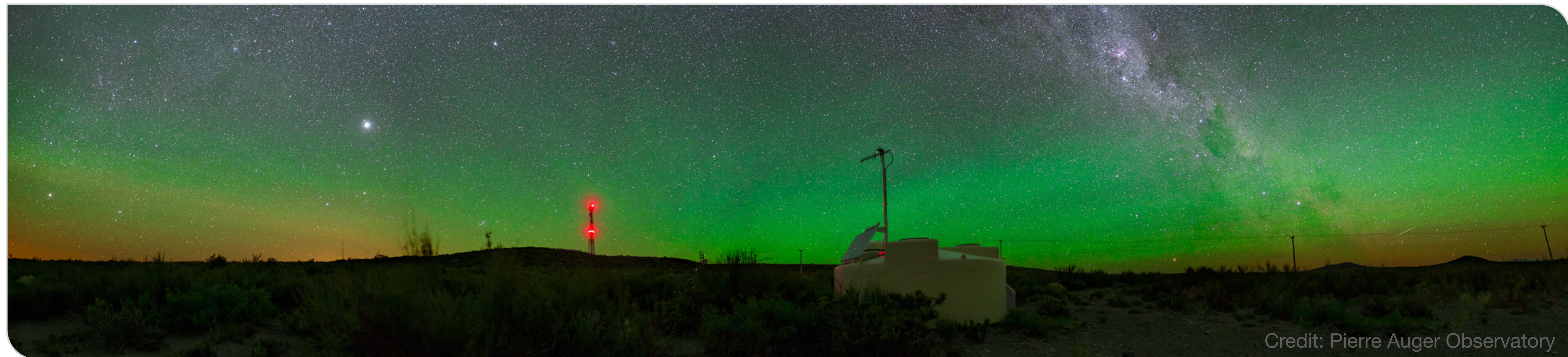


Inference of the Mass Composition of Cosmic Rays with energies between 3 and 100 EeV using the data of the Pierre Auger Observatory and Deep Learning

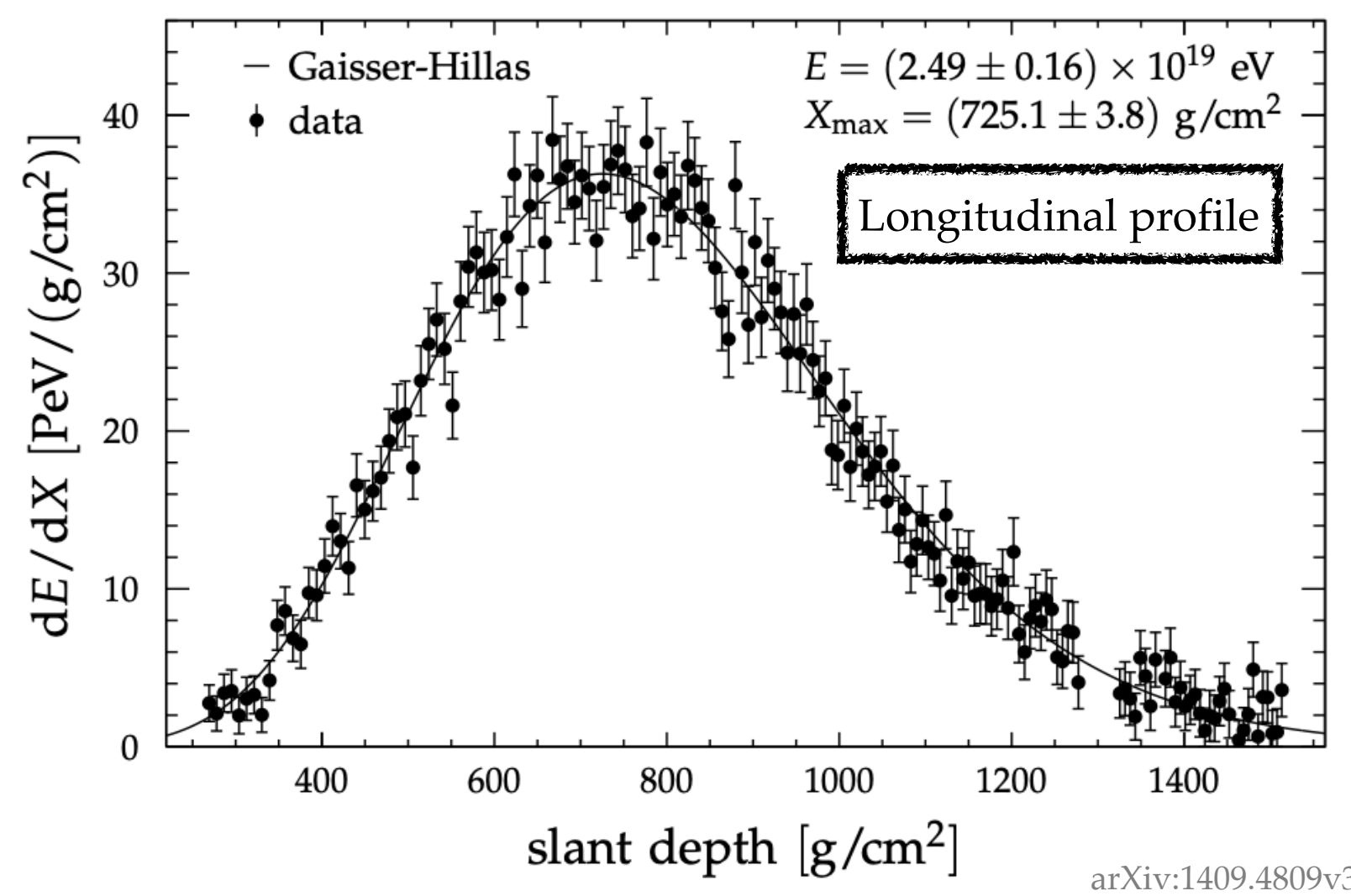
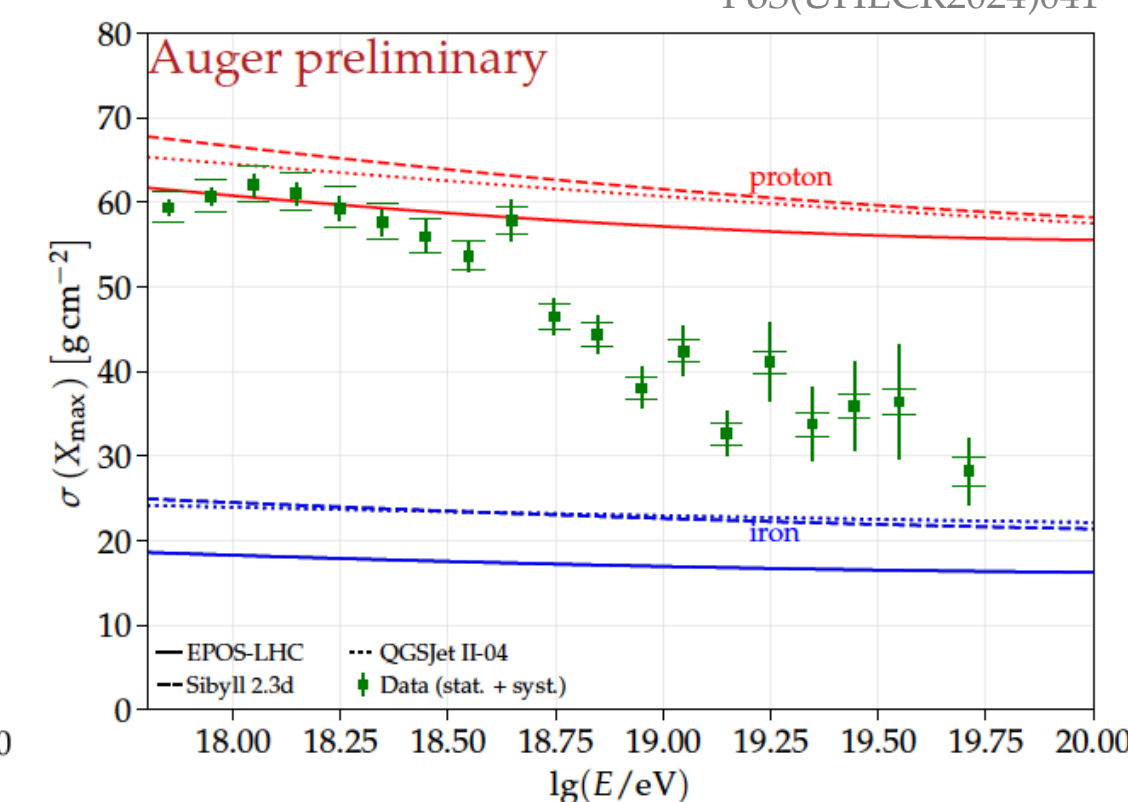
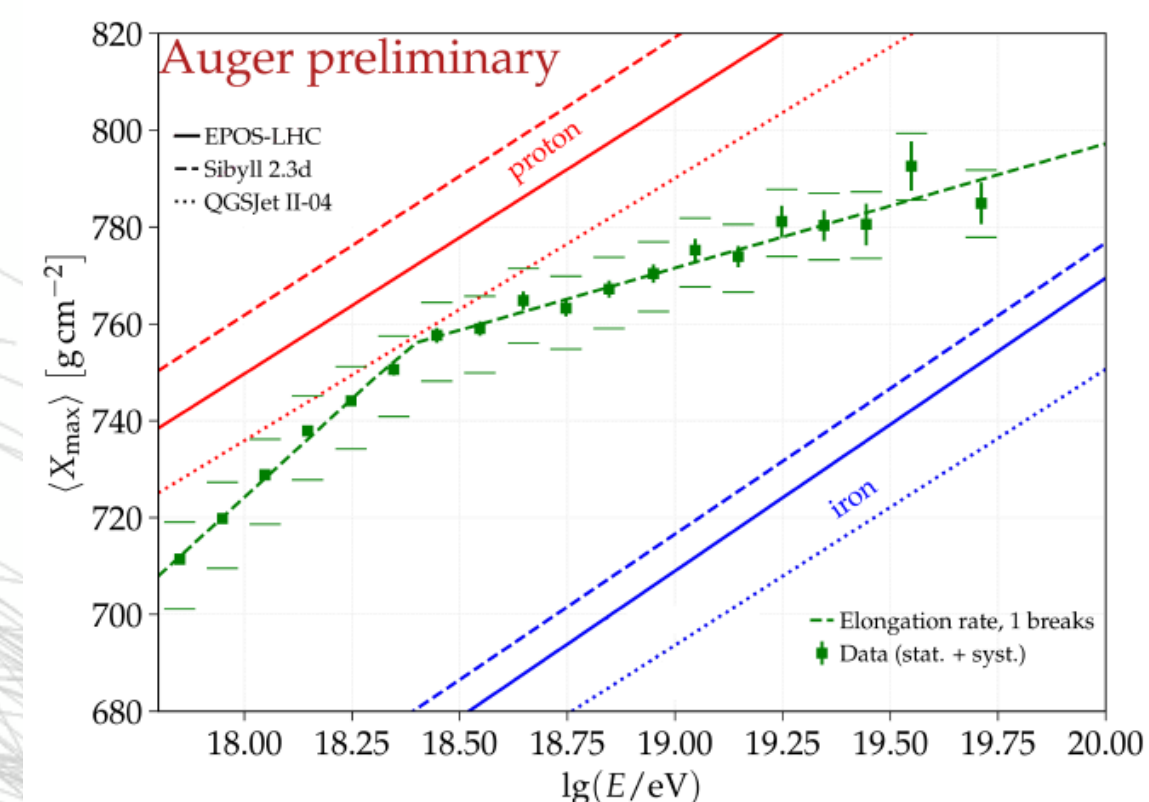
Berenika Čermáková for the Pierre Auger Collaboration



Credit: Pierre Auger Observatory

Introduction

- Ultrahigh-energy cosmic rays (UHECRs)
 - Extraterrestrial charged particles with ultra-high energies (above 10^{18} eV)
 - Acceleration mechanisms and sources are unknown



- Determination of cosmic ray (CR) mass
 - Degeneracy of mass and energy
 - Due to scarcity only indirect detection possible
 - Estimation relies on quantities statistically encoding mass (number of muons, depth of the shower maximum $X_{\max}$)

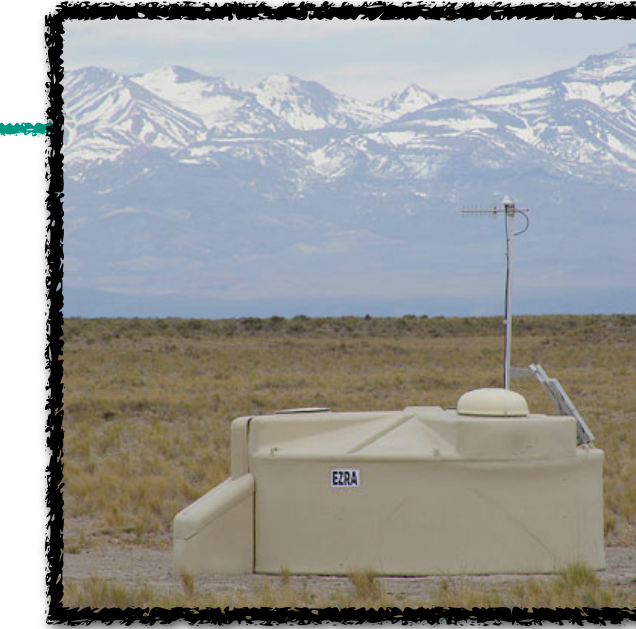
Pierre Auger Observatory



- Malargue, Argentina
- 3 000 km²
- Surface Detector:
 - 1 600 stations
 - Multiple sub-detectors
 - Main detector: water-Cherenkov detectors in a triangular grid
- Fluorescence Detector:
 - Overlooking the SD
 - 15% duty cycle
 - Direct X_{max} measurement
 - Longitudinal profile

