

Low-energy parameterizations of X_{\max} statistics

Carmelo Evoli, on behalf of the Auger-L'Aquila group

Gran Sasso Science Institute, L'Aquila (Italy)
INFN/Laboratori Nazionali del Gran Sasso (LNGS), Assergi (Italy)

April 4, 2024



Based on GAP-2020-058

Motivations

- ▷ Our goal is to broaden the scope of the combined fit in order to cover lower energies down to $E \gtrsim 10^{15}$ eV → Galactic-ExtraGalactic transition
- ▷ Existing parametrizations for calculating cosmic ray composition using X_{\max} statistics are constrained to energies above $E \gtrsim 10^{17}$ eV (see GAP2020_058)
- ▷ It's crucial to assess whether these current models remain accurate for energies down to 10^{15} eV. Should discrepancies arise, we must consider refining our parametrization models to ensure reliable composition in the knee energy spectrum.

Conex simulations

From GAP2020_058:

- ▷ Conex version: version 4.37 for EPOS-LHC and QGSJet II-04, version 7.30 for Sibyll 2.3d
- ▷ Hadronic models used Sibyll 2.3d, EPOS-LHC and QGSJet II-04
- ▷ The log energy range $17 \rightarrow 20$ in 13 fixed lg E bins with $\Delta \log E = 0.25$
- ▷ Number of showers 5.4k - 7.7k / bin
- ▷ The Xmax used to build the distributions is taken from the **XmxEdX** branch of the CONEX file
- ▷ Primary nuclei: H, He, N, Si, Ca and Fe

New simulations at CNAF:

- ▷ Conex version: **version 7.60**
- ▷ Hadronic models used Sibyll 2.3d, EPOS-LHC, QGSJet II-04, and **DPM-JET III**
- ▷ The log energy range **$15 \rightarrow 20.5$** in 23 fixed lg E bins with $\Delta \log E = 0.25$
- ▷ Number of showers **$10k / \text{bin} \rightarrow 100k / \text{bin}$** ^{goal}
- ▷ The Xmax used to build the distributions is taken from the **XmxEdX** branch of the CONEX file
- ▷ Primary nuclei: H, He, N, Si, and Fe

Definitions

► Mean:

$$\langle x \rangle = \frac{1}{N} \sum_{i=0}^N x_i \quad (1)$$

► Variance:

$$\sigma_x^2 = \frac{1}{N} \sum_{i=0}^N |x_i - \langle x \rangle|^2 \quad (2)$$

► Standard Deviation:

$$\sigma_x = \sqrt{\sigma_x^2} \quad (3)$$

► Error of the Mean:

$$\epsilon = \frac{\sigma_x}{\sqrt{N}} \quad (4)$$

► Error of the Standard Deviation:

$$\rho = \frac{\sigma_x}{\sqrt{2N}} \quad (5)$$

X_{\max} parametrizations

- ▷ We model X_{\max} as a function of $x \equiv \log(E/E_0)$ and $y \equiv \ln A$
- ▷ GAP parametrization (4 free parameters):

$$\begin{aligned} p'_0(y) &= p_0 + \alpha y \\ p'_1(y) &= p_1 + \beta y \\ f(x, y) &= p'_0(y) + p'_1(y)x \end{aligned}$$

which can be re-written as

$$f(x, y) = (p_0 + p_1 x) + (\alpha + \beta x)y \equiv f_p(x) + f_E(x)y$$

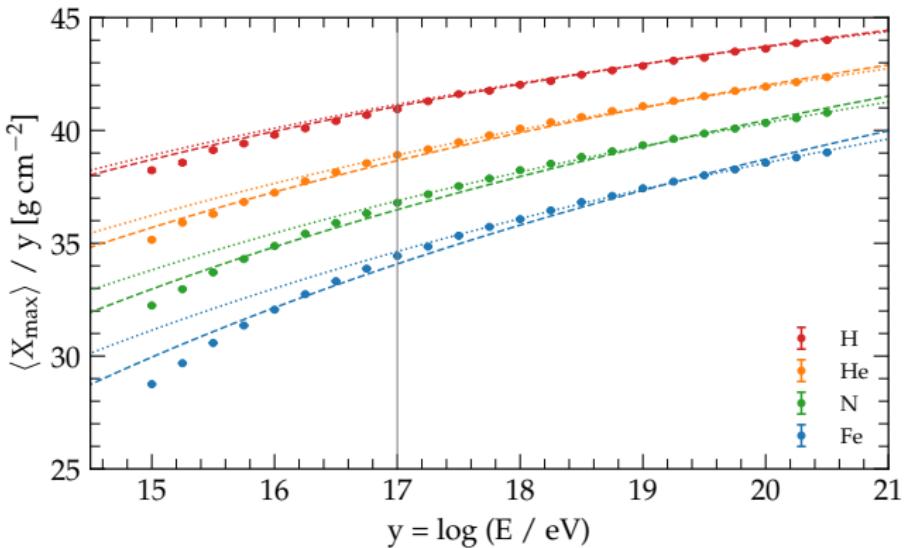
- ▷ EXT parametrization (6 free parameters):

$$\begin{aligned} p'_0(y) &= p_0 + \alpha y \\ p'_1(y) &= p_1 + \beta y \\ p'_2(y) &= p_2 + \gamma y \\ f(x, y) &= p'_0(y) + p'_1(y)x + p'_2(y)x^2 \end{aligned}$$

which can be re-written as

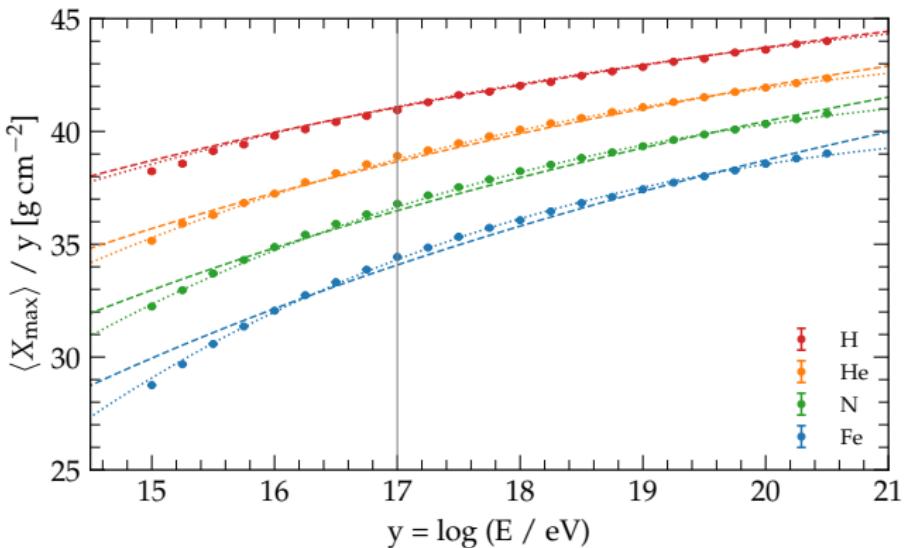
$$f(x, y) = (p_0 + p_1 x + p_2 x^2) + (\alpha + \beta x + \gamma x^2)y \equiv f_p(x) + f_E(x)y$$

X_{\max} parametrizations



- ▷ Dots: CONEX simulations using Sibyll-23d. Error bars show the **std error of the mean** over N simulations
- ▷ In the y-axis we show X_{\max}/y to emphasize the deviation from the linear trend
- ▷ Dashed lines: best fit of GAP (linear in y) parametrization assuming $E_{\min} = 10^{15}$ eV
- ▷ Dotted lines: best fit of GAP (linear in y) parametrization assuming $E_{\min} = 10^{17}$ eV

