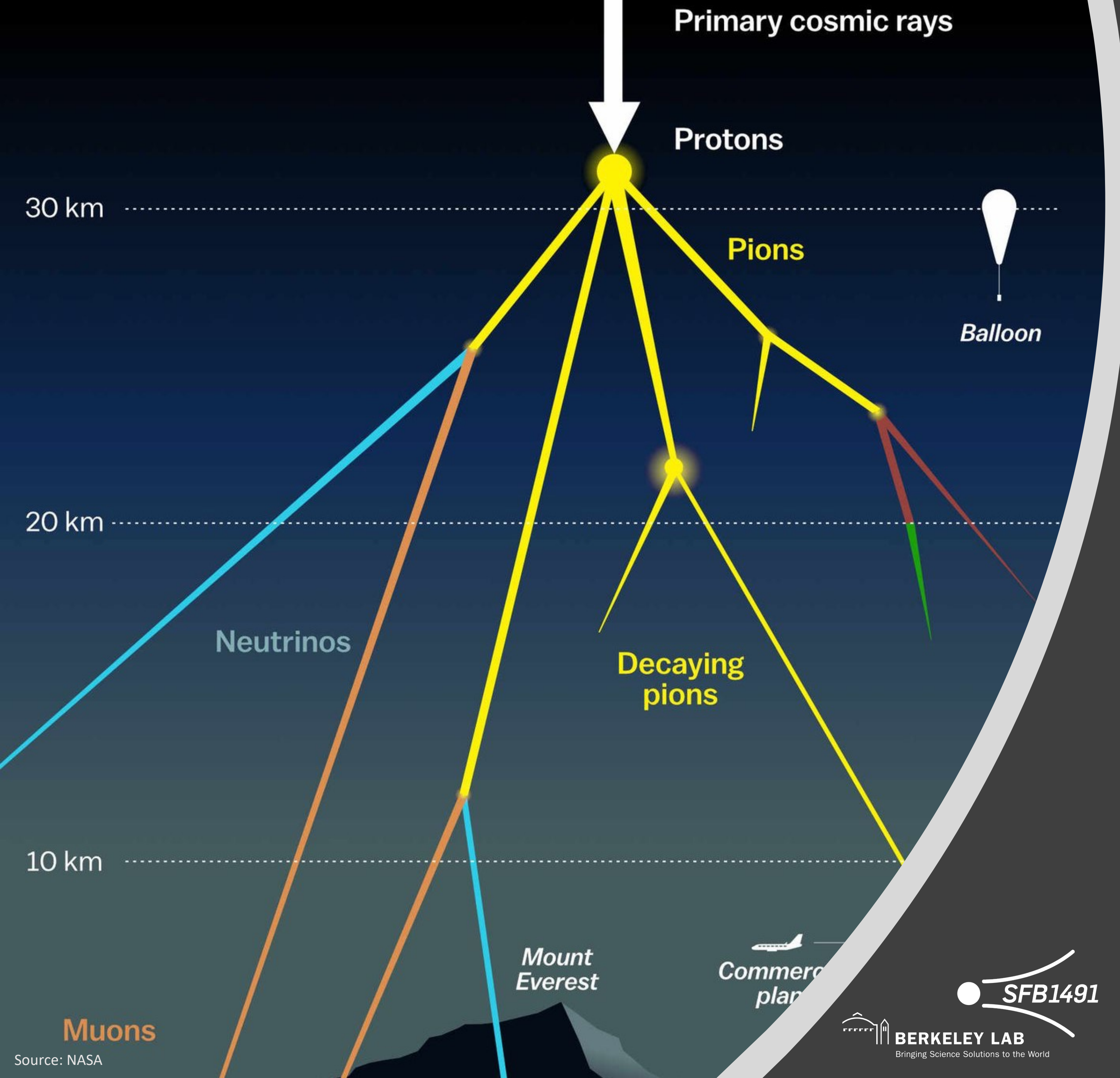


Source: NASA

tu dortmund university

ICECUBE NEUTRINO OBSERVATORY

Unfolding the Atmospheric Muon Flux with IceCube

Pascal Gutjahr

wiki page

CRWG: Dennis
Collab: Anatoli
Tech: Karolin

IceCube Meeting Utah

October 7, 2025

Introduction

Motivation

- Characterize muon flux components and depth intensity
- Probe hadronic interactions in the atmosphere
- Constrain uncertainties and longstanding questions like the “muon puzzle”

Long term

- Enhance understanding of hadronic processes relevant in distant astrophysical sources (e.g., supernovae, AGNs, …)

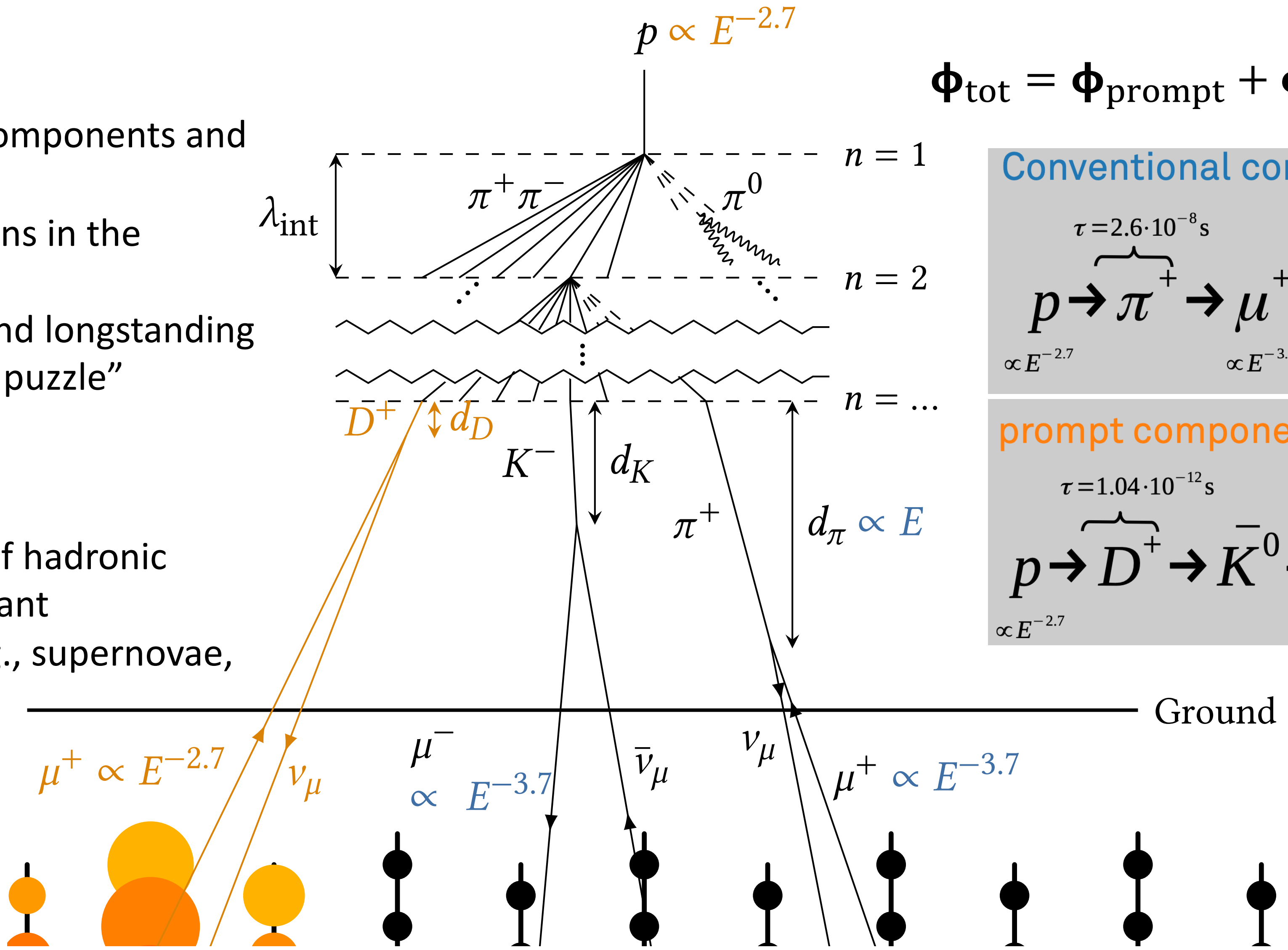
$$\Phi_{\mathrm{tot}} = \Phi_{\mathrm{prompt}} + \Phi_{\mathrm{conventional}}$$

Conventional component:

$$\underset{\propto E^{-2.7}}{p} \rightarrow \overbrace{\pi^+}^{\tau = 2.6\cdot 10^{-8}\,\mathrm{s}} \rightarrow \underset{\propto E^{-3.7}}{\mu^+} + \nu_\mu$$

prompt component:

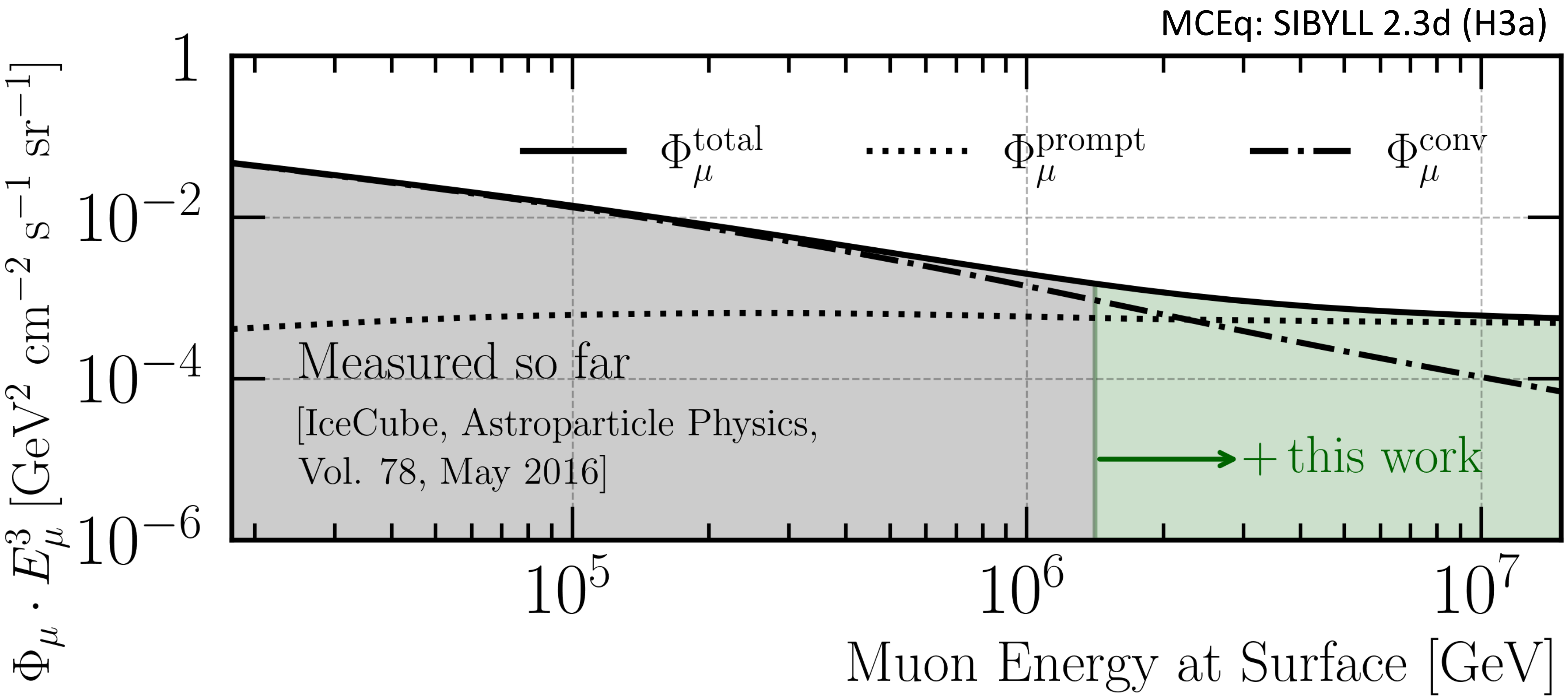
$$\underset{\propto E^{-2.7}}{p} \rightarrow \overbrace{D^+}^{\tau = 1.04\cdot 10^{-12}\,\mathrm{s}} \rightarrow \bar{K}^0 + \underset{\propto E^{-2.7}}{\mu^+} + \nu_\mu$$



Credit: Ludwig Neste

Conventional Muon: Parent is pion or kaon

Goal: Measure Muon Flux at Surface



Overview

Simulation

Dataset
- CORSIKA 77500 Ehist
- SIBYLL 2.3d
- 10 TeV – 100 EeV
- Snowstorm

Validation
- Tag prompt / conv particles
- MCEq comparisons

Reconstruction

DNN reconstructions
- Muon energies
- Muon direction
- Track geometry

Selection

Level 3
- L2 muon filter
- Add labels

Level 4
- 500 TeV bundle energy cut at surface
- Add DNN reconstructions

Level 5
- Data-MC quality cuts

Final Level
- Leadingness cut

Unfolding

Muon spectrum
- Unfold muon flux at surface

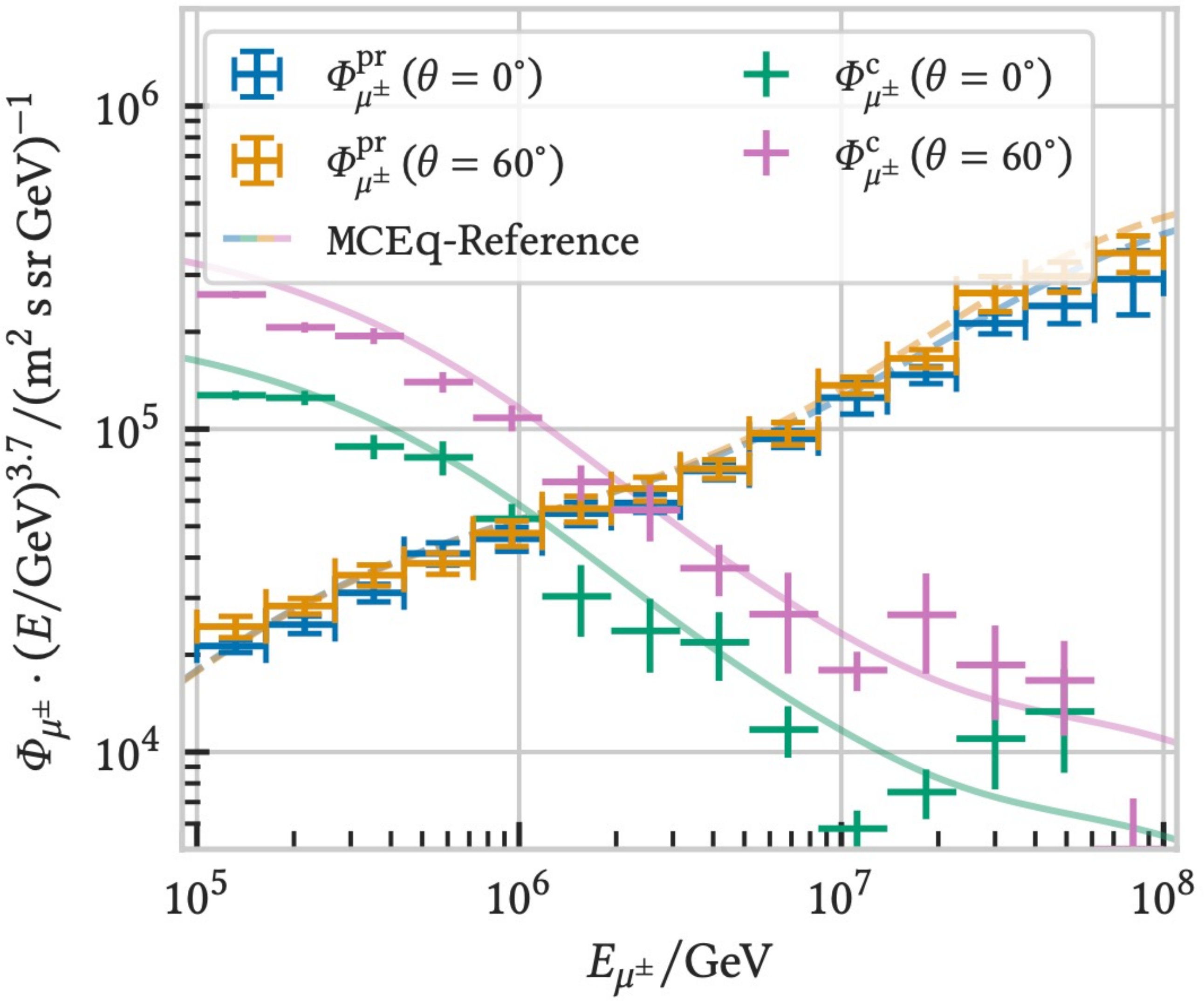
New CORSIKA simulation

with extended history option for information about the parent particles

6 Datasets: 22875, 22774—8

Primary Energy: 600 GeV – 1e11 GeV

CORSIKA 7 vs. MCEq



MCEq: tool to numerically solve the cascade equations that describes the evolution of particle densities as they propagate through a gaseous, dense medium
https://github.com/mceq-project/MCEq

➢ Good agreement for inclusive flux

Python package developed – PANAMA

- Execute CORSIKA 7 (multi core)
- Read DAT files → pandas DataFrames
- Parse EHIST option
- Calculate primary weightings

https://github.com/The-Ludwig/PANAMA

Neste & Gutjahr et al., EPJ C, 85, 759 (2025)

Reconstruction

Machine Learning (CNN)

Input data per DOM

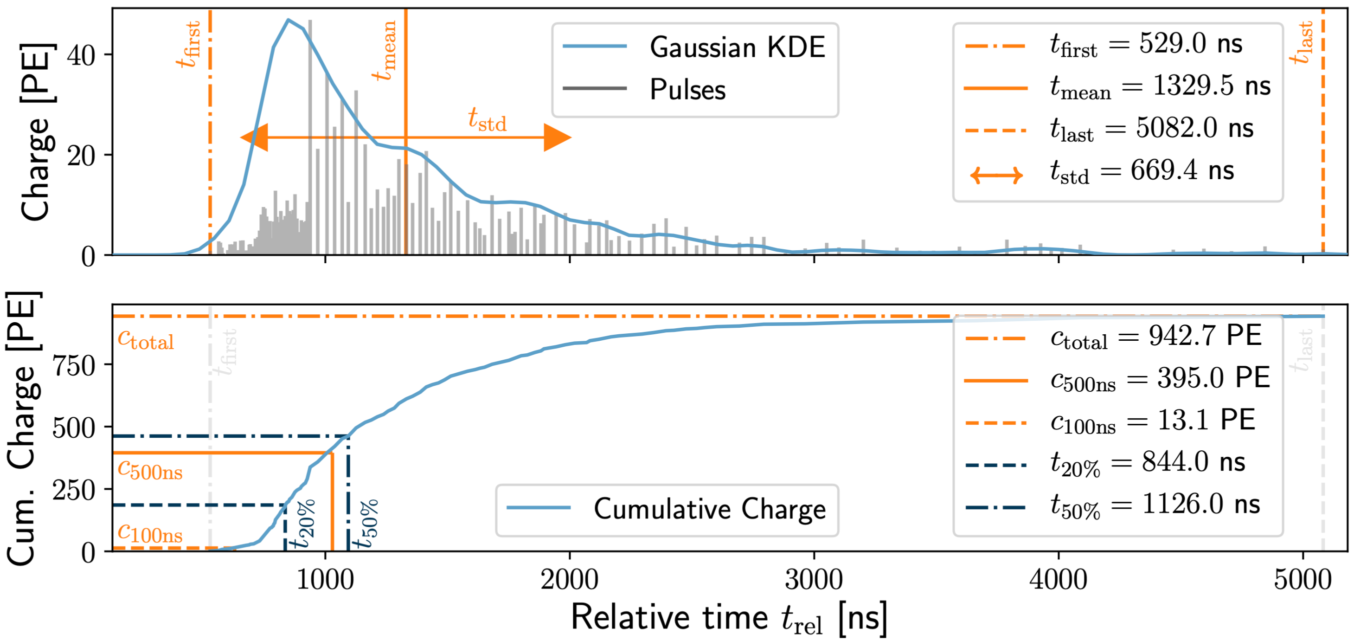
10.1088/1748-0221/16/07/P07041

3 inputs

- c_{total}: Total charge
  - Sum of charge
- t_{first}: Relative time of first pulse
  - Relative to total time offset, calculated as the charge weighted mean time of all pulses
- t_{std}: Standard deviation of first pulse
  - Charge weighted standard deviation of pulse times relative to total time offset

9 inputs

- t_{last}: Relative time of last pulse
  - Relative to total time offset, calculated as the charge weighted mean time of all pulses
- $t_{20\%}$: Relative time of 20% charge
  - Relative to total time offset, calculated as the charge weighted mean time of all pulses
- $t_{50\%}$: Relative time of 50% charge
  - Relative to total time offset, calculated as the charge weighted mean time of all pulses
- t_{mean}: Mean time
  - Charge weighted mean time of all pulses relative to total time offset
- c_{500ns}: Charge at 500ns
  - Sum of charge after 500ns
- c_{100ns}: Charge at 100ns
  - Sum of charge after 100ns

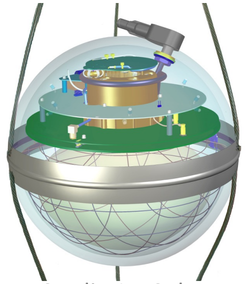


Input pulse cleaning

- 6000 ns

Training data

- Different simulations for robustness



Credit: IceCube

Reconstructed properties

Energy

- entry_energy: Leading muon energy at the detector entry
- bundle_energy_at_entry: Muon bundle energy at the detector entry
- muon_energy_first_mctree: Leading muon energy at surface
- bundle_energy_in_mctree: Muon bundle energy at surface

Network 1

Network 0: Muon bundle energy at surface (fast pre-cut)

Track geometry

- Length: Propagation length of muon in the ice
- LengthInDetector: Propagation length of muon in the detector
- center_pos_x: Closest x position of muon to center of the detector
- center_pos_y: Closest y position of muon to center of the detector
- center_pos_z: Closest z position of muon to center of the detector
- center_pos_t: Time of closest approach to the center of the detector
- entry_pos_x: x position of muon at the detector entry
- entry_pos_y: y position of muon at the detector entry
- entry_pos_z: z position of muon at the detector entry
- entry_pos_t: Time of muon at the detector entry