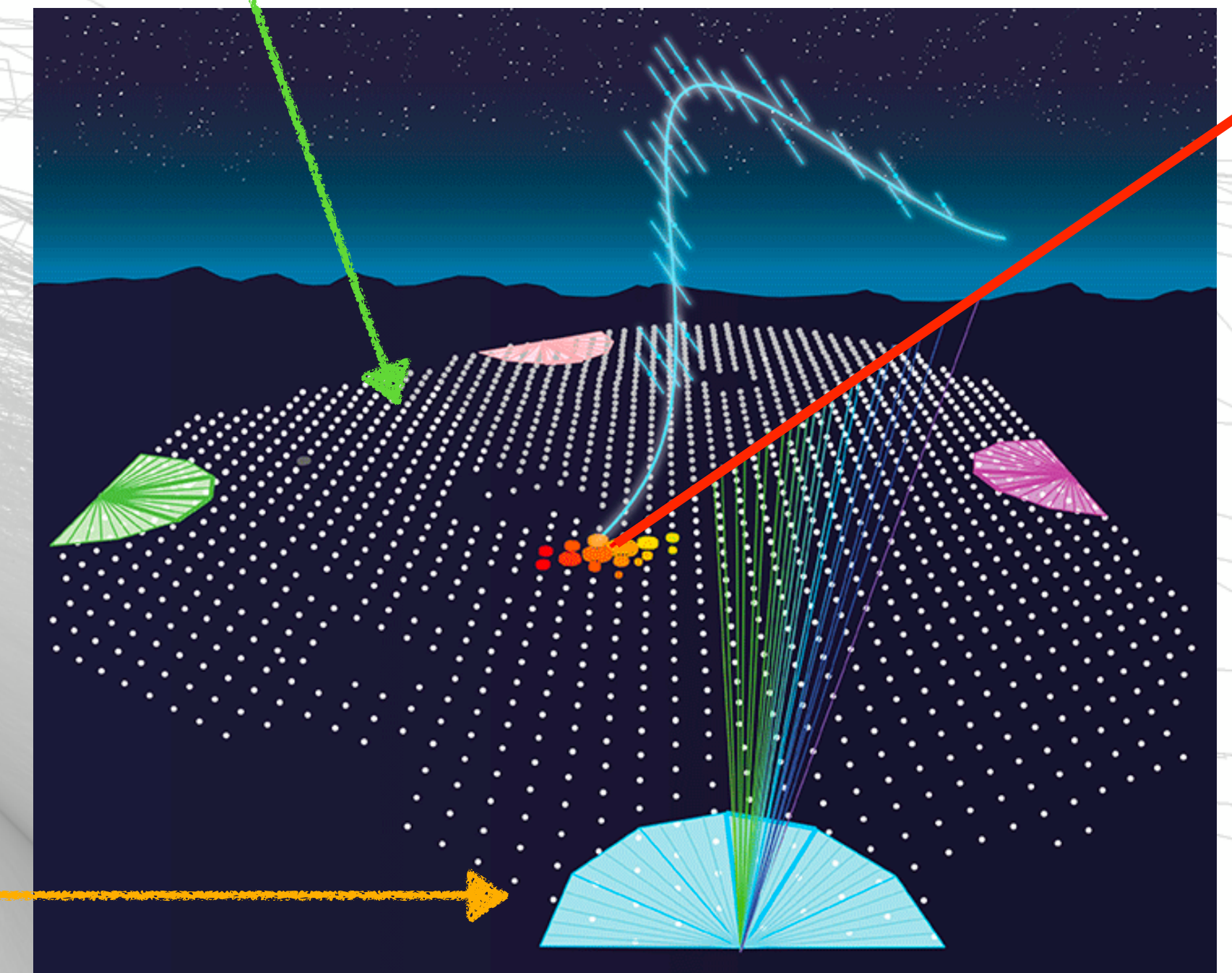


Fluorescence Detector

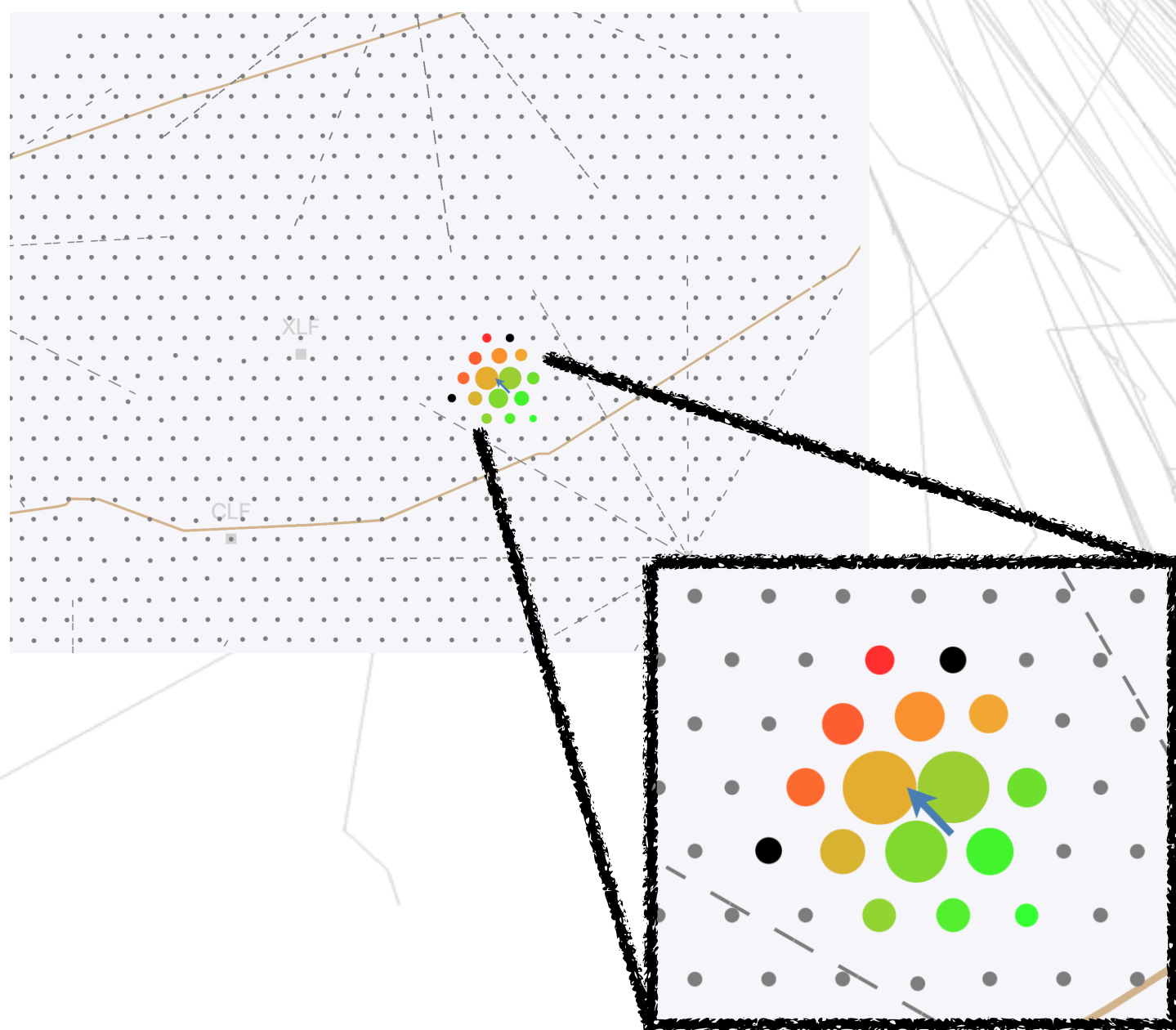


Water-Cherenkov
Detector

- 100 % duty cycle
- Air-shower footprint
- Muon traces measured with 3 PMTs

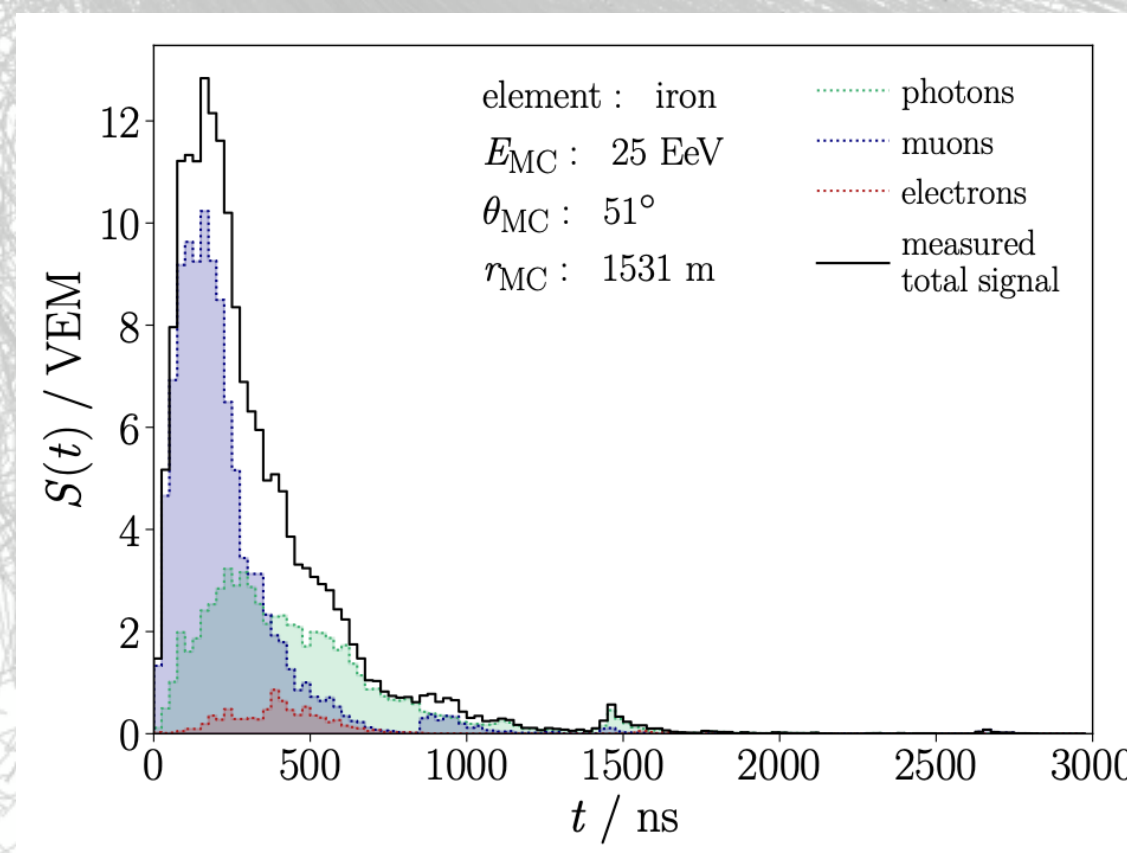
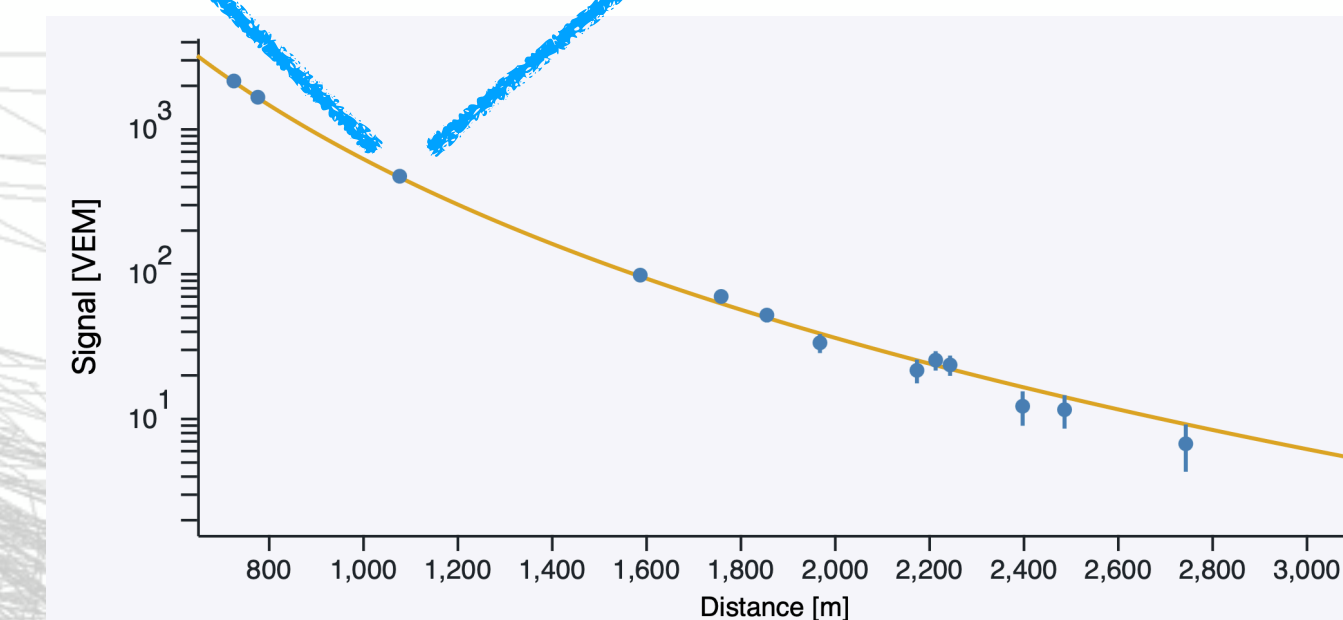
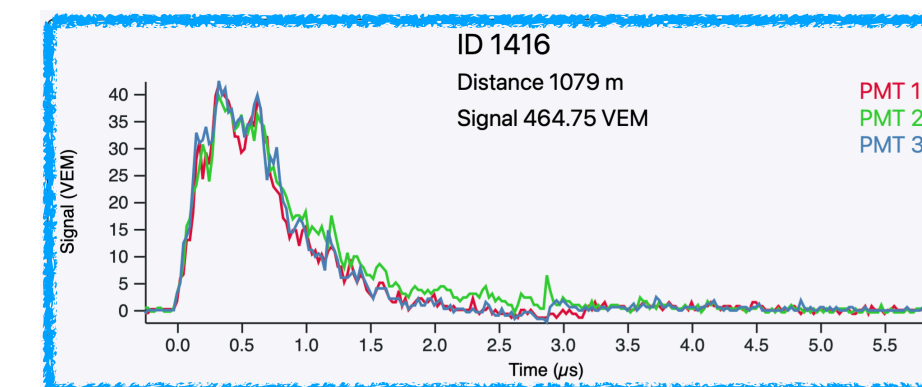


Surface detector signal



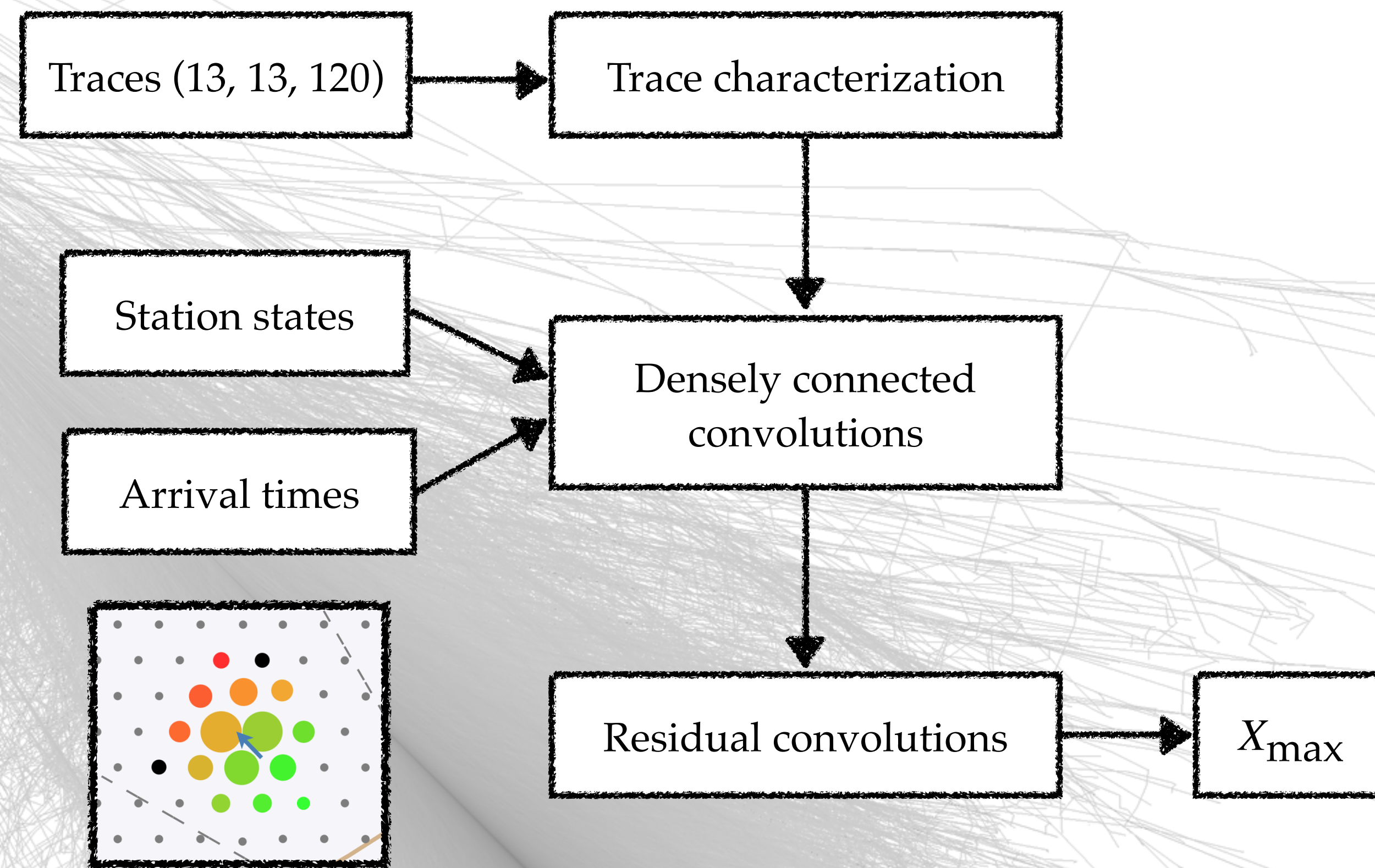
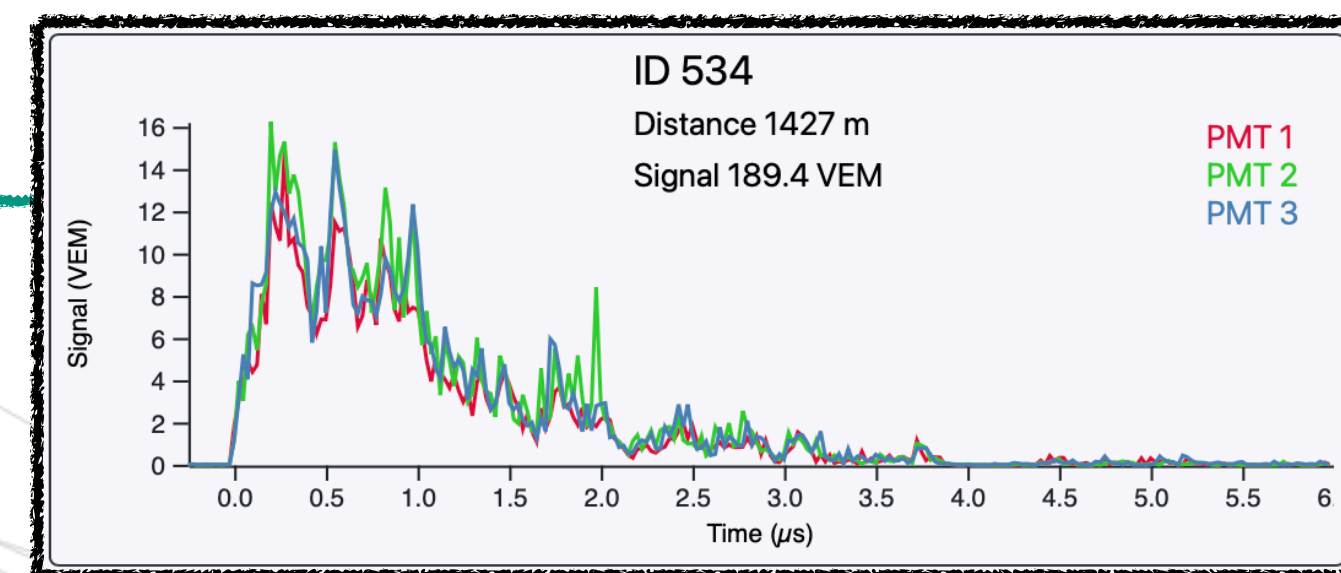
- Shower geometry
- Arrival times