X_{\max} parametrizations



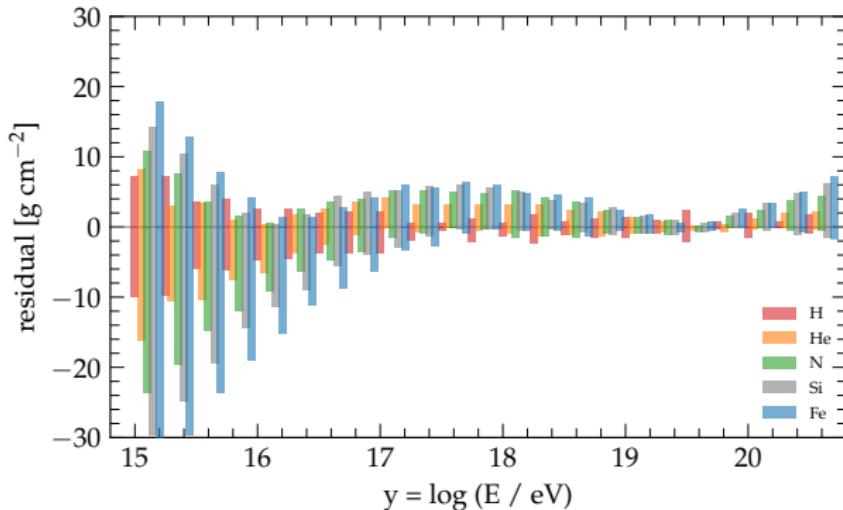
- ▷ Dots: CONEX simulations using Sibyll-23d. Error bars show the **std error of the mean** over N simulations
- ▷ In the y-axis we show X_{\max}/y to emphasize the deviation from the linear trend
- ▷ Dashed lines: best fit of GAP (linear in y) parametrization assuming $E_{\min} = 10^{15}$ eV
- ▷ Dotted lines: best fit of EXT (2nd order in y) parametrization assuming $E_{\min} = 10^{15}$ eV

X_{\max} parametrizations

Fit parameters based on Sibyll-2.3D.

	D_0	D_1	D_2	α	β	γ
GAP-full	815.69 ± 0.11	58.76 ± 0.06	-	-26.37 ± 0.03	1.57 ± 0.02	-
GAP-hi	815.83 ± 0.11	58.09 ± 0.11	-	-26.19 ± 0.03	0.67 ± 0.03	-
EXT	816.21 ± 0.12	58.01 ± 0.11	-0.37 ± 0.04	-25.67 ± 0.04	0.56 ± 0.03	-0.46 ± 0.01

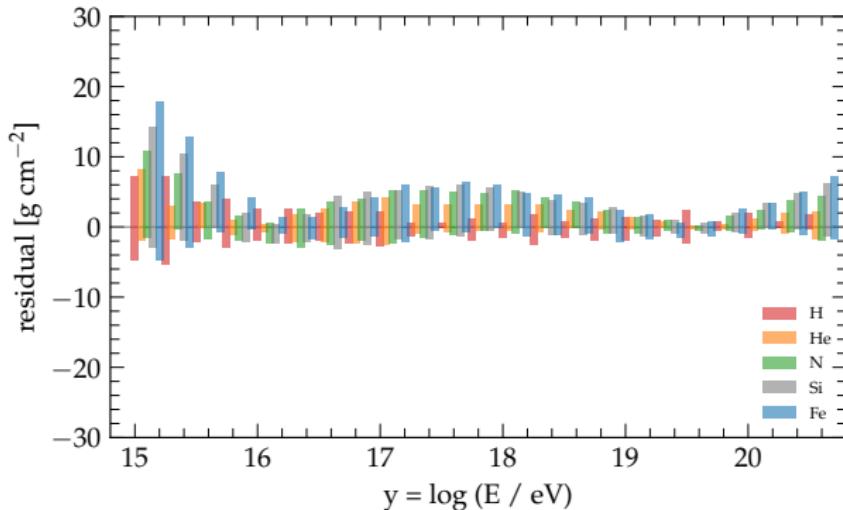
X_{\max} residuals



- Positive plane: residuals of the GAP parametrization assuming $E_{\min} = 10^{15}$ eV
- Negative plane: residuals of the GAP parametrization assuming $E_{\min} = 10^{17}$ eV
- Mean residuals in g/cm². Si was not included in the fit:

	H	He	N	Si	Fe
GAP-full	2.15	2.45	3.64	4.39	4.94
GAP-hi	3.07	2.95	4.56	5.71	7.11

X_{\max} residuals



- ▷ Positive plane: residuals of the GAP parametrization assuming $E_{\min} = 10^{15}$ eV
- ▷ Negative plane: residuals of the EXT parametrization assuming $E_{\min} = 10^{15}$ eV
- ▷ Mean residuals in g/cm². Si was not included in the fit:

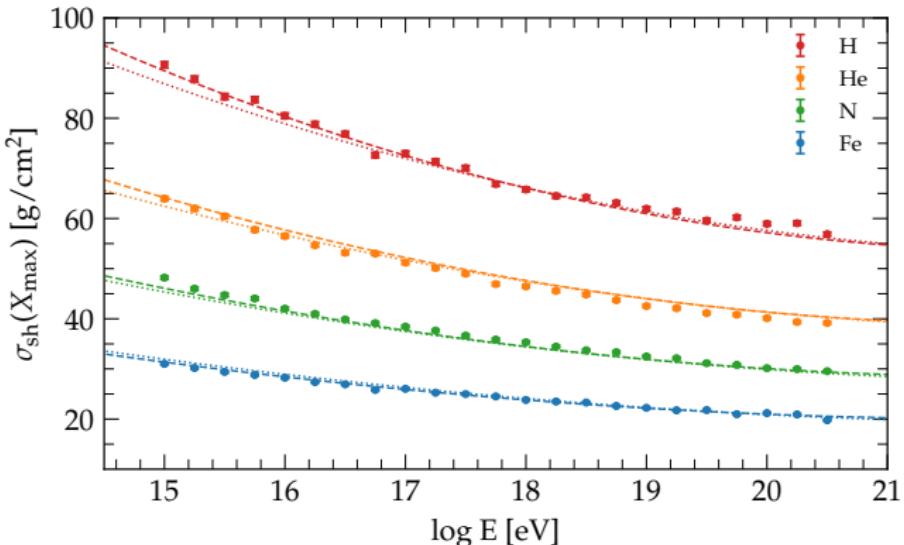
	H	He	N	Si	Fe
GAP	2.15	2.45	3.64	4.39	4.94
EXT	1.96	0.98	1.21	1.35	1.37

$\sigma(X_{\max})$ parametrizations

- ▷ We model $\sigma(X_{\max})$ as a function of $x \equiv \log(E/E_0)$ and $y \equiv \ln A$
- ▷ GAP parametrization (6 free parameters):

$$\begin{aligned} a'_0(x) &= a_0 + a_1 x \\ p(x) &= p_0 + p_1 x + p_2 x^2 \\ f(x, y) &= p(x) [1 + a'_0(x)y + b_0 y^2] \end{aligned}$$

$\sigma(X_{\max})$ parametrizations



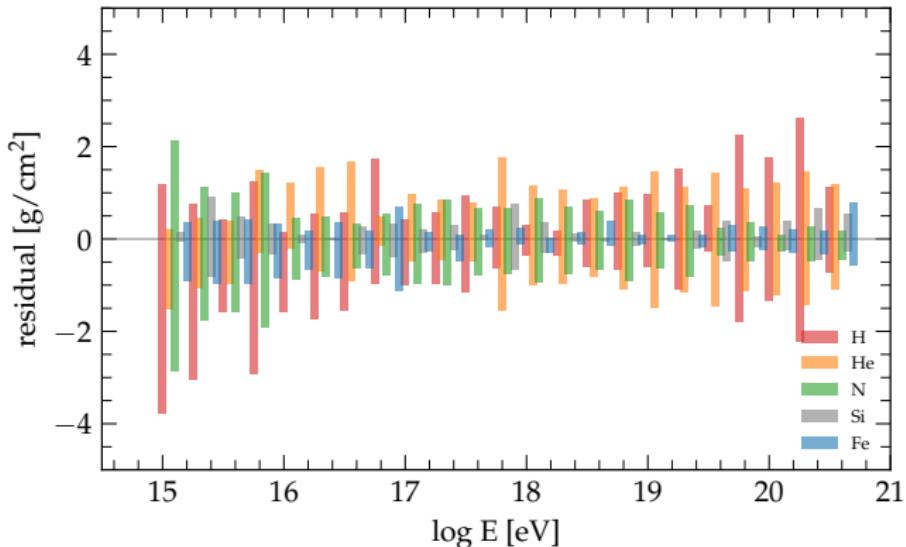
- ▷ Dots: CONEX simulations using Sibyll-23d. Error bars show the **error of the std deviation** over N simulations
- ▷ Dashed lines: best fit of GAP parametrization assuming $E_{\min} = 10^{15}$ eV
- ▷ Dotted lines: best fit of GAP parametrization assuming $E_{\min} = 10^{17}$ eV

$\sigma(X_{\max})$ parametrizations

	p ₀	p ₁	p ₂	a ₀	a ₁	b
GAP-full	61.0 ± 0.1	-4.5 ± 0.1	0.67 ± 0.02	-0.223 ± 0.001	0.0008 ± 0.0001	0.0161 ± 0.0002
GAP-hi	61.3 ± 0.1	-4.3 ± 0.1	0.53 ± 0.07	-0.228 ± 0.001	-0.0002 ± 0.0002	0.0173 ± 0.0003

Fit parameters based on Sibyll-2.3D.

