



Universidade Federal do Rio de Janeiro
Centro de Ciências Matemáticas e da Natureza
Observatório do Valongo
Programa de Pós Graduação em Astronomia



Study of anisotropy, energy and chemical composition of ultra-high energy cosmic rays measured by the Pierre Auger Observatory

Cynthia Ahiezer Vizcarra Ventura

PhD Thesis

Advisor:

Rogério Menezes de Almeida
Universidade Federal Fluminense
EEIMVR - Brasil

Advisor:

João R. T. de Mello Neto
Universidade Federal do Rio de Janeiro
Observatório do Valongo - Brasil

Rio de Janeiro, Brazil 2021

Study of anisotropy, energy and chemical composition of ultra-high energy cosmic rays measured by the Pierre Auger Observatory

Cynthia Ahiezer Vizcarra Ventura

Tese de Doutorado submetida ao Programa de Pós Graduação em Astronomia, Observatório do Valongo, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutora em Ciências (Astronomia).

Orientadores:

Rogério Menezes de Almeida

João Torres de Mello Neto

Rio de Janeiro
December, 2021

CIP - Catalogação na Publicação

VC996s Vizcarra Ventura, Cynthia Ahiezer
Study of anisotropy, energy and chemical composition of ultra-high energy cosmic rays measured by the Pierre Auger Observatory / Cynthia Ahiezer Vizcarra Ventura. -- Rio de Janeiro, 2021. 173 f.

Orientador: Rogério Menezes de Almeida.
Coorientador: João Ramos Torres de Mello Neto.
Tese (doutorado) - Universidade Federal do Rio de Janeiro, Observatório do Valongo, Programa de Pós Graduação em Astronomia, 2021.

1. UHECR. 2. Mass composition. 3. Anisotropy. 4. Energy spectrum. I. Menezes de Almeida, Rogério, orient. II. Ramos Torres de Mello Neto, João, coorient. III. Título.

Study of anisotropy, energy and chemical composition of ultra-high energy cosmic rays measured by the Pierre Auger Observatory

Cynthia Abiezer Vizcarra Ventura

Dr. João Ramos Torres de Mello Neto (Orientador)

Dr. Rogerio Menezes de Almeida (Coorientador)

Tese de Doutorado submetida ao Programa de Pós-Graduação em Astronomia, Observatório do Valongo, da Universidade Federal do Rio de Janeiro - UFRJ, como parte dos requisitos necessários à obtenção do título de Doutora em Ciências - Astronomia.

Aprovada por:

Dr^a. Rita de Cássia dos Anjos – (Avaliadora Externa) – UFPR

Dr. Carola DobrigkeitChinellato – (Avaliador Externo) – UNICAMP

Dr. Arman Esmaili Taklimi – (Avaliador Externo) – PUC/RJ

Dr.^a Heloísa Maria Castro Boechat – (Avaliador Interno) – OV/UFRJ

Dr. Rogerio Menezes de Almeida – (Coorientador) - UFF



Presidente, Dr. João Ramos Torres de Mello Neto (Orientador) – IF e OV/UFRJ

Esta defesa realizou-se através de videoconferência, completamente remota, em conformidade com as Resoluções do CEPG n.º 1 de 16 de março de 2020 e n.º 2 de 24 de abril de 2020.

Novembro de 2021

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

A toda mi familia

Acknowledgments

My name is on the first page, but I could not have written this work without the unconditional support from a lot of people, therefore, I want to dedicate these lines to express my gratitude.

First of all, I would like to thank my supervisors, Rogério Menezes and João de Mello Neto, for giving me the opportunity to join their research group and to carry out this work on this interesting research topic in the Pierre Auger Collaboration. I specially thank Rogério for his constant support, guidance, ideas and discussions which helped to me bring my PhD thesis in its final shape. I am grateful for all your advice and encouragement during my candidature and for always finding the time to follow my progress. I also owe a special thanks to João for his enthusiastic support and active participation in discussing the details of the analysis carried out in this work.

I would like to thank all the members of the Auger-Rio-Lorena group: Bruno Lago, Diego Correia, Fernando Catalani, Jaime Souza, Rafael Batista, for all the knowledge of physics and computing that taught me and for all the hours of discussions about my work, for quickly answering every question I had, but also for their patience and sense of humor. It has been great working with all of you. I also wish to extend my thanks to Marcio Müller for his suggestions proved useful in part of my research.

Thanks to my friends and colleagues in Valongo Observatory, for all the fun conversations and all the great initiatives to organize social activities. Many thanks to Astrid, Heidy, Luize, Yanna, Douglas, Vanessa, Kim, Fredi, Veronica and others for the good conversations, trips, dinner, when I was stuck with my own work. It was amazing meeting so many wonderful people.

Thanks to the members of the PhD committee and the administrative staff of the Valongo Observatory, for your support and dedication for conducting this PhD defense by video-conference in a pandemic scenario.

Thank you to Shyam Sundar, for always encouraging me to keep going and for always being there, and for the endless supply of coffee.

Thanks to everyone else I forgot to mention who crossed my path and led me to where I am now. Finally, keeping the best for last: my family, without you I would simply not be here. Thank you for teaching me to never stop learning.

Resumo

Nas últimas duas décadas, com análises dos dados colhidos pela Colaboração Pierre Auger, foi observado um grande avanço na compreensão sobre os raios cósmicos de energia ultra alta. Apesar destes resultados indicarem fortemente que os raios cósmicos de mais altas energias têm origem extragaláctica, a identificação de suas fontes ainda não foi possível.

Uma vez que a composição química dos raios cósmicos com energias acima de 10^{18} eV se torna mais pesada à medida que a energia aumenta e, considerando que o nosso conhecimento sobre os campos magnéticos galáctico e extragaláctico é escasso, a melhor maneira de identificar as fontes dessas partículas é através da combinação das diferentes medidas realizadas pela Colaboração Auger. Portanto, o principal objetivo deste trabalho é contribuir para a obtenção de um cenário global acerca destes raios cósmicos ultraenergéticos, combinando informações relacionadas à composição química da partícula primária, espectro de energia e direções de chegada.

A principal contribuição desta tese é o desenvolvimento de um código independente para o ajuste combinado da composição química e espectro de energia dos eventos medidos pela Colaboração Pierre Auger. Além disso, inserimos a informação sobre a anisotropia em larga escala, reportada para eventos com energia acima de 8×10^{18} eV. Embora seja observada uma discrepância entre as previsões do modelo e os dados na amplitude do dipolo para eventos com energias acima de 32×10^{18} eV, a ferramenta desenvolvida e disponibilizada para a Colaboração Auger é de importância fundamental a fim de estabelecermos uma imagem global do Universo a altas energias.

Adicionalmente, motivados pelo recente resultado reportado pela Colaboração Auger, no qual é mostrado que é possível obtermos informação da composição química do raio cósmico evento a evento usando redes neurais e dados do detector de superfície, estimamos o ganho no poder de detecção do estimador de correlação cruzada entre as direções dos eventos e as posições de possíveis fontes, quando utilizamos amostras contendo núcleos mais leves. Calculamos, também, mapas de probabilidades relacionados à composição química dos eventos na esfera celeste e mostramos que a técnica de *shuffling* pode ser utilizada para os eventos medidos pelo Detector de Fluorescência do Observatório.

Com o conhecimento adquirido na construção dos mapas de composição química, exemplificamos o ganho obtido no poder de detecção quando combinamos dados direcionais de intensidade e dados direcionais de composição em comparação com a análise de apenas dados direcionais de composição.

Finalmente, assumindo que os raios cósmicos com energias acima de 39×10^{18} eV, significativamente correlacionados com a posição de galáxias *starbursts*, sejam núcleos de carga Z e energia E , procuramos por um excesso de eventos nas mesmas direções a energias acima de E/Z , devido a prótons originados nas mesmas fontes.

Esperamos que as análises descritas neste trabalho contribuam para o melhor entendimento dos raios cósmicos de energia ultra alta e seu papel no Universo.

Abstract

In the last two decades, with the analysis of data collected by the Pierre Auger Collaboration, a major advance in our understanding of ultra high energy cosmic rays has been observed. Although these results strongly indicate that the cosmic rays of higher energies have an extragalactic origin, identifying of their sources has not yet been possible.

Since the mass composition of cosmic rays with energies above 10^{18} eV becomes heavier as the energy increases and, considering that our knowledge about the galactic and extragalactic magnetic fields is scarce, the best way to identify the sources of these particles is through the combination of the different measures performed by the Auger Collaboration. Therefore, the main goal of this work is to contribute to obtaining a global scenario about these ultra-energetic cosmic rays by combining information related to the chemical composition of the primary particle, energy spectrum, and arrival directions.

The main contribution of this thesis is the development of an independent code for the combined fit of the chemical composition and energy spectrum of the events measured by the Pierre Auger Collaboration. In addition, we added information about large-scale anisotropies reported for events with energy above 8×10^{18} eV. Although a discrepancy is observed between the model's predictions and the data on the dipole amplitude for events with energies above 32×10^{18} eV, the tool developed and made available to the Auger Collaboration is of prime importance in order to establish a global image of the Universe at high energies.

Additionally, motivated by the recent result reported by the Auger Collaboration, in which it is shown that it is possible to obtain information on the chemical composition

of the cosmic rays on an event-by-event basis by using neural networks and data from the surface detector, we estimated the gain in the detection power of the cross-correlation estimator between the directions of the events and the positions of possible sources, when using samples containing lighter nuclei. We also calculated probability maps related to the chemical composition of events in the celestial sphere and showed that the shuffling technique can be used for the events measured by the Fluorescence Detector of the observatory.

With the knowledge obtained during the construction of the chemical composition maps, we exemplified the gain in the detection power when we combine directional intensity data and directional composition data in comparison with the analysis of directional composition data only.

Finally, assuming that the cosmic rays with energies above 39×10^{18} eV, significantly correlated with the position of starbursts galaxies, are nuclei of charge Z and energy E , we searched for an excess of events in the same directions at energies above E/Z , due to protons originated from the same sources.

We hope that the analyzes described in this work will contribute to a better understanding of ultra-energetic cosmic rays and their role in the Universe.

List of Figures

2.1	The all particle cosmic ray energy spectrum.	7
2.2	The fraction of cosmic rays that have been created with an energy above E_{th}	12
2.3	Energy loss lengths for different processes	14
2.4	The Hillas plot classifying potential sources of UHECRs according to their size and the magnetic field strength.	17
3.1	Schematic view showing the positions of the detectors at the site of the Pierre Auger Observatory	20
3.2	Surface Detector station (SD) of Auger	22
3.3	FD Building at Los Leones	23
3.4	Schematic of the main components of an FD	24
3.5	Extension to observatory, HEAT	25
3.6	Schematic view of the geometric reconstruction of the shower.	26
3.7	Scheme of the shower front estimation.	27
3.8	Combined energy spectrum, adjusted to a theoretical flux.	29
3.9	Combined.	30
3.10	X_{max} and $\sigma(X_{max})$ as a function of energy.	32
3.11	Sky map in equatorial coordinates.	33
3.12	Energy dependence of the dipolar amplitude measured above 4 EeV	33

LIST OF FIGURES

3.13	Scan of test statistic (TS) as a function of the energy threshold for SBGs and γ -AGNs with and without take into account the attenuation of the intensity due to energy losses	35
3.14	Observed excess map for starburst galaxies.	36
3.15	Deviance $\sqrt{D - D_{min}}$, as function of γ and $\log_{10}(R_{cut}/V)$	38
3.16	The Auger energy spectrum at the top of the Earth's atmosphere	39
4.1	Distributions of the number of pairs obtained from one specific simulation of an isotropic cosmic-ray sky containing 1,000 events.	43
4.2	Example of Landy-Szalay cross-correlation estimator between a simulated isotropic dataset and positions of starburst galaxies.	44
4.3	χ^2 distributions obtained for a scenario with a fraction of proton $f_p = 5\%$ and different fractions of selected events $f_{X_{max}}^{sel}$	47
4.4	Simulated X_{max} distributions	48
4.5	Detection power	49
5.1	Mean X_{max} map for events with energy above $10^{18.7}$ eV using a top-hat angular window of 20° . The color scale unity is in g/cm^2	51
5.2	Probability composition sky map obtained for $E > 10^{18.7}$ eV.	53
5.3	Sky map of cosmic-ray composition for $E \geq 10^{18.7}$ eV obtained by using a different test statistic and a top-hat filter of 30° of radius.	53
5.4	Left: example of the ADC signal as a function of the time bin for a given triggered pixel of a real event. Right: distribution of the signal variance for all triggered events.	55
5.5	Left: example of the ADC signal as a function of the time bin for a given triggered pixel of a real event. Right: signal variance distribution for all triggered events.	55
5.6	Mean zenith angle as a function of the variance. The variance bins are of different size in order to contain the same number of events.	56

5.7	At the left, we have the zenith angle distribution of events according to different time windows during the FD shifts. Right, average zenith angle distribution according to each time bin.	57
6.1	Behavior of the proton fraction as a function of the right ascension, $x(\alpha)$, for different values of κ	60
6.2	Event count rate (in the left panel) and $\langle X_{max} \rangle$ (in the right panel) as a function of right ascension obtained from one simulated sky with $\kappa = 1$. . .	61
6.3	Functions $f(\alpha)$ and $g(\alpha)$ corresponding to the distributions of Fig. 6.2 for one simulated sky with $\kappa = 1$	61
6.4	Detection power as a function of κ . The black line corresponds to a scenario without taking into account the detector resolution of the experiment. . .	63
6.5	Detection power as a function of κ for several right ascension bin sizes. . .	63
6.6	Detection power as a function of κ considering samples measured by the FD and SD.	64
6.7	Map of the proton fraction x in for $\kappa = 2$ in Galactic coordinates.	65
6.8	Example of $f(\vec{n})$ and $g(\vec{n})$ functions, in Galactic coordinates, resulting from one specific simulation, considering angular windows with 20° of radius. . .	66
6.9	Detection power as a function of κ for different top-hat filter radii.	67
6.10	Detection power as a function of κ for different top-hat filter radii.	67
7.1	The directions of the 23 brightest nearby starburst galaxies.	70
7.2	Coverage maps in galactic coordinates. Left: $E \geq 3.0$ EeV Right: $E \geq 1.5$ EeV.	71
7.3	Energy loss length of the helium as a function of energy, which indicates the bands of each process of loss of energy.	74

LIST OF FIGURES

8.1 *Left:* Minimum values of $\sqrt{D - D_{min}}$ as a function of $(\gamma, \log_{10}(R_{cut}/V))$ obtained using the package Minuit. *Right:* Energy spectrum of UHECRs at Earth (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2015 dataset. 83

8.2 X_{max} distributions from the model for the best fit (histogram) and data (profile histogram) for different energy bins. ICRC 2015 dataset. 84

8.3 *Left:* Minimum values of $\sqrt{D - D_{min}}$ as a function of $(\gamma, \log_{10} R_{cut}/V)$ obtained using the package Minuit. *Right:* Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. 85

8.4 X_{max} distributions from the model for the best fit (histogram) and data (profile histogram) for different energy bins. ICRC 2019 dataset. 86

8.5 *Left:* Minimum values of $\sqrt{D - D_{min}}$ as a function of $(\gamma, \log_{10}(R_{cut}/V))$ obtained using the package Minuit. *Right:* Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. PRL 2020 dataset. 87

8.6 X_{max} distributions from the model for the best fit (histogram) and data (profile histogram) for different energy bins. ICRC 2015 dataset. 88

8.7 Expected flux maps $\Phi_{LM,j}(\hat{n})$ for detected events with $E > 32$ EeV resulting from the injection of different nuclei with $(\gamma, R_{cut}) = (-1.0, 18.2)$ at sources distributed as our local matter. 92

8.8 Expected flux maps $\Phi_{LM,j}(\hat{n})$ for detected events with $E > 32$ EeV resulting from the injection of different nuclei with $(\gamma, R_{cut}) = (-1.0, 20)$ at sources distributed as our local matter. 93

8.9	(a) Minimum values of $\sqrt{D - D_{min}}$ as a function of $(\gamma, \log_{10} R_{cut}/V)$ obtained using Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges: H (red), He (gray), N (green), Si (cyan) and Fe (blue). (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red).	96
8.10	X_{max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins.	97
8.11	(a) All Deviances ($D_J, D_{X_{max}}, D_{dip}$) divided by the number of experimental points as a function of points for values of $\sigma_0 = 10^\circ, 20^\circ$ and 25° . (b) $\log_{10}(R_{cut})$ as a function of γ for three values of σ_0 considered. (c) Dipolar amplitudes as a function of energies for the three considered values of σ_0 . (CTGE)	99
8.12	Deviance of anisotropy divided by the number of experimental points D_{dip}/N as a function of σ_0 by performing a scan in σ_0 . The minimum is obtained for $\sigma_0 = 33^\circ$ corresponding to a $D_{dip}/N = 2.95$	100
8.13	<i>Left:</i> The deviance of the anisotropy as a function of σ_0 considering source evolution for $m = -3, m = 0$ and $m = +3$. <i>Right:</i> The dipole amplitudes as a function of energy for the corresponding optimal value of σ_0 , for each case of source evolution m	101
8.14	Dipole amplitudes as a function of energy expected from the CTGS model (in blue) and measured by the Pierre Auger Collaboration (in red) in the scenario with a X_{max} shift $\delta = -20 \text{ g /cm}^2$ and $\sigma_0 = 10^\circ$	104

LIST OF FIGURES

B.1 (a) Minimum values of $\sqrt{D - D_{min}}$ as a function of $(\gamma, \log_{10} R_{cut}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red) using $\sigma_0 = 10^\circ$ and considering no source evolution . . . 125

B.2 X_{max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset. 126

B.3 (a) Minimum values of $\sqrt{D - D_{min}}$ as a function of $(\gamma, \log_{10} R_{cut}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red) using $\sigma_0 = 25^\circ$ and considering no source evolution . . . 127

B.4 X_{max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset. 128

B.5 (a) Minimum values of $\sqrt{D - D_{min}}$ as a function of $(\gamma, \log_{10} R_{cut}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red) using $\sigma_0 = 20^\circ$ and considering source evolution $m = -3$ 129

B.6 X_{max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset. 130

B.7 (a) Minimum values of $\sqrt{D - D_{min}}$ as a function of $(\gamma, \log_{10} R_{cut}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red) using $\sigma_0 = 20^\circ$ and considering source evolution $m = +3$	131
B.8 X_{max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset.	132
B.9 (a) Minimum values of $\sqrt{D - D_{min}}$ as a function of $(\gamma, \log_{10} R_{cut}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red).	133
B.10 X_{max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset.	134
B.11 (a) Minimum values of $\sqrt{D - D_{min}}$ as a function of $(\gamma, \log_{10} R_{cut}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red).	135
B.12 X_{max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset.	136

LIST OF FIGURES

B.13 (a) Minimum values of $\sqrt{D - D_{min}}$ as a function of $(\gamma, \log_{10} R_{cut}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red). 137

B.14 X_{max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset. 138

B.15 (a) Minimum values of $\sqrt{D - D_{min}}$ as a function of $(\gamma, \log_{10} R_{cut}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red). 139

B.16 X_{max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset. 140

B.17 (a) Minimum values of $\sqrt{D - D_{min}}$ as a function of $(\gamma, \log_{10} R_{cut}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red). 141

B.18 X_{max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset. 142

B.19 (a) Minimum values of $\sqrt{D - D_{min}}$ as a function of $(\gamma, \log_{10} R_{cut}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red).	143
B.20 X_{max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset.	144
B.21 (a) Minimum values of $\sqrt{D - D_{min}}$ as a function of $(\gamma, \log_{10} R_{cut}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red).	145
B.22 X_{max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset.	146

List of Tables

4.1	The 23 selected starburst galaxies	45
7.1	Total observed and background event numbers for different energy bins. The uncertainty in the number of background events, N_{bkg} , was evaluated assuming Poissonian statistics.	72
7.2	Total, observed and background event numbers for different energy bins for M83 (a), NGC 253 (b) and NGC 4945 (c).	73
7.3	Total observed and background event numbers for different energy bins. The uncertainty in the number of background events, N_{bkg} , was evaluated assuming Poissonian statistics.	74
7.4	Total, observed and background event numbers for different energy bins for M83, NGC 4945 and NGC 253.	76
8.1	Summary of the γ and R_{cut} parameter estimation for the data sets con- sidered. The values of the deviances for the best-fit parameters are also presented. For the ICRC 2015 dataset, we included the values for the two minima.	85
8.2	Summary of the abundance parameter estimation for the data sets con- sidered. For the ICRC 2015 dataset, we included the values for the two minima.	87
8.3	Best-fit results obtained by including anisotropies in the fit for different smearing parameters σ_0 (10° , 20° and 25°).	95

8.4	Best-fit results in the scenario obtained by including anisotropies in the fit EPOS-LHC with smearing (20°).	100
8.5	Best-fit results in the scenario including anisotropies in the fit EPOS-LHC by considering a systematic shift on X_{\max} with (20°).	102
8.6	Best-fit results obtained with Sibyll 2.3d hadronic interaction model, considering or not a shift in the X_{\max} distribution with different smearing parameters (10° , 20°).	104
A.1	Parameters of acceptance and resolution used to fit the X_{\max} distributions. (all values in g/cm^2 , except f , which is adimensional).	123

Contents

Acknowledgments	x
Resumo	xii
Abstract	xiv
List of Figures	xv
List of Tables	xxiv
Contents	xxvi
1 Introduction	1
2 Overview of Ultra-High Energy Cosmic Rays	4
2.1 Cosmic-ray spectrum	5
2.2 Acceleration Mechanisms	8
2.3 Propagation of Ultra-High Energy Cosmic Rays	10
2.3.1 Radiation field interactions	10
2.3.2 Energy loss processes	13
2.3.3 Magnetic fields deflection	13
2.4 Candidate sources	15
2.5 Chemical composition	17

3	The Pierre Auger Observatory	19
3.1	The detectors of the Pierre Auger Observatory	19
3.1.1	The Surface Detector	20
3.1.2	The Fluorescence Detector	22
3.2	Hybrid reconstruction	25
3.3	Main results of the Pierre Auger Observatory	28
3.3.1	Energy spectrum	28
3.3.2	Mass composition	30
3.3.3	Arrival directions	31
3.3.4	Combined Fit results of spectral and chemical composition data measured by Pierre Auger Observatory	36
4	Sensitivity of cross-correlation studies by using X_{\max} information measured by the Surface Detector of the Pierre Auger Observatory	40
4.1	Landy-Szalay cross-correlation estimator	41
4.2	Detection sensitivity	44
5	Study of the mass composition as a function of the arrival direction	50
5.1	On the use of the shuffling technique for FD events	54
6	On combining directional intensity and direction composition	58
6.1	Directional data in terms of right ascension only	59
6.2	Directional data in terms of sky maps	64
7	Correlation Studies of UHECRs events and starburst sources at lower energies	68
7.1	Data analysis	69
7.2	Conclusions	75
8	Combined fit of the energy spectrum, mass composition and arrival directions detected at the Pierre Auger Observatory	77

CONTENTS

8.1	The Rio Combined Fit package	78
8.1.1	Energy spectrum	78
8.1.2	X_{\max} distributions	79
8.1.3	The deviance	80
8.1.4	The datasets	81
8.1.5	Running the code	82
8.1.6	Results	82
8.1.7	Conclusions	89
8.2	Combined fit with addition of anisotropy information	89
8.2.1	Results	95
8.2.2	Conclusions	103
9	Conclusions	106
	Bibliography	109
	Appendix	121
A	X_{\max} distributions used in the global combined fit	121
A.1	Gumbel Distributions	121
A.2	Detector Acceptance	122
A.3	Resolution parameters	122
B	Results in the best-fit scenario	124
B.1	Running the CODE	147
B.1.1	The input file	147

Chapter 1

Introduction

Ultra-high energy cosmic rays (UHECRs), particles with energies above 10^{18} eV (1 EeV = 10^{18} eV), are messengers of astrophysical sources carrying fundamental information about the Universe. Although these high energy particles have been detected since the 1960s, their sources, chemical composition, mechanism of acceleration at the sources, and how they propagate to Earth are still open questions.

A giant experiment was built aiming to answer these questions and has been producing important results: the Pierre Auger Observatory, located in the province of Mendoza, Argentina. Despite the significant results achieved, a global picture with respect to those questions is still not established.

The investigation about the cosmic-ray sources through the study of the arrival direction of the particles is completely dependent on the primary mass composition because heavier particles have an average larger deflection during the propagation on the Galactic and extragalactic magnetic field. **Therefore, the main goal of this thesis is to contribute to obtaining a comprehensive scenario about the UHECRs by combining information related to the mass compositions, energy and arrival directions of those particles.**

The thesis is structured as follows. In chapter 2, an overview of the UHECR field is presented, discussing possible source candidates for the acceleration of these particles as

well as energy loss processes and deflections by magnetic fields during their propagation from sources to Earth. In chapter 3, we describe the Pierre Auger Observatory, its detection techniques and the most important results. Subsequently, the studies conducted in this work are presented in the following chapters.

Part of the difficulty in identifying the sources of UHECRs comes from the fact that the information on the chemical composition of primary cosmic ray is obtained statistically from the distribution of the atmospheric depth in which the energy deposition by the atmospheric shower is maximum, X_{\max} . At the Pierre Auger Observatory, this information is obtained by the fluorescence telescopes, which operate only on clear nights, without moonlight, corresponding to $\sim 13\%$ of the operating time of the experiment. We recently showed in [1] that it is possible to obtain the same information in an event-by-event basis by using deep neural networks and the surface detectors of the experiment (which operate 100% of the time). This possibility opens up a very interesting window for anisotropy studies. In this way, we exemplify, in chapter 4, the gain in detection efficiency of standard cross-correlation analyses, such as the Landy-Szalay estimator [2], when we select a sample containing only events with the highest X_{\max} values (corresponding to the lightest charged particles, whose deflections in magnetic fields are smaller on average).

In chapter 5, as an important cross-check of the composition anisotropy reported in [3], we build mass composition maps and compute the statistical significance of larger or smaller $\langle X_{\max} \rangle$ through the celestial sphere in a completely independent way. Besides, a useful strategy to compute the statistical significance of these composition maps is to perform the *shuffling* technique, which implicitly assumes that the zenith angle distribution of detected events is time-independent. Since there are different sources of background light during the period of the Fluorescence Detector data, this assumption can be no longer valid. Therefore, we also investigate in this chapter the applicability of the shuffling technique to events detected by the Fluorescence Detector studying the behavior of the mean zenith angle distribution as a function of the ADC count signal variance.

After studying directional intensity data in chapter 4 and directional composition data in chapter 5, we exemplify, in chapter 6, the net gain obtained by **combining both**

information in comparison to the standalone analysis of directional composition data only (as done in chapter 5). For this, we used a cross-correlation estimator in order to characterize whether possible similarities between these data are significant or not. Sensitivity curves as a function of the composition modulation are obtained.

In chapter 7, motivated by the indication of anisotropy in the arrival directions of events with energies above 39 EeV [4] through comparison to the flux pattern of extragalactic gamma-ray sources, we search for excesses of events at lower energies around the directions of the starburst galaxies by considering the proposition of Waxman and Lemoine [5]. According to it, assuming that the cosmic ray acceleration depends just on the particle rigidities and neglecting the energy losses during the cosmic-ray propagation, an anisotropy at highest energy due to nuclei of charge Z and energy E should be followed by an anisotropy in the same direction of the sky at energies Z times smaller (E/Z) due to protons originated in the same sources. No significant excess of events was found in any of the energy bins considered.

The main contribution of this thesis is presented in chapter 8, in which we perform a combined fit of the energy spectrum, mass composition, and arrival direction measurements obtained by the Pierre Auger Observatory. Best fits are obtained for different source evolutions, magnitudes of magnetic deflections, and hadronic interaction models. A summary of the conclusions as well as a short outlook on the future improvements of the analyses performed in this thesis are described in chapter 9.

Chapter 2

Overview of Ultra-High Energy Cosmic Rays

The discovery of cosmic rays was undoubtedly a breakthrough in Science. By 1912, in an experiment on board of a balloon, Victor Hess observed that the ionization rate of the Earth's atmosphere at a height of 5300 meters was about three times higher than at sea level. This finding led Hess to conclude that the excess rate of ionization at higher altitudes is due to the radiation coming from outer space. Hess received the Nobel Prize in Physics in 1936 for this discovery. Between the 30s and 50s of the 20th century, it is interesting to know that before the construction of particle accelerators, the main source of the discovery of new particles was through cosmic rays. In 1932, the positron (e^+) was discovered by Carl Anderson when he studied cosmic-ray trajectories in a Wilson chamber with a Pb plate inside [6]. Later, in 1937, Carl Anderson and Seth Neddermeyer discovered the muon by analyzing cosmic-ray traces by inserting a Geiger counter into a Wilson chamber [7]. The discovery of pion was made in 1947 by Cesar Lattes, G. Occhialini and C. Powell via exposing emulsion plates to cosmic radiation at higher altitudes on Mount Chacaltaya in Bolivia [8]. In 1938, Pierre Auger used two ionizing radiation detectors spatially separated and detected the simultaneous arrival of a high flux of particles that provided a first means to estimate the energy of cosmic rays. In this experiment, Auger discovered the cascades of

secondary particles and nuclei resulting from the collision of primaries like protons, helium, heavier nuclei, e^+ and e^- with air molecules, called extensive air showers. In this pioneering experiment, Pierre Auger was capable of measuring energies to an upper limit of 10^{15} eV. The experiment occupied an area of 8 km^2 , with 19 scintillators positioned at a distance of $\sim 1 \text{ km}$ between them [9].

Later in 1962, John Linsley developed the arrays of large detectors which were used for the first detection of an ultra-high energy cosmic ray of 100 EeV (10^{20} eV) at the Volcano Ranch experiment in New Mexico [10]. Since that time, several experiments have been dedicated to the detection of high energy cosmic rays. Even today, the only way to study particles at energies higher than 10^{17} eV is in UHECR experiments. Its study allows us to establish characteristic properties of propagation in the intergalactic medium, identify the sources capable of providing this energy, and delve into theoretical models about the processes linked to acceleration mechanisms, provide information on high-energy hadronic interactions and elucidate the nature of the primary particles. After more than 100 years, it still remains one of the most fascinating mysteries in astrophysics. This chapter is dedicated to show the technical details of our current understanding of cosmic-ray phenomenon.

2.1 Cosmic-ray spectrum

The term cosmic rays these days is used to describe essentially high-energy charged particles¹ that travel through the Universe at very nearly the speed of light. They play an important role in astrophysics on several scales, with cosmic-ray energies computable on Earth ranging from 10^6 eV up to 10^{20} eV (from about 1 particle/ m^2/s at low energies, to 1 particle/ $\text{km}^2/\text{century}$ at the highest energies). It is remarkable that the highest energies observed are several orders of magnitude higher than the energies currently accessible in human-made particle accelerators. In Fig. 2.1, the energy spectrum, i.e. the flux rate at Earth as a function of energy is shown, which is a powerful tool to understand cosmic rays

¹Actually, neutral particle such as photons, neutrinos and neutrons are also cosmic rays. In this thesis, however, we will use the term “cosmic rays” only to refer to charged particles.

(sources, acceleration mechanism) through the different features.

Generally the flux can be described by a power law, and can be roughly expressed as:

$$J(E) = \frac{dN}{dEdAd\Omega dt} \propto E^{-\gamma} \text{m}^{-2} \text{s}^{-1} \text{sr}^{-1} \text{eV}^{-1}, \quad (2.1)$$

In this expression, γ is the spectral index and dN is the number of particles arriving at Earth per energy dE , surface dA , time dt and solid angle $d\Omega$. At the lowest energies, the flux of cosmic rays is high enough to be directly observed at high altitudes or in space, allowing for precise composition measurements. In these regions the spectrum is well described by a power law of spectral index of $\gamma \sim 2.7$. At higher energies, cosmic rays have to be detected indirectly, through ground detectors, studying cascades of secondary particles generated in the atmosphere. We are now mainly interested in higher energies, $E \geq 10^{18}$ eV. From these energies on, the flux becomes too low that even the ground detectors find it difficult to accumulate sufficient statistics to measure the spectrum. The changes in the spectral index γ in different energy regions of the spectrum are thought to be due to changes in composition, propagation or sources. Fig 2.1 shows some features in the spectrum, which are interpreted in terms of astrophysical phenomena [12]. The energy spectrum exhibits three main features described below:

Knee

At around $E \sim 4 \times 10^{15}$ eV, the spectrum first steepens, and shows a change in the spectral slope from $\gamma \sim 2.6$ to ~ 3.1 . This is assumed to be the maximum acceleration energy galactic sources can reach for protons [13]. After that, a second knee is seen around $\sim 10^{17}$ eV, and the power law becomes steeper. Composition studies have suggested a transition from a lighter to a heavier nucleus. Therefore, the second knee could indicate the maximum energy of the flux of the heavy galactic component [14, 15].

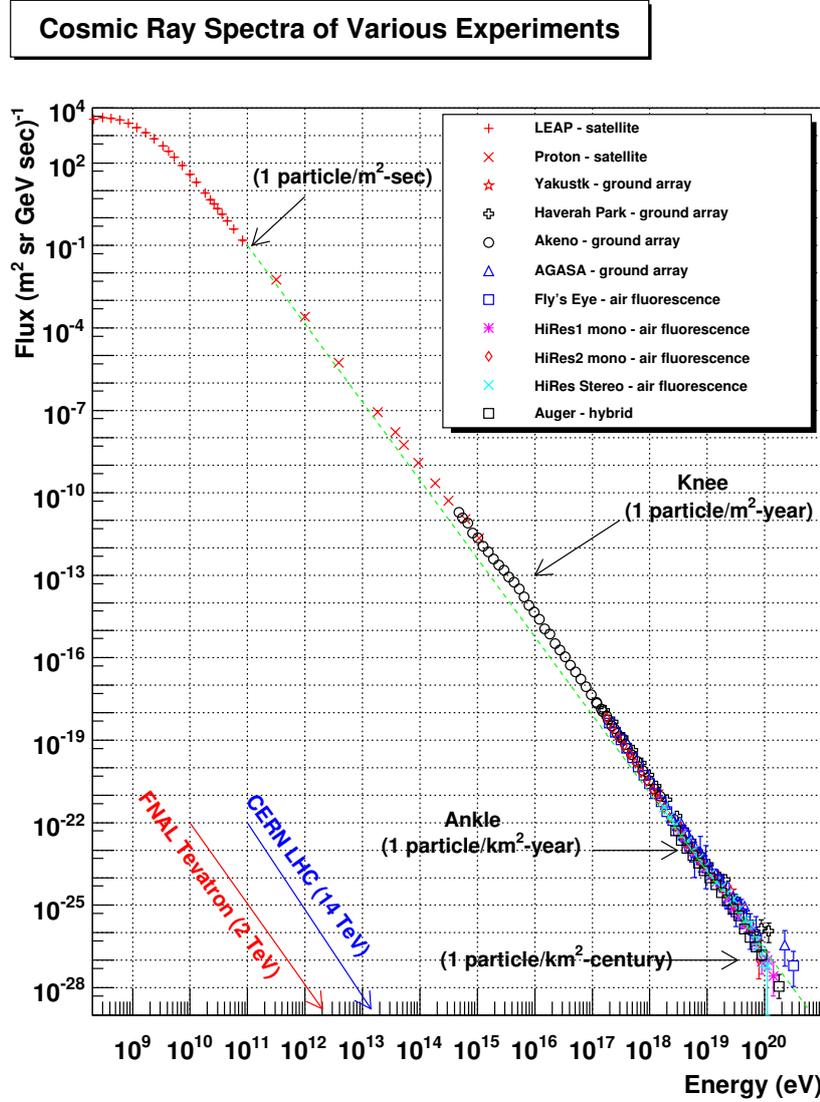


Figure 2.1: The all particle cosmic-ray energy spectrum measured by various experiments. This is shown in order to highlight the changes of the spectral index in the spectrum. [11]

Ankle

A flattening of the spectrum is observed at $E \sim 4 \times 10^{18}$ eV where γ becomes smaller again $\gamma \sim 2.6$ (measured by Pierre Auger Observatory more details in section 3.3). At this point, an important question arises related to at which energy the Galactic-extragalactic transition takes place. The ankle can be interpreted as the transition from Galactic to

extragalactic origin of the cosmic rays [16]. However, it has also been proposed that the ankle emerges due to the onset of e^+e^- pair production of extra-galactic protons with the cosmic microwave background, in the so-called dip model [17], or by photo-disintegration of heavy nuclei near their sources [18].

Suppression

Finally, an interesting region is the end of the spectrum, at energies above $E \sim 4 \times 10^{19}$ eV, where a suppression in the flux of cosmic rays has been observed by the Pierre Auger Observatory (more details in section 3.3). This suppression has two possible explanations: the theoretically predicted GZK cutoff ² (a prediction of a flux suppression due to resonant pion-photoproduction of UHECRs nuclei with photonbackgrounds, described in more details in section 2.3) or as due to the extragalactic sources reaching a limit to their acceleration capabilities (maximum rigidity). The phenomenological reason for this suppression is still on debate and unclear. To distinguish the two scenarios it is necessary to study the mass composition of elements that arrive at Earth.

2.2 Acceleration Mechanisms

When studying ultra-high energy cosmic rays, the big question involved is about the acceleration mechanisms that are able to boost the particle to such a high energy. Cosmic rays of low energies ($E < 10^{17}$ eV) can be accelerated by what we call the “Fermi mechanism”, proposed by Enrico Fermi in 1949 [19]. The astrophysical processes responsible for the acceleration of the particles that make up the UHECRs ($E > 10^{18}$ eV) are still unknown, and what exists today are theoretical models that seek to describe them. Historically, there are two classes of acceleration processes that we can consider: the first one is called bottom-up model, which is currently the most accepted model, and the second one is the top down model, which was discarded according to data obtained by the Pierre Auger Collaboration.

²GZK - referred to the famous Greisen, Zatsepin and Kuzmin cutoff.

The top-down model was proposed to explain the energy spectrum measured by the AGASA experiment, which was not compatible with the flux suppression for energies higher than $10^{19.5}$ eV [20]. The absence of the GZK suppression could be explained by a continuous distribution of sources originating particles of very high energy close to the Earth. This suggests that the particles for these models would originate from remnants of the early Universe with energies greater than 10^{21} eV [21]. In this scenario, UHECRs are the decay product of more energetic exotic particles, such as topological defects from the early Universe, then generating a cosmic ray which traveled to Earth. One of the main constraints in such theories is the explanation as to how these particles, which were created in the early Universe, would still be abundant today. The reason that discarded this model is that, in theory, there would be a high flux of photons and neutrinos, not observed by the Auger Observatory [22, 23]. Thus, with the results described by the Pierre Auger Observatory, the Top-Down models lost relevance as candidates for sources.

The bottom-up models, on the other hand, propose that UHECRs are the result of standard particles or nuclei, such as protons, being incrementally accelerated up to the highest energies. Most of them are stochastic models. In 1949, Enrico Fermi introduced the idea and proposed that charged particles gradually gain energy through multiple interactions on massive clouds with magnetic fields [19]. These clouds move randomly and depending on the direction the particle enters the cloud, it can lose or gain energy. The probability of a head-on collision give as more frequent result an energy gain. Therefore, it is possible to find that the average energy gain is $\langle \Delta E \rangle / E \propto \beta^2 = (v/c)^2$. This mechanism is called the second-order Fermi acceleration. Shock fronts are assumed to occur in several astrophysical environments, such as supernova remnants, jets of active galactic nuclei, gamma ray bursts in starburst galaxies, or in accretion shocks around massive galaxy cluster. As it is a second-order process, this energy gain model does not have a good efficiency to reach very high energies and could not describe the density of cosmic rays considering realistic acceleration times.

To describe another acceleration process, the model was modified in 1950, being also known as diffusive shock acceleration [24], in which charged particles are accelerated

through multiple crossings of astrophysical shocks, gaining some energy at every crossing. It can be shown that at every crossing the particle gains $\langle \Delta E \rangle / E \propto \beta = (v/c)$, named first-order Fermi mechanism. The spectrum expected in the first-order Fermi mechanism, in the test particle limit, for non-relativistic shock acceleration, is a power law with a fixed index $\gamma \sim 2$ and $\gamma \sim 2.2$ to 2.3 for relativistic shock acceleration [25]. For this, an acceleration site needs to preserve the particles inside the magnetic field region.

2.3 Propagation of Ultra-High Energy Cosmic Rays

After the primary particle leaves the acceleration sites, it propagates through the Universe, and may possibly reach Earth. Along this path, the primary suffers interactions with matter from the intergalactic medium (IGM) or interstellar medium (ISM) or radiation fields, such as the cosmic microwave background (CMB) and are deflected in magnetic fields according to their charge. Therefore, when we study the propagation of UHECRs, we need to consider their energy losses by several processes.