- Signal traces from 3 PMTs in each WCD
- Total recorded signal depends on the energy of the primary CR and the distance from the shower core (reconstructed shower axis)
- Shape depends on the different shower components
 - Electromagnetic
 - Muonic

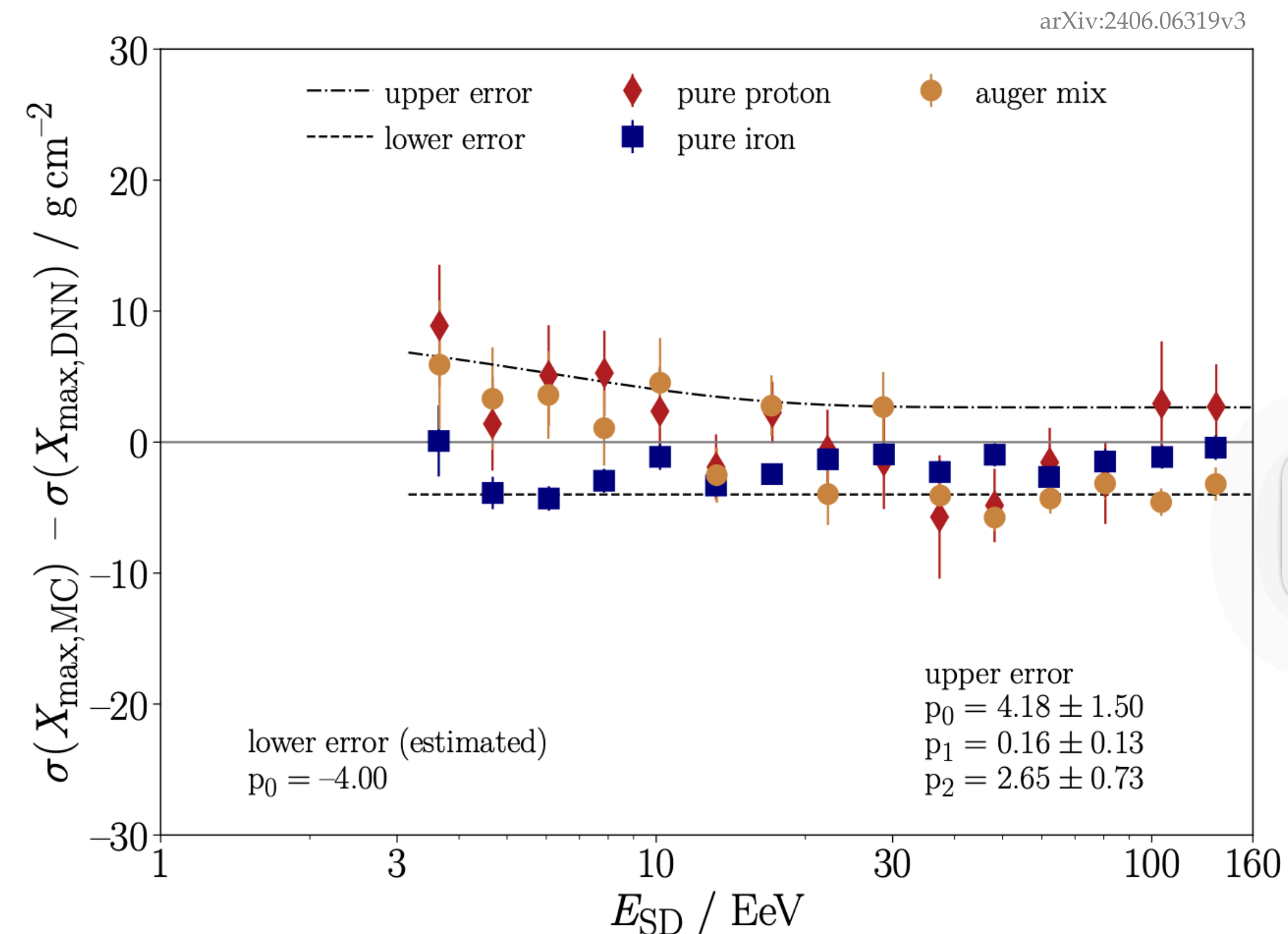
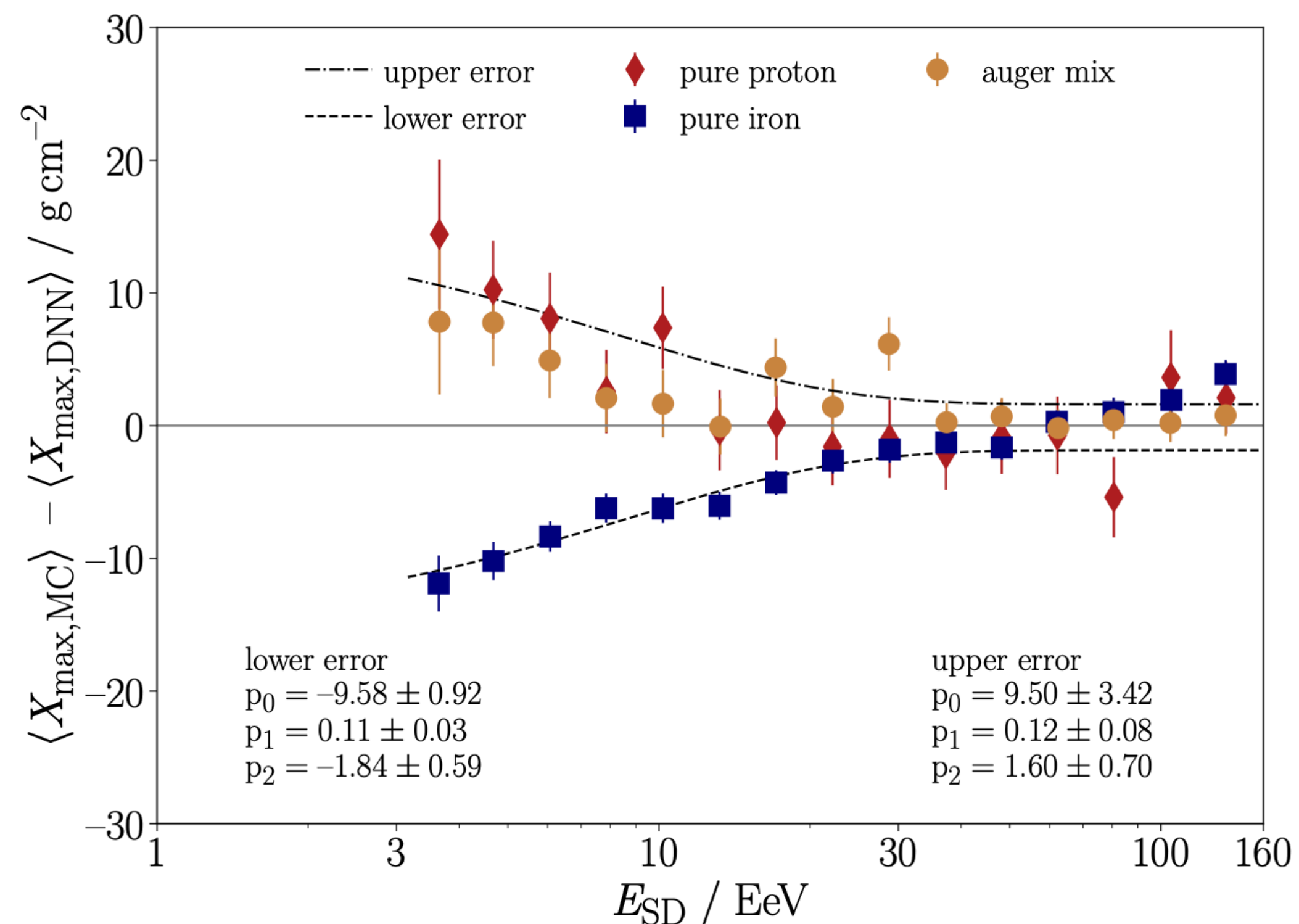


Mass estimation with DNN

- $X_{\max}$ reconstruction from a shower footprint
- Triangular to Cartesian grid - 13x13 stations
- Input normalization
- Training:
 - 400,000 simulated detector responses
 - CORSIKA with the EPOS LHC hadronic interaction model
 - Proton, helium, oxygen, iron
 - Energy between 1 and 160 EeV, spectral index $\gamma = -1$
 - Zenith angles $\theta < 65^\circ$



Performance on simulations



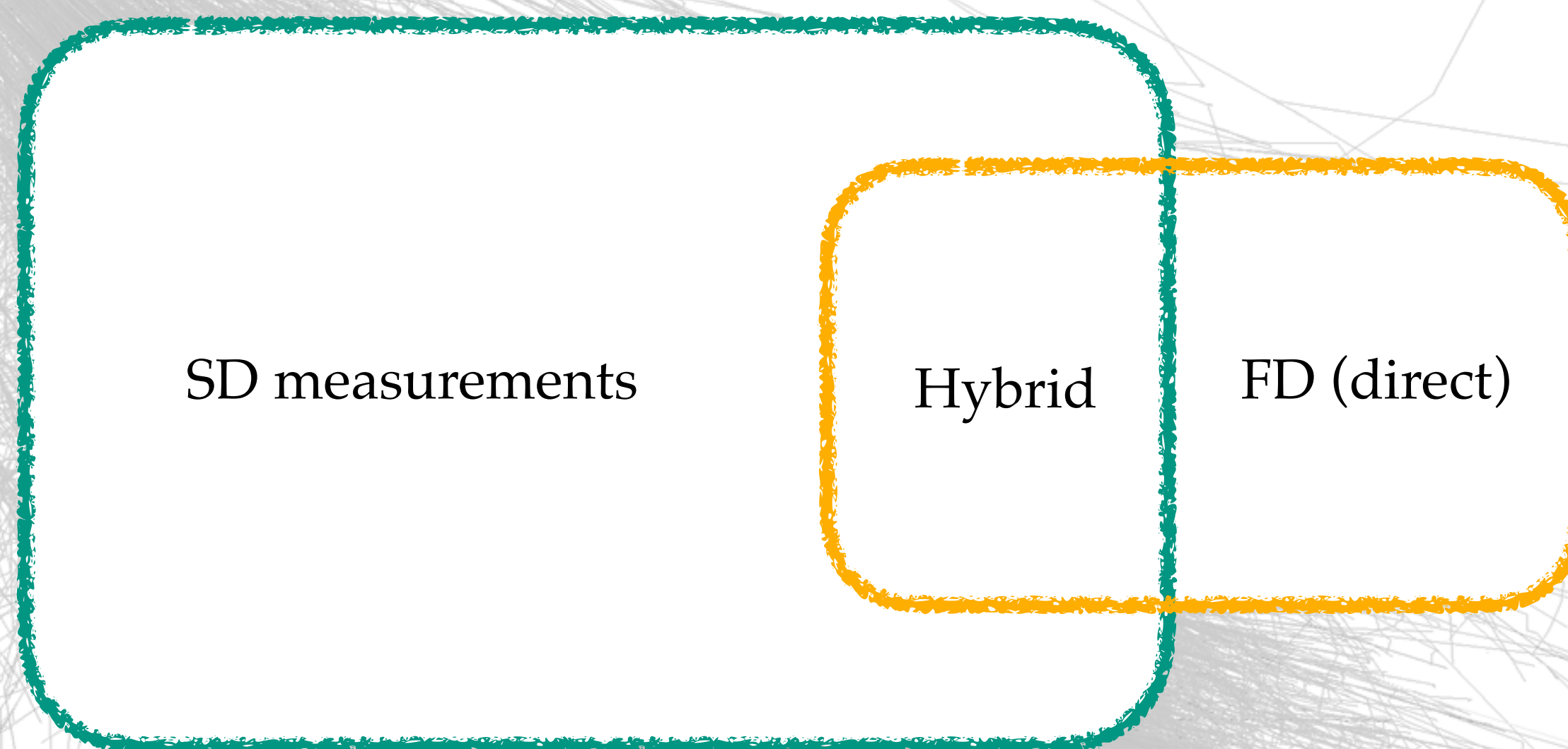
SD and Hybrid (SD+FD) datasets

■ SD dataset:

- Vertical events $\theta < 60^\circ$
- $\lg E/\text{eV} > 18.5$
- All stations in the hexagon around the hottest station (with the largest signal)
- Analysis-specific selection (high-quality SD cut, detector aging effects, core distance from the hottest station...)
- 48,824 events after cuts