$\sigma(X_{\max})$ residuals



- Positive plane: residuals of the GAP parametrization assuming $E_{\min} = 10^{15}$ eV
- Negative plane: residuals of the GAP parametrization assuming $E_{\min} = 10^{17}$ eV
- Mean residuals in g/cm^2 . Si was not included in the fit:

	H	He	N	Si	Fe
GAP full	0.97	1.08	0.68	0.29	0.26
GAP hi	1.34	0.93	0.92	0.25	0.45

Understanding Anomalous Shower Profiles

- ▷ Most air showers caused by high-energy cosmic rays follow a predictable pattern, showing a single, clearly defined maximum in their longitudinal development.
- ▷ A small subset exhibits deviations from this norm, with some showing significantly different profiles, including cases with two distinct maxima.
- ▷ We use simulations to analyze the frequency of these anomalous profiles as a function of the primary energy.
- ▷ We limit our investigation to primary protons (where anomalies are most pronounced) and apply a single interaction model for the moment.

Modeling the Longitudinal Shower Profile

- ▷ Gaisser–Hillas¹: A foundational 3-parameter model defines the normalized ($N_{\max} = 1$) longitudinal profile of a shower. The function is given by:

$$y(x) = \left(1 + R \frac{x'}{L}\right)^{R^{-2}} \exp\left(-\frac{x'}{RL}\right) \quad (6)$$

where

$$x' = x - x_{\max} \quad (7)$$

- ▷ We evaluate the accuracy of our model using a cost function

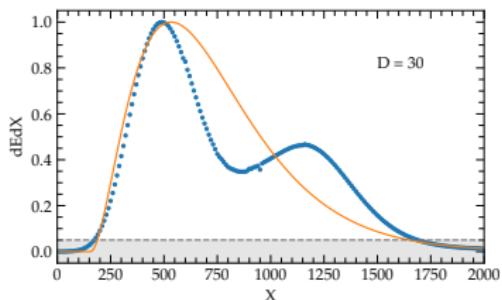
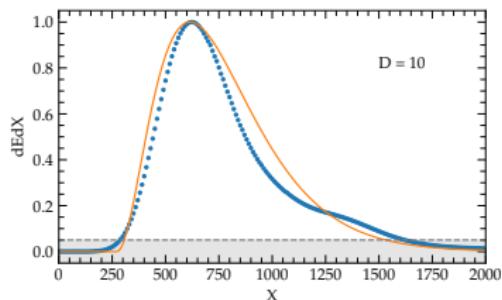
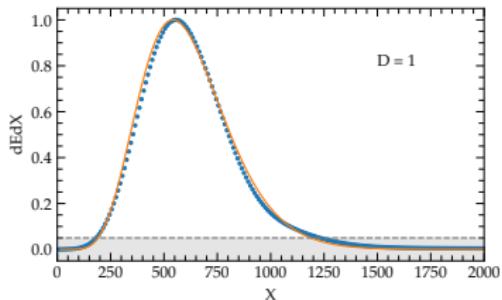
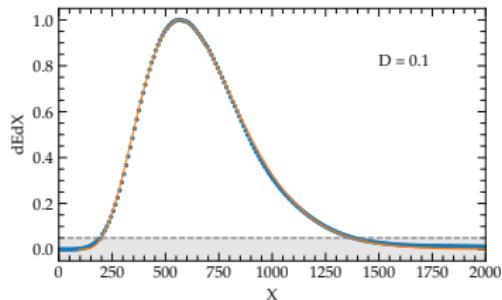
$$D = \sum_{y_i > 5\%} \frac{[y(x_i) - y_i]^2}{y_i^2} \quad (8)$$

where y_i represents the normalized profile simulated by CONEX.

- ▷ To find the best fit parameters, (x_{\max}, R, L) , we perform a minimization of D starting from approximately 100 randomly selected initial points in the parameter space.

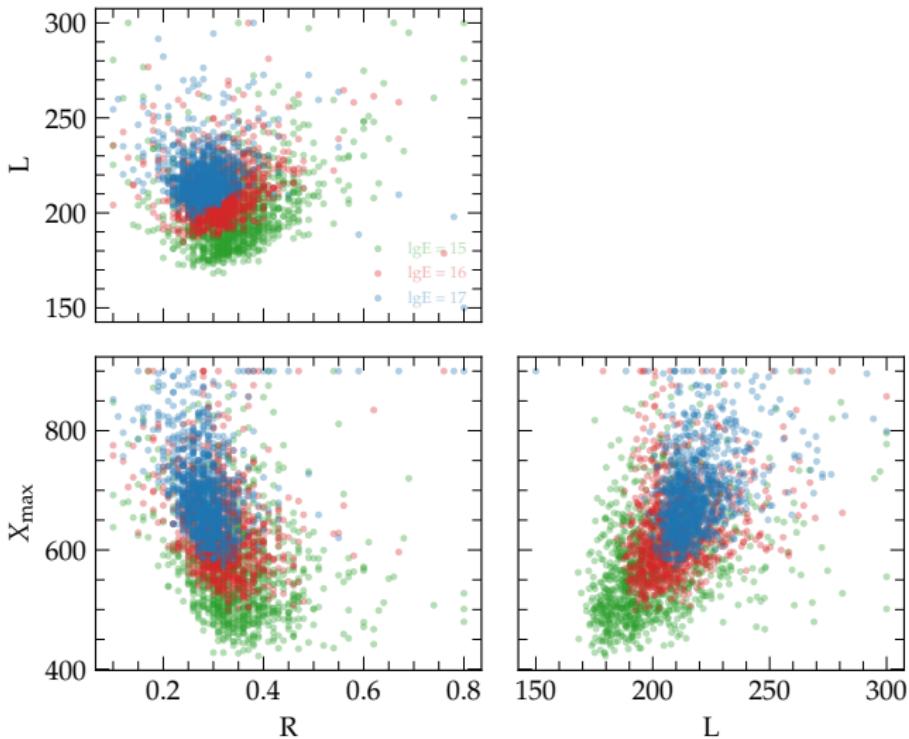
¹https://en.wikipedia.org/wiki/Gaisser–Hillas_function

Shower examples

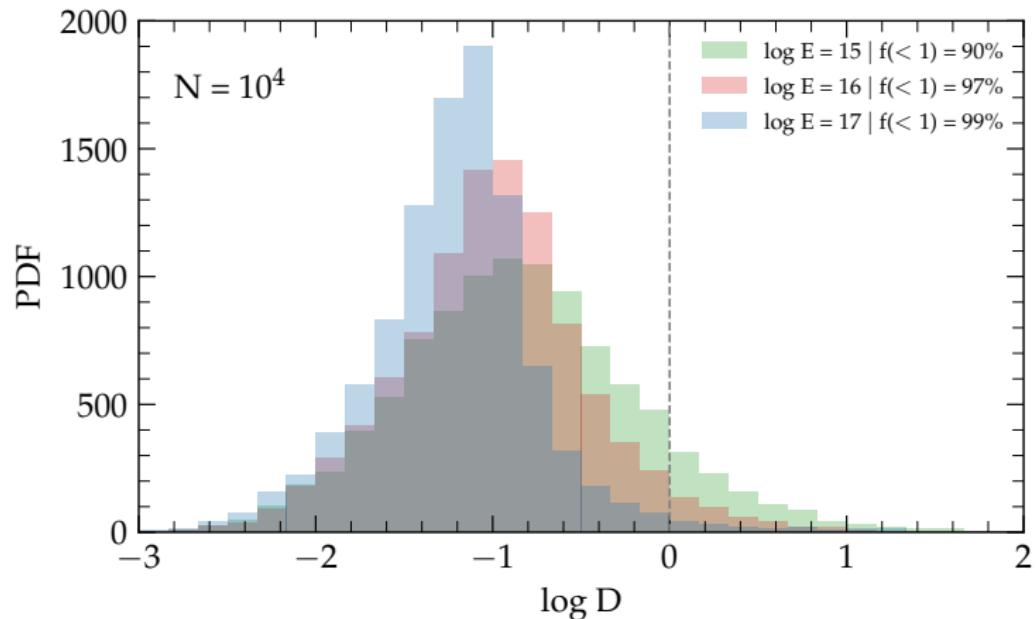


- ▷ We present four shower profiles with different best-fit values of $D = 0.1, 1, 10, 30$
- ▷ The cost function excludes the region where $y < 5\%$
- ▷ We conclude that setting a threshold of $D < 1$ to distinguish non-anomalous from anomalous profiles might be very conservative.

