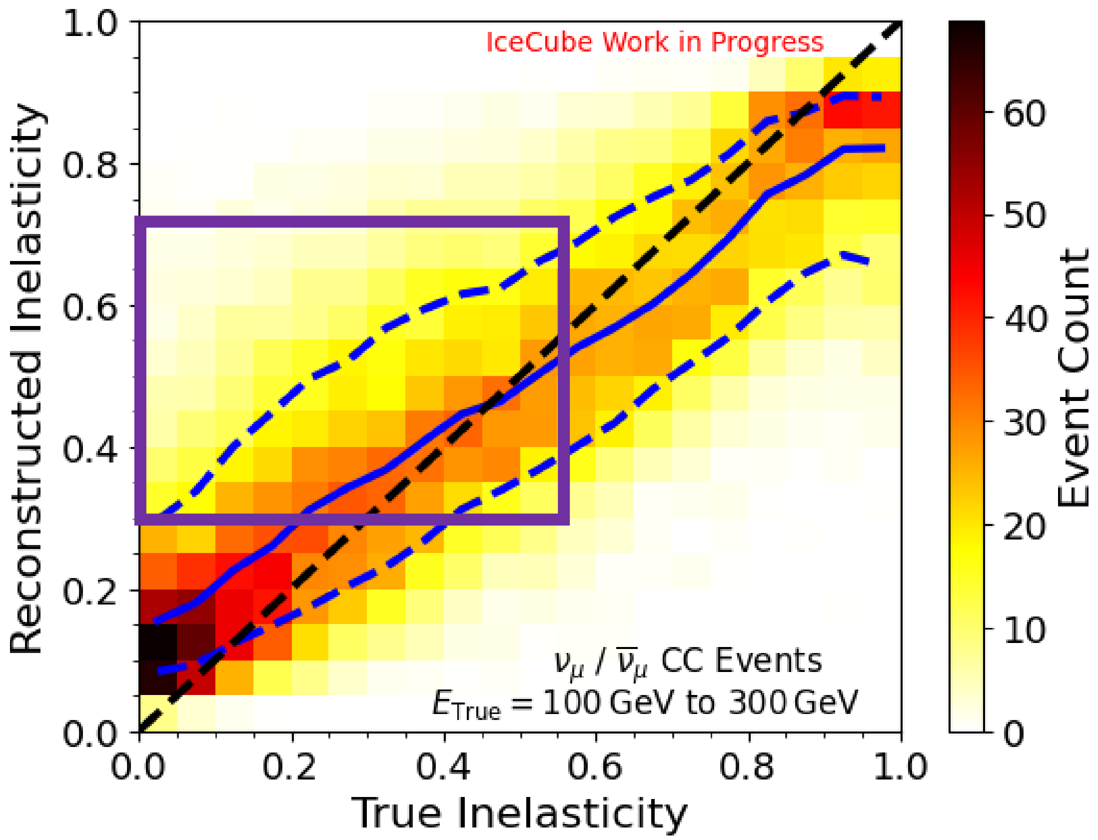
- At GeV energies, the outgoing muon is a minimum ionizing particle ⇒ average track length is directly proportional to muon energy
  - Determining inelasticity could then also be determined through topology, so a natural candidate for our CNN

Other Things I Tried That Didn't Help

- Jason Koskenan suggested I try having the NN output a beta probability distribution instead of a measured y
  - NN outputs the concentrations, α and β, of the beta distribution
  - The loss you minimize is $-Log(Beta(y_{true}; \alpha, \beta))$
  - Your measurement is then the mode of the distribution
- I tried incorporating spatial information through additional channels per image, and then you convolve with a 4-dimensional kernel
  - Kind of strange since I'm only adding information