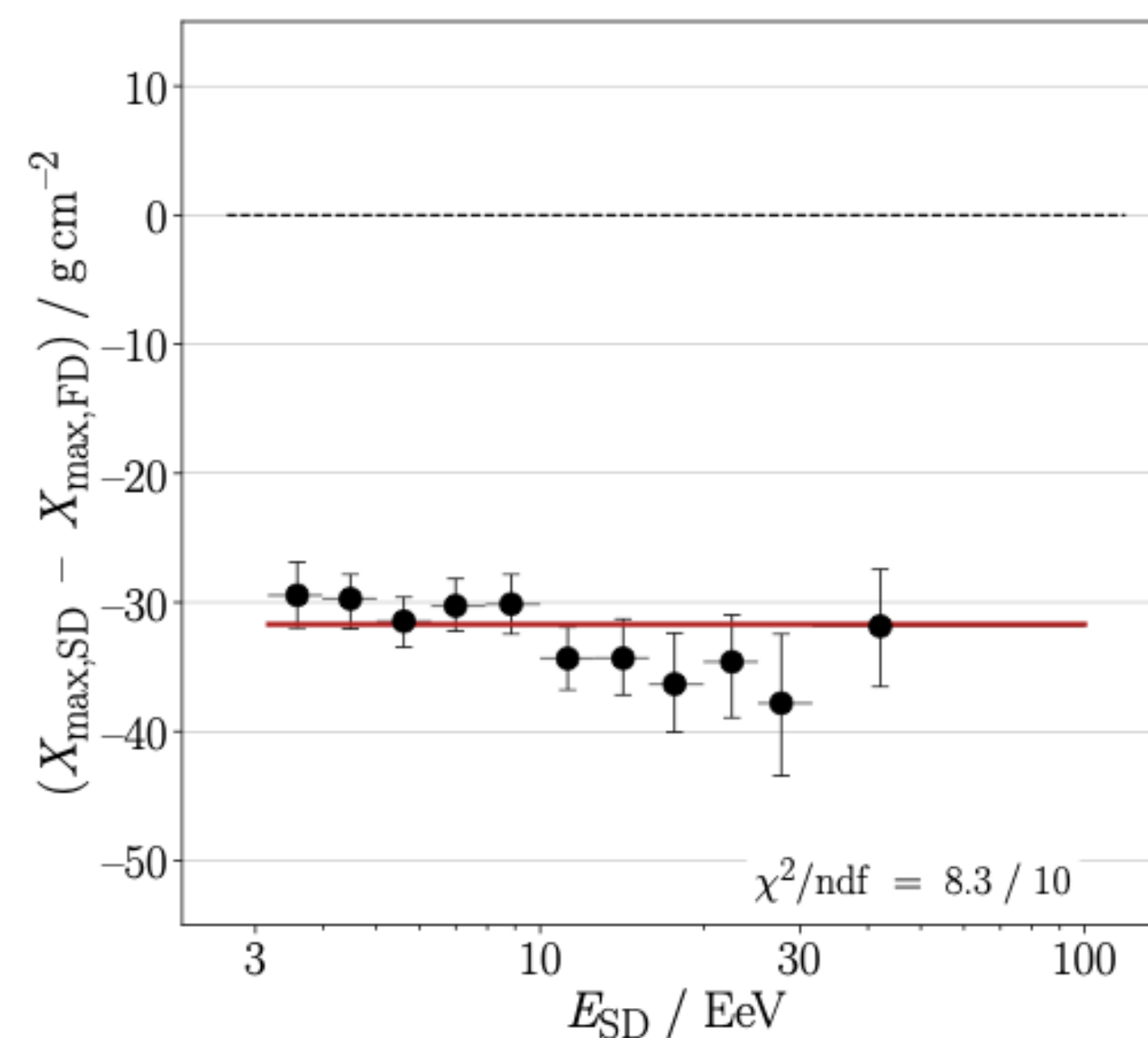
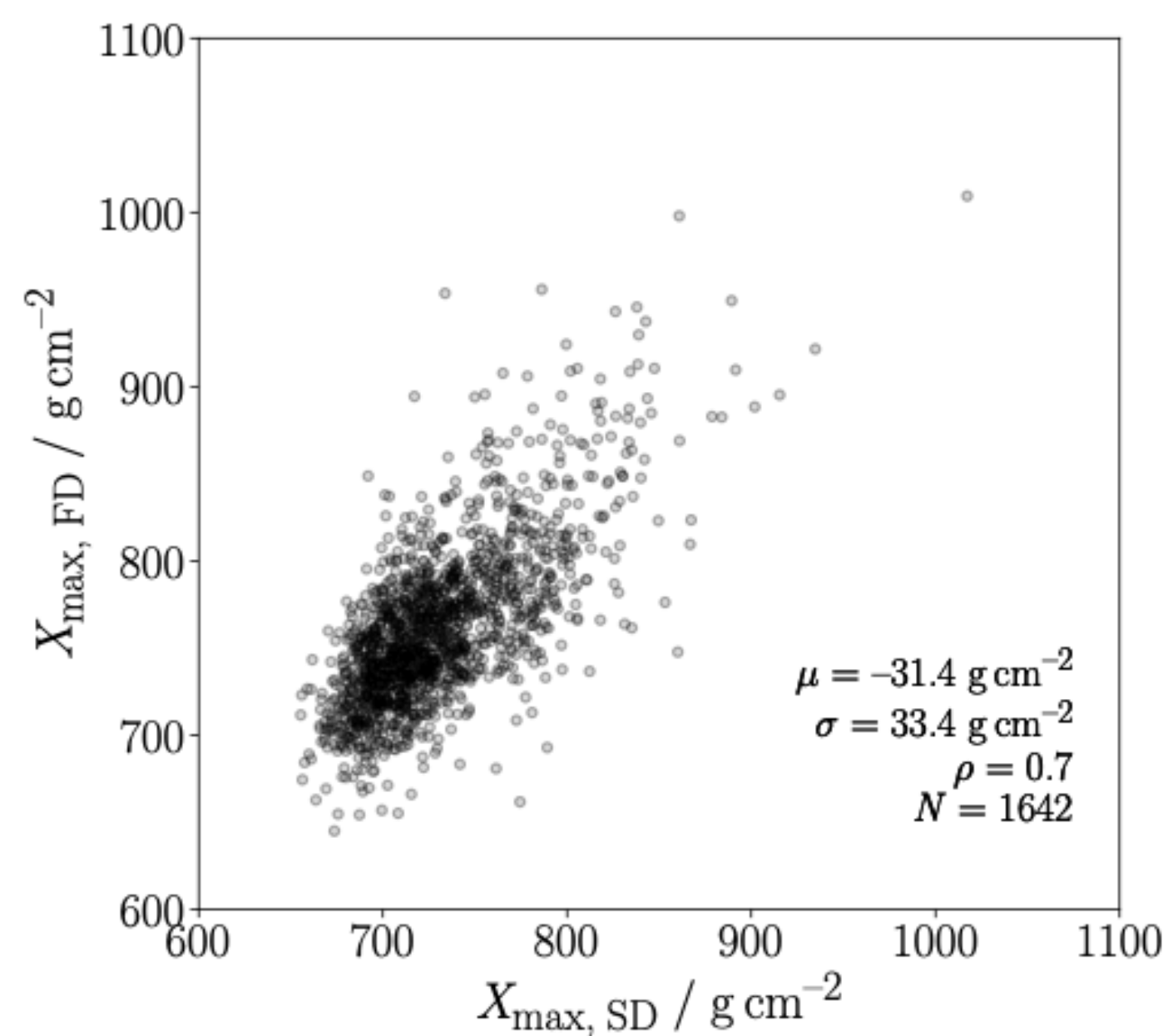
■ Hybrid dataset:

- Events measured with both SD and FD
- Allow for comparison between the direct measurement and NN prediction
- 1,642 events after high-quality selection



Calibration with hybrid events

- NN predictions show linear correlation to FD measurements of 0.7
- Absolute bias of $-31.4 \pm 0.8 \text{ g cm}^{-2}$
(currently available had. int. models underpredict the muonic component)
- Constant bias over energy: calibration to account for the effects mentioned above

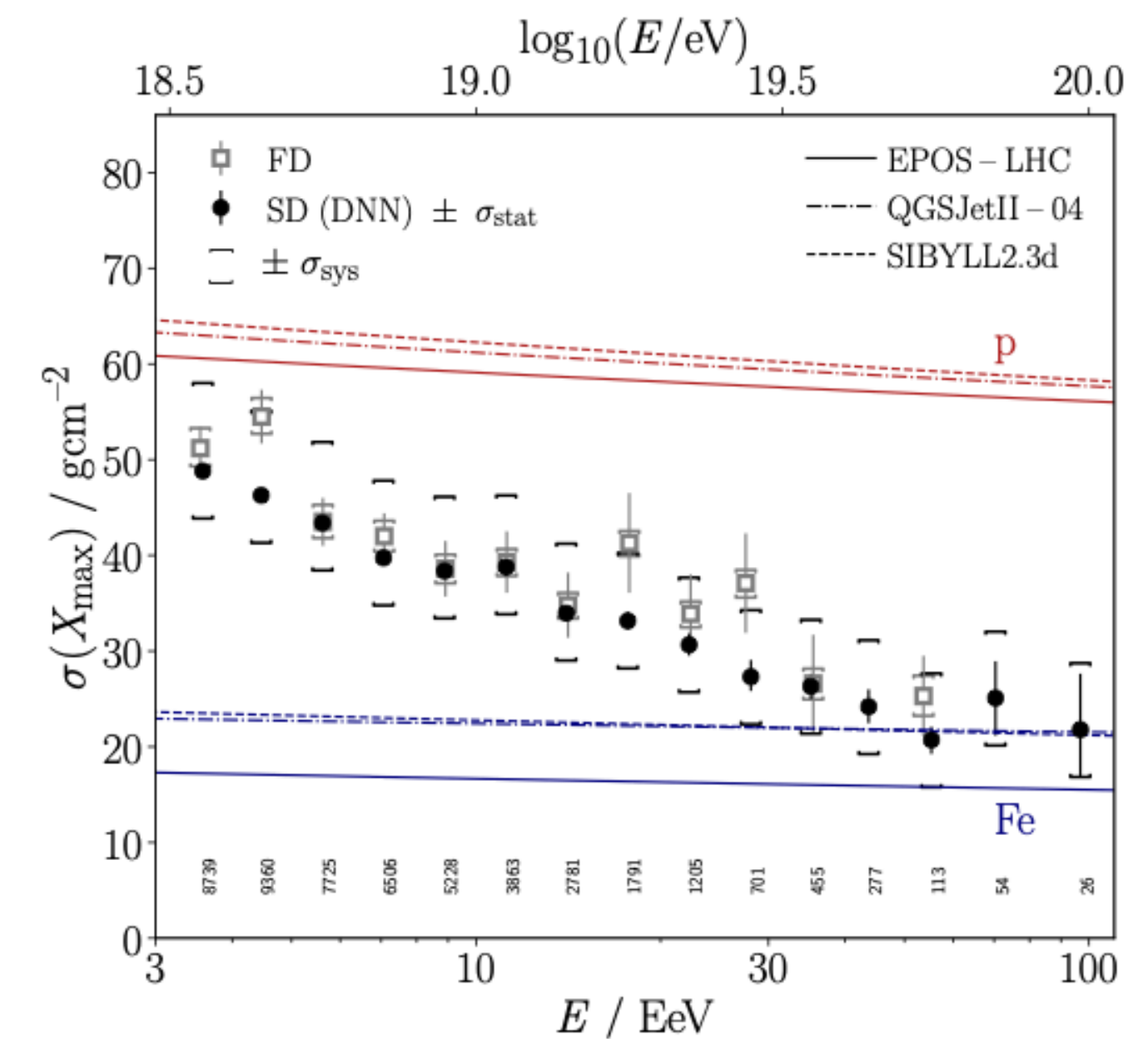
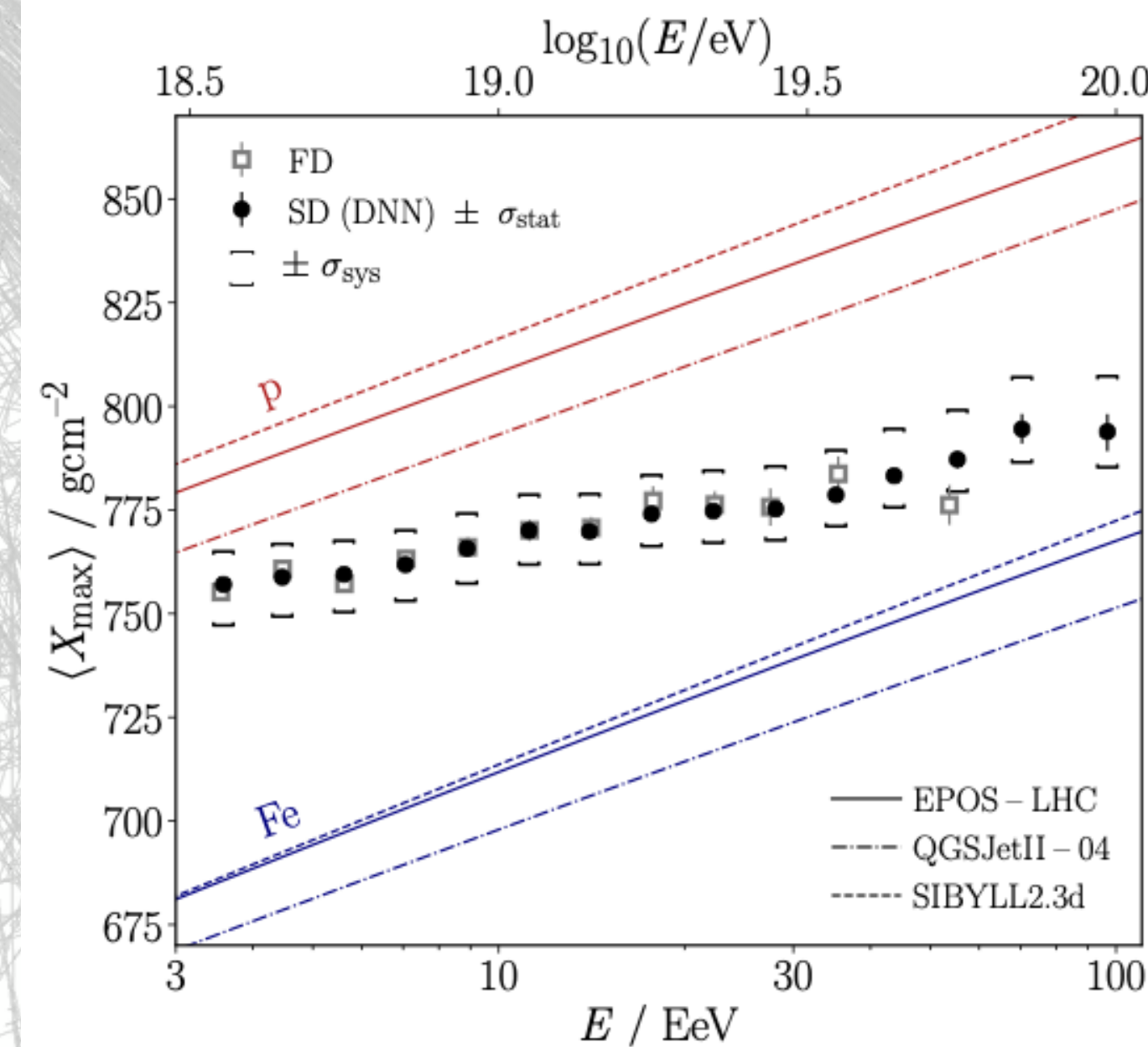


arXiv:2406.06319v3

Investigating the mass composition

- Mass composition of UHECRs connected to the energy evolution of $\langle X_{\max} \rangle$ and $\sigma(X_{\max})$
- Using the SD dataset (48,824 events) compared to the full FD dataset (9,822 events)
- $\langle X_{\max} \rangle$ and $\sigma(X_{\max})$ of SD dataset predictions in agreement with the FD dataset
- Change of $\langle X_{\max} \rangle$ per decade of energy described by the *elongation rate*