2.3.1 Radiation field interactions

Even the largest voids of our Universe are filled with photons of different wavelength, which can interact with cosmic rays. At the highest energies, the dominant radiation field is the CMB. Outside our Galaxy, space is also filled with photons emitted by stars, called the extragalactic background light (EBL). This radiation field consists mostly of the light emitted by star formation processes and AGNs which is still travelling through the Universe at wavelengths longer than it was emitted, due to redshift energy losses, absorption by dust. The EBL includes the ultraviolet (UV), optical and infrared bands (IR) of the electromagnetic spectrum. The number density of this radiation is much smaller than the CMB but still plays a roll for cosmic rays since it limits the propagation of UHECRs. While cosmic rays travel through the Universe, they interact with the cosmic microwave background and with the infrared photon backgrounds.

Cosmic microwave background radiation is a type of electromagnetic radiation that fills the Universe, being an important source of information about the early Universe. In the early stages of the Universe, radiation and hot plasma were dominant, filling the cosmos with a uniform mist that became opaque at high redshifts. Successively, the Universe became colder by expansion, and when it cooled down enough, stable atoms began to form. These atoms were no longer able to absorb thermal photons, and consequently, the Universe became transparent. The CMB photons produced at this stage propagated throughout the Universe with their energy decreasing with the expansion. As well, infrared photon background is extragalactic in origin and isotropic on large scales. This is because the infrared photon background is described by a spectrum formed from the evolution and luminosity from sources, along with dust and cosmological formation processes of stars and galaxies [26]. As the formation of cosmic infrared radiation is directly linked to the formation of stars in the Universe, limits of this radiation can be used to provide links in the history of the formation and evolution of galaxies.

The interaction of cosmic rays with the CMB is named after Greisen, Zatsepin and Kuzmin (GZK) [27, 28], and involves the excitation of a ultra-high energy proton at the Delta resonance that quickly decays to a proton and a neutral pion, or a neutron and a positive pion as:

$$p + \gamma_{CMB} \longrightarrow \Delta^+ \longrightarrow p + \pi^0 \tag{2.2}$$

$$p + \gamma_{CMB} \longrightarrow \Delta^+ \longrightarrow n + \pi^+, \tag{2.3}$$

where γ_{CMB} is a photon from the CMB and Δ^+ is the unstable delta baryon. The photo-pion production with the CMB depends on the energy per nucleon so that it also applies to nuclei. However, before nuclei reach this energy, they undergo photo-disassociation with the CMB and infrared background photons. This is a process in which some nucleons are ejected from the nucleus.

For protons with energy increasing above $10^{19.7}$ eV, the attenuation length³ quickly shortens from 110 Mpc due to the onset of photo-pion production with the CMB. The result of this process is that there should be a cut-off of the cosmic-ray spectrum at an energy of about $10^{19.7}$ eV. This predicted end to the cosmic-ray spectrum is called the GZK-limit. It also has profound consequences for the origin of the cosmic rays, and it would not be possible to observe protons with energies above the GZK limit coming from distant sources. Due to their attenuation, their sources should most likely be located relatively close to the Earth, as is shown in Figure 2.2.

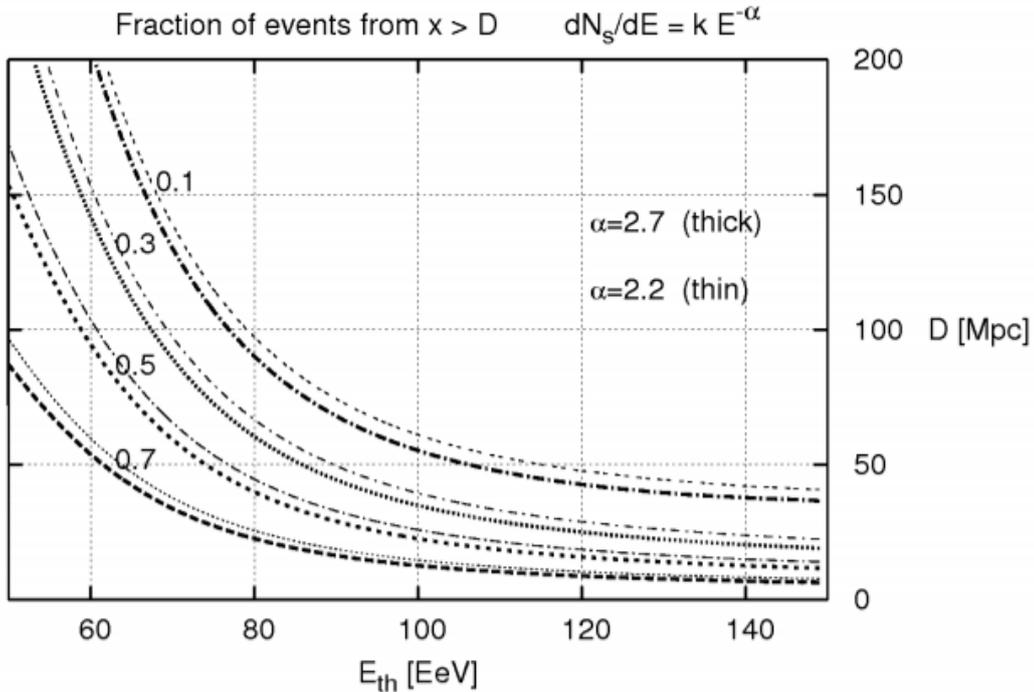


Figure 2.2: The fraction of cosmic rays that have been created with an energy above E_{th} and still have an energy above E_{th} after travelling a distance D from their source. The sources are assumed to be isotropically distributed and two assumptions of the spectral index α of the source are shown. The graph shows the attenuation of protons [29].

Photo-disintegration processes have threshold energies that vary with the mass of the nucleus.

³Attenuation length is defined as the distance that the particle travels in average before losing $1/e$ of its energy.

2.3.2 Energy loss processes

- Electron pair production. This process can occur when photons with energies above ~ 1.02 MeV interact with charged particles producing electron-positron pairs, $N + \gamma \rightarrow N + e^+ + e^-$, with N the UHECR nucleus, γ the background photon, e^+ the positron and e^- the electron (this is known as the Bethe-Heitler process). This interaction has a very short mean free path but leads to a very small fractional energy loss $\sim 0.1\%$ at the threshold. For UHE protons, pair-production is the dominant energy loss interaction for $E \leq 50$ EeV.
- Energy loss due to the expansion of the Universe. Our Universe is very well described by the Λ CDM model [30], in which the Universe is assumed to be homogeneous and isotropic, in accordance with the observations at length scales > 100 Mpc, and has been expanding from an initial singularity (the Big Bang) about 13.8 billion years ago. Protons lose energy predominantly due to the adiabatic expansion of the Universe, at a rate given by

$$-\frac{1}{E} \frac{dE}{dt} = H(z) = H_0[\Omega_M(1+z)^3 + \Omega_\Lambda]^{1/2} \quad (2.4)$$

if a flat Universe is assumed. Here, Ω_M and Ω_Λ are the dimensionless density parameters denoting the “matter density” and “dark energy density” of the Universe. At $z = 0$, the loss length to this process for protons is given by ~ 4000 Mpc. The propagation processes are not the same for protons compared to heavier nuclei as shown in Fig. 2.3.

2.3.3 Magnetic fields deflection

Besides the attenuation in these photon fields, UHECRs can be composed of charged elementary particles or nuclei. Their trajectories can be modified by the Galactic magnetic

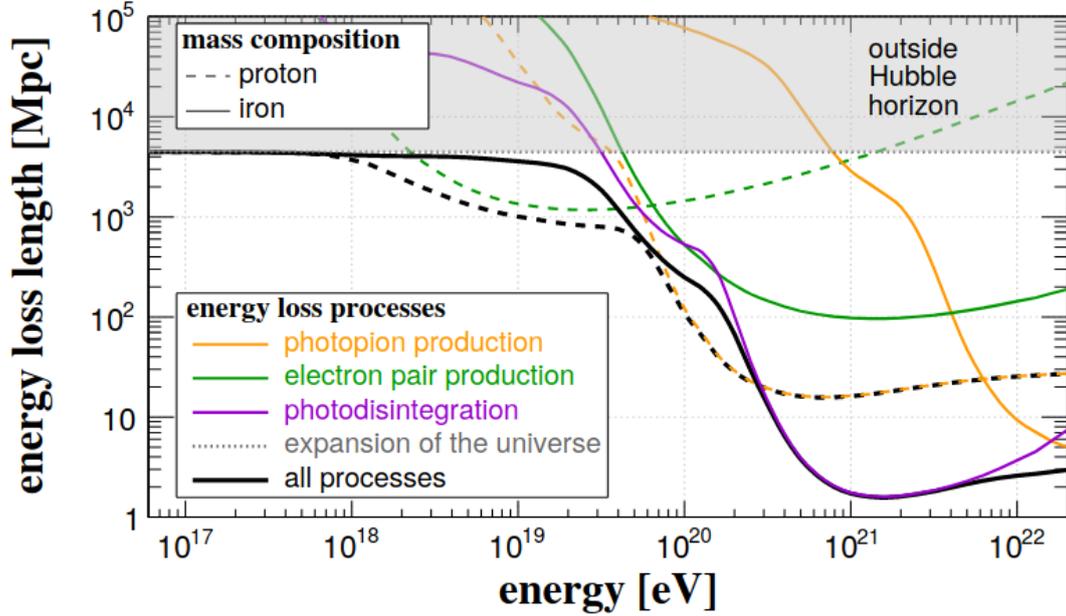


Figure 2.3: Energy loss lengths as a function of the energy for different interactions of UHECR: photo-pion production (orange), electron pair production (green), photodisintegration (purple), adiabatic expansion of the Universe (gray) and total (black). [31]

field (GMF) in the interstellar medium and extragalactic magnetic fields (EGMF) filling the space between galaxies, during their propagation through the Universe. Therefore, the arrival direction of a cosmic ray on Earth is not necessarily the same as the direction to its source in the sky, and the magnitude of this deflection depends on the rigidity of the cosmic ray. These particles are deflected by the magnetic fields with a gyroradius of:

$$r_g = \frac{p}{|q|B_{\perp}} \simeq \frac{E/c}{ZeB_{\perp}} \quad (2.5)$$

where q is a charged particle moving with velocity \vec{v} , and momentum \vec{p} and B_{\perp} denotes the magnetic field component perpendicular to the path of the cosmic ray. Thus, the circular radius induced by the magnetic field is the Larmor radius. Nuclei with higher charge are more affected since the deflection depends on the rigidity $R = E/Z$ of the cosmic ray.

It is well established that the Milky Way generates a Galaxy-wide magnetic field and this field tends to trap charged cosmic rays within. The Galactic magnetic field has a com-

ponent with strength of the order of $\sim 1 \times 10^{-6}$ G [32]. The dominant component is the regular one, with a coherence length of the order of 1 kpc, which can significantly deflect even EeV protons. Currently, the most used GMF model is the JF12 model [32], which describes a magnetic field following the spiral structure of our Galaxy and an additional irregular field which acts as a random component.

As concerns the extragalactic magnetic field, it is poorly understood. It is known that outside the Milky Way the magnetic field uncertainties are significantly larger, and estimations are characterised by a mean field strength B and a correlation length l_c of the field. The best way to study the expected deflection using EGMF models is through magneto-hydrodynamic simulations [33]. When a particle with charge Ze and energy E propagates through this field over a distance d that is much larger than l_c , its deflection is given by [34]

$$\theta(E, d) = 0.8^\circ Z \left(\frac{E}{10^{20}\text{eV}} \right)^{-1} \left(\frac{d}{10\text{Mpc}} \right)^{1/2} \left(\frac{l_c}{1\text{Mpc}} \right)^{1/2} \left(\frac{B}{10^{-9}\text{G}} \right). \quad (2.6)$$

Cosmological structure formation models predict values of $10^{-9} \leq B \leq 10^{-8}$ G along filaments [35]. For regions inside galaxy clusters, the average field strength can be larger. For an average electron density of 10^{-3} cm^{-3} and length scale of about 1 Mpc, the expected field strength is of the order of 1 nG, which agrees with observed rotation measure dispersions of $\sim 100 \text{ rad m}^{-2}$ [36].

2.4 Candidate sources

One of the goals of the Pierre Auger Observatory is to identify the sources of ultra-high-energy cosmic rays. However, after seventeen years of operation of the observatory, no point sources of UHECRs have been discovered so far. Compared to heavier elements, lighter elements suffer smaller deflection by magnetic fields. Due to the GZK limit, particles with energies exceeding the GZK energy are limited to about ~ 100 Mpc. This limit reduces the number of possible sources and favors the directional correlations of source locations and arrival directions of cosmic rays. It was realized that there could only be a few potential

sources that are likely to be able to accelerate cosmic rays up to the highest energies. In 1984, Hillas [37] calculated the maximum acceleration energy for a source with a given size L and magnetic field strength B for a cosmic ray with charge number Z :

$$\left(\frac{E}{10^{15}\text{eV}} \right) < \frac{Z\beta}{2} \left(\frac{L}{\text{pc}} \right) \left(\frac{B}{\mu\text{G}} \right), \quad (2.7)$$

where β is the velocity of the shock in units of the speed of light, c . Based on this criteria, the Hillas plot for ultra-high-energy cosmic rays is constructed in Figure 2.4. This Hillas plot can be used to rule out systems which do not have the combination of field strength and size necessary to reach the observed energies. As well, in the diagram, several possible sources for the observed cosmic rays are admissible in the Hillas plot.

The two solid diagonal lines show the source conditions where the minimum product of BL required to accelerate protons (red line) or iron nuclei (blue line) to 10^{20} eV for a fast shock where β is assumed to be 1. On the left of the diagram, classes of object of the lines do not satisfy the Hillas criterion. As shown with the dashed diagonal lines, the required product of BL is higher for slower shocks ($\beta = 0.01$) [38]. The Hillas diagram reveals that in normal galaxies, supernovae (SNe), and stars that drive massive magnetized winds, the confinement condition is not satisfied. The most powerful accelerators are active galactic nuclei (AGN) and gamma-ray bursts (GRB).

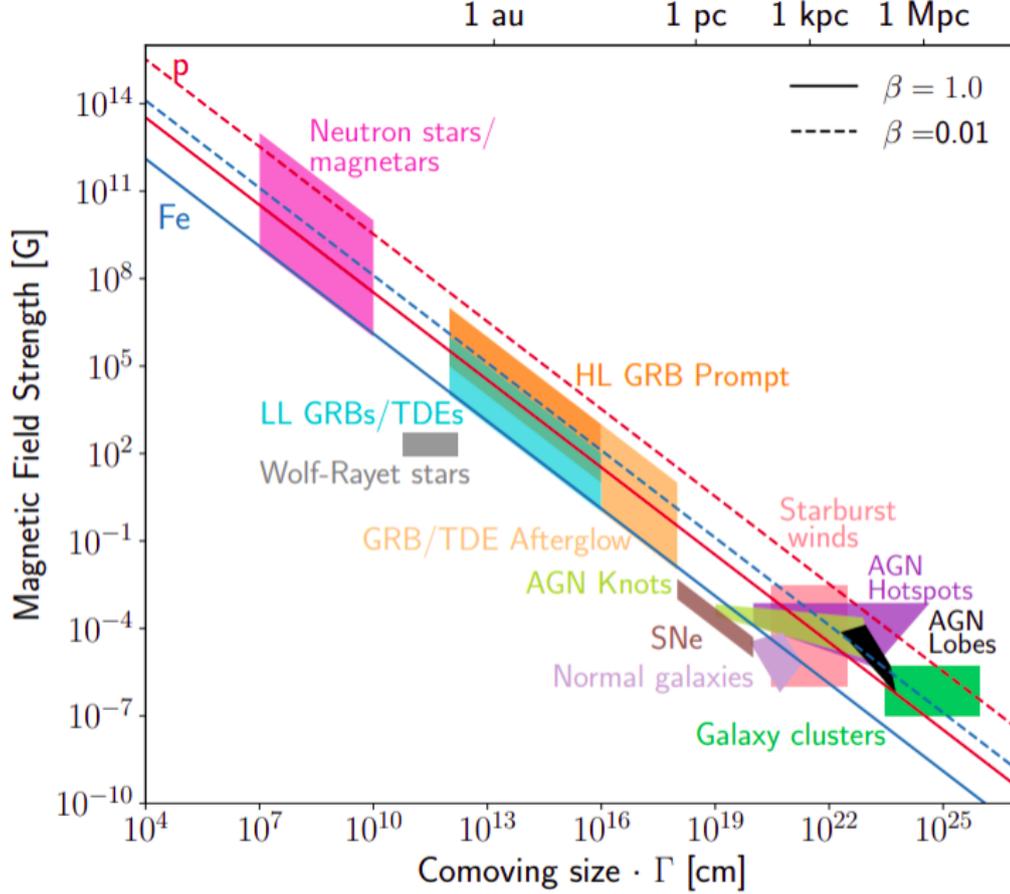


Figure 2.4: A Hillas plot classifying potential sources for UHECRs according to their size of the region in which the acceleration takes place and the magnetic field strength. It shows the maximum energy of a charged particle that can be confined in a region of size L , with a magnetic field B . For a certain particle type, it depends on the velocity of the particle, $\beta = v/c$. Setting $\beta = 1$, the source properties for confining proton (red) and iron (blue) nuclei accelerated to 10^{20} eV are indicated by the diagonal lines. Possible sources must be in the upper right region above the line. [38]

2.5 Chemical composition

To explore the origin of cosmic rays at ultra-high energies and their paths through the Universe requires the knowledge of the chemical composition. Unfortunately, the nature of the primary particles is not a directly accessible quantity and is obtained through indirect measurements. It is possible that cosmic rays observed on Earth do not have the same composition as those originating from the source. As mentioned, the initial composition

of particles leaving the source (primary particles) can be modified on the way to Earth (secondary particles). Currently, we have a consensus that index changes of the power-law spectrum are due to changes in the chemical composition and sources of cosmic rays. For energies below 10^{14} eV, the elemental abundance in the flux of cosmic rays can be measured directly using detectors above the atmosphere. This is not possible at higher energies due to the steep decrease of the flux with growing energy. Therefore, it is extremely important to study their energy spectrum according to the composition of the observed flux to provide theoretical models that explain how UHECRs are generated. Moreover, observables sensitive to energy and composition of the primary cosmic rays must be carefully analyzed.

The composition of UHECRs is determined from the nature of the particle that collided at the top of the atmosphere and its shower. A parameter used for this study is the depth of the maximum development of the extensive air shower (X_{\max}). As it differs for different primary species, it is indicated to discriminate particles by their primary mass. However, fluctuations in the early state of the air-shower development prevent an event-by-event discrimination. Thus, the mass composition has to be inferred from the distribution of the average depth of shower maximum. The relationship of the mean X_{\max} with respect to energy E is [39].

$$\langle X_{\max} \rangle = X_0 + D \log \left(\frac{E}{E_0 A} \right), \quad (2.8)$$

with X_0 being the mean depth for proton showers at energy E_0 and D being the elongation rate, which refers to the change of X_{\max} with $\log E$. It is expected that heavy nuclei interact and develop earlier in the atmosphere than protons [40] since the interaction probability increases with the number of nucleons. Therefore, showers produced by iron primaries reach their maximum development (X_{\max}) higher in the atmosphere. Additionally, for the same reason, fluctuations in the first interaction depth are significantly smaller for iron primaries than for protons. Therefore, the X_{\max} distribution becomes narrower for iron primaries by a factor of three with respect to that of protons [41]. Taking this into account, X_{\max} is a helpful observable to study the mass composition of primary cosmic rays.

Chapter 3

The Pierre Auger Observatory

In this chapter, we present a description, main features and current state of the Pierre Auger Observatory, which is the world's largest observatory for high energy cosmic rays at present. The Pierre Auger Observatory is located in the Pampa Amarilla, in the province of Mendoza, Argentina, designed to measure the energy spectrum, the chemical composition and the arrival directions distribution of cosmic rays with energies above 10^{17} eV. Its construction began in 2002, and the first data collection was started in 2004, way before its construction was completed in 2008. The observatory has a large area of detection that covers 3000 km^2 in the Southern hemisphere. To detect UHECR, the Pierre Auger Observatory was designed as a hybrid detector, which is a combination of an array of particle detectors and a set of fluorescence telescopes. The hybrid design is important in order to have complementary detection that employs two independent methods to study high energies cosmic rays above 10^{18} eV, providing cross-checks.

3.1 The detectors of the Pierre Auger Observatory

The geographical location of the observatory and its detectors at the site Pierre Auger Observatory are shown in [3.1](#). At present, different types of detectors are in use at the Pierre Auger observatory, which are discussed here briefly.

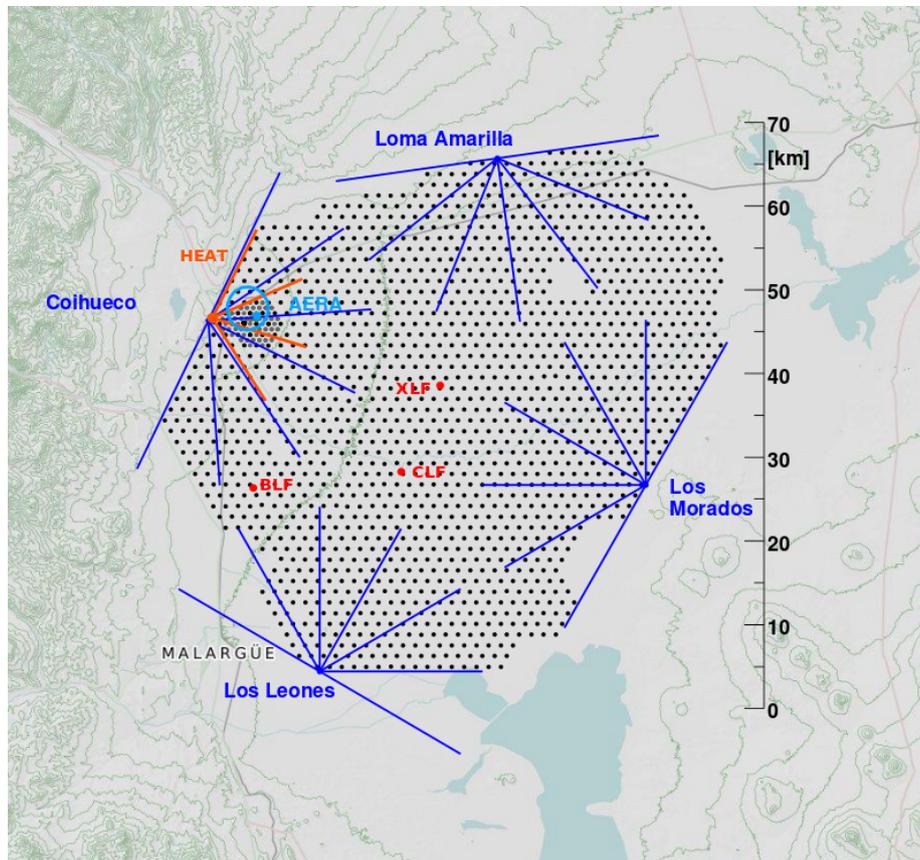


Figure 3.1: Schematic view showing the positions of the detectors at the site of the Pierre Auger Observatory. The black dots represent the 1600 surface detector stations installed along 3000 km^2 , and the blue lines show the field of view of the fluorescence telescopes, each site with the field of view of its six telescopes. The Pierre Auger Observatory extensions at the proximity of the Coihueco FD site include the HEAT high elevation fluorescence telescopes for the hybrid detection of lower energy atmospheric showers developing over the densely placed surface detectors (the infill array) [42].

3.1.1 The Surface Detector

The surface detector array (SD) is composed of 1600 water Cherenkov detectors, arranged on a triangular grid with 1500 m separation between the neighbor stations, forming a network of 3000 km^2 . The Cherenkov detectors are represented by black dots in Figure 3.1, being fully efficient in the detection of primary particles with energies above $3 \times 10^{18} \text{ eV}$.

As of 2008, another smaller network, namely the infill array with an area of 23.5 km^2 ,

consisting of 61 SD stations spaced 750m, was added and configured to detect primary particles with energies down to 3×10^{17} eV.

Each SD station consists of a water tank (Fig 3.2) of 3.6 m diameter and 1.2 m height, filled with 12000 liters of a highly purified water, enclosed in a bag of Tyvek, a material allowing the efficient reflection of Cherenkov photons which are produced when charged particles pass through the water at a speed $v \geq c/n$, where n is the refractive index of the water.

The tanks are called water Cherenkov detector stations because the molecules of the medium emit Cherenkov light. The shower particles excite and polarize the medium and the molecules will lose energy by re-emitting radiation. The particle speed exceeds the speed of light in the water, and the Cherenkov effect occurs, producing light. The Cherenkov light is collected by three photomultipliers (PMTs) which are symmetrically distributed at a distance of 1.20 m from the center of the tank, and they look downwards through windows of clear polyethylene into the water. These PMTs then emit analog signals at the last dynode and at the anode, which are read by a front-end card, placed in a box above the tanks, and converted into digital signals by FADC (Fast Analog Digital Converter), with a sampling frequency of 40 MHz, i.e., one sample every 25 ns. The digitized signals are sent to the station trigger block. The detector trigger time, crucial for event reconstruction, is measured with a precision of ~ 8 ns using the Global Positioning System (GPS) units. The output signals which pass a certain trigger level are transmitted to the closest FD location by a radio antenna fixed on the detector surface, from where they are transferred via communication towers to the Central Data Acquisition System (CDAS) building in Malargüe for further processing (higher trigger levels).

The SD has a duty cycle of about 100%, so that it provides the vast majority of the events. However, unlike the fluorescence detector, it cannot directly observe an air shower development. The goal of the SD is to measure the lateral distribution function (LDF¹) of the shower which is useful to reconstruct the primary particle energy [43].

¹LDF is a function that describes the number of particles as a function of the distance to the shower core, defined as the point where the shower axis meets the ground.

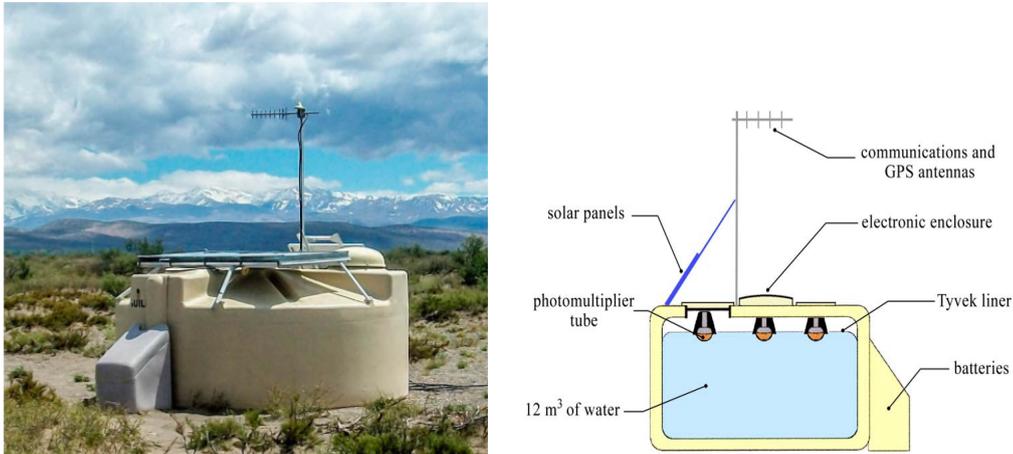


Figure 3.2: Parts of a detector tank. **Left:** Tank installed in the Pierre Auger Observatory. **Right:** Illustration of the main components of a tank and their respective positions. [42].

3.1.2 The Fluorescence Detector

The Fluorescence Detector (FD) consists of fluorescence telescopes located in four buildings, each building hosting six independent telescopes. The FD buildings are arranged at the borders of the surface detector array to monitor the atmosphere above the SD network. Each telescope includes the following components as described below.

- A circular diaphragm of 1.1 m radius, on which are placed a filter selecting the wavelength band $\sim 300 - 400$ nm in the ultraviolet range to reduce the background noise of the sky, and a corrective lens increasing the light collection surface.
- A spherical mirror with a radius of 3.5 m; the lens and the mirror are designed following Schmidt optics to ensure sharp images and uniform illumination over a very large field of approximately $30^\circ \times 30^\circ$. It is therefore a very different optic from that of conventional telescopes, which allows with six telescopes to cover 180° in azimuth (with even a slight overlap of the fields of view of the different telescopes) and about 30° in height (at above the horizon).
- The focus, a camera composed of 440 hexagonal PMTs, each covering a portion of the sky $\sim 1.5^\circ$ in diameter, as shown in Figure 3.4.

The air-fluorescence light from nitrogen enters through a large window of the UV-passing filter and the Schmidt optics corrector ring. Each telescope has a mirror that overlooks a camera of 440 pixels and photomultiplier light sensors. These pixels digitize the light pulses every 100 ns and then trigger levels are applied hierarchically to detect and measure the amount of light emitted by the nitrogen molecules excited by the cosmic shower.

These buildings are on hills located at strategic points on the edges of the SD. Each building received a name equal to that of the hill where it is located such as Los Leones, Coihueco, Loma Amarilla and Los Morados. Figure 3.3 shows an example of FD site. The telescopes are protected in a clean climate controlled building. For example, a schematic of the fluorescence telescope is depicted in the right plot of Figure 3.4.



Figure 3.3: FD building at Los Leones during the day. Behind the building there is a communication tower. This photo was taken during daytime when shutters were opened because of maintenance. [44]

In the region of Pampa Amarilla, the atmosphere is very clear, and there is a

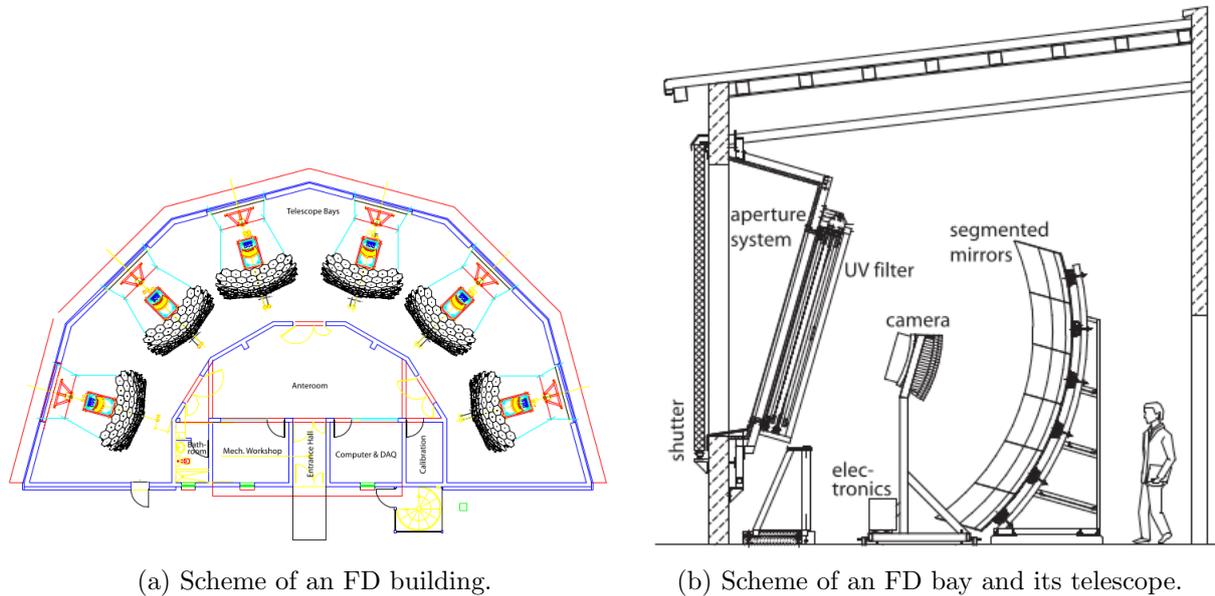


Figure 3.4: Schematic of the main components of an FD. [44].

very low light pollution from the artificial light produced in Malargüe. A total of 27 fluorescence telescopes measure the details of the longitudinal development of the extensive air showers with high accuracy, focusing on the ultraviolet radiation generated by the interaction between the shower particles and the nitrogen molecules along their trajectory. Through information compiled by the FD, it is possible to determine the energy of the primary particle and the depth in the atmosphere at which the production of particles of the shower reaches a maximum value, called X_{\max} , measured from the top of the atmosphere. As we will see later, the X_{\max} is an important parameter connected to mass composition. The FD can only take data in dark and clean nights. It also cannot take data in rainy or windy weather. Consequently, it has a duty cycle of about 15%.

HEAT

In addition, the site Coihueco houses three more telescopes in one station apart, a subset known as HEAT (High Elevation Auger Telescopes) is the FD equivalent to the SD-750 of the standard SD and extends the measurements for observing lower energies up

to 10^{17} eV. As the energy decreases, the amount of fluorescence light emitted is reduced and the shower has to land closer to a telescope in order to measure it. The field of view shrinks when approaching the telescope and as a result the longitudinal profile is detected only partially. By combining the three telescopes with the 180 m distant telescopes of Coihueco during the reconstruction, an additional virtual FD is built, called HeCo (virtual FD by combining Coihueco and HEAT).



Figure 3.5: HEAT, the low energy extension of the FD [44].

3.2 Hybrid reconstruction

In the following, we will describe the reconstruction of the shower properties from the data taken by the different detectors.

- **The FD Shower Reconstruction**

The reconstruction is based on FD data with additional timing information from the SD. First, to reconstruct the shower geometry of the shower, it is necessary to identify the shower detector plane (SDP), i.e., the plane that includes the location of FD telescope and the line of the shower axis, see Fig. 3.6. Once the SDP has been determined, the second step is the determination of the shower direction contained in the SDP, through the data

of arrival of the signal in each of the pixels (t_i) from the point S_i , as shown in Fig. 3.6. The shower axis can be characterized by two parameters: the perpendicular distance R_p from the camera to the track and the angle χ_0 that the track makes with the horizontal line in the SDP. Each pixel which observes the track has a pointing direction which makes an angle χ_i with the horizontal line. Let T_0 be the time when the shower front on the axis passes the point of closest approach R_p to the camera. The timing of the signal in the i -th pixel can then be expected to be

$$t(\chi_i) = T_0 + \frac{R_p}{c} \tan \frac{\chi_0 - \chi_i}{2}. \quad (3.1)$$

The free parameters T_0 , R_p and χ_0 are estimated through the fit, so that with the SDP the shower geometry is fully determined.

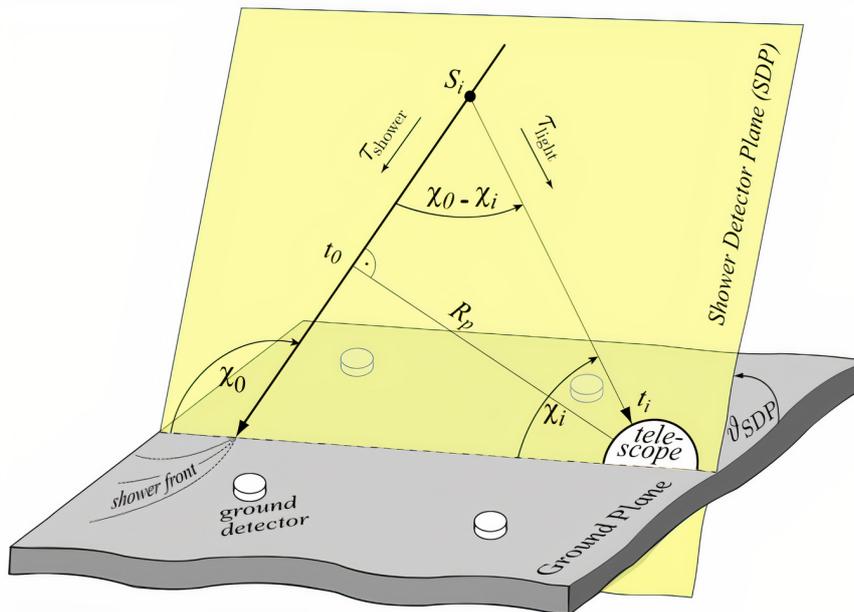


Figure 3.6: Geometry of an FD event within the shower detector plane [45].

- **The SD Shower Reconstruction**

The shower geometry (i.e., the shower size and its arrival direction) can be reconstructed from the relative arrival times of SD signals. Thus, the arrival direction of the shower is

approximated assuming a spherical shower front

$$c(t_i - t_0) = |\vec{x}_{sh} - \vec{x}_i|, \quad (3.2)$$

with c as the speed of light, t_i is the arrival time of the shower front at a station i with position \vec{x}_i . The parameters t_0 and \vec{x}_{sh} are the start time and position of the shower. A scheme of the geometry reconstruction with the ground array is shown in Fig. 3.7.

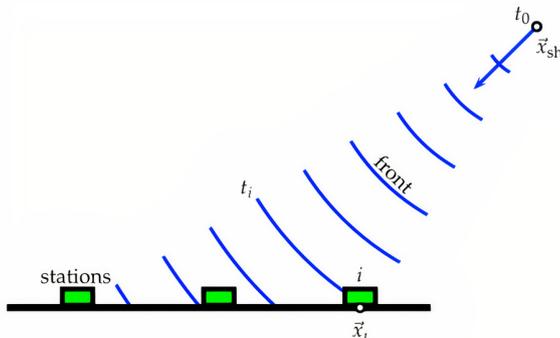


Figure 3.7: Scheme of the shower front estimation [44].

The impact point of the shower axis on the ground can be found by using the log likelihood method and simultaneously fitting the LDF²

$$S(r) = S(r_{opt}) \left(\frac{r}{r_{opt}} \right)^\beta \left(\frac{r + r_s}{r_{opt} + r_s} \right)^{\beta + \gamma}, \quad (3.3)$$

where S is the signal at the perpendicular distance to the shower axis r . The r_s is fixed to 700 m, and r_{opt} denotes the optimum distance based on the array spacing. The averaged slope parameters β and γ are fitted on selected data samples and then parameterized as a function of the zenith angle θ . Since the SD array in Pierre Auger Observatory has a spacing of 1500 m, the optimal distance is 1000 m, S_{1000} [46]. The parameter S_{1000} is an estimator of the shower size and the signal value obtained at an optimal distance from the core of the shower where fluctuations are minimal. The shower size S_{1000} has a zenith

²LDF is an empirical function, which describes the evolution of the signal with the distance r from the shower core. The distance r is the perpendicular distance between a station and the shower axis.

dependence, the value of $S(r_{opt})$ decreases with the zenith angle θ since showers which are more inclined traverse a larger amount of atmosphere before reaching the ground and are thus more attenuated. To correct for this effect, a Constant Intensity Cut (CIC) is used [47]. This results in an attenuation function that can be used to convert $S(1000)$ to a zenith independent shower size. The $S(r_{opt})$ is converted to $S(\theta_{ref})$. This method provides an attenuation curve $f_{CIC}(\theta)$ and can be used to convert S_{1000} by

$$S_{38} = \frac{S(1000)}{f_{CIC}(\theta)}. \quad (3.4)$$

The S_{38} parameter represents the signal intensity of an air shower with an zenith angle of 38° at 1000 m from the air shower axis. This variable is correlated with the energy measured by the Fluorescence Detector and is used to estimate the energies of the particles detected by the Surface Detectors.

3.3 Main results of the Pierre Auger Observatory

Since the beginning of its operation (~ 17 years), the Pierre Auger Observatory has already reported several results regarding the energy spectrum, chemical composition and arrival directions of the UHECRs. These results have greatly broadened our understanding of ultra-high energy cosmic rays. In the following, the most important and recent results are presented.