Network 2

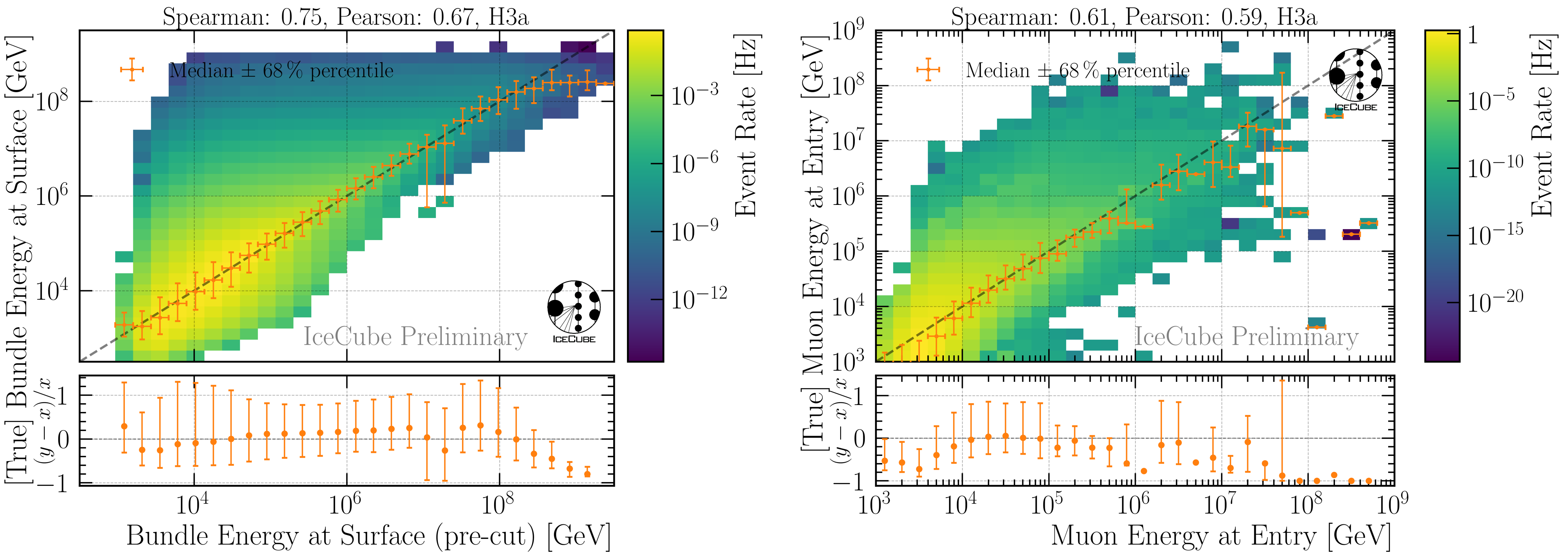
Direction

- zenith: Zenith angle of muon
- azimuth: Azimuth angle of muon

Network 3

pascal.gutjahr@tu-dortmund.de

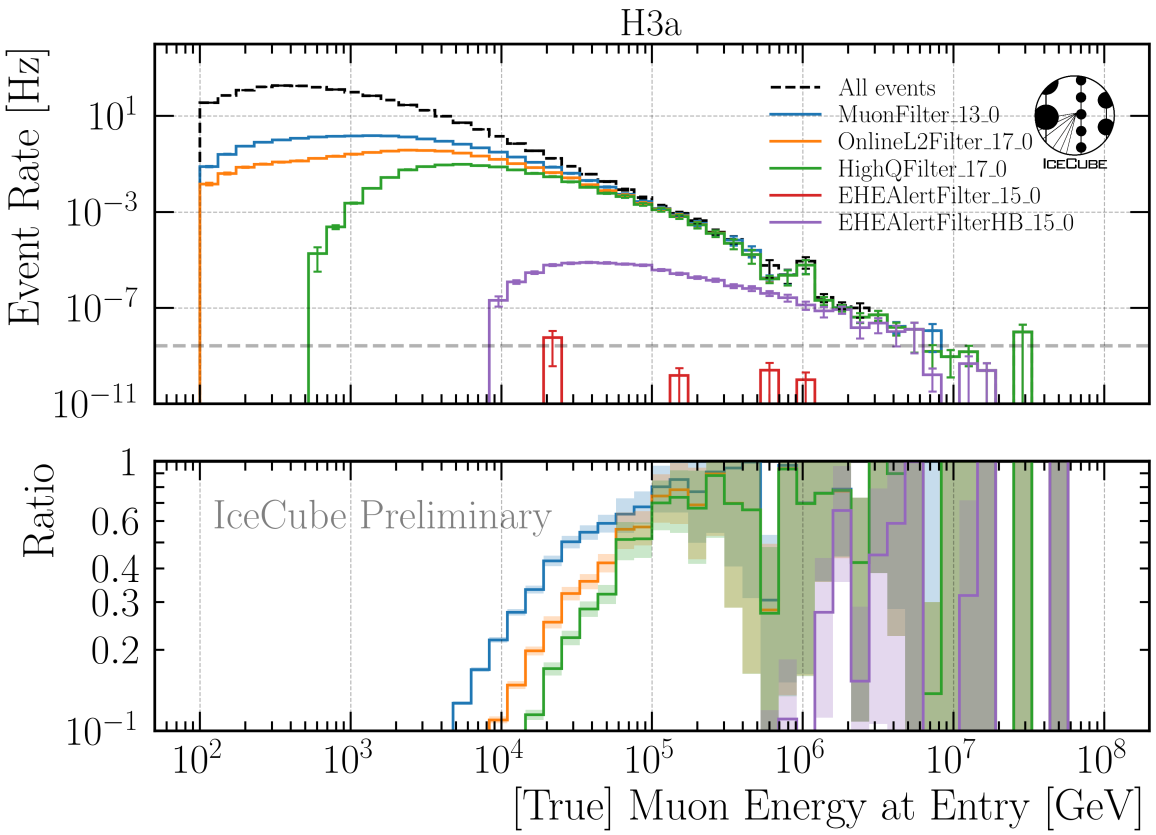
Energy Reconstructions

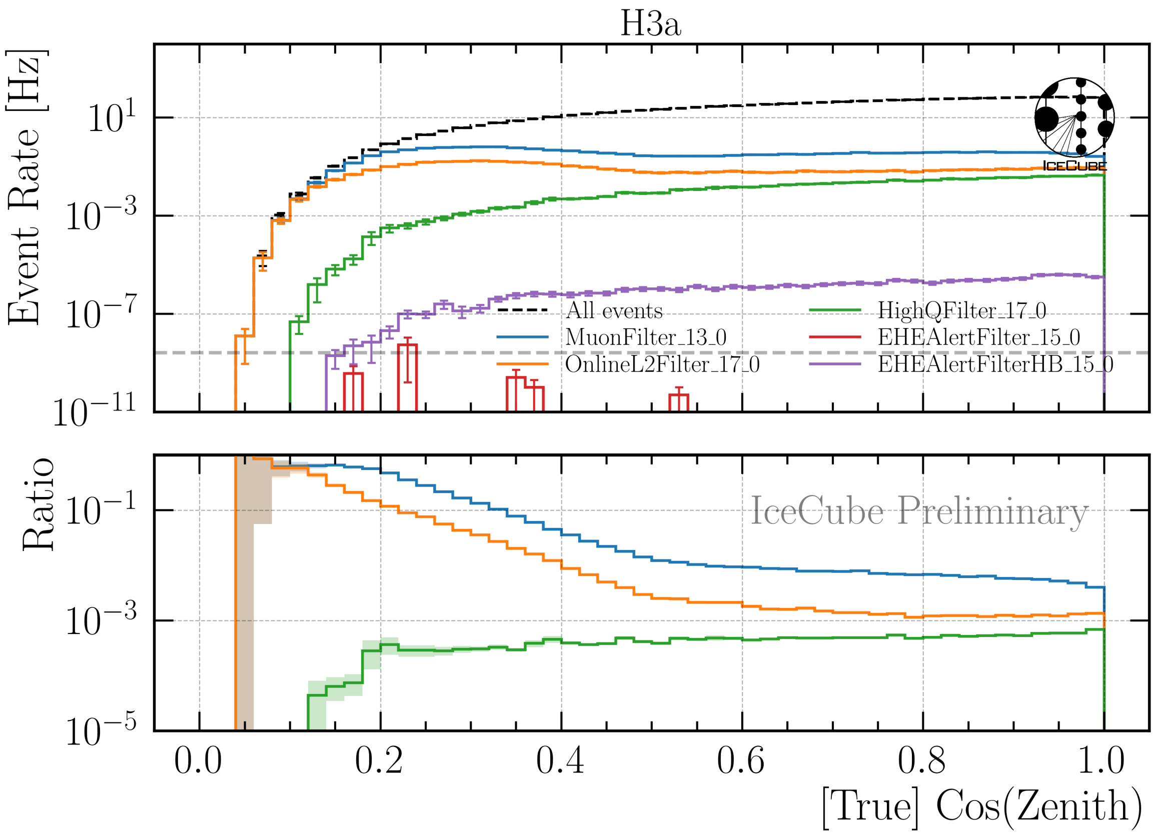


- Sufficient energy reconstructions
- Tight 68 % intervals, with outliers

Selection

Level 3: Muon Filter

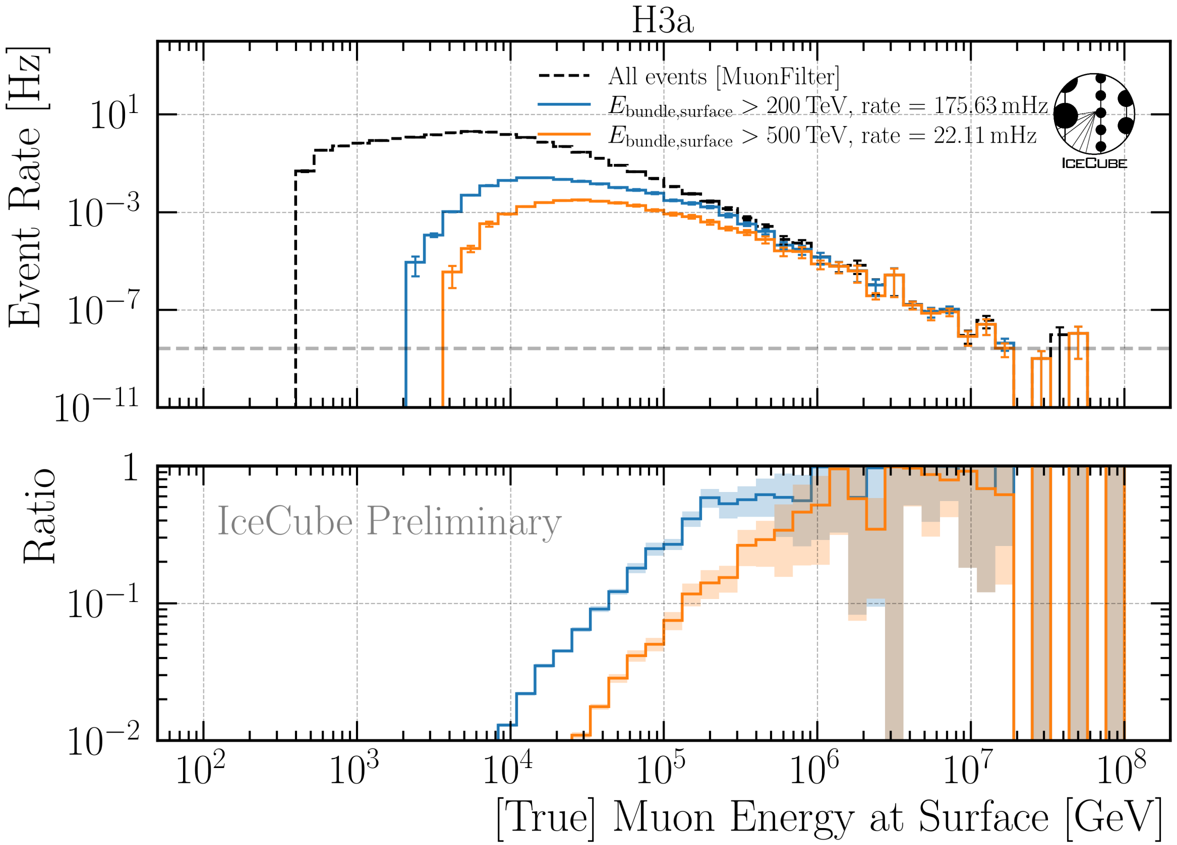


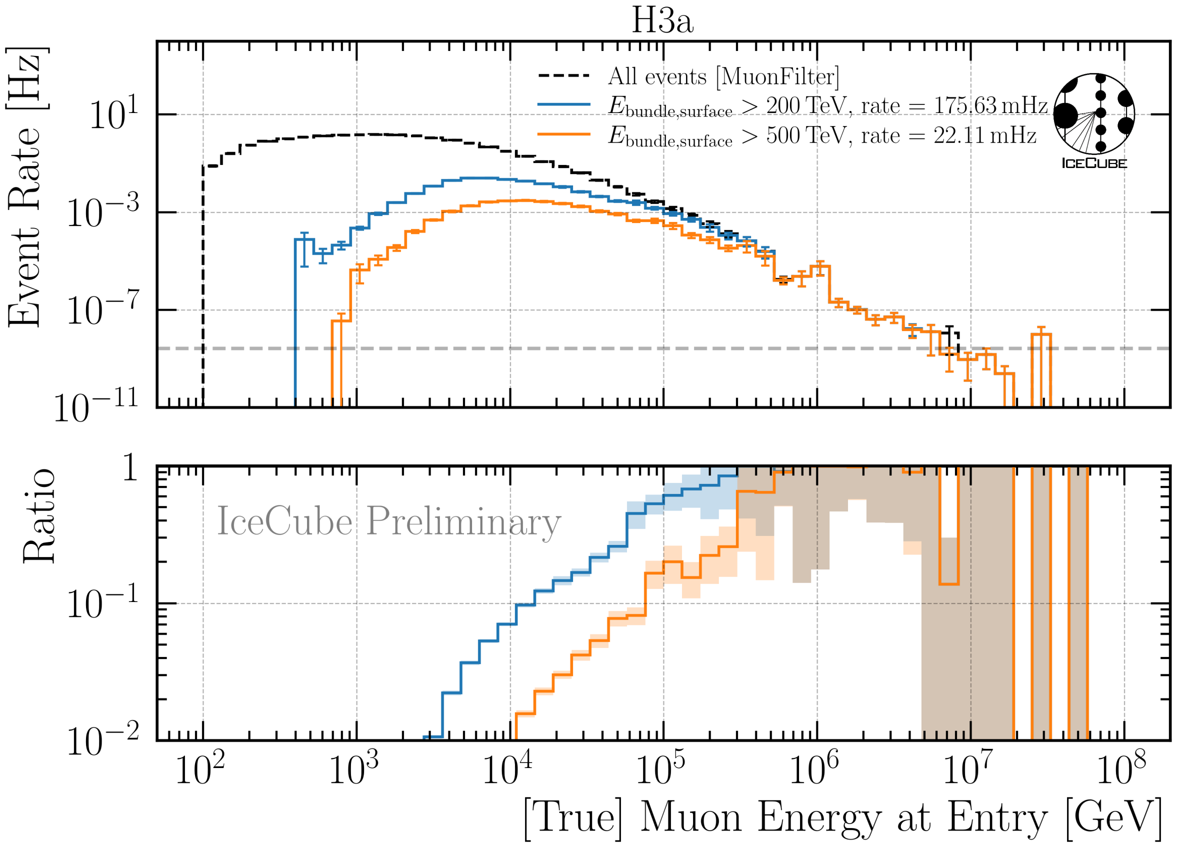


Muon filter: zenith-dependent charge and quality cut

➢ Choose muon filter to select as many high-energy muons as possible

Level 4: Energy Cut

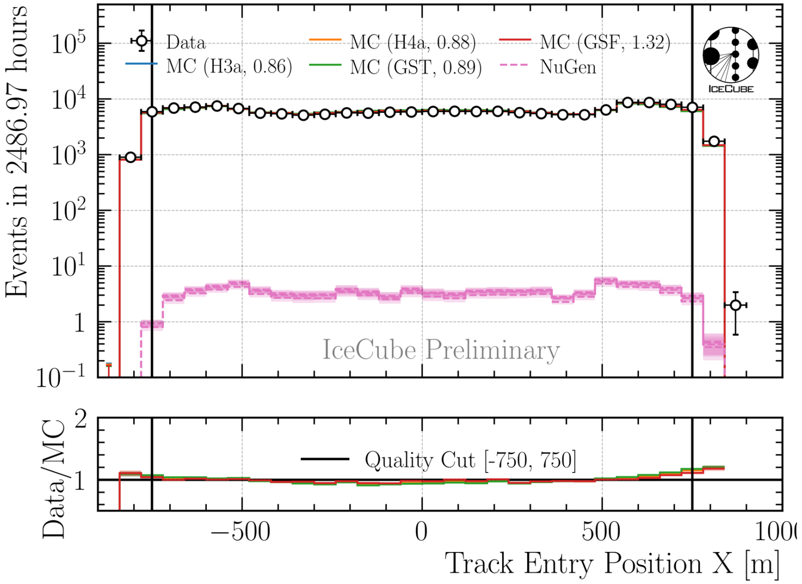


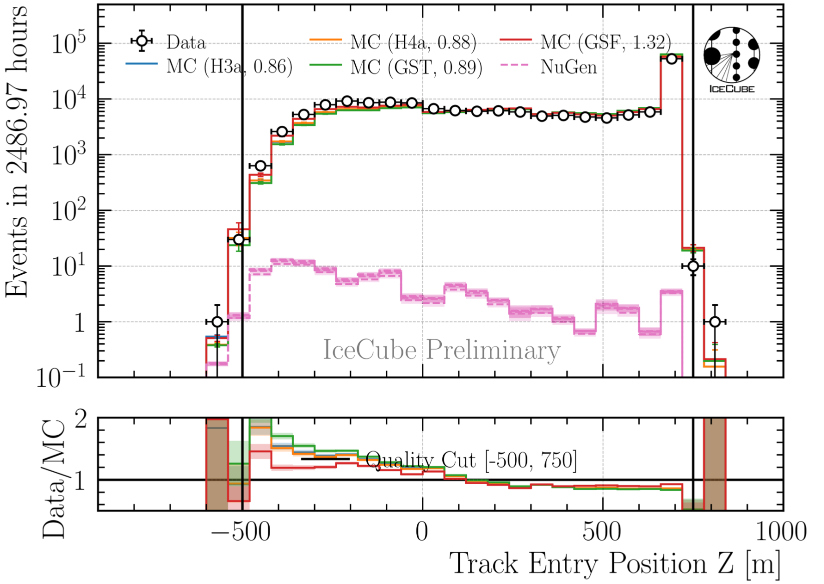


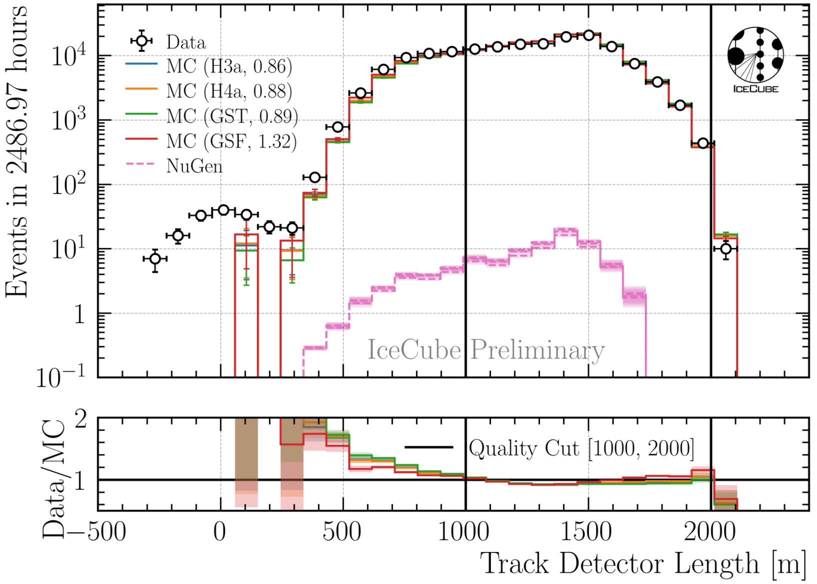
- 6 billion events expected in 10 years → computationally not feasible
- focus on high-energetic events
- Remove low-energy muons: bundle energy at surface > 500 TeV

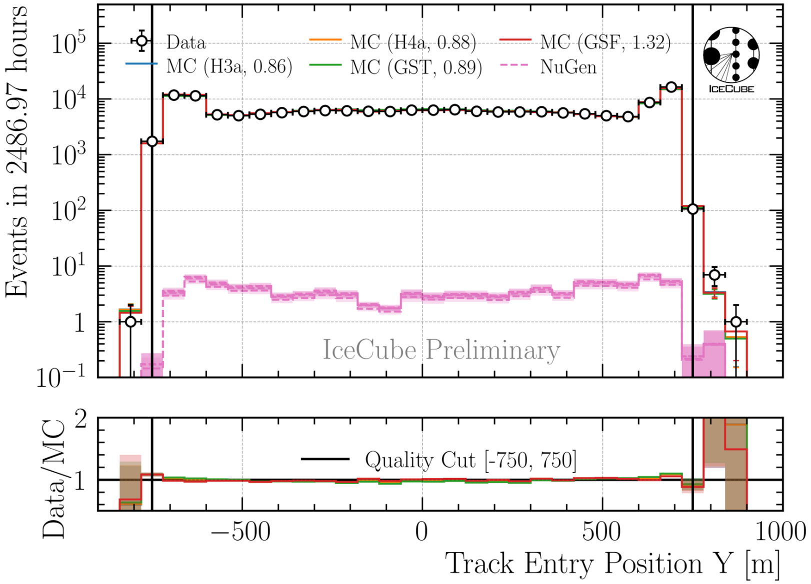
Level 5: Data/MC Quality Cuts

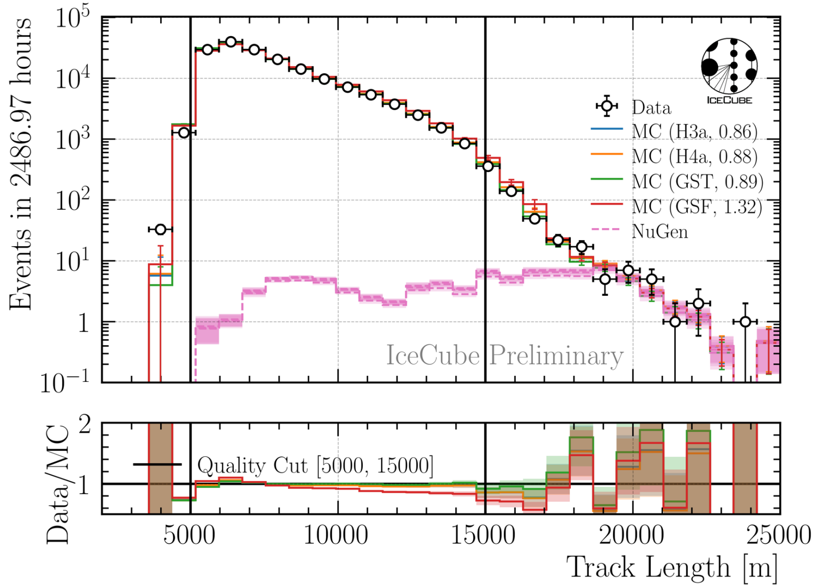
- Remove outliers
- Depth—dependent slope (no analysis relevance)
- CR—model impact on energy reconstruction
- 23 quality cuts in total

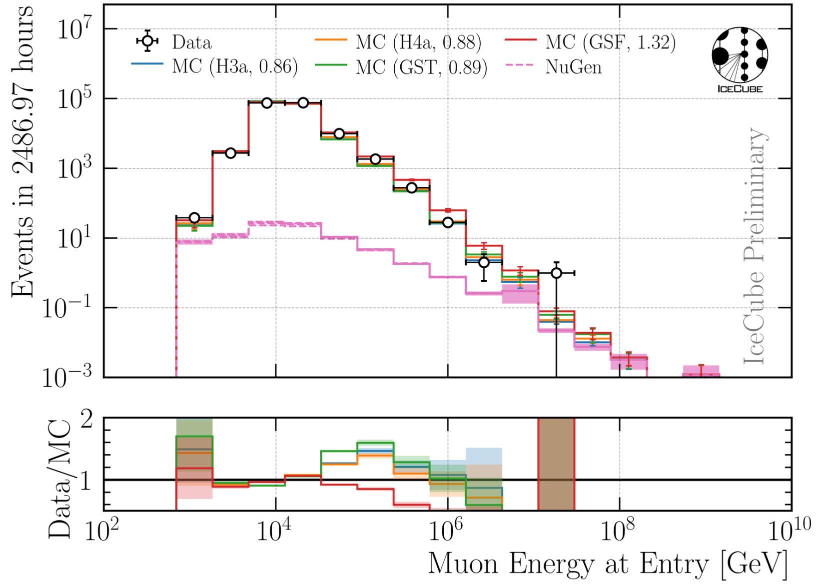








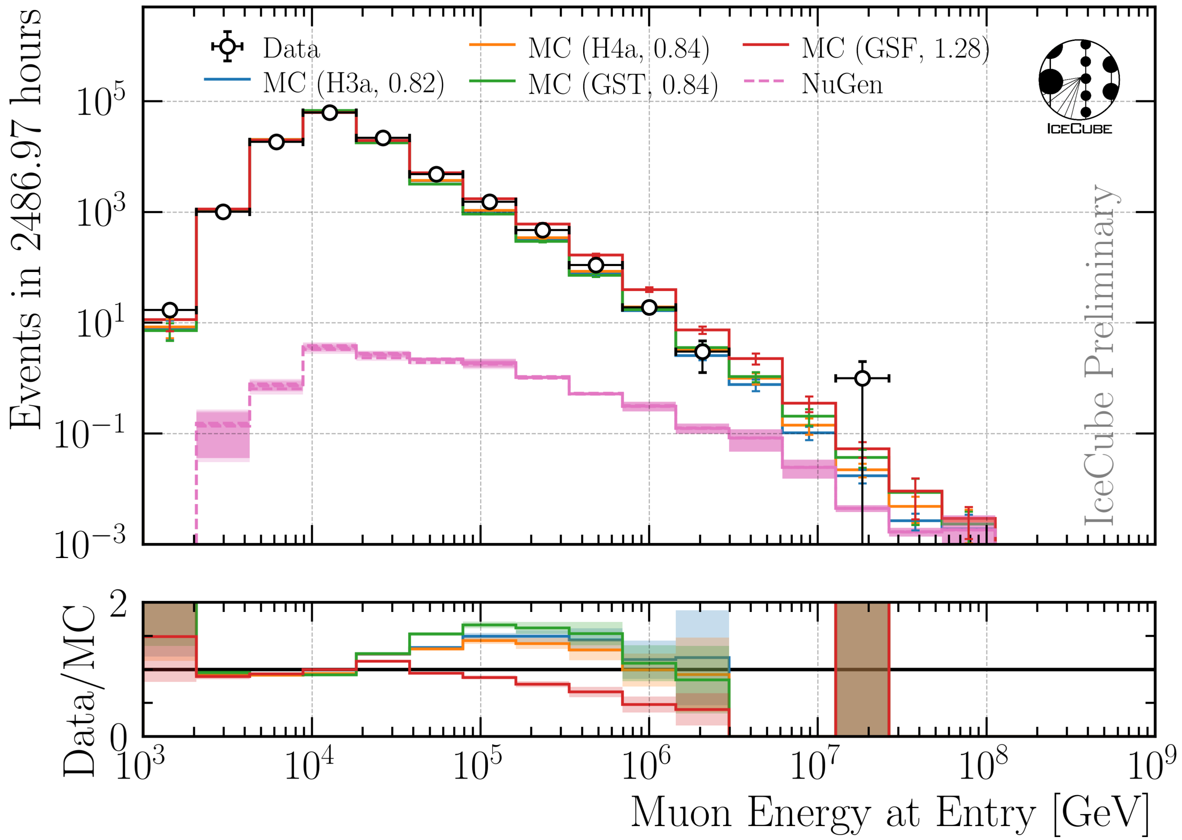




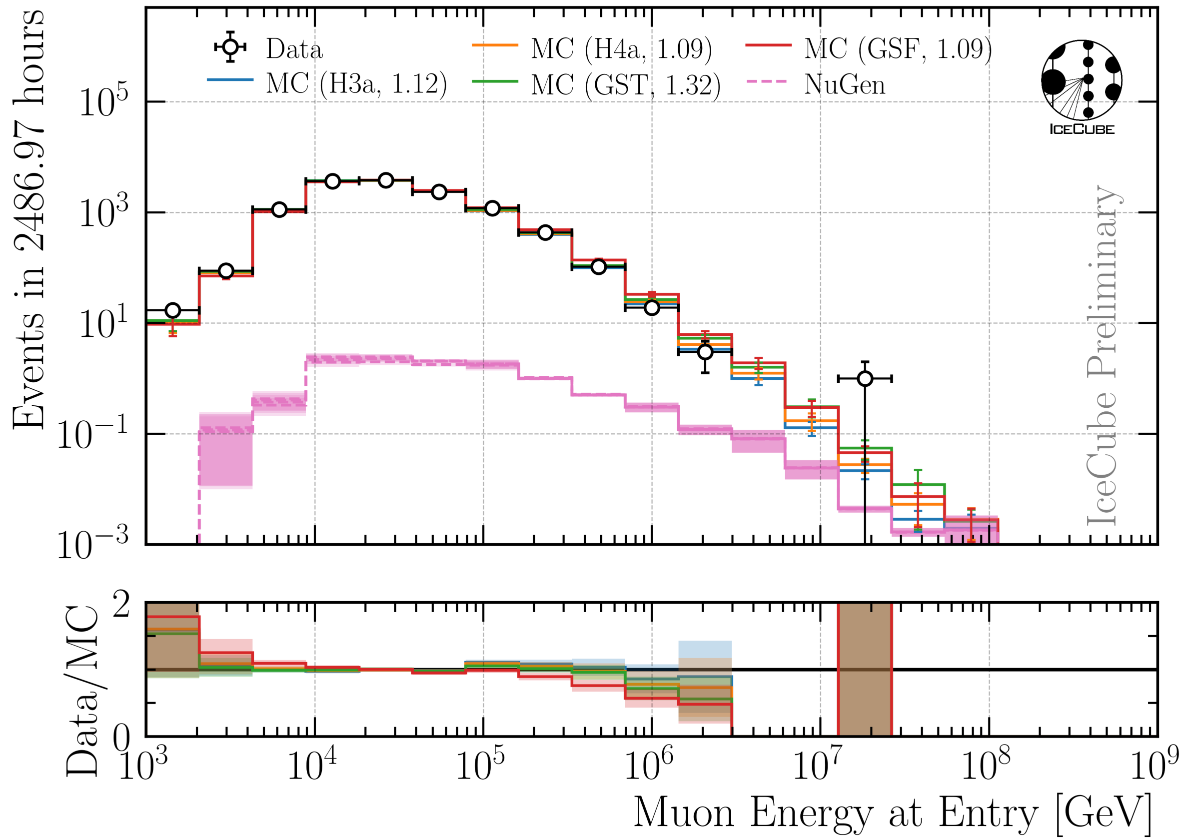
Final Level: Leadingness > 40 %

$$L = \frac{\text{Leading Energy at Entry}}{\text{Bundle Energy at Entry}}$$

No leadingness cut



Leadingness > 40 %



➢ Improve Data/MC by leadingness cut

pascal.gutjahr@tu-dortmund.de

Unfolding

Unfolding in a nutshell

measured proxy

detector response

background

$$g(y) = \int_{E_0}^{E_1} A(E_\mu, y) f(E_\mu)\, \mathrm{d}E_\mu + b(y)$$

true energy distribution

1. Discretized form: $\overbrace{\vec{g} = A\vec{f}}^{\text{folding}} \leftrightarrow \overbrace{\vec{f} = A^{-1}\vec{g}}^{\text{unfolding}}$

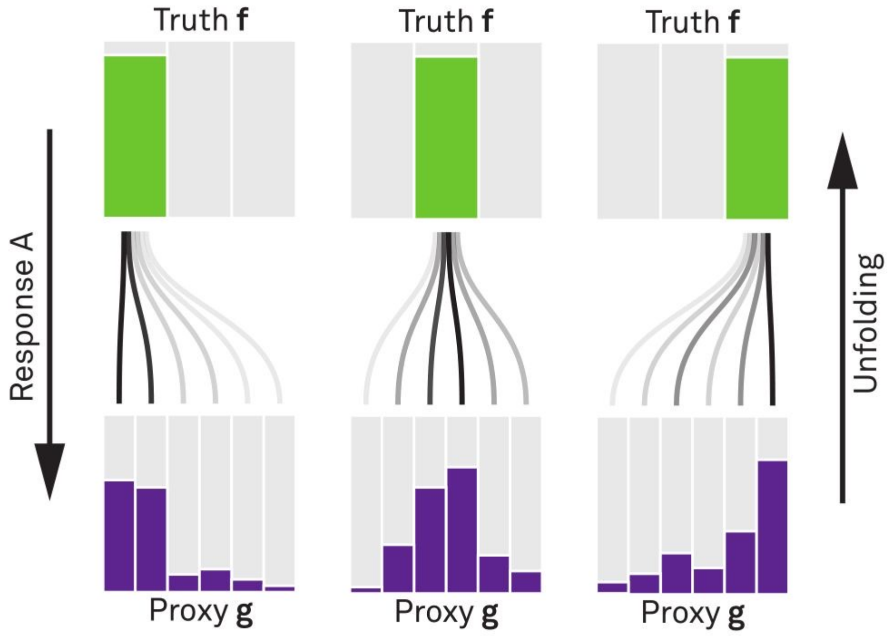
2. Maximum likelihod method:

$$\mathcal{L}(\vec{g}|\vec{f}) = \prod_{j=1}^{M} \frac{\lambda_j^{g_j}}{g_j!} \exp(-\lambda_j)$$

$$= \prod_{j=1}^{M} \frac{(A\vec{f})_j^{g_j}}{g_j!} \exp(-(A\vec{f})_j)$$

3. Thikonov regularization:

$$t(\vec{f}) = -\frac{1}{2}(C\vec{f})^T (\tau 1)^{-1} (C\vec{f})$$



Credit: T. Hoinka

4. Maximize $\log(\mathcal{L}(\vec{g}|\vec{f})) + t(\vec{f})$ with respect to $\vec{f}$ using Markov Chain Monte Carlo (MCMC) or Minuit

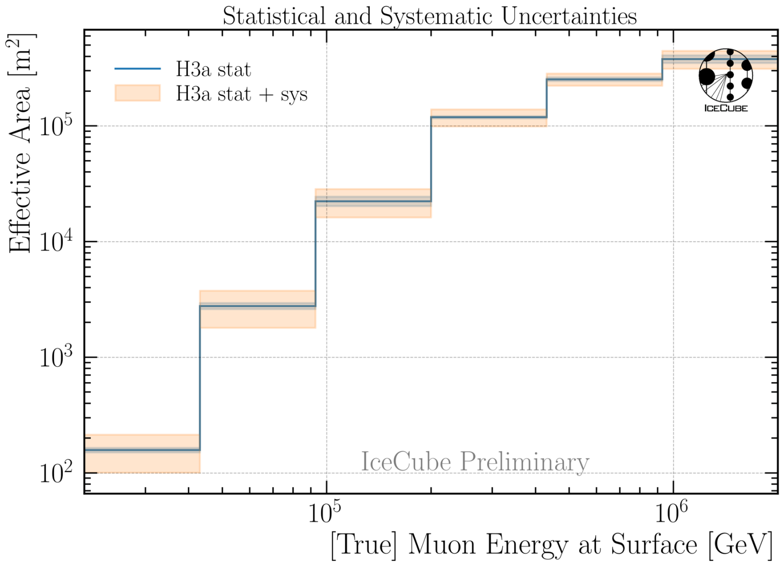
GitHub **funfolding** by M. Börner

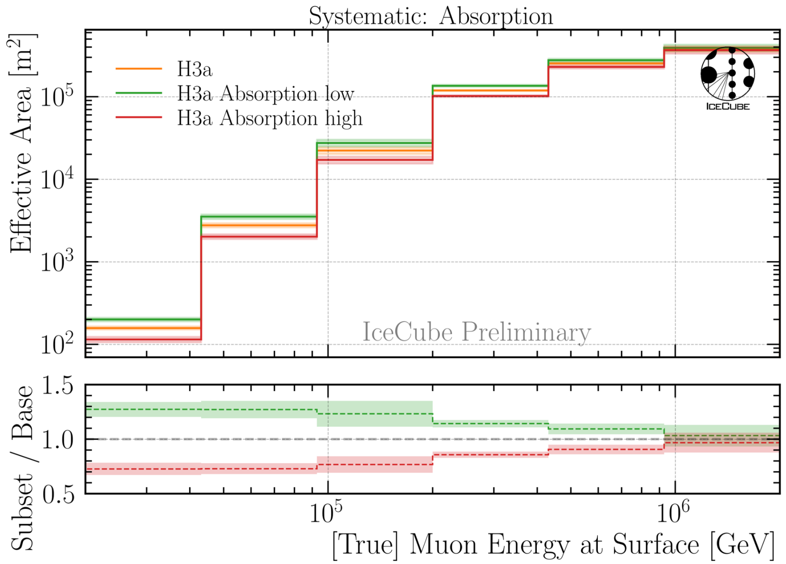
Muon Flux Unfolding

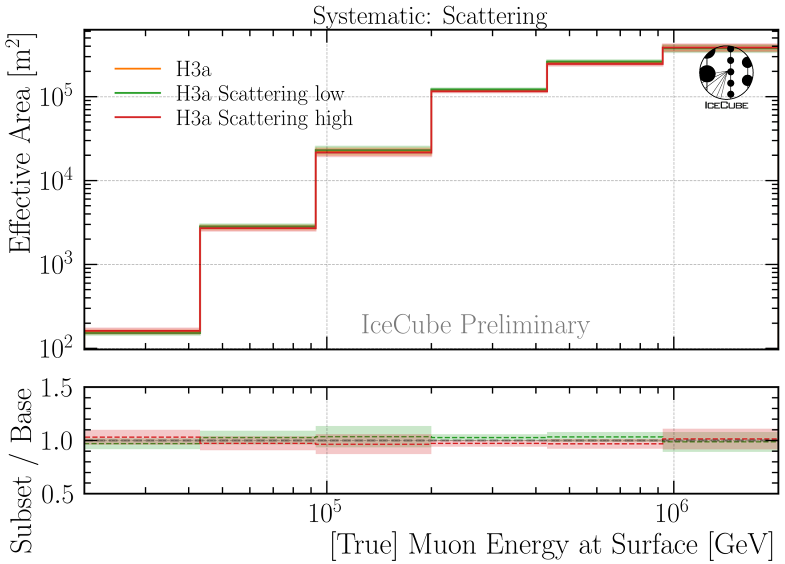
- ❑ Effective area
- ❑ Systematics
- ❑ Proxy that correlates with target
- ❑ Regularization
- ➢ Unfolding