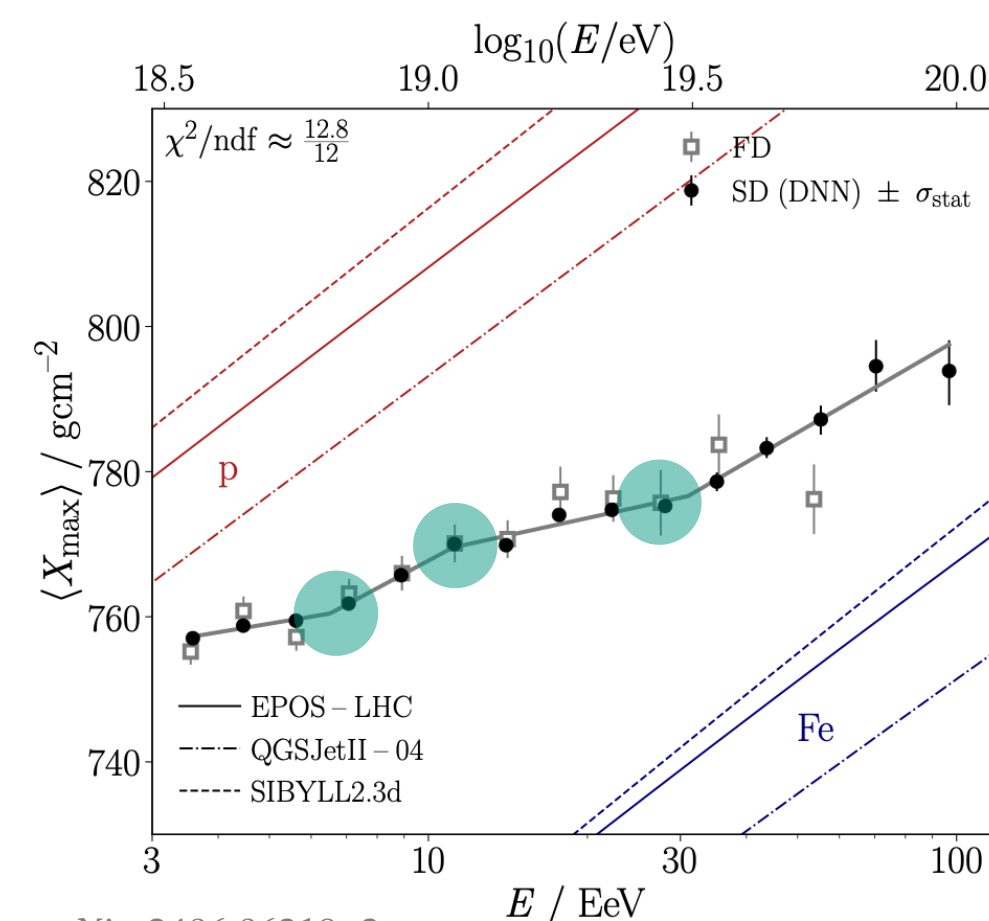
$$D_{10} = \frac{d\langle X_{\max} \rangle}{d \lg E} = \hat{D}_{10} = \left(1 - \frac{d\langle \ln A \rangle}{d \ln E} \right)$$
- Average elongation rate of the SD data agrees with the FD measurements



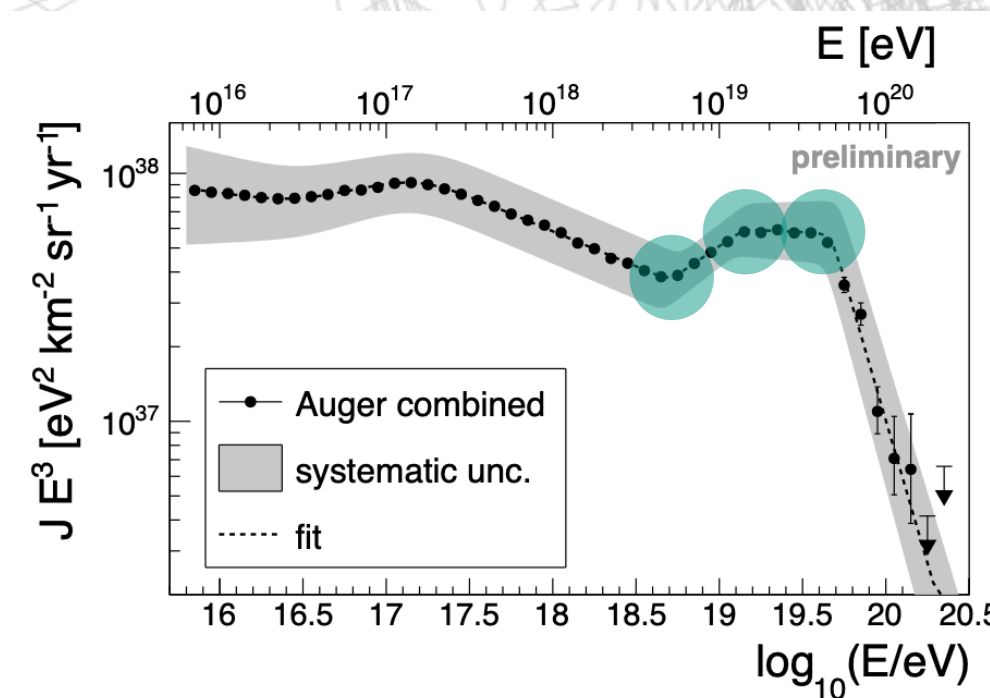
arXiv:2406.06319v3

Interpretation of the results

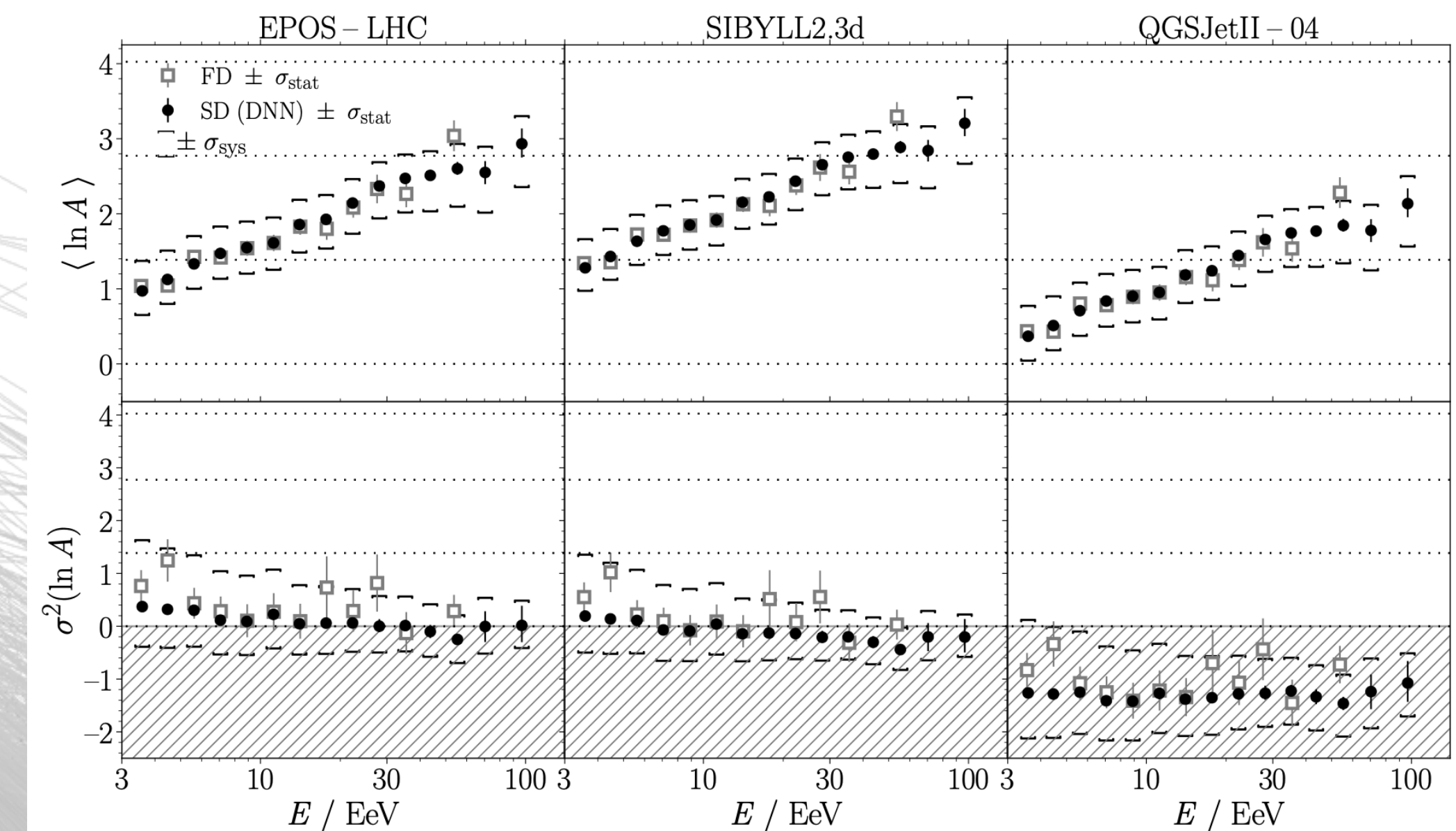
- Simple linear model doesn't describe the elongation rate well
- Model with 3 breaks with 4.4σ significance
 - Non-trivial composition changes
 - Light to heavier composition with energy
 - Breaks close to the features in the energy spectrum



arXiv:2406.06319v3



arXiv:2501.01736v1



arXiv:2406.06319v3

- Interpretation with different had. int. models

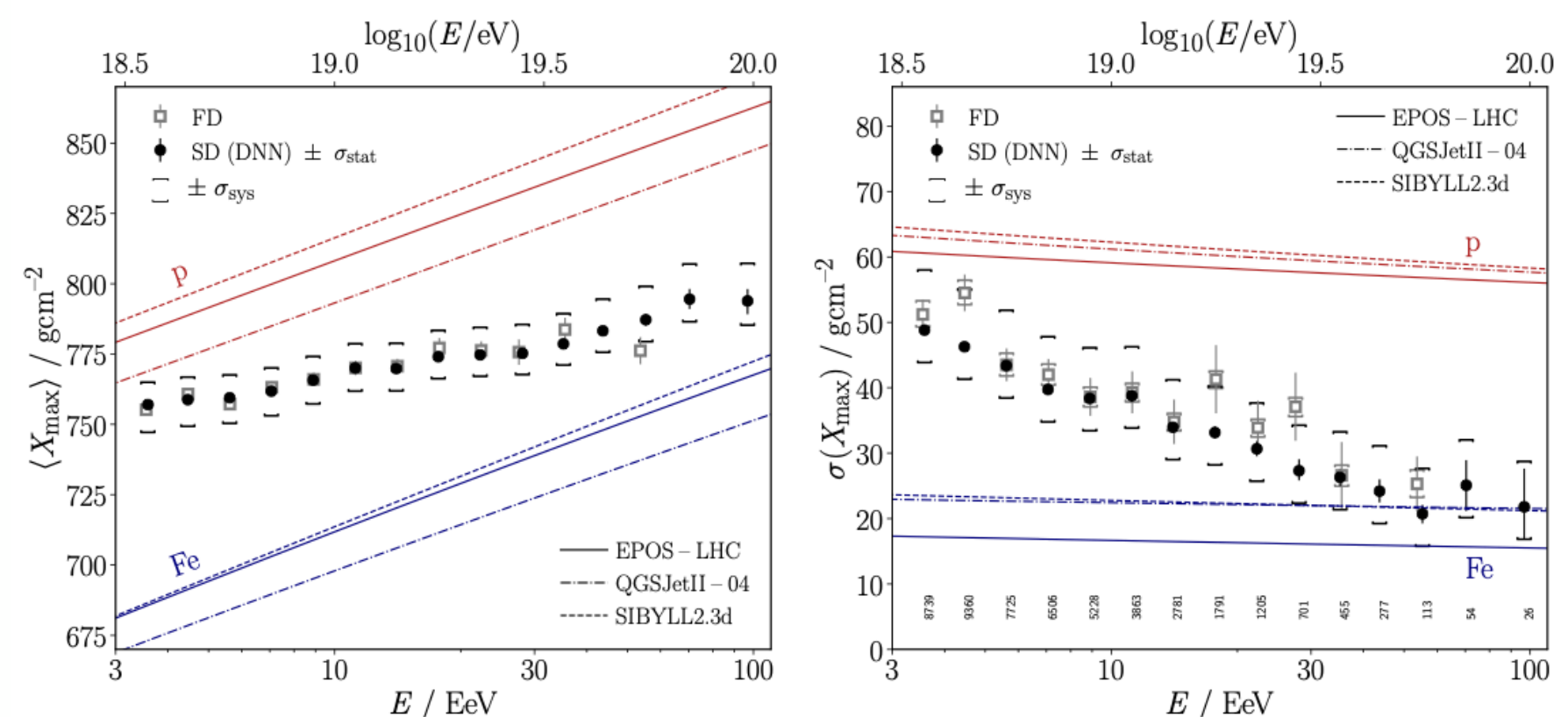
$$\langle \ln A \rangle = \frac{\langle X_{\max} \rangle - \langle X_{\max} \rangle_p}{f_E}, f_E = f_E(E, D, \xi, \delta)$$

JCAP02(2013)026

- Small fluctuations of $\sigma^2(\ln A)$: dominated by a single type of nucleus

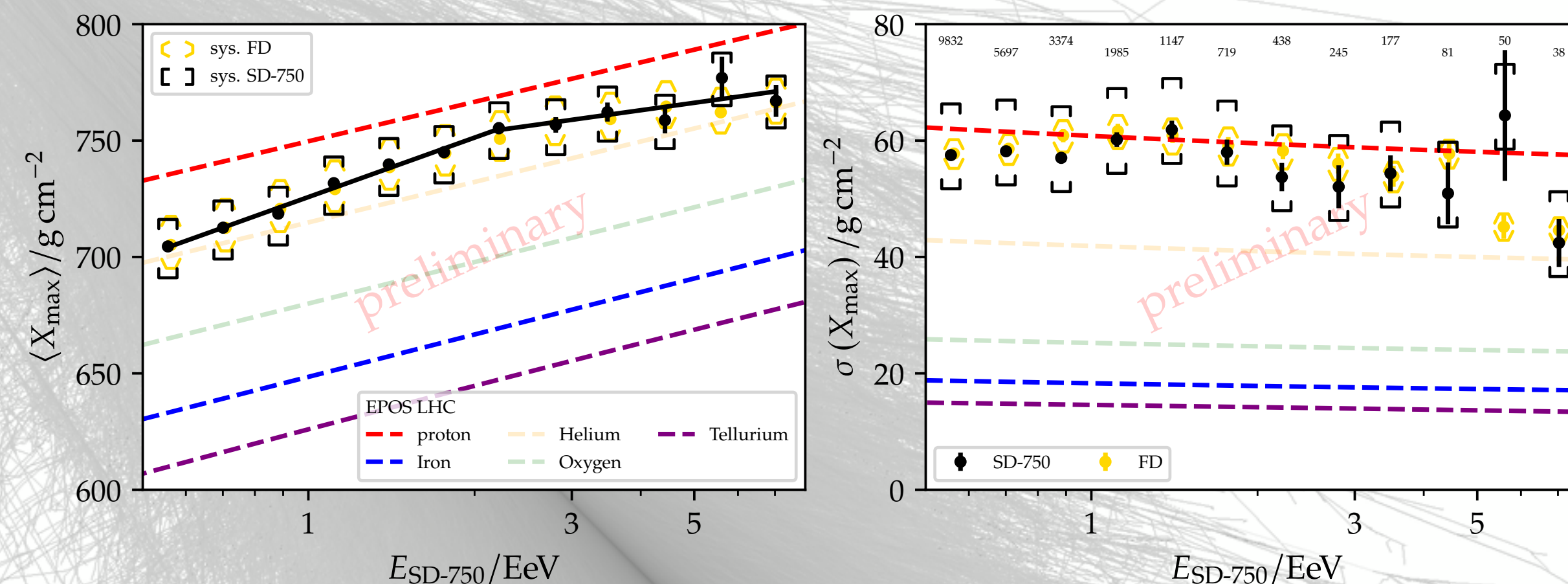
Summary and outlook

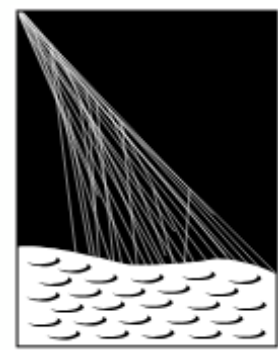
- Using DNN for $X_{\max}$ predictions of events with energies between 3 EeV and 100 EeV
- Excellent agreement of $\langle X_{\max} \rangle$ with fluorescence detector measurements
- Heavier and purer composition with increasing energy confirmed
- Found 3-break model describing the elongation rate



■ Outlook:

- Ongoing AugerPrime upgrade: including WCD upgrade
- Improvements in ML-based analysis strategies
- Extending the study to lower energies (see ICRC 2025)