3.3.1 Energy spectrum

The UHECRs energy spectrum, which is the measurement of cosmic-ray flux as a function of energy, is important for probing cosmic-ray sources since their different structures can indicate changes in their origin or nature. The latest results obtained by combining all the data sets in Fig. 3.8, with energies from $10^{16.5}$ eV to $10^{20.2}$ eV, have achieved an unprecedented precision. The most recent measurements from the Pierre Auger Observatory can be seen in Fig. 3.8 with the high-statistics data from the SD. The spectrum is ob-

tained by using events with zenith angle $\theta < 60^\circ$ (vertical events³) and energy $E > 2.5 \times 10^{18}$ eV. Due to the behavior of the spectrum, a power law $J(E) \propto E^{-3}$ can be approximated, and the measured flux is multiplied by a factor E^3 in order to better identify the spectrum features. In Fig. 3.8 we can see different features. A hardening of the spectrum from

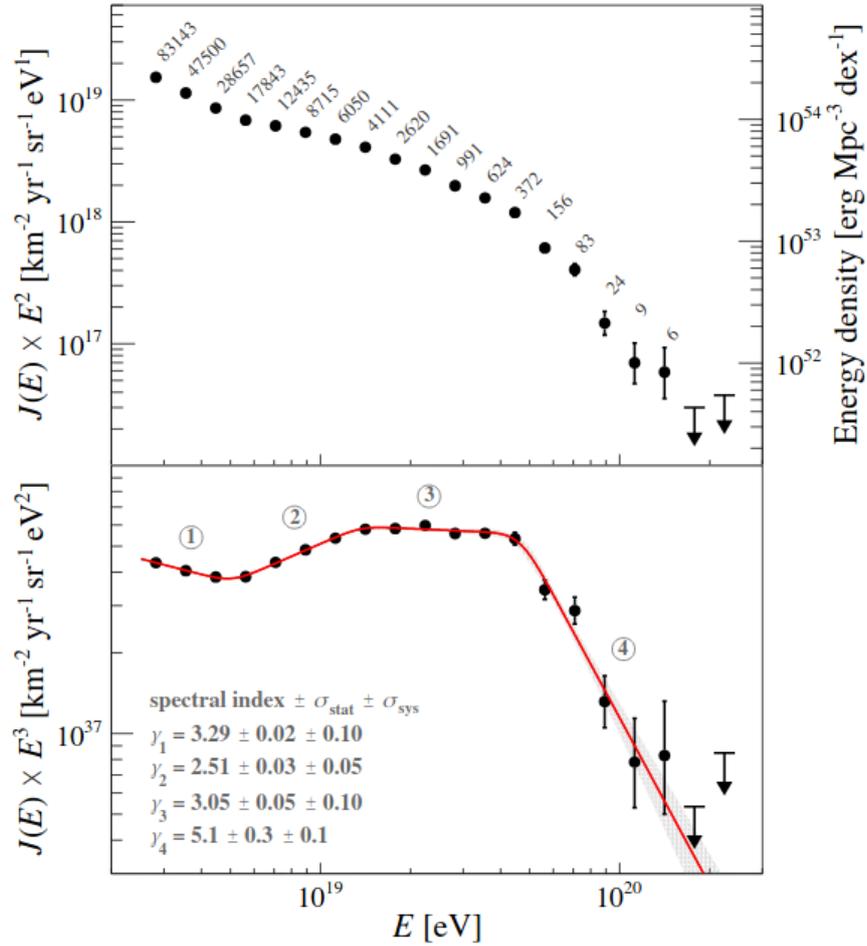


Figure 3.8: **Top:** Combined energy spectrum, adjusted to a theoretical flux scaled by E^2 . The number of detected events is indicated for each energy bin. The black dots represent the data, the error bars represent the statistical uncertainties.

Bottom: Energy spectrum scaled by E^3 fitted with a sequence of power laws (red line). Each number identify the energy intervals where the spectrum is described by a power law with a spectral index. The shaded band indicates the statistical uncertainty of the fit. Upper limits are at the 90% confidence level [12].

³Based on 215030 events detected between the 1st of January 2004 and the 31st of August 2018 for an exposure of $60400 \pm 1810 \text{ km}^2 \cdot \text{sr} \cdot \text{yr}$.

$\gamma_1 = 3.29 \pm 0.02 \pm 0.10$ to $\gamma_2 = 2.51 \pm 0.03 \pm 0.05$, at about $E_{12} = (5 \pm 0.1 \pm 0.8) \times 10^{18}$ eV is referred to as the ankle. The softening at $E_{23} = (13 \pm 1 \pm 2) \times 10^{18}$ eV, from γ_2 to $\gamma_3 = 3.05 \pm 0.05 \pm 0.10$, is a new feature observed by the Observatory; followed by an even stronger softening at $E_{34} = (46 \pm 3 \pm 6) \times 10^{18}$ eV with $\gamma_4 = 5.1 \pm 0.3 \pm 0.1$ which is commonly referred to as the cutoff. By considering the results of a combined fit of the energy and mass composition measured with the Auger Observatory [48], this new feature in the spectrum results from the interplay between the contributions of helium and carbon-nitrogen-oxygen components injected at the source to the flux at Earth, taking into account their distinct cutoff energies and photo-disintegration processes. Figure 3.9 shows the energy density obtained from this benchmark scenario in comparison with the Auger energy spectrum.

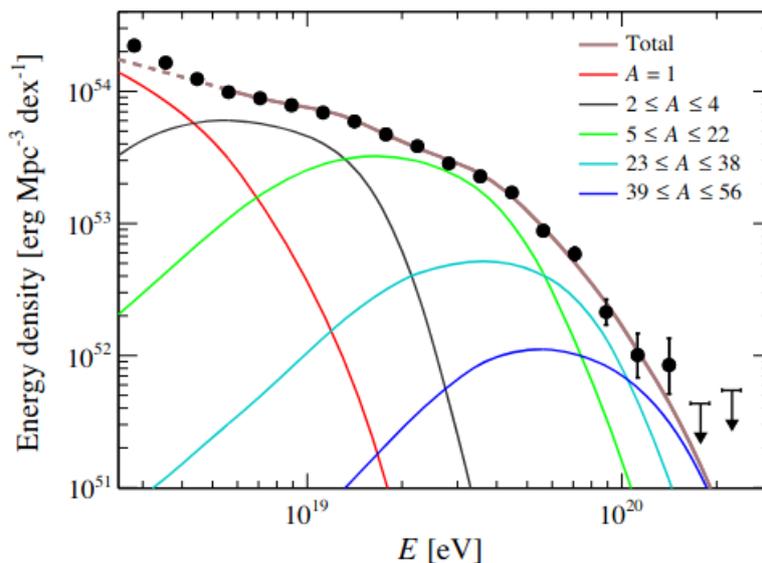


Figure 3.9: Figure of the energy density obtained in a benchmark scenario with five different mass components. The dashed curve shows the energy range that is not used in the fit and where an additional component is needed for describing the spectrum [12].

3.3.2 Mass composition

The chemical composition of the primary particle is determined mainly by the distribution of X_{\max} (measured in g/cm²) obtained with the FD. The first two moments of

this distribution, i.e., its mean value of the maximum shower depth, $\langle X_{\max} \rangle$, and measures of the fluctuation⁴ of X_{\max} , $\sigma(X_{\max})$, respectively, are shown in Figure 3.10, as well as the theoretical prediction of models for the X_{\max} energy dependence. Different models of hadronic interactions at higher energies were used, such as EPOS-LHC [49], Sibyll2.1 [50] and QGSJetII-04 [51].

The composition of the primary particle tends to be predominantly lighter up to $\sim 10^{18.3}$ eV, and a possible transition to heavier elements is observed at higher energies (above the ankle). We can probably exclude that the particles arriving at Earth at these energies are predominantly protons, and the composition tends towards intermediate-heavy masses [52]. The variance suggests a light or mixed composition at low energy. On the other hand, it is compatible with an intermediate-heavy composition at higher energies. It is important to mention that, due to the limited number of hybrid events, the X_{\max} distributions provided by the FD reach up to $\sim 10^{19.7}$ eV, making it impossible to infer information on the mass composition at higher energies. As mentioned, we have focused on the X_{\max} measurements obtained with the FD, but also the mean value of X_{\max} can be estimated with the SD [1]. With a larger number of events (larger statistics), it allows to expand the mass composition inferences up to 10^{20} eV, although increasing systematic uncertainties and providing only the $\langle X_{\max} \rangle$.

3.3.3 Arrival directions

Another important observable of UHECRs is the distribution of their arrival directions over the sky. As mentioned earlier, the cosmic rays are mostly charged particles, so the information of their trajectory is lost due to the deflection by magnetic fields, leading to random-walk like trajectories. The resistance of UHECRs to deflections will depend on the Energy E and charge of the particle and is expressed in term of the rigidity, $R = E/Z$. The Pierre Auger Collaboration has conducted several searches for non-uniformities and anisotropies in the arrival direction distributions of UHECRs using different techniques

⁴ σ is the standard deviation.

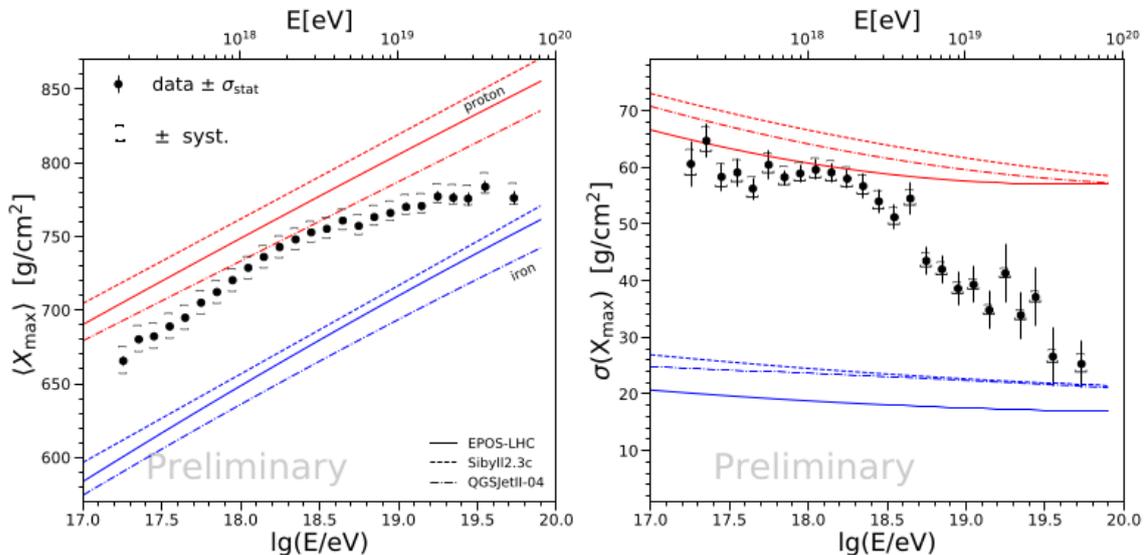


Figure 3.10: X_{\max} and $\sigma(X_{\max})$ as a function of energy. **Left:** values of the averages of X_{\max} as a function of energy. Also shown are the predictions of the models EPOS-LHC, Sibyll2.1 and QGSJetII-04. **Right:** The fluctuations of the X_{\max} measurements. In both figures, the last point represents the value corresponding to all events with energy $E > 10^{19.7}$ eV [52].

such as blind searches for overdensities, auto-correlation, correlation with astrophysical sources and with the directions of events detected by different experiments, as well as harmonic analysis.

The searches for possible anisotropies performed by the Pierre Auger Collaboration are best revealed using the technique of the harmonic analysis of the counting rate, both in right ascension and in azimuth angle, which is sensitive to non-uniformity in declination. Recently, the analysis of large scale anisotropies reported in [53] was updated using data collected until 2020, December 31. For energies above 8 EeV, we observe a significant dipolar amplitude of the first harmonic in right ascension, detected at 6.6σ of confidence level [54]. The amplitude of the 3D dipolar component that was determined at this energy bin is $d = 7.3\%$. The dipole direction $\sim 115^\circ$ away from the Galactic center, supporting the hypothesis that these particles are of an extragalactic origin. The result of the analysis is shown in the panel of figure 3.11.

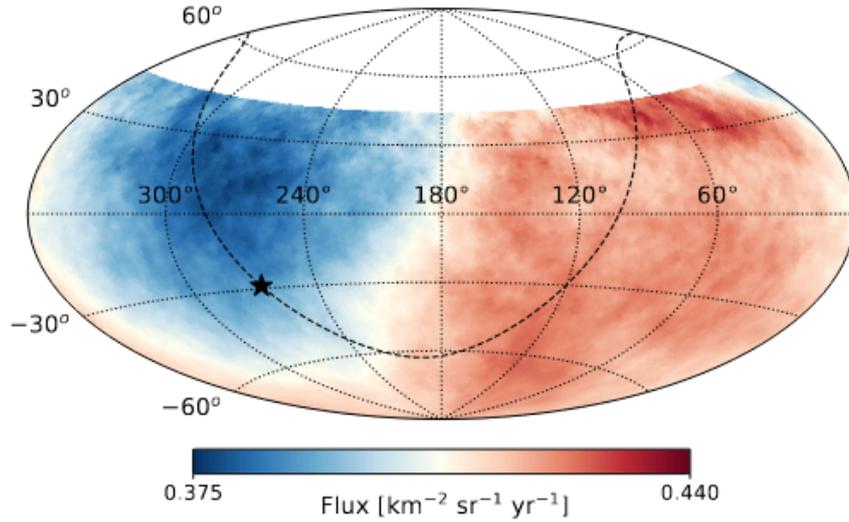


Figure 3.11: Sky map in equatorial coordinates. It shows the cosmic-ray flux with energies above 8 EeV in equatorial coordinates averaged with a top-hat window of 45° radius. The Galactic plane is represented by a dashed line, and the Galactic center is indicated by a star [54].

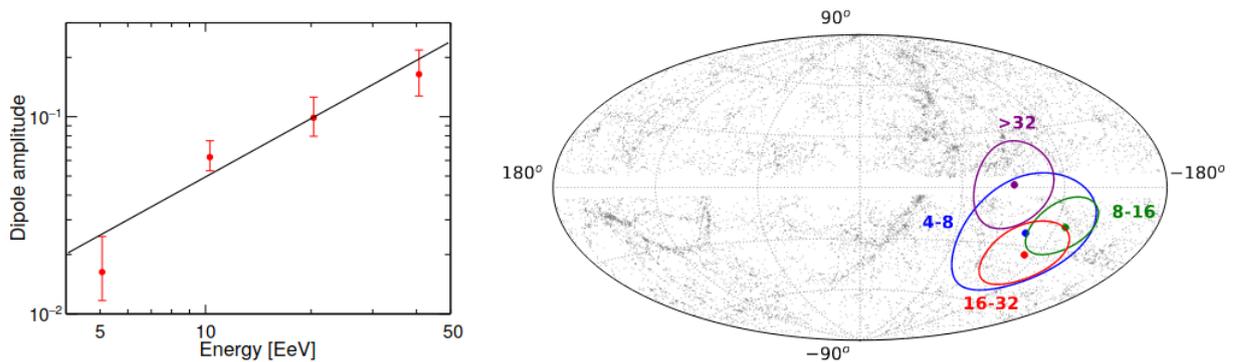


Figure 3.12: **Left:** Evolution of the dipole amplitude as a function of energy measured above 4 EeV. **Right:** The reconstructed dipole directions in different energy bins and corresponding 68% C. L. uncertainty in Galactic coordinates. The dots indicate the positions of 2MRS galaxies within 100 Mpc [54].

An update was performed in [54] by splitting the highest energy bin in: [4,8] EeV, [8,16] EeV, [16,32] EeV and $E \geq 32$ EeV. The reconstructed amplitude as a function of the energy, considering different energy bins is shown in the figure 3.12. The amplitude of the dipole increases with energy above 4 EeV, as expected from the predictions. Although there is no clear trend in the change of the dipole direction as a function of energy, they are consistent with an extragalactic origin in all bins.

Another interesting analysis reported in [4] also provides evidence for anisotropy in the arrival direction of UHECRs, which is an indication of anisotropy at intermediate scales for events with energies above 40 EeV. In this analysis, a comparison between the distributions of the sources of selected catalogs and the measured arrival directions of UHECRs was done. The flux from single sources was taken into account, based on the assumption that the UHECR flux is proportional to the non-thermal electromagnetic flux. The detected events by the Auger collaboration were analyzed in comparison with two classes of extragalactic sources: the nearby active galactic nuclei (AGNs) and starburst galaxies (SBGs) from the Fermi-LAT source catalogs.

In this study, the two extragalactic gamma-ray populations are composed of 17 blazars and radiogalaxies with γ -ray fluxes used as a proxy for the UHECR flux and selected 23 starburst galaxies with a radius flux greater than 0.3 Jy. The attenuation of the flux due to energy losses during the propagation of the cosmic ray up to the Earth was also taken into account.

The UHECR model is derived as the sum of an isotropic component and an anisotropic contribution from the sources, with anisotropic fraction and a smearing angle that takes into account the deflection of the cosmic-ray trajectory due to magnetic fields as free parameters of the model. Figure 3.13 shows the test statistic (TS) based on the likelihood ratio between model and isotropy as a function of the energy. In this analysis, a maximum likelihood analysis was performed to maximize the test statistics while varying two parameters: the search radius and the anisotropic fraction. The test statistic evolution is found for 39 EeV for the starburst galaxies and 60 EeV for the AGNs. The best-fit parameters for the

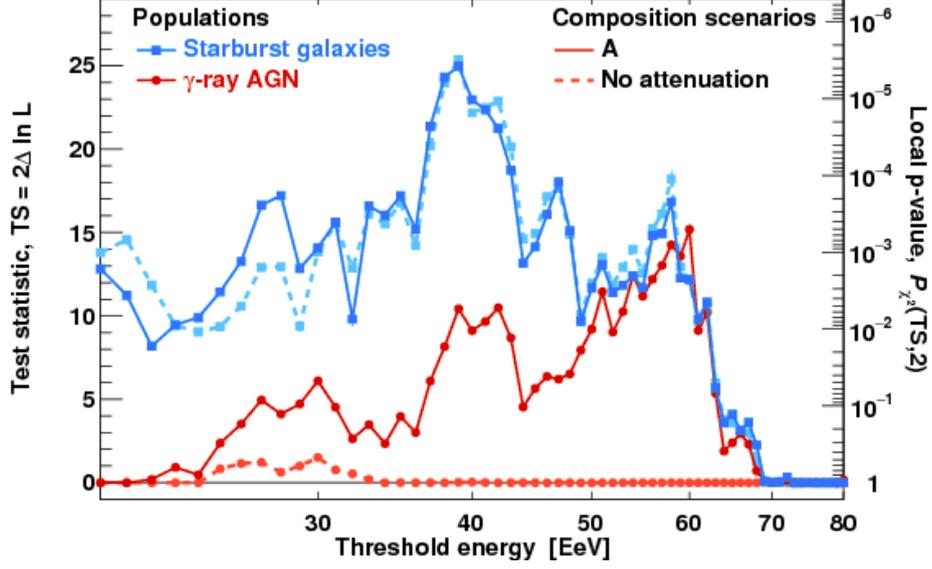


Figure 3.13: Scan of test statistic (TS) as a function of the energy threshold for SBGs and γ -AGNs with and without take into account the attenuation of the intensity due to energy losses [4].

starburst galaxy case are 10% of anisotropic fraction and a smearing angle of 13° , where this value gives the highest statistical significance. The TS obtained is 24.9 corresponding to a significance of $\sim 4.5\sigma$ in the correlation between arrival directions and positions of the SBGs. The corresponding values for the γ -ray AGN case are 7% and an angular scale of 7° with a TS of 15.2 ($\sim 2.7\sigma$). In particular, SBGs are interesting objects to study and provide an astrophysical interpretation to the Auger data due to the larger significance. Fig. 3.14 presents the observed excess map for the starburst galaxy population considering events with energies above 39 EeV. Additionally, the three strongest starburst galaxies namely M83, NGC4945, and NGC253 are indicated.

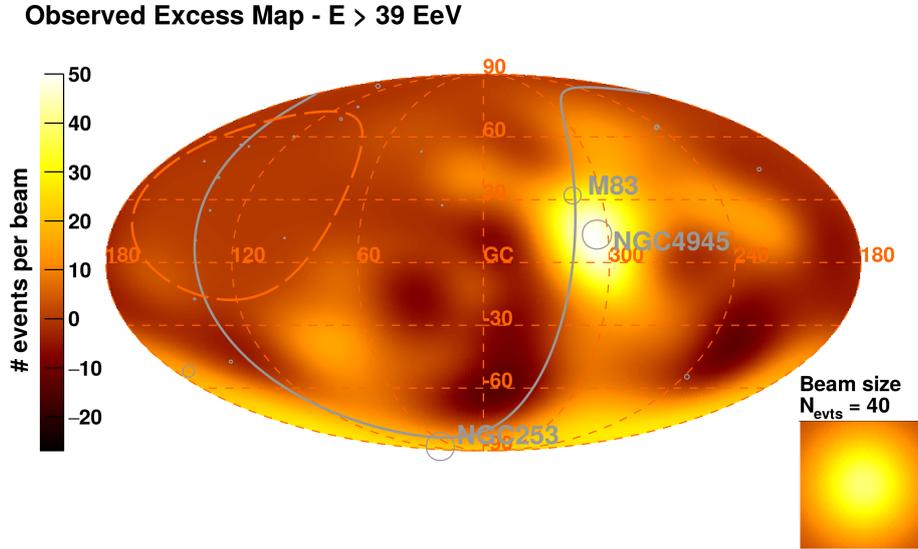


Figure 3.14: Sky maps in equatorial coordinates. Observed excess map for starburst galaxies. The Galactic plane is shown as a solid gray line. The orange dashed line delimits the field of view of the array [4].

3.3.4 Combined Fit results of spectral and chemical composition data measured by Pierre Auger Observatory

An attempt to simultaneously fit the Auger spectrum and X_{max} measurements for energies range 5×10^{18} eV up to highest energies at the Pierre Auger Observatory was published in 2017 [48]. This analysis was performed assuming a simple astrophysical scenario of UHECRs with identical sources uniformly distributed in a comoving volume, in which several nuclear components are injected at the sources with a power-law spectrum and with the maximal energy of the sources modeled with an exponential cut-off. In this analysis, five representative masses are considered at the sources: hydrogen (H^1), helium (He^2), nitrogen (N^{14}), silicon (Si^{28}) and iron (Fe^{56}). For each nuclear species A , the injected flux J_A is described by a power law of energy, modified by a broken exponential cut-off.

$$J_A(E_{\text{inj},i}) = \frac{dN_{\text{inj},i}}{dE} = \begin{cases} J_0 \cdot a_i \cdot \left(\frac{E_{\text{inj},i}}{E_0}\right)^{-\gamma}, & E_{\text{inj},i}/Z_i < R_{\text{cut}} \\ J_0 \cdot a_i \cdot \left(\frac{E_{\text{inj},i}}{E_0}\right)^{-\gamma} \exp\left(1 - \frac{E_{\text{inj},i}}{Z_i R_{\text{cut}}}\right), & E_{\text{inj},i}/Z_i \geq R_{\text{cut}} \end{cases}$$

where $E_{\text{inj},i}$ is the energy of the nucleus i at the source, J_0 is a normalization factor of the spectrum at energy $E_0 = 10^{18}$ eV⁵, a_i is the fraction of the injected element of mass i so that $\sum_i a_i = 1$, γ is the spectral index of the flux, Z_i is the atomic number of the nucleus i and R_{cut} is the maximum rigidity at the source. By considering a benchmark with PSB cross-sections, Gilmore EBL model and EPOS-LHC hadronic interaction model [48], the best fit obtained occurs for $\gamma = 0.96$, $R_{\text{cut}} = 10^{18.68}$ V. Figure 3.15 shows the deviance from D_{min} as a function of (γ, R_{cut}) . The best fit can be seen as part of a long valley, extending to lower values of γ and R_{cut} , approximately along the curve shown. Although it presents a much greater deviance, there is also another local minimum for $\gamma \approx 2$. The impact of the different cross-sections, EBL spectrum and hadronic interaction models on the fit was also studied. In general, as results, the minimum occurs for a hard spectral index and a low rigidity cutoff. More details about the combined fit analysis will be given in Chapter 8, where we present our independent implementation of the fit. The distribution of the energy spectra. mean and variance of the X_{max} corresponding to the best fit are shown in Figure 3.16.

⁵The energy at which the ejected mass fractions are defined is arbitrarily chosen to be lower than the estimated cutoff energy for protons.

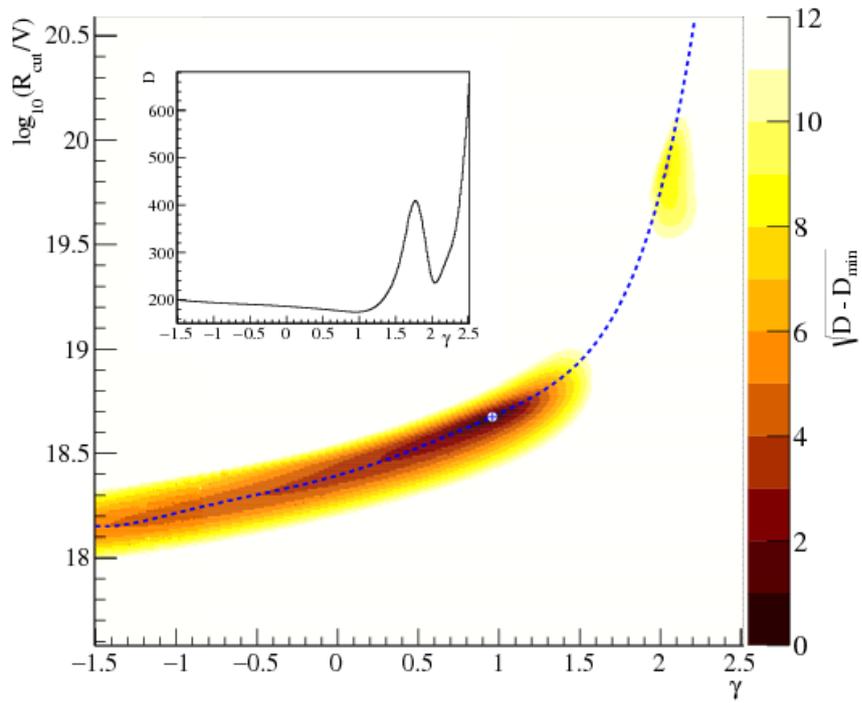


Figure 3.15: Deviance $\sqrt{D - D_{min}}$, as function of γ and $\log_{10}(R_{cut}/V)$. The color diagram indicates the confidence levels: 1σ , 2σ , etc. In detail, D values are shown along the dashed curve. From [48].

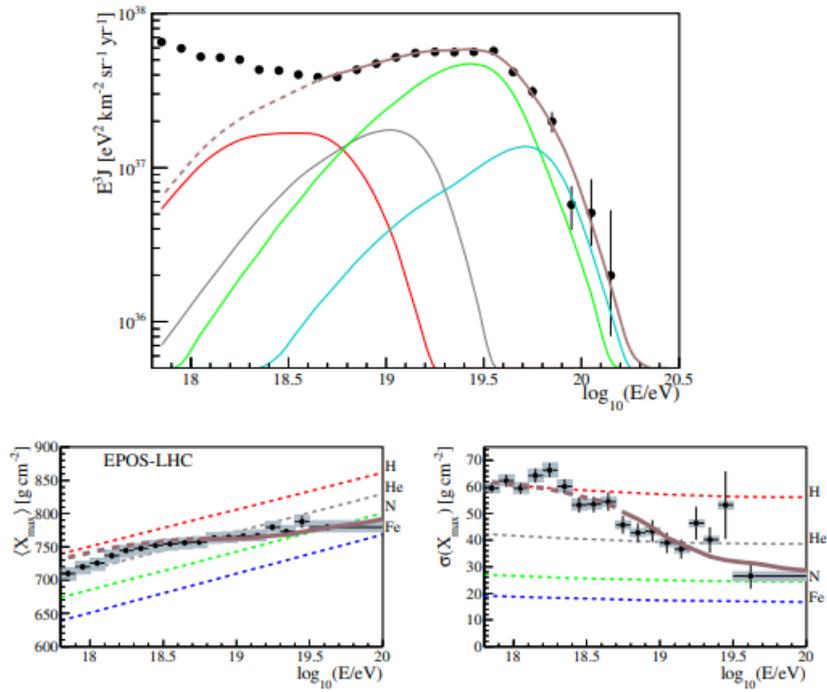


Figure 3.16: The Auger energy spectrum at the top of Earth’s atmosphere, fitted with a simulated energy spectrum following the model used in [48]. The simulated energy spectra (colored lines) are obtained by performing a combined fit of a simple astrophysical model to the energy spectrum and the values of X_{\max} . The all particle simulated spectrum (brown line) is the sum of the mass-restricted energy spectra: in red $A = 1$, in grey $2 \leq A \leq 4$, in green $5 \leq A \leq 22$, in cyan $23 \leq A \leq 38$ and in blue $A \geq 39$. From [48].

Chapter 4

Sensitivity of cross-correlation studies by using X_{\max} information measured by the Surface Detector of the Pierre Auger Observatory

The study of anisotropies through the cross-correlation between the arrival directions of UHECRs and the positions of astrophysical objects is an important element to reveal the sources of those cosmic particles. Since most cosmic rays are charged particles, they are deflected along their path through the Galactic and extragalactic magnetic fields. The limited knowledge about the magnitude of these fields and the chemical composition of the cosmic rays makes the identification of the sources of the ultra-energetic cosmic rays a very difficult task. Part of this difficulty comes from the fact that the information about the chemical composition of the primary cosmic ray is obtained statistically from the distribution of the atmospheric depth at which the maximum deposition of energy by the extensive air shower, X_{\max} , occurs. At the Pierre Auger Observatory, this information is measured by the fluorescence telescopes, which operate only on clear nights, without moonlight, corresponding to $\sim 13\%$ of the operating time of the experiment. The Pierre

Auger Collaboration recently has shown that it is possible to obtain the same information in an event-by-event basis by using deep neural networks and data from the surface detectors (which operate almost 100% of the time) of the Pierre Auger Observatory. This opens a very interesting window for anisotropy studies. In this chapter, we study the gain in detection power of the Landy-Szalay anisotropy cross-correlation estimator [2] when we select samples of the highest X_{max} values, corresponding to the lightest elements. As the magnetic deflection in the cosmic-ray trajectory is proportional to its electrical charge, the selection of such samples, composed by lighter nuclei, must present a higher correlation with the direction of the sources that produced them, being quite promising for detecting anisotropies in the cosmic-ray arrival directions and, possibly, unveil their origins.

4.1 Landy-Szalay cross-correlation estimator

The 2-point cross-correlation function is a method designed to measure the degree of similarities between the distributions in the sky of two sets of astrophysical objects. The Landy-Szalay [2] cross-correlation method is based on the estimator:

$$\omega(\theta) = \frac{N_{D_1 D_2}(\theta) - N_{D_1 R_2}(\theta) - N_{R_1 D_2}(\theta) + N_{R_1 R_2}(\theta)}{N_{R_1 R_2}(\theta)}, \quad (4.1)$$

where D_1 denotes the data set of interest (the arrival directions measured by the Pierre Auger Observatory, in our case), D_2 is a reference catalog of possible astrophysical sources and R_1 and R_2 are isotropic simulated sky maps with the same exposures of D_1 and D_2 , respectively. The notation $N_{X_1 X_2}(\theta)$ indicates the number of pairs between the two samples, X_1 and X_2 , as a function of the angular distance θ . For simplicity of notation, we will refer to $N_{X_1 X_2}(\theta)$ only by $N_{X_1 X_2}$.

It is interesting to note that this estimator fluctuates around zero if the data set D_1 or the catalog D_2 is isotropic. In this case, $D_1 = R_1$ (or $D_2 = R_2$) by definition, and Eq.

4.1 is

$$\omega(\theta) = \frac{\overbrace{N_{R_1 D_2} - N_{R_1 D_2}}^{\approx 0} - \overbrace{N_{D_1 R_2} + N_{D_1 R_2}}^{\approx 0}}{N_{R_1 R_2}} \approx 0. \quad (4.2)$$

For a better comprehension, we show in figure 4.1 the distributions of the number of pairs obtained from one specific simulation of an isotropic cosmic-ray sky containing 1,000 events (D_1) generated by taking into account the exposure of the Pierre Auger Observatory. To test the cross-correlation, we used the positions of starburst galaxies (D_2) [4]. The isotropic mock samples (R_1 and R_2) were generated as explained above with 10,000 directions sampled for each one. Figure 4.2 shows the resulting $\omega(\theta)$ obtained for this specific simulation. We can see that it fluctuates around zero as expected.

A possible choice of a global estimator for the cross-correlation between samples D_1 and D_2 that takes into account the statistical penalization for searches in different angular scales θ is given by the reduced χ^2_ν as

$$\chi^2_\nu = \frac{1}{N_{bins}} \sum_1^{N_{bins}} \left(\frac{\omega(\theta) - \overbrace{\langle \omega(\theta) \rangle_{iso}}^{=0}}{\sigma_{iso}(\theta)} \right)^2, \quad (4.3)$$

where $\langle \omega(\theta) \rangle_{iso}$ and σ_{iso} are the expected value and standard deviation of $\omega(\theta)$ obtained from isotropic skies. N_{bins} is the number of bins for searches in different angular scales θ . For simplicity, from now on, we will only refer to χ^2_ν as χ^2 .

The statistical test is done by comparing the χ^2 obtained for a given data set with the $\chi^2_{iso,99\%}$, defined such as $\int_{-\infty}^{\chi^2_{iso,99\%}} d\chi^2 p(\chi^2_{iso}) = 0.99$, where $p(\chi^2_{iso})$ is the reduced χ^2 probability density function. If χ^2 obtained from the data is greater than $\chi^2_{iso,99\%}$, the isotropy hypothesis is rejected.

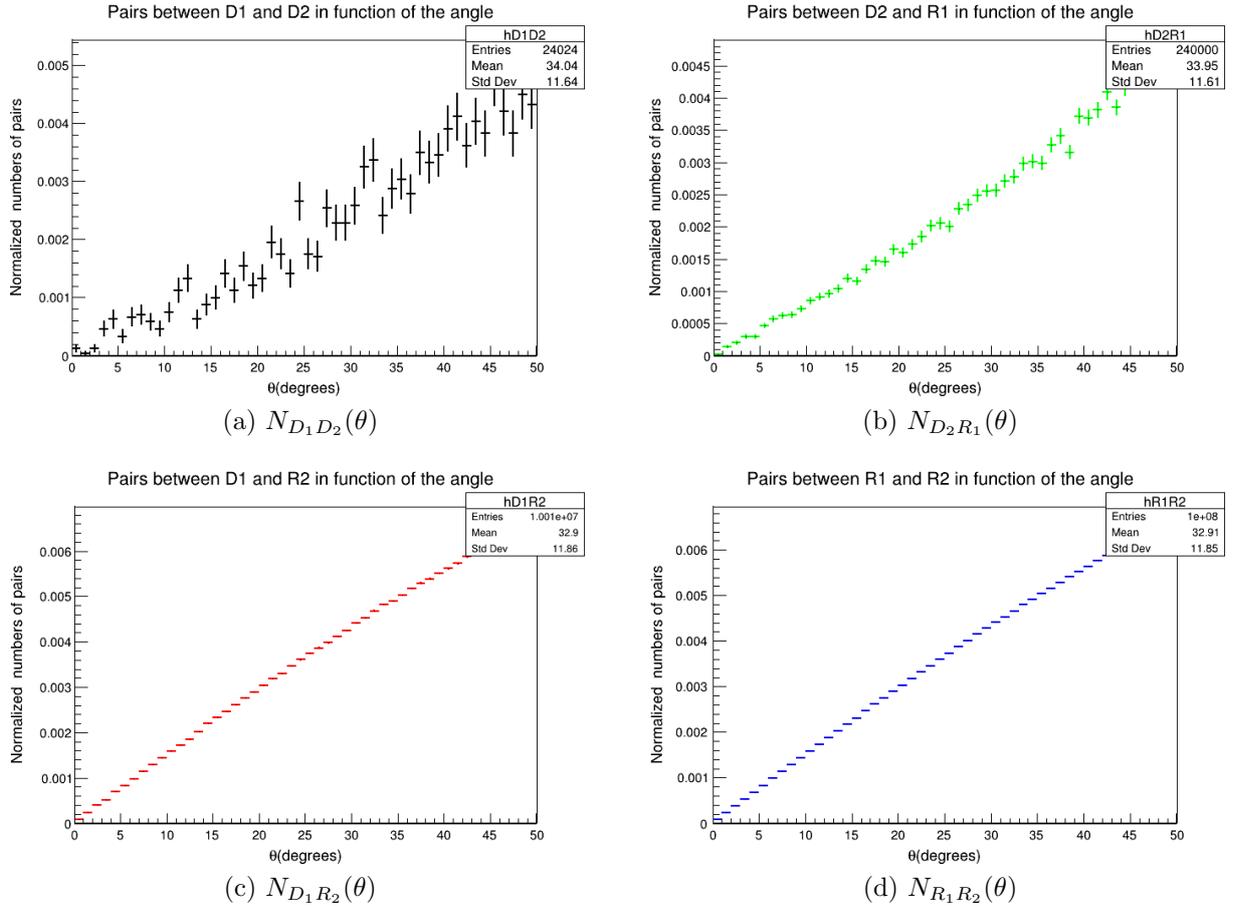


Figure 4.1: Distributions of the number of pairs obtained from one specific simulation of an isotropic cosmic ray sky containing 1,000 events (D_1) generated by taking into account the exposure of the Pierre Auger Observatory and the positions of starburst galaxies (D_2). R_1 and R_2 correspond to isotropic mock samples generated with exposures according D_1 and D_2 .

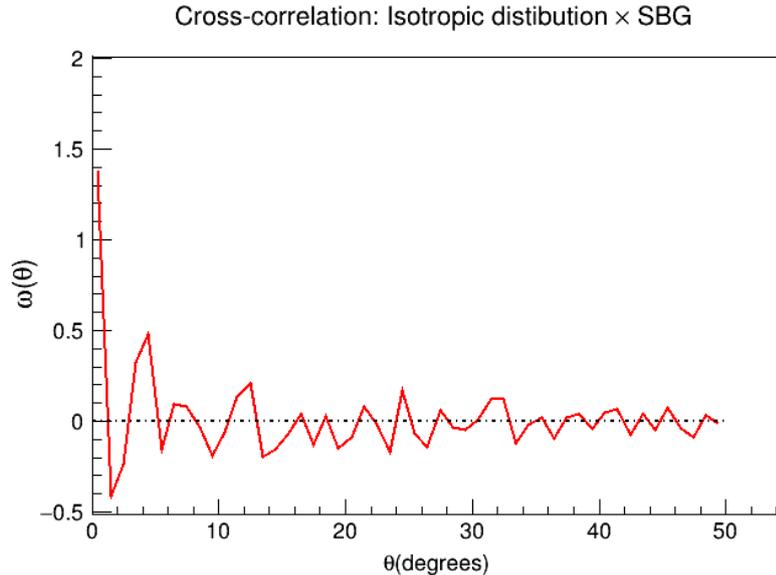


Figure 4.2: Example of Landy-Szalay cross-correlation estimator between a simulated isotropic dataset and positions of starburst galaxies. The red line is the value of the estimator $\omega(\theta)$ as a function of the angular scale θ obtained for the data. The black dotted line shows the expected value in the isotropic case.

4.2 Detection sensitivity

The goal of this section is to compute the gain in detection power when it is possible to use a subsample of events, selected by the highest X_{max} values, corresponding to the lightest nuclei. Since the magnetic deflection of the cosmic-ray trajectory is proportional to its electrical charge, we expect that such selection presents a higher correlation with the direction of the sources that produced them in comparison with the cross-correlation considering the whole sample.

Our toy model, inspired in [4], consists of a mix of events coming from the SBG listed in Table 4.1, containing the positions of the 23 starburst galaxies, in galactic coordinates and from the background. The events injected at the SBG are protons while the background is composed of an equal mix of nitrogen N and iron Fe nuclei. Particles are simulated with energies above 40 EeV following an energy spectrum $dN/dE \propto E^{-\gamma}$, with $\gamma = -5$, in accordance with Auger the energy spectrum [12, 55] in the considered energy range.

Table 4.1: The 23 selected starburst galaxies, where the positions are given in Galactic coordinates, and their flux weights, as used in [4], normalized in such a way that the sum of all are 100%.

SBGs	$[l]$	$[b]$	Distance D [Mpc]
NGC 253	97.4	-88	2.7
M82	141.4	40.6	3.6
NGC 4945	305.3	13.3	4
M83	314.6	32	4
IC 342	138.2	10.6	4
NGC 6946	95.7	11.7	5.9
NGC 2903	208.7	44.5	6.6
NGC 5055	106	74.3	7.8
NGC 3628	240.9	64.8	8.1
NGC 3627	242	64.4	8.1
NGC 4631	142.8	84.2	8.7
M51	104.9	68.6	10.3
NGC 891	140.4	-17.4	11
NGC 3556	148.3	56.3	11.4
NGC 660	141.6	-47.4	15
NGC 2146	135.7	24.9	16.3
NGC 3079	157.8	48.4	17.4
NGC 1068	172.1	-51.9	17.9
NGC 1365	238	-54.6	22.3
Arp 299	141.9	55.4	46
Arp 220	36.6	53	80
NGC 6240	20.7	27.3	105
Mkn 231	121.6	60.2	183

Energy losses are neglected during the protons propagation because the SBG are very close to our Galaxy. A more complete approach, taking into account the energy losses, will be treated in a future work. Besides, a smearing in the arrival direction of protons is performed by using a Von-Mises Fischer distribution with an angular radius corresponding to 13° , according to [4]. Nitrogen and iron nuclei directions are isotropically distributed in the sky. The entire sample is generated by taking into account the exposure of the Pierre Auger Observatory, and the depth of shower maximum X_{max} for each particle is generated according to the Gumbel parameterizations, considering the EPOS-LHC model [49].