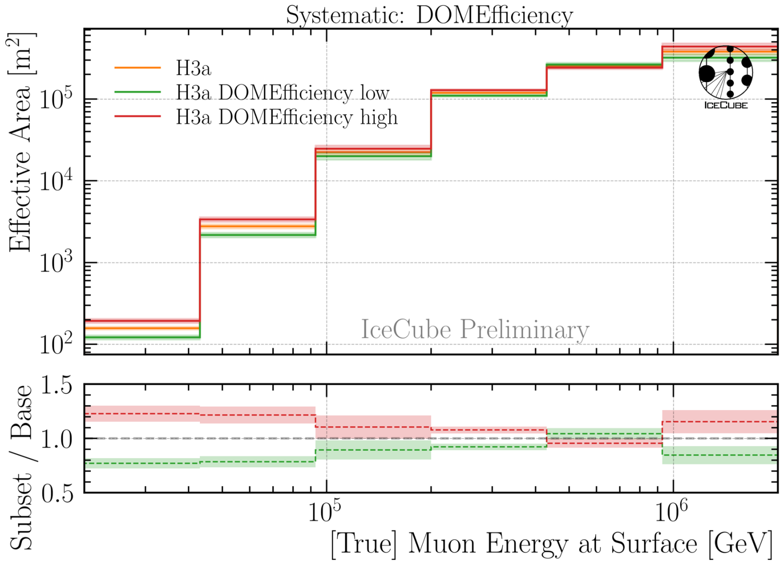
Ice & Detector Systematics

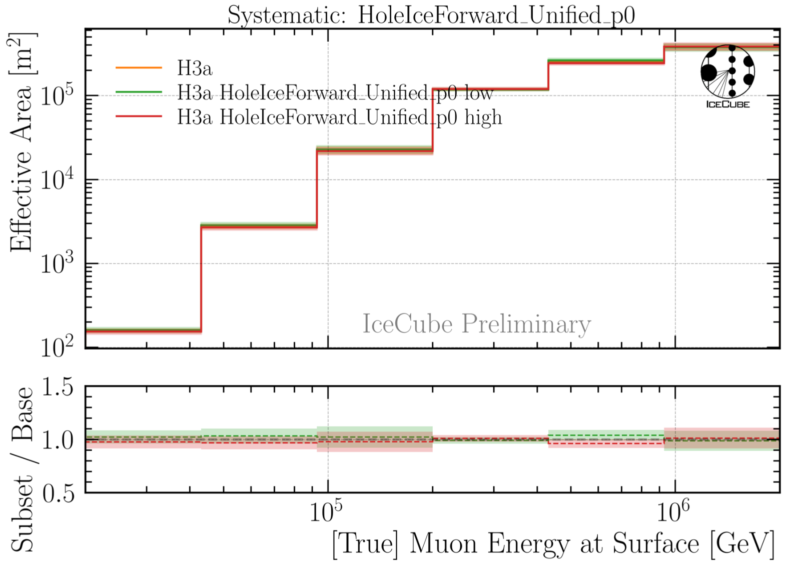
$$\sigma_{\mathrm{sys},i} = \max(A_{\mathrm{low}}, A_{\mathrm{high}}) - A_{\mathrm{base}}$$

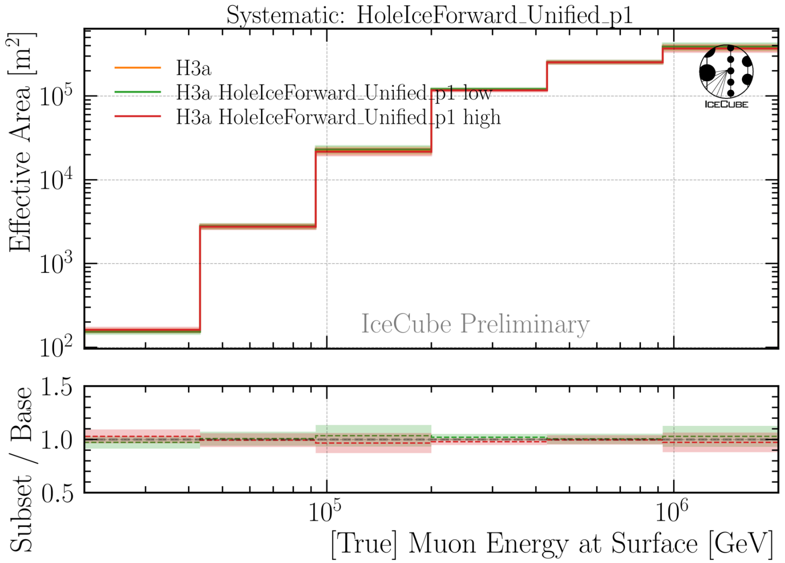










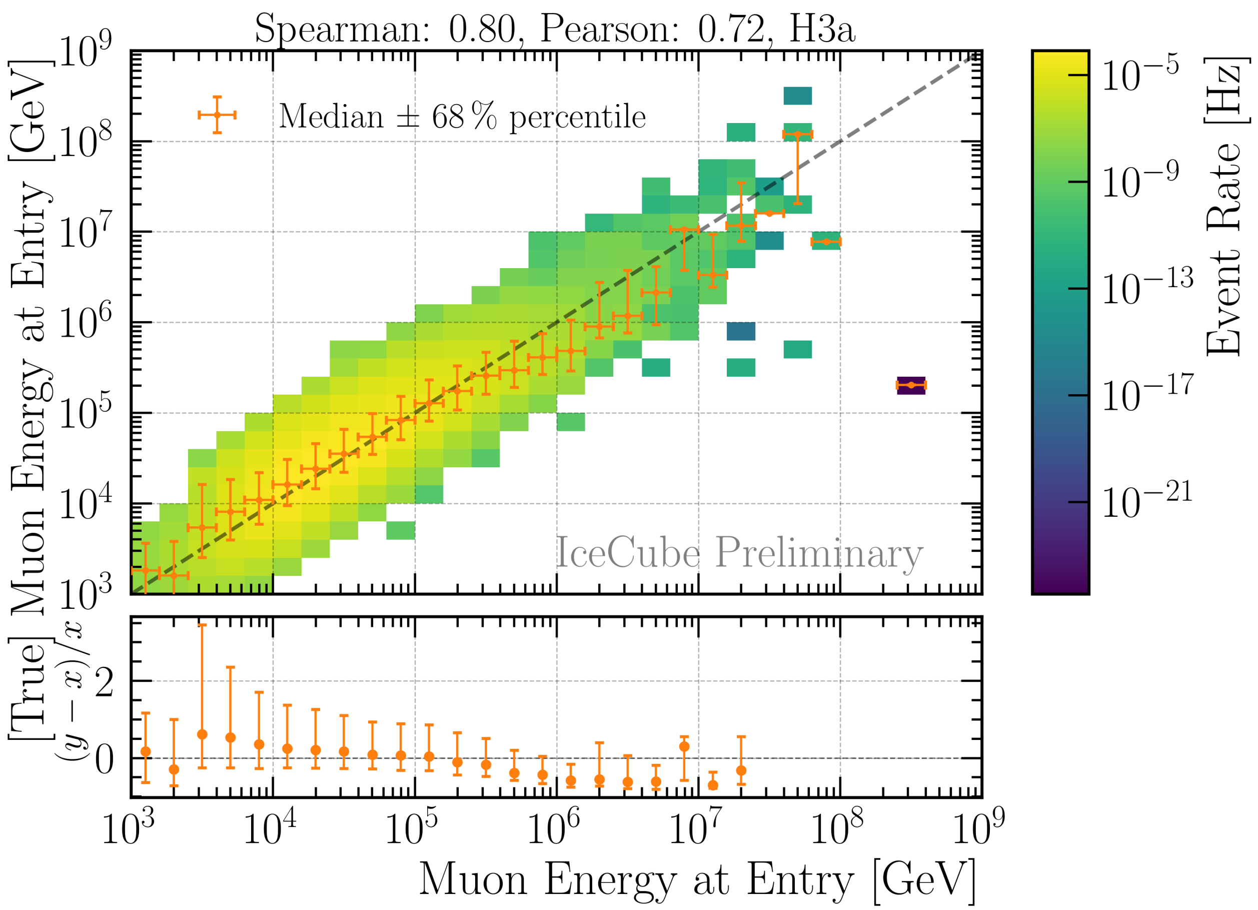


- Baseline: entire set
- Subset: above/below center

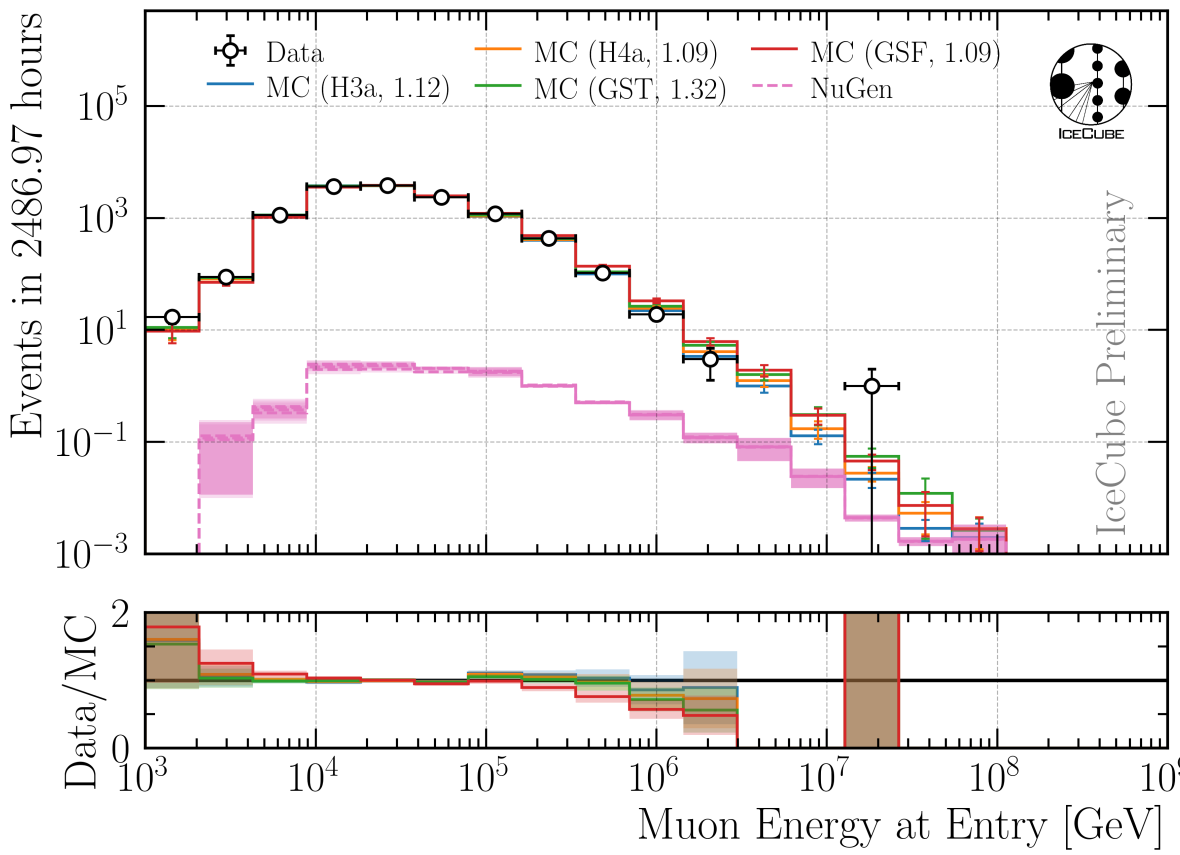
$$\sigma_{\mathrm{tot}} = \sqrt{\sigma_{\mathrm{stat}}^2 + \sum \sigma_i^2}\,,\ i\text{: scat, abs, DOME, holeice p0, p1}$$

Reconstruction and Data—MC: Leading Muons

- Good reconstruction of leading muon energy → proxy

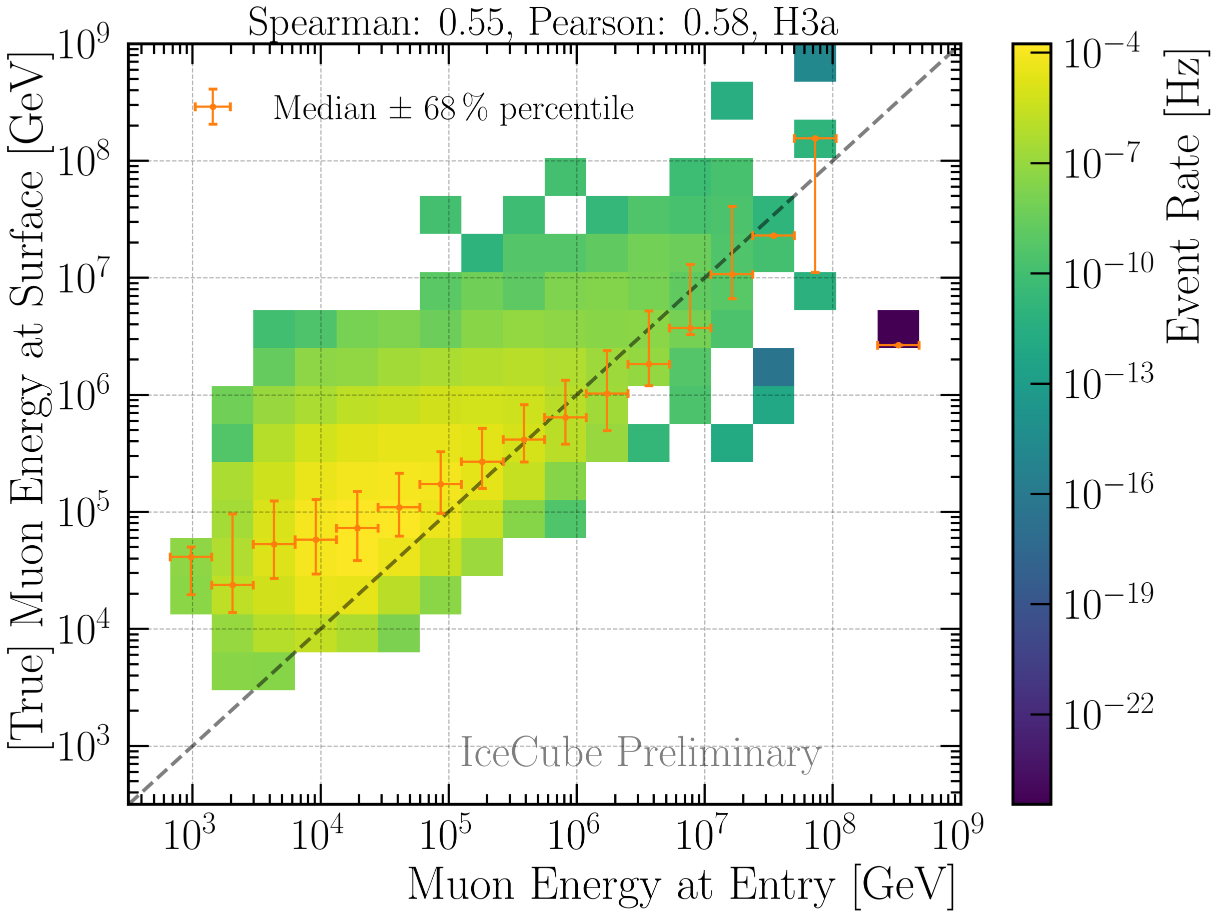


- Good data—MC agreement
- Global offset → upscale MC by 12%

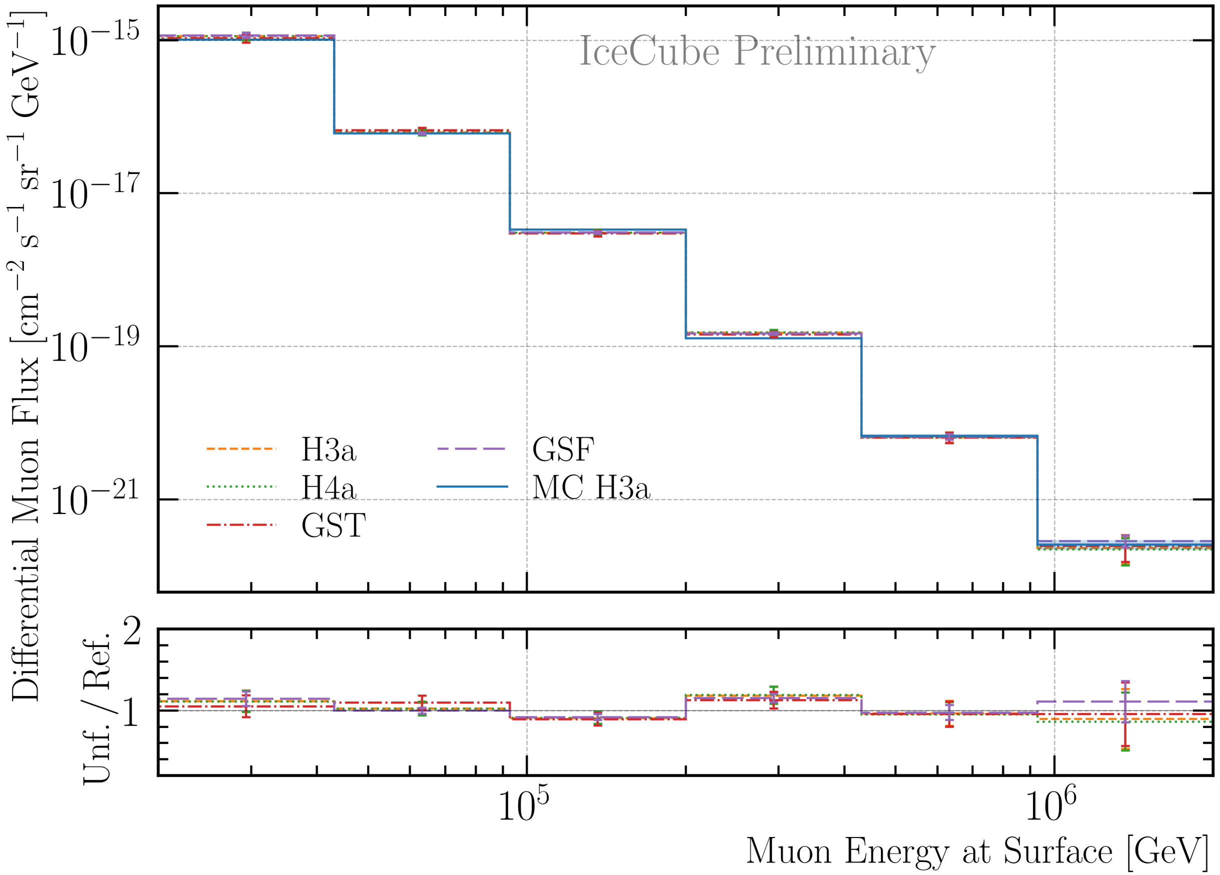


Correlation and Robustness Test: Leading Muons

- Correlation between proxy and target



- Build unfolding matrix on H3a, H4a, GST & GSF
- Unfold H3a as test set
- Results are within uncertainties → robust

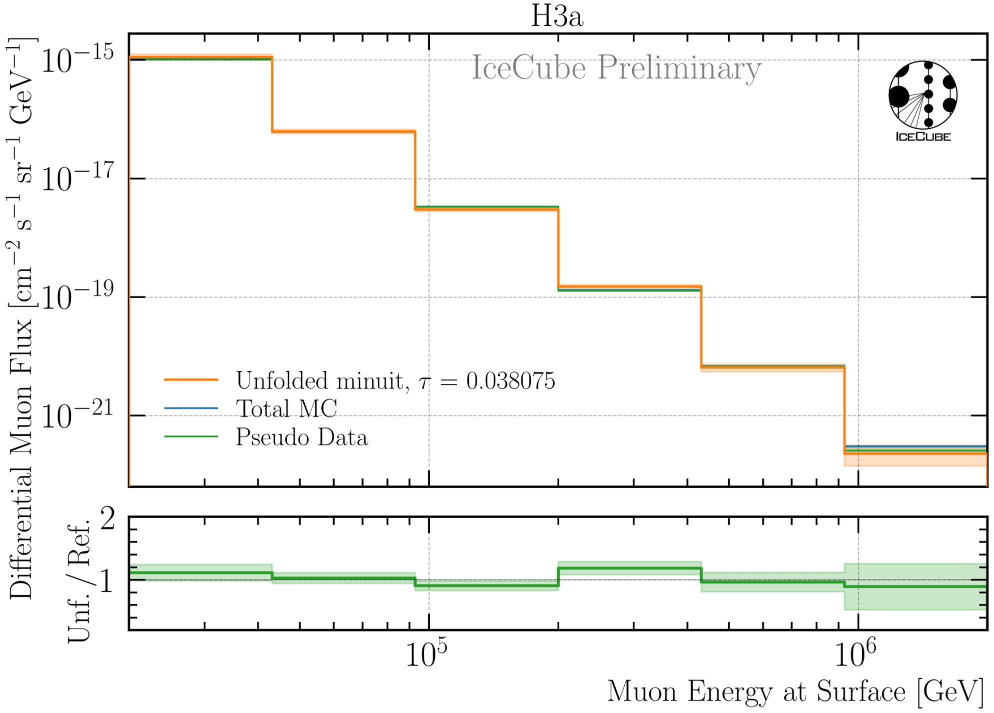


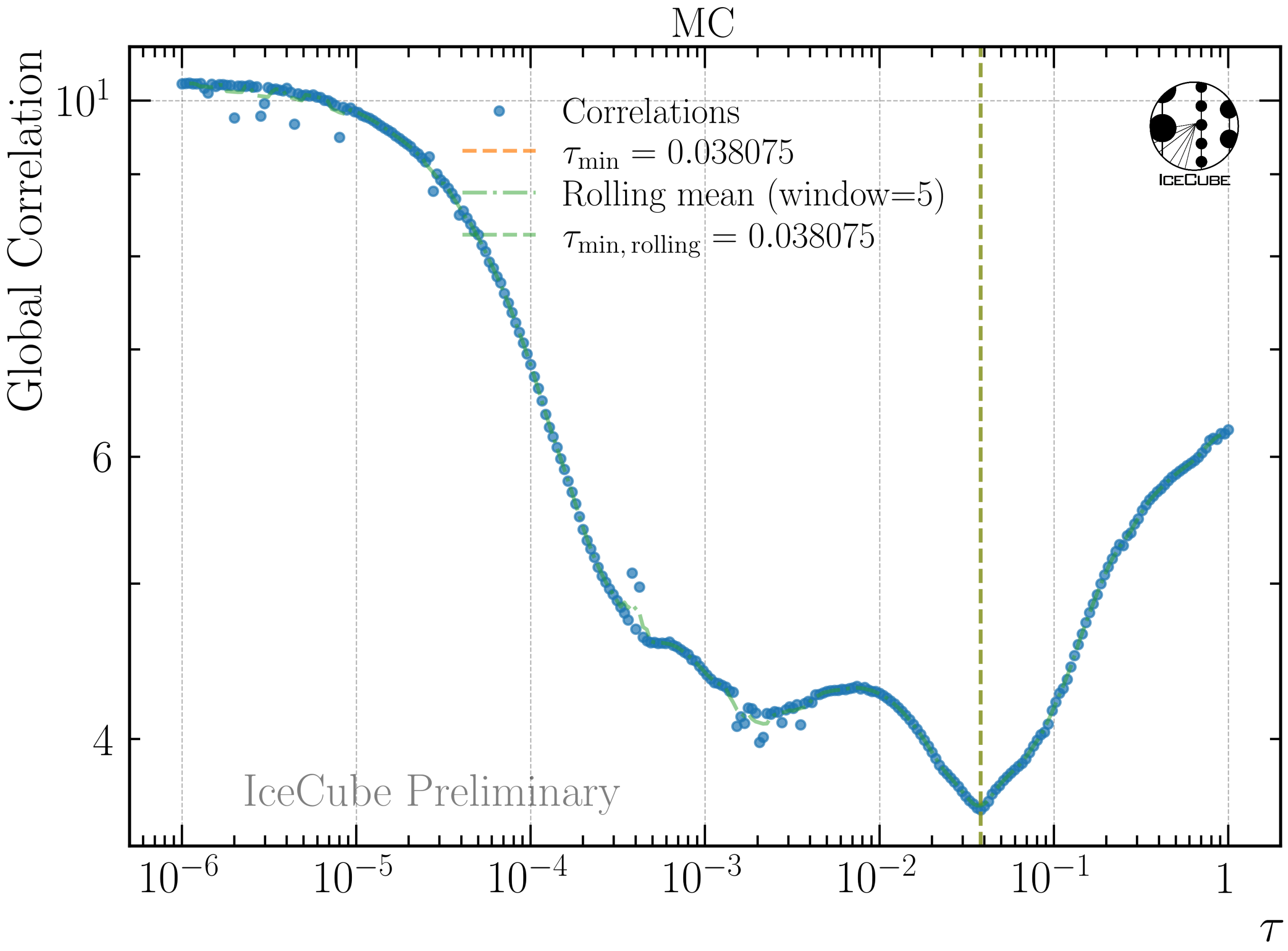
Burnsample

Determine Regularization

- Find regularization τ with minimal bin-to-bin correlation
- LLH minimization (unfolding) provides full covariance matrix V