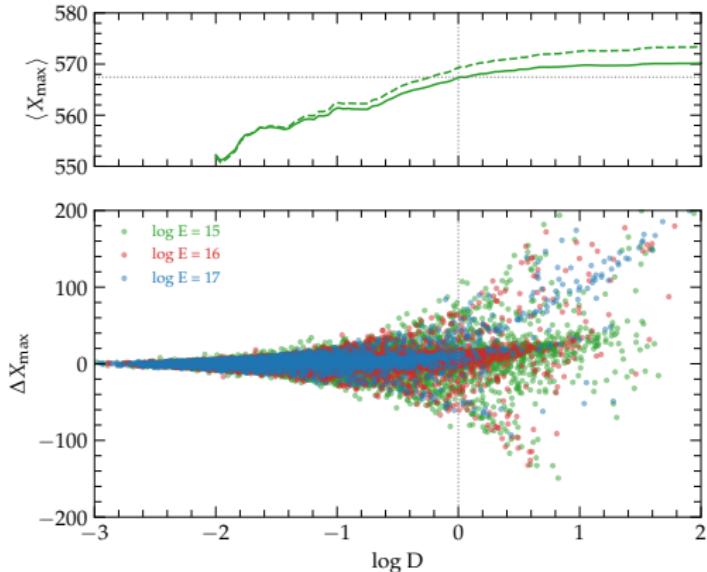
Exploring the Parameter Space of Shower Profiles



The Fraction of Anomalous Profiles

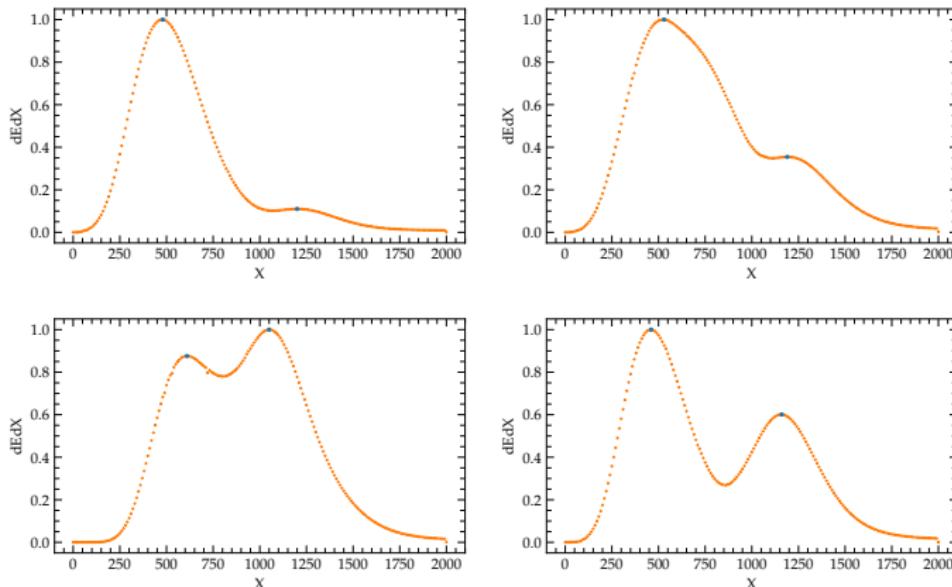


Shower inspection



- ▷ Upper plot: the mean X_{\max} as a function of the maximum D allowed
 - ▶ Solid line: the X_{\max} is computed from the G-H fit
 - ▶ Dashed line: the X_{\max} is taken from the CONEX file (2nd order polyfit around the max value)
- ▷ Lower plot: the ΔX_{\max} is the difference between the X_{\max} 's above as a function of D
- ▷ We conclude that the **maximum** impact of AP for proton showers at $E = 10^{15}$ is $\lesssim 2 \text{ gr/cm}^2$

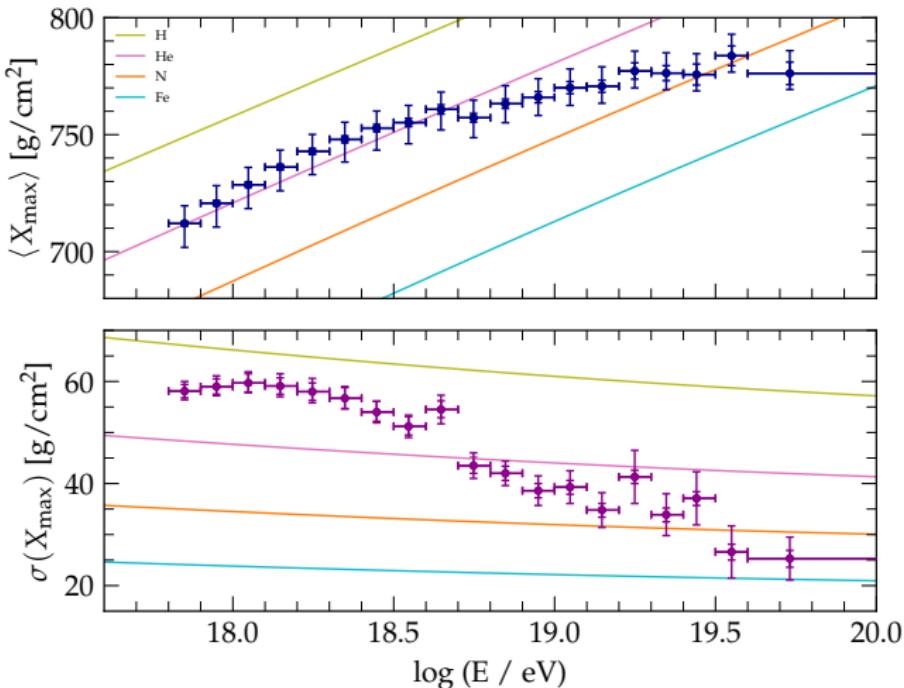
Work in progress: Anomalous Profiles with Double-Peak identification



- ▷ An alternative approach we tested consists in identify all local maxima by simple comparison of neighbouring values
- ▷ We tag as AP those with more than 1 peak → we show 4 examples where 2 peaks have been identified
- ▷ With this definition the fraction of AP is:

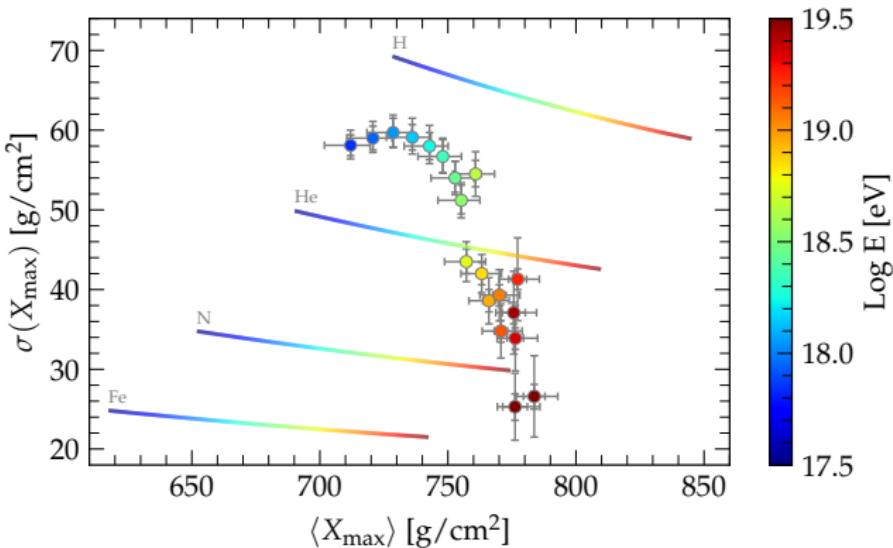
Energy	15	16	17
fraction [%]	0.65	0.17	0.03