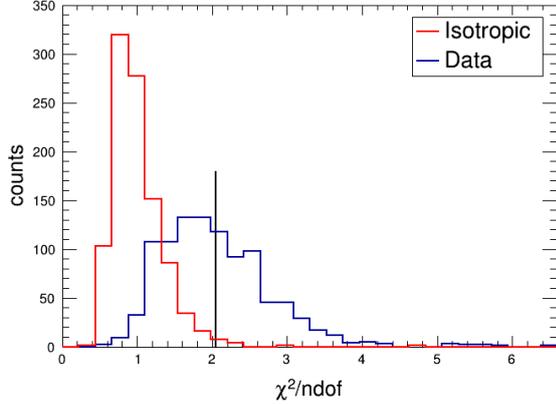
Each simulated cosmic-ray sky is composed of 1,000 events, a fraction f of which consists of protons coming from SBG and the remaining, $1 - f$, of nitrogen and iron isotropically distributed. The sensitivity of the cross-correlation estimator will be studied as a function of the fraction of protons $f_p = f$ and the fraction of selected events $f_{X_{max}}^{sel}$ corresponding to the highest X_{max} values. For each case, we performed 1,000 cosmic-ray sky simulations. The detection power, P , is computed by the fraction of skies that have a χ^2 larger than $\chi_{iso,99\%}^2$, i.e.,

$$P = \frac{N(\chi^2 > \chi_{iso,99\%}^2)}{N_{total}}, \quad (4.4)$$

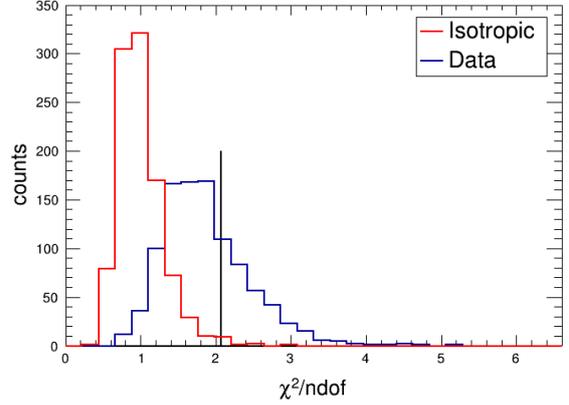
where $N(\chi^2 > \chi_{iso,99\%}^2)$ is the number of cosmic-ray skies with χ^2 larger than $\chi_{iso,99\%}^2$ and N_{total} is the total number of simulated skies for a given scenario.

Figure 4.3 shows, as illustration, the χ^2 distributions obtained for a scenario with a fraction of protons $f_p = 5\%$ and different fractions of selected events $f_{X_{max}}^{sel}$. Notice that the separation from the χ^2 isotropic distribution (reflected in the detection power) is larger when one considers a subset of the data selected by the highest values of X_{max} . Interestingly, the sensitivity of the cross-correlation estimator is larger when we select only $f_{X_{max}}^{sel} = 3\%$ than $f_{X_{max}}^{sel} = 5\%$. The detection power magnitude results from a competition between the size and the proton purity of the sample. Since there is a large overlap between the X_{max} distributions of protons and background (N + Fe) events (see figure 4.4), a more impure proton sample is obtained when one selects $f_{X_{max}}^{sel} = 5\%$ than $f_{X_{max}}^{sel} = 3\%$, justifying

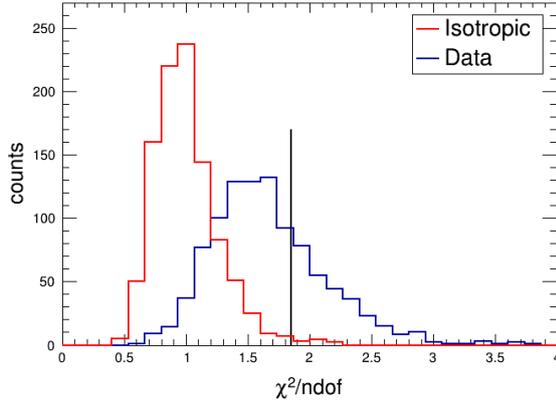
the largest detection power for $f_{X_{max}}^{sel} = 3\%$.



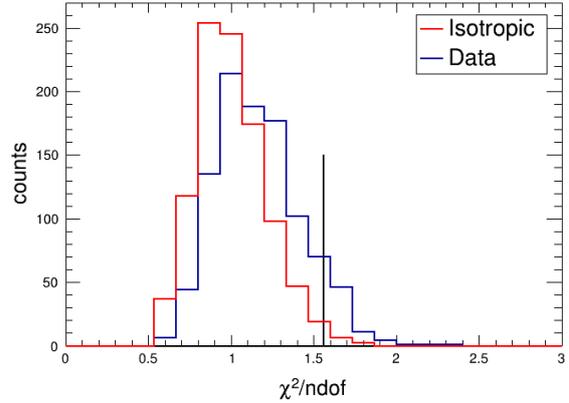
(a) Red: χ^2 isotropic distribution data. Blue: χ^2 distribution obtained by selecting a subsample consisting only of events with the highest 3% X_{max} values ($f_{X_{max}}^{sel} = 3\%$).



(b) Red: χ^2 isotropic distribution data. Blue: χ^2 distribution obtained by selecting a subsample consisting only of events with the highest 5% X_{max} values ($f_{X_{max}}^{sel} = 5\%$).



(c) Red: χ^2 isotropic distribution data. Blue: χ^2 distribution obtained by selecting a subsample consisting only of events with the highest 10% X_{max} values ($f_{X_{max}}^{sel} = 10\%$).



(d) Red: χ^2 isotropic distribution data. Blue: χ^2 distribution obtained by using the whole sample.

Figure 4.3: χ^2 distributions obtained for a scenario with fraction of proton $f_p = 5\%$ and different fraction of selected events $f_{X_{max}}^{sel}$ (blue curves). The distribution in red corresponds to isotropic sky maps. The black lines indicate the $\chi_{iso,99\%}$ threshold to accept (χ^2 smaller than $\chi_{iso,99\%}$) or reject the isotropy hypothesis (χ^2 larger than $\chi_{iso,99\%}$).

In Fig. 4.5, we show the resulting detection power as a function of the proton fraction f_p for different fractions of selected events $f_{X_{max}}^{sel}$. The net gain when we use a small subset containing the highest X_{max} depends of the proton fraction, ranging from a factor of ~ 2 to ~ 4 . It is noteworthy to mention that applications of analyses such as this

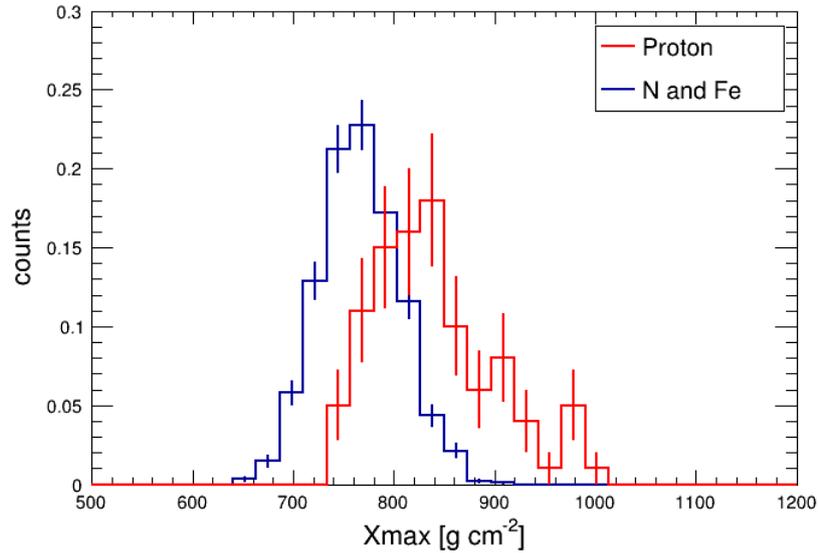


Figure 4.4: X_{max} distributions of protons and a mix of N and Fe considering the EPOS-LHC model.

one to real data sets are not allowed at this moment by the Pierre Auger Collaboration. This will only be possible after exhausting all tests of performances, including different estimators and parameters. This work is one of the several efforts on this task. More detailed studies, combining anisotropy studies in combination with X_{max}^{SD} measurements will continue to be explored in future works.

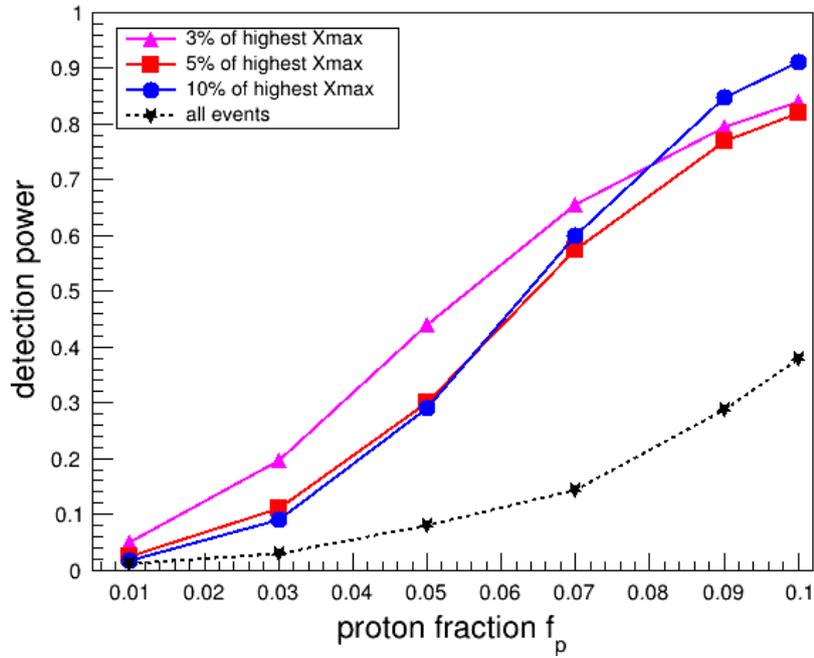


Figure 4.5: Detection power as a function of the proton fraction f_p for different fractions of selected events $f_{X_{max}}^{sel}$. The lines indicate the different values of the highest X_{max} considered.

Chapter 5

Study of the mass composition as a function of the arrival direction

Studies about composition maps [56, 57, 58] have attracted the attention of the Auger Collaboration. The Pierre Auger Collaboration tested the hypothesis of an anisotropy laying along the Galactic plane which depends on the mass of primary cosmic rays. Splitting the FD data into on and off-plane regions using the Galactic latitude of each event to form two distributions in X_{\max} , it was found that the distribution of X_{\max} from the on-plane region has a $9.1 \pm 1.6_{-2.2}^{+2.1}$ g/cm² shallower mean and a $5.9 \pm 2.1_{-2.5}^{+3.5}$ g/cm² narrower width than that of the off-plane region, considering an energy threshold of $10^{18.7}$ eV and a galactic latitude splitting at $|b| = 30^\circ$ [3].

The analysis described in this chapter was performed as a cross-check of these results, by computing the statistical significance of larger or smaller $\langle X_{\max} \rangle$ through the celestial sphere in a completely independent way. For this goal, we build mass composition maps, composed by the average of the primary mass composition evaluated within angular windows of a given size for each direction in the celestial sphere. The statistical significance of larger or smaller values of $\langle X_{\max} \rangle$ for each angular window are computed by considering probability maps in order to identify regions in the sky in which the primary mass composition is lighter or heavier.

The data set used in this analysis is composed of 2241 events collected by the fluorescence detector of the Pierre Auger Observatory with energy above $10^{18.7}$ eV. The data are corrected by field of view¹ cuts, X_{max} bias, and energy bias [59]. Since the average $\langle X_{max} \rangle$ increases as a function of the energy for the same primary particle, and we use a wide energy bin, the X_{max} of the events are normalized by the expected average X_{max} of iron nuclei according to the EPOS-LHC model [49] as

$$X_{max}^{norm}(E) = X_{max}(E) - \langle X_{max} \rangle(E)_{Fe,EPOS}. \quad (5.1)$$

The composition map is done in the following way. A top-hat window of 20° is scanned over all the pixels of the map. For each pixel, we compute the average $\langle X_{max} \rangle$ of the events inside the top-hat window. Figure 5.1 shows the obtained mean X_{max} map in g/cm^2 . In order to verify if a large (low) value of $\langle X_{max} \rangle$ in the map is because of

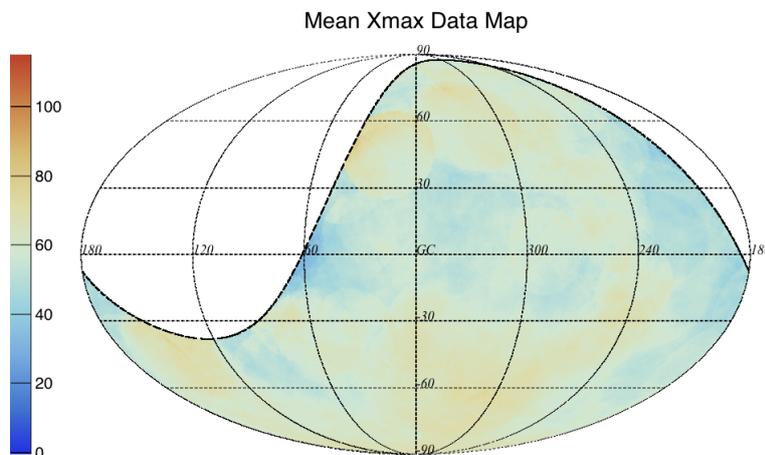


Figure 5.1: Mean X_{max} map for events with energy above $10^{18.7}$ eV using a top-hat angular window of 20° . The color scale unity is in g/cm^2 .

the presence of a lighter (heavier) element around a given direction or just a statistical fluctuation, we decided to build probability maps. This can be achieved by using the

¹A shower is reconstructed accurately only if its X_{max} is within the detector FoV (field of view). Shallow or deep events are more likely to have their X_{max} values outside the FoV and are excluded from the analysis.

shuffling (time-scrambling) technique, which consists of the average sampling of different realizations of event sets through random permutations of actual event times and arrival directions. The idea behind the method is that by scrambling the event time without changing the zenith and azimuth coordinates, one generates a different direction in the celestial sphere while keeping at the same time the same local acceptance. We generated 1000 mock isotropic composition maps by shuffling the arrival times and depths of the maximum of the events. The probability map is defined as

$$P_i = \frac{N(\langle X_{\max}^{\text{in-shuffling}} \rangle > \langle X_{\max}^{\text{in-data}} \rangle)}{N_{\text{shuffling}}}, \quad (5.2)$$

i.e., the probability for each pixel i is given by the number of mock isotropic skies with average $\langle X_{\max} \rangle$ for the same direction larger than the one from the data, $N(\langle X_{\max}^{\text{in-shuffling}} \rangle > \langle X_{\max}^{\text{in-data}} \rangle)$ divided by the total number of isotropized skies $N_{\text{shuffling}}$. The probability map obtained for $E > 10^{18.7}$ eV is shown in figure 5.2. Notice that the color axis corresponds to $1 - P$ so that values close to 0 indicate heavy composition (blue color) while close to 1 indicate light composition (red color). From the analysis of the figure, we can see that there is an excess of heavier particles correlated with the Galactic plane. Notice the similarity with Fig. 5.3 of [3], in which a composition map is presented by using a different test statistic and a top-hat filter of 30° of radius.

Cross-checks of the results by different groups are very important within a large collaboration such as the Pierre Auger Collaboration. Besides its importance, the know-how acquired in this task was very important to the study that will be presented in Chapter 6, where we assess the net gain obtained by combining directional intensity data with directional composition data in comparison to the standalone analysis of directional composition data only. Additionally, as a by-product of this cross-check, we performed a *service work*² related to the monitoring of the Fluorescence Detector. This was motivated by the fact that caution must be taken when the scrambling technique is applied to FD events. First,

²A *service work* is a very important task performed within large collaborations. It is intended to monitor the detectors or the quality of the data set and usually is not directly related to main physics analysis.

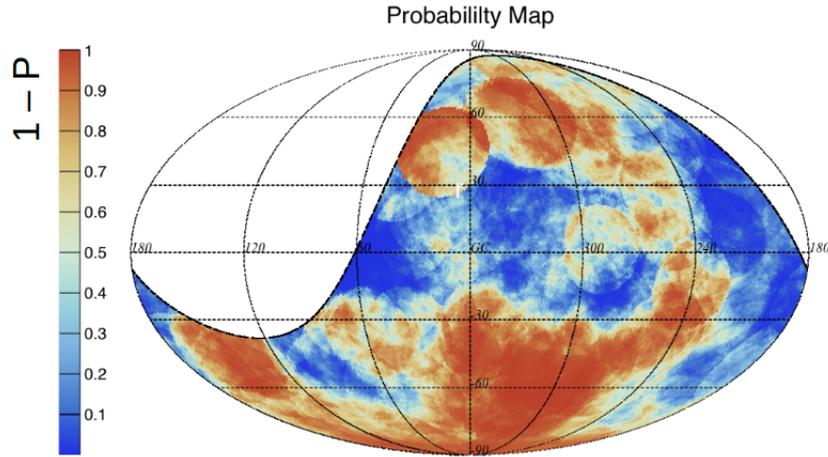


Figure 5.2: Probability composition sky map obtained for $E > 10^{18.7}$ eV.

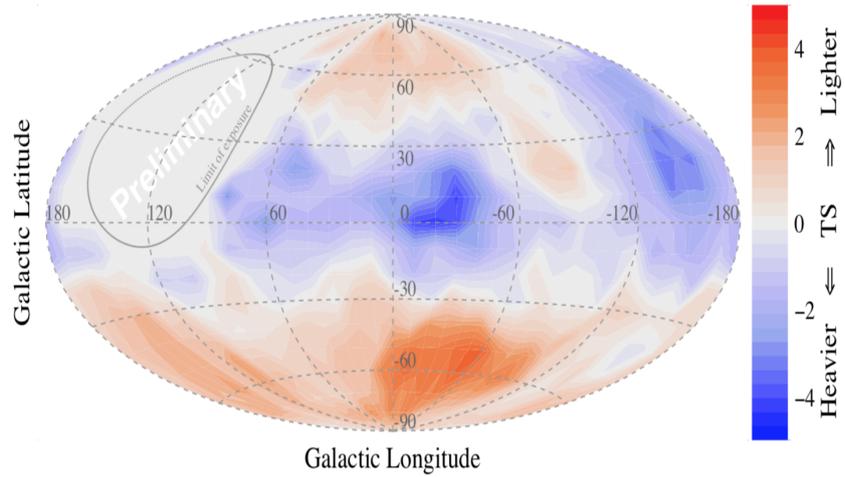


Figure 5.3: Sky map of cosmic ray composition for $E \geq 10^{18.7}$ eV obtained by using a different test statistic and a top-hat filter of 30° of radius [3].

the shuffling of the local times should be applied only within events detected by the same FD station since different telescopes cover different ranges of azimuth angles. Second, the shuffling technique implicitly assumes that the zenith angle distribution is independent of the time. Considering that there are different sources of background light during the period of the FD data taking, it is necessary to check if there is any bias correlated with the luminosity in the zenith angle distribution. This study is presented in the next section.

5.1 On the use of the shuffling technique for FD events

The shuffling technique implicitly assumes that the zenith angle distribution is time-independent. Since there are different sources of background light during the period of the Fluorescence Detector data taking, this assumption may be no longer valid. Motivated by the probability maps we have built before, we study in this section the applicability of the *shuffling* technique to events detected by the Fluorescence Detector. As the variance of the ADC³ counts is proportional to the background light, we first study the behavior of the mean zenith angle distribution as a function of the signal variance. Finally, the zenith angle distribution is analyzed as a function of the detection time.

Since the variance of the ADC counts is proportional to the background light, the idea is to use events that pass the quality selection cuts [59] to study the behavior of the mean zenith angle as a function of the variance. This is done by using the information stamped in the first 200 time bins of all triggered pixels of a given event.

For this analysis, we used 25,591 events detected from September 2004 until December 2015. We have evaluated the variance of the ADC signal counting with respect to all triggered pixels for each event. The left panel of figure 5.4 shows an example of the ADC signal as a function of the time bin for a given triggered pixel of a real event while the right panel of the same figure shows the distribution of the signal variance for all triggered events.

The scatter plot of the zenith angle as a function of the signal variance is shown

³Analog-to-digital converter.

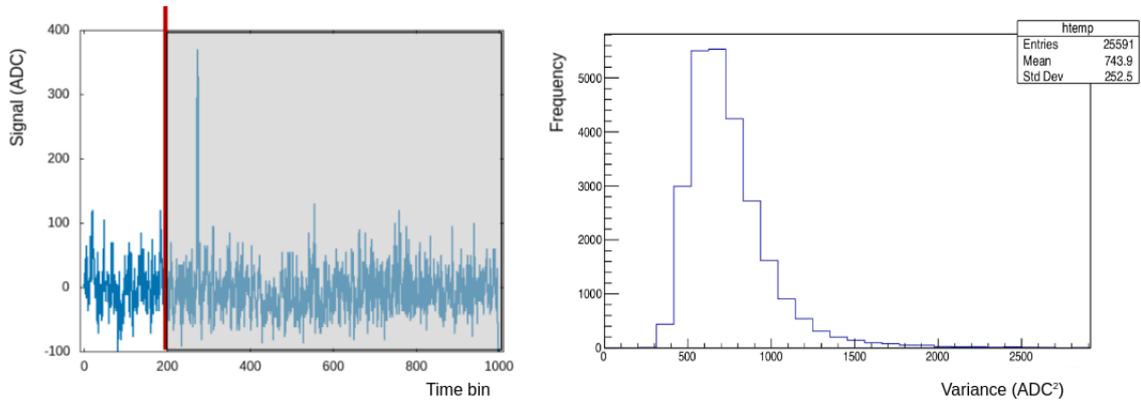


Figure 5.4: Left: example of the ADC signal as a function of the time bin for a given triggered pixel of a real event. Right: distribution of the signal variance for all triggered events.

in the left panel of figure 5.5. As we can see from that plot and from the right panel of the figure 5.4, there are few events with large variance values. The right panel of figure 5.5 shows the mean zenith angle $\langle\theta\rangle$ as a function of the variance. One can notice a slight increase of the mean zenith angle with increasing variance up to ~ 1500 ADC². The increase of the mean zenith angle stops at higher values of the variance changing to a flat behavior. The error bars for the highest variance bins are very large because of the corresponding small number of events.

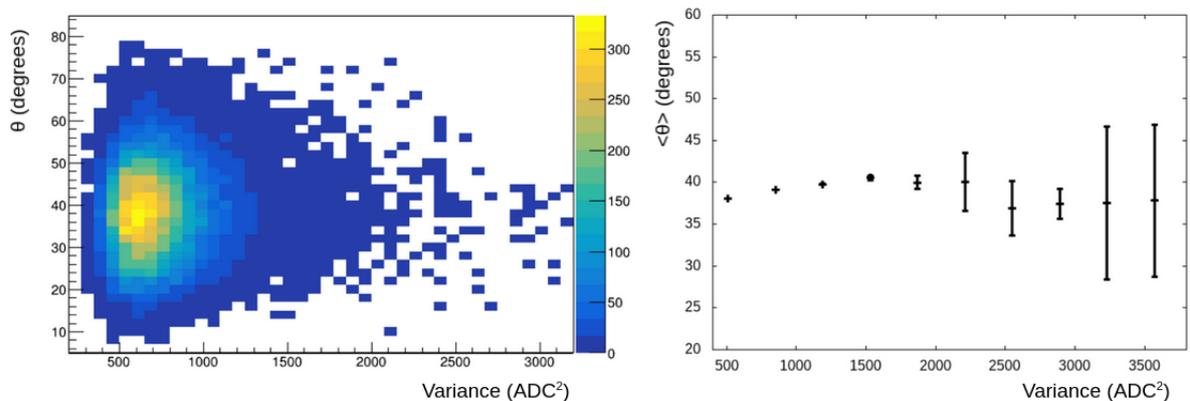


Figure 5.5: Left: example of the ADC signal as a function of the time bin for a given triggered pixel of a real event. Right: signal variance distribution for all triggered events.

In order to avoid bins with too few events, we plotted the zenith angle as a function of the variance by using bins of different size with the same number of events. This is shown in figure 5.6. From this plot, we can see really clearer the slight increase of the mean zenith angle from $\sim 37.5^\circ$ to $\sim 40^\circ$ as the variance increases. The reason for this bias will be addressed elsewhere through the investigation of bias in X_{max} and χ_0 ⁴ distributions as a function of the signal variance.

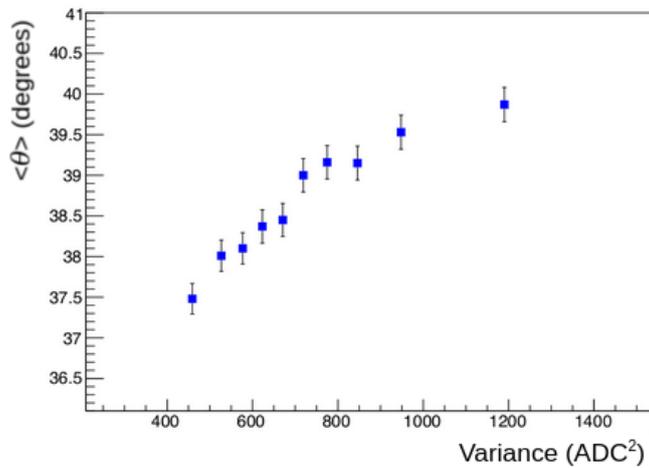


Figure 5.6: Mean zenith angle as a function of the variance. The variance bins are of different size in order to contain the same number of events.

The left panel of figure 5.7 shows the zenith angle distribution from different bins of UTC time. One can see that all are consistent, indicating that there is no change in zenith angle distribution as a function of time. The same conclusion is obtained from the analysis of the right panel of figure 5.7, where the mean zenith $\langle \theta \rangle$ is shown as a function of the time. A flat behavior of the mean zenith angle as a function of time is observed, with a $\chi^2/\text{ndf} = 13.04/10$ obtained for a constant function fit. Therefore the *shuffling* technique can be applied for events detected by the FD since the zenith angle distribution is independent of the detection time.

In conclusion, we have studied the behavior of the zenith angle distribution of FD events as a function of the variance of the ADC counts. A slight increase is observed in the

⁴The definition of χ_0 is given in section 3.2.

mean zenith from $\sim 37.5^\circ$ until $\sim 40^\circ$ for signal variances smaller than 1500 ADC^2 . Above that value, the mean zenith angle distribution seems to have a flat behavior, although this conclusion is dependent on the reduced statistic for the highest bins of variance. On the other hand, we have shown that the zenith angle distribution is independent of the detection time, indicating that the *shuffling* technique can be safely applied for events detected by the FD.

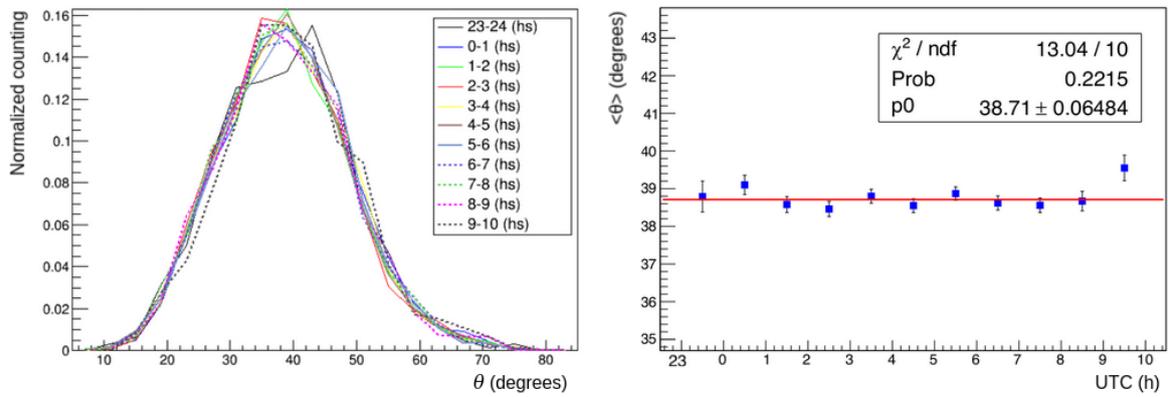


Figure 5.7: At the left, we have the zenith angle distribution of events according to different time windows during the FD shifts. Right, average zenith angle distribution according to each time bin.

Chapter 6

On combining directional intensity and direction composition

The aim of this chapter is to assess the net gain obtained by combining directional intensity data with directional composition data in comparison to the standalone analysis of directional composition data only. Instead of searching for enhanced anisotropies based on a subset of events enriched in light elements (as done in chapter 5), our approach is based on performing a cross-correlation between intensity and composition data in order to characterize whether possible similarities between these data are significant or not, and ultimately, whether the non-uniformities in the directional intensity are preferentially shaped by lighter elements or not. Our model considers the data set we have at our disposal above 8 EeV, namely 30,000 SD events from which we have information about energy and arrival direction, and 3,000 FD events, providing also X_{\max} information. Besides, we also consider the possibility of using X_{\max} information for each SD event, considering the recent publication [1]. Directional data are characterised in two ways: in terms of right ascension only or using sky-maps. We will present sensitivity curves as a function of the composition modulation.

6.1 Directional data in terms of right ascension only

In this section, the directional data is characterised in terms of right ascension only. Our toy model consists of a “mock signal data set” that is generated by shaping a 4.5% first harmonic amplitude in right ascension to the SD sample, with phase $\alpha_0 = 100^\circ$, in accordance with [53, 54]. With respect to the composition, we consider a simplified scenario, composed of a mix of protons and CNO elements only. The fraction of helium is set to zero. Values of X_{\max} are randomly generated following the Gumbel parametrizations [60] at a mean energy of 10 EeV, considering the EPOS-LHC model [49]. The mass of each event, $A = 1$ or $A = 14$, is drawn at random according to the right ascension-dependent proton fraction x calculated as

$$x(\alpha) = \frac{\langle X_{\max}^{\text{tot}} \rangle + \langle \delta X_{\max}(\alpha) \rangle - \langle X_{\max}^{\text{CNO}} \rangle}{\langle X_{\max}^{\text{prot}} \rangle - \langle X_{\max}^{\text{CNO}} \rangle}, \quad (6.1)$$

with $\langle X_{\max}^{\text{tot}} \rangle = (\langle X_{\max}^{\text{prot}} \rangle + \langle X_{\max}^{\text{CNO}} \rangle) / 2$. The right ascension-dependent shift $\langle \delta X_{\max}(\alpha) \rangle$ is determined in a given bin of right ascension $\Delta\alpha$ by the available statistics $\Delta N(\alpha)$ and by an amplitude factor κ describing the significance of the first harmonic modulation of $\langle X_{\max}(\alpha) \rangle$ relative to $\langle X_{\max} \rangle$:

$$\langle \delta X_{\max}(\alpha) \rangle = \frac{\kappa \cos(\alpha - \alpha_0) \sigma(X_{\max}^{\text{tot}})}{\sqrt{\Delta N(\alpha)}}. \quad (6.2)$$

For instance, for $\kappa = 3$, $\langle \delta X_{\max}(\alpha) \rangle$ corresponds to a higher average of X_{\max} with 3σ confidence level at the right ascension α_0 (since $\cos(\alpha - \alpha_0) = 1$) associated with a smaller average of X_{\max} with 3σ confidence level at the opposite direction in right ascension. For a better comprehension, we can write the proton fraction as

$$x(\alpha) = 0.5 + A \cos(\alpha - \alpha_0), \quad (6.3)$$

with $A \approx 0.12\kappa$. Figure 6.1 illustrates the behavior of the proton fraction $x(\alpha)$ for different values of κ .

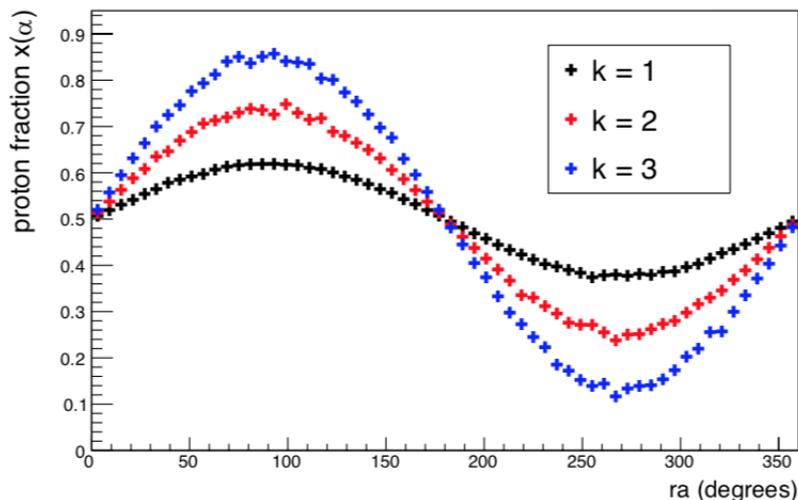


Figure 6.1: Behavior of the proton fraction as a function of the right ascension, $x(\alpha)$, for different values of κ .

Sensitivity curves will be a function of this parameter. Once a signal data set is drawn, the cross-correlation of the $dN/d\alpha$ and $d\langle X_{\max}\rangle/d\alpha$ distributions is estimated using a scalar product-like quantity:

$$c = \frac{\int d\alpha f(\alpha) g(\alpha)}{\sqrt{\int d\alpha |f(\alpha)|^2 \int d\alpha |g_{\text{sat}}(\alpha)|^2}}, \quad (6.4)$$

where the f and g functions are defined so that the scale of c naturally ranges from around -1 and $+1$:

$$f(\alpha) = \frac{dN}{d\alpha} - \left\langle \frac{dN}{d\alpha} \right\rangle_{\alpha}, \quad g(\alpha) = \frac{d\langle X_{\max}\rangle}{d\alpha} - \left\langle \frac{d\langle X_{\max}\rangle}{d\alpha} \right\rangle_{\alpha}. \quad (6.5)$$

The notation g_{sat} stands for the maximum deviation of $\langle X_{\max}(\alpha) \rangle$ relative to $\langle X_{\max} \rangle$, obtained for $x = 1$. The choice for this normalisation allows for maximising the cross-correlation parameter c when an excess is entirely caused by proton showers. Fig. 6.2 shows, as an example, the event count rate (in the left panel) and the $\langle X_{\max} \rangle$ (in the right panel) as a function of right ascension obtained from one simulated sky with $\kappa = 1$. The corresponding values of $f(\alpha)$ and $g(\alpha)$ are shown in Fig. 6.3.

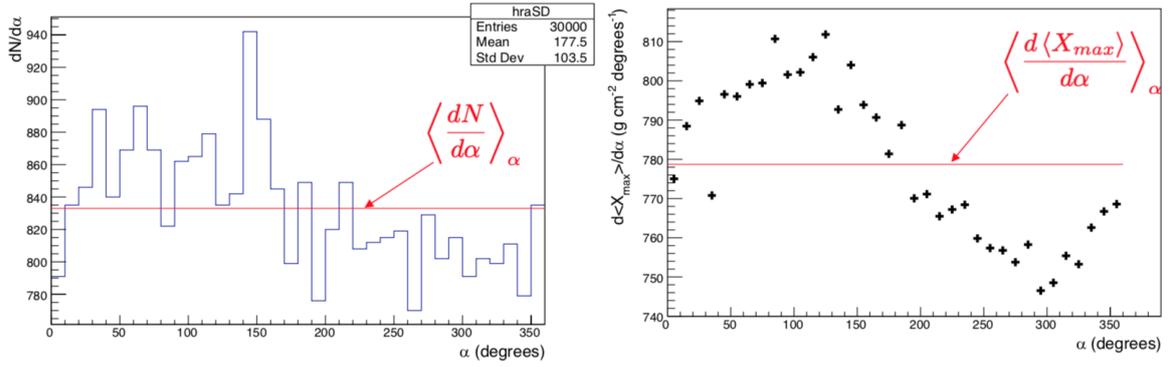


Figure 6.2: Event count rate (in the left panel) and $\langle X_{max} \rangle$ (in the right panel) as a function of right ascension obtained from one simulated sky with $\kappa = 1$.

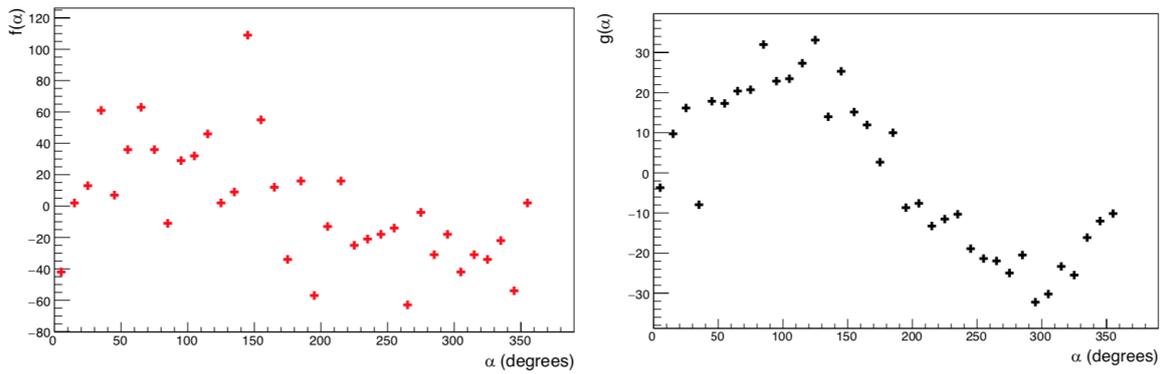


Figure 6.3: Functions $f(\alpha)$ and $g(\alpha)$ corresponding to the distributions of Fig. 6.2 for one simulated sky with $\kappa = 1$.

The degree of correlation between the event count rate and $\langle X_{max} \rangle$ as a function of right ascension is then obtained by comparing the obtained value of c to the distribution of c values obtained for isotropic samples with $\langle \delta X_{max}(\alpha) \rangle = 0$. The p -value for c corresponds to the fraction of background samples whose values of c are larger than the one obtained. The detection power of this cross-correlation proposal is obtained by counting the fraction of the simulated skies whose values of c are larger than the value of c_{99}^{iso} , corresponding to 99% C.L.. In other words, c_{99}^{iso} is defined such that only 1% of the isotropic simulations presents values of $c > c_{99}^{iso}$.

The detection power is presented in Fig. 6.4 for various values of κ . They were obtained by considering 10,000 simulations for each value of κ and a bin size of 6° . The black line corresponds to a scenario without taking into account the detector resolution of the experiment. The blue one only accounts for a detector resolution in X_{max} of $\Delta X_{max} = 16 \text{ g/cm}^2$, while the red curve takes into account both the X_{max} detector resolution and the dipole uncertainty in right ascension $\Delta\alpha_0 = 10^\circ$. One can see that detection power is 99% for the modest value of $\kappa = 1.6$, corresponding to a fraction of protons $x = 69\%$ in the right ascension corresponding to the maximum of the dipole, and that the detector resolution has a minor effect in the detection power $\sim 5\%$.

We also studied the effect of the right ascension bin size on the detection power. The detection power with 99% C.L. as a function of κ for several bin sizes is presented in Fig. 6.5. We can see that the detection power increases with the size of the bin in right ascension and saturates at the bin size $\sim 20^\circ$.

These results considered that the X_{max} information resulted from FD measurements. As we mentioned before, we can also use the X_{max}^{SD} information obtained on an event-by-event basis by the SD of the Observatory. We adopted a conservative X_{max}^{SD} resolution of 40 g/cm^2 , repeated the same exercise, and computed the detection power by considering samples with all SD events (30,000 events) and with half of the events (15,000 events). The motivation behind this is that due to the selection criteria, the analysis required to infer the X_{max}^{SD} of the SD events can discard some of them. Figure 6.6 presents a comparison between the detection power acquired by the FD and SD samples as a function of κ . A slightly

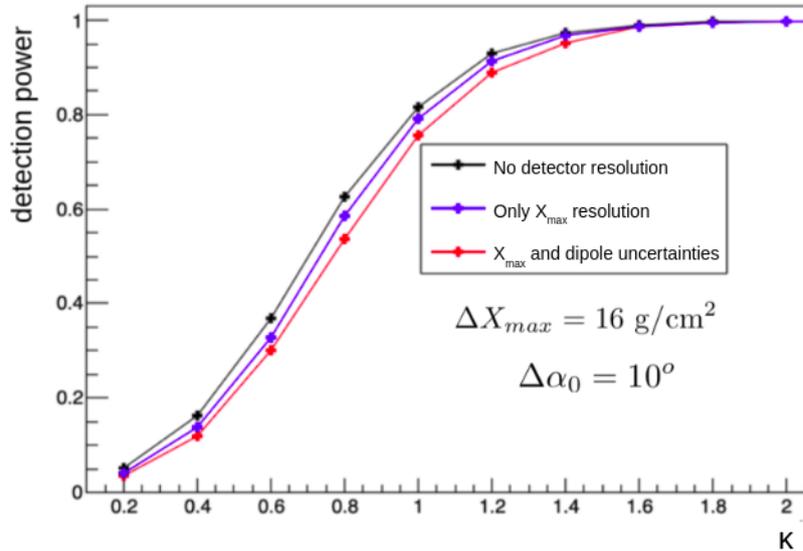


Figure 6.4: Detection power as a function of κ . The black line corresponds to a scenario without taking into account the detector resolution of the experiment. The blue only takes into accounts in X_{max} detector resolution of $\Delta X_{max} = 16 \text{ g/cm}^2$ while the red curve takes into account both the X_{max} detector resolution and the dipole uncertainty in right ascension $\Delta\alpha_0 = 10^\circ$.