Impact of >300 GeV Events

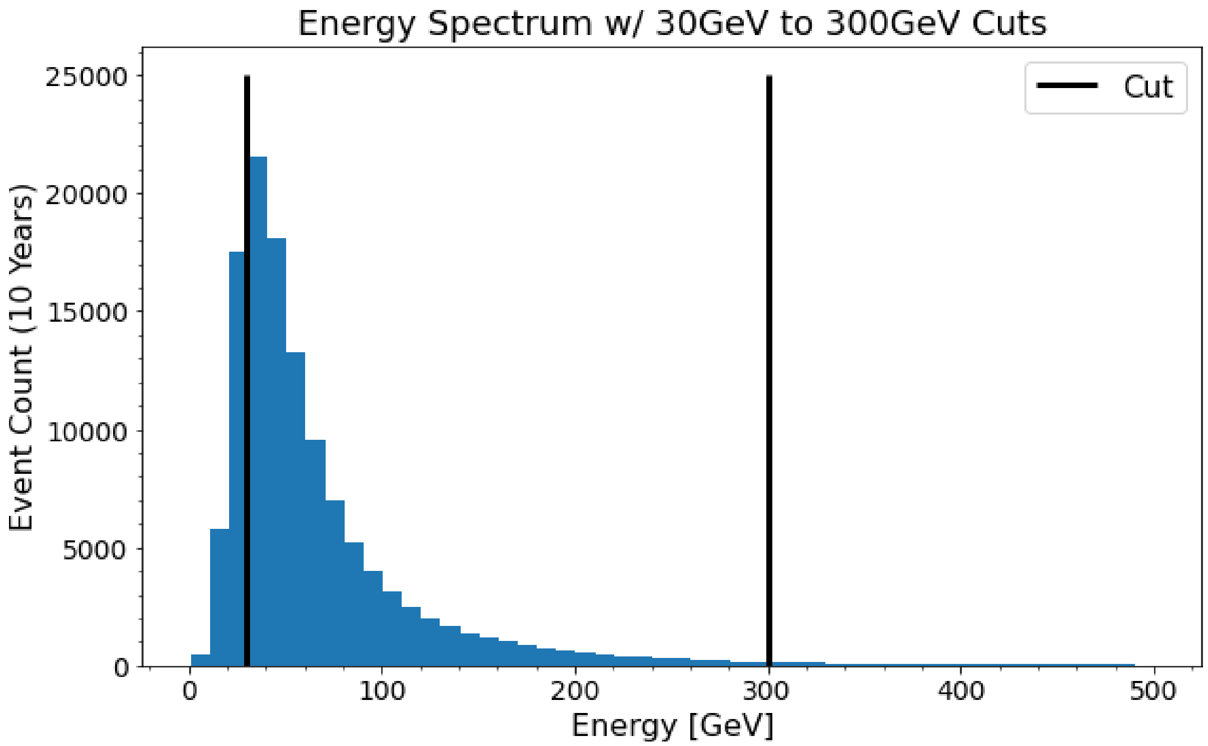




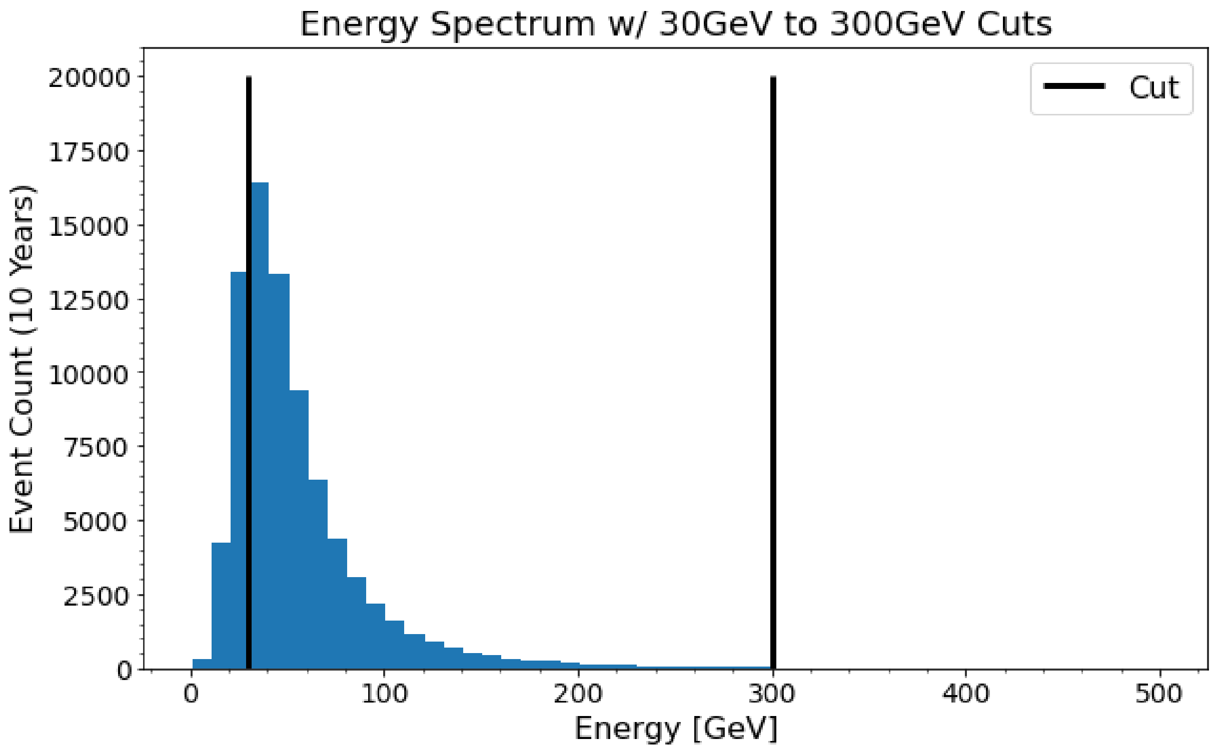
- **Better fit at lower inelasticity values**