➢ Minimize global correlation $\rho = \sum_{j > i} V_{ij}$





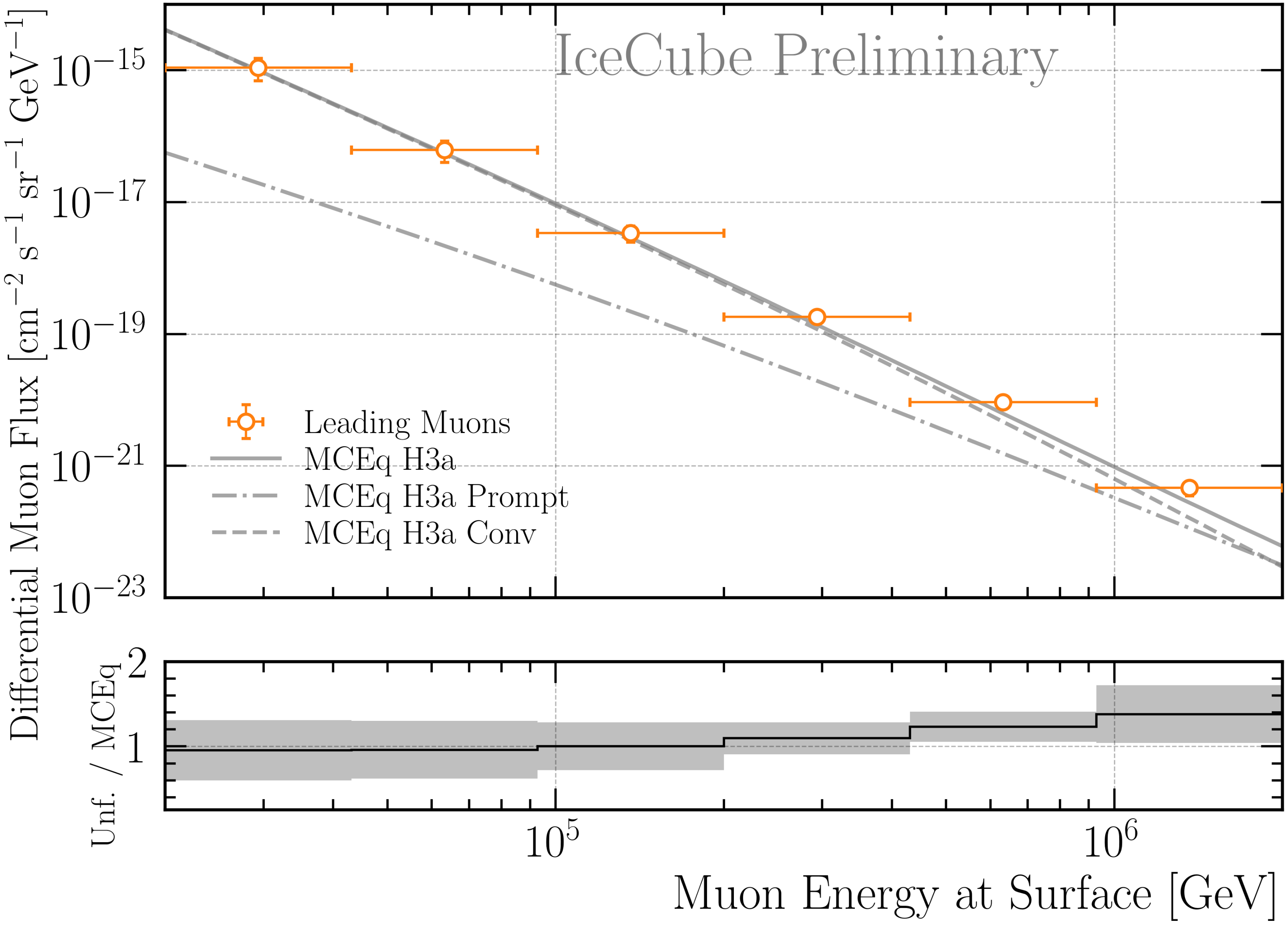
Unfold MC truth → ratio should be 1

Burnsample

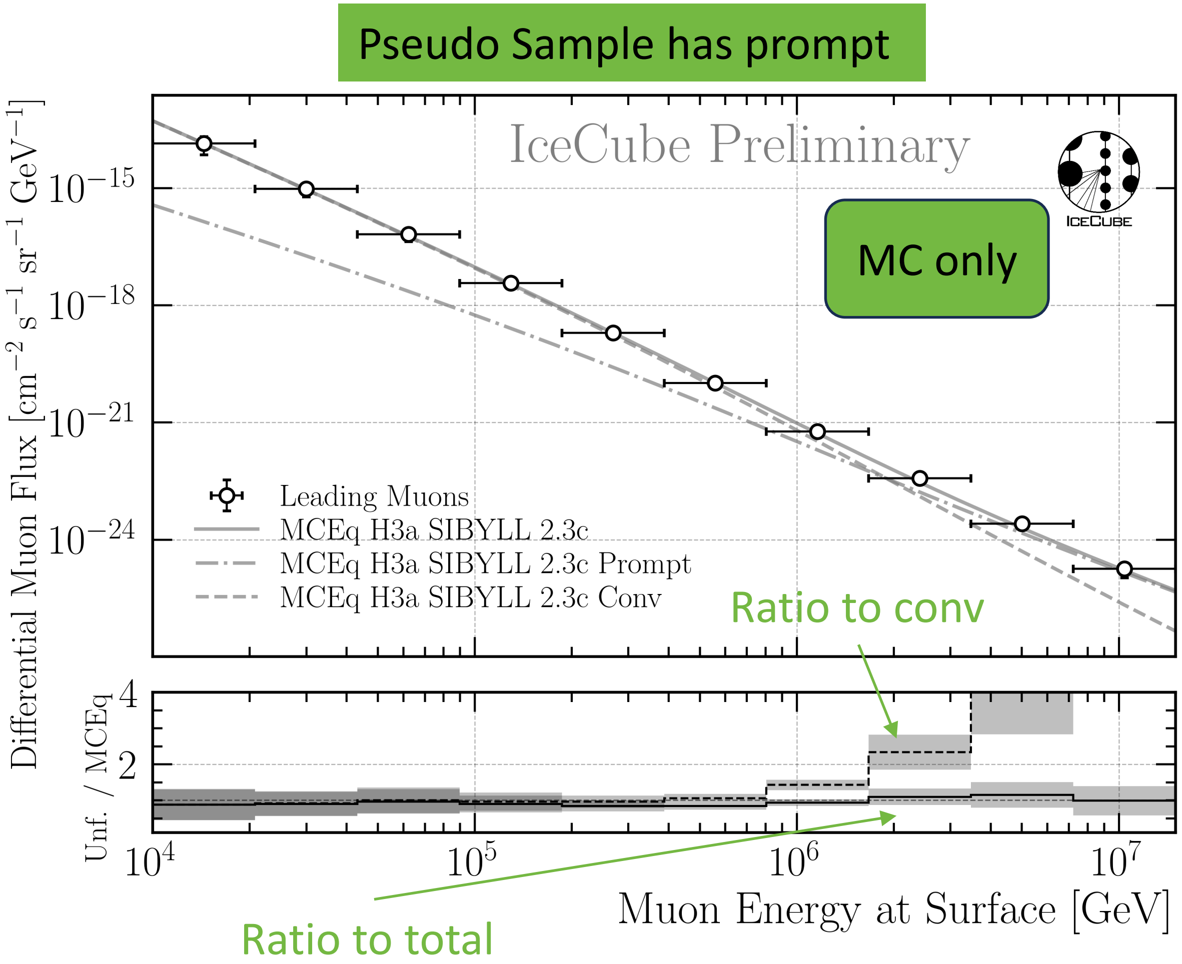
Burnsample Unfolding

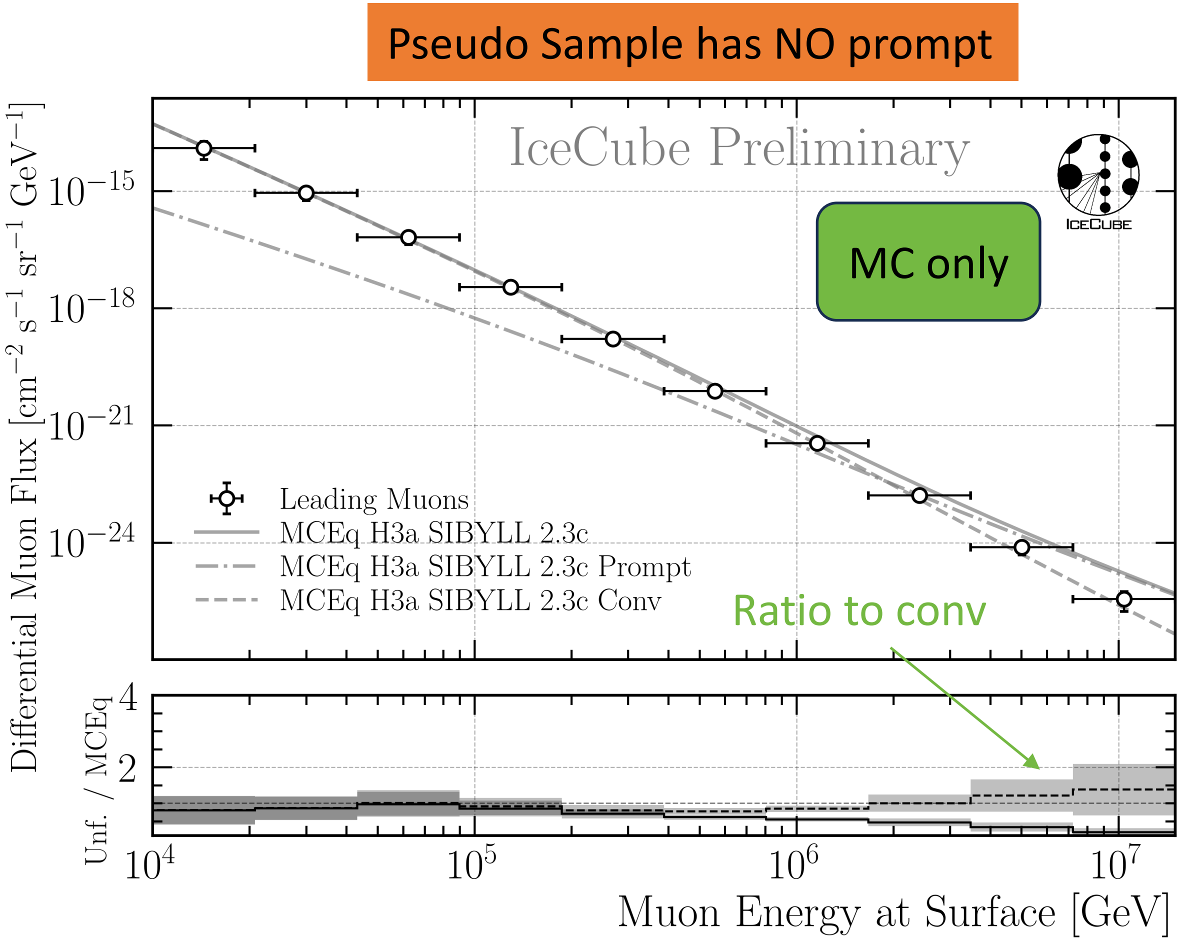
Leading muons

- 2487 h IceCube data
- 12754 events
- Agrees with MCEq



Prediction: 12 Years of IceCube Data Unfolding



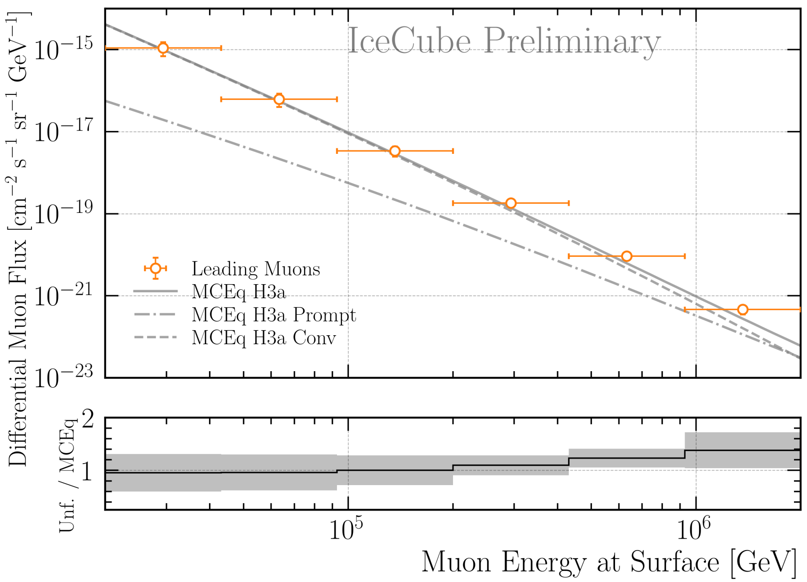


- Unfolding is sensitive to the prompt component

*with systematic uncertainties from effective area

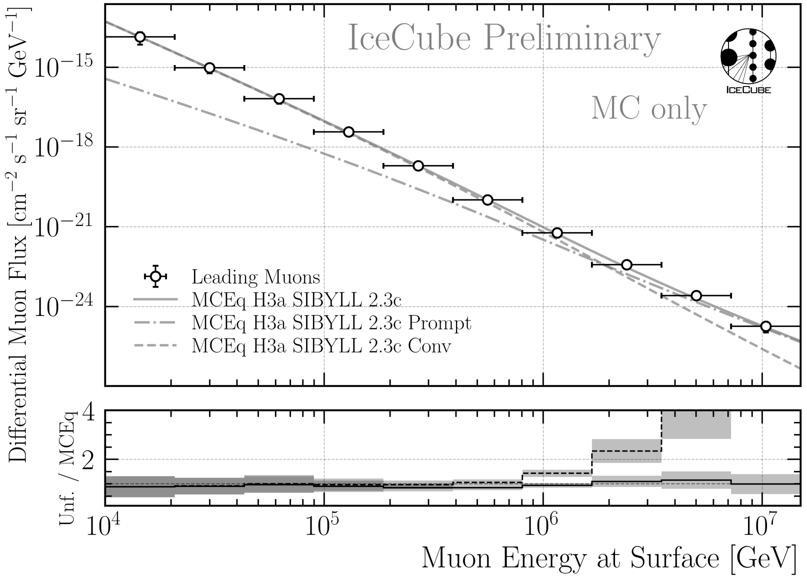
Conclusion & Outlook

- New CORSIKA EHIST simulations
  - Prompt tagging agrees with MCEq
- DNN-based reconstructions
- Leading muon selection
- Burnsample unfolding agrees with MCEq prediction



Outlook

- Estimate unfolding uncertainty
- Unfold 12 years of data



Future

- Forward folding to fit prompt and conv normalization

Thank you for your attention!

Backup

Unblinding Proposal

Unblinding plans

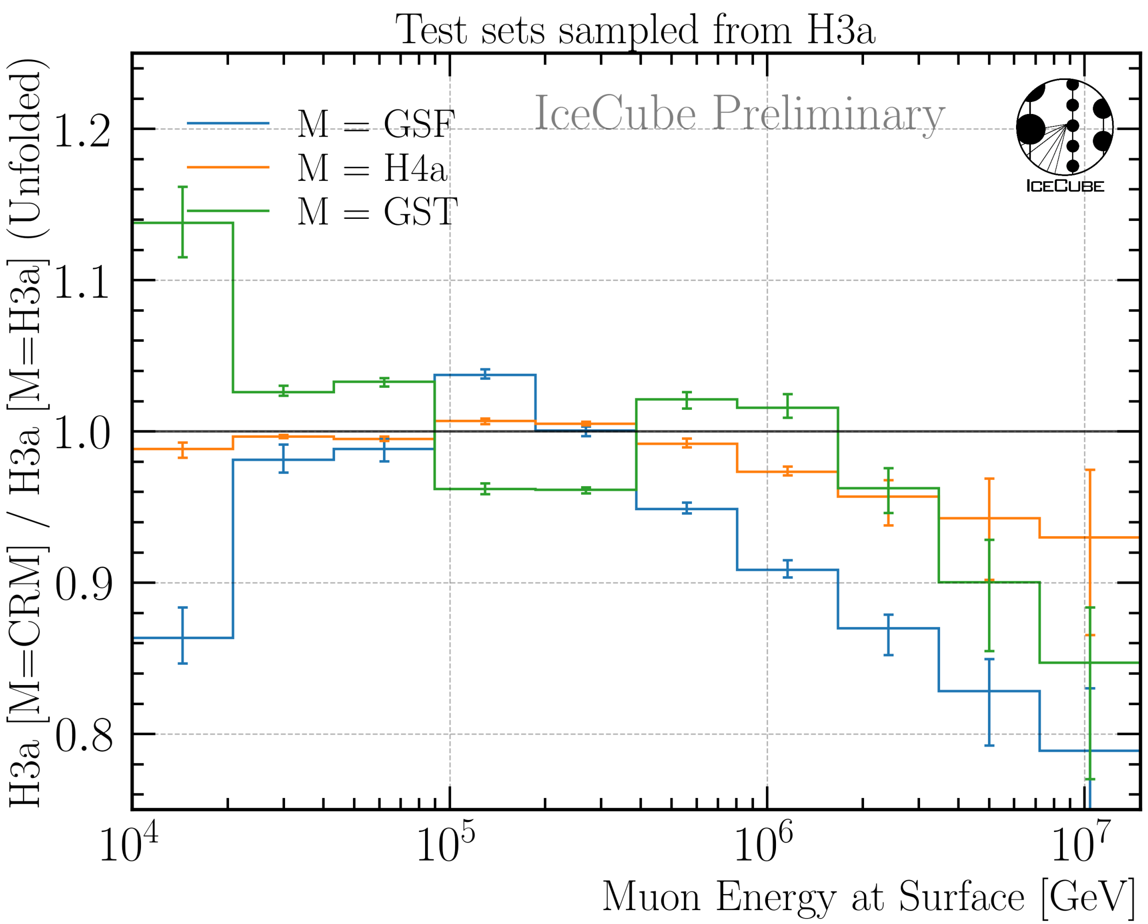
- Unfold 12 years of data for IC86 from 2011 to 2022
- Determine regularization strength on data
- Perform unfolding on data

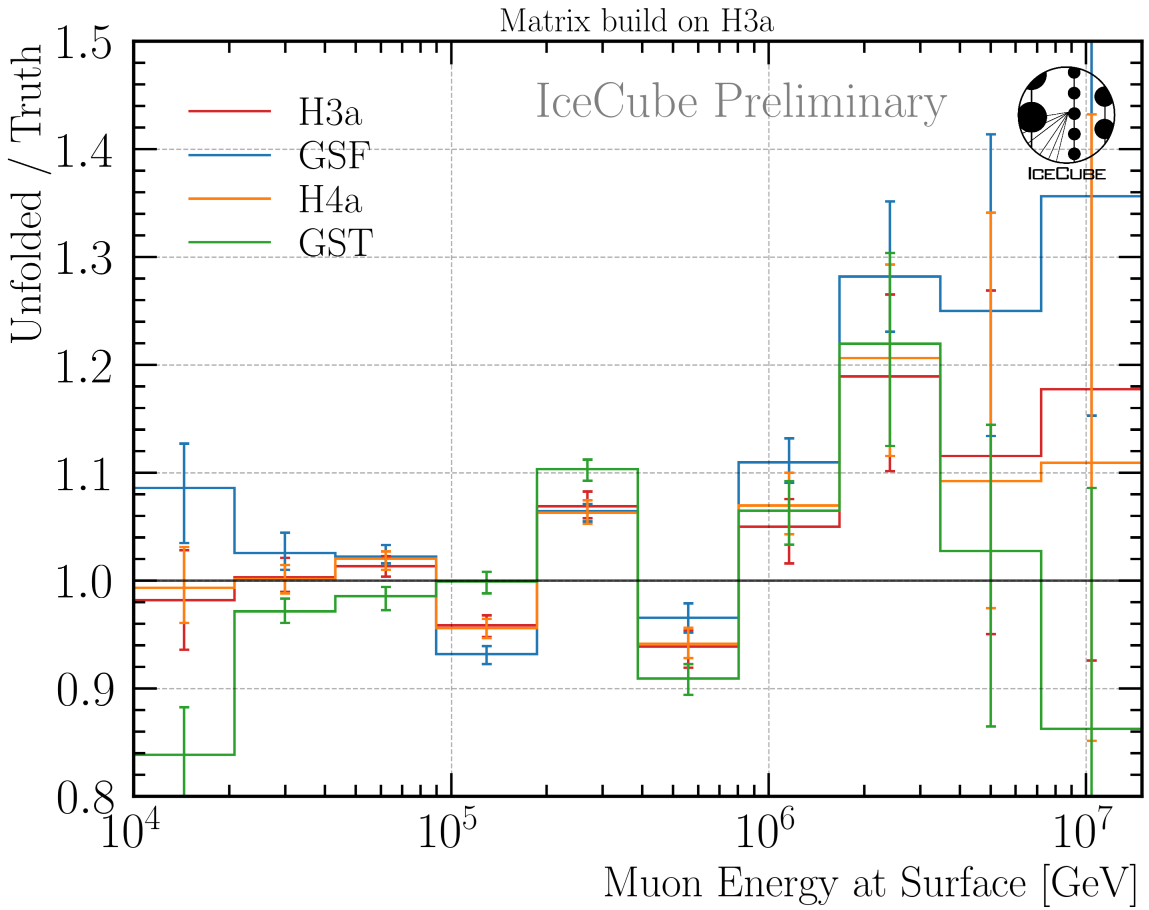
Post-unblinding checks

- Check the data-MC agreement for proxy variable (leading energy)
  - pre-fit (on the entire Snowstorm dataset)
  - post-fit (re-weight to fitted systematics)
- Divide total dataset in 3 subsets á 4 years to analyze systematic impacts over the years
- Unfold spectrum with all 4 CR model matrices

Thank you for your attention!