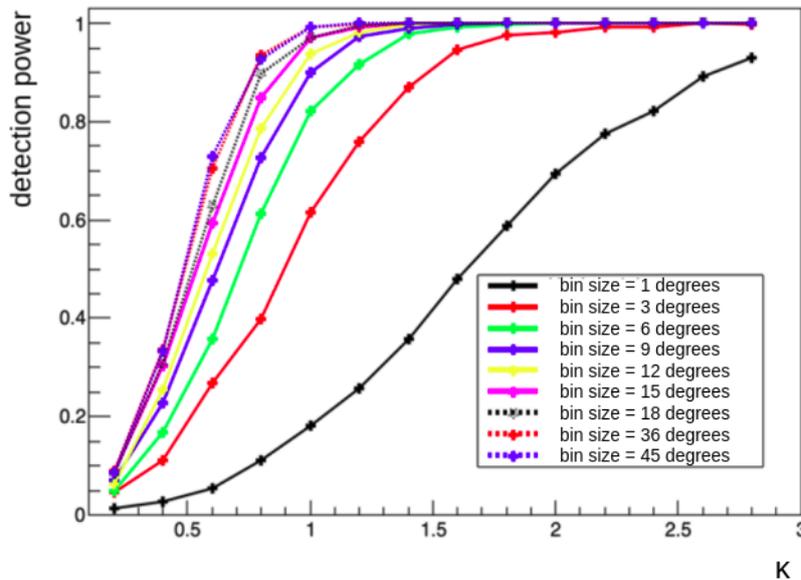


Figure 6.5: Detection power as a function of κ for several right ascension bin sizes.

better sensitivity is obtained by using the X_{max}^{SD} information of all SD events. Notice also that the X_{max}^{SD} resolution by using deep neural network, as reported in [1], is ~ 25 g/cm², implying in a larger detection power.

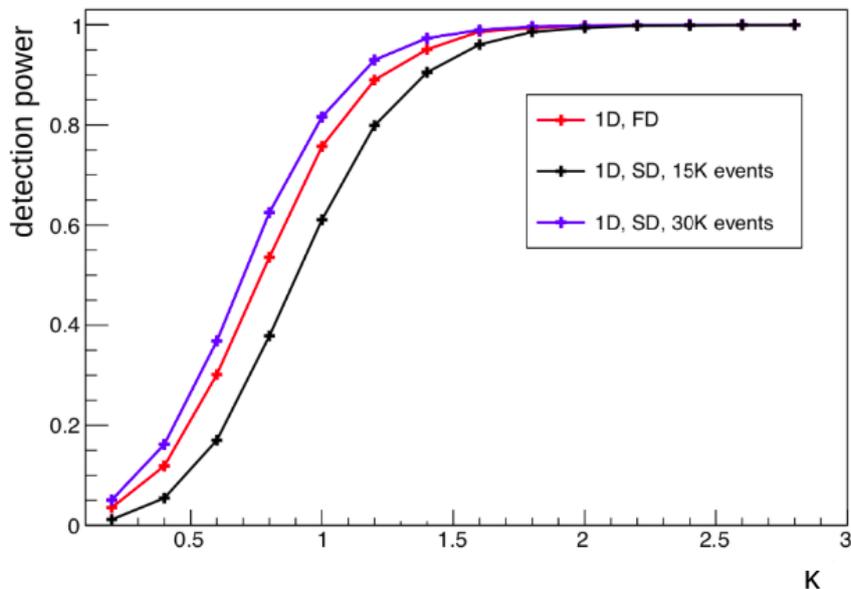


Figure 6.6: Detection power as a function of κ considering samples measured by the FD and SD.

6.2 Directional data in terms of sky maps

In this section, we extend the previous analysis to a 3D case, by considering sky maps of the event count rate and composition maps. Our model consists in injecting a 6.5% dipolar flux modulation in the event count rate with dipolar direction $\hat{n}_d = (\alpha_d, \delta_d) = (100^\circ, -24^\circ)$, corresponding to Galactic coordinates $(l_d, b_d) = (233^\circ, -13^\circ)$, in agreement with Auger measurements [53] for $E > 8$ EeV. Now the fraction of protons x is written as a function of the angular distance between the dipole direction \hat{n}_d and a given direction of the sky \hat{n} as

$$x(\hat{n}) = 0.5 + A \cos(\Psi), \quad (6.6)$$

with $\cos(\Psi) = \hat{n} \cdot \hat{n}_d$ and $A \approx 0.12\kappa$. An illustration of a map of the proton fraction x is presented in Fig. 6.7 for $\kappa = 2$ in Galactic Coordinates.

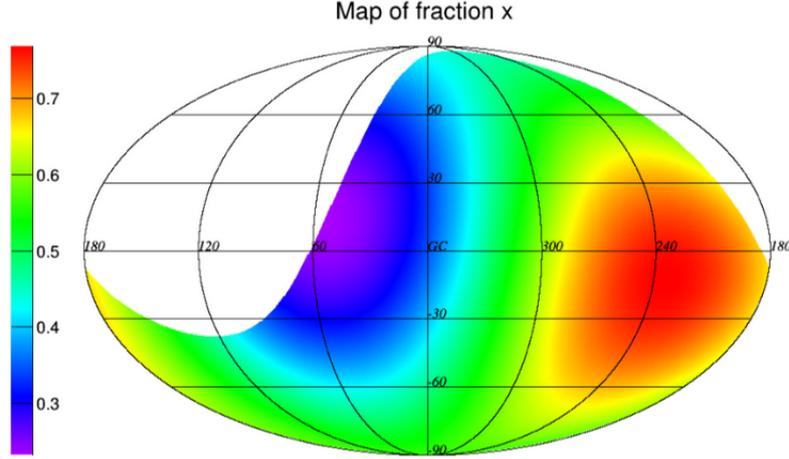


Figure 6.7: Map of the proton fraction x in for $\kappa = 2$ in Galactic coordinates.

The cross-correlation estimator, extended to the 3D case, is given by:

$$c = \frac{\int d\Omega f(\vec{n})g(\vec{n})}{\sqrt{\int d\Omega [f(\vec{n})]^2 \cdot \int d\Omega [g(\vec{n})]^2}}, \quad (6.7)$$

where $f(\vec{n})$ is written as a function of the event map $\Phi(\vec{n})$ and the coverage map $\Phi^{BG}(\vec{n})$,

$$f(\vec{n}) = \frac{\Phi(\vec{n}) - \Phi^{BG}(\vec{n})}{\sqrt{\Phi^{BG}(\vec{n})}}, \quad (6.8)$$

and $g(\vec{n})$ is

$$g(\vec{n}) = \frac{d \langle X_{\max} \rangle}{d\Omega} - \left\langle \frac{dX_{\max}}{d\Omega} \right\rangle. \quad (6.9)$$

The quantities presented in Eq. 6.8 and 6.9 are computed within angular windows of a given radius. Fig. 6.8 presents an example of $f(\vec{n})$ and $g(\vec{n})$ functions, in Galactic coordinates, resulting from one specific simulation, considering angular windows with 20° of radius. From now on, we will call this process as a top-hat filter, since only events within the angular windows are used to compute $f(\vec{n})$ and $g(\vec{n})$. For instance, a value of 4

in the $f(\vec{n})$ sky map indicates that there is a local excess of events with 4σ of C.L. at the corresponding location in the celestial sphere. On the other hand, a value of 15 in the $g(\vec{n})$ sky map indicates that, for this specific direction in the sky, the average of X_{\max} within a circle of 20° of radius is 15 g/cm^2 larger than the average computed for the whole sky.

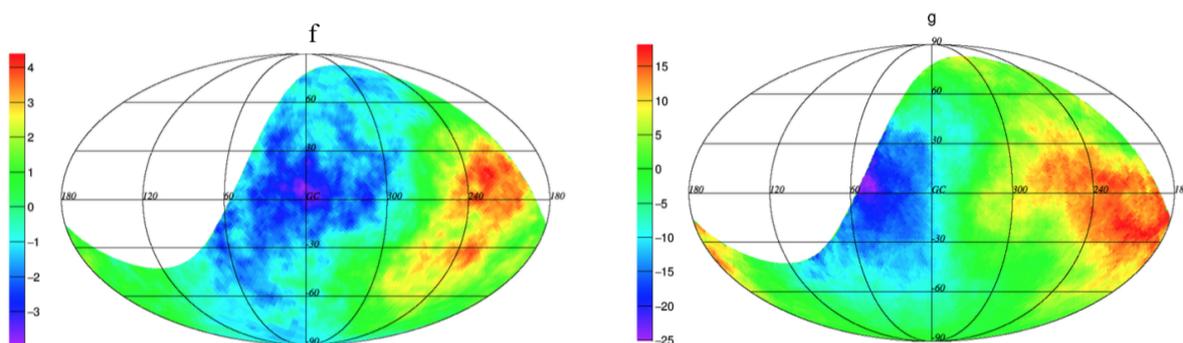


Figure 6.8: Example of $f(\vec{n})$ and $g(\vec{n})$ functions, in Galactic coordinates, resulting from one specific simulation, considering angular windows with 20° of radius.

We again performed 10,000 MC simulations of cosmic-ray skies in order to assess the detection power as a function of κ by considering X_{\max} measurements as obtained by the FD without taking into account its resolution. The maps are convoluted with a top-hat filter of a specific angle. Figure 6.9 presents the detection power as a function of κ for different top-hat filter radii. One can see that the detection power saturates at a top-hat filter of $\sim 20^\circ$.

Finally, in Fig. 6.10, we present a comparison between the detection power obtained by using directional data only in right ascension (1D) with the one obtained by using sky maps (3D). We conclude that both have similar detection power. A probable reason for that is the small value of the z component of the dipole d_z considered in our toy model ($\delta_d = -24^\circ$). As a consequence, the gain in using 3D data is not satisfactory. A larger detection power is expected in the case of a dipolar flux modulation with a larger d_z component. This should be considered in the future since the aim of this study is to search for correlation between directional flux and directional mass composition data in several energy bins.

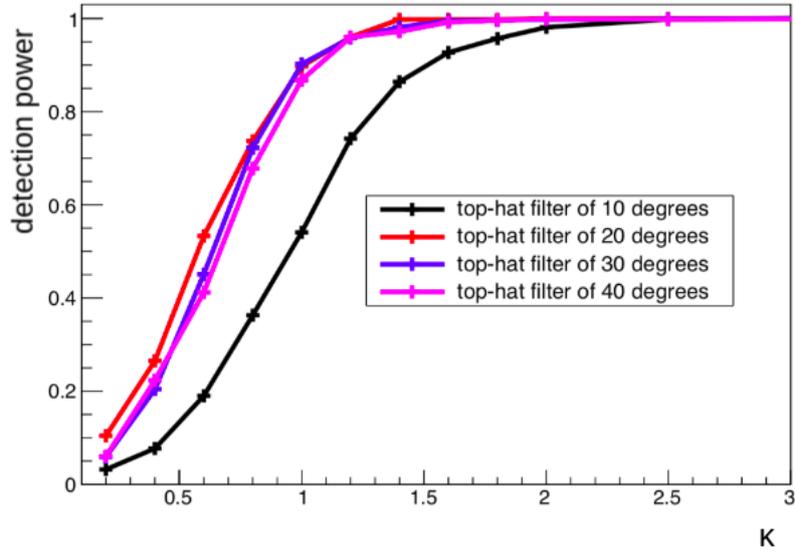


Figure 6.9: Detection power as a function of κ for different top-hat filter radii.

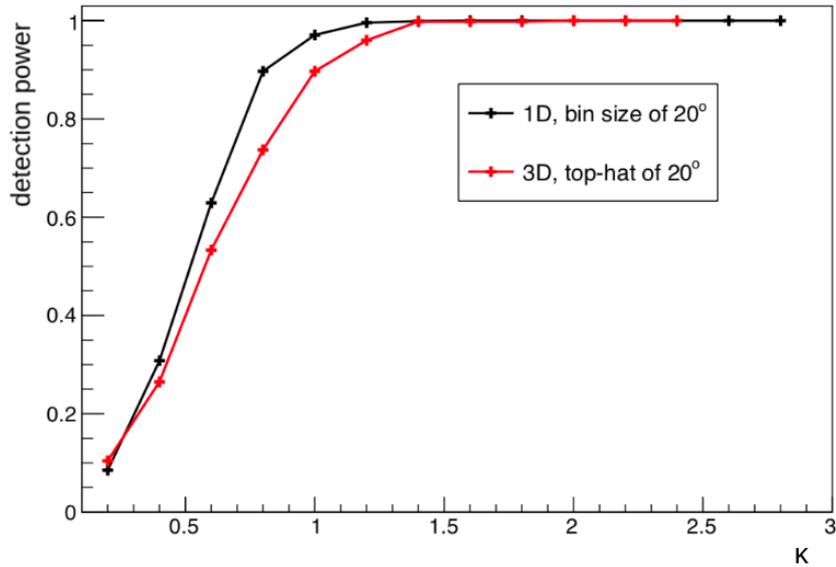


Figure 6.10: Detection power as a function of κ for different top-hat filter radii.

Chapter 7

Correlation Studies of UHECRs events and starburst sources at lower energies

Recently the Pierre Auger Collaboration reported an indication of anisotropy in the arrival directions of cosmic rays with energy above 39 EeV on an intermediate angular scale. As described in chapter 3.3.3, this result is based on the search for a correlation between Auger data and two classes of sources detected by Fermi-LAT - active galactic nuclei (AGN) and starburst galaxies, by using their gamma-ray and radio flux as weights for their relative cosmic-ray luminosity. The UHECR model was built as a sum of an isotropic and an anisotropic component, the last one originated from the selected sources. A maximum likelihood analysis, with anisotropic fraction f and search radius Ψ as free parameters, was performed to optimize the degree of correlation between the sky map and the cosmic-ray events. For the gamma-ray AGNs, a 2.7σ excess was found at an angular scale of $\Psi = 6.9_{-2.3}^{+3.9}^\circ$ and for star forming galaxies the best-fit search reached a maximum at 4.0σ at an angular scale of $\Psi = 13_{-3.0}^{+4.0}^\circ$.

Inspired in a previous correlation between events with energy above 55 EeV and AGNs from the VCV catalog [5], searches for overdensities at lower energies were performed [61]

in the AGNs directions based on the proposition of Lemoine and Waxman [5]. According to it, assuming that the cosmic-ray acceleration depends just on the particle rigidities and neglecting the energy losses during the cosmic-ray propagation, an anisotropy at high energy due to nuclei of charge Z and energy E should be followed by an anisotropy in the same direction of the sky at energies Z times smaller (E/Z) due to protons originated in the same sources. Taking into account the energy bins relative to $Z = 2, 6, 13$ and 26 , no evidence for overdensities was found.

In this work, assuming that the correlation of the arrival directions of events with energy above 39 EeV and the starburst galaxies reported in [4] is originated by nuclei of charge $Z = 2, 6, 13$ or 26 , we search for overdensities within 21° around the starburst directions at corresponding energy thresholds of $E_{th}/Z = 19.5$ EeV, 6.5 EeV, 3.0 EeV and 1.5 EeV. This choice of angular scale is justified because, in the starburst analysis [4], the maximum deviation from isotropy is at $\Psi = 13^\circ$ using a Gaussian filter, whose corresponding top-hat radius is given approximately by $\sim 1.5 \times 13^\circ = 21^\circ$ [62].

7.1 Data analysis

The data set used in this work is composed of events detected with the surface detector (SD) of the Pierre Auger Observatory from 1 January 2004 to 31 May 2017 with zenith angles θ up to 60° . We consider a quality cut that requires at least five of the six water-Cherenkov detectors surrounding the station with the largest signal were operational at the time the event was recorded. Due to the supposition that the correlation reported in [4] is originated by nuclei with $Z = 2, 6, 13$ or 26 , we selected events with energies above $E_{th}/Z = 19.5$ EeV, 6.5 EeV, 3.0 EeV and 1.5 EeV. In Figure 7.1 the directions of the 23 brightest nearby starburst galaxies used in this analysis (red stars with circles of 20°) are shown together with the arrival directions of the events with energies above $E_{th}/Z = 19.5$ EeV, 6.5 EeV, 3.0 EeV and 1.5 EeV. A list containing the positions of the 23 starburst galaxies, in galactic coordinates, is presented in table 4.1. It also presents the distance of each source and its Flux weight of the source by Fermi-LAT used in [4]. For each

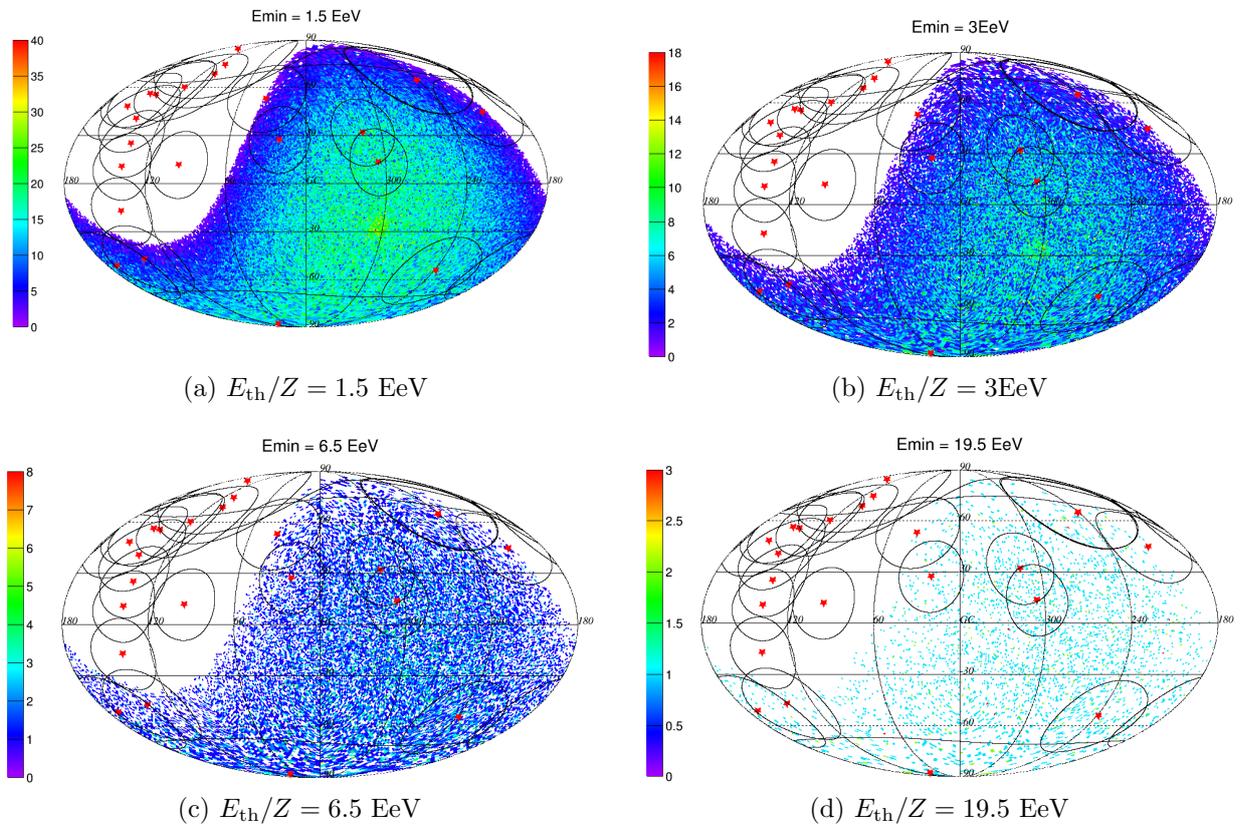


Figure 7.1: The directions of the 23 brightest nearby starburst galaxies used in this analysis (red stars) are shown together with the arrival directions of the cosmic rays with energy above $E_{\text{th}}/Z = 19.5 \text{ EeV}$, 6.5 EeV , 3.0 EeV and 1.5 EeV . Maps are in Galactic coordinates.

energy bin, we search for overdensities within circles of 21° radius centered around each star-forming galaxy position by counting the number of observed (N_{obs}) and background (N_{bkg}) events. The number of events expected inside the angular windows centered around the starburst galaxies by random correlations taking into account the detector exposure, N_{bkg} , is estimated as

$$N_{bkg} = \frac{(N_{tot} - N_{obs})x}{1 - x}. \quad (7.1)$$

In this equation, N_{tot} is the total number of events, N_{obs} is the number of events inside the angular windows and x is the fraction of the sky weighted by the observatory's exposure within 21° around the sources. For energies above 3 EeV, the trigger efficiency is 100% for vertical events and the exposure is determined analytically [41]. On the other hand, the exposure for $E > 1.5$ EeV was evaluated using the shuffling method [63]. Figure 7.2 presents the coverage maps in Galactic coordinates for $E \geq 3$ EeV, in the left panel, and $E \geq 1.5$ EeV, in the right panel. For all energy bins, the fraction of the sky weighted by the observatory's exposure is $x = 0.32$. All the uncertainties in the number of background

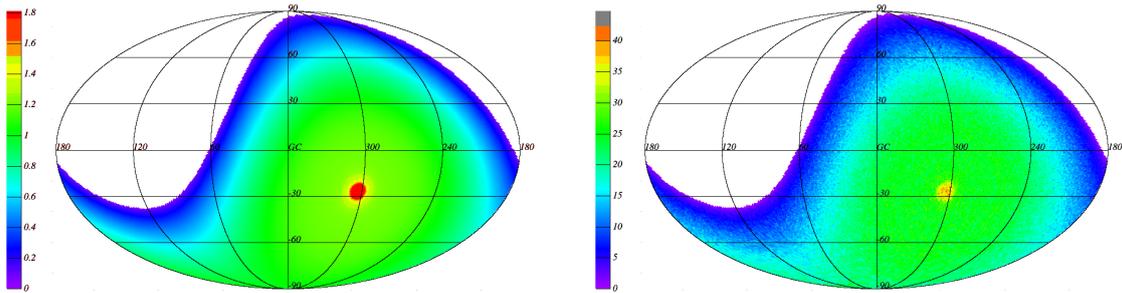


Figure 7.2: Coverage maps in galactic coordinates. **Left:** $E \geq 3.0$ EeV **Right:** $E \geq 1.5$ EeV.

events were evaluated assuming Poissonian statistic. We found a significance of 3.33σ for energies above 39 EeV as expected since this is the energy threshold that maximizes the significance of the likelihood study [4]. Nevertheless, notice that, differently from the later, the sources are not weighted by their fluxes in the present study. Besides, we also observed

CHAPTER 7. CORRELATION STUDIES OF UHECRS EVENTS AND STARBURST SOURCES AT LOWER ENERGIES

an excess above 3σ for $E_{\text{th}}/Z = 19.5$ EeV, corresponding to $Z=2$. This excess is correlated with the result above 39 EeV. It can be seen that the significance is reduced to $s = 2.10\sigma$ if the highest energy events are discarded. It is important to say that the detection of helium nuclei coming from starburst galaxies at $E > 39$ EeV is expected, given the relative small distance of these SBG to us. Figure 7.3 presents the mean energy loss length χ_{loss} in Mpc of helium, which indicates the bands of each process of loss of energy.

Table 7.1: Total observed and background event numbers for different energy bins. The uncertainty in the number of background events, N_{bkg} , was evaluated assuming Poissonian statistics.

Z	E_{min} [EeV]	N_{tot}	N_{obs}	N_{bkg}	$(N_{\text{obs}} - N_{\text{bkg}})/(\sqrt{N_{\text{bkg}}})$
	39.0	579	215	171 ± 13	3.33
2	19.5	3649	1242	1133 ± 34	3.25
2	19.5 - 39.0	3071	1027	962 ± 31	2.1
6	6.5	31805	10219	10158 ± 101	0.60
13	3.0	134209	43160	42847 ± 207	1.51
26	1.5	594161	189709	190330 ± 436	-1.42

It is worth to continue the monitoring of this region possibly by searching for multiplet events.

We also searched for excess of lower energy events around the three most significant SBGs sources for the analysis in [4], NGC 4945 (ℓ, b) = (305.3°, 13.3°), NGC 253 (ℓ, b) = (97.4°, -88°) and M83 (ℓ, b) = (314.6°, 32°). Thus, for energy $E \geq 3$ EeV the fraction of the sky weighted by the observatory's exposure is $x = 0.0574$ for M83, $x = 0.0550$ for NGC 253, and $x = 0.0648$ for NGC 4945. On the other hand, for $E < 3$ EeV it is $x = 0.0582$ for M83, $x = 0.0555$ for NGC 253 and $x = 0.0657$ for NGC 4945. Table 7.2 summarizes the results for the three SBGs with the major impact in the likelihood study [4]. The excess of 4.01σ for energies above 39 EeV in the direction of NGC 4945 is a clear indication that this source has a major impact on the likelihood study [4]. All other results are consistent with background expectations. Additionally, we performed an analysis including inclined events detected from 01/01/2004 to 31/08/2016 for $E_{\text{th}} > 19.5$ EeV. The results are presented in

CHAPTER 7. CORRELATION STUDIES OF UHECRS EVENTS AND STARBURST SOURCES AT LOWER ENERGIES

Table 7.2: Total, observed and background event numbers for different energy bins for M83 (a), NGC 253 (b) and NGC 4945 (c).

SBGs			M83		
Z	E_{\min} [EeV]	N_{tot}	N_{obs}	N_{bkg}	$(N_{\text{obs}} - N_{\text{bkg}})/(\sqrt{N_{\text{bkg}}})$
	39.0	579	41	33 ± 6	1.44
2	19.5	3649	220	209 ± 15	0.77
6	6.5	31805	1832	1825 ± 43	0.16
13	3.0	134209	7793	7698 ± 88	1.08
26	1.5	594161	34420	34590 ± 186	-0.91

(a)

SBGs			NGC 253		
Z	E_{\min} [EeV]	N_{tot}	N_{obs}	N_{bkg}	$(N_{\text{obs}} - N_{\text{bkg}})/(\sqrt{N_{\text{bkg}}})$
	39.0	579	41	31 ± 6	1.73
2	19.5	3649	215	200 ± 14	1.07
6	6.5	31805	1743	1750 ± 42	-0.16
13	3.0	134209	7526	7373 ± 86	1.78
26	1.5	594161	33208	32962 ± 182	1.35

(b)

SBGs			NGC 4945		
Z	E_{\min} [EeV]	N_{tot}	N_{obs}	N_{bkg}	$(N_{\text{obs}} - N_{\text{bkg}})/(\sqrt{N_{\text{bkg}}})$
	39.0	579	60	36 ± 6	4.01
2	19.5	3649	272	234 ± 15	2.48
6	6.5	31805	2075	2060 ± 45	0.33
13	3.0	134209	8784	8691 ± 93	1.0
26	1.5	594161	38737	39057 ± 198	-1.62

(c)

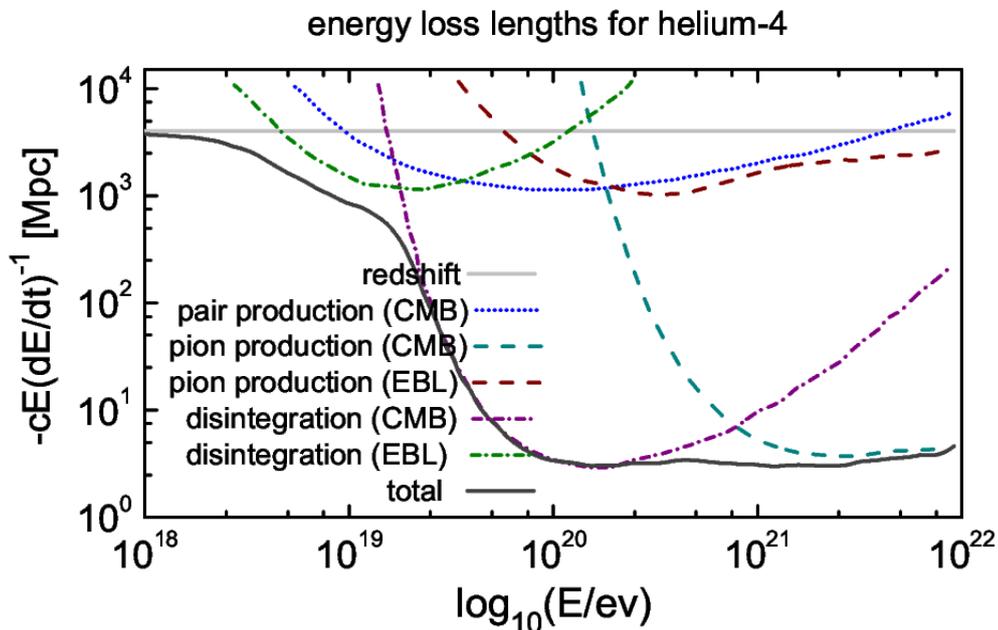


Figure 7.3: Energy loss length of the helium as a function of energy, which indicates the bands of each process of loss of energy [64, 65].

table 7.3. The fraction of the sky weighted by the exposure of the observatory is $x = 0.33$. An excess of events above 3σ for $E > 19.5$ EeV was obtained, and as said before, this excess is correlated with the one observed at $E_{th}/Z > 39$ EeV. The significance is reduced to 2.29σ if the highest energy events are discarded.

Table 7.3: Total observed and background event numbers for different energy bins. The uncertainty in the number of background events, N_{bkg} , was evaluated assuming Poissonian statistics.

Z	E_{min} [EeV]	N_{tot}	N_{obs}	N_{bkg}	$(N_{obs} - N_{bkg})/(\sqrt{N_{bkg}})$
	39.0	796	292	248 ± 16	2.78
2	19.5	5234	1814	1685 ± 41	3.16
2	19.5 - 39.0	4439	1523	1436 ± 38	2.29

In table 7.4, we also included inclined events considering the three SBGs with great impact on the likelihood analysis. The excess around NGC 4945 for energies above 39 EeV increased to 5.67σ with the inclusion of inclined events. Besides, an excess of events above

3σ for $E > 19.5$ EeV was obtained, but, as said before, this excess is correlated with the one observed at $E_{\text{th}}/Z = 39$ EeV. The significance is reduced to 1.06σ if the highest energy events are discarded. All other results are consistent with background fluctuations.

7.2 Conclusions

We searched for excess of low energy events in the direction of starburst galaxies by considering the proposition of Lemoine and Waxman [5] which says that an anisotropy in high energy due to nuclei of charge Z and energy E should be followed by an anisotropy in the same direction of the sky at energies Z times smaller (E/Z) due to protons originated in the same sources. An excess of events with statistical significance above 3σ for $E \geq 19.5$ EeV was observed. This energy threshold is associated with He ($Z = 2$) nuclei coming from starburst galaxies at energies above 39 EeV. However, this excess is correlated with the one observed at $E > 39$ EeV, decreasing the significance to 2.10σ if the highest energy events are discarded. However, notice that, differently from the likelihood study, the sources are not weighted in the present study. It is also important to say that the detection of helium nuclei coming from SBGs is expected, given the relative small distances of these sources to us. It is worth continuing the monitoring of this region possibly by searching for multiplet events, patterns of aligned cosmic-ray directions in the sky map caused by magnetic deflections during the trajectory from the source to Earth with the highest energy events pointing closer their source positions. Additionally, we searched for excesses due to the most significant SBGs sources (M83/ 100 NGC 253/ NGC 4945). Interesting excesses of events around NGC 4945 were observed. The increase in the statistics with the detection of new events will help to understand the results described in this work.

CHAPTER 7. CORRELATION STUDIES OF UHECRS EVENTS AND STARBURST SOURCES AT LOWER ENERGIES

Table 7.4: Total, observed and background event numbers for different energy bins for M83, NGC 4945 and NGC 253.

SBGs			M83		
Z	E_{\min} [EeV]	N_{tot}	N_{obs}	N_{bkg}	$(N_{\text{obs}} - N_{\text{bkg}})/(\sqrt{N_{\text{bkg}}})$
	39.0	796	53	40 ± 6	1.99
2	19.5	5234	263	270 ± 16	-0.42
2	19.5 - 39.5	4439	210	230 ± 15	-1.29

(a)

SBGs			NGC 253		
Z	E_{\min} [EeV]	N_{tot}	N_{obs}	N_{bkg}	$(N_{\text{obs}} - N_{\text{bkg}})/(\sqrt{N_{\text{bkg}}})$
	39.0	796	50	39 ± 6	1.8
2	19.5	5234	281	257 ± 16	1.47
2	19.5 - 39.5	4439	231	219 ± 15	0.83

(b)

SBGs			NGC 4945		
Z	E_{\min} [EeV]	N_{tot}	N_{obs}	N_{bkg}	$(N_{\text{obs}} - N_{\text{bkg}})/(\sqrt{N_{\text{bkg}}})$
	39.0	796	84	46 ± 7	5.67
2	19.5	5234	368	312 ± 18	3.16
2	19.5 - 39.5	4439	284	267 ± 16	1.06

(c)

Chapter 8

Combined fit of the energy spectrum, mass composition and arrival directions detected at the Pierre Auger Observatory

The results presented in the Auger Combined fit papers [48, 66] show that the analysis is sensitive to several choices made, such as: (1) the code used for the simulation (CRPropa [67] or SimProp [68]), (2) the interaction model for photo-disintegration (TALYS [69], PSB [70] or Geant4 [71]), (3) the assumed extragalactic background light spectrum (Gilmore [72] or Domínguez [73]) and (4) the hadronic interaction model (EPOS-LHC [49], QGSJetII-04 [74] or Sibyll 2.1 [75]). The degree of sensitivity of the combined fit analysis is reflected on the different results obtained independently by different groups within the Pierre Auger Collaboration [76] and [77]. In order to help on that task, we focused on the CRPropa3, TALYS and Gilmore models, developed an independent software package and performed the analysis for different datasets for both energy spectrum and X_{\max} distributions. In section 8.1, we summarize our implementation of the analysis and the results obtained so far. The package is also described with brief comments on the implementation,

model, datasets, simulation and inputs. Besides, notice that the results obtained in [48] were obtained by assuming a more simplistic assumption of uniformly distributed sources and do not take into account the significant large-scale anisotropy [53, 78, 54] for events with energies above 8 EeV measured by the Auger Collaboration. Aiming to establish a global picture about the cosmic-ray astrophysics, we insert the anisotropy information on the global combined fit, trying to describe the UHECR energy spectrum, shower depth distribution and arrival directions at the same time. The approach to performing this global fit as well, as the results obtained, are presented in section 8.2.

8.1 The Rio Combined Fit package

The package aims to fit the astrophysical parameters at the sources and does not include tools for the simulation stage. It is written in C++, uses ROOT [79] libraries for several calculations, and can be run in parallel. In order to ease the comprehension, we included a documentation created from the source files through Doxygen [80]. The code has been made available to the Pierre Auger Collaboration at the phenomenology task wiki page¹. In the following subsection, we comment on some features of our implementation, as well as the results obtained, adding suitable changes to optimize the fit analysis.

8.1.1 Energy spectrum

As mentioned in section 3.3.4, we assume a power-law injection spectrum at sources with a rigidity-dependent broken exponential cutoff:

$$\frac{dN_{\text{inj},i}}{dE} = \begin{cases} J_0 a_i \left(\frac{E_{\text{inj},i}}{E_0} \right)^{-\gamma}, & E_{\text{inj},i}/Z_i < R_{\text{cut}} \\ J_0 a_i \left(\frac{E_{\text{inj},i}}{E_0} \right)^{-\gamma} \exp \left(1 - \frac{E_{\text{inj},i}}{Z_i R_{\text{cut}}} \right), & E_{\text{inj},i}/Z_i \geq R_{\text{cut}}, \end{cases} \quad (8.1)$$

where $E_{\text{inj},i}$ is the energy of the nucleus i at the source, J_0 is a normalization factor of the spectrum, a_i is the relative abundance of the nucleus i at the source, γ is the spectral index

¹<https://www.auger.unam.mx/AugerWiki/Phenomenology/Tools>

of the flux, Z_i is the atomic number of the nucleus i and R_{cut} is the maximum rigidity at the source. Therefore, the free parameters to be fit are: J_0 , γ , R_{cut} and four of the five abundances a_i , since $a_{\text{H}} + a_{\text{He}} + a_{\text{N}} + a_{\text{Si}} + a_{\text{Fe}} = 1$. We simulated the propagation of 4×10^6 particles of each type of nucleus from the sources to the Earth assuming a simple astrophysical scenario by using the CRPropa3 code [67]. The simulated injected energy of events at source follows a flat distribution in $\log(E_{\text{inj}}/\text{eV})$, i.e., $dN/dE \propto E^{-1}$. Therefore the power low energy spectrum given by eq. 8.1 was obtained by assigning a weight $\propto E_{\text{inj}}^{1-\gamma}$ to each nucleus at Earth. These values were used to built the energy spectrum for each type of injected nucleus. The sum of these histograms (using the abundances a_i as weights) provides the energy spectrum that is used to evaluate a generalized χ^2 defined by deviance, $D(J)$, through a comparison with the Auger data. The fit is performed in $\log(E/\text{eV})$ bins of 0.1 width within $18.7 \leq \log(E/\text{eV}) \leq 20.2$ (15 data points).

8.1.2 X_{max} distributions

We use the values of X_{max} measured with the FD detectors in $\log_{10}(E/\text{eV})$ bins of 0.1 width from $10^{17.8}$ eV to $10^{19.6}$ eV and one larger bin with energies between $10^{19.6}$ eV to 10^{21} eV. Each X_{max} distribution is binned in intervals of $20 \text{ g} \cdot \text{cm}^{-2}$. For the X_{max} simulations, for each injected nucleus we fill 2D histograms with axis $\log_{10}(E/\text{eV})$ (energy at Earth) and A (mass number). These histograms are also weighted accordingly to take into account the parameters to be fit. The model for the X_{max} distributions is obtained from these histograms and the Gumbel distributions [60]. The Gumbel function $g(X_{\text{max}} | E, A)$ describes the X_{max} distribution expected for nuclei with mass number A and energy E (for more details see Appendix A). For the fits presented in this section, we use the hadronic interaction model EPOS-LHC[49]. The Gumbel parameterization is used to infer the fractions of the representative primaries with mass numbers $A = 1, 4, 14, 28, 56$ in the X_{max} distributions. We consider the distributions of all mass number from $A = 1$ to $A = 56$ from the simulation. To compare the values of simulated X_{max} with those measured ones, the Gumbel distributions are corrected for detection effects to give the

expected model probability evaluated at the logarithmic average of the observed event energies, for a given mass distribution at the detection. Finally, we obtain $g(X_{\max}^{\text{rec}}|E)$, which is the expected reconstructed X_{\max} distribution generated by a given combination of mass number fractions at energy E , taking into account the detector response as described in Appendix A.

As mentioned before, several nuclei are injected at the sources, resulting in a huge number of particles at Earth, specially secondaries from the photo-disintegration of heavier nuclei. Reading such simulations and performing the corresponding calculations is very time consuming. As an alternative, the simulations are used to fill 3D histograms whose axis are: A (mass number), $\log_{10}(E/\text{eV})$ (energy at Earth), $\log_{10}(E_{\text{inj}}/\text{eV})$ (energy at source); which are used in the analysis. With bin widths of 0.01 for the energies and 1 for the mass number, this approach shows negligible impact on the results and great increase in performance. For use in the code available in the Auger phenomenology task, the simulations have to be in a file named `sim_hists.root` containing the 3D histograms `A_E_Es_z=XX`, with $XX \in \{1,2,7,14,26\}$.