PIERRE
AUGER
OBSERVATORY

Back-up

Data selection

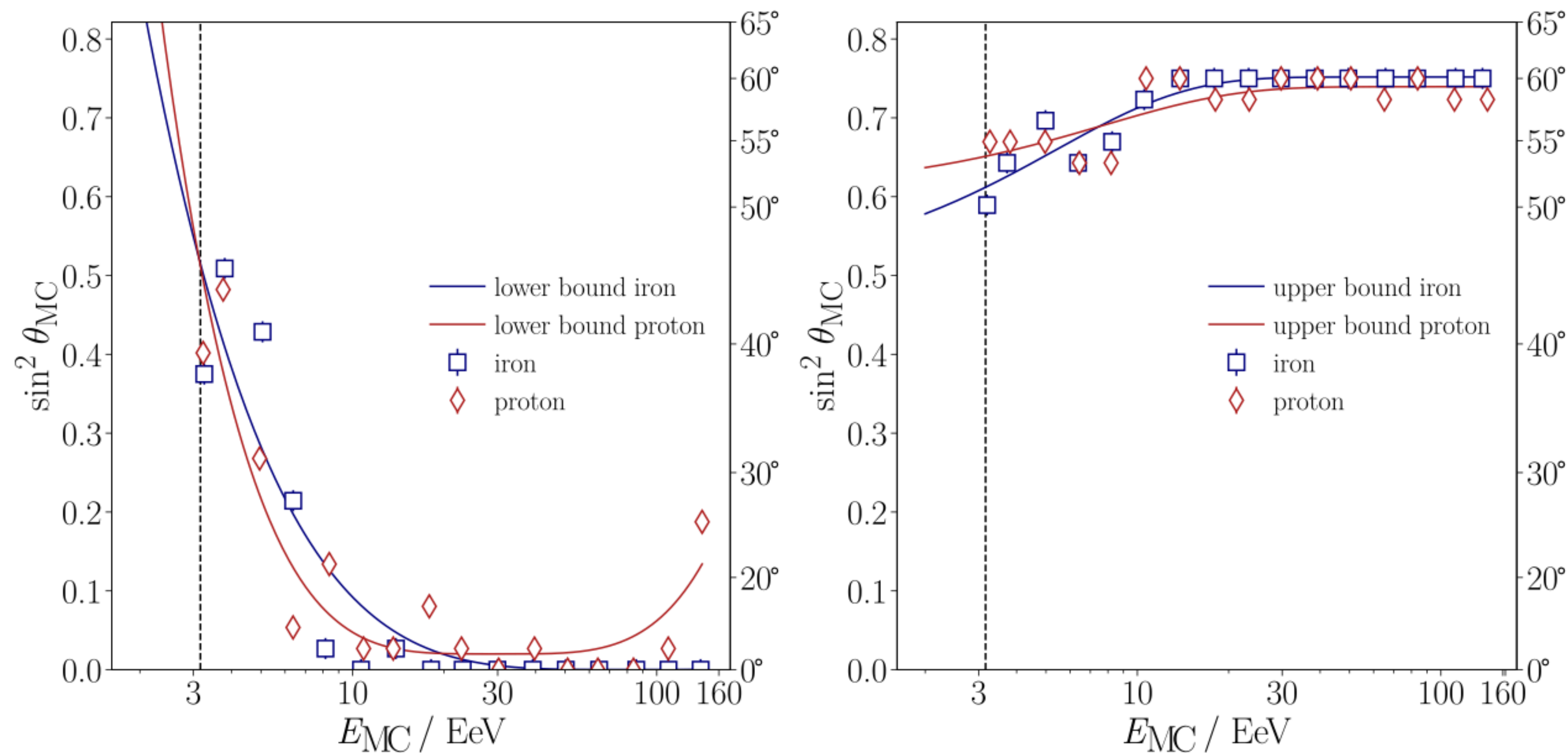
TABLE I: Basic and analysis-specific selections (separated by a line) for the SD data set.

Cut	Events	ϵ (%)
reconstructed vertical event ($\theta < 60^\circ$)	5,994,712	—
is 6T5	4,858,291	81.1
$\log_{10}(E/\text{eV}) > 18.5$	133,167	2.7
hardware status	129,403	97.2
station start slot	128,308	99.2
$2.75 < A/P < 3.45$	126,033	98.2
$\bar{S}_{\text{tot}} > 5$ VEM in surrounding hexagon	125,828	99.8
$350 \text{ m} < \text{core distance} < 1000 \text{ m}$	101,392	80.6
fiducial SD cut	48,824	48.1

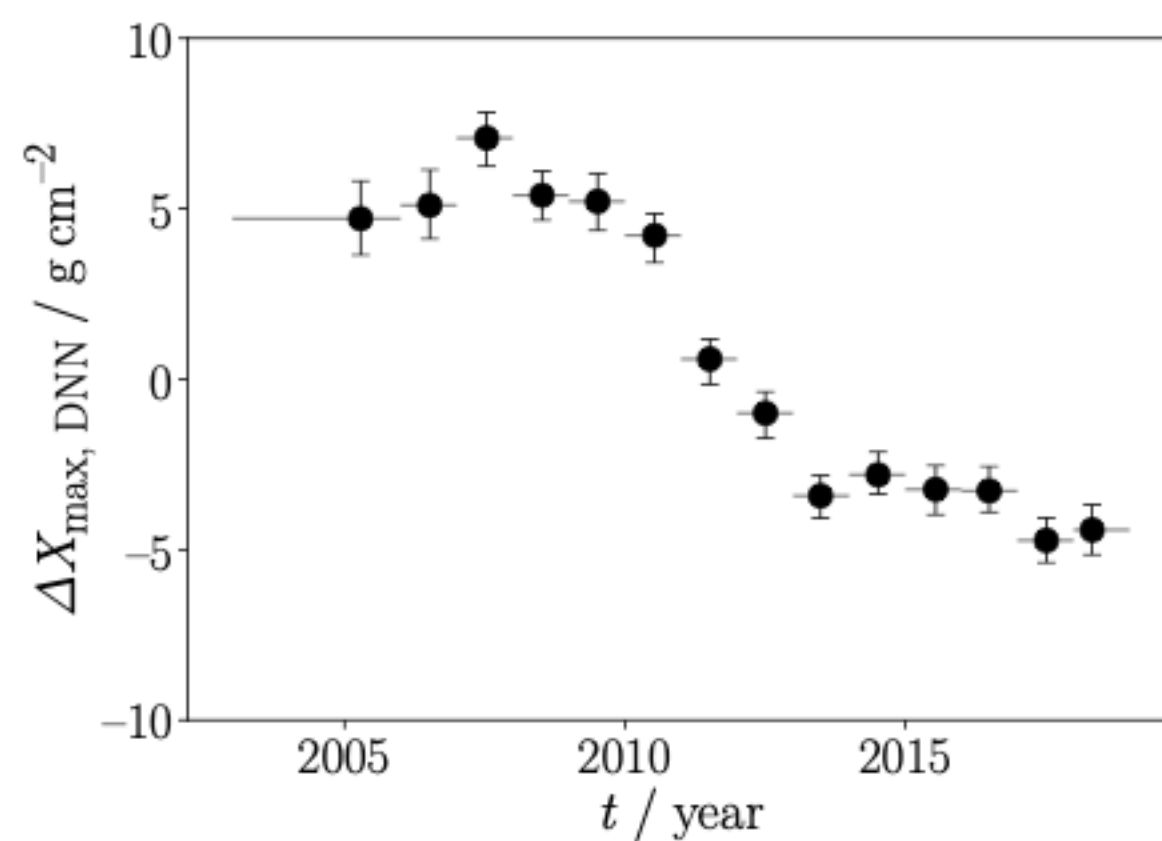
TABLE II: Selections for hybrid data.

Cut	Events	ϵ (%)
number of events	25,076	—
telescope cuts	19,733	78.7
hardware status	16,916	85.7
aerosols/clouds	9,822	58.1
hybrid geometry	9,157	93.2
fiducial FoV cut	3,497	38.2
profile cuts	3,331	95.3
passed SD selection	3,086	92.6
analysis-specific cuts	1,642	53.2

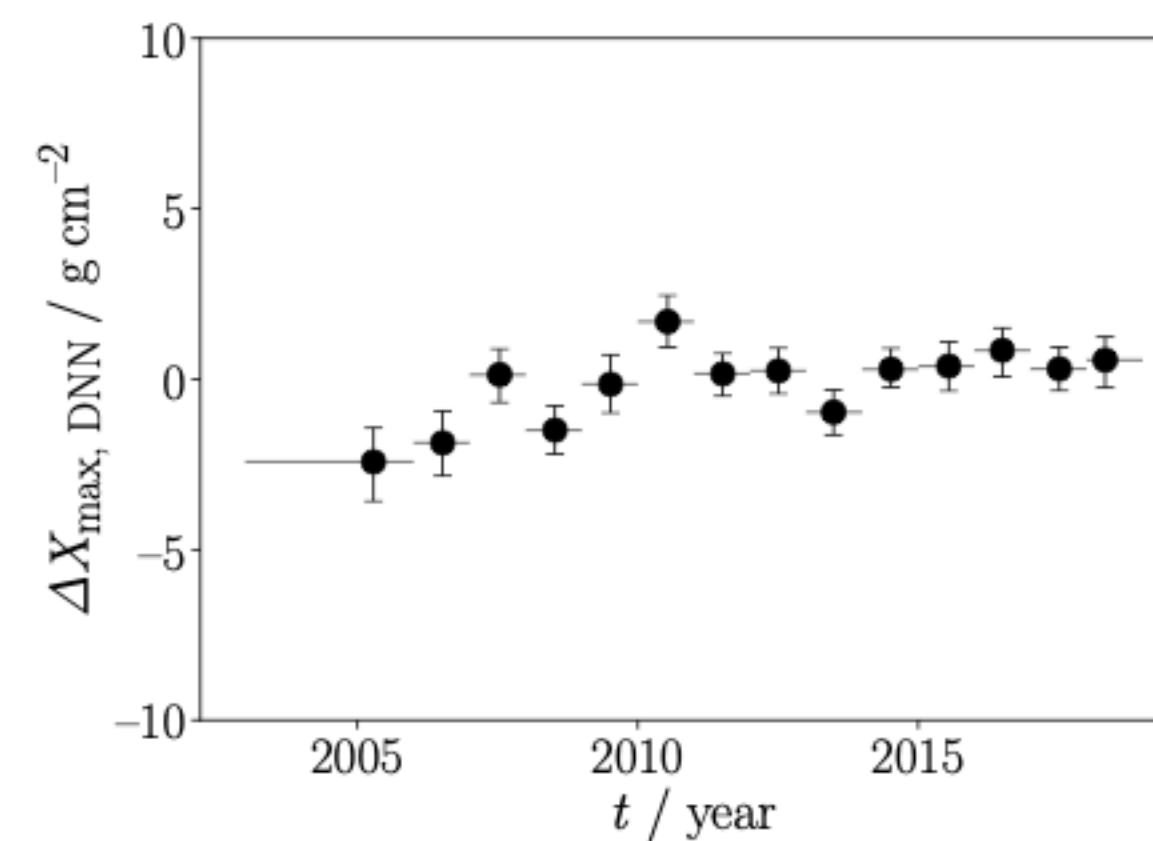
SD fiducial selection



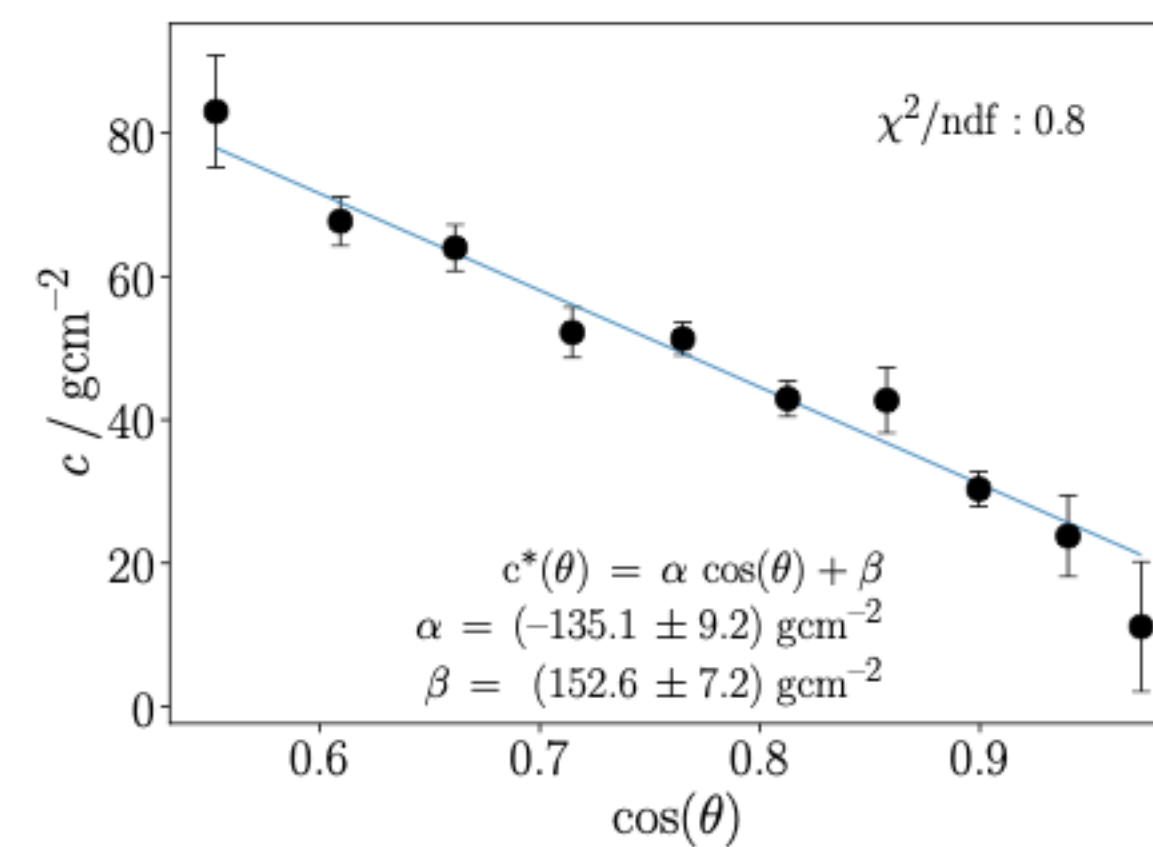
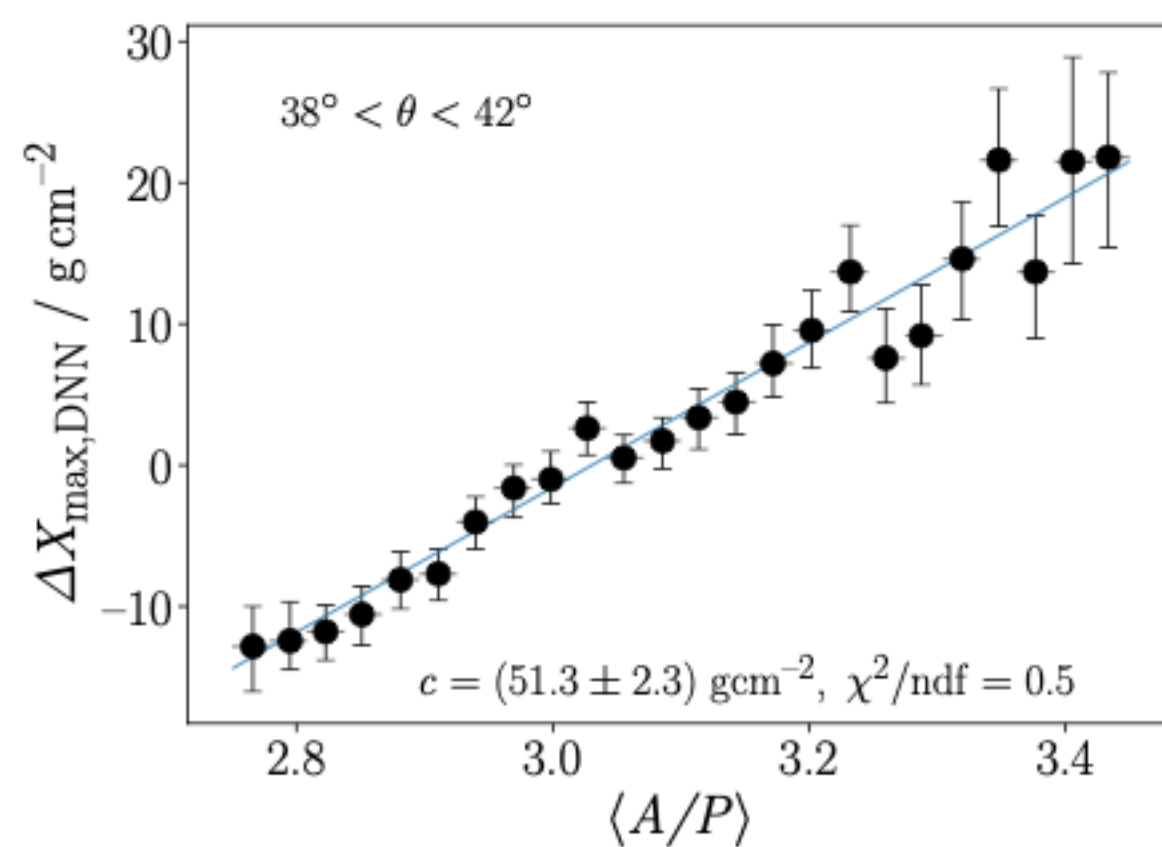
Data-driven corrections