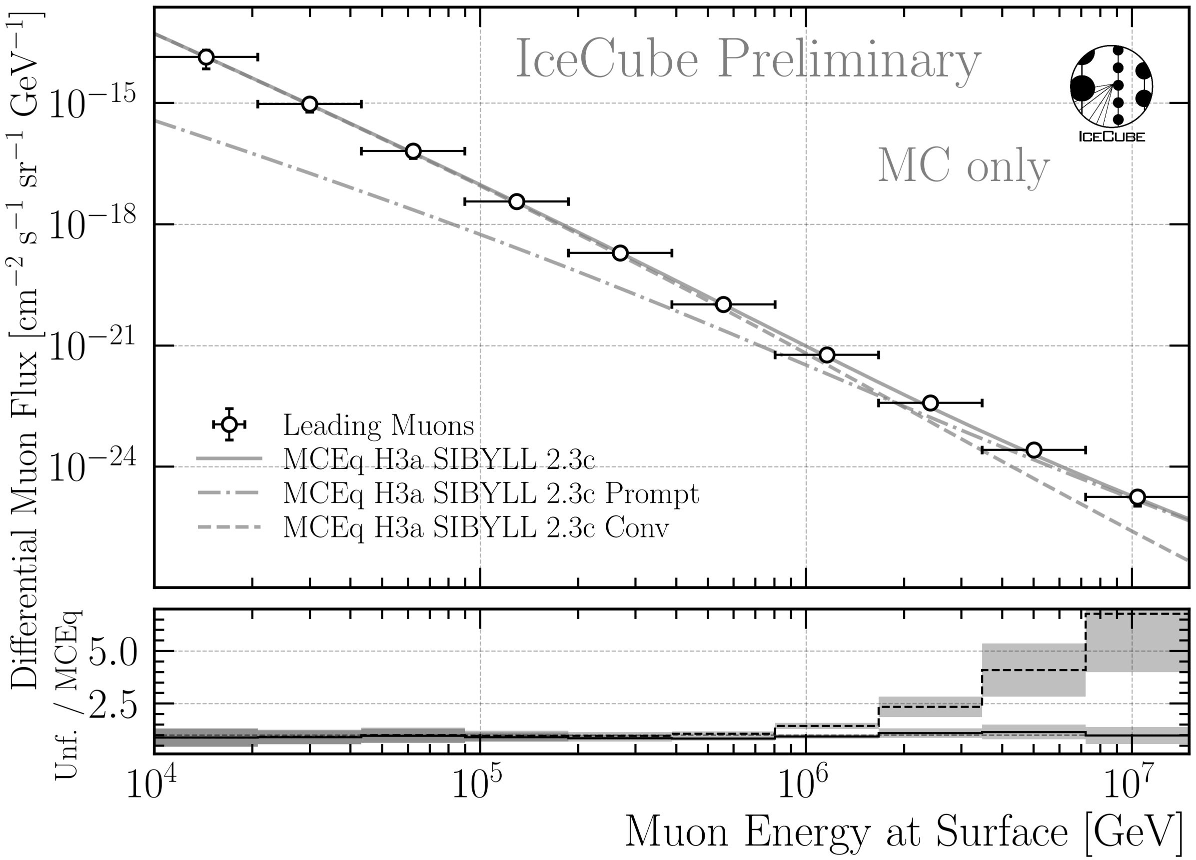
CR Model Impact and Unfolding Bias



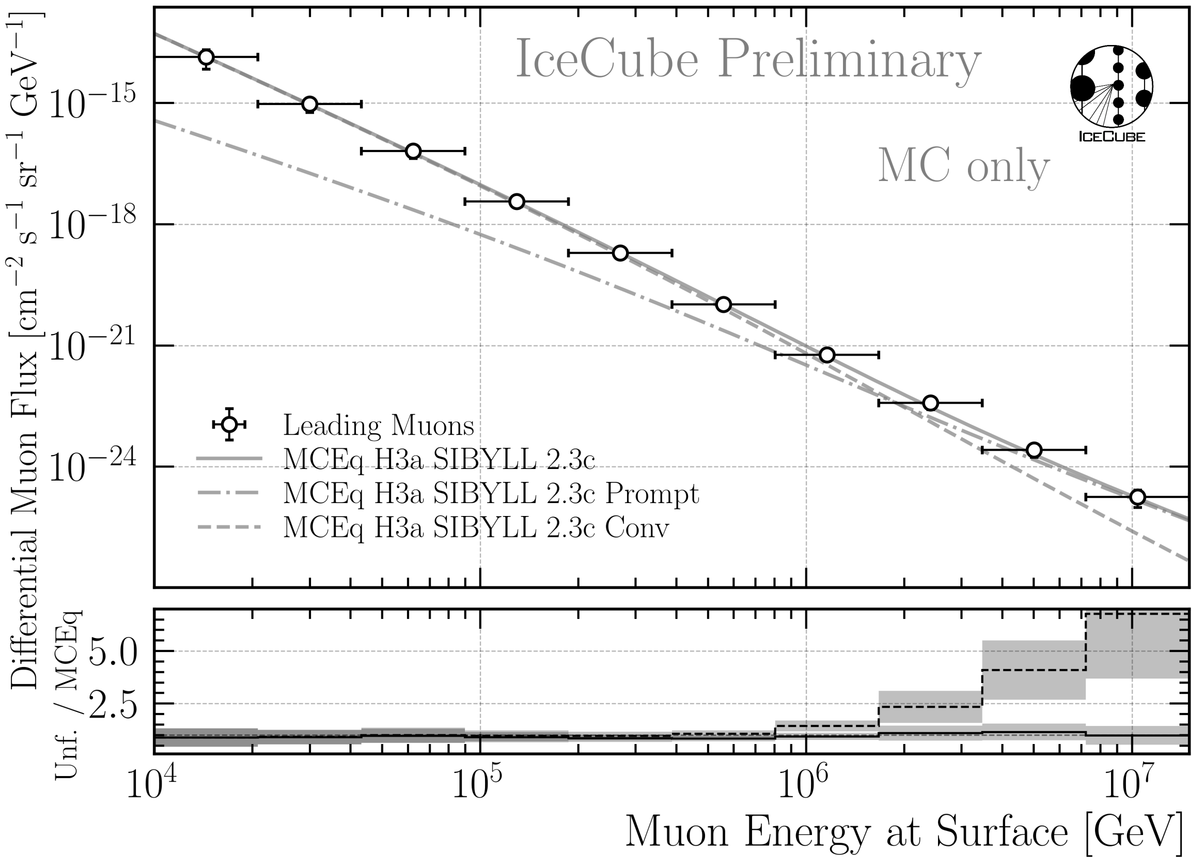


12 Years Prediction: All Uncertainties

minuit + effective area uncertainty

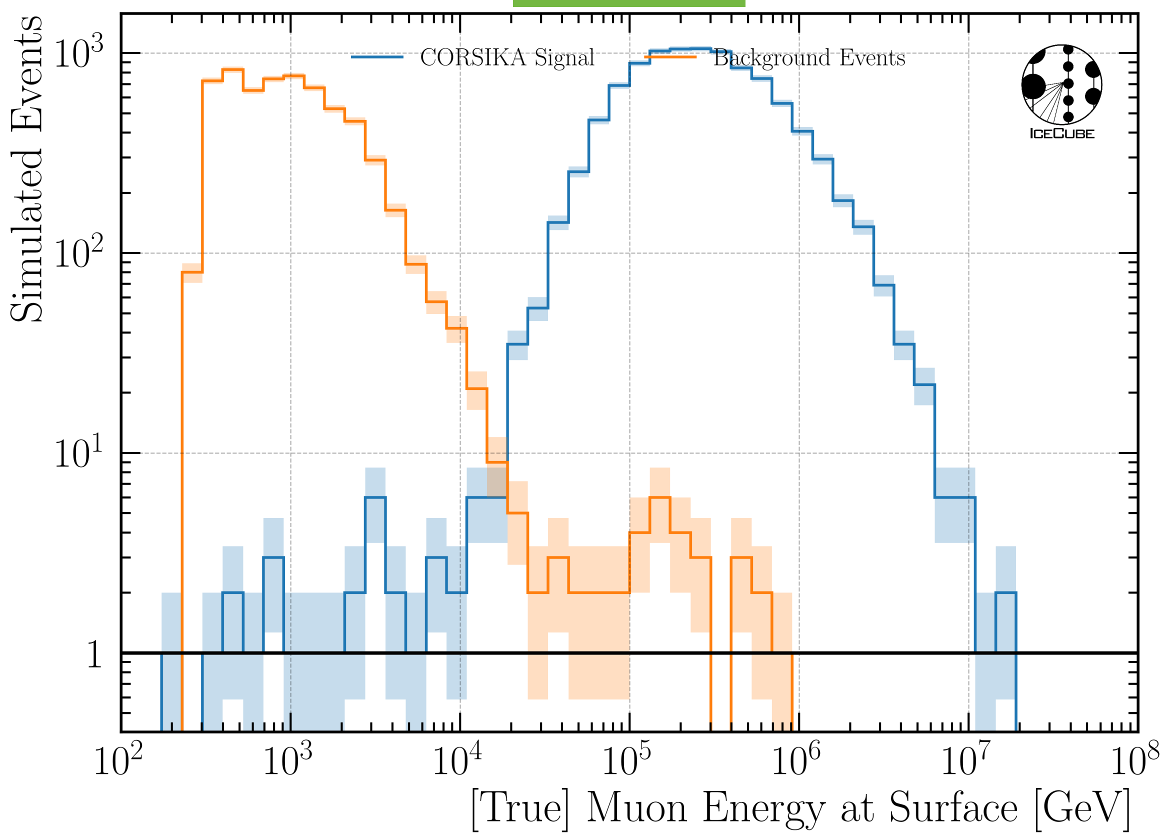


+ CR model uncertainty + unfolding bias

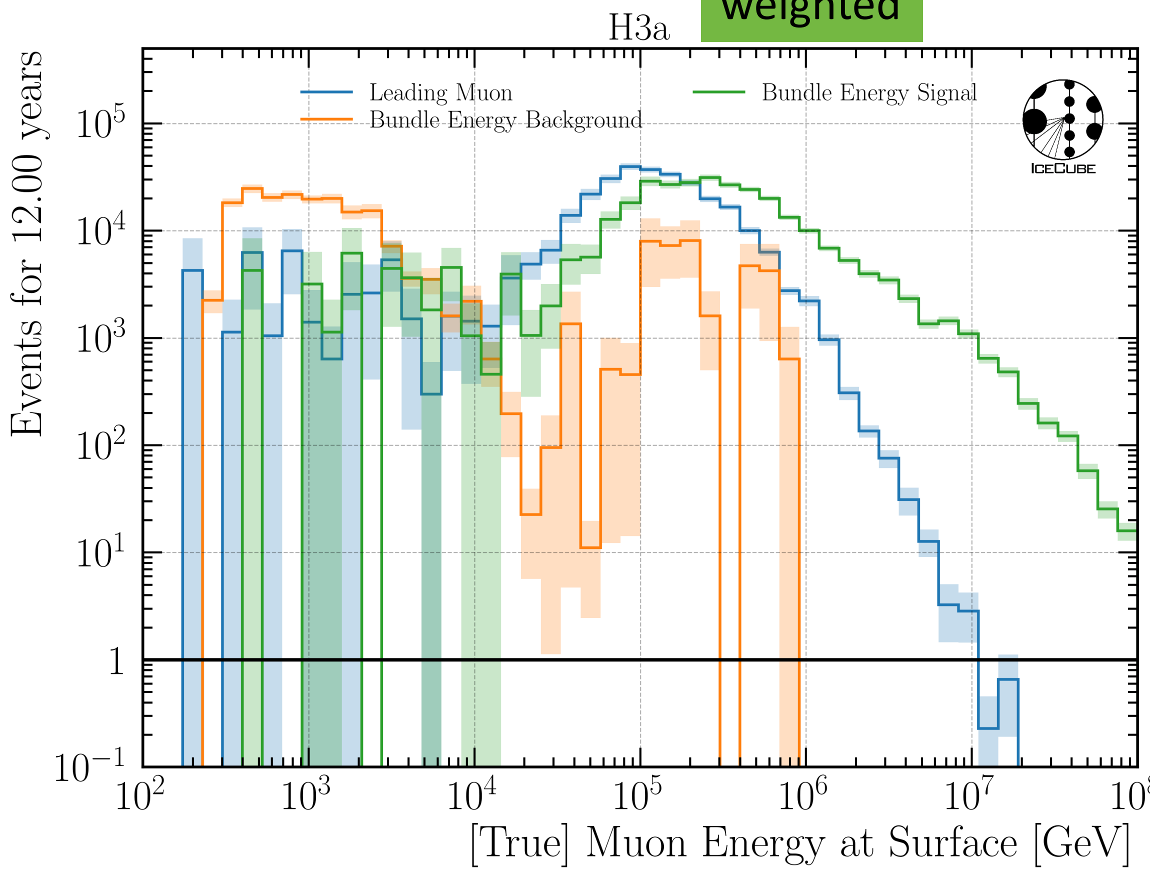


Muon Energy at Surface

simulated



weighted

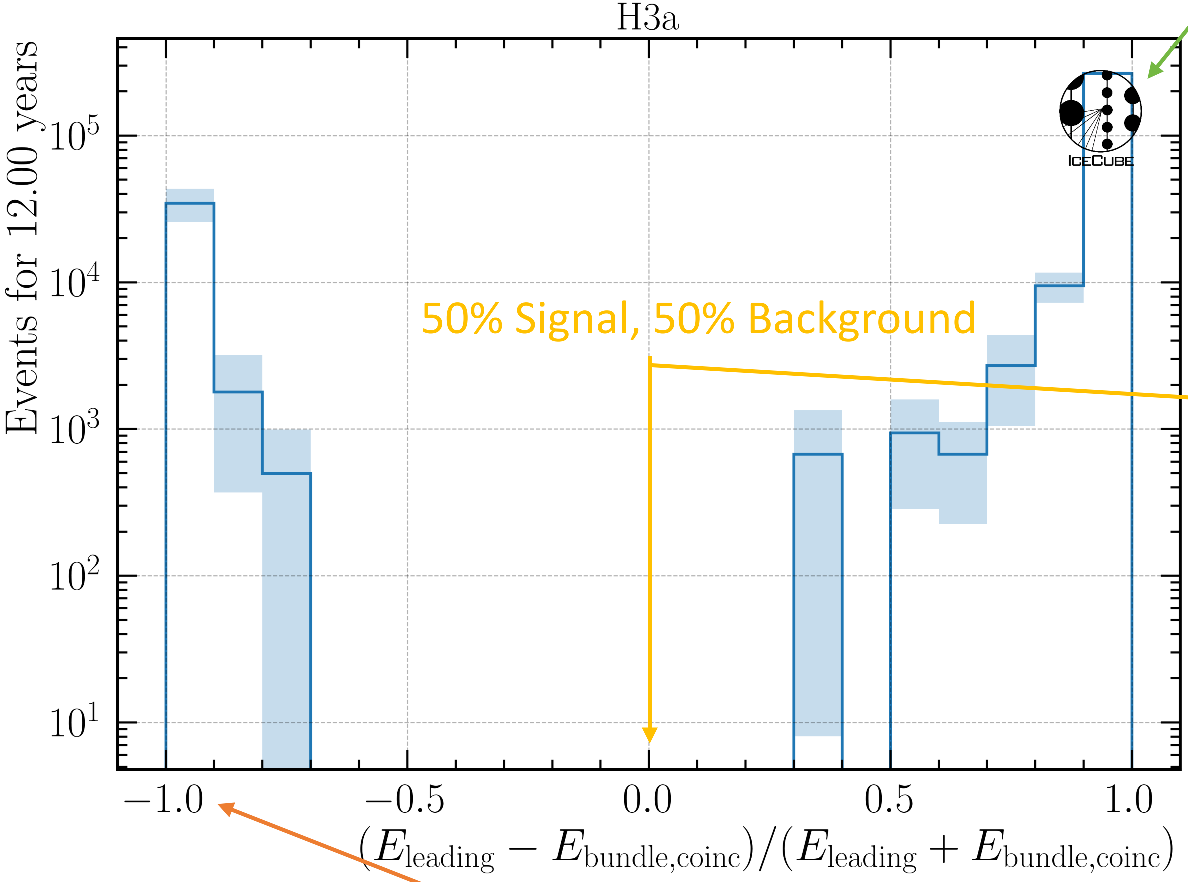


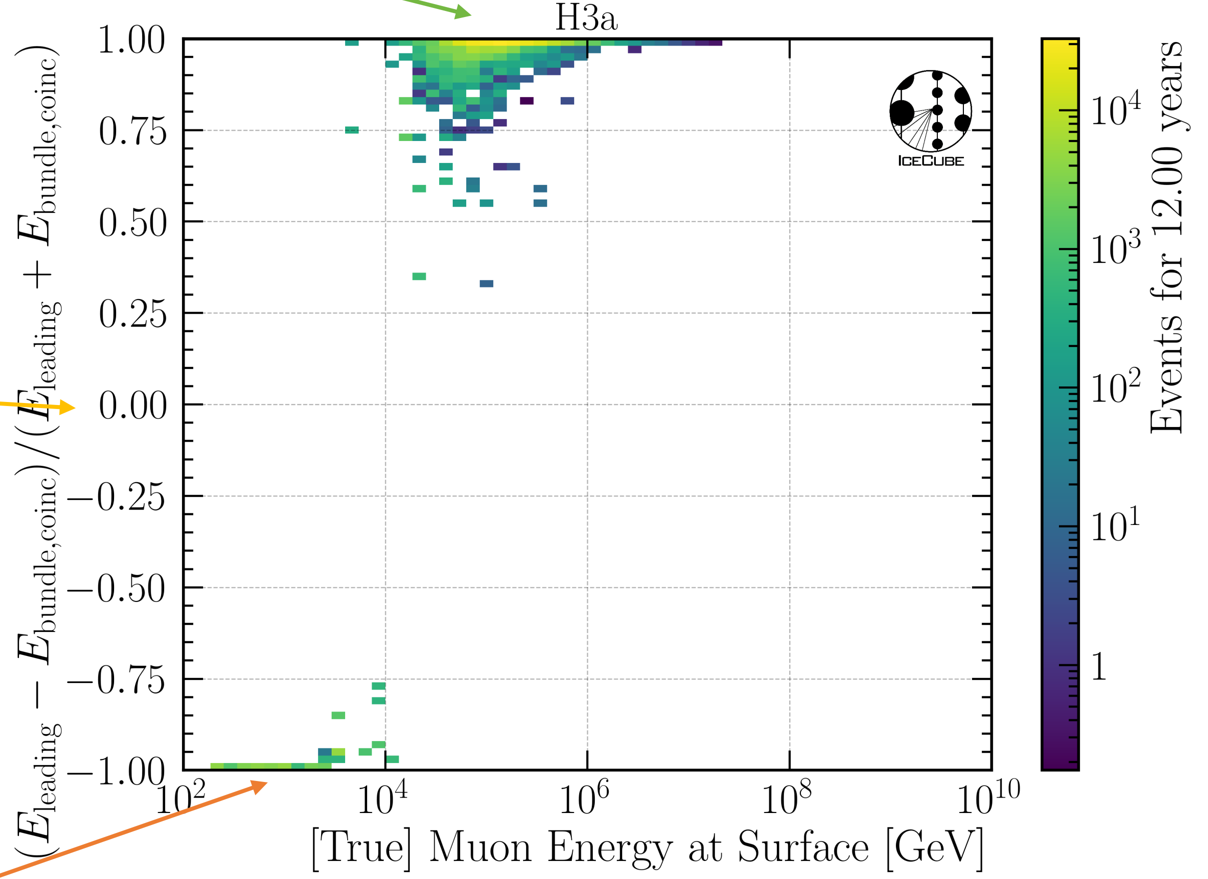
- Leading Muon: most energetic muon in tree
- Bundle Energy Signal: sum of all muons of ”signal” primary
- Bundle Energy Background: sum of all muons of coincident primaries

Coincidence: Muons on Event Level

100% Signal

50% Signal, 50% Background





100% Background

- above 10 TeV, no background dominated events

Estimate Rates with H3a

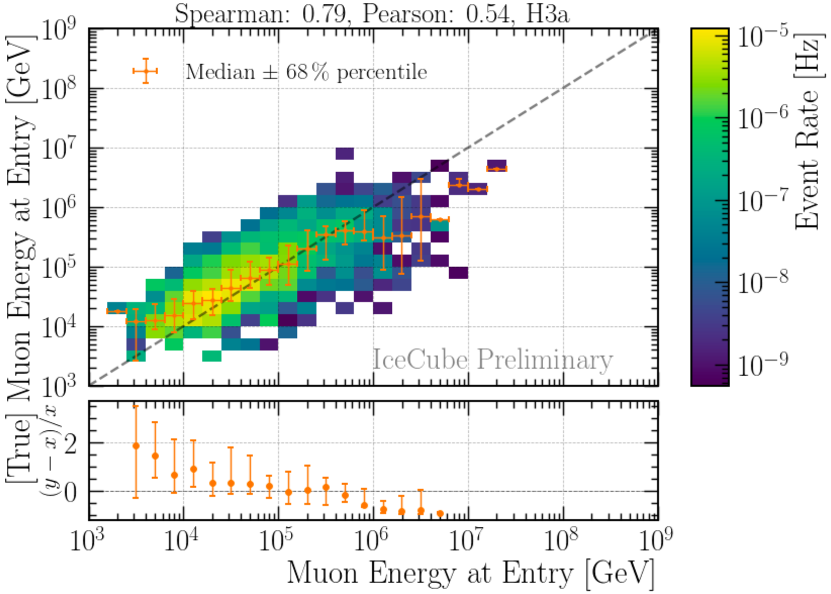
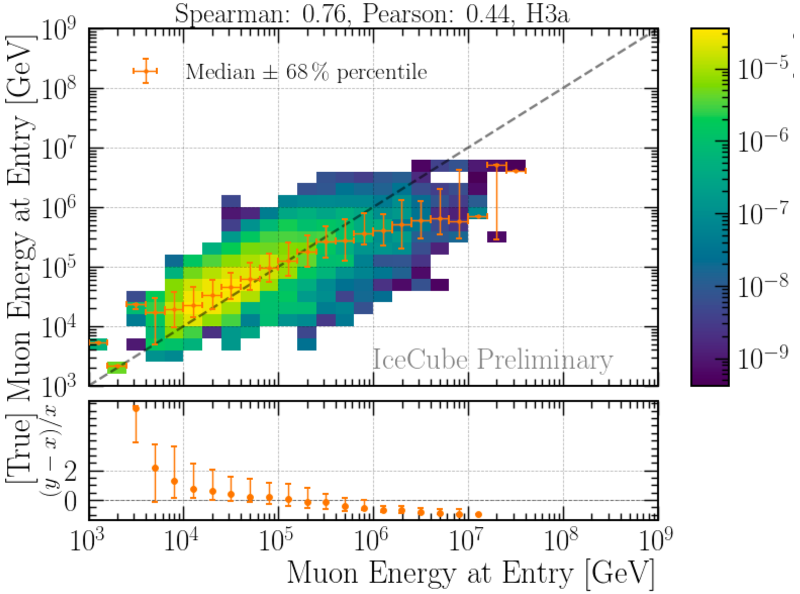
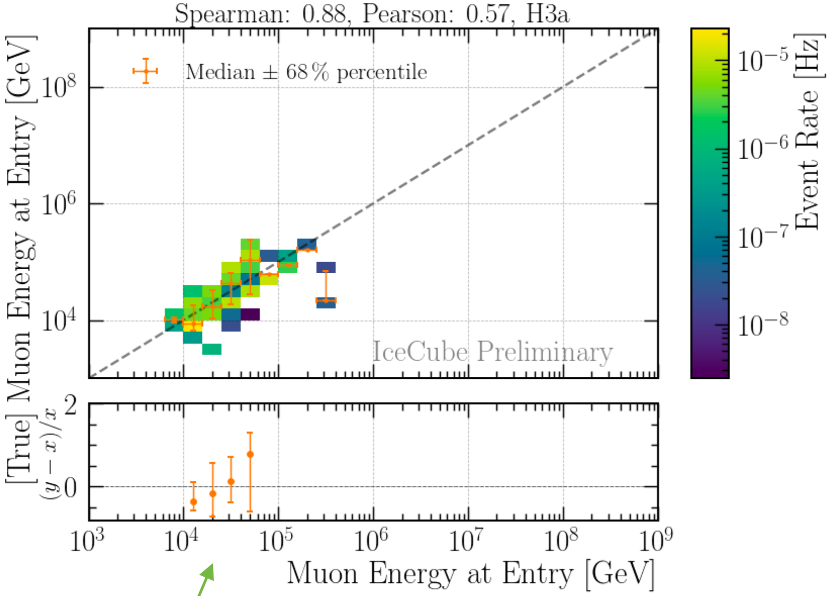
- Signal rate: 0.84 mHz (total event rate)
- Background rate: 0.68 mHz (event rate with at least 1 coincident primary)

- Signal rate: 0.74 mHz (leading muon energy at surface > 10 TeV)
- Background rate: 0.58 mHz (event rate with at least 1 coincident primary & lead. muon E. > 10 TeV)

- Signal rate: 0.74 mHz (leading muon energy at surface > 10 TeV)
- Background rate: 0.02 mHz (event rate with at least 1 coincident primary & lead. muon E. > 10 TeV & coincident muon bundle energy at surface hast at least 10% of leading muon energy)

- ➢ The light of the background events overlaps with the signal → no chance to separate
- ➢ A 10% bundle energy contribution would shift the measured light up by roughly 10% → within the uncertainties of the energy prediction
- ➢ Networks have been trained on MC with coincident events → they are able to subtract a little light in case they assume there is a background event, however, this is not quantified

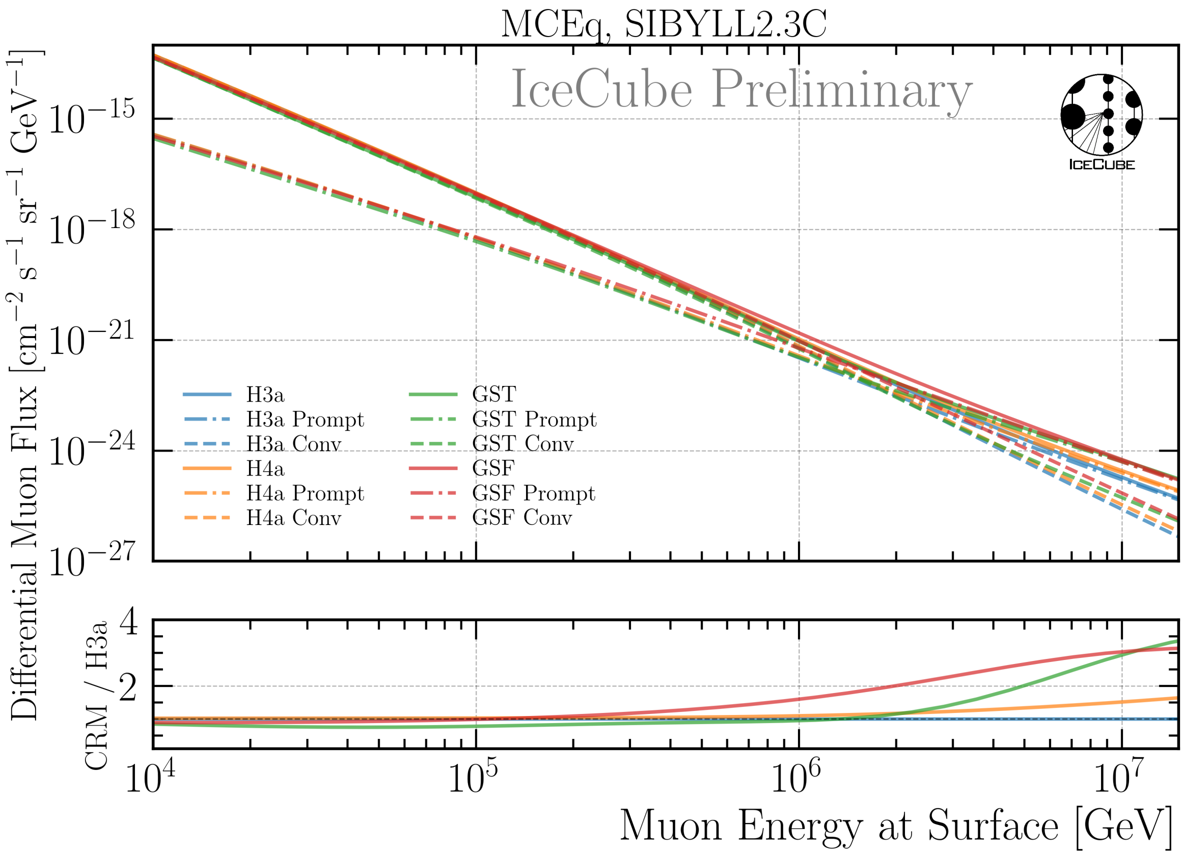
Reconstructions: Coincident Events

no background event	$\frac{\text{bundle energy at surface}}{\text{leading energy at surface}} < 0.1$	$\frac{\text{bundle energy at surface}}{\text{leading energy at surface}} > 0.1$
		

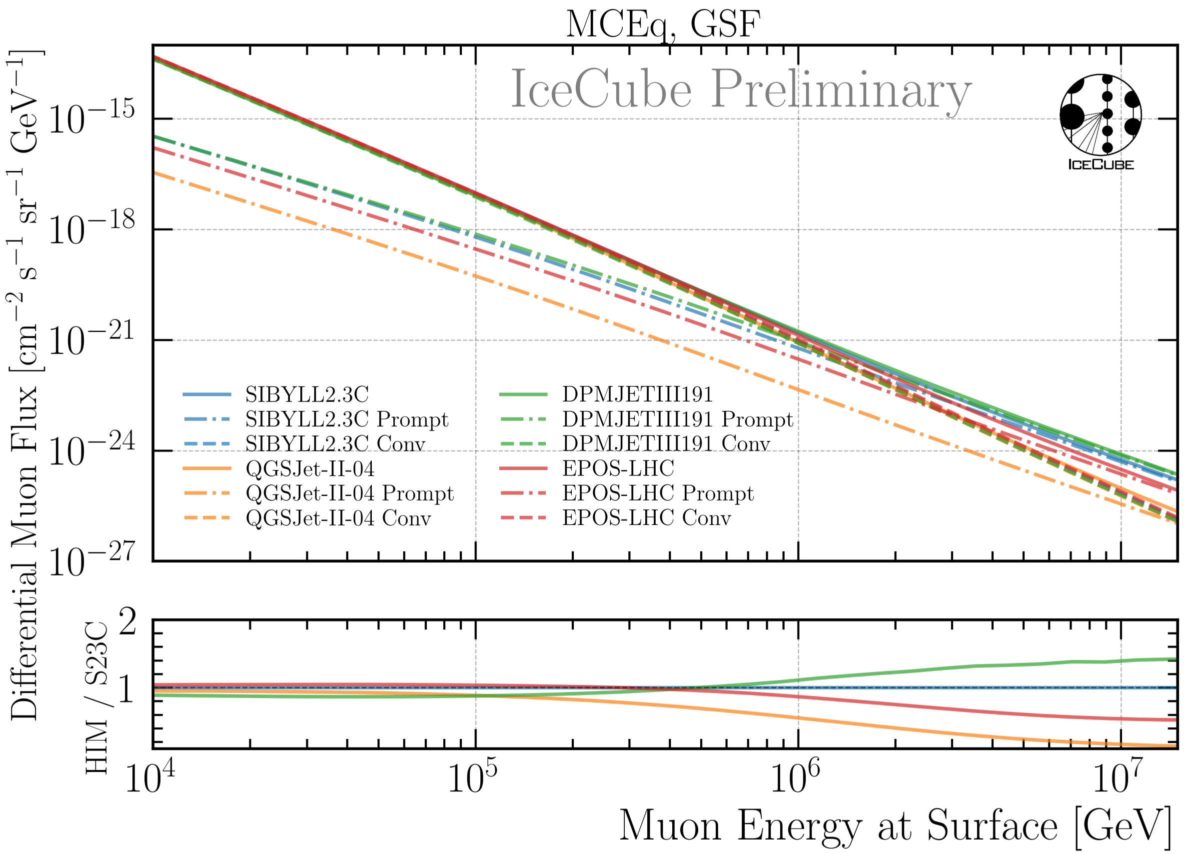
- sufficient energy reconstruction for events with background

Estimate Impact of Hadronic Interaction Model

Different CR Models



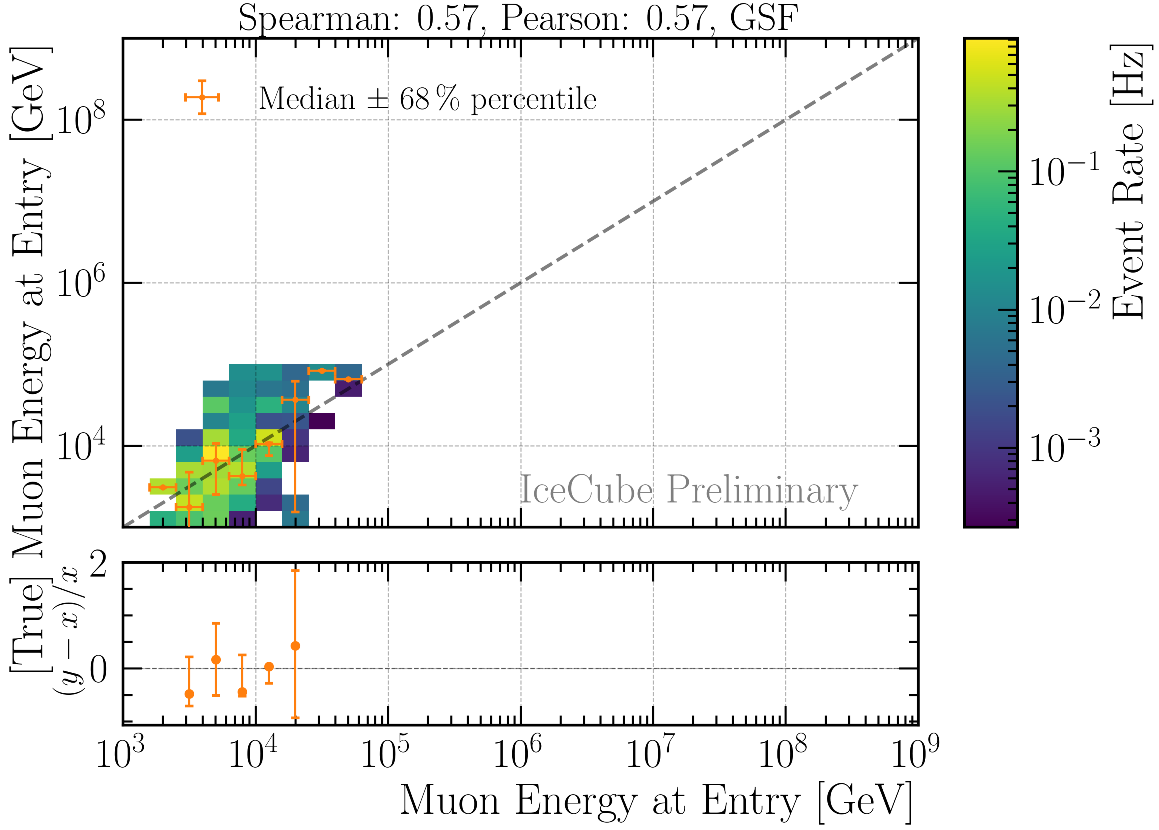
Different Haronic Models



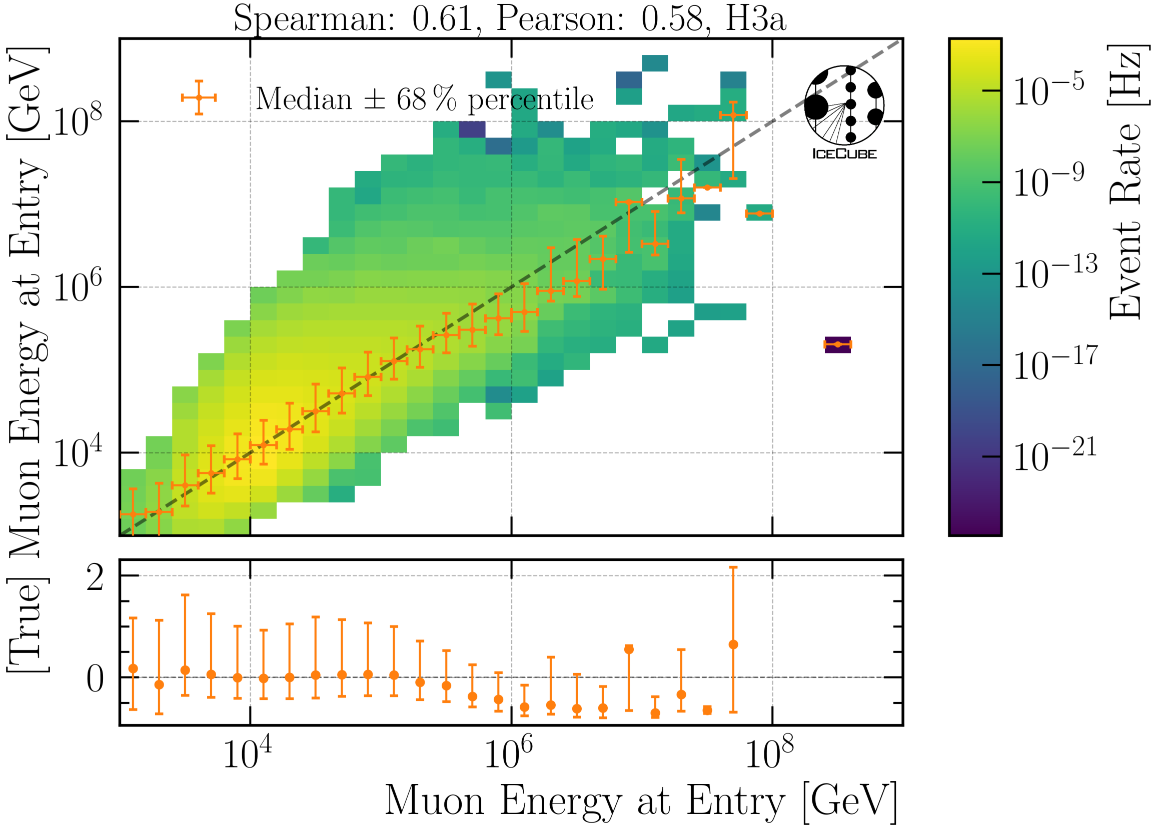
- Not all hadronic models include charm
- Impact of HIM smaller than CRM
  - CRM impact below 3%
  - Hadronic Model impact is negligible

Reconstruction: EPOS-LHC vs SIBYLL 2.3d [Level 5]

EPOS-LHC



SIBYLL 2.3d



- IceTop simulations for proton and iron
- 23198 and 23201
- Primary energy > 1e7

Event Rate to Flux → Effective Area

- Unfolding estimates an event rate
- Transfer event rate to flux

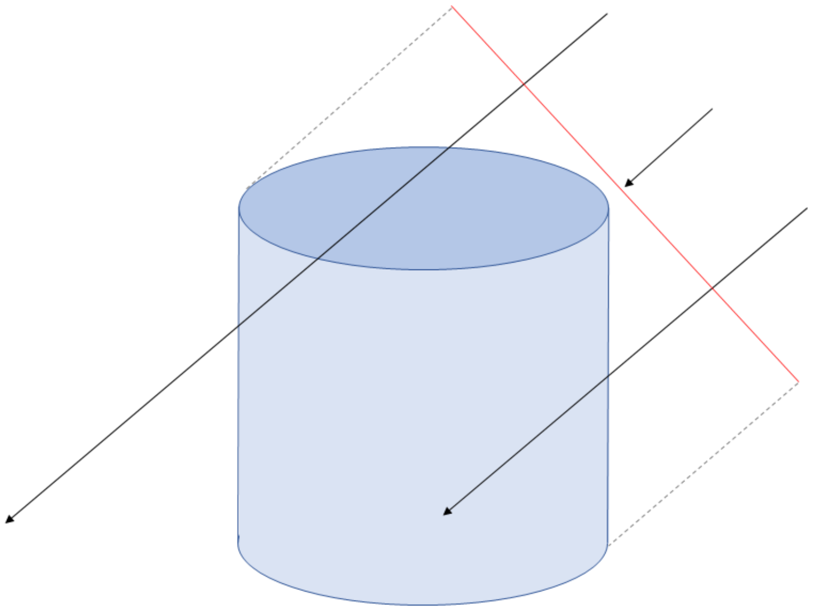
$$\Phi_i = \frac{N_i}{T \cdot \Delta L_i \cdot \Omega_i \cdot A_{\mathrm{eff},i}}$$