8.1.3 The deviance

The best-fit parameters are obtained through the minimization of the deviance defined as

$$D \equiv D_J + D_{X_{\max}} = -2 \ln \frac{\mathcal{L}_J}{\mathcal{L}_J^{\text{sat}}} - 2 \ln \frac{\mathcal{L}_{X_{\max}}}{\mathcal{L}_{X_{\max}}^{\text{sat}}} \quad (8.2)$$

where \mathcal{L}_J is the likelihood coming from the spectrum data and $\mathcal{L}_{X_{\max}}$ is the likelihood from the X_{\max} data. In both cases the, \mathcal{L}^{sat} denotes a model that perfectly describes the data. Since the Auger energy spectrum is an unbiased measurement of the flux corrected with the detector response, the correspondent likelihood is obtained in two different ways. In the first case, before comparing model and measurement, a forward folding procedure [81, 82] is applied to each simulated energy spectrum and the deviance D_J is defined by using Poissonian statistics as

$$D_J = -2 \sum_i \mu_i - n_i + n_i \ln \left(\frac{n_i}{\mu_i} \right), \quad (8.3)$$

where n_i corresponds to the observed number of events in the i -th energy bin, and μ_i denotes the corresponding expected number of events obtained from the simulations. In the second case, the measured energy spectrum is unfolded (J^{data}), the likelihood is a product of Gaussian distributions, and the associated deviance is given by

$$D_J = \sum_i \left(\frac{J_i^{data} - J_i^{model}}{\sigma_i^{data}} \right)^2. \quad (8.4)$$

In relation to $\mathcal{L}_{X_{max}}$, it is given by considering that the probability of observing a X_{max} distribution $\vec{k}_i = (k_{i1}, k_{i2}, \dots)$ in the i -th energy bin is given by a multinomial distribution, so that:

$$\mathcal{L}_{X_{max}} = \prod_i n_i! \prod_x \frac{1}{k_{ix}!} (G_{ix}^{model})^{k_{ix}}. \quad (8.5)$$

G_{ix}^{model} is the probability to observe an event in the X_{max} bin x , obtained from the Gumbel distributions corrected for detector effects, given by

$$G_i^{model}(X_{max}^{rec}) = \sum_A a_A \cdot g(X_{max}^{rec} | E_i, A). \quad (8.6)$$

8.1.4 The datasets

The spectrum data for the ICRC 2015 [43], ICRC 2019 [83] and PRL 2020 [12] are hard-coded and can be set on the input file. For the ICRC 2015 spectrum, it is possible to perform the analysis using the forward folding procedure. On the other hand, the X_{max} data is read from ASCII files in a directory specified on the input file. We used the X_{max} data from [59] and [52].

8.1.5 Running the code

There are two main modes to run the code: (i) *scanning* mode and (ii) *full optimization* mode. The *scanning* mode follows the approach described in [48]. A scan is performed on the values of γ and $\log_{10}(R_{\text{cut}}/V)$, and the abundances a_i are optimized using MINUIT. Note that although the normalization J_0 is a free parameter, it can be easily optimized analytically [84]. For the case of Poisson distributions, the expression for the optimal J_0^* is:

$$J_0^* = \frac{\sum_i n_i}{\sum_i \tilde{\mu}_i}. \quad (8.7)$$

where $\tilde{\mu}_i$ is the normalized model prediction, and n_i is the number of detected cosmic rays in the energy bin i . The range of the scan and the increment steps of both parameters can be set via an input file (see appendix (B.1) for more details). In the *full optimization* mode, either γ or $\log_{10}(R_{\text{cut}}/V)$ can be optimized using MINUIT with some features being set through the input file. Once the fit is performed, it is possible to plot the energy spectrum and the X_{max} distributions by using a ROOT-based plotting routine.

Furthermore, the package includes a shared library that can be used in python and other languages. This feature expands the range of tools that can be used in the Combined Fit analysis as a whole. For example it, can be easily adapted and used with packages like PyMC3 [85] for straightforward MCMC analysis.

8.1.6 Results

We simulated cosmic rays at the source for each injected nucleus (H, He, N, Si and Fe) using the CRPropa3. The photo-disintegration interaction model in this case was the TALYS [69], and the EBL model is the one from Gilmore [72]. With these simulations in hand, we built the histograms described before and performed the analysis of distinct datasets.

The scan was performed in steps of 0.02 for both γ and $\log_{10}(R_{\text{cut}}/V)$. As usual,

the abundances at the sources were obtained through the Minuit minimization. In order to obtain the parameter estimation, we performed a re-scan of the likelihood around the best fit for each pair γ and $\log_{10} R_{\text{cut}}/V$. For this re-scan, we used steps of 0.02 for the abundances at the source. Hence, the parameter estimation was obtained through the marginalization of the posterior probability assuming flat priors for the parameters.

In the following, we present the results we obtained for each dataset and a summary can be found on tables 8.1 and 8.2.

ICRC 2015 data

As a sanity check, we performed the analysis of the ICRC 2015 data, under the same assumptions made in [48] and found no significant difference between our results and the ones reported in the aforementioned work. The values of $\sqrt{D - D_{\text{min}}}$ for the optimal abundances as a function of γ and $\log_{10}(R_{\text{cut}})$ are shown in the left panel of figure 8.1. We also found two minima with similar values of deviance, that can be seen on the plot. On the right panel of figure 8.1, we show the energy spectrum obtained for the best fit. The X_{max} distributions are shown in figure 8.2. The best-fit parameters are presented on tables 8.1 and 8.2.

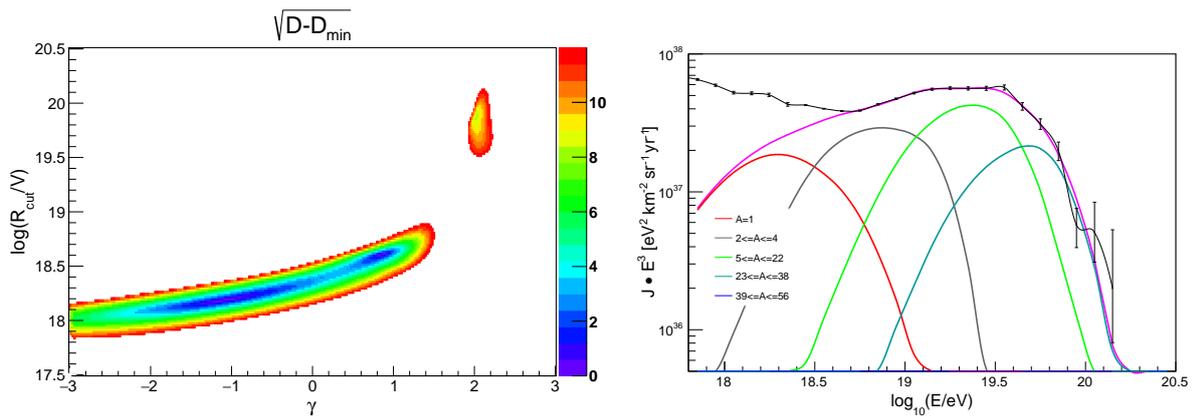


Figure 8.1: *Left*: Minimum values of $\sqrt{D - D_{\text{min}}}$ as a function of $(\gamma, \log_{10}(R_{\text{cut}}/V))$ obtained using the package Minuit. *Right*: Energy spectrum of UHECRs at Earth (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2015 dataset.

CHAPTER 8. COMBINED FIT OF THE ENERGY SPECTRUM, MASS COMPOSITION AND ARRIVAL DIRECTIONS DETECTED AT THE PIERRE AUGER OBSERVATORY

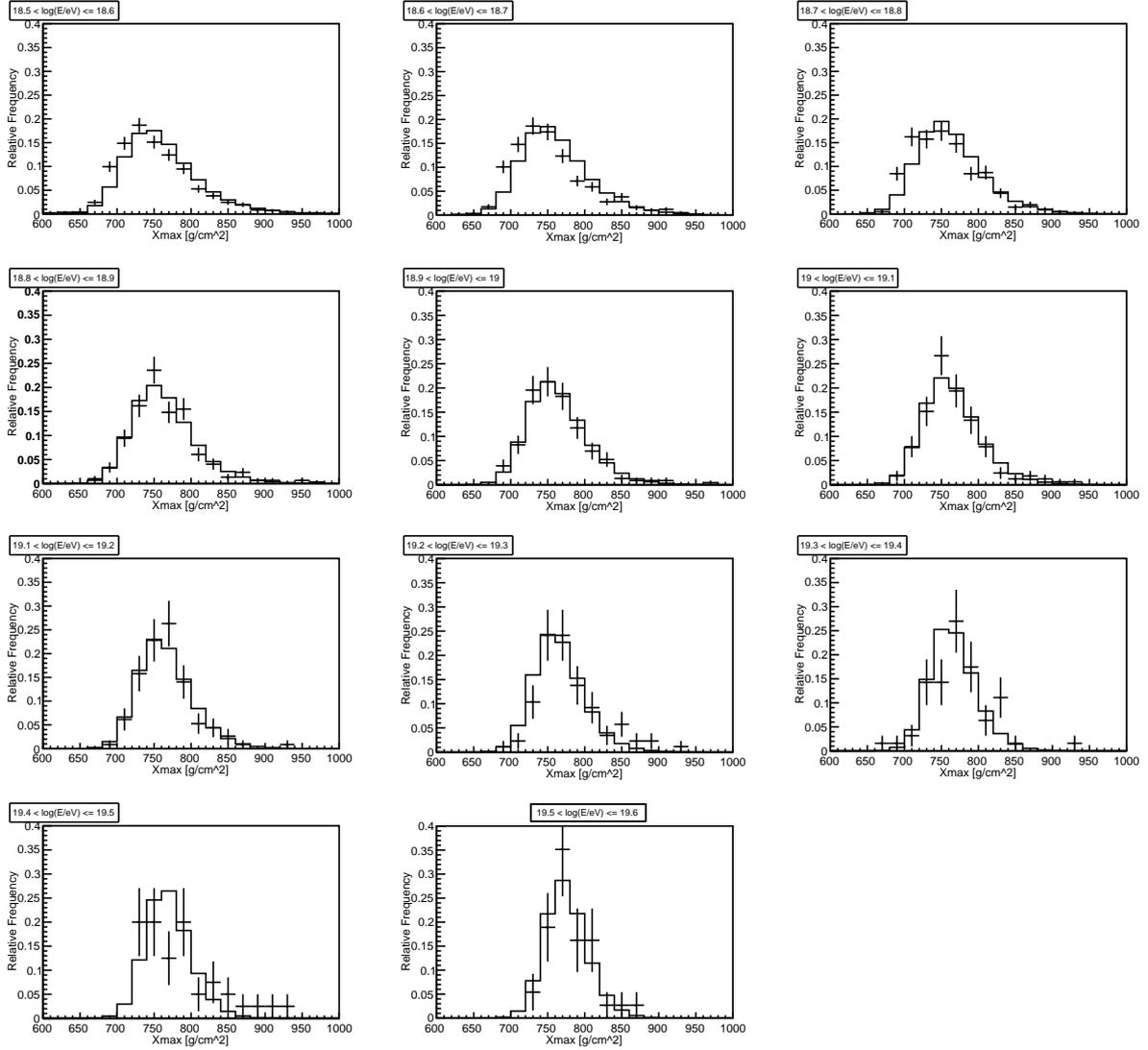


Figure 8.2: X_{\max} distributions from the model for the best fit (histogram) and data (profile histogram) for different energy bins. ICRC 2015 dataset.

ICRC 2019 data

The values of the deviance are shown in figure 8.3. For this dataset, the *second minimum* (with positive γ) does not appear, and the values of the deviance from both the spectrum and the X_{\max} show a significant increase when compared to the ICRC 2015 result (see tables 8.1 and 8.2). The corresponding best-fit spectrum is shown in the right panel of figure 8.3 and the X_{\max} distributions in figure 8.4.

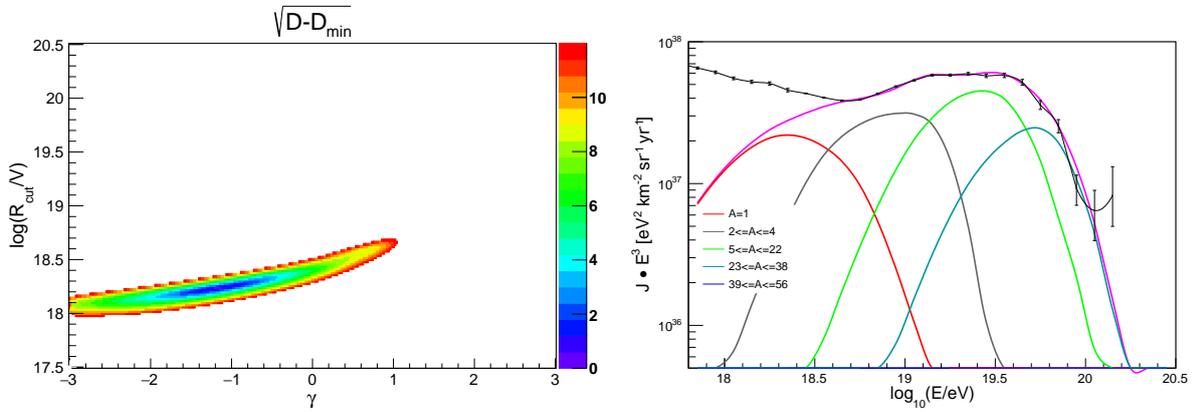


Figure 8.3: *Left*: Minimum values of $\sqrt{D - D_{\min}}$ as a function of $(\gamma, \log_{10} R_{\text{cut}}/V)$ obtained using the package Minuit. *Right*: Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset.

Table 8.1: Summary of the γ and R_{cut} parameter estimation for the data sets considered. The values of the deviances for the best-fit parameters are also presented. For the ICRC 2015 dataset, we included the values for the two minima.

	γ	$\log_{10}(R_{\text{cut}}/V)$	D_{total}	D_J	$D_{X_{\max}}$
ICRC 2015	$0.81^{+0.11}_{-0.08}$	$18.59^{+0.03}_{-0.01}$	190.45	27.00	163.45
ICRC 2015	$-0.88^{+0.35}_{-0.44}$	$18.22^{+0.04}_{-0.07}$	190.63	8.52	180.11
ICRC 2019	$-1.07^{+0.25}_{-0.21}$	$18.25^{+0.03}_{-0.03}$	303.03	25.64	277.39
PRL 2020	$-1.16^{+0.24}_{-0.22}$	$18.24^{+0.03}_{-0.03}$	301.43	27.93	273.50

CHAPTER 8. COMBINED FIT OF THE ENERGY SPECTRUM, MASS COMPOSITION AND ARRIVAL DIRECTIONS DETECTED AT THE PIERRE AUGER OBSERVATORY

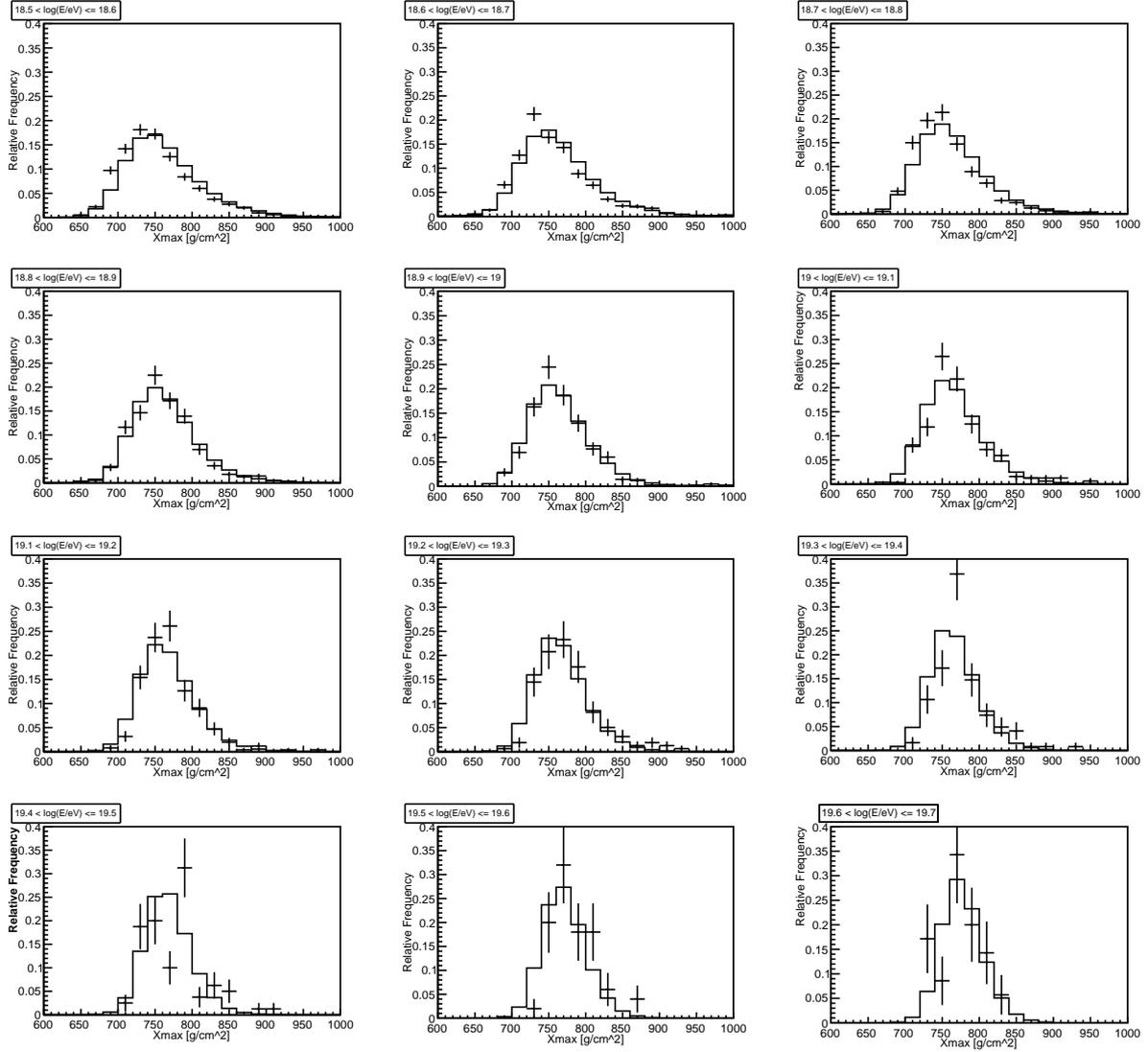


Figure 8.4: X_{\max} distributions from the model for the best fit (histogram) and data (profile histogram) for different energy bins. ICRC 2019 dataset.

Table 8.2: Summary of the abundance parameter estimation for the data sets considered. For the ICRC 2015 dataset, we included the values for the two minima.

	a_{H}	a_{He}	a_{N}	a_{Si}	a_{Fe}
ICRC 2015	$0.08^{+0.02}_{-0.08}$	$0.15^{+0.03}_{-0.15}$	$0.70^{+0.15}_{-0.08}$	$0.09^{+0.02}_{-0.02}$	$0.00^{+0.01}_{-0.00}$
ICRC 2015	$0.65^{+0.19}_{-0.07}$	$0.33^{+0.06}_{-0.17}$	$0.02^{+0.01}_{-0.02}$	$0.00^{+0.00}_{-0.00}$	$0.00^{+0.01}_{-0.00}$
ICRC 2019	$0.66^{+0.14}_{-0.08}$	$0.33^{+0.08}_{-0.13}$	$0.02^{+0.01}_{-0.01}$	$0.00^{+0.00}_{-0.00}$	$0.00^{+0.01}_{-0.00}$
PRL 2020	$0.57^{+0.18}_{-0.11}$	$0.41^{+0.10}_{-0.17}$	$0.02^{+0.01}_{-0.01}$	$0.00^{+0.01}_{-0.00}$	$0.00^{+0.01}_{-0.00}$

PRL 2020 data

We also performed the analysis for the PRL 2020 spectrum (with the ICRC 2019 X_{max} data). In this case, the result is similar to the ICRC 2019 one, with the value of γ being more negative, the maximum rigidity slightly lower and the abundances at the source also favoring lighter elements. The plots of the deviance and best-fit spectrum are presented in figure 8.5, and the X_{max} distributions are shown in figure 8.6.

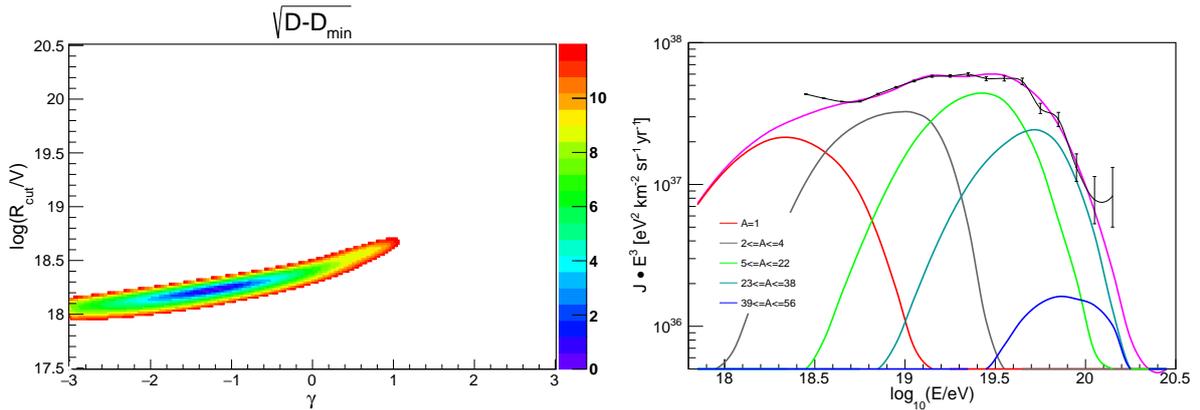


Figure 8.5: *Left*: Minimum values of $\sqrt{D - D_{\text{min}}}$ as a function of $(\gamma, \log_{10}(R_{\text{cut}}/V))$ obtained using the package Minuit. *Right*: Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. PRL 2020 dataset.

CHAPTER 8. COMBINED FIT OF THE ENERGY SPECTRUM, MASS COMPOSITION AND ARRIVAL DIRECTIONS DETECTED AT THE PIERRE AUGER OBSERVATORY

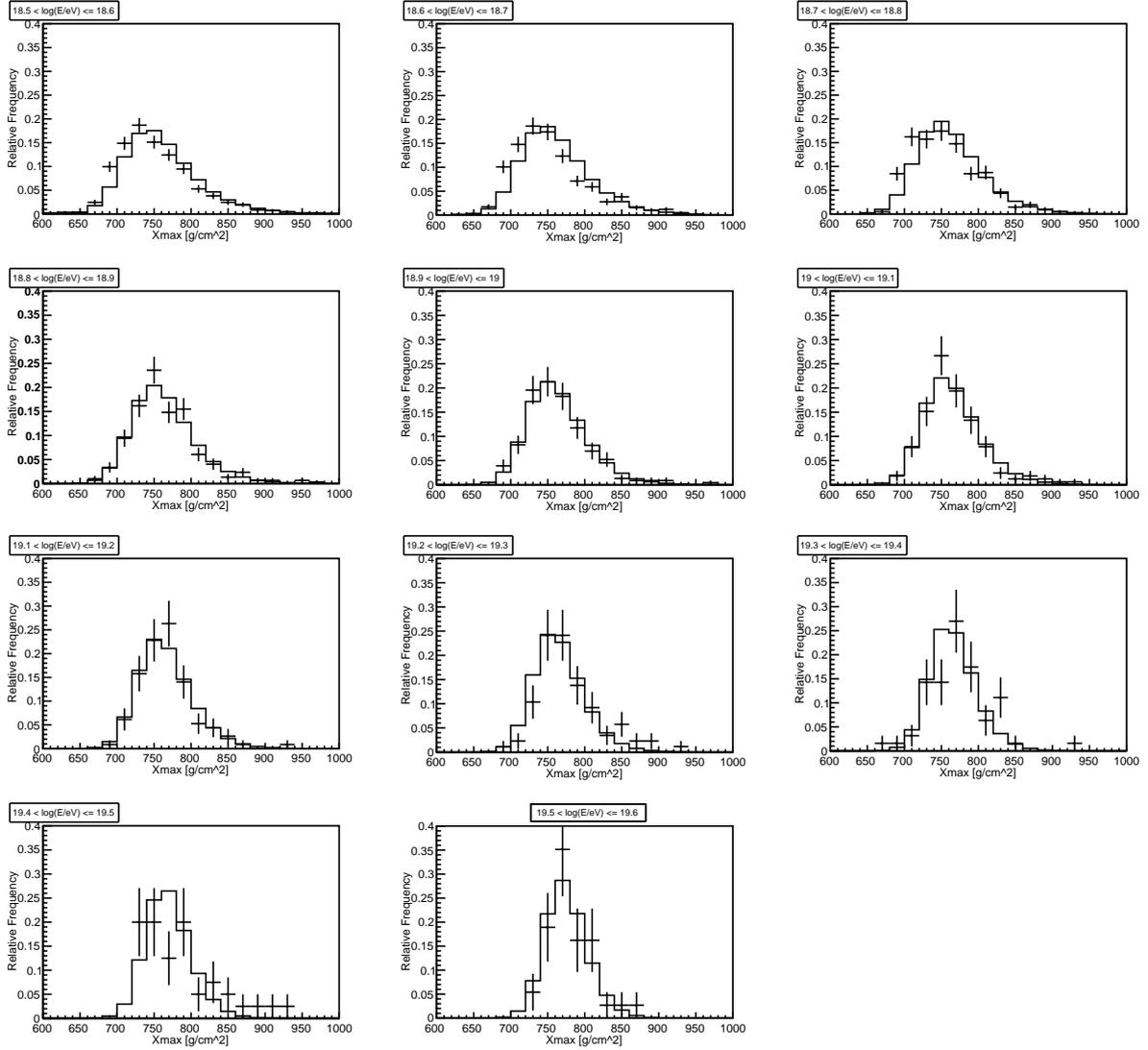


Figure 8.6: X_{\max} distributions from the model for the best fit (histogram) and data (profile histogram) for different energy bins. ICRC 2015 dataset.

8.1.7 Conclusions

We developed a package with the goal of contributing to the combined fit task and have shown that it reproduces the results presented for the ICRC 2015 dataset. We were able to reproduce the two minima for the ICRC 2015 data with compatible best-fit values and the increase on the deviance values for the newer data, although there is some discrepancy in the values of γ when compared to [76].

8.2 Combined fit with addition of anisotropy information

The combined fit of the energy spectrum and mass composition measured by the Pierre Auger Collaboration, described in the previous section, start from a simplistic assumption of sources uniformly distributed and does not take into account the significant large-scale anisotropy [53, 78, 54] measured by the Auger Collaboration. Therefore, in this chapter, we present a first trial to include such information in the global fit. The Auger anisotropy measurement that we want to fit is the dipole amplitude as a function of energy. They were reported for 4 energy bins, $4 \leq E/\text{EeV} < 8$, $8 \leq E/\text{EeV} < 16$, $16 \leq E/\text{EeV} < 32$ and $E \geq 32 \text{ EeV}$ [54], and are presented in Fig. 3.12. However, since the Auger Combined fit is performed for energies above 5 EeV, the first anisotropy bin in our analysis corresponds to the energy range $5 \leq E/\text{EeV} < 8$.

Instead of considering sources uniformly distributed, we now consider that the cosmic-ray sources follow the local matter distribution for distances $r \leq 285 \text{ Mpc}$, being uniformly distributed for $r > 285 \text{ Mpc}$. Our local matter distribution representation follows a catalog of stellar masses reported in [86], based on the 2MASS spectroscopic and photometric redshift surveys.

We neglect the effects of the local source positions on the energy spectrum and X_{max} distributions at Earth, so that the fits concerning both measurements are computed in the same way described in section 8.1. Therefore, the only impact of the local sources

considered in our model is over the expected dipole amplitudes as a function of energy at Earth. In order to take it into account, we add a deviance term related to the anisotropy measurements D_{dip} in the total deviance D , so that Eq. 8.2 is now written as:

$$D \equiv D_J + D_{X_{\text{max}}} + D_{\text{dip}}, \quad (8.8)$$

with D_{dip} given by

$$D_{\text{dip}} = \sum_{i=1}^{i=4} \left(\frac{d_{\text{mod},i} - d_{\text{data},i}}{\sigma(d_{\text{data},i})} \right)^2. \quad (8.9)$$

In this equation, $d_{\text{data},i}$ and $\sigma(d_{\text{data},i})$ correspond to the dipole amplitude and its uncertainty measured by the Auger Collaboration, while $d_{\text{mod},i}$ is the expected dipole amplitude considering a given astrophysical model with respect to an energy bin i . The expected sky map for a given set of input parameters is given by

$$\Phi(\hat{n}) = (1 - f_{\text{iso}})\Phi_{LM}(\hat{n}) + f_{\text{iso}}\Phi_{\text{iso}}(\hat{n}), \quad (8.10)$$

where $\Phi_{LM}(\hat{n})$ corresponds to the flux of cosmic rays originated from sources distributed as our local matter, and $\Phi_{\text{iso}}(\hat{n})$ corresponds to the isotropic flux originated from the contributions of the uniformly distributed sources with $r > 285$ Mpc. f_{iso} is the fraction of the measured flux that comes from sources located at $r > 285$ Mpc. It depends on γ , R_{cut} and nuclei abundances at sources and, for each combination of those parameters, it is computed by

$$f_{\text{iso}}(\gamma, R_{\text{cut}}, a_i) = \frac{\int_{285\text{Mpc}}^{\infty} \frac{dN}{dr} dr}{\int_0^{\infty} \frac{dN}{dr} dr}, \quad (8.11)$$

where dN/dr is the histogram of the source distances of the simulated events detected at Earth.

Since $\Phi_{LM}(\hat{n})$ depends on the same parameters, we performed the following approach. First, for each nucleus, we simulated the extragalactic propagation of 10^6 particles

injected at sources located in our local Universe by using the CRPropa3 [67] code and the catalog of sources reported in [86]. Since the simulated energy of events at the source follows a flat distribution in $\log(E_{\text{inj}}/\text{eV})$, the corresponding spectrum taking into account γ and R_{cut} was obtained by assigning a weight $\propto E_{\text{inj}}^{1-\gamma}$ to each simulated nucleus that arrived at Earth. Besides, in order to account for the magnetic deflection during the particle propagation, the arrival direction of each event is randomized by using a von Mises-Fisher distribution [87] corresponding to a Gaussian smearing of resolution $\sigma = \frac{10\sigma_0}{R}$. Since $R = \frac{E}{Z}$ is the particle magnetic rigidity at the entry in our Galaxy, σ_0 corresponds to the deflection that a proton with energy of 10 EeV would experience in average. Due to the uncertainties in the magnitude of the Galactic and extragalactic magnetic fields, we performed the combined fits for $\sigma_0 = 10^\circ$, $\sigma_0 = 20^\circ$ and $\sigma_0 = 25^\circ$. By scanning γ and R_{cut} in spaces of 0.04 and 0.02, respectively, we produced a total of 22,801 $\Phi_{LM,j}(\hat{n})$ maps for each nucleus and energy bin. Figures 8.7 and 8.8 show, as illustration, the model flux maps $\Phi_{LM,j}(\hat{n})$ obtained for different nuclei considering $(\gamma, R_{\text{cut}}) = (-1, 18.2)$ and $(\gamma, R_{\text{cut}}) = (-1, 20.0)$, respectively, for $E > 32$ EeV and $\sigma_0 = 20^\circ$.

Finally, for a given set of parameters $(\gamma, R_{\text{cut}}, a_i)$, Φ_{LM} is computed as

$$\Phi_{LM}(\hat{n})(\gamma, R_{\text{cut}}, a_i) = \frac{\sum_{j=1}^5 a_j \times \mathcal{N}_j \times \Phi_{LM,j}(\hat{n})}{\sum_{j=1}^5 a_j \times \mathcal{N}_j}. \quad (8.12)$$

The term \mathcal{N}_j accounts for the different number of events at Earth for each nucleus injected at the sources resulting from the nuclear fragmentation. The model dipole amplitudes for each energy bin, $d_{\text{mod},i}$, necessary to compute the anisotropy deviance D_{dip} (eq. 8.9), are obtained from the model sky map $\Phi(\hat{n})$ (eq. 8.10) after computing $\Phi_{LM}(\hat{n})$ (eq. 8.12). However, before obtaining the dipole amplitudes, it is necessary to multiply $\Phi(\hat{n})$ by the exposure of the Pierre Auger Observatory $\omega(\hat{n})$ to mimic the cosmic-ray sky as observed in the experiment.

To reconstruct the dipole from an observed sky map with partial sky exposure, one can use, for instance, the Rayleigh analysis or the K-inverse method [88]. Both methods give similar results. The Rayleigh one is preferred in the Auger analysis of real data because

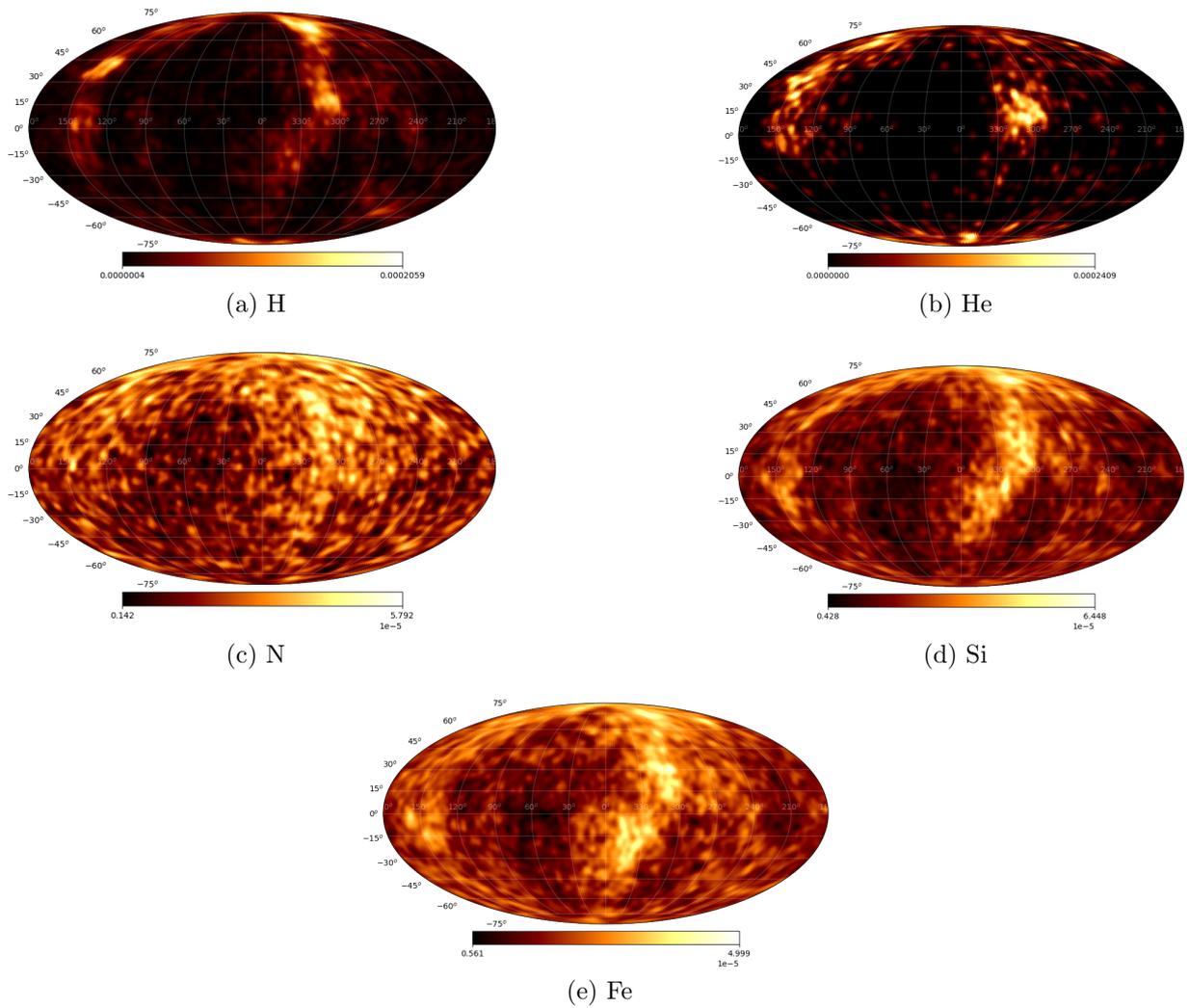


Figure 8.7: Expected flux maps $\Phi_{LM,j}(\hat{n})$ for detected events with $E > 32$ EeV resulting from the injection of different nuclei with $(\gamma, R_{\text{cut}}) = (-1.0, 18.2)$ at sources distributed as our local matter.

CHAPTER 8. COMBINED FIT OF THE ENERGY SPECTRUM, MASS
COMPOSITION AND ARRIVAL DIRECTIONS DETECTED AT THE PIERRE
AUGER OBSERVATORY

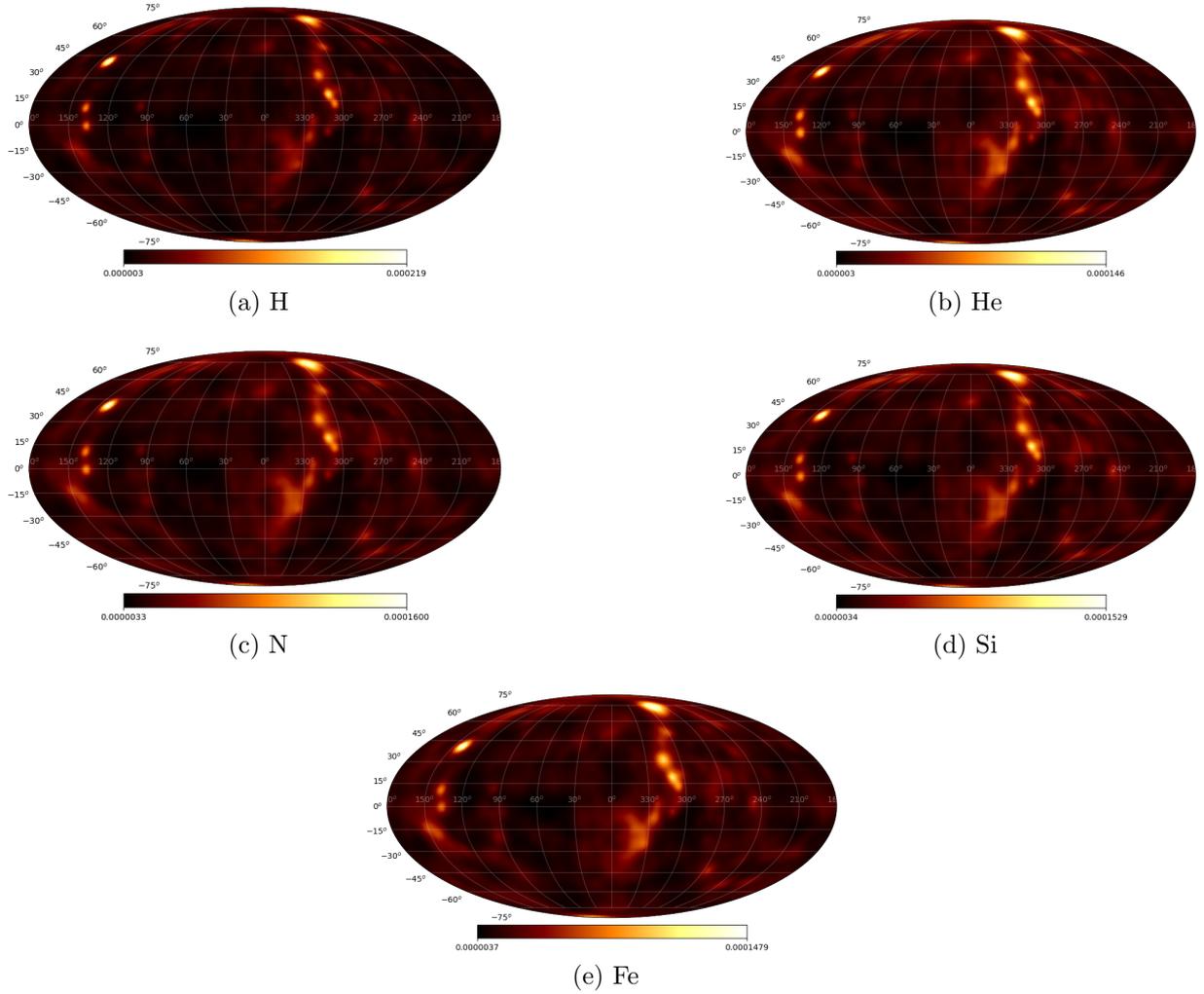


Figure 8.8: Expected flux maps $\Phi_{LM,j}(\hat{n})$ for detected events with $E > 32$ EeV resulting from the injection of different nuclei with $(\gamma, R_{\text{cut}}) = (-1.0, 20)$ at sources distributed as our local matter.

it is less sensitive to the systematics of the experiment. In our case, we decided to use the K-inverse method to reconstruct the dipole since we work with a full-sky model and it is more straightforward. We can write the cosmic-ray flux in a given direction \vec{n} in terms of spherical harmonics $Y_{\ell m}(\vec{n})$

$$\Phi(\vec{n}) = \sum_{\ell=0}^{\ell_{max}} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\vec{n}), \quad (8.13)$$

where the coefficients $a_{\ell m}$ have all the information about a possible anisotropy presented in the cosmic-ray flux. In case of a partial sky coverage, they are obtained from [89]

$$a_{\ell m} = \sum_{\ell'=0}^{\ell_{max}} \sum_{m'=-\ell'}^{\ell'} [K_{\ell_{max}}^{-1}]_{\ell m}^{\ell' m'} \tilde{a}_{\ell' m'}, \quad (8.14)$$

where the $\tilde{a}_{\ell' m'}$ are the pseudo-harmonic coefficients, and the deconvolution Kernel matrix truncated to an upper bound ℓ_{max} is given by

$$[K_{\ell_{max}}^{-1}]_{\ell m}^{\ell' m'} = \int_{\Delta\Omega} d\Omega Y_{\ell' m'}(\vec{n}) \omega(\vec{n}) Y_{\ell m}^*(\vec{n}). \quad (8.15)$$

In ref. [88], it is shown that the resolution in $a_{\ell m}$ increases exponentially as we increase the ℓ_{max} of the expansion, being this reason responsible for $\ell_{max} = 1$ used in the Auger analysis. Once one has obtained the a_{1m} coefficients, the computation of the dipole amplitude d is straightforward by

$$d = \frac{\sqrt{3(a_{10}^2 + a_{11}^2 + a_{1-1}^2)}}{a_{00}}. \quad (8.16)$$

With this result, we have all the necessary information to compute the total deviance D , given by Eq. 8.8.