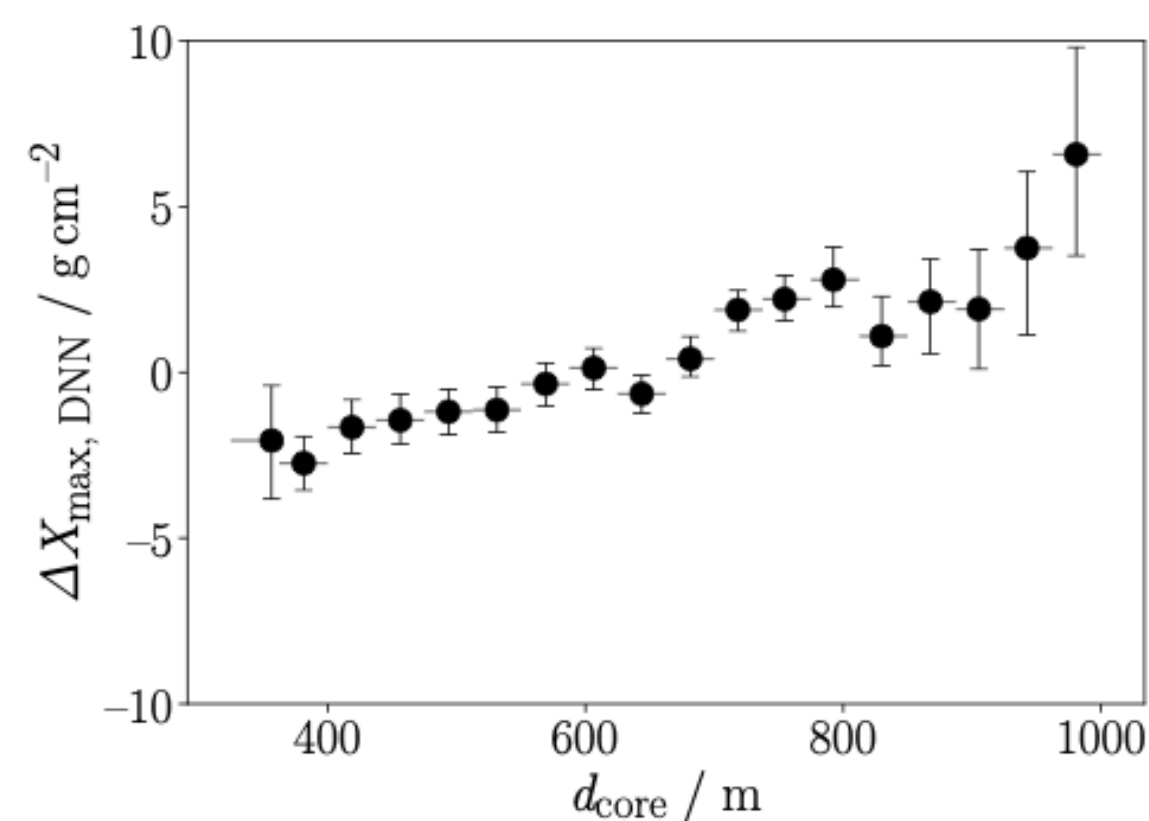
(a)



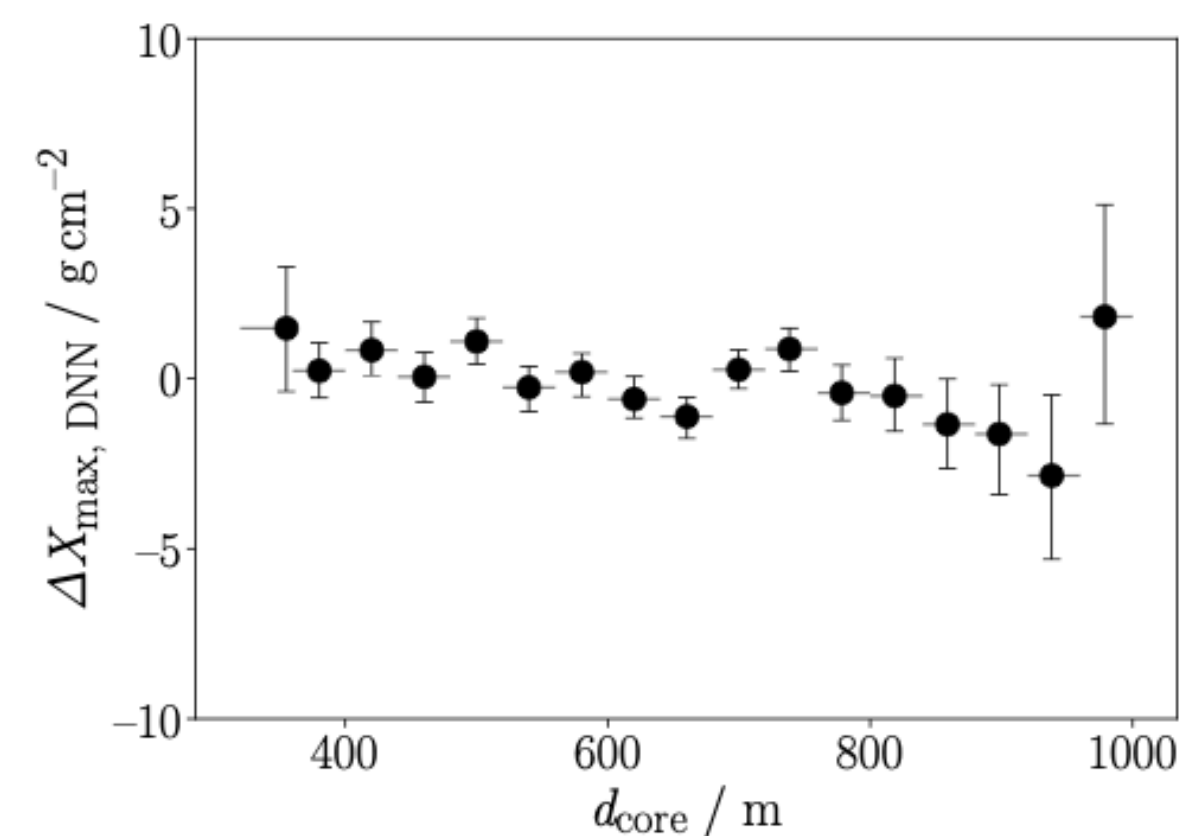
(b)



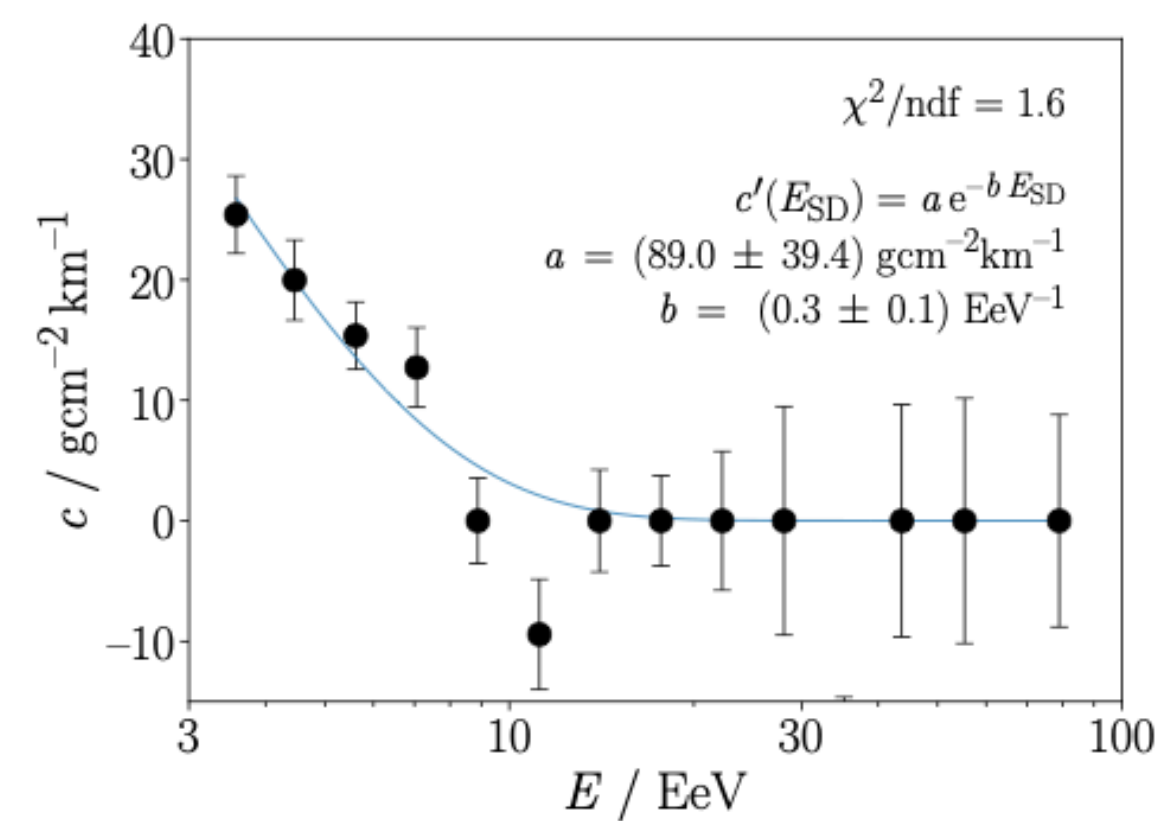
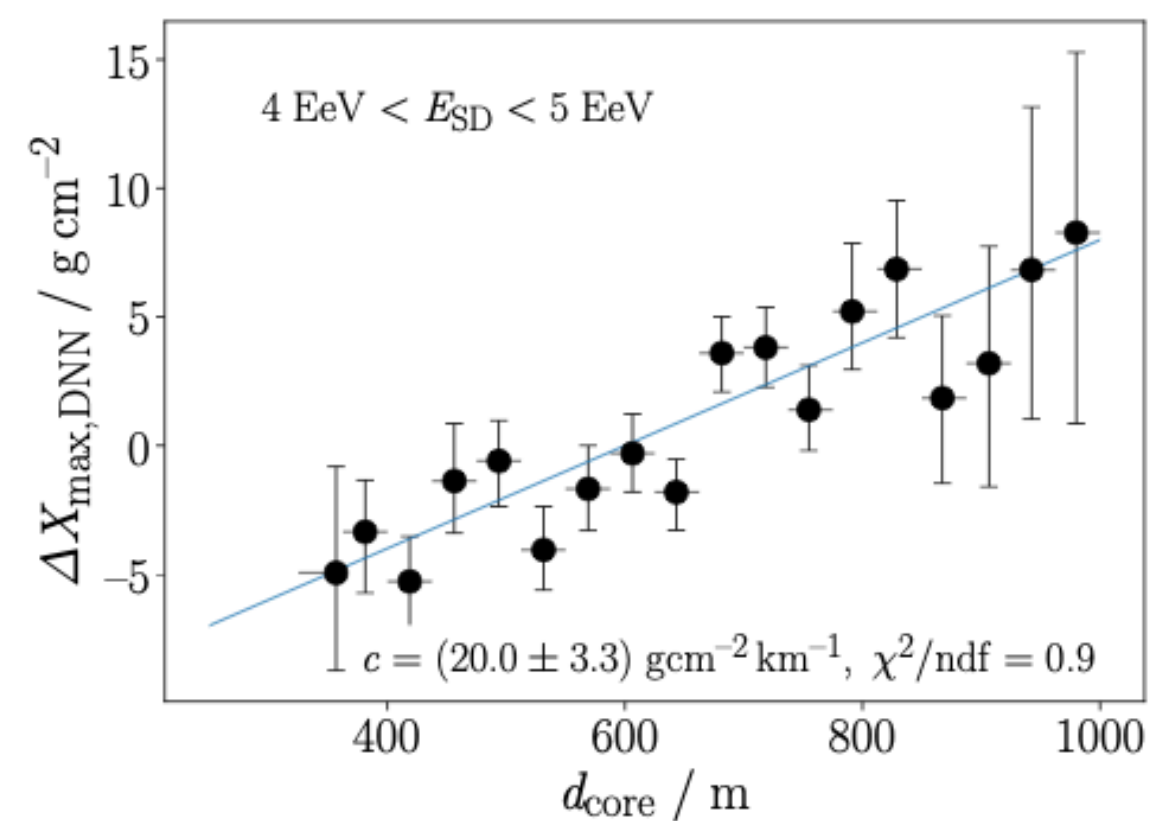
Data-driven corrections

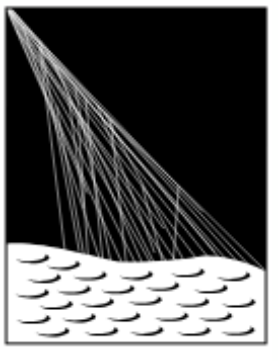


(a)



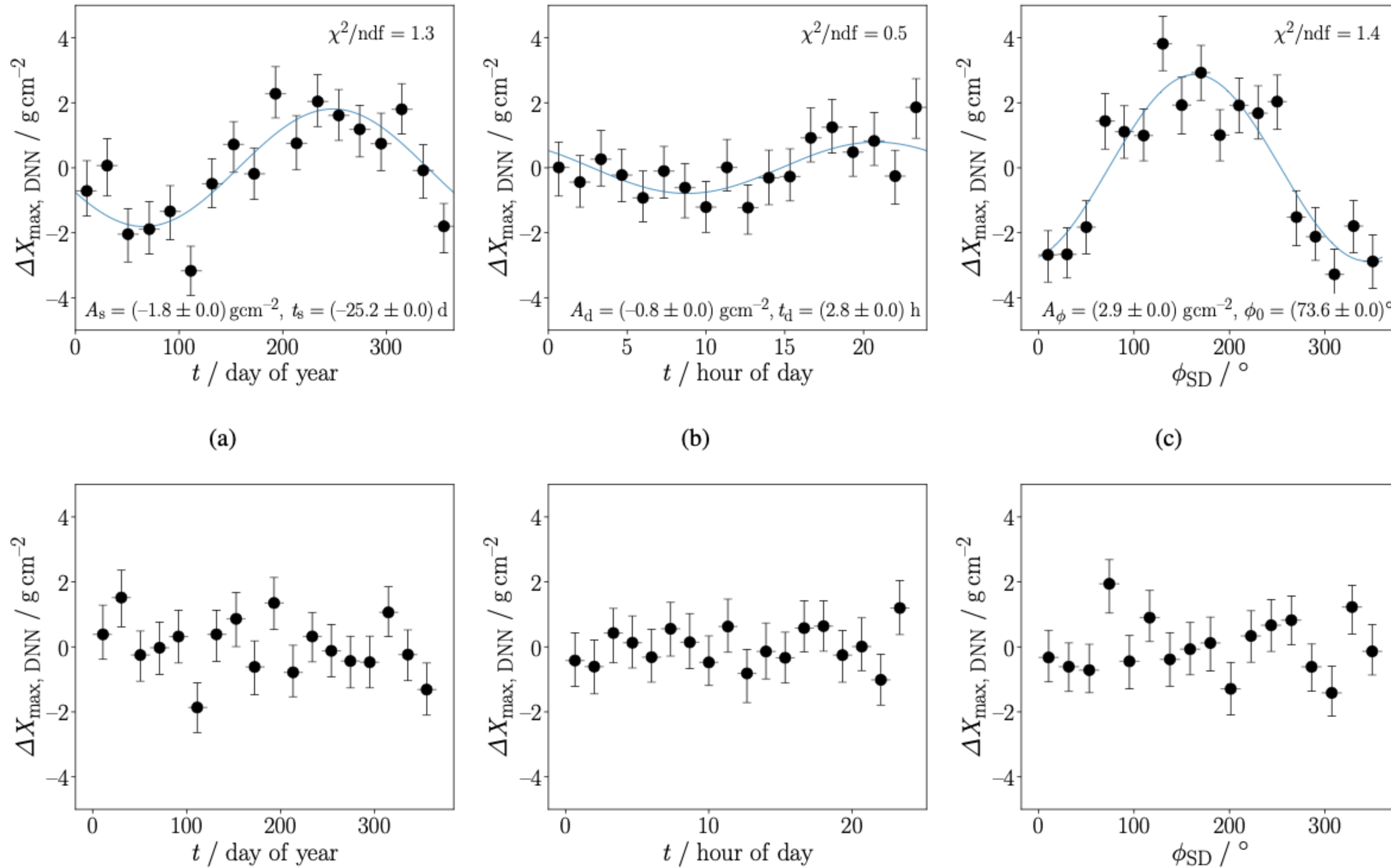
(b)