- with solid angle

$$\Omega_i = 2\pi \cdot (\cos\Theta_{\mathrm{min},i} - \cos\Theta_{\mathrm{max},i})$$

- and effective area

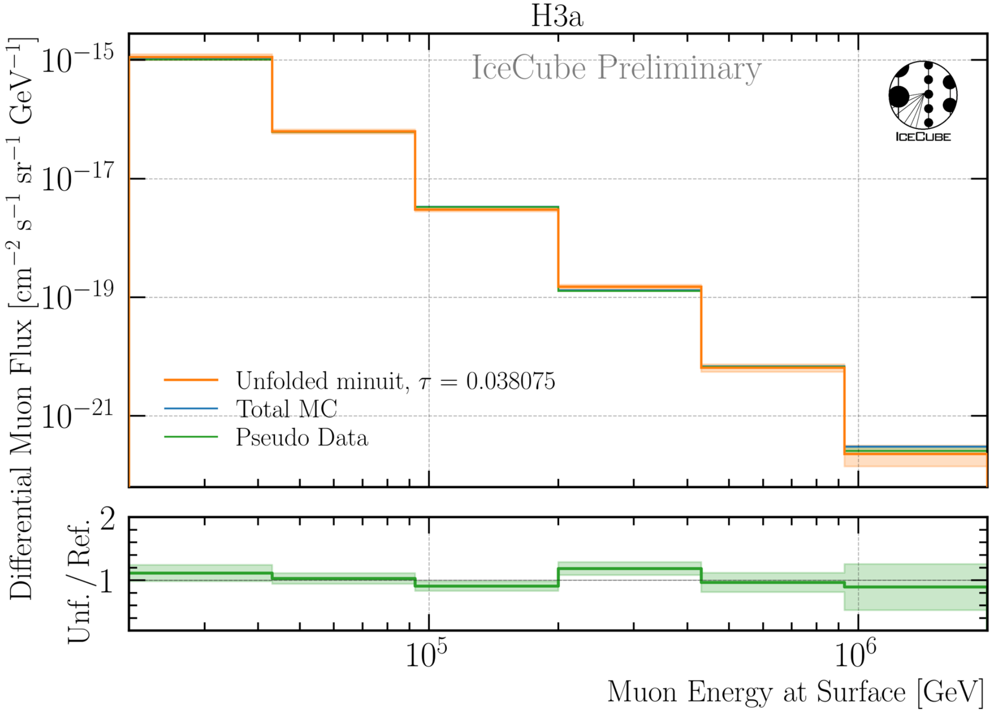
$$A_{\mathrm{eff}} = A_{\mathrm{sim}} \frac{N_{\mathrm{sel}}}{N_{\mathrm{gen}}}$$

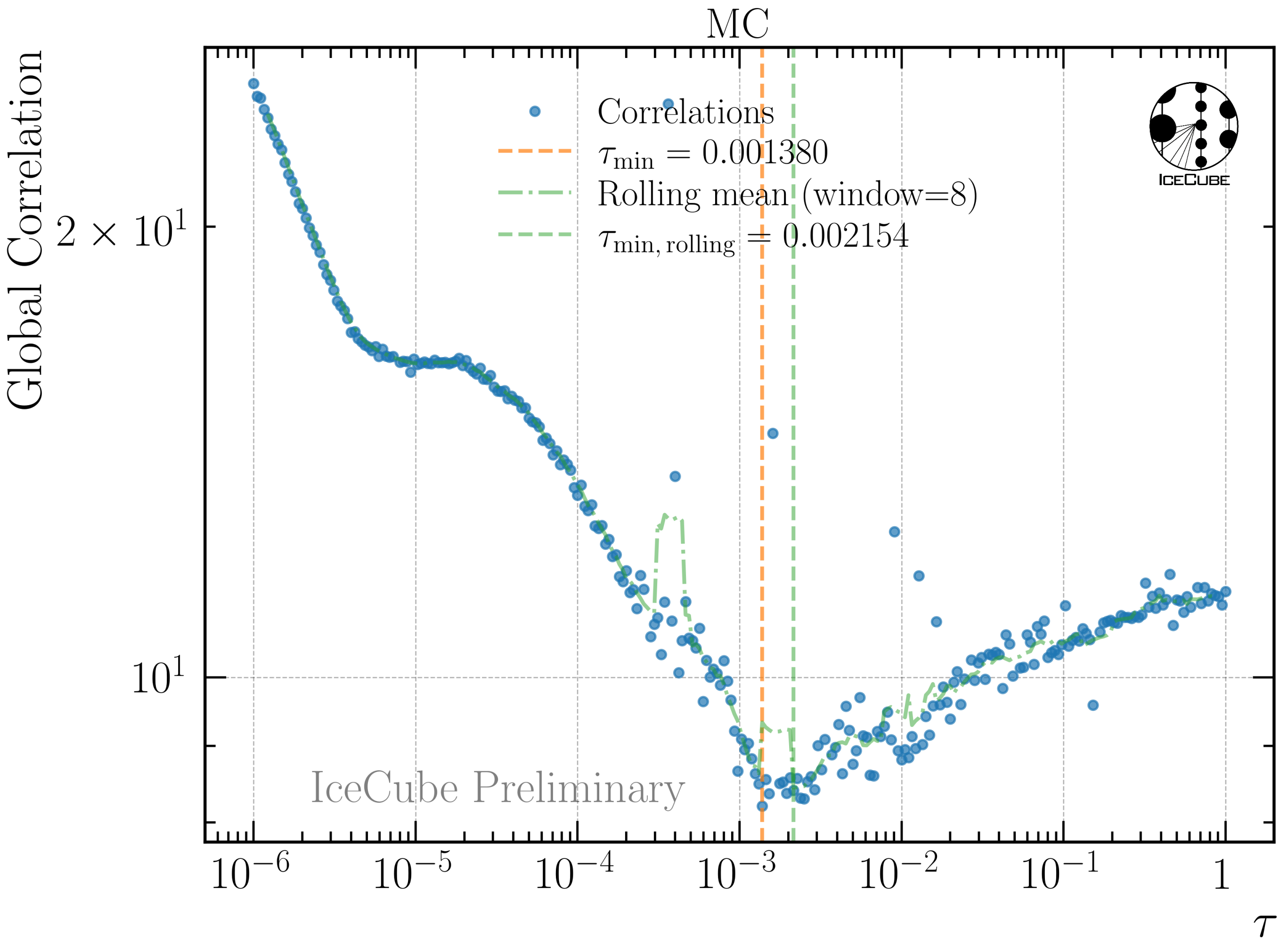


Determine Regularization

- Find regularization τ with minimal bin-to-bin correlation
- LLH minimization (unfolding) provides full covariance matrix V

➢ Minimize global correlation $\rho = \sum_{i>i} V_{ij}$





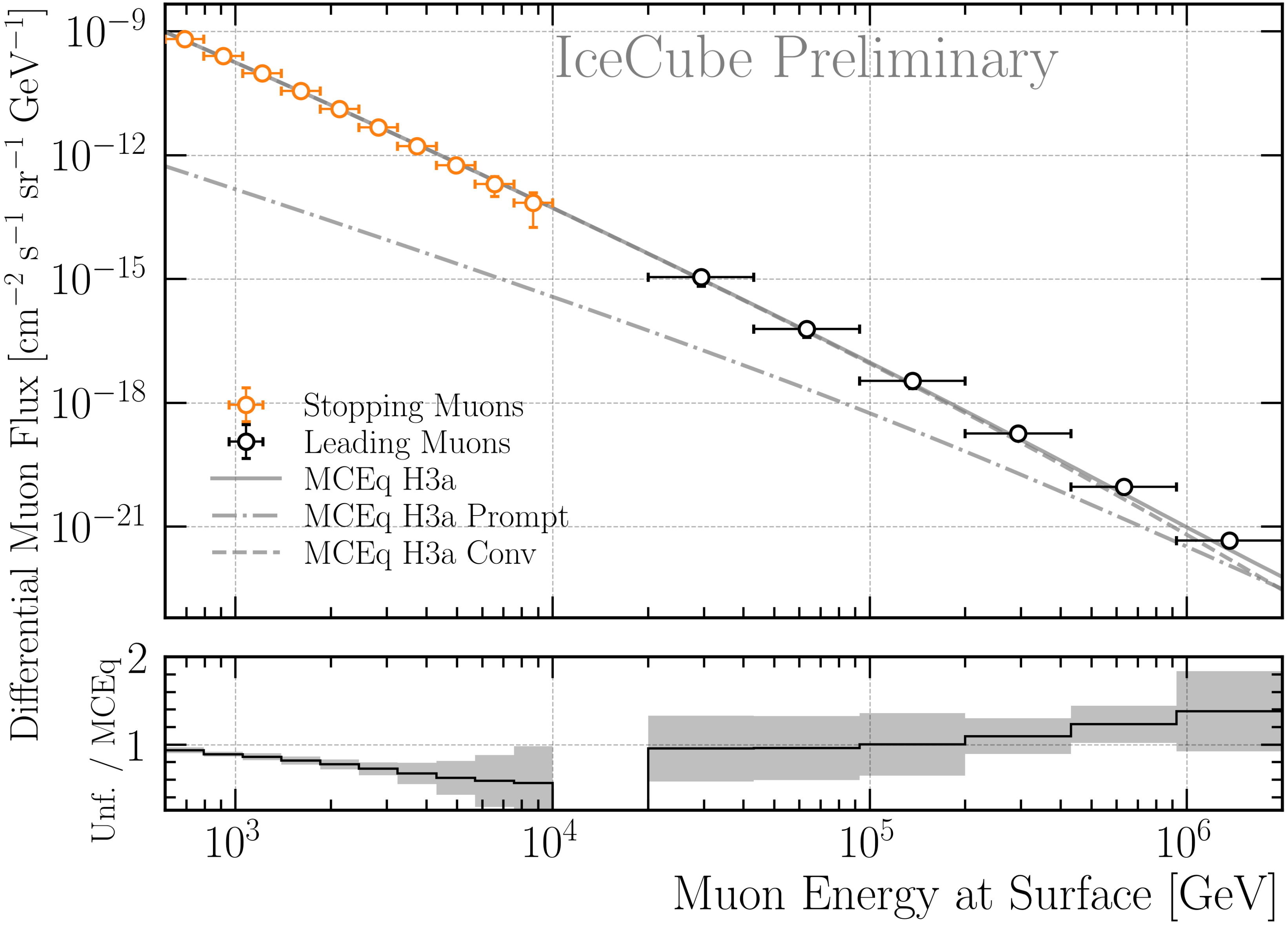
Unfold MC truth → ratio should be 1

Burnsample

Muon Flux Unfolding

Leading muons
- 2487 h IceCube data
- 12754 events
- Agrees with MCEq

Stopping muons
- 47 min IceCube data
- 32943 events
- Below MCEq prediction

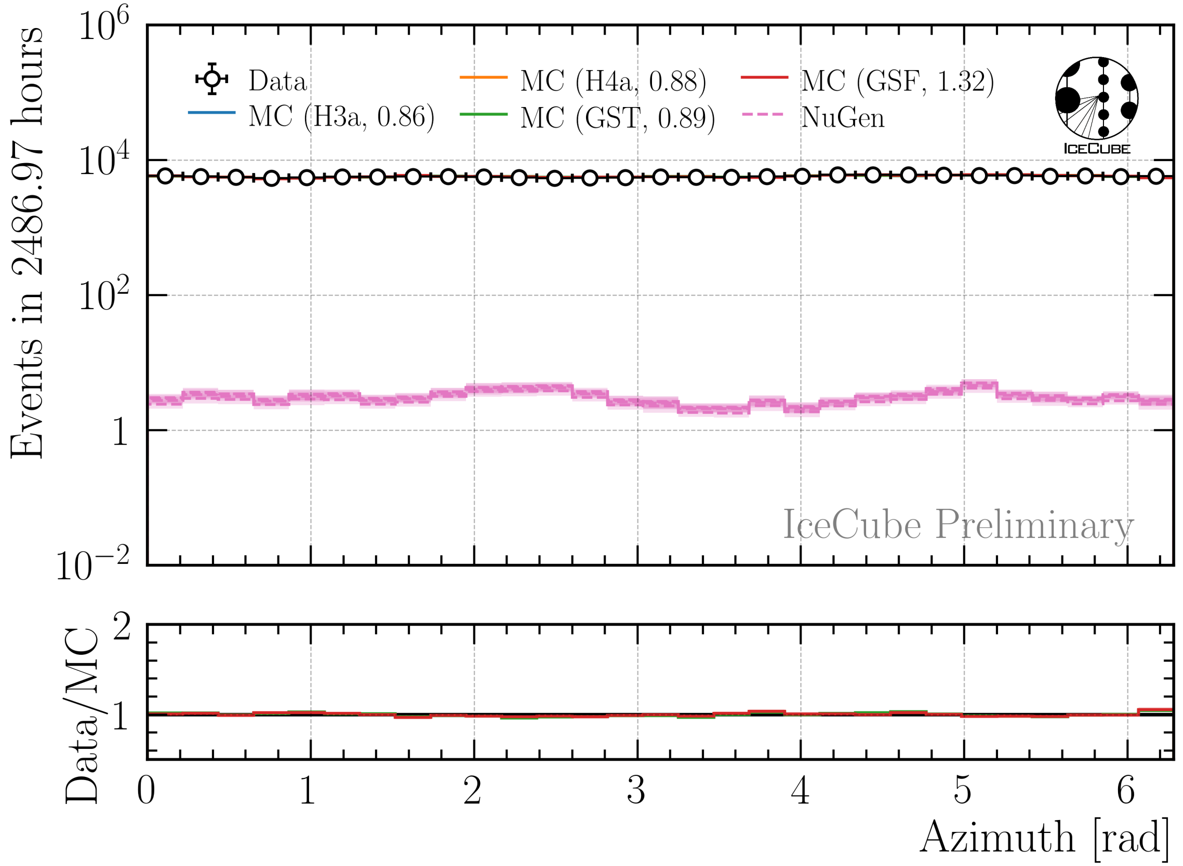


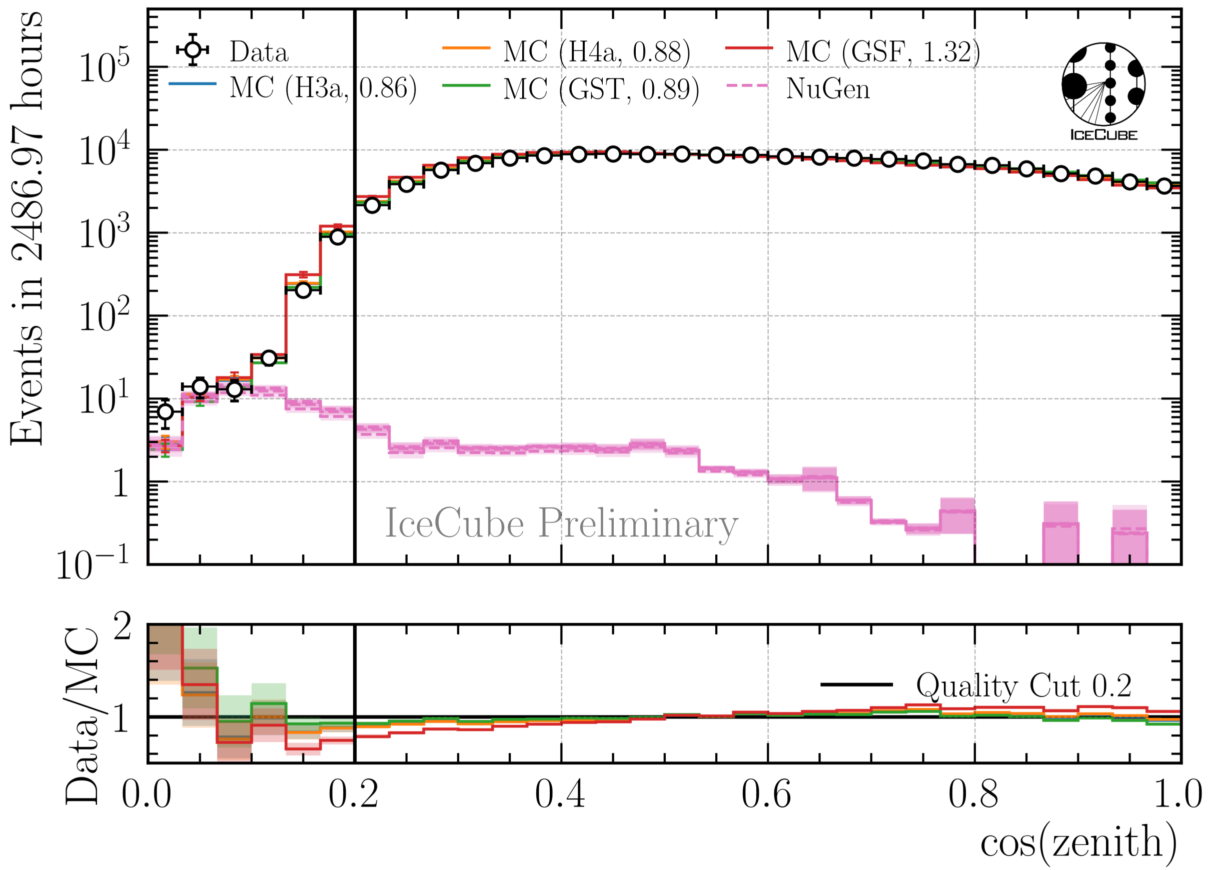
*Stopping muon analysis done by Lucas Witthaus

Level 5: Data/MC Quality Cuts

Neutrino weighting:
SPL: n = 1.8, γ = 2.52

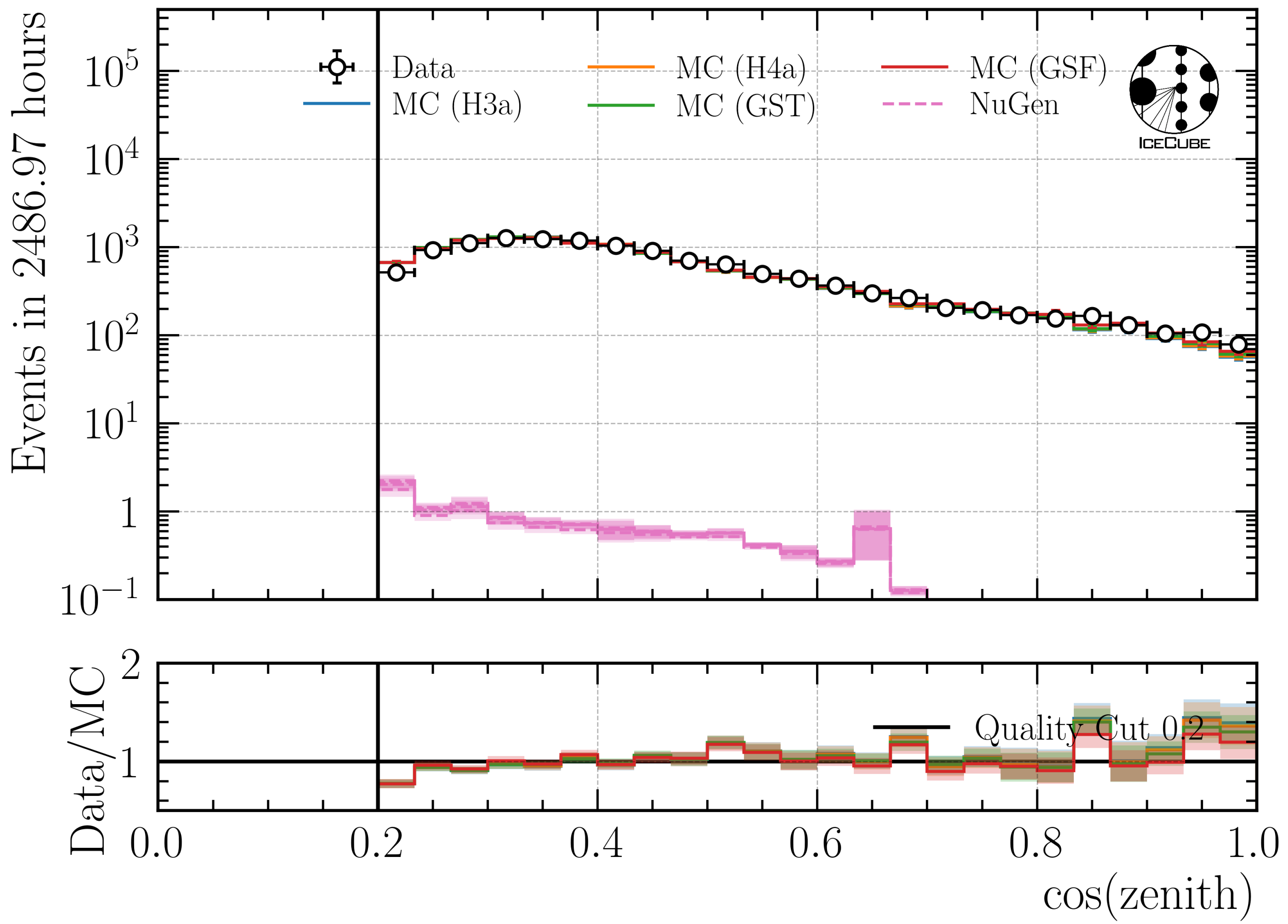
IceCube Collaboration PoS ICRC2023 1064





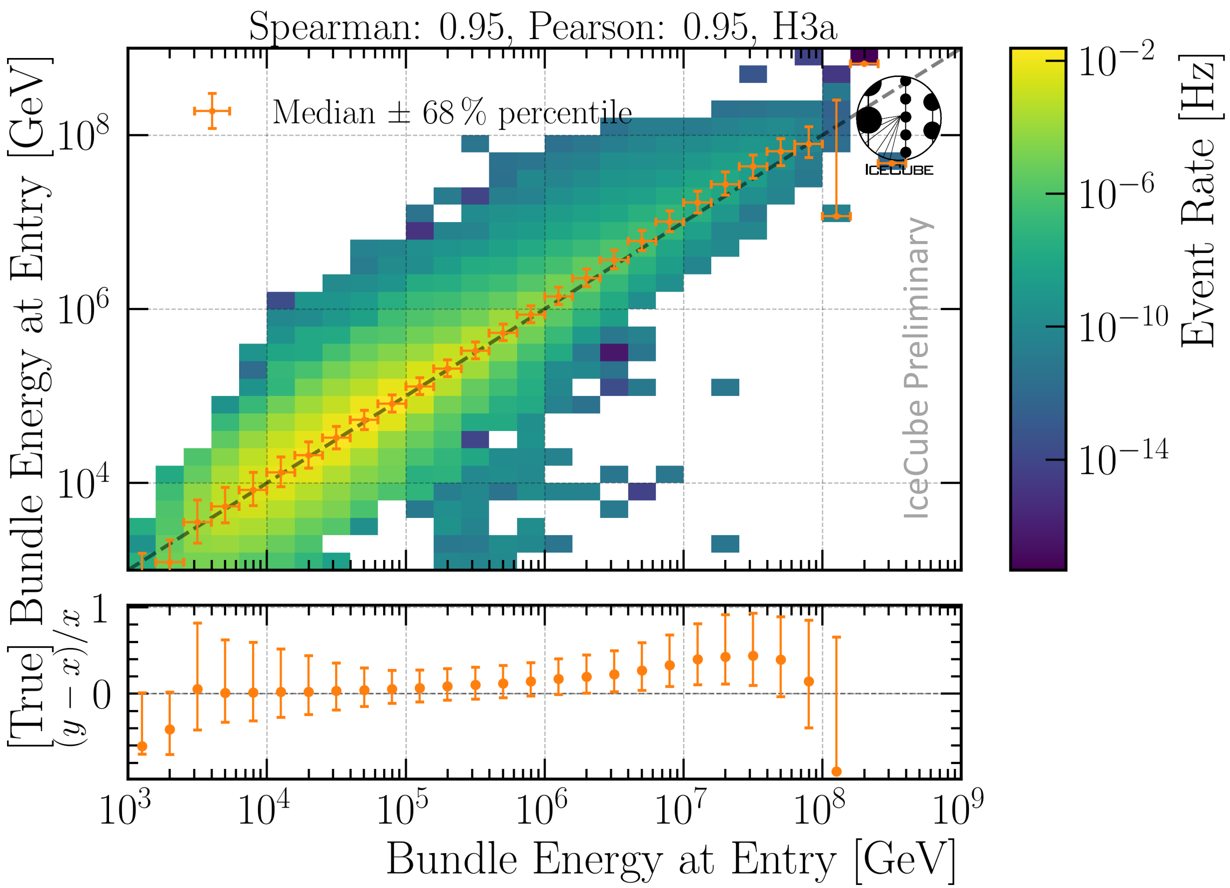
- Good directional reconstruction
- Cut: cos(zenith) > 0.2

Final Level cos(zenith)