8.2.1 Results

For the analysis described in this section, we used the energy spectrum and X_{\max} distributions reported in ICRC 2019 [83]. We choose the CTGE model as our reference, i.e., we use events simulated with CRPropa3 [67], with the TALYS [69] photo-disintegration cross-section and the extragalactic background light spectrum described by Gilmore [72]. The conversion from nuclei with energies E arriving at Earth to X_{\max} distributions is performed by using the Gumble distributions parametrized by the EPOS-LHC [49] hadronic interaction model and, the event arrival directions are randomly smeared by using $\sigma_0 = 10^\circ, 20^\circ$ and 25° . Table 8.3 summarizes the results of fit.

Table 8.3: Best-fit results obtained by including anisotropies in the fit for different smearing parameters σ_0 ($10^\circ, 20^\circ$ and 25°).

Smearing	10°	20°	25°
γ	-1.40	-1.28	-1.36
$\log_{10}(R_{\text{cut}}/V)$	18.18	18.20	18.20
$a_H(\%)$	85.70	83.10	84.50
$a_{He}(\%)$	13.60	16.00	14.80
$a_N(\%)$	0.66	0.82	0.71
$a_{Si}(\%)$	0.01	0.02	0.02
$a_{Fe}(\%)$	0	0	0
$D_{X_{\max}}(N_{X_{\max}})$	283.07 (121)	282.68 (121)	279.17 (121)
$D_J(N_J)$	62.46 (15)	59.46 (15)	60.19 (15)
$D_{\text{dip}}(N_{\text{dip}})$	93.94 (4)	24.44 (4)	11.03 (4)
$D_{\text{total}}(N)$	439.47 (140)	366.58 (140)	350.39 (140)

The values of $\sqrt{D - D_{\min}}$ for the optimal abundances as a function of γ and $\log_{10}(R_{\text{cut}})$ are shown in the panel (a) of figure 8.9. In the panel (b) of the same figure, we show the energy spectrum obtained for the best fit while in panel (c) we present

the corresponding dipole amplitudes as a function of energy. The X_{\max} distributions corresponding to the best fit are shown in figure 8.10.

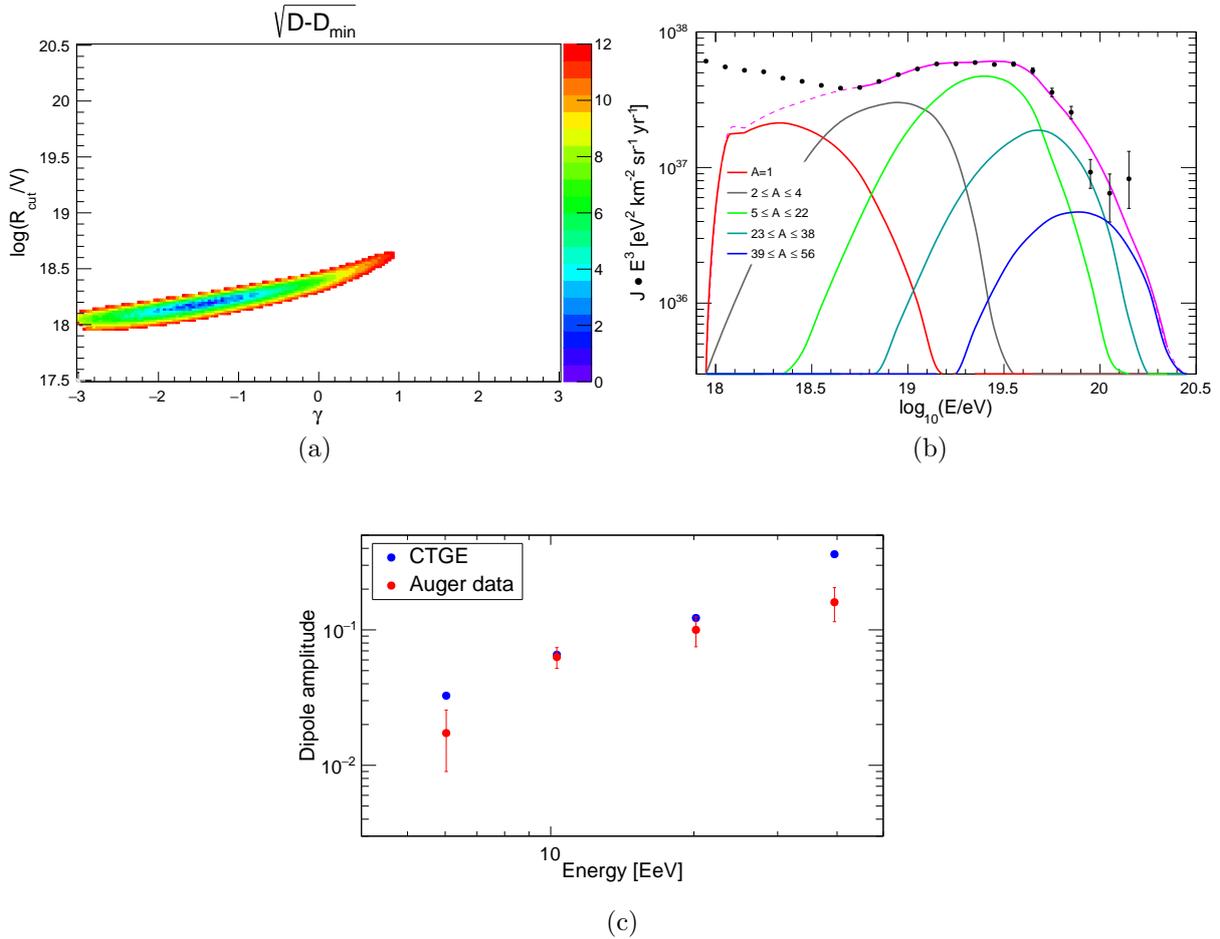


Figure 8.9: (a) Minimum values of $\sqrt{D - D_{\min}}$ as a function of $(\gamma, \log_{10} R_{\text{cut}}/V)$ obtained using Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges: H (red), He (gray), N (green), Si (cyan) and Fe (blue). (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red).

We notice that while the position of the best fit is not very different from the one without taking into account the dipole amplitude measurements, the deviance of the energy spectrum D_j worsens (from $D_j = 25.64$ to $D_j = 59.46$) in order to accommodate better the anisotropy results. The deviance of X_{\max} , $D_{X_{\max}}$, is almost the same, while the deviance with respect to the anisotropy is $D_{\text{dip}}/N_{\text{dip}} \sim 6$. In general, the combined fit results have a

CHAPTER 8. COMBINED FIT OF THE ENERGY SPECTRUM, MASS
COMPOSITION AND ARRIVAL DIRECTIONS DETECTED AT THE PIERRE
AUGER OBSERVATORY

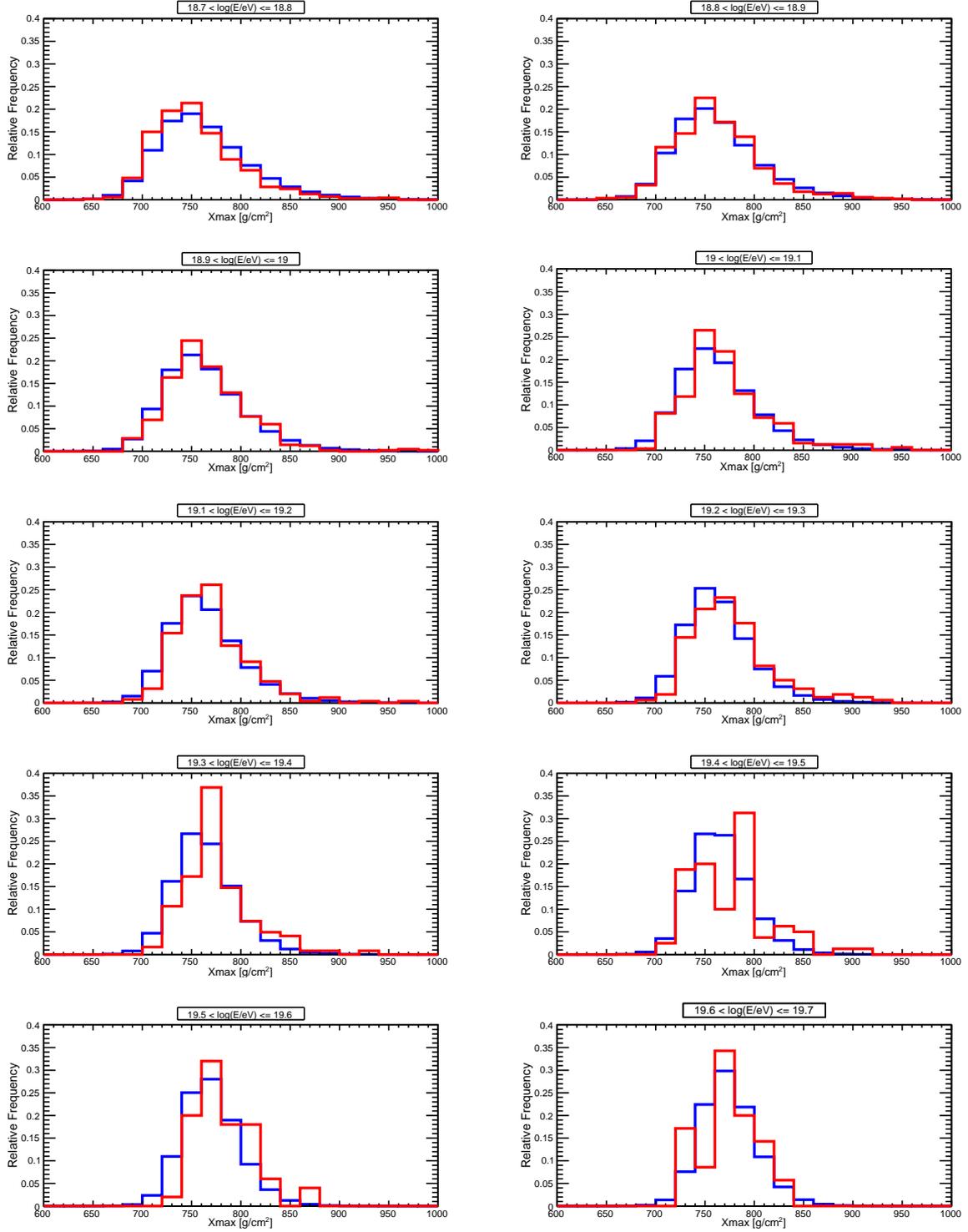


Figure 8.10: X_{\max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins.

good description to the dipole amplitudes for the most significant bins, $8 \leq E/\text{EeV} < 16$ and $16 \leq E/\text{EeV} < 32$, and to the last bin $E > 32 \text{ EeV}$ predict a higher dipole amplitude than the one corresponding to the data. We notice that, in this energy range, the main contribution to the flux comes from injected nitrogen. Since the best-fit maximum rigidity cut $\log_{10}(R_{\text{cut}}/V) = 18.2$ is too low, the contribution of nearby sources is high, implying in a strong large-scale anisotropy. Table 8.3 also shows the obtained results by considering the smearing parameters $\sigma_0 = 10^\circ$ and $\sigma_0 = 25^\circ$. Figure 8.11 shows the deviances divided by the number of experimental points (a), the changes in the best-fit position (b) and the dipolar amplitudes as a function of energy for the three considered values of σ_0 . One can see that the deviances D_J and $D_{X_{\text{max}}}$ as well as other parameters of best fit are essentially unaffected when we change σ_0 .

Therefore, considering this result, assuming that the σ_0 value does not change the best-fit parameters, we computed the deviance of anisotropy divided by the number of experimental points $D_{\text{dip}}/N_{\text{dip}}$ as a function of σ_0 by performing a scan in σ_0 from $\sigma_0 = 10^\circ$ to $\sigma_0 = 70^\circ$. The results can be seen in figure 8.12. The minimum is obtained for $\sigma_0 = 33^\circ$, corresponding to a $D_{\text{dip}}/N_{\text{dip}} = 2.95$.

Effect of the Source Evolution

The results presented before were obtained by assuming no cosmological evolution for the populations of extragalactic sources. We perform the fit also assuming a distribution of sources whose comoving emissivity evolves with redshift z as $\dot{\epsilon} = \dot{\epsilon}_0(1+z)^m$, where m is the source evolution parameter. The emissivity, by definition, accounts for both the effects of source density and luminosity. Beyond $m = 0$, already reported, we have also considered the cases of $m = -3$ and $m = +3$, representative values for Tidal Disruption Events (TDE) [90, 91] and star formation rate (SFR) evolutions [92, 93], respectively. Table 8.4 summarizes the best-fit parameters obtained for the different source evolution cases and $\sigma_0 = 20^\circ$. One can see that, in order to compensate the larger amount of low (high) energy particles, a positive (negative) source evolution produces a hardening (softening) of the

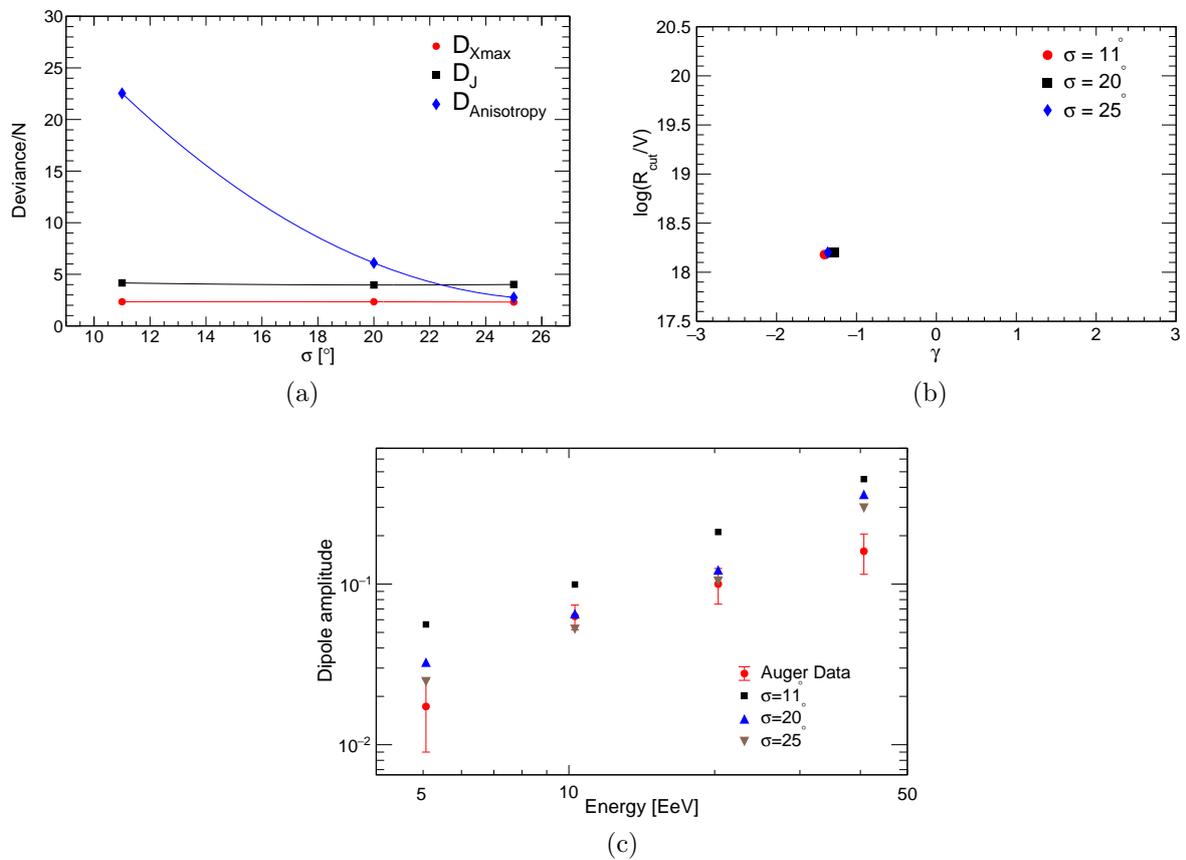


Figure 8.11: (a) All Deviances (D_J , $D_{X_{\max}}$, D_{dip}) divided by the number of experimental points as a function of points for values of $\sigma_0 = 10^\circ$, 20° and 25° . (b) $\log_{10}(R_{\text{cut}})$ as a function of γ for three values of σ_0 considered. (c) Dipolar amplitudes as a function of energies for the three considered values of σ_0 . (CTGE)

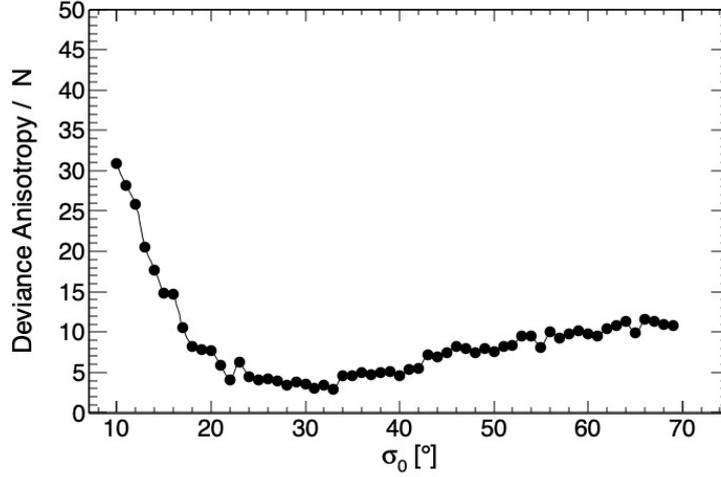


Figure 8.12: Deviance of anisotropy divided by the number of experimental points D_{dip}/N as a function of σ_0 by performing a scan in σ_0 . The minimum is obtained for $\sigma_0 = 33^\circ$ corresponding to a $D_{\text{dip}}/N = 2.95$.

Table 8.4: Best-fit results in the scenario obtained by including anisotropies in the fit EPOS-LHC with smearing (20°).

Source evolution	-3	0	+3
γ	-0.56	-1.28	-2.16
$\log_{10}(R_{\text{cut}} / V)$	18.28	18.20	18.14
$a_H(\%)$	71.00	83.10	83.00
$a_{He}(\%)$	27.00	16.00	16.80
$a_N(\%)$	1.90	0.82	0.60
$a_{Si}(\%)$	0.01	0.02	0.1
$a_{Fe}(\%)$	0	0	0
$D_{X_{\text{max}}}(N_{X_{\text{max}}})$	293.17 (121)	282.68 (121)	279.17 (121)
$D_J(N_J)$	55.97 (15)	59.46 (15)	60.19 (15)
$D_{\text{dip}}(N_{\text{dip}})$	41.27 (4)	24.44 (4)	11.03 (4)
$D_{\text{total}}(N)$	390.41 (140)	366.58 (140)	350.39 (140)

energy spectrum at the source, in agreement with previous results [48, 66]. The minimum of the total deviance D is obtained for $m = +3$ essentially because of the better fit of the dipole amplitude measurements. We also performed a scan in different values of σ_0 , keeping the same best-fit parameters for each value of m . The left panel of figure 8.13 presents the deviance of the anisotropy as a function of σ_0 for $m = -3$, $m = 0$ and $m = +3$. The dipole amplitudes as a function of energy for the corresponding optimal value of σ_0 , for each m , are presented in the right panel of the same figure. We can see that the best fit, after considering the scan in σ_0 , is obtained for $m = -3$ and $\sigma_0 = 32^\circ$, corresponding to $D/N = 2.13$. As for the case without source evolution, the main discrepancy between data and model is the very high dipole amplitude expected for the model for $E > 32$ EeV due to the small rigidity cutoff.

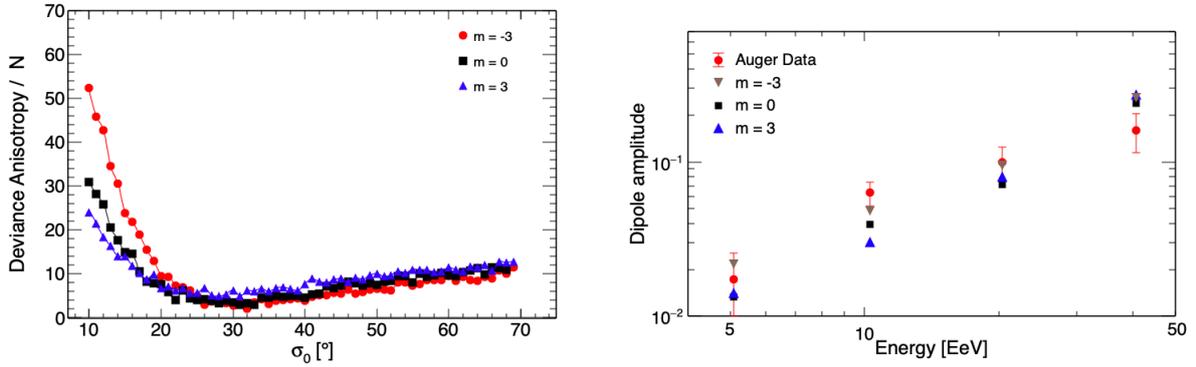


Figure 8.13: *Left*: The deviance of the anisotropy as a function of σ_0 considering source evolution for $m = -3$, $m = 0$ and $m = +3$. *Right*: The dipole amplitudes as a function of energy for the corresponding optimal value of σ_0 , for each case of source evolution m .

Systematic uncertainties on X_{\max}

The systematic uncertainties of instrumental origin affect both the measurements of the energy and the X_{\max} . The uncertainty on the energy scale is assumed to be $\Delta E/E = 14\%$ [55], while the one in X_{\max} ranges from 6 to 9 g/cm² [59]. In general, the systematic uncertainties on both measurements are taken into account by shifting all the measured

energies and X_{\max} values by one systematic standard deviation in each direction. This is a very important investigation and remains to be done in a near future.

In this work, we only model uncertainties on the reconstructed X_{\max} arising either from the systematic uncertainties on the measured value or from variations due to the hadronic model considered by shifting the Gumbel model by a quantity δ . This quantity is a new parameter in our combined fit. In this way, $G_{j,x}^{model} \rightarrow G_{j,x+\delta}^{model}$, where δ represents a shift of δ in the X_{\max} bin. Besides, large deviations from $\delta = 0$ g/cm² are penalized by including a new term in the deviance, $D_\delta = \delta^2/\sigma_{syst}^2$, with $\sigma_{syst} = 8$ g/cm². We summarize the best-fit results including the systematic uncertainties in X_{\max} , for different values of source evolution, in Table 8.5.

Table 8.5: Best-fit results in the scenario including anisotropies in the fit EPOS-LHC by considering a systematic shift on X_{\max} with (20°).

Source evolution	-3	0	+3
γ	1.44	1.12	0.76
$\log_{10}(R_{\text{cut}}/V)$	18.56	18.58	18.44
$a_H(\%)$	0.30	0.01	63.50
$a_{He}(\%)$	43.80	29	0.30
$a_N(\%)$	36.50	49.90	28.00
$a_{Si}(\%)$	17.50	20.7	8.0
$a_{Fe}(\%)$	2	0.2	0.8
$D_{X_{\max}}(N_{X_{\max}})$	233.71 (121)	234.70 (121)	246.57 (121)
$D_J(N_J)$	51.29 (15)	41.60 (15)	59.51(15)
$D_{\text{dip}}(N_{\text{dip}})$	20.91 (4)	27.9 (4)	25.70 (4)
$D_{\text{total}}(N)$	305.91 (140)	304.21(140)	331.78 (140)
δ (g/cm ²)	-16.21	-13.90	-17.63

One can notice a better description of the X_{\max} distributions, reflected by smaller

value of $D_{X_{\max}}$ in comparison with the results without shifting X_{\max} . On the other hand, it would be necessary to shift the X_{\max} distribution by $\sim -16 \text{ g / cm}^2$, which corresponds to $\sim 2\sigma_{\text{sys}}$. So, considering a gate function for the description of the systematic uncertainty, this result is not satisfactory. Furthermore, the dipole amplitude measurements by Auger are equally not well described by the models, as one can notice from the high values of D_{dip} .

Effect of the choice of hadronic interaction model

We also studied the impact on the fit results of changing the hadronic interaction model. For this goal, only for the case without source evolution ($m = 0$), we used the Sibyll 2.3d [75] hadronic interaction model instead of EPOS-LHC. We refer to this model as CTGS. We also considered the cases of $\sigma_0 = 10^\circ$ and $\sigma_0 = 20^\circ$, as well as the possible systematic shift in X_{\max} distributions. Table 8.6 summarizes the results of the best fit for both smearing parameters. One can notice an increase in the contribution of He at sources with a worsening of the energy spectrum deviance. The anisotropy, on the other hand, is best fitted in a scenario with a X_{\max} shift $\delta = -20 \text{ g / cm}^2$ and $\sigma_0 = 10^\circ$. Figure 8.14 presents the dipole amplitudes as a function of energy for this specific case.

8.2.2 Conclusions

In this section, we investigated the impact of adding anisotropy information to the Auger global combined fit. In order to take into account our poor knowledge about the magnitude of the Galactic and extragalactic magnetic fields, we investigated the effect of the smearing angle deflection in the anisotropy deviance. Besides, we also studied the impact of the source evolution, hadronic interaction model and a possible shift of the X_{\max} distributions in the combined fit. Although the position of the best fit in the plane $(\gamma, \log_{10}(R_{\text{cut}}))$ does not change too much, the deviance of the energy spectrum worsens in order to accommodate better the fit of the arrival direction measurements. Although the combined fit results in a good description of dipole amplitudes for the most significant

CHAPTER 8. COMBINED FIT OF THE ENERGY SPECTRUM, MASS COMPOSITION AND ARRIVAL DIRECTIONS DETECTED AT THE PIERRE AUGER OBSERVATORY

Table 8.6: Best-fit results obtained with Sibyll 2.3d hadronic interaction model, considering or not a shift in the X_{\max} distribution with different smearing parameters (10° , 20°).

Smearing	10°	20°	10° with X_{\max} shift	20° with X_{\max} shift
γ	-1.64	-0.6	1.48	1.48
$\log_{10}(R_{\text{cut}} / V)$	18.04	18.24	18.48	18.52
$a_H(\%)$	51.91	49.00	14.30	0.27
$a_{He}(\%)$	46.87	47.80	28.23	43.34
$a_N(\%)$	1.20	3.20	33.40	24.58
$a_{Si}(\%)$	0.02	0.20	18.80	28.00
$a_{Fe}(\%)$	0.0	0.10	5.0	3.80
$D_{X_{\max}}(N_{X_{\max}})$	290.58 (121)	295.58 (121)	221.93 (121)	225.19 (121)
$D_J(N_J)$	88.09 (15)	60.63 (15)	59.35 (15)	52.84 (15)
$D_{\text{dip}}(N_{\text{dip}})$	42.87 (4)	19.76 (4)	17.66 (4)	23.28 (4)
$D_{\text{total}}(N)$	421.540 (140)	375.97 (140)	299.02 (140)	301.20 (140)
δ (g/cm ²)	-	-	-20	-19.14

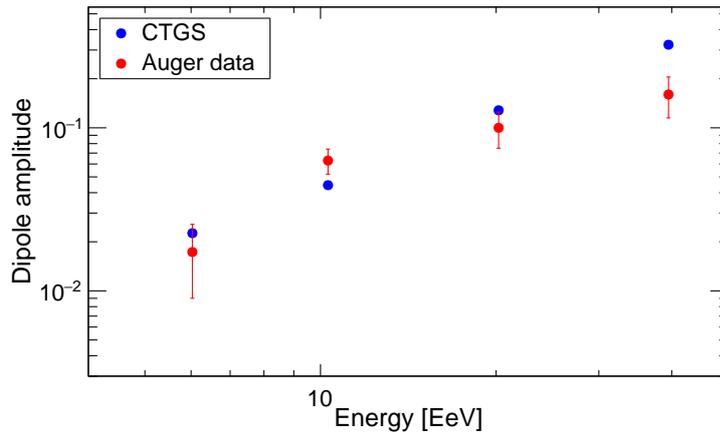


Figure 8.14: Dipole amplitudes as a function of energy expected from the CTGS model (in blue) and measured by the Pierre Auger Collaboration (in red) in the scenario with a X_{\max} shift $\delta = -20$ g /cm² and $\sigma_0 = 10^\circ$.

bins, $8 \leq E/\text{EeV} < 16$ and $16 \leq E/\text{EeV} < 32$, it predicts a higher dipole amplitude than the one corresponding to the data for $E > 32$ EeV. Furthermore, the differences between observation and model for $5 \leq E/\text{EeV} < 8$ could be explained by a possible superposition of Galactic and an extragalactic flux at this energy range, in such a way that the predicted extragalactic large-scale anisotropy would be canceled. The tension for $E > 32$ EeV, on the other hand, could be reduced by considering, for instance, a more realistic description of the Galactic magnetic field, a larger abundance of silicon and iron at sources or the emergence of a dominant source at these energies.

Possible extensions of this analysis include a realistic descriptions of the Galactic magnetic field, source overdensity as in [66], source emissivity proportional to its electromagnetic counterpart, larger statistics for X_{max} measurements at the highest energies (possibly by using information collected by the SD) and addition of anisotropy measurements for events with energies above 40 EeV.

Despite of this, the addition of anisotropy information to the combined fit, as well as the availability of the analysis code for collaboration, represents a necessary and fundamental step forward to a better comprehension of the Auger measurements, aiming to establish a global picture of our Universe at the highest energies.

Chapter 9

Conclusions

Since the beginning of this PhD study, in 2017, it was already clear that, due to the increasing cosmic-ray mass composition with energies and considering that the magnetic deflection in the particle trajectory is proportional to its electrical charge, only the combination of different measurements performed by the Pierre Auger Collaboration could make it possible the identification of the ultra-high energy cosmic-ray sources. From this perspective, all the analyses presented at this thesis aim to enhance our comprehension of UHECRs by using at the same time information about the cosmic-ray mass composition, energy and arrival direction.

The main contribution of our study is the development of an independent code to perform the combined fit of the energy spectrum and mass composition measured with the Pierre Auger Observatory. This code has been made available for the use and development of the entire Auger Collaboration. Additionally, we inserted in the fit the observed large-scale anisotropy reported for events with energies above 8 EeV and study the impact on the deviance of different source evolutions, magnetic deflections and hadronic interaction models. For this, we consider that the UHECRs sources are distributed as the local matter for distances $r < 285$ Mpc and homogeneously distributed otherwise. In general, for all explored scenarios, although the dipole expectations for $8 \leq E/\text{EeV} < 16$ and $16 \leq E/\text{EeV} < 32$ are in good agreement with the observed anisotropies, the deviance of the fit

is too high, being the main reason for that the very high dipole amplitude predicted by the model for $E > 32$ EeV. This tension could be reduced by considering, for instance, a more realistic description of the Galactic magnetic field, a larger abundance of Silicon and Iron at sources or the emergence of a dominant source at these energies. **Nevertheless, the inclusion of anisotropy information to the combined fit and the availability of the analysis code for the Auger Collaboration are of prime importance to establish a complete picture over these high energy particles.**

Furthermore, we estimated the increase in the power detection of the Landy-Szalay cross-correlation estimator when one uses a small sub set of data containing the highest X_{\max} events. The gain depends on the proton fraction ranging from a factor of ~ 2 to ~ 4 .

Additionally, we built mass composition maps and computed the statistical significance of larger or smaller $\langle X_{\max} \rangle$ through the celestial sphere. As a by-product of this study, we have shown that the zenith angle distribution of the cosmic rays detected by the Fluorescence Detector is independent of the detection time, indicating that the *shuffling* technique can be safely applied for those events.

From the knowledge obtained by using compositions maps, we exemplified the net gain obtained by combining both the directional intensity data and directional composition data in comparison to the standalone analysis of directional composition data only. For this, we used a cross-correlation estimator in order to characterize whether possible similarities between these data are significant or not. As a result, we observed that the power detection is 99% even for modest cases in which the significance of the first harmonic modulation of $\langle X_{\max}(\alpha) \rangle$ relative to $\langle X_{\max} \rangle$ is 1.6σ .

Finally, assuming that the cosmic rays with energies above 39 EeV, correlated with starburst galaxies, are nuclei of charge Z and energy E , according to the Waxmann and Lemoine proposition, we should observe an anisotropy in the same direction of the sky at energies Z times smaller (E/Z), due to protons originated in the same sources. Considering this, we searched for excess of events around the starburst directions at energy thresholds of $E_{\text{th}}/Z = 19.5$ EeV, 6.5 EeV, 3.0 EeV and 1.5 EeV, corresponding to $Z = 2, 6, 13$ or 26 at energies above 39 EeV. No significant excess of events was observed.

Recently the Pierre Auger Collaboration has shown that it is possible to obtain information about the cosmic-ray mass composition in a event-by-event basis by using deep neural networks and the surface detectors of the observatory, which operates almost 100% of the time. This opens a very interesting window for anisotropy studies expanding the importance of the results presented here.

In summary, the work described in this thesis was reported in 5 internal notes of the Pierre Auger Collaboration [94, 95, 96, 97, 98] and discussed in several meetings of the Pierre Auger Collaboration [99, 100, 101, 102, 103, 104]. The importance of code availability for the Collaboration can be observed, for instance, by noticing that the new feature in the energy spectrum is interpreted [12] in terms of a scenario that best fits the Auger measurements. The cross-checks performed within the Pierre Auger Collaboration are equally, being the paper [66] an example of our contribution. Finally, we expect that the analyzes described in this work contribute to a better comprehension about the UHECRs and its role in the Universe at the highest energies.

Bibliography

- [1] A. Aab *et al.*, “Deep-learning based reconstruction of the shower maximum x_{\max} using the water-cherenkov detectors of the pierre auger observatory,” *Journal of Instrumentation*, vol. 16, p. P07019, jul 2021. [2](#), [31](#), [58](#), [64](#)
- [2] S. Landy and A. Szalay, “Bias and variance of angular correlation functions,” *The Astrophysical Journal*, vol. 412, pp. 64–71, 1993. [2](#), [41](#)
- [3] E. Mayotte *et al.*, “Indication of a mass-dependent anisotropy above $10^{18.7}$ eV in the hybrid data of the Pierre Auger Observatory,” in *37th International Cosmic Ray Conference*, vol. 37 of *ICRC2021*, p. 12, July 2021. [2](#), [50](#), [52](#), [53](#)
- [4] A. Aab *et al.*, “Indication of anisotropy in the arrival directions of ultra-high-energy cosmic rays through comparison to the flux pattern of extragalactic gamma-ray sources,” *The Astrophysical Journal Letters*, vol. 853, p. L29, 2018. [3](#), [34](#), [35](#), [36](#), [42](#), [44](#), [45](#), [46](#), [69](#), [71](#), [72](#)
- [5] M. Lemoine and E. Waxman, “Anisotropy vs chemical composition at ultra-high energies,” *Journal of Cosmology and Astroparticle Physics*, vol. 11, p. 009, Nov. 2009. [3](#), [68](#), [69](#), [75](#)
- [6] C. D. Anderson, “The apparent existence of easily deflectable positives,” *Science*, vol. 76, no. 1967, pp. 238–239, 1932. [4](#)
- [7] S. H. Neddermeyer and C. D. Anderson, “Note on the nature of cosmic-ray particles,” *Physical Review Letters*, vol. 51, pp. 884–886, May 1937. [4](#)

BIBLIOGRAPHY

- [8] C. M. G. Lattes, G. P. S. Occhialini, and C. F. Powell, “Observations on the Tracks of Slow Mesons in Photographic Emulsions. 1,” *Nature*, vol. 160, pp. 453–456, 1947. [4](#)
- [9] P. Auger, P. Ehrenfest, R. Maze, J. Daudin, and R. A. Fréon, “Extensive cosmic-ray showers,” *Reviews of Modern Physics*, vol. 11, pp. 288–291, July 1939. [5](#)
- [10] J. Linsley, “Evidence for a primary cosmic-ray particle with energy 10^{20} -eV,” *Physical Review Letters*, vol. 10, pp. 146–148, 1963. [5](#)
- [11] Energy spectrum, “<http://www.physics.utah.edu/~whanlon/spectrum.html>,” Accessed 2021. [7](#)
- [12] A. Aab, P. Abreu, and et al., “Features of the energy spectrum of cosmic rays above 2.5×10^{18} ev using the pierre auger observatory,” *Physical Review Letters*, vol. 125, no. 12, 2020. [6](#), [29](#), [30](#), [44](#), [81](#), [108](#)
- [13] B. Peters, “Primary cosmic radiation and extensive air showers,” *Il Nuovo Cimento*, vol. 22, pp. 800–819, Nov. 1961. [6](#)
- [14] W. D. Apel *et al.*, “Ankle-like feature in the energy spectrum of light elements of cosmic rays observed with KASCADE-Grande,” *Physical Review D*, vol. 87, p. 081101, Apr. 2013. [6](#)
- [15] Telescope Array Collaboration, “The Cosmic Ray Energy Spectrum between 2 PeV and 2 EeV Observed with the TALE Detector in Monocular Mode,” *Astrophysical Journal*, vol. 865, p. 74, Sept. 2018. [6](#)
- [16] R. Aloisio, V. Berezhinsky, and P. Blasi, “Ultra high energy cosmic rays: implications of Auger data for source spectra and chemical composition,” *Journal of Cosmology and Astroparticle Physics*, vol. 10, p. 020, Oct. 2014. [8](#)
- [17] V. Berezhinsky, “Extragalactic cosmic rays and their signatures,” *Astroparticle Physics*, vol. 53, p. 120–129, Jan 2014. [8](#)

- [18] M. Unger, G. R. Farrar, and L. A. Anchordoqui, “Origin of the ankle in the ultrahigh energy cosmic ray spectrum, and of the extragalactic protons below it,” *Physical Review D*, vol. 92, Dec 2015. [8](#)
- [19] E. Fermi, “On the origin of the cosmic radiation,” *Physical Review*, vol. 75, pp. 1169–1174, Apr 1949. [8](#), [9](#)
- [20] AGASA, “<http://www-akeno.icrr.u-tokyo.ac.jp/AGASA/>,” Accessed 2021. [9](#)
- [21] V. Berezhinsky, M. Kachelrieß, and A. Vilenkin, “Ultrahigh energy cosmic rays without greisen-zatsepin-kuzmin cutoff,” *Physical Review Letters*, vol. 79, pp. 4302–4305, Dec 1997. [9](#)
- [22] J. Abraham, P. Abreu, *et al.*, “Upper limit on the diffuse flux of ultrahigh energy tau neutrinos from the pierre auger observatory,” *Physical Review Letters*, vol. 100, p. 211101, May 2008. [9](#)
- [23] J. Abraham *et al.*, “Upper limit on the cosmic-ray photon flux above 10^{19} eV using the surface detector of the Pierre Auger Observatory,” *Astroparticle Physics*, vol. 29, pp. 243–256, 2008. [9](#)
- [24] E. Fermi, “Galactic Magnetic Fields and the Origin of Cosmic Radiation.,” *Astrophysical Journal*, vol. 119, p. 1, Jan. 1954. [9](#)
- [25] A. Achterberg, Y. A. Gallant, J. G. Kirk, and A. W. Guthmann, “Particle acceleration by ultrarelativistic shocks: theory and simulations,” *Monthly Notices of the Royal Astronomical Society*, vol. 328, pp. 393–408, 12 2001. [10](#)
- [26] M. G. Hauser and E. Dwek, “The cosmic infrared background: Measurements and implications,” *Annual Review of Astronomy and Astrophysics*, vol. 39, no. 1, pp. 249–307, 2001. [11](#)
- [27] K. Greisen, “End to the cosmic-ray spectrum?,” *Physical Review Letters*, vol. 16, pp. 748–750, Apr 1966. [11](#)