Comparison with Auger data



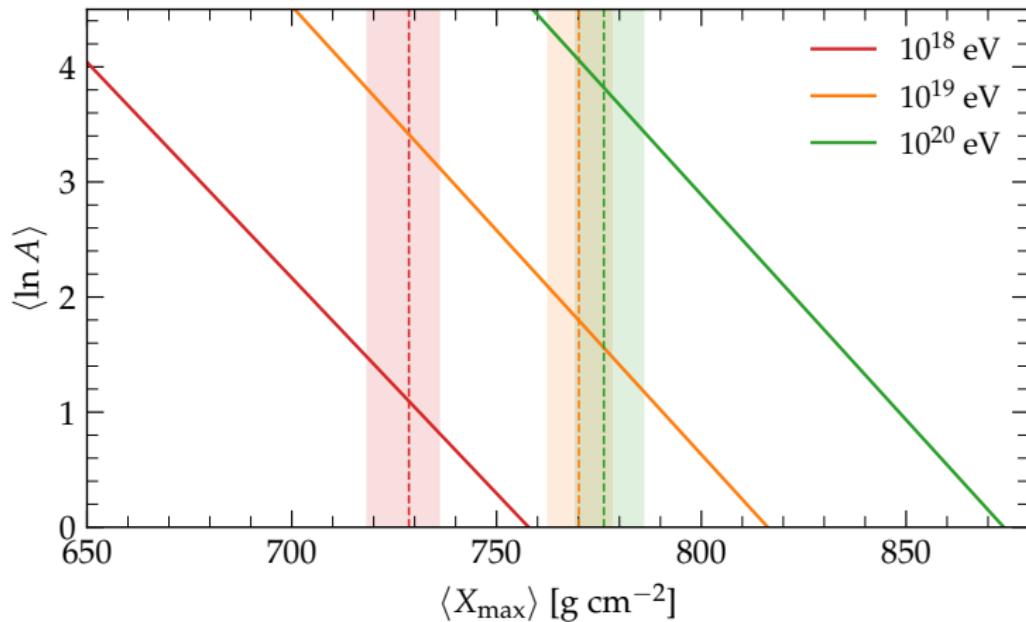
Based on Sibyll-2.3D.
P.Auger Coll., ICRC 2019, internal use only

Comparison with Auger data



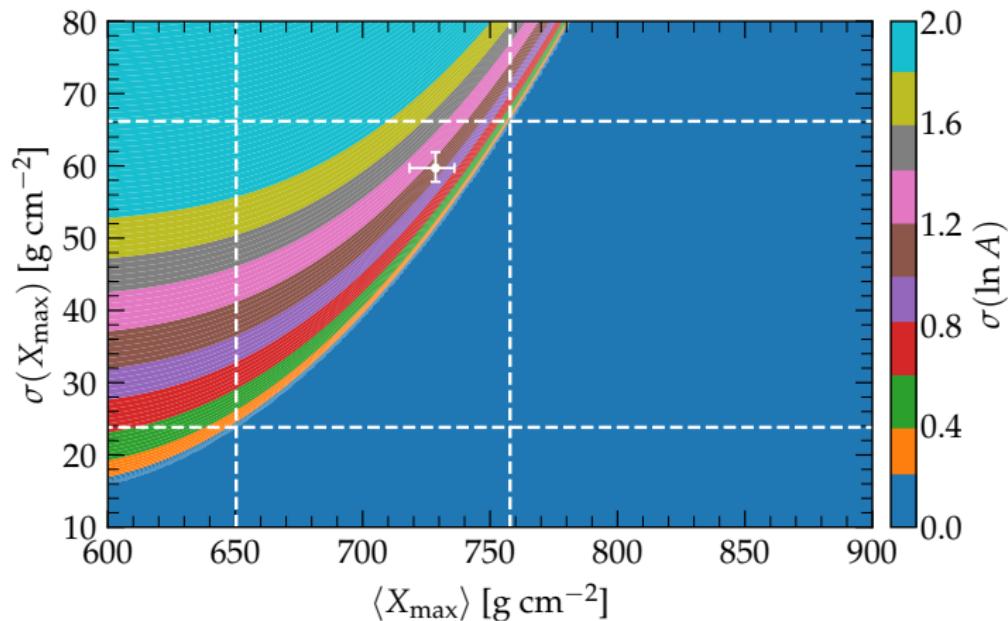
Based on Sibyll-2.3D.
P.Auger Coll., ICRC 2019, internal use only