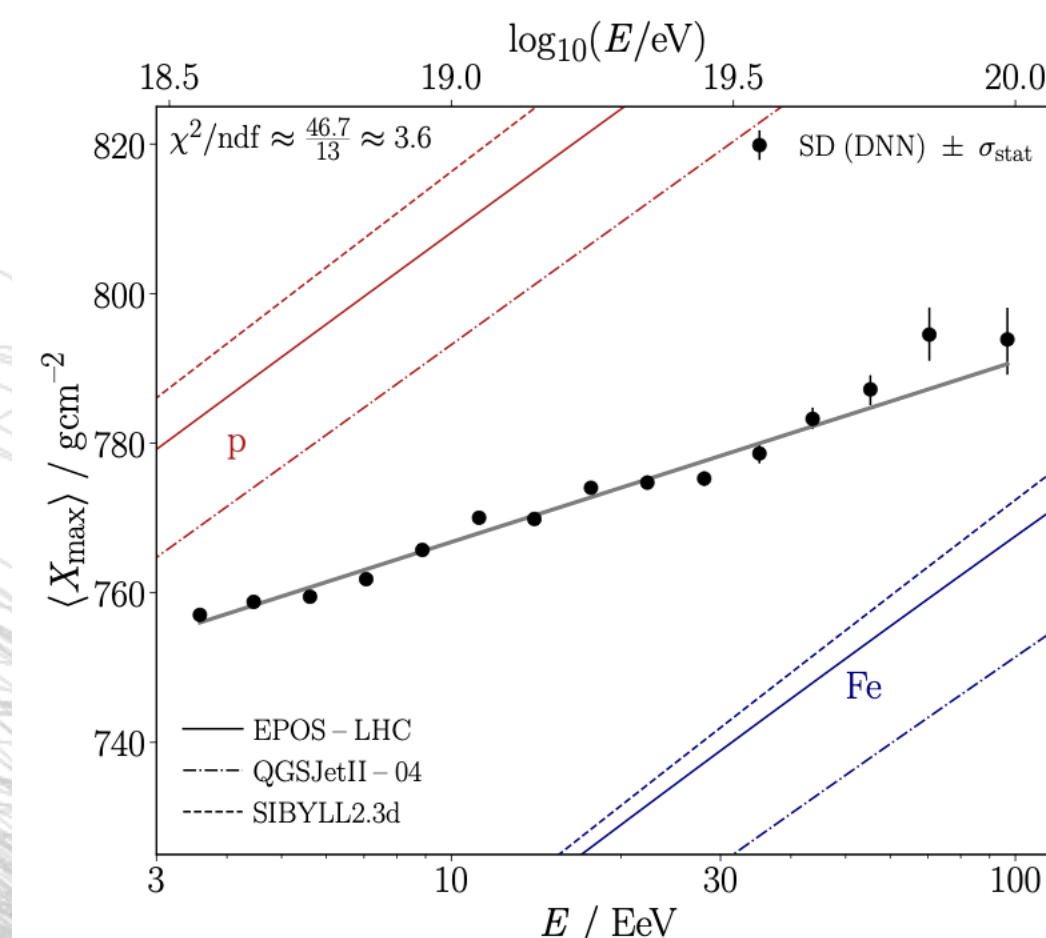
PIERRE
AUGER
OBSERVATORY

Data-driven corrections

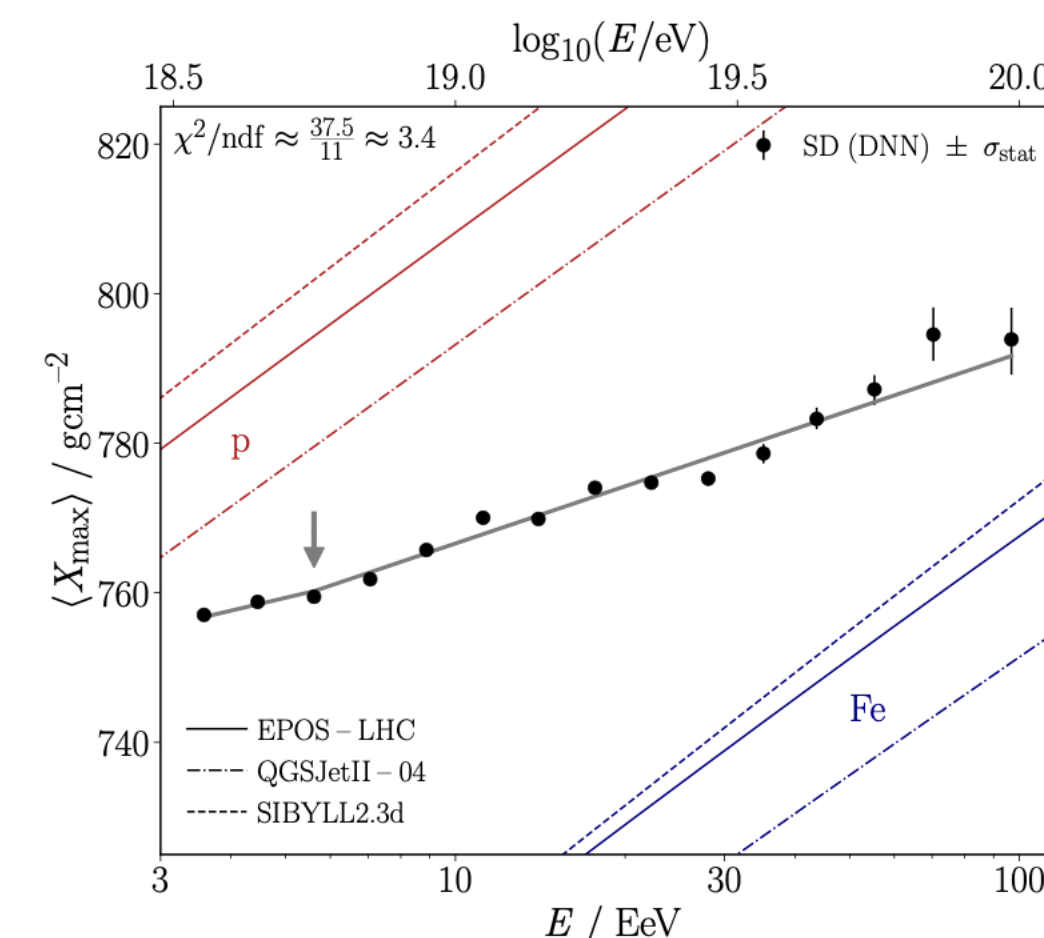


Elongation rate models

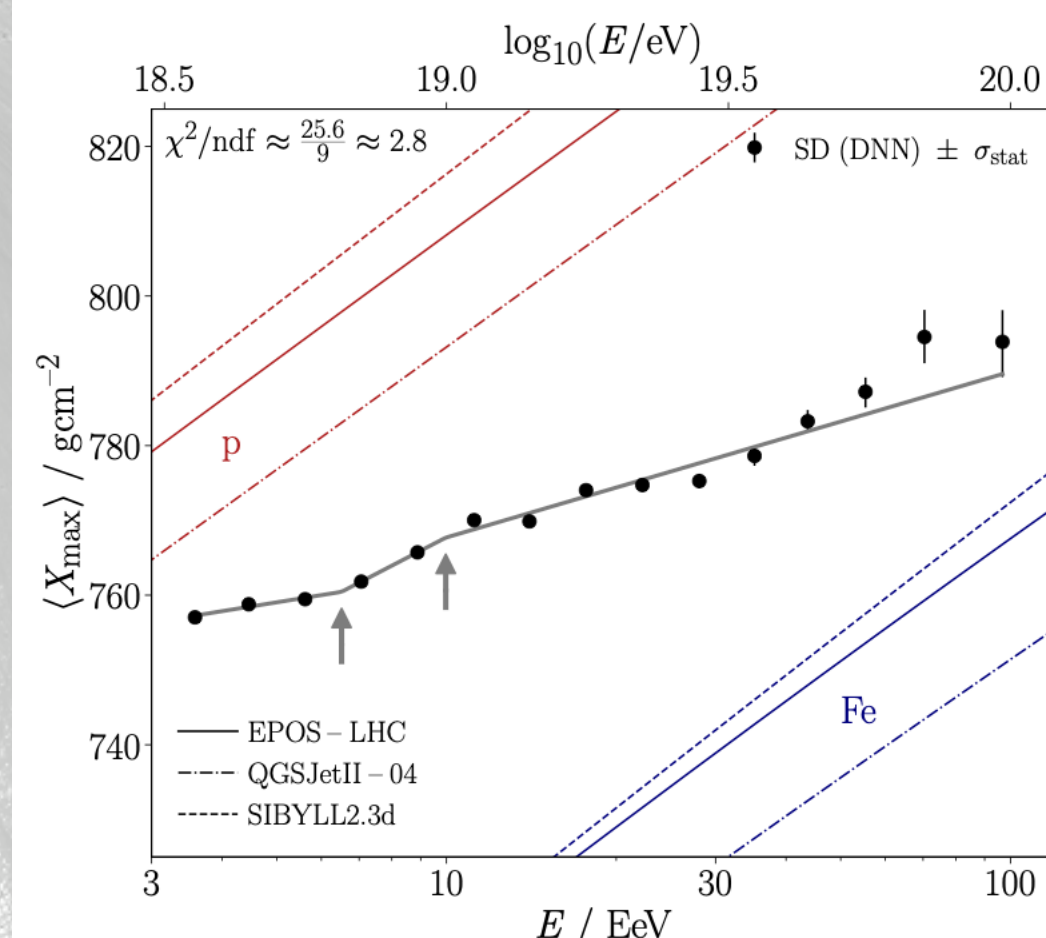
Parameter	Const. elong.	1-break model	2-break model	3-break model	Energy spectrum
	Val $\pm \sigma_{\text{stat}} \pm \sigma_{\text{sys}}$	Val $\pm \sigma_{\text{stat}} \pm \sigma_{\text{sys}}$	Val $\pm \sigma_{\text{stat}} \pm \sigma_{\text{sys}}$	Val $\pm \sigma_{\text{stat}} \pm \sigma_{\text{sys}}$	Val $\pm \sigma_{\text{stat}} \pm \sigma_{\text{sys}}$
$b / \text{g cm}^{-2}$	$743 \pm 5 \pm 13$	$743 \pm 5 \pm 13$	$750.5 \pm 4 \pm 13$	$750.5 \pm 3 \pm 13$	
$D_0 / \text{g cm}^{-2} \text{decade}^{-1}$	$24 \pm 1 \pm 4$	$23 \pm 2 \pm 12$	$12 \pm 6 \pm 5$	$12 \pm 5 \pm 6$	
E_1 / EeV		$35 \pm 12 \pm 16$	$6.5 \pm 0.9 \pm 1$	$6.5 \pm 0.6 \pm 1$	$4.09 \pm 0.1 \pm 0.8$
$D_1 / \text{g cm}^{-2} \text{decade}^{-1}$		$39 \pm 14 \pm 12$	$39 \pm 12 \pm 10$	$39 \pm 5 \pm 14$	
E_2 / EeV			$10 \pm 2 \pm 3$	$11 \pm 2 \pm 1$	$14 \pm 1 \pm 2$
$D_2 / \text{g cm}^{-2} \text{decade}^{-1}$			$22 \pm 3 \pm 8$	$16 \pm 3 \pm 6$	
E_3 / EeV				$31 \pm 5 \pm 3$	$47 \pm 3 \pm 6$
$D_3 / \text{g cm}^{-2} \text{decade}^{-1}$				$42 \pm 9 \pm 12$	



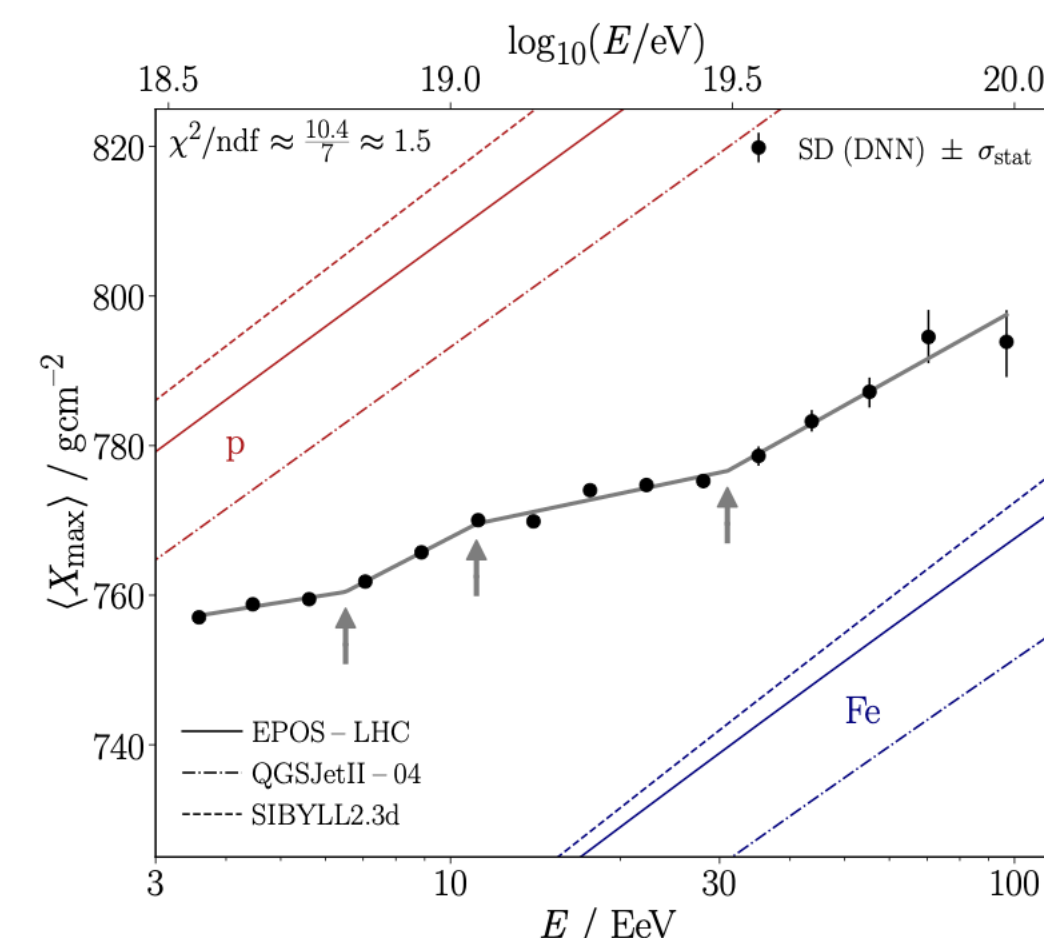
(a)



(b)



(c)



(d)