Cut Results

Before containment cuts



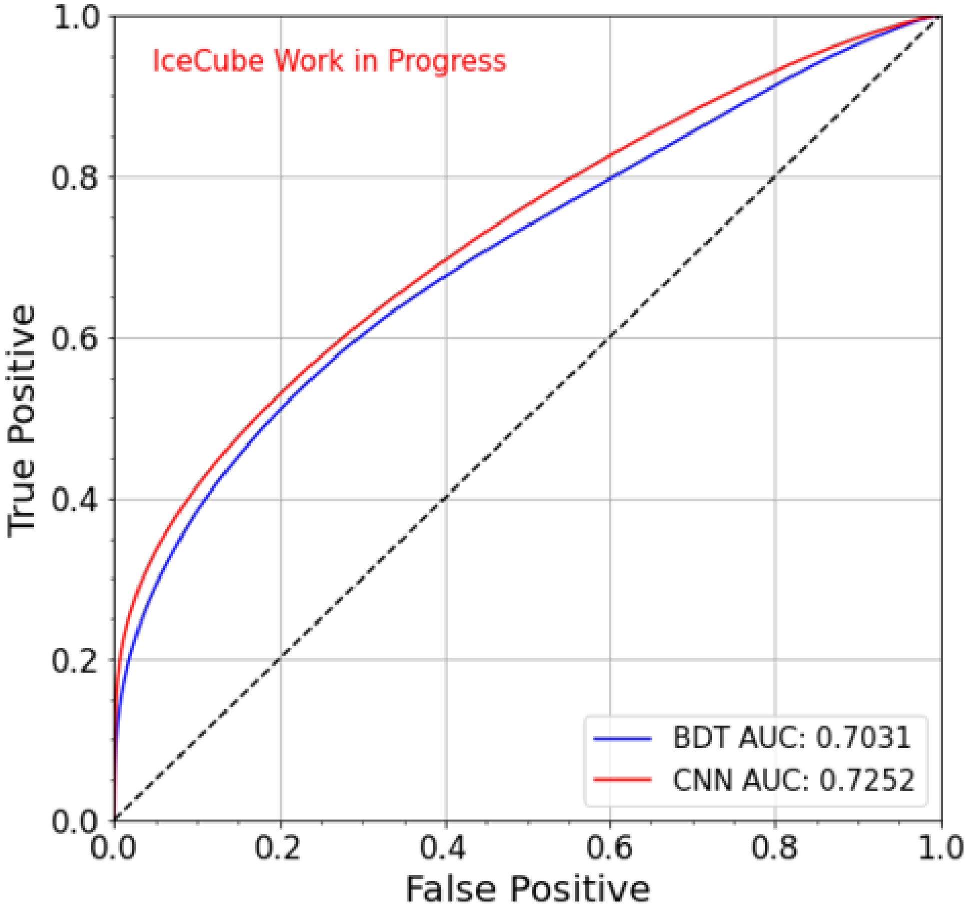
After containment cuts



- **With the new cuts we also see a significant reduction in the tail of the distribution towards high energies**

PID Results

- Receiver Operating Characteristic (ROC) Curves for the 2D CNN and BDT
- Plots include numu, nue, and nutau events
    - Nutau events are labeled as cascade
- Events are weighted with the atmospheric weights (assume 10 years of livetime)



PID Results

- Histogram of test set events binned by assigned track score
- Same events as the previous plot
- Events are weighted with the atmospheric weights (assume 10 years of livetime)