Comparison with Auger data



Based on Sibyll-2.3D.
P.Auger Coll., ICRC 2019, internal use only

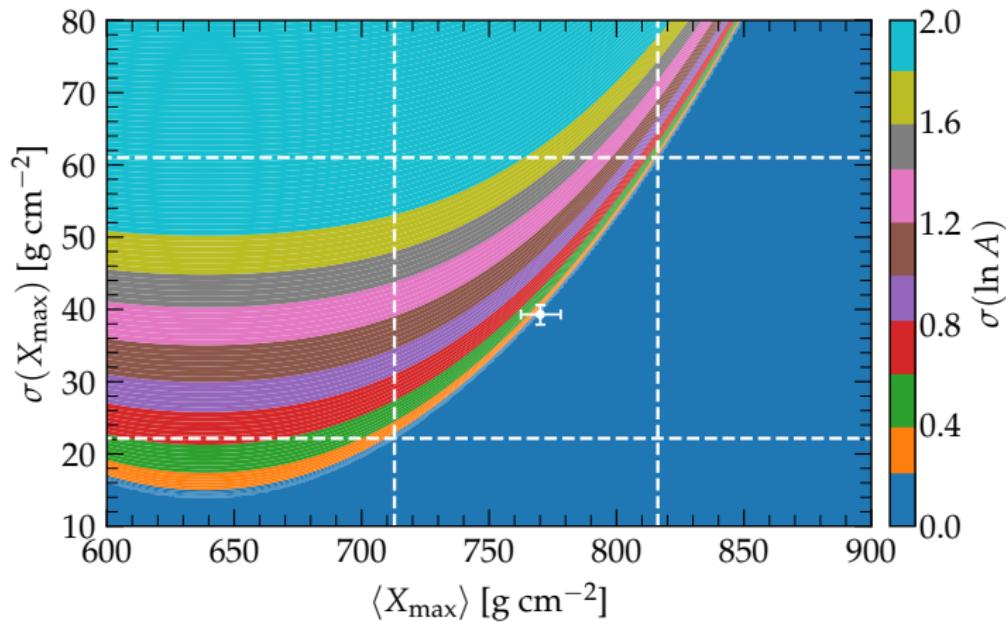
Comparison with Auger data



$$E = 10^{18} \text{ eV}$$

P.Auger Coll., ICRC 2019, internal use only

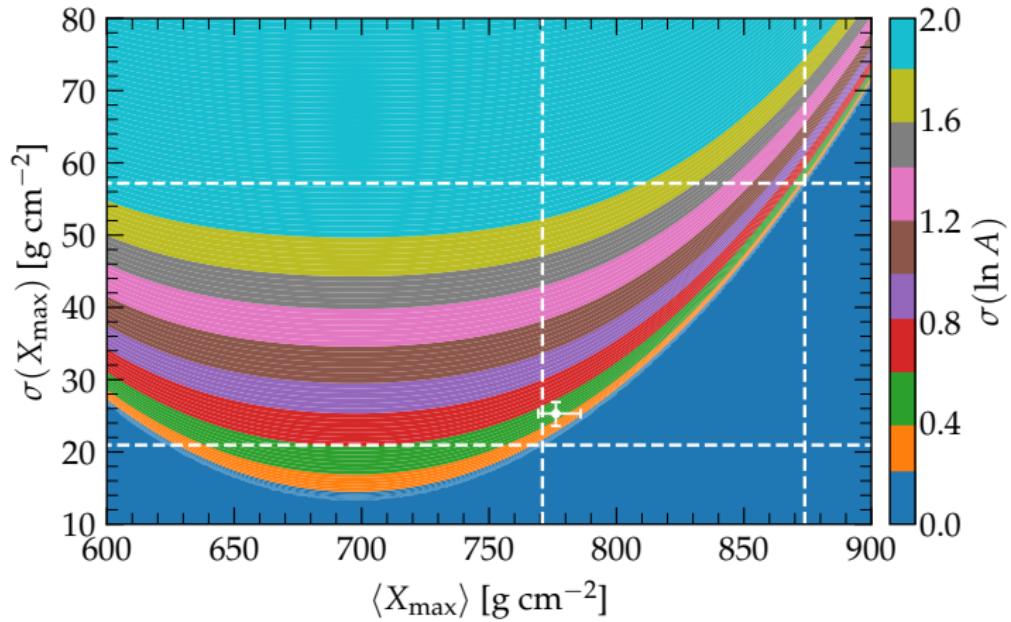
Comparison with Auger data



$$E = 10^{19} \text{ eV}$$

P.Auger Coll., ICRC 2019, internal use only

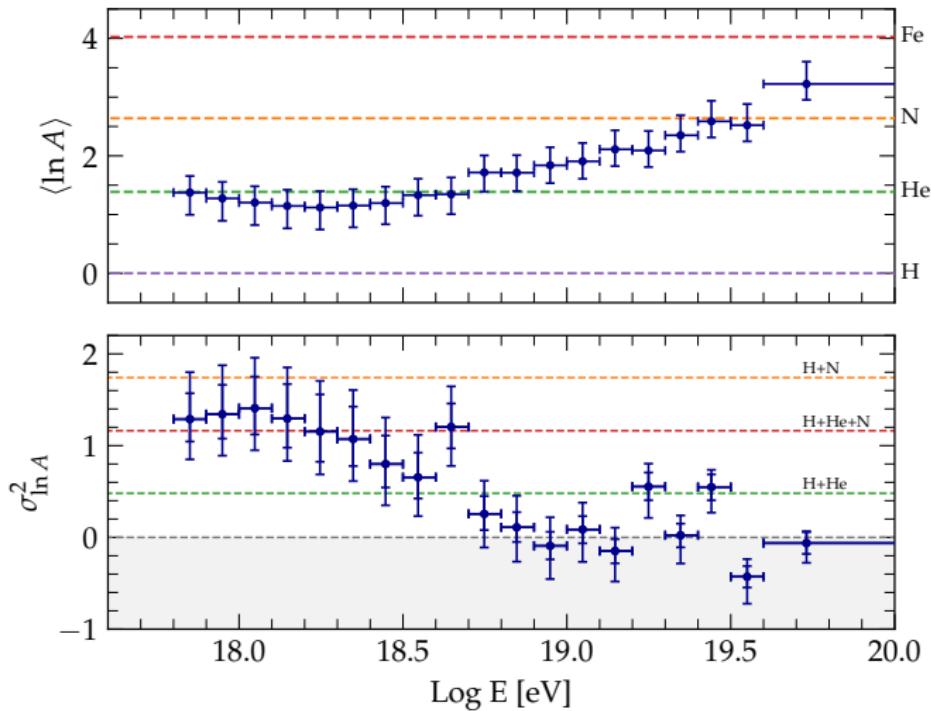
Comparison with Auger data



$$E = 10^{20} \text{ eV}$$

P.Auger Coll., ICRC 2019, internal use only

$\ln A$ moments in Auger data



Based on Sibyll-2.3D.

P.Auger Coll., ICRC 2019, internal use only

Variance of lnA in Auger data

- There are two independent sources of fluctuations: the intrinsic shower-to-shower fluctuations and the lnA dispersion arising from the mass distribution:

$$\sigma^2(x_{\max}) = \langle \sigma_{\text{sh}}^2 \rangle + \left(\frac{d\langle x_{\max} \rangle}{d \ln A} \right)^2 \sigma_{\ln A}^2 = \langle \sigma_{\text{sh}}^2 \rangle + f_E^2 \sigma_{\ln A}^2 \quad (9)$$

- We assume a parameterization for σ_{sh}^2 as follows

$$\sigma_{\text{sh}}^2(\ln A) = \sigma_p^2 [1 + a \ln A + b (\ln A)^2] \quad (10)$$

therefore

$$\langle \sigma_{\text{sh}}^2 \rangle = \sigma_p^2 [1 + a \langle \ln A \rangle + b \langle (\ln A)^2 \rangle] \quad (11)$$

- After substitution

$$\sigma^2(x_{\max}) = \sigma_p^2 [1 + a \langle \ln A \rangle + b \langle (\ln A)^2 \rangle] + f_E^2 \sigma_{\ln A}^2 \quad (12)$$

- We apply the definition

$$\langle (\ln A)^2 \rangle = \sigma_{\ln A}^2 + \langle \ln A \rangle^2 \quad (13)$$

- As a result we arrive at

$$\sigma_{\ln A}^2 = \frac{\sigma^2(x_{\max}) - \sigma_{\text{sh}}^2(\langle \ln A \rangle)}{b \sigma_p^2 + f_E^2} \quad (14)$$

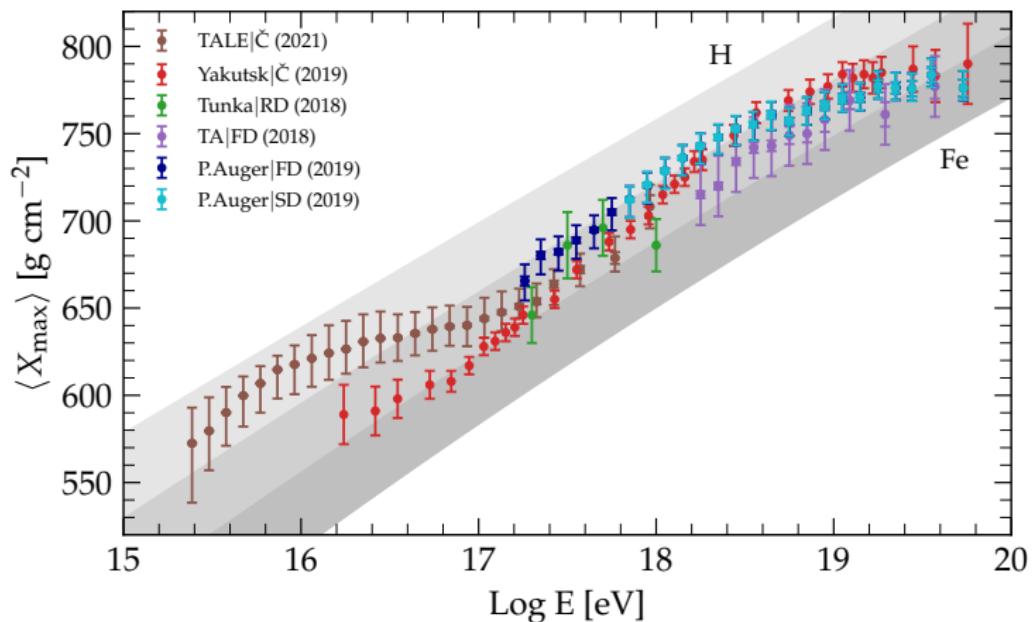
Measurements of the mean of the Xmax distribution

id	Experiment	Mode	Ref.	Table	comment
54	PAO	FD	ICRC 2019	↙	PAO public data
180	PAO	RD	UHECR 2023	✗	
193	PAO	SD	ICRC 2019	↙	PAO internal data(?)
75	TA	FD	ApJ 2018	↙	Tab. 4 but data corrected as done in Ref. 2(?)
143	TALE	Č	ApJ 2021	↙	Tab. 5 is bias-corrected(?)
178	Tunka	Č	ICRC 2021	✗	
182	Tunka	RD	PRD 2018	↙	Tab. 3
179	Yakutsk	Č	ASR 2019	↙	Tab. 2+3
183	Yakutsk	RD	ICRC 2019	✗	
181	LOFAR	radio	PRD 2021	✗	
832	Hi-Res/Mia		ApJ 2001	✗	

Datasets cited in the Snowmass paper.