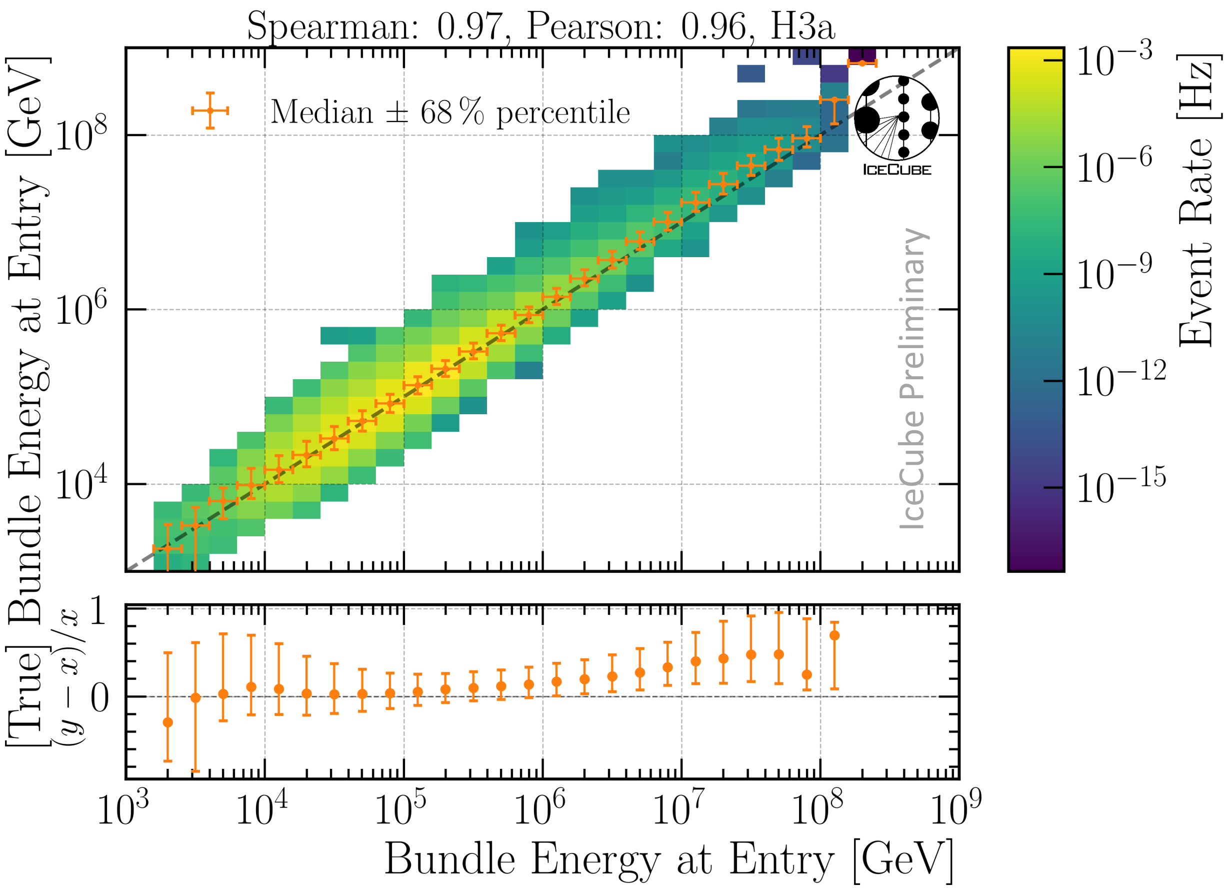


Bundle Energy Reconstruction

Level 4

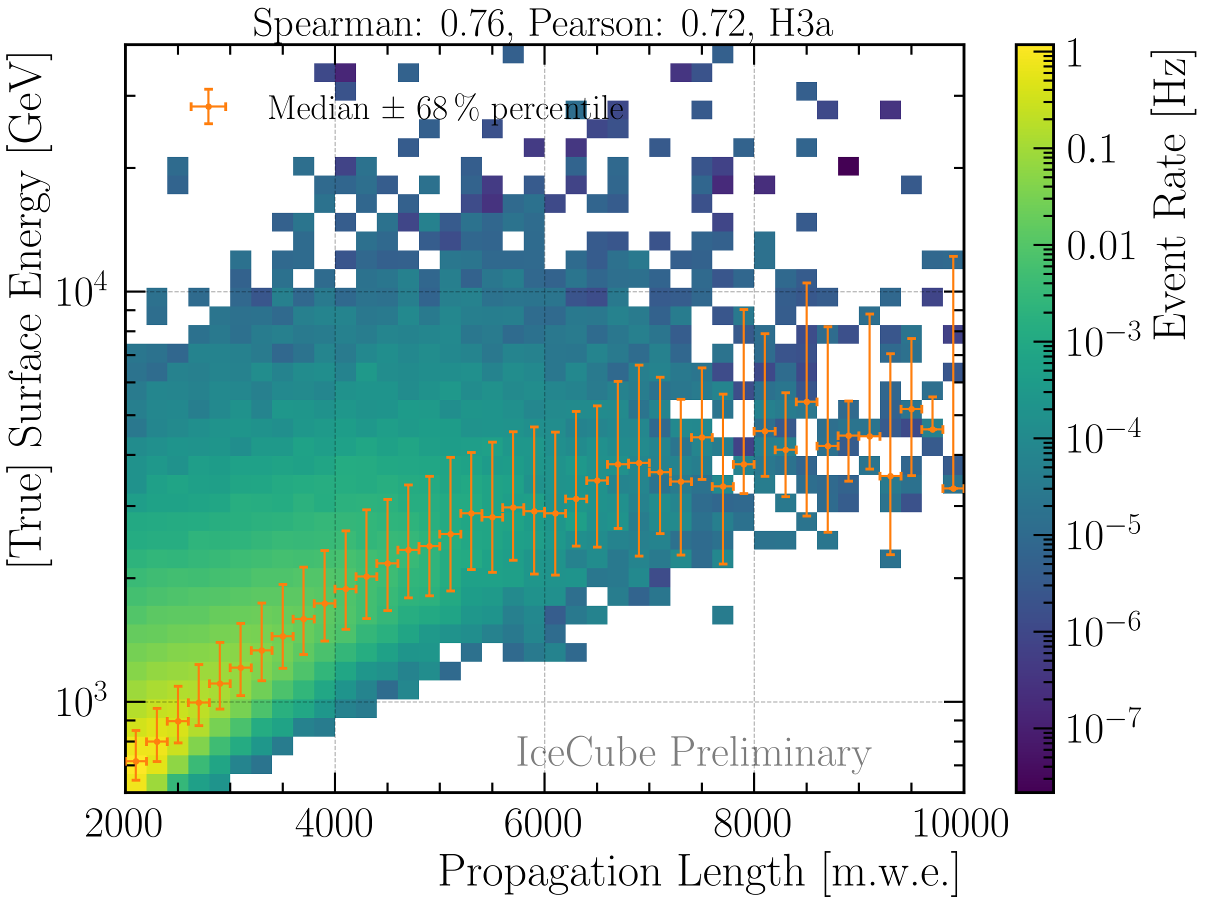


Level 5

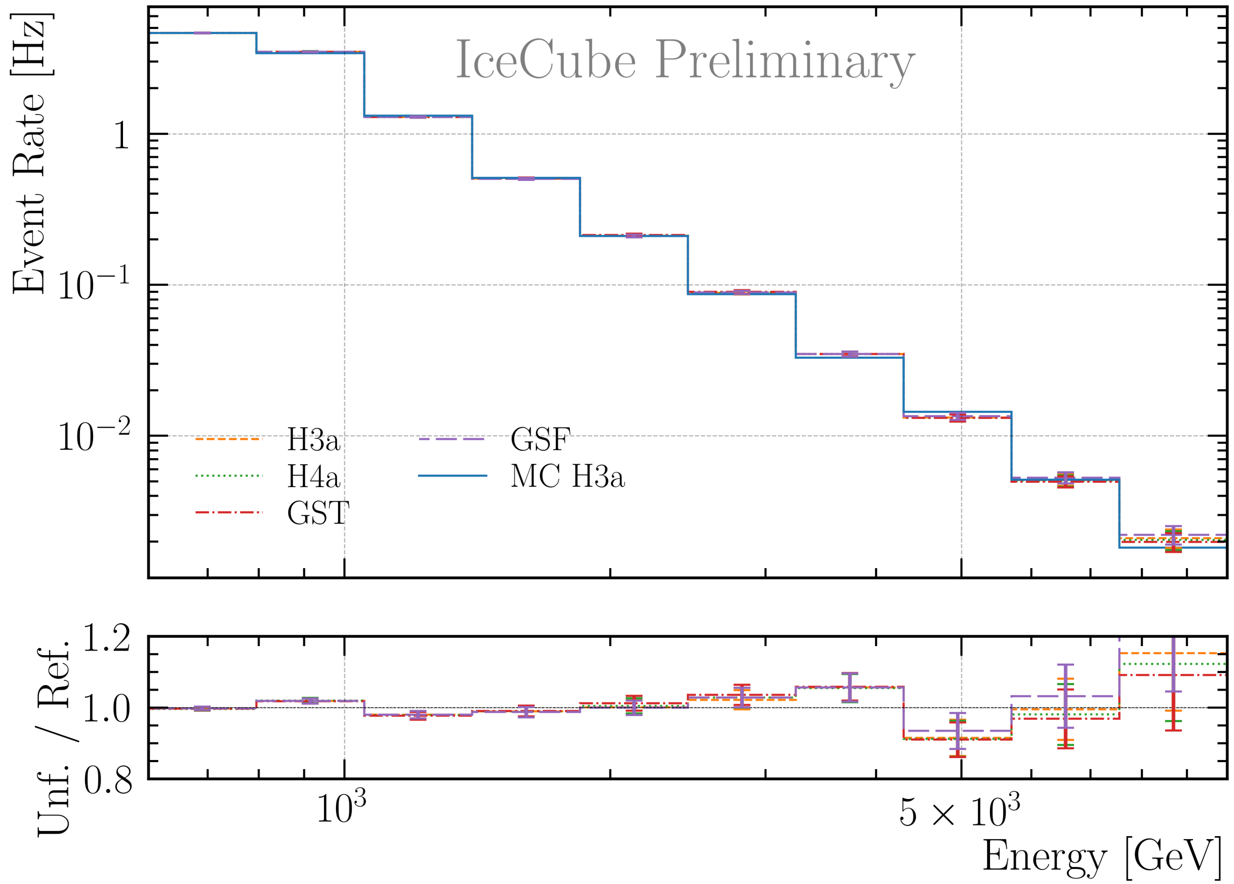


Correlation and Robustness Test: Stopping Muons

- Correlation between proxy and target

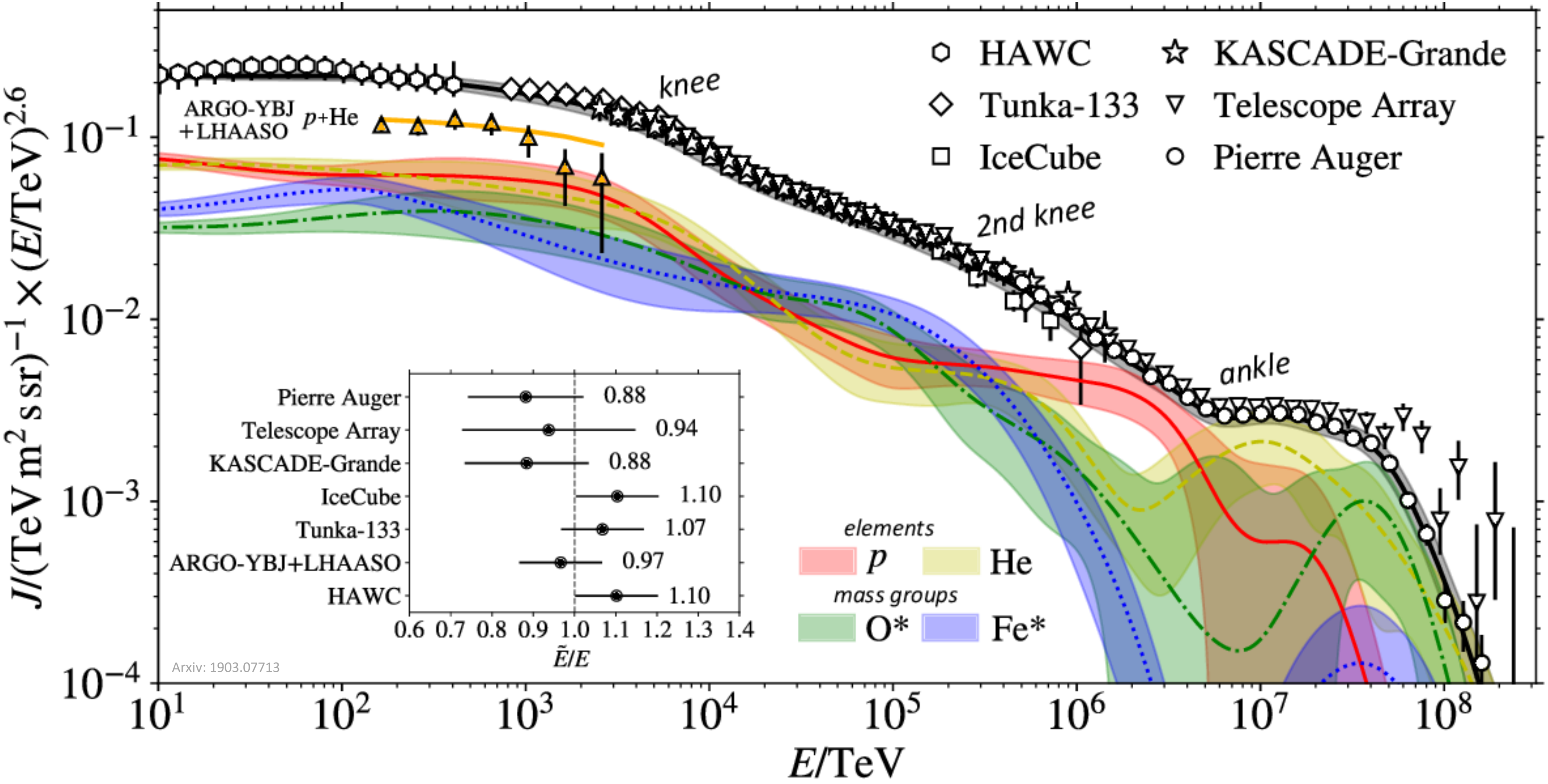


- Build unfolding matrix on H3a, H4a, GST & GSF
- Unfold H3a as test set
- Results are within uncertainties → robust

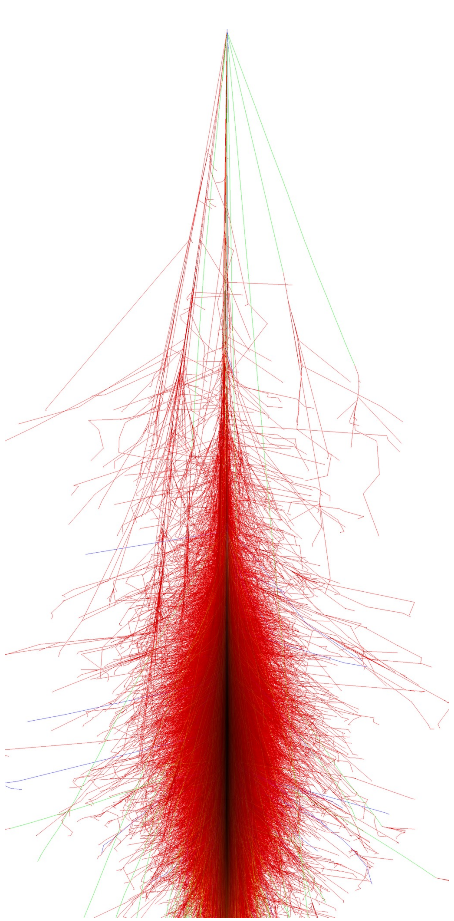
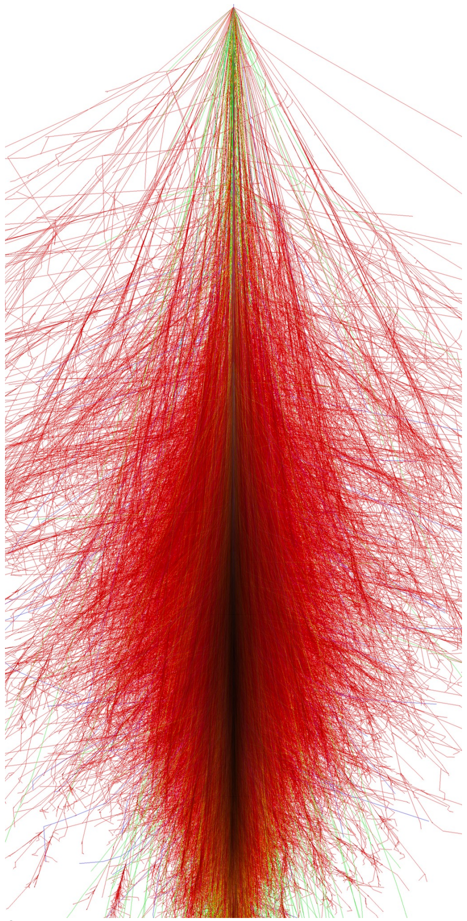


pascal.gutjahr@tu-dortmund.de

Cosmic ray flux



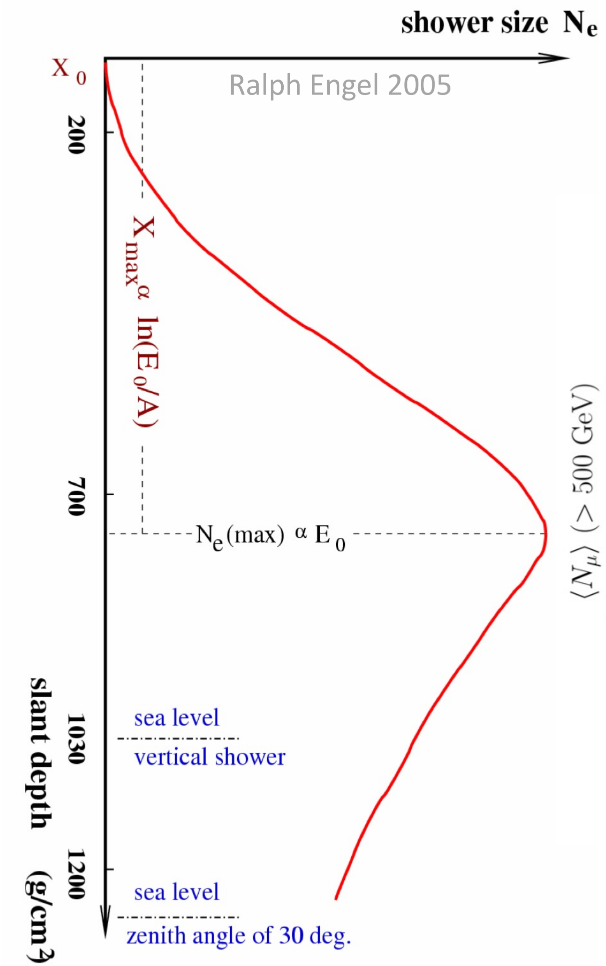
Air shower – 10 TeV

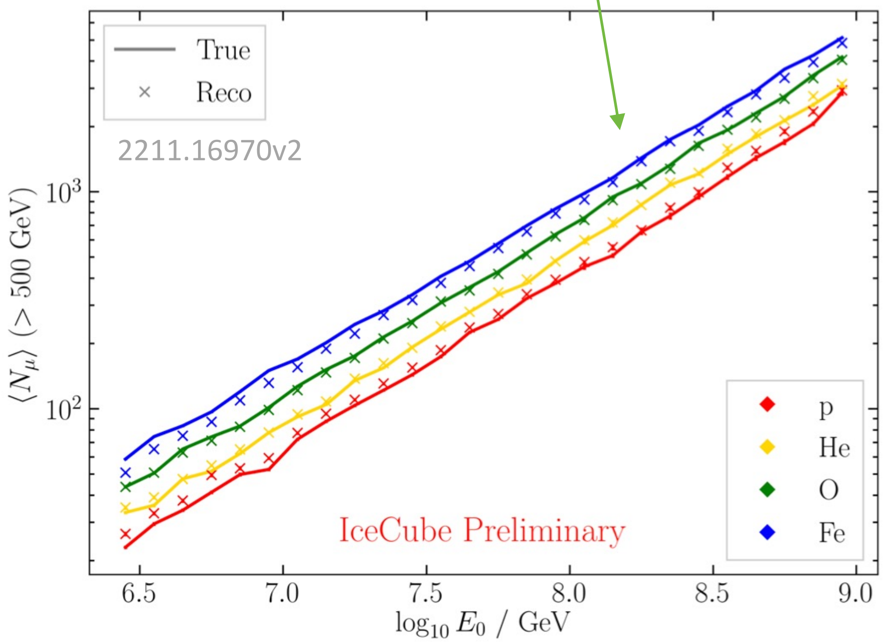
www.iap.kit.edu/corsika

Proton

Iron

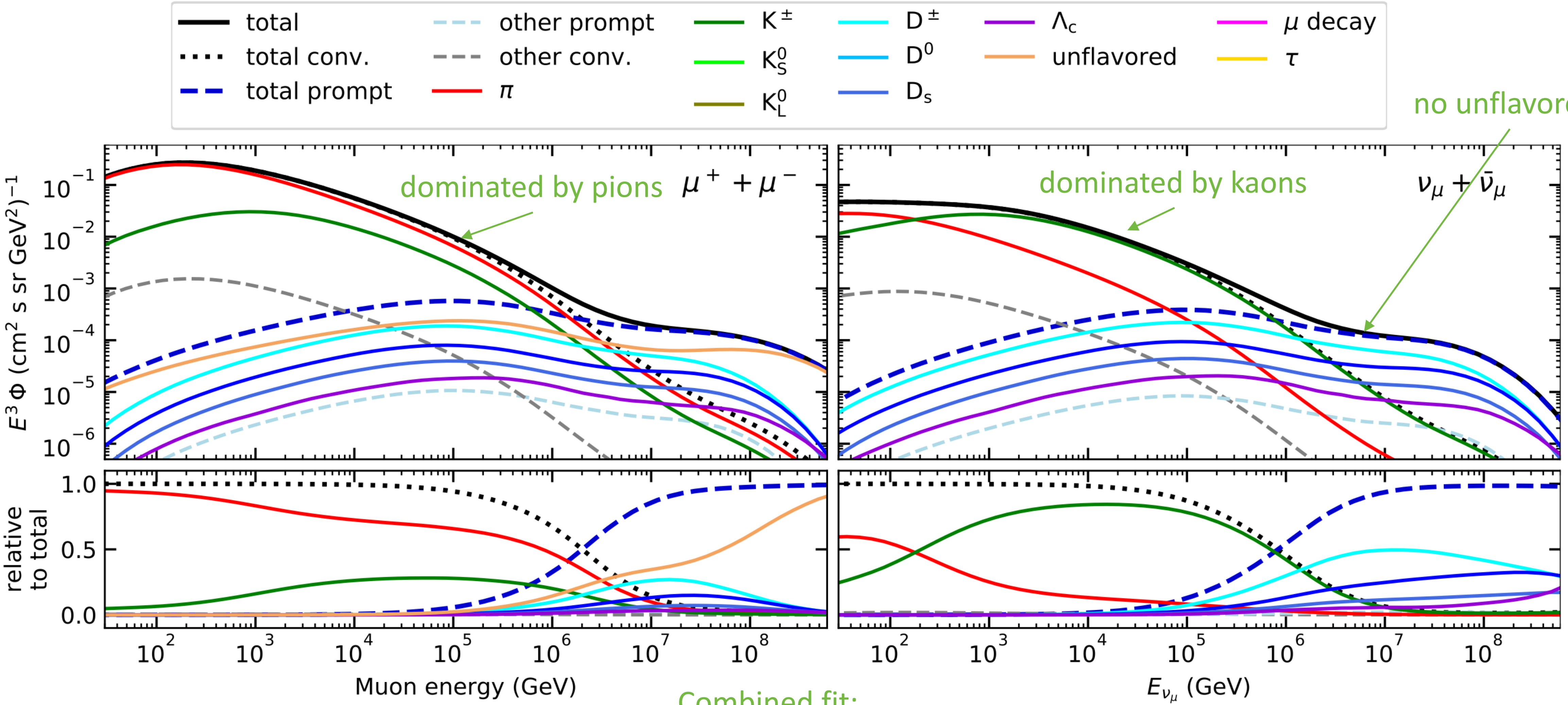


"muon bundles"



Prompt atmospheric muons and neutrinos

10.1103/PhysRevD.100.103018



Combined fit:

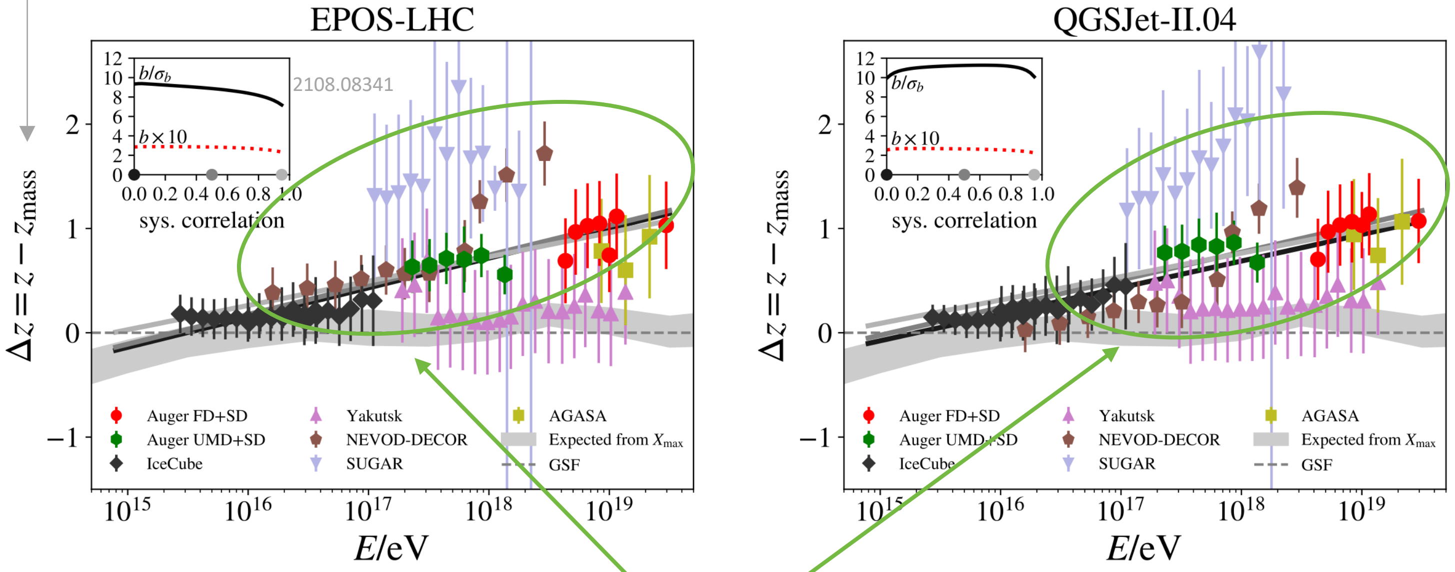
- handle on pion/kaon ratio
- handle on charmed mesons

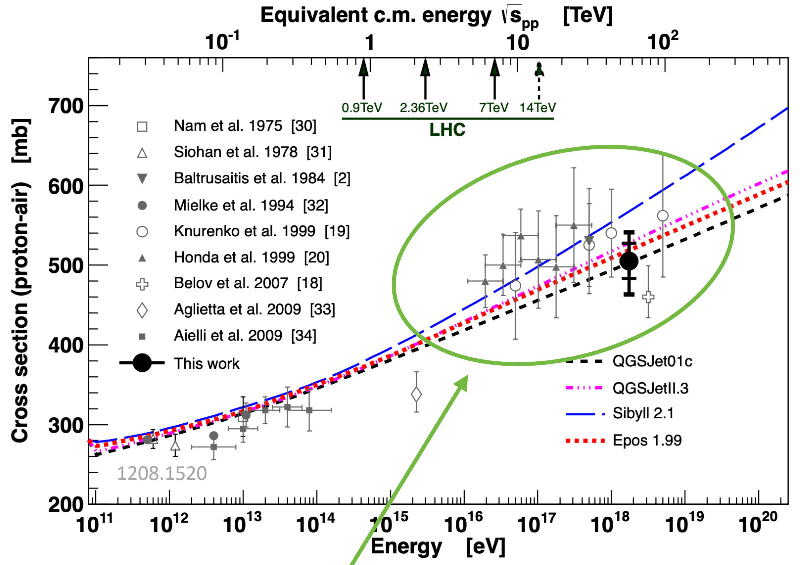
Muon puzzle and model uncertainties

"muon number"

Expected z
("muon number")

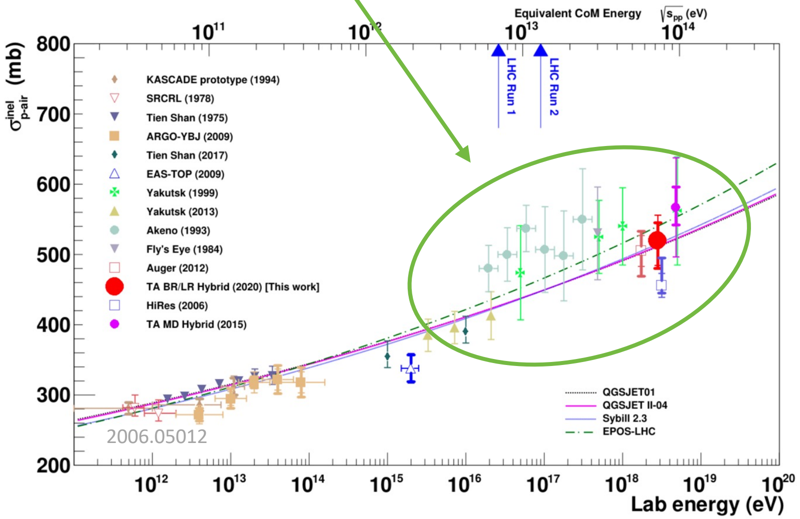
$$z = \frac{\ln\langle N_\mu\rangle - \ln\langle N_\mu\rangle_p}{\ln\langle N_\mu\rangle_{\mathrm{Fe}} - \ln\langle N_\mu\rangle_p}$$





2012

➢ Uncertainties at E > 10 PeV

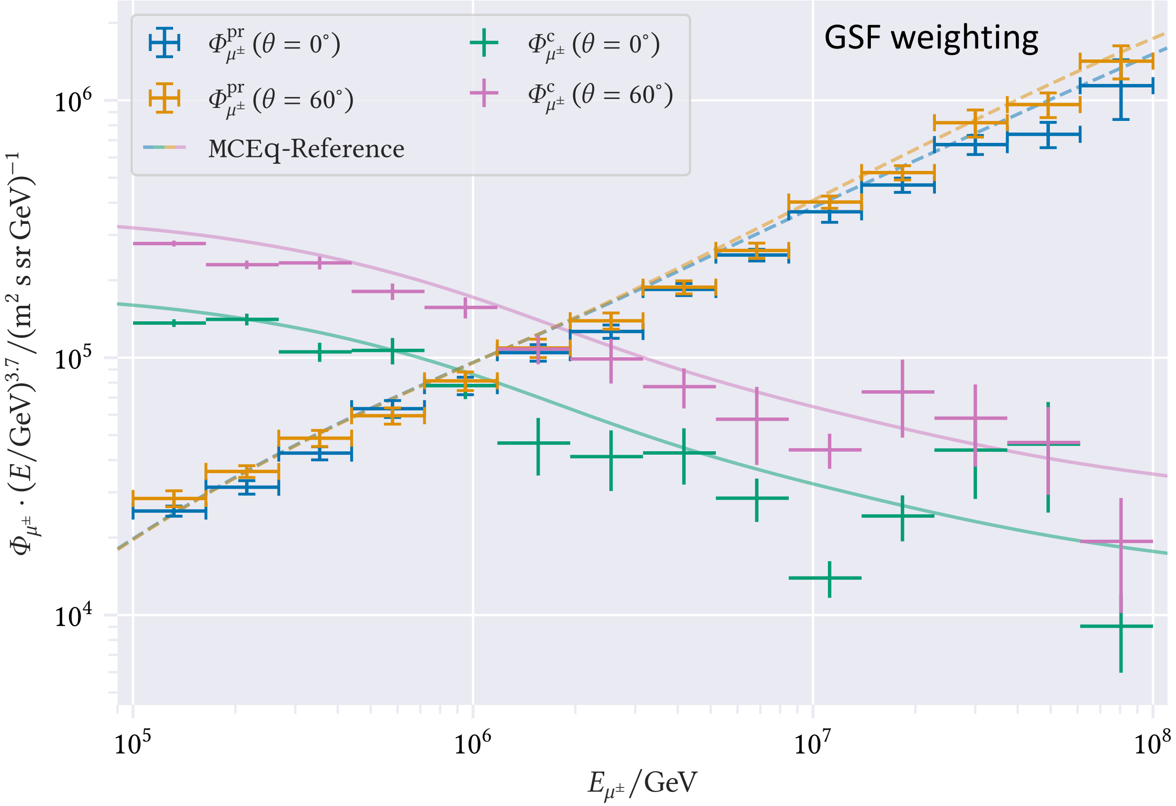


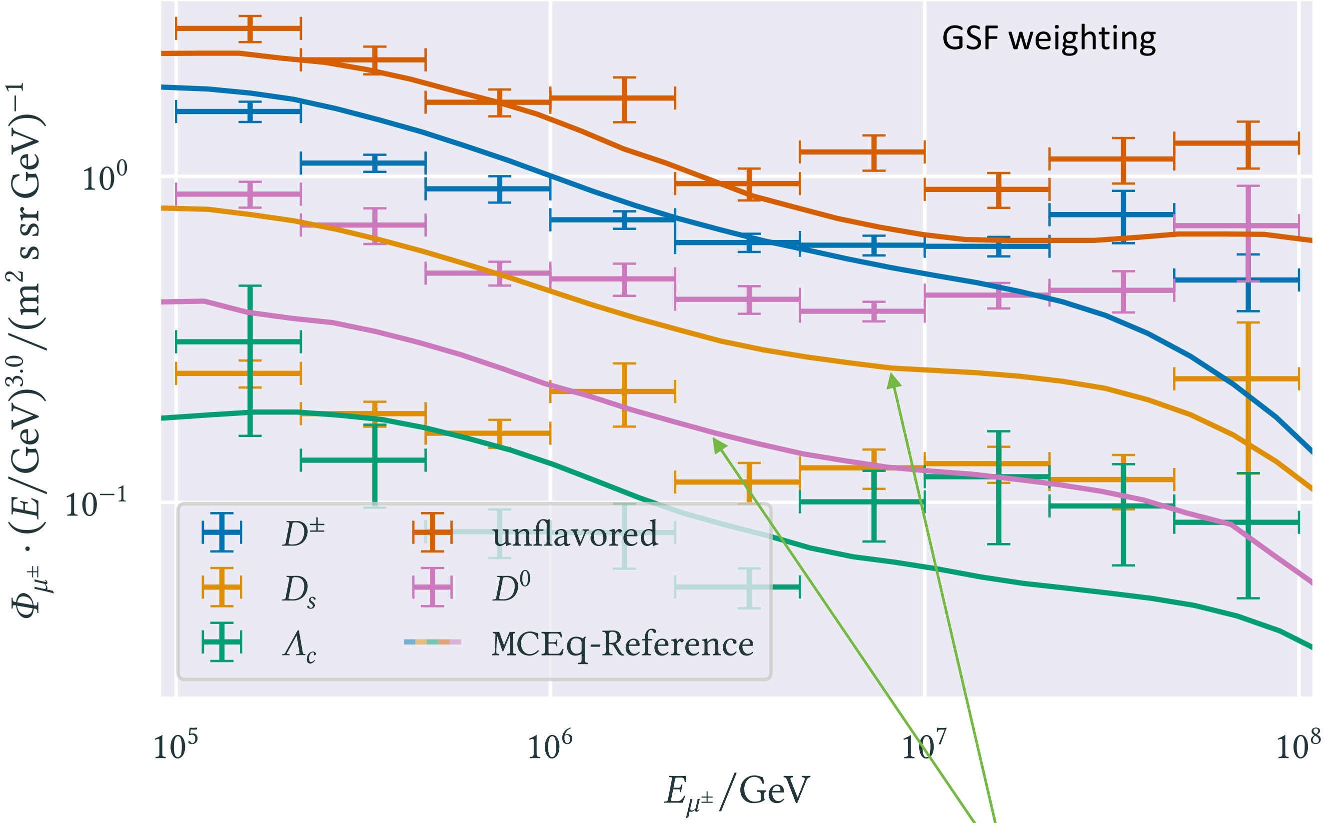
2020

➢ More muons measured than simulated for E > 40 PeV ~ cms 8 TeV

CORSIKA 7 tagging and MCEq comparison

MCEq: tool to numerically solve the cascade equations that describes the evolution of particle densities as they propagate through a gaseous, dense medium
https://github.com/mceq-project/MCEq