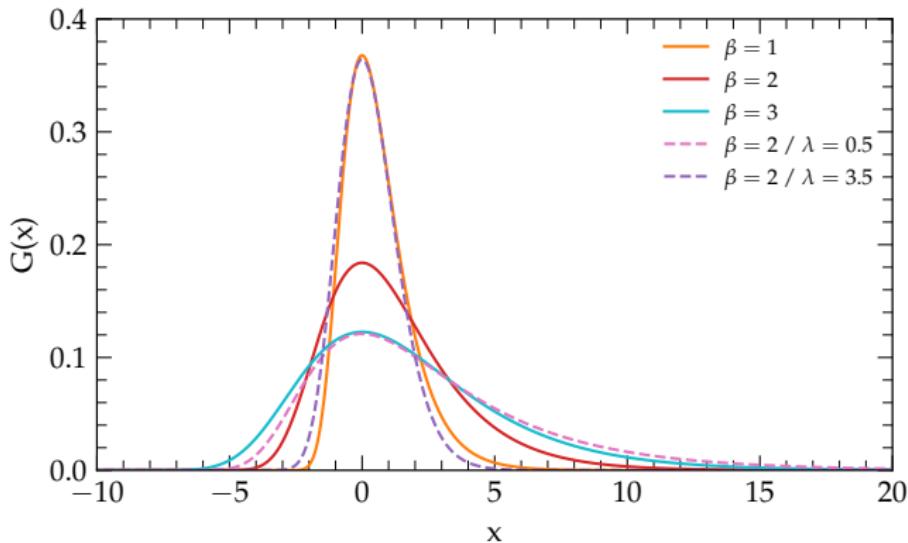
Mean InA in other datasets

Work in progress...



Based on Sibyll-2.3D.

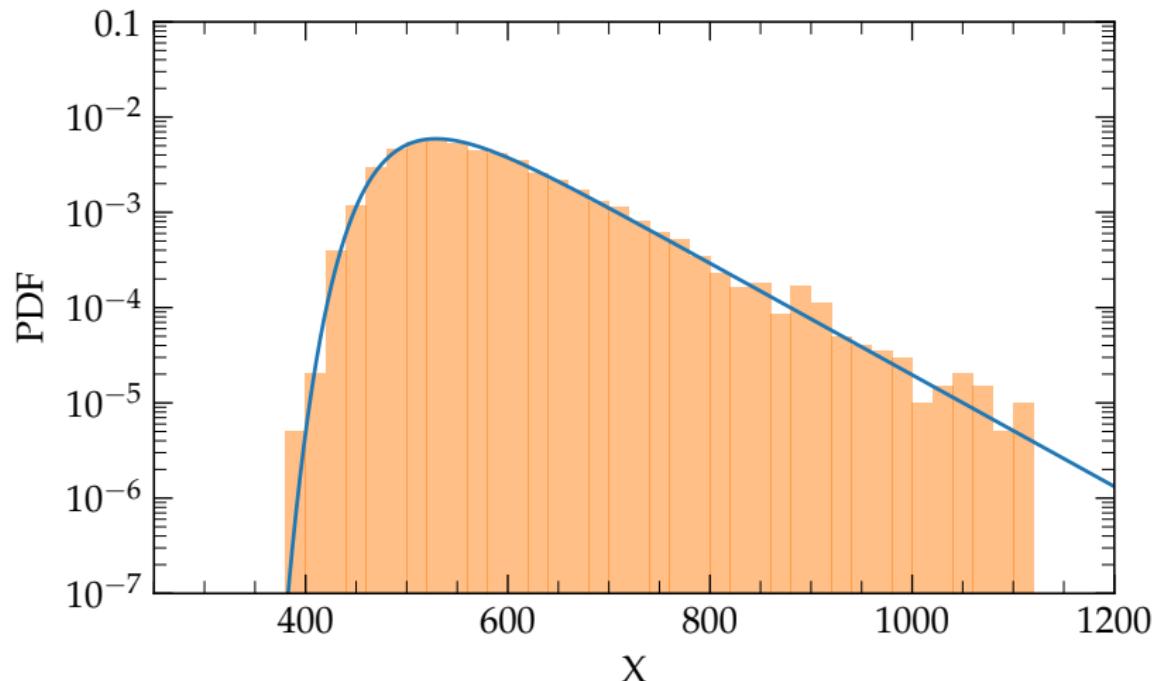
Gumbel function



$$z = \frac{x - \mu}{\beta} \quad , \quad G(z) = \frac{1}{\beta} \frac{\lambda^\lambda}{\Gamma(\lambda)} e^{-\lambda(z+e^{-z})}$$

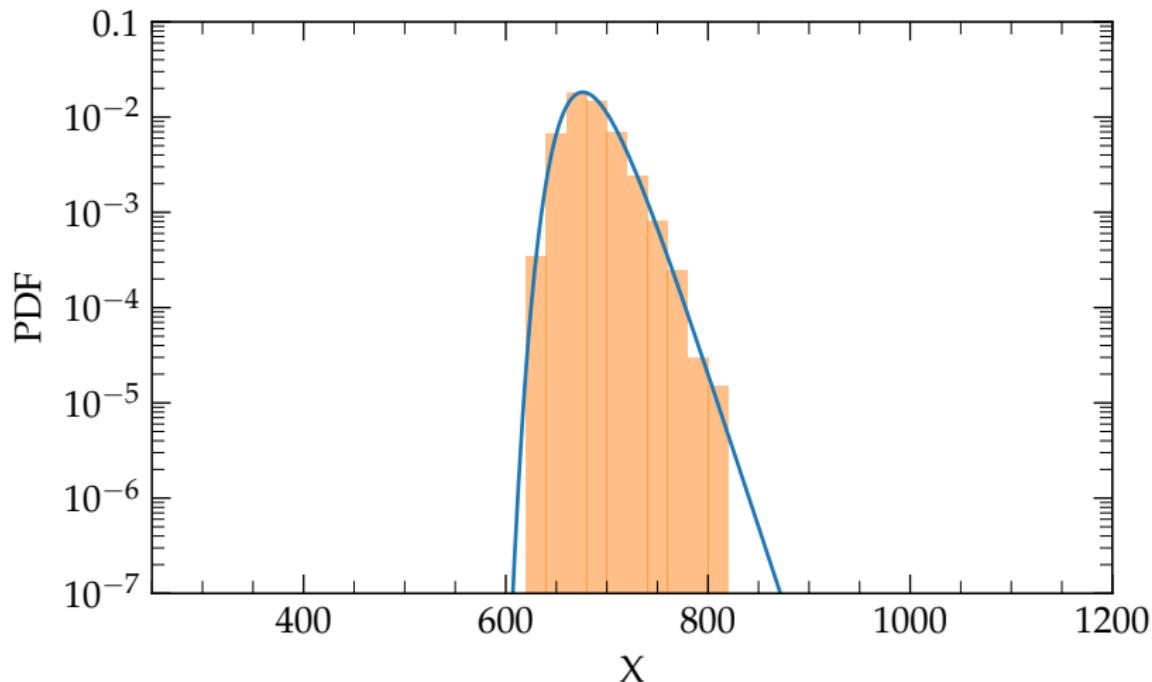
Goal: parametrize $\mu = \mu(x, y)$, $\beta = \beta(x, y)$, $\lambda = \lambda(x, y)$

Gumbel function



Mass: H, Energy: 10^{16} eV

Gumbel function



Mass: H, Energy: 10^{20} eV

Parametrizations

- ▷ μ as a function of $x \equiv \log(E/E_0)$ and $y \equiv \ln A$

$$\begin{aligned} p_0^\mu(y) &= a_0^\mu + a_1^\mu y + a_2^\mu y^2 \\ p_1^\mu(y) &= b_0^\mu + b_1^\mu y + b_2^\mu y^2 \\ p_2^\mu(y) &= c_0^\mu + c_1^\mu y + c_2^\mu y^2 \\ \mu(x, y) &= p_0^\mu(y) + p_1^\mu(y)x + p_2^\mu(y)x^2 \end{aligned}$$