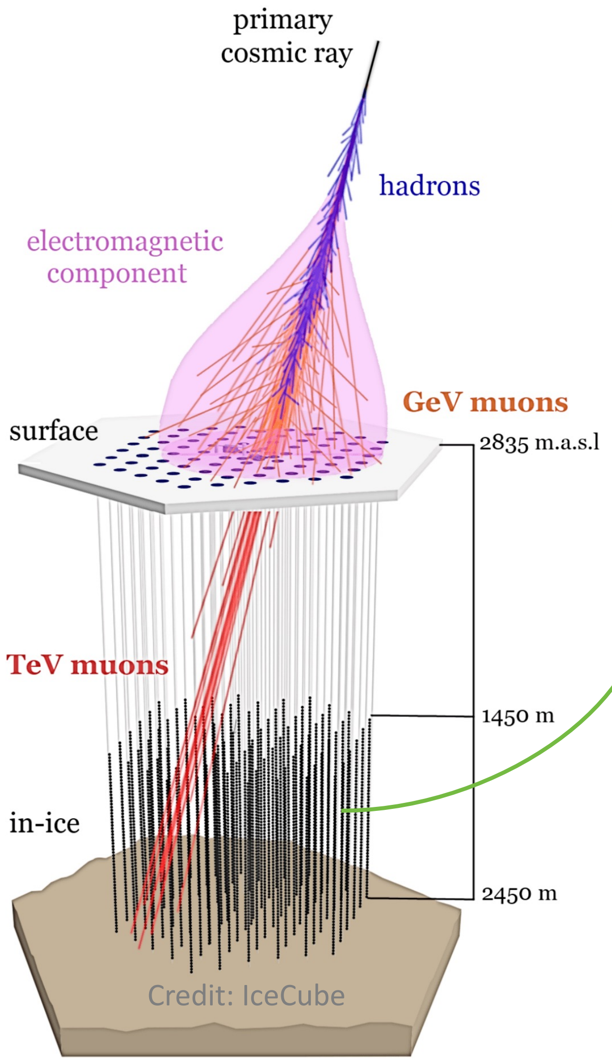


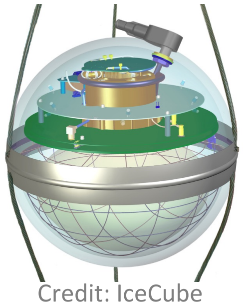
- Good agreement in total prompt and conv muon flux

- D^0 and D_s are swapped here but this is fixed in MCEq

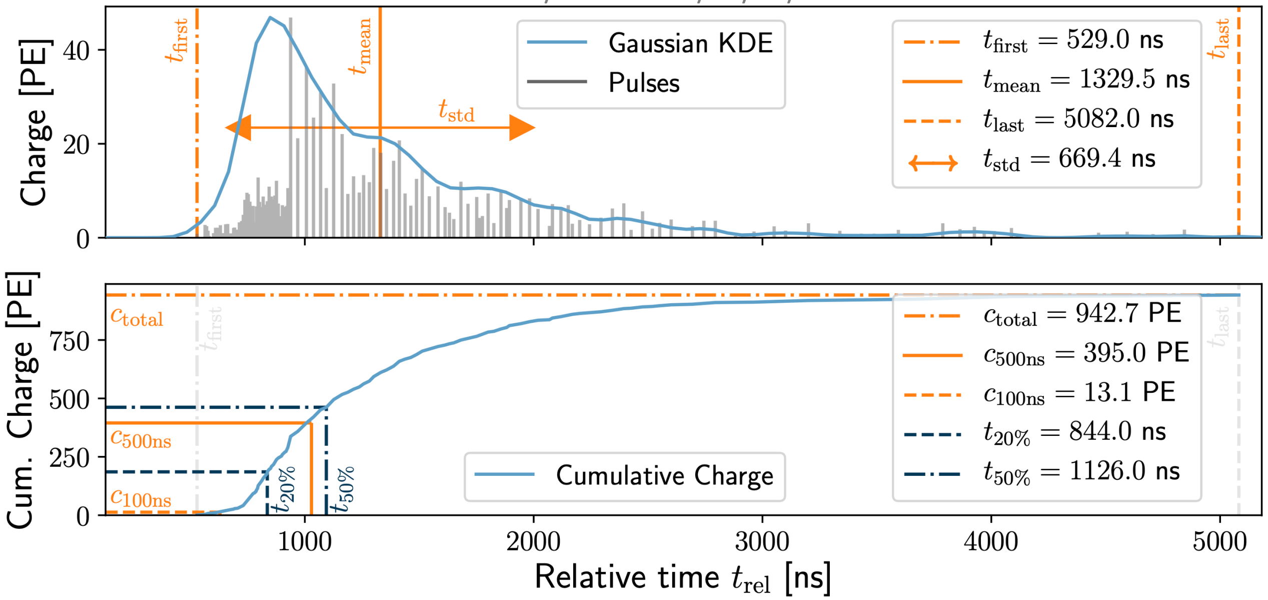
mention panama

Machine Learning Reconstructions





10.1088/1748-0221/16/07/P07041



Convolutional Neural Network

Inputs

- c_{total}: Total charge
  - Sum of charge
- t_{first}: Relative time of first pulse
  - Relative to total time offset
- t_{std}: Standard deviation of first pulse
  - Charge weighted standard deviation of pulse times

Outputs

- Direction
- Stopping point
- Entry point
- Energy at entry/surface
- …

Level5: quality cuts

containment cuts	>	<
length in detector	1000 m	2000 m
entry pos x, y	-750 m	750 m
entry pos z	-500 m	750 m
center pos x, y	-550 m	550 m
center pos z	-650 m	650 m

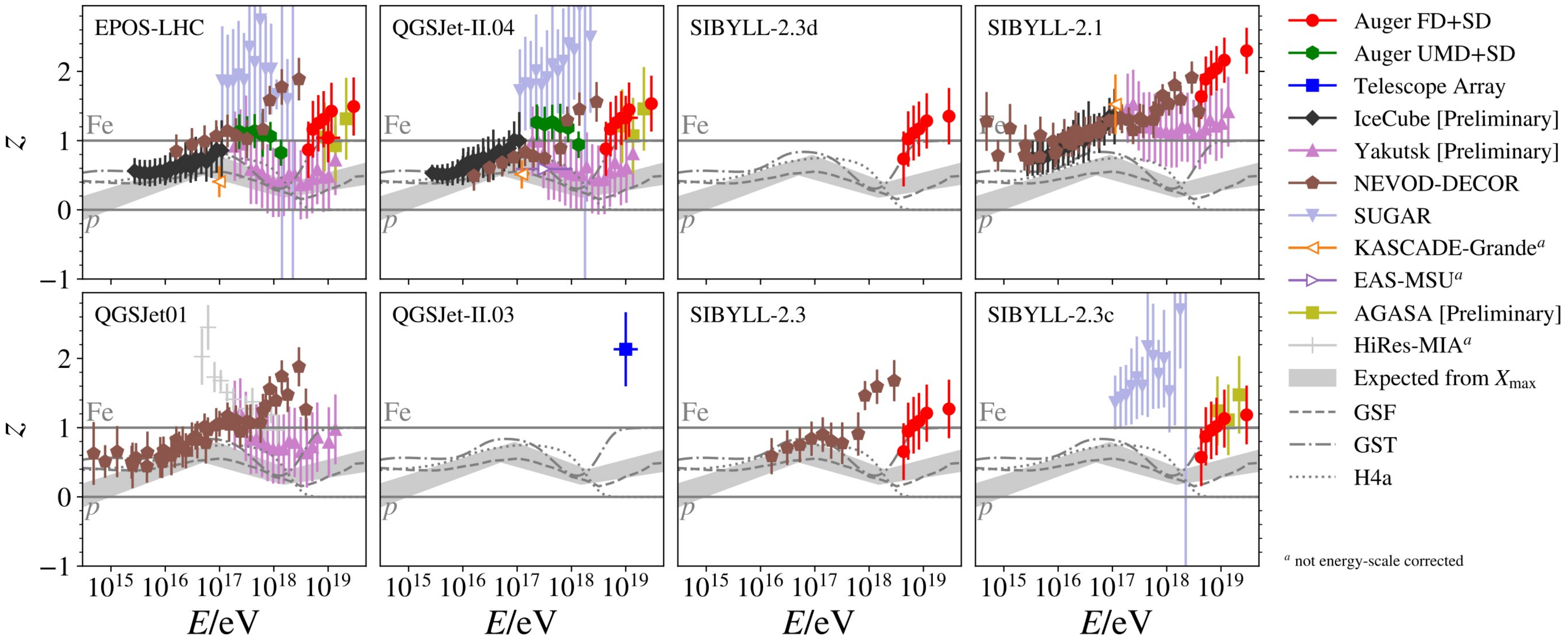
neutrino cuts	>	<
cos(zenith)	0.2	
length	5000 m	15000 m

uncertainty cuts	<
bundle energy at entry	0.9 log10(GeV)
bundle energy at surface	2.0 log10(GeV)
zenith	0.1 rad
azimuth	0.2 rad
entry pos x , y, z	42 m
center pos x, y, z	50 m
entry pos time	200 ns
center pos time	600 ns
length in detector	160 m
length	2000 m

The Muon Puzzle

"muon number"

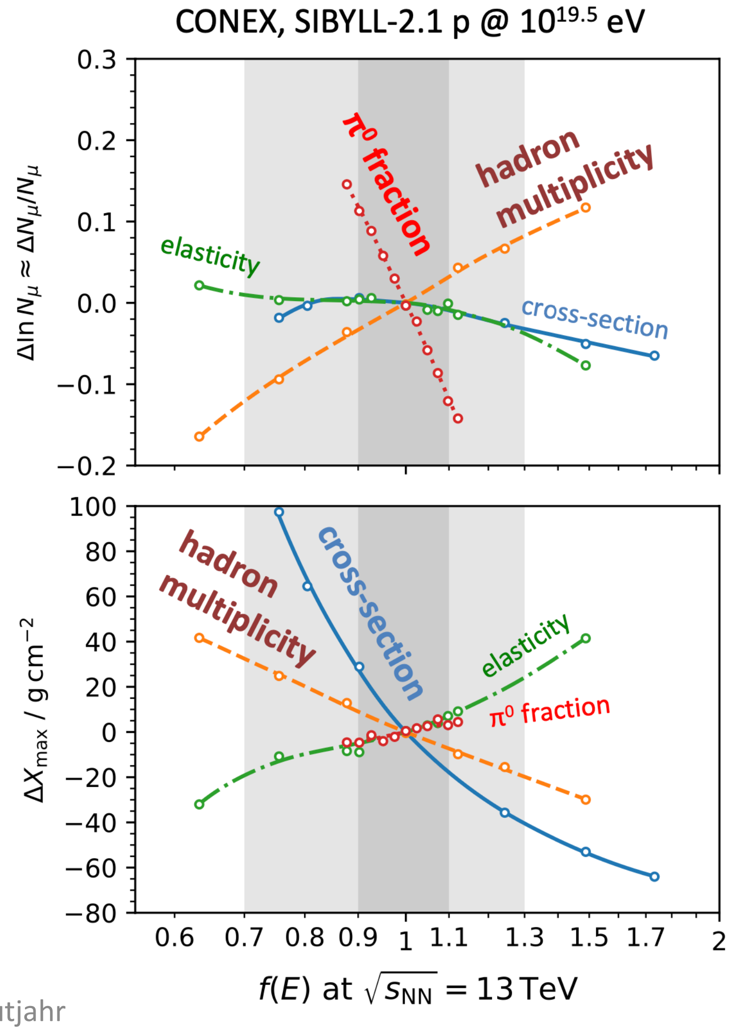
$$z = \frac{\ln\langle N_\mu\rangle - \ln\langle N_\mu\rangle_p}{\ln\langle N_\mu\rangle_{\mathrm{Fe}} - \ln\langle N_\mu\rangle_p}$$



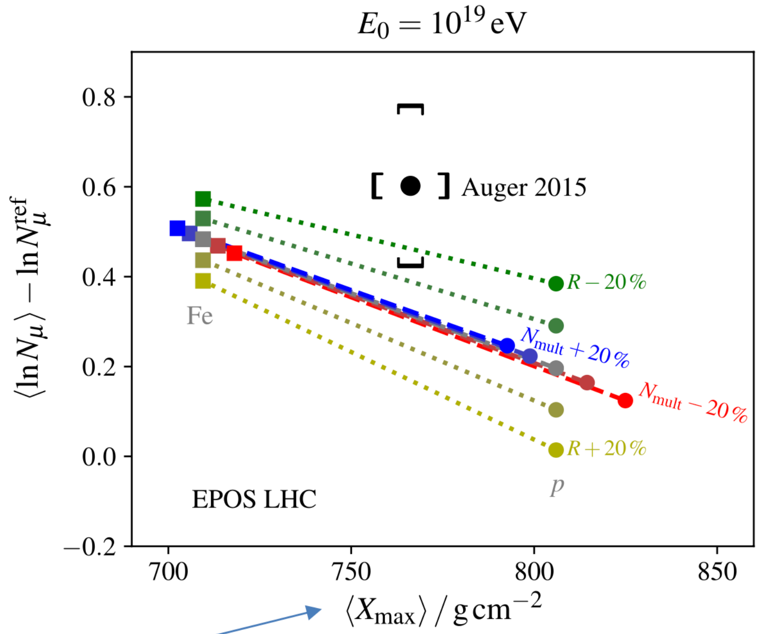
Arxiv: 2108.08341

Possible Solutions

R. Ulrich, R. Engel, M. Unger, PRD 83 (2011) 054026



S. Baur, HD, M. Perlin, T. Pierog, R. Ulrich, K. Werner, arXiv:1902.09265

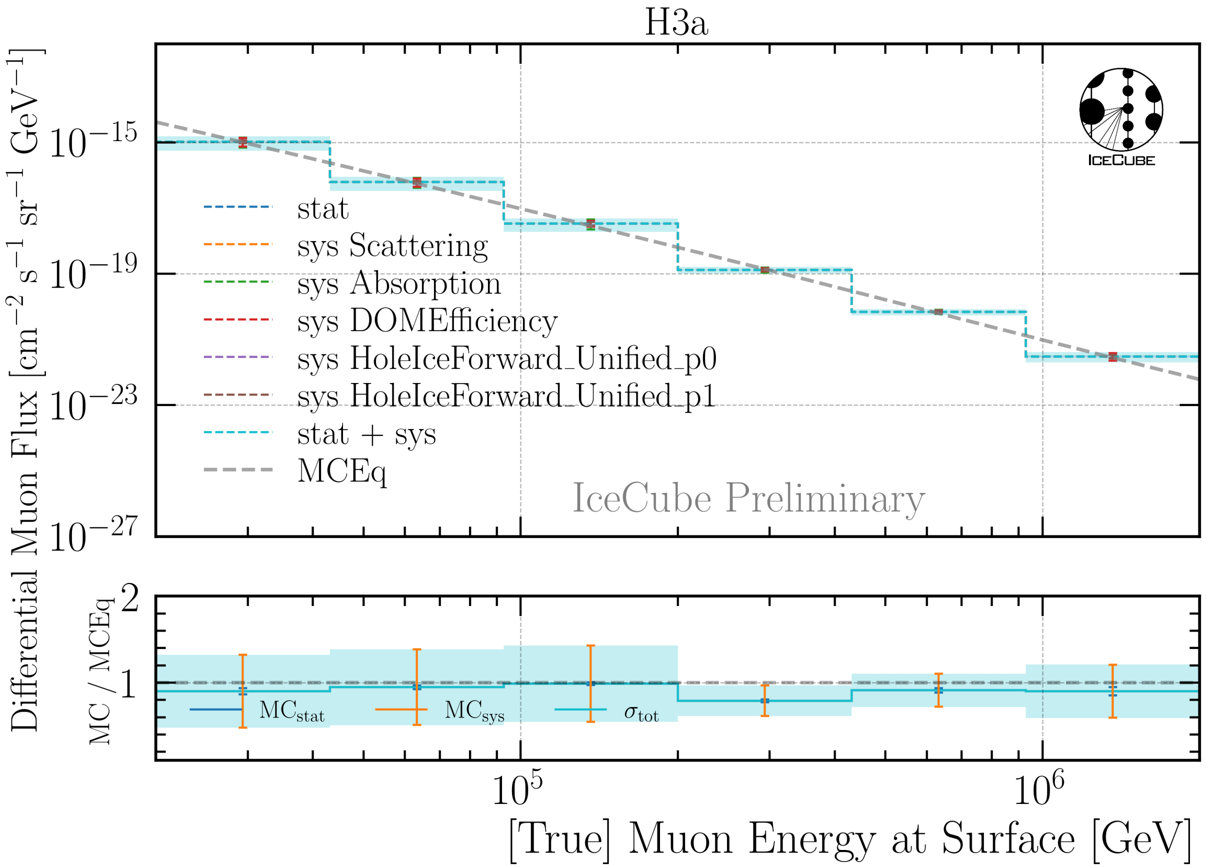


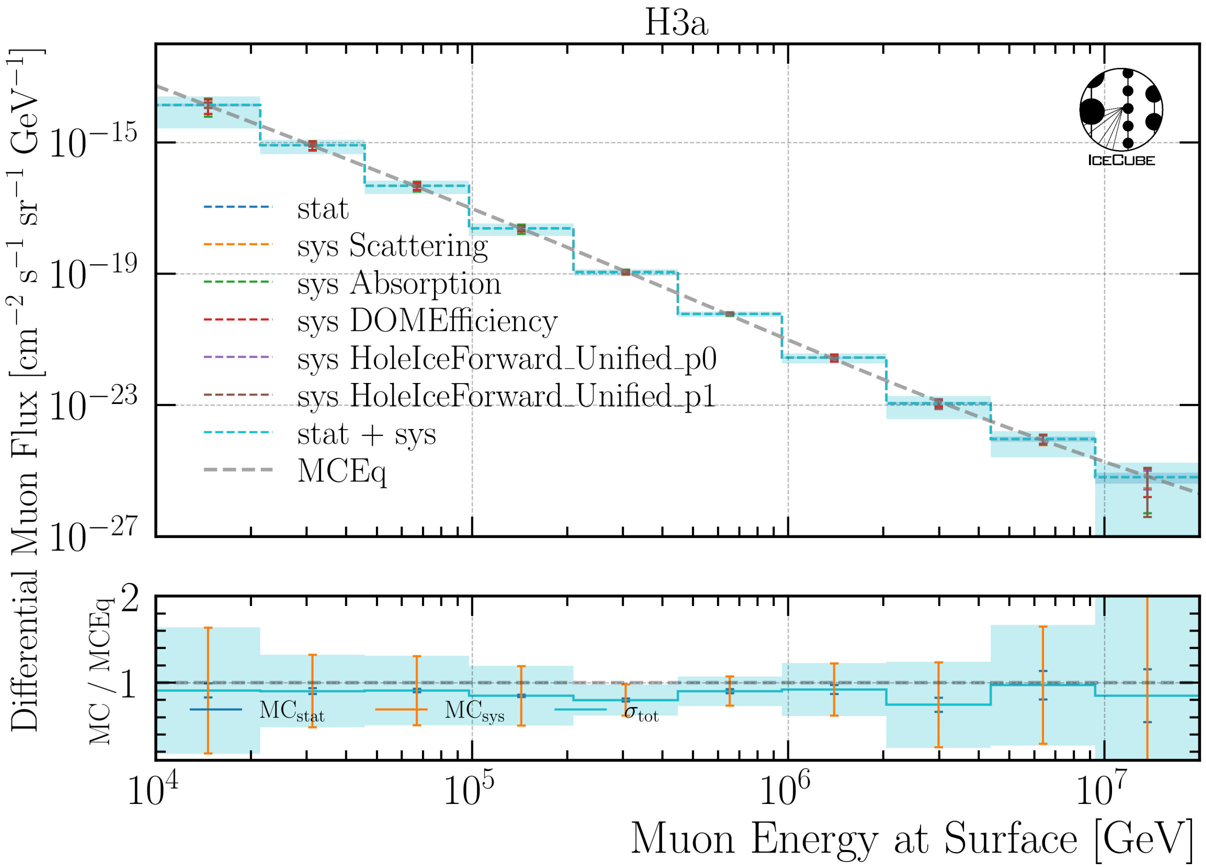
$$R = \frac{E_{\pi^0}}{E_{\text{other hadrons}}}$$

- Only changes to R can solve muon puzzle
- Small changes have large effect,
  R needs to be known to about 5 %

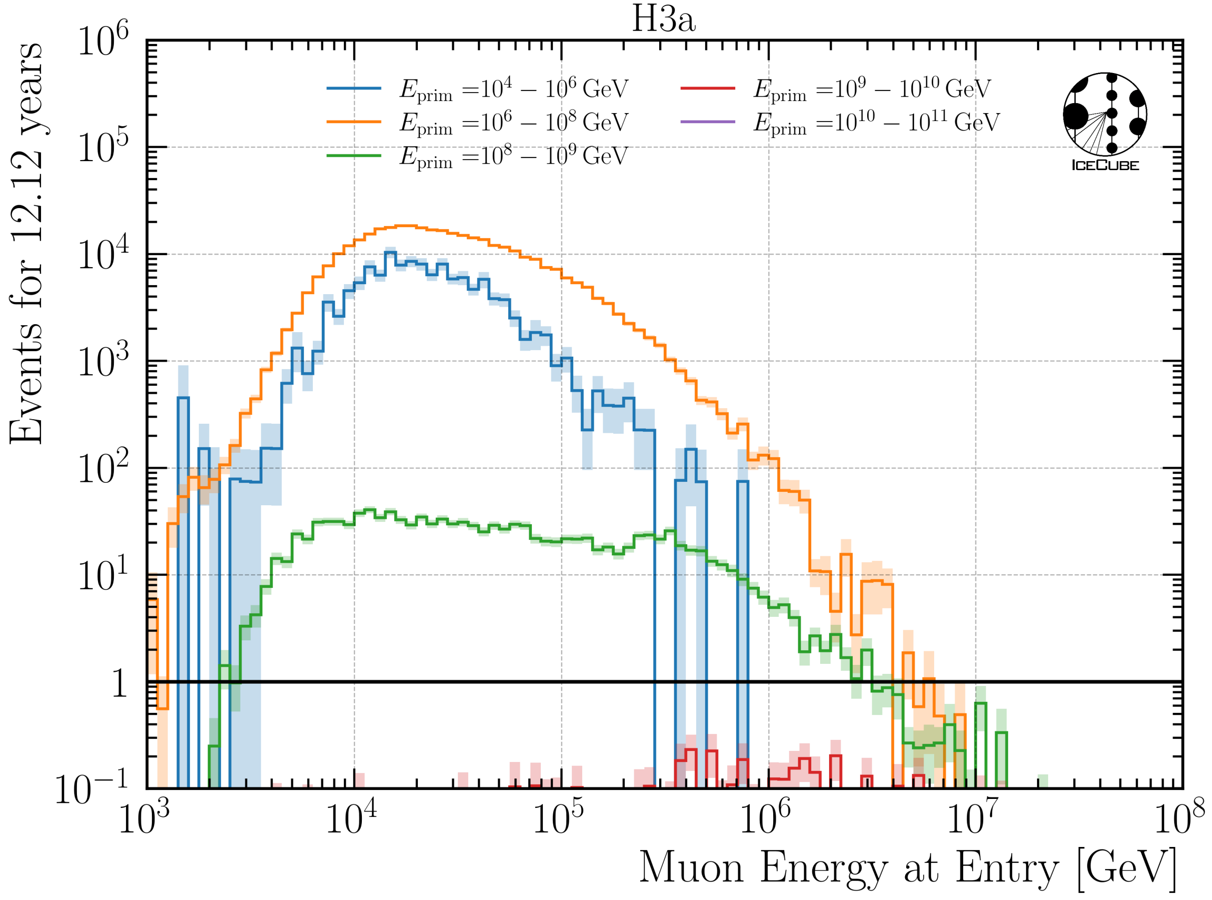
Slide taken from Hans Dembinski

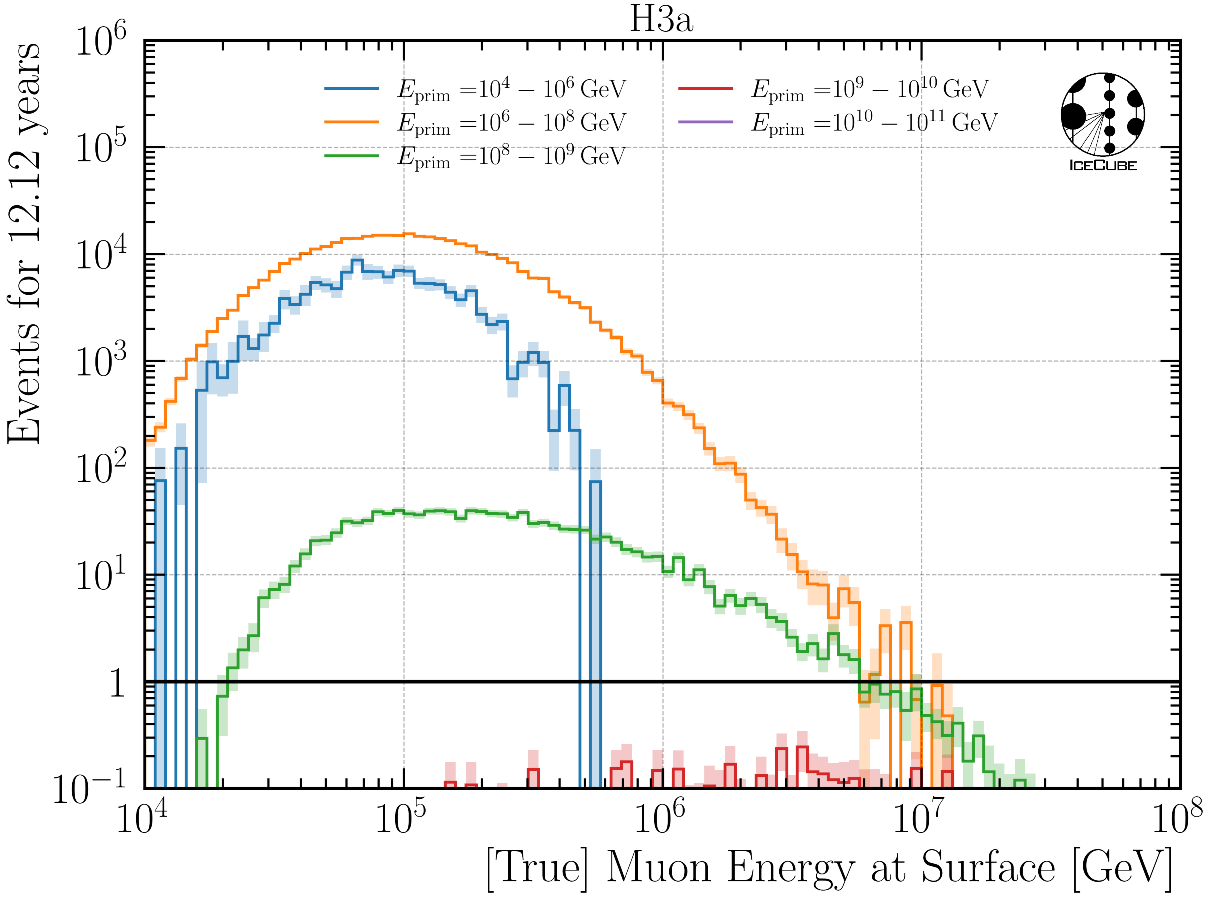
MCEq vs CORSIKA





MC Statistics: Fine binning





MC Statistics: Unfolding binning