- ▷ β as a function of $x \equiv \log(E/E_0)$ and $y \equiv \ln A$

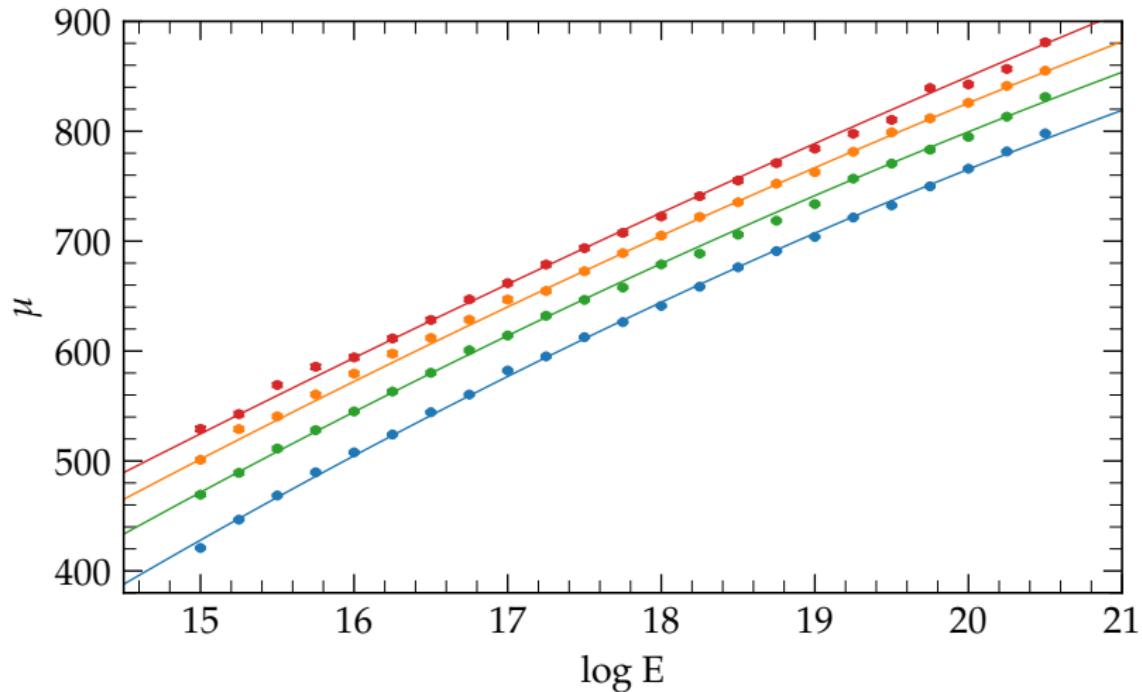
$$\begin{aligned} p_0^\beta(y) &= a_0^\beta + a_1^\beta y + a_2^\beta y^2 \\ p_1^\beta(y) &= b_0^\beta + b_1^\beta y + b_2^\beta y^2 \\ \beta(x, y) &= p_0^\beta(y) + p_1^\beta(y)x + p_2^\beta(y)x^2 \end{aligned}$$

- ▷ λ as a function of $x \equiv \log(E/E_0)$ and $y \equiv \ln A$

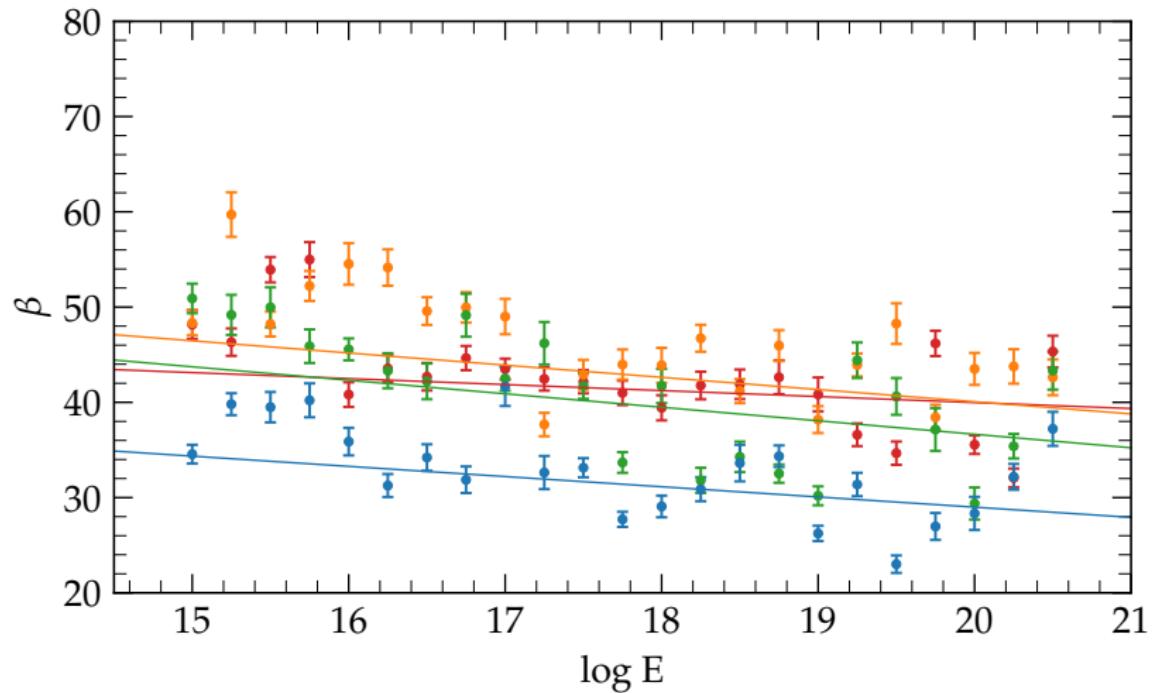
$$\begin{aligned} p_0^\lambda(y) &= a_0^\lambda + a_1^\lambda y + a_2^\lambda y^2 \\ p_1^\lambda(y) &= b_0^\lambda + b_1^\lambda y + b_2^\lambda y^2 \\ \lambda(x, y) &= p_0^\lambda(y) + p_1^\lambda(y)x + p_2^\lambda(y)x^2 \end{aligned}$$

- ▷ 21 free parameters. Cross fingers...

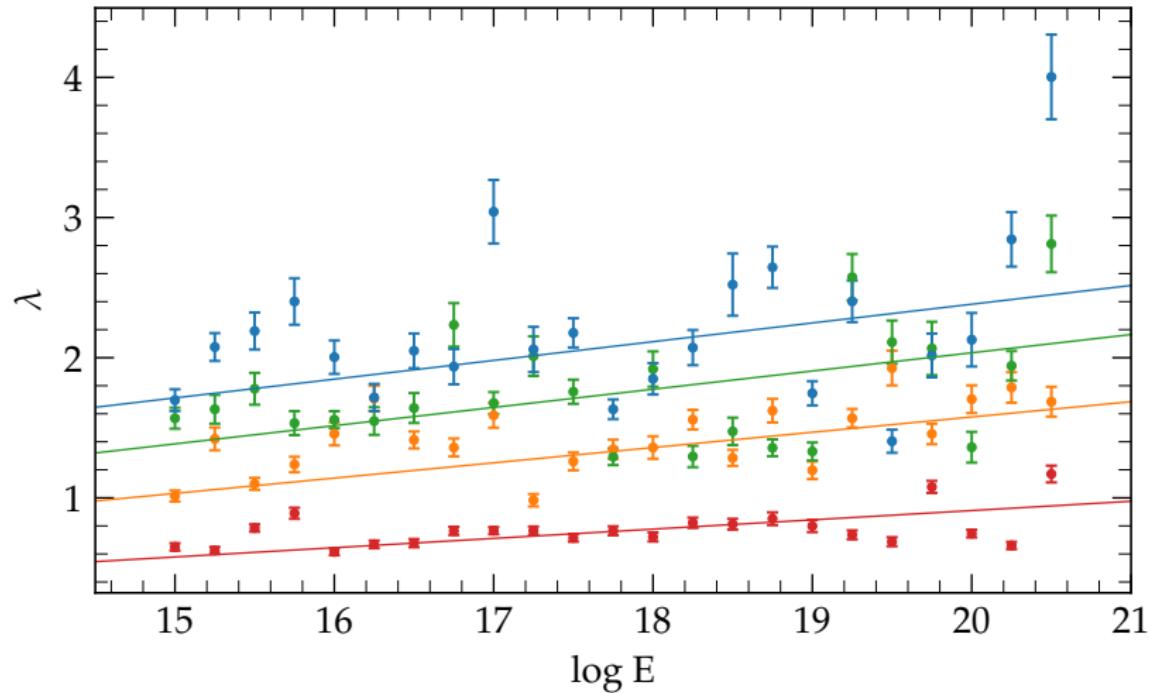
Gumbel function



Gumbel function



Gumbel function



Next Steps

- ▷ Calculating the parametrization parameters for 4 distinct HIMs → assessing the variations among these models.
- ▷ We plan to augment our dataset of CONEX simulations by 10x. This increase in data volume aims to minimize the scatter in Gumbel parametrizations.
- ▷ We are actively testing additional parametrization methods rather than Gumbel functions (see, e.g., Luan B. Arbeletche, Vitor de Souza, Astroparticle Physics 116, 2020, 102389)
- ▷ We aim at providing parametrizations and simulated X_{\max} templates public online on ZENODO