BIBLIOGRAPHY

- [28] G. T. Zatsepin and V. A. Kuz'min, "Upper Limit of the Spectrum of Cosmic Rays," *Soviet Journal of Experimental and Theoretical Physics Letters*, vol. 4, p. 78, Aug. 1966. [11](#)
- [29] D. Harari, S. Mollerach, and E. Roulet, "On the ultrahigh energy cosmic ray horizon," *Journal of Cosmology and Astroparticle Physics*, vol. 11, p. 012, 2006. [12](#)
- [30] E. W Kolb and M. Stanley Turner, *The early universe*, vol. 69. New York Westview Press, 1994. [13](#)
- [31] R. Alves Batista, *On the cosmological propagation of high energy particles in magnetic fields*. PhD thesis, U. Hamburg, Dept. Phys., 2015. [14](#)
- [32] R. Jansson and G. R. Farrar, "A New Model of the Galactic Magnetic Field," *The Astrophysical Journal*, vol. 757, p. 14, Sept. 2012. [15](#)
- [33] S. Das, H. Kang, D. Ryu, and J. Cho, "Propagation of ultra-high-energy protons through the magnetized cosmic web," *The Astrophysical Journal*, vol. 682, pp. 29–38, jul 2008. [15](#)
- [34] P. Bhattacharjee and G. Sigl, "Origin and propagation of extremely high-energy cosmic rays," *Physics Reports*, vol. 327, p. 109–247, Apr 2000. [15](#)
- [35] P. P. Kronberg, "Extragalactic radio sources, igm magnetic fields, and agn-based energy flows," *Astronomische Nachrichten*, vol. 327, no. 5-6, pp. 517–522, 2006. [15](#)
- [36] T. E. Clarke, P. P. Kronberg, and H. Böhringer, "A new radio-x-ray probe of galaxy cluster magnetic fields," *The Astrophysical Journal*, vol. 547, pp. L111–L114, feb 2001. [15](#)
- [37] A. M. Hillas, "The Origin of Ultra-High-Energy Cosmic Rays," *Annual Review of Astronomy and Astrophysics.*, vol. 22, pp. 425–444, Jan. 1984. [16](#)

- [38] R. Alves Batista, J. Biteau, M. Bustamante, K. Dolag, R. Engel, K. Fang, K.-H. Kampert, D. Kostunin, M. Mostafa, K. Murase, and et al., “Open questions in cosmic-ray research at ultrahigh energies,” *Frontiers in Astronomy and Space Sciences*, vol. 6, Jun 2019. [16](#), [17](#)
- [39] P. Abreu *et al.*, “Interpretation of the Depths of Maximum of Extensive Air Showers Measured by the Pierre Auger Observatory,” *Journal of Cosmology and Astroparticle Physics*, vol. 02, p. 026, 2013. [18](#)
- [40] K.-H. Kampert and M. Unger, “Measurements of the cosmic ray composition with air shower experiments,” *Astroparticle Physics*, vol. 35, no. 10, pp. 660 – 678, 2012. [18](#)
- [41] P. Sommers and M. Diaz, “Cosmic rays above the knee region science issues and open questions observational evidence,” *AIP Conference Proceedings*, vol. 1123, no. 1, pp. 190–196, 2009. [18](#), [71](#)
- [42] A. Aab *et al.*, “The Pierre Auger Observatory Upgrade - Preliminary Design Report,” *arXiv*, 4 2016. [20](#), [22](#)
- [43] I. Valino, “The flux of ultra-high energy cosmic rays after ten years of operation of the Pierre Auger Observatory,” *34th International Cosmic Ray Conference*, vol. ICRC2015, p. 271, 2016. [21](#), [81](#)
- [44] A. Aab *et al.*, “The Pierre Auger Cosmic Ray Observatory,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. A798, pp. 172–213, 2015. [23](#), [24](#), [25](#), [27](#)
- [45] M. Risse, K.-H. Kampert, and D. Kuempel, “Geometry reconstruction of fluorescence detectors revisited,” *30th International Cosmic Ray Conference.*, vol. ICRC2007, 2007. [26](#)

- [46] D. Newton, J. Knapp, and A. A. Watson, “The Optimum Distance at which to Determine the Size of a Giant Air Shower,” *Astroparticle Physics*, vol. 26, pp. 414–419, 2007. [27](#)
- [47] T. K. Gaisser and A. M. Hillas, “Reliability of the Method of Constant Intensity Cuts for Reconstructing the Average Development of Vertical Showers,” in *International Cosmic Ray Conference*, vol. 8 of *ICRC*, p. 353, Jan. 1977. [28](#)
- [48] Pierre Auger Collaboration, “Combined fit of spectrum and composition data as measured by the Pierre Auger Observatory,” *Journal of Cosmology and Astroparticle Physics*, vol. 1704, no. 04, p. 038, 2017. [Erratum: JCAP1803,no.03,E02(2018)]. [30](#), [36](#), [37](#), [38](#), [39](#), [77](#), [78](#), [82](#), [83](#), [101](#)
- [49] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner, “EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider,” *Physical Review*, vol. C92, no. 3, p. 034906, 2015. [31](#), [46](#), [51](#), [59](#), [77](#), [79](#), [95](#)
- [50] E.-J. Ahn, R. Engel, T. K. Gaisser, P. Lipari, and T. Stanev, “Cosmic ray interaction event generator sibyll 2.1,” *Physical Review D*, vol. 80, p. 094003, Nov 2009. [31](#)
- [51] S. Ostapchenko, “Lhc data on inelastic diffraction and uncertainties in the predictions for longitudinal extensive air shower development,” *Physical Review D*, vol. 89, p. 074009, Apr 2014. [31](#)
- [52] A. Yushkov, “Mass composition of cosmic rays with energies above $10^{17.2}$ eV from the hybrid data of the Pierre Auger Observatory,” in *36th International Cosmic Ray Conference*, vol. 36 of *ICRC2019*, July 2019. [31](#), [32](#), [81](#)
- [53] A. Aab *et al.*, “Observation of a Large-scale Anisotropy in the Arrival Directions of Cosmic Rays above 8×10^{18} eV,” *Science*, vol. 357, no. 6537, pp. 1266–1270, 2017. [32](#), [59](#), [64](#), [78](#), [89](#)

- [54] R. de Almeida *et al.*, “Large-scale and multipolar anisotropies of cosmic rays detected at the Pierre Auger Observatory with energies above 4 EeV,” in *37th International Cosmic Ray Conference*, vol. 37 of *ICRC2021*, July 2021. [32](#), [33](#), [34](#), [59](#), [78](#), [89](#)
- [55] A. Aab *et al.*, “Measurement of the cosmic ray energy spectrum above 2.5×10^{18} ev using the pierre auger observatory,” *Physical Review D*, vol. 102, p. 062005, Sept. 2020. [44](#), [101](#)
- [56] E. Mayotte and K. H. Kampert, “Composition sky mapping,” in *Auger Malargue Meeting - Argentina*, November 2018. [50](#)
- [57] E. Mayotte, “Composition anisotropy update: Using full sky realmc to remove fiducia fov cuts,” in *Auger Malargue Meeting - Argentina*, March, 2019. [50](#)
- [58] E. Mayotte, “Composition anisotropies update: v3r99 update, significance with mc, seasonal variation, and gaussian sampling for maps,” in *Auger Nijmegen Meeting*, June, 2019. [50](#)
- [59] The Pierre Auger Collaboration, “Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above $10^{17.8}$ eV,” *Physical Review*, vol. D90, no. 12, p. 122005, 2014. [51](#), [54](#), [81](#), [101](#), [122](#)
- [60] M. De Domenico, M. Settimo, S. Riggi, and E. Bertin, “Reinterpreting the development of extensive air showers initiated by nuclei and photons,” *Journal of Cosmology and Astroparticle Physics*, vol. 1307, p. 050, 2013. [59](#), [79](#), [121](#)
- [61] The Pierre Auger Collaboration *et al.*, “Anisotropy and chemical composition of ultra-high energy cosmic rays using arrival directions measured by the pierre auger observatory,” *Journal of Cosmology and Astroparticle Physics*, vol. 2011, no. 06, p. 022, 2011. [68](#)
- [62] P. Billoir and A. Letessier-Selvon, “A procedure to produce excess, probability and significance maps and to compute point-sources flux upper limits,” *GAP-2005-055*, 2005. [69](#)

- [63] G. L. Cassiday *et al.*, “Evidence for 10^{18} eV neutral particles from the direction of cygnus x-3,” *Physical Review Letters*, vol. 62, p. 383, 1989. [71](#)
- [64] O. Deligny and J. Biteau in *Pierre Auger online collaboration meeting- Malargüe XXXVII*, 2015. [74](#)
- [65] R. P. C. Junior and M. A. Leigui de Oliveira, “A Numerical Model for the Propagation of Ultra-High Energy Cosmic Rays through Extragalactic Magnetic Fields,” *35th International Cosmic Ray Conference.*, vol. ICRC2017, p. 477, 2018. [74](#)
- [66] Guido, Eleonora and others, “Combined fit of the energy spectrum and mass composition across the ankle with the data measured at Pierre Auger Observatory,” in *37th International Cosmic Ray Conference*, vol. 37 of *ICRC2021*, p. 12, July 2021. [77](#), [101](#), [105](#), [108](#)
- [67] R. Alves Batista *et al.*, “CRPropa 3 - a Public Astrophysical Simulation Framework for Propagating Extraterrestrial Ultra-High Energy Particles,” *Journal of Cosmology and Astroparticle Physics*, vol. 1605, no. 05, p. 038, 2016. [77](#), [79](#), [91](#), [95](#)
- [68] R. Aloisio, D. Boncioli, A. di Matteo, A. Grillo, S. Petrera, and F. Salamida, “Sim-Prop v2r3: Monte Carlo simulation code of UHECR propagation,” *arXiv e-prints*, p. arXiv:1602.01239, Feb. 2016. [77](#)
- [69] A. J. Koning, S. Hilaire, and M. C. Duijvestijn, “Talys: Comprehensive nuclear reaction modeling,” *AIP Conference Proceedings*, vol. 769, no. 1, pp. 1154–1159, 2005. [77](#), [82](#), [95](#)
- [70] J. L. Puget, F. W. Stecker, and J. H. Bredekamp, “Photonuclear interactions of ultrahigh energy cosmic rays and their astrophysical consequences,” *The Astrophysical Journal*, vol. 205, pp. 638–654, Apr. 1976. [77](#)
- [71] J. Allison *et al.*, “Geant4 developments and applications,” *IEEE Transactions on Nuclear Science.*, vol. 53, p. 270, 2006. [77](#)

- [72] R. C. Gilmore, R. S. Somerville, J. R. Primack, and A. Domínguez, “Semi-analytic modelling of the extragalactic background light and consequences for extragalactic gamma-ray spectra,” *Monthly Notices of the Royal Astronomical Society*, vol. 422, p. 3189–3207, Apr 2012. [77](#), [82](#), [95](#)
- [73] A. Dominguez, J. R. Primack, D. Rosario, F. Prada, R. Gilmore, *et al.*, “Extragalactic Background Light Inferred from AEGIS Galaxy SED-type Fractions,” *Monthly Notices of the Royal Astronomical Society.*, vol. 410, pp. 2556–2578, 01 2011. [77](#)
- [74] S. Ostapchenko, “Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model,” *Physical Review D*, vol. 83, p. 014018, Jan. 2011. [77](#)
- [75] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev, “Hadronic interaction model Sibyll 2.3d and extensive air showers,” *Physical Review D*, vol. 102, no. 6, p. 063002, 2020. [77](#), [103](#)
- [76] E. Guido, “Combined fit results for the PRL spectrum paper,” in *Pierre Auger online collaboration meeting*, 2020. [77](#), [89](#)
- [77] T. Winchen, D. Wittkowski, and K.-H. Kampert, “Combined fit results for the PRL spectrum paper,” in *Pierre Auger online collaboration meeting*, 2020. [77](#)
- [78] A. Aab *et al.*, “Large Scale Distribution of Ultra High Energy Cosmic Rays Detected at the Pierre Auger Observatory With Zenith Angles up to 80°,” *The Astrophysical Journal*, vol. 802, no. 2, p. 111, 2015. [78](#), [89](#)
- [79] R. Brun and F. Rademakers, “ROOT: An object oriented data analysis framework,” *Nuclear Instruments and Methods in Physics Research Section A*, vol. A389, pp. 81–86, 1997. [78](#)
- [80] “<https://www.doxygen.nl/>.” [78](#)

- [81] J. Abraham *et al.*, “Measurement of the energy spectrum of cosmic rays above 10^{18} eV using the Pierre Auger Observatory,” *Physics Letters B*, vol. B685, pp. 239–246, 2010. [80](#)
- [82] A. Aab and P. Abreu and M. Aglietta and Ahn, and Pierre Auger Collaboration, “Measurement of the cosmic ray spectrum above 4×10^{18} eV using inclined events detected with the Pierre Auger Observatory,” *Journal of Cosmology and Astroparticle Physics*, Aug. 2015. [80](#)
- [83] V. Verzi, “Measurement of the energy spectrum of ultra-high energy cosmic rays using the Pierre Auger Observatory,” in *36th International Cosmic Ray Conference (ICRC2019)*, vol. 36 of *International Cosmic Ray Conference*, p. 8, July 2019. [81](#), [95](#)
- [84] A. di Matteo, *Ultra-high-energy cosmic ray phenomenology: Monte Carlo simulations and experimental data*. PhD thesis, Università degli Studi dell’Aquila - Dipartimento di Scienze Fisiche e Chimiche, 2015. GAP 2016-011 - section 5.2. [82](#)
- [85] <https://docs.pymc.io/>. [82](#)
- [86] J. Biteau, “Stellar Mass and Star Formation Rate within a Billion Light-years,” *The Astrophysical Journal Supplement Series*, vol. 256, p. 15, Sept. 2021. [89](#), [91](#)
- [87] K. V. Mardia and P. J. Zemroch, *Applied Statistics: The von Mises Distribution Function*, vol. 24. Journal of the Royal Statistical Society: Series C (Applied Statistics), 1975. [91](#)
- [88] P. Billoir and O. Deligny, “Estimates of multipolar coefficients to search for cosmic ray anisotropies with non-uniform or partial sky coverage,” *Journal of Cosmology and Astroparticle Physics*, vol. 02, p. 009, 2008. [91](#), [94](#)
- [89] P. Abreu, M. Aglietta, M. Ahlers, E. J. Ahn, I. F. M. Albuquerque, D. Allard, I. Allekotte, J. Allen, P. Allison, and et al., “Large scale distribution of arrival dir-

- ections of cosmic rays detected above 10^{18} eV at the Pierre Auger Observatory,” *The Astrophysical Journal Supplement Series*, vol. 203, p. 34, Dec 2012. [94](#)
- [90] C. Guépin, K. Kotera, E. Barausse, K. Fang, and K. Murase, “Ultra-High Energy Cosmic Rays and Neutrinos from Tidal Disruptions by Massive Black Holes,” *Astronomy & Astrophysics*, vol. 616, p. A179, 2018. [Erratum: *Astron. Astrophys.* 636, C3 (2020)]. [98](#)
- [91] D. Biehl, D. Boncioli, C. Lunardini, and W. Winter, “Tidally disrupted stars as a possible origin of both cosmic rays and neutrinos at the highest energies,” *Scientific Reports*, vol. 8, no. 1, 2018. [98](#)
- [92] H. Yuksel, M. D. Kistler, J. F. Beacom, and A. M. Hopkins, “Revealing the High-Redshift Star Formation Rate with Gamma-Ray Bursts,” *The Astrophysical Journal Letters*, vol. 683, pp. L5–L8, 2008. [98](#)
- [93] B. E. Robertson, R. S. Ellis, S. R. Furlanetto, and J. S. Dunlop, “Cosmic Reionization and Early Star-forming Galaxies: a Joint Analysis of new Constraints From Planck and the Hubble Space Telescope,” *The Astrophysical Journal Letters*, vol. 802, no. 2, p. L19, 2015. [98](#)
- [94] C. Ventura, F. Catalani, R. M. de Almeida, *et al.*, “Search for correlation between starburst galaxies and low energy cosmic rays,” *GAP-2018-012*, 2018. [108](#)
- [95] C. Ventura, F. Catalani, R. M. de Almeida, *et al.*, “Search for correlation between starburst galaxies and low energy cosmic rays: Part 2,” *GAP-2018-060*, 2018. [108](#)
- [96] M. B. Muller, C. Ventura, *et al.*, “On the using of the shuffling technique for FD events,” *GAP-2019-053*, 2019. [108](#)
- [97] B. Lago, C. Ventura, *et al.*, “A new ROOT based combined fit analysis,” *GAP-2020-047*, 2020. [108](#)

BIBLIOGRAPHY

- [98] C. Ventura, R. de Almeida, *et al.*, “On combining directional intensity and direction composition,” *GAP-2021-055*, 2021. 108
- [99] Collaboration Meeting, “<https://indico.nucleares.unam.mx/event/1408/>,” March - 2019. 108
- [100] CR Phenomenology task, “<https://www.auger.unam.mx/AugerWiki/Phenomenology>,” 2019. 108
- [101] Auger Collaboration Meeting November 2019, “<https://indico.nucleares.unam.mx/event/1409/contributions>,” 2019. 108
- [102] Auger Collaboration Meeting, “<https://indico.nucleares.unam.mx/event/1472/contributions>,” 2019. 108
- [103] Teleconferences of the Arrival Directions Task, “<https://www.auger.unam.mx/AugerWiki/ArrivalDirections/ArrivalDirectionsTeleconferences>,” 2019. 108
- [104] Teleconferences of the Mass Composition, “<https://www.auger.unam.mx/AugerWiki/MassComposition>,” 2019. 108
- [105] T. Pierog *et al.*, “First results of fast one-dimensional hybrid simulation of EAS using CONEX,” *Nuclear Physics B Proceedings Supplements*, vol. 151, pp. 159–162, 2006. 122
- [106] S. Petrerá and F. Salamida, “Update of the Xmax parameterizations for post-LHC hadronic models,” *GAP-2018-21*, 2018. 122

Appendix A

X_{\max} distributions used in the global combined fit

Here, we will describe the parametrization used to X_{\max} distributions. Gumbel parametrization: First, we parametrize the X_{\max} distribution before detector effects as a Gumbel function.

A.1 Gumbel Distributions

The X_{\max} distributions at each log energy bin and for each mass are parameterized using Gumbel probability distribution functions [60]. These functions are denoted as $g(X_{\max} | \lg E, A)$

$$g(X_{\max} | \lg E, A) = \frac{\lambda^\lambda}{\sigma \Gamma(\lambda)} \exp\left(-\lambda \frac{X_{\max} - \mu}{\sigma} - \lambda \exp\left(-\frac{X_{\max} - \mu}{\sigma}\right)\right) \quad (\text{A.1})$$

where

$$\mu = p_{0\mu} + p_{1\mu} \log_{10}(E/E_0) + p_{2\mu} \log_{10}^2(E/E_0) \quad (\text{A.2})$$

$$\sigma = p_{0\sigma} + p_{1\sigma} \log_{10}(E/E_0) \quad (\text{A.3})$$

$$\lambda = p_{0\lambda} + p_{1\lambda} \log_{10}(E/E_0) \quad (\text{A.4})$$

$$p_{0\mu} = a_{\mu 0} + a_{\mu 1} \ln_A + a_{\mu 2} \ln_A^2 \quad (\text{A.5})$$

$$p_{1\mu} = b_{\mu 0} + b_{\mu 1} \ln_A + b_{\mu 2} \ln_A^2 \quad (\text{A.6})$$

$$p_{2\mu} = c_{\mu 0} + c_{\mu 1} \ln_A + c_{\mu 2} \ln_A^2 \quad (\text{A.7})$$

The chosen reference energy is $E_0 = 10^{19}$ eV. The values of $a_{\mu j}, b_{\mu j}, c_{\mu j}$ were obtained by fitting this Gumbel parametrization to X_{max} distributions obtained via CONEX simulations [105]. The parameters λ, σ, μ depend on the mass and the energy of the detected nuclei and differ for each hadronic model. These parameters are reported in [106].

A.2 Detector Acceptance

A Gumbel distribution probability function has to be multiplied by the acceptance function $\mathcal{A}(X_{\text{max}}, E)$. The probability that a shower with energy E and maximum depth X_{max} occurs within the detector area it will be detected, using the parametrization from [59].

$$\mathcal{A}(X_{\text{max}}, E) = \begin{cases} \exp\left(-\frac{x_1 - X_{\text{max}}}{\lambda_1}\right), & X_{\text{max}} < x_1, \\ 1, & x_1 \leq X_{\text{max}} \leq x_2 \\ \exp\left(-\frac{X_{\text{max}} - x_2}{\lambda_2}\right), & X_{\text{max}} > x_2 \end{cases} \quad (\text{A.8})$$

A.3 Resolution parameters

We convolve the result by the detector resolution function, parameterized in [59].

$$\mathcal{R}(X_{\text{max}}^{\text{rec}} - X_{\text{max}} | E) = f \cdot G(X_{\text{max}}^{\text{rec}} | X_{\text{max}}, \sigma_1) + (1 - f) \cdot G(X_{\text{max}}^{\text{rec}} | X_{\text{max}}, \sigma_2), \quad (\text{A.9})$$

where $G(x|\mu, \sigma)$ is a Gaussian distribution with mean μ and standard deviation σ , and the parameters f, σ_1 and σ_2 are functions of E and are tabulated in A.1. The probability that an event with energy E and mass number A is detected with reconstructed maximum

APPENDIX A. X_{MAX} DISTRIBUTIONS USED IN THE GLOBAL COMBINED FIT

depth X_{rec} is

$$p(X_{\text{rec}}|E, A) = \int_0^{+\infty} \mathcal{R}(X_{\text{max}}^{\text{rec}} - X_{\text{max}}|E) \mathcal{A}(X_{\text{max}}, E) g_{\text{tot}}(X_{\text{max}}|\log E, A) dX_{\text{max}}. \quad (\text{A.10})$$

In practice, we compute $A(X_{\text{max}}, E)g_{\text{tot}}(X_{\text{max}}|E)$ at the centre X_i of the X_{max} bins, and use

$$p(X_j|E) = \sum_i \mathcal{A}(X_i|E)g_{\text{tot}}(X_i|E) \int_{X_j-X_i-0.5\Delta X}^{X_j-X_i+0.5\Delta X} \mathcal{R}(X|E) dX, \quad (\text{A.11})$$

Table A.1: Parameters of acceptance and resolution used to fit the X_{max} distributions. (all values in g/cm^2 , except f , which is adimensional).

	acceptance				resolution		
$\log_{10}(\text{E/eV})$	x_1	x_2	λ_1	λ_2	σ_1	σ_2	f
17.8-17.9	586.1	881.3	108.5	94.8	17.54	33.73	0.6168
17.9-18.0	592.3	883.1	133.4	100.9	16.67	32.86	0.6255
18.0-18.1	597.3	884.9	157.8	106.9	15.87	31.95	0.6342
18.1-18.2	601.1	886.7	182.0	113.0	15.12	31.01	0.6430
18.2-18.3	603.7	888.5	206.0	119.1	14.44	30.04	0.6518
18.3-18.4	605.1	890.3	229.9	125.2	13.81	29.05	0.6606
18.4-18.5	605.3	892.1	253.2	131.3	13.26	28.09	0.6694
18.5-18.6	604.3	893.9	275.9	137.4	12.78	27.15	0.6781
18.6-18.7	602.0	895.8	298.7	143.5	12.35	26.25	0.6869
18.7-18.8	598.5	897.6	321.3	149.7	11.98	25.41	0.6958
18.8-18.9	593.9	899.4	343.1	155.7	11.69	24.68	0.7045
18.9-19.0	587.9	901.3	365.4	162.0	11.45	24.05	0.7135
19.0-19.1	580.8	903.1	386.5	168.0	11.28	23.57	0.7222
19.1-19.2	572.6	904.9	407.3	174.1	11.17	23.25	0.7310
19.2-19.3	562.9	906.7	428.2	180.3	11.10	23.08	0.7399
19.3-19.4	552.4	908.5	448.1	186.3	11.08	23.07	0.7485
19.4-19.5	541.2	910.2	466.9	192.0	11.09	23.19	0.7568
19.5-19.6	526.8	912.2	488.6	198.7	11.12	23.45	0.7664
19.6-21.0	499.9	915.5	523.2	209.8	11.22	24.10	0.7823

Appendix B

Results in the best-fit scenario

In section 8.2.1 we reported the best fit for different scenarios.

Figure B.1 shows the results of the best fit for $\sigma_0 = 10^\circ$, reported in Table 8.3, and the X_{\max} distributions corresponding to the best fit are shown in figure B.2.

Figures B.3 and B.4 show the results of the best fit for $\sigma_0 = 25^\circ$, reported in Table 8.3. Figure B.5 and B.6 show the results of the best fit using the scenario CTGE for $m = -3$, reported in Table 8.4.

Figure B.7 and B.8 show the results of the best fit using the scenario CTGE for $m = +3$, reported in Table 8.4.

Now, we will show in Figure B.9 and B.10 the results for the best fit using the scenario CTGE considering a shift of δ in the X_{\max} for no source evolution, $\sigma_0 = 20^\circ$, reported in Table 8.5.

Figure B.11 and B.12 show the results of the best fit using the scenario CTGE considering shift of δ in the X_{\max} for $m = -3$, $\sigma_0 = 20^\circ$, reported in Table 8.5.

Figure B.13 and B.14 show the results of the best fit using the scenario CTGE considering shift of δ in the X_{\max} for $m = +3$, $\sigma_0 = 20^\circ$, reported in Table 8.5.

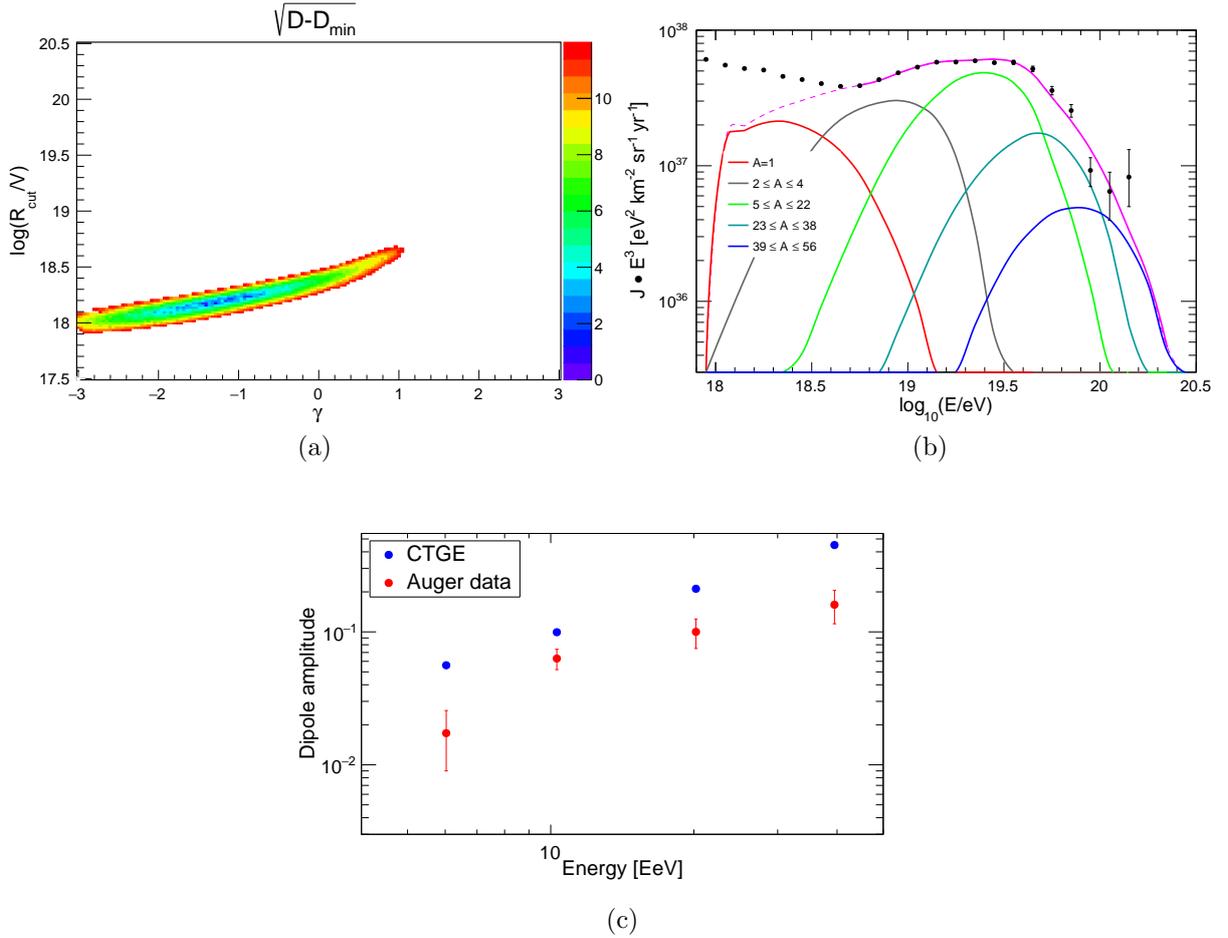


Figure B.1: (a) Minimum values of $\sqrt{D - D_{\min}}$ as a function of $(\gamma, \log_{10} R_{\text{cut}}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red) using $\sigma_0 = 10^\circ$ and considering no source evolution

APPENDIX B. RESULTS IN THE BEST-FIT SCENARIO

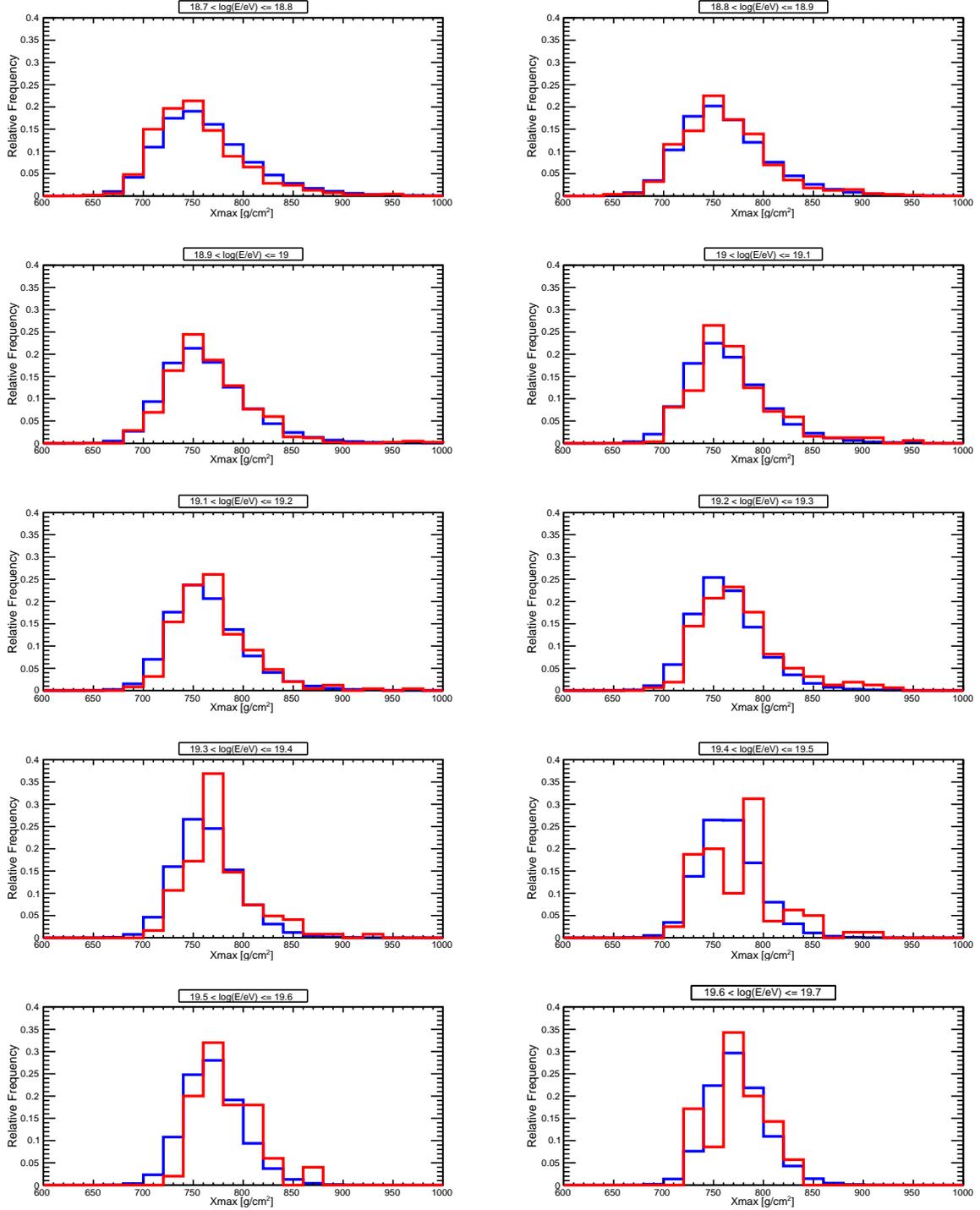


Figure B.2: X_{\max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset.

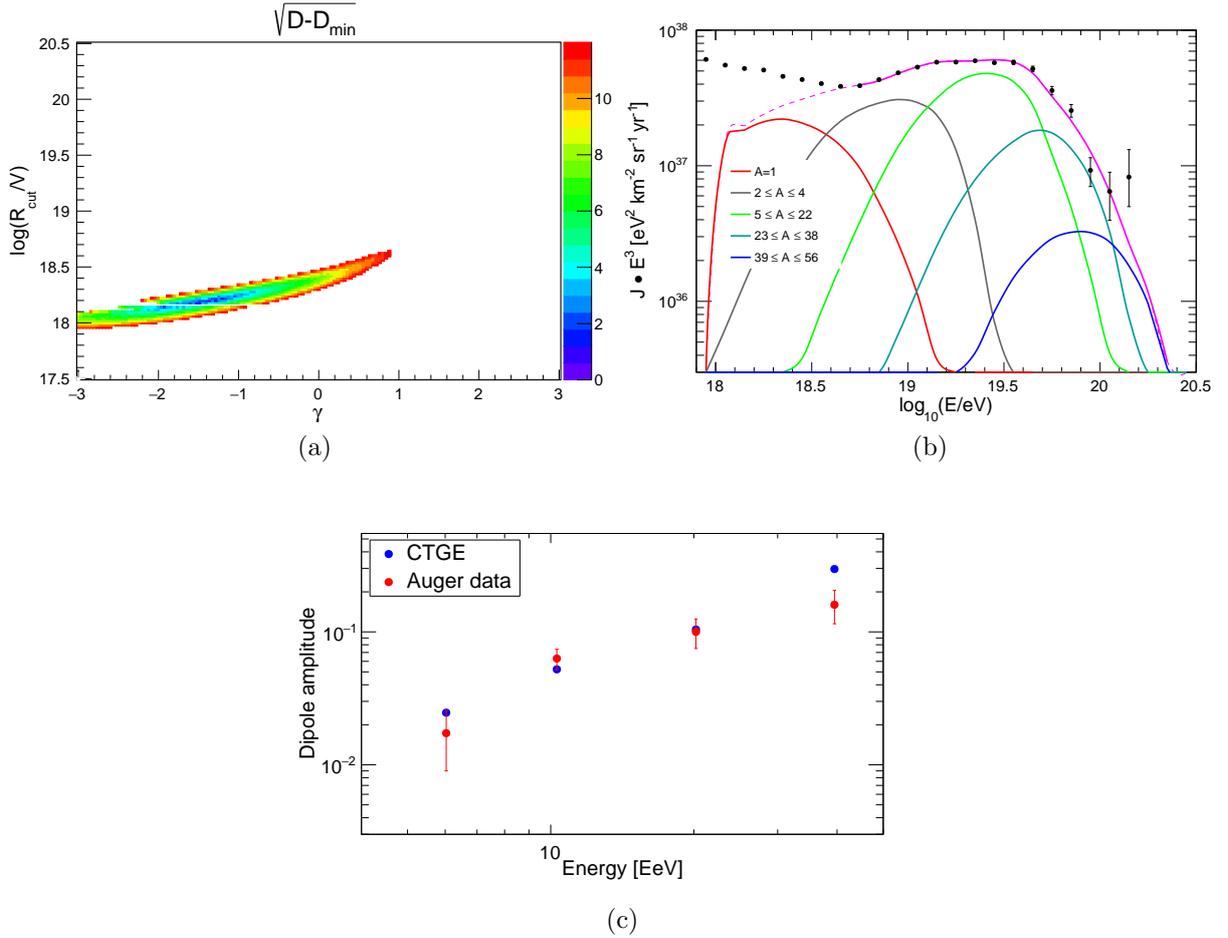


Figure B.3: (a) Minimum values of $\sqrt{D - D_{\min}}$ as a function of $(\gamma, \log_{10} R_{\text{cut}}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red) using $\sigma_0 = 25^\circ$ and considering no source evolution

APPENDIX B. RESULTS IN THE BEST-FIT SCENARIO

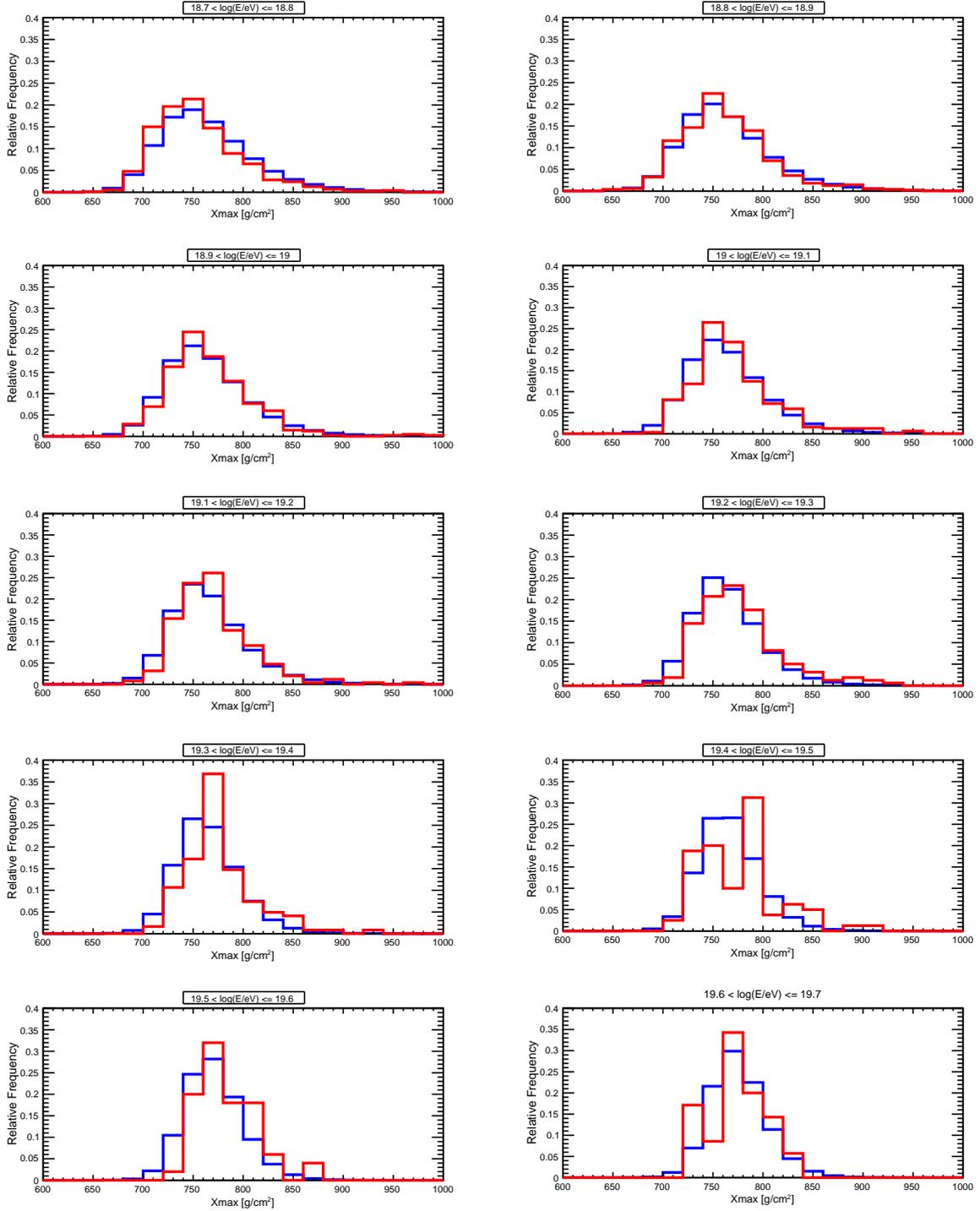


Figure B.4: X_{\max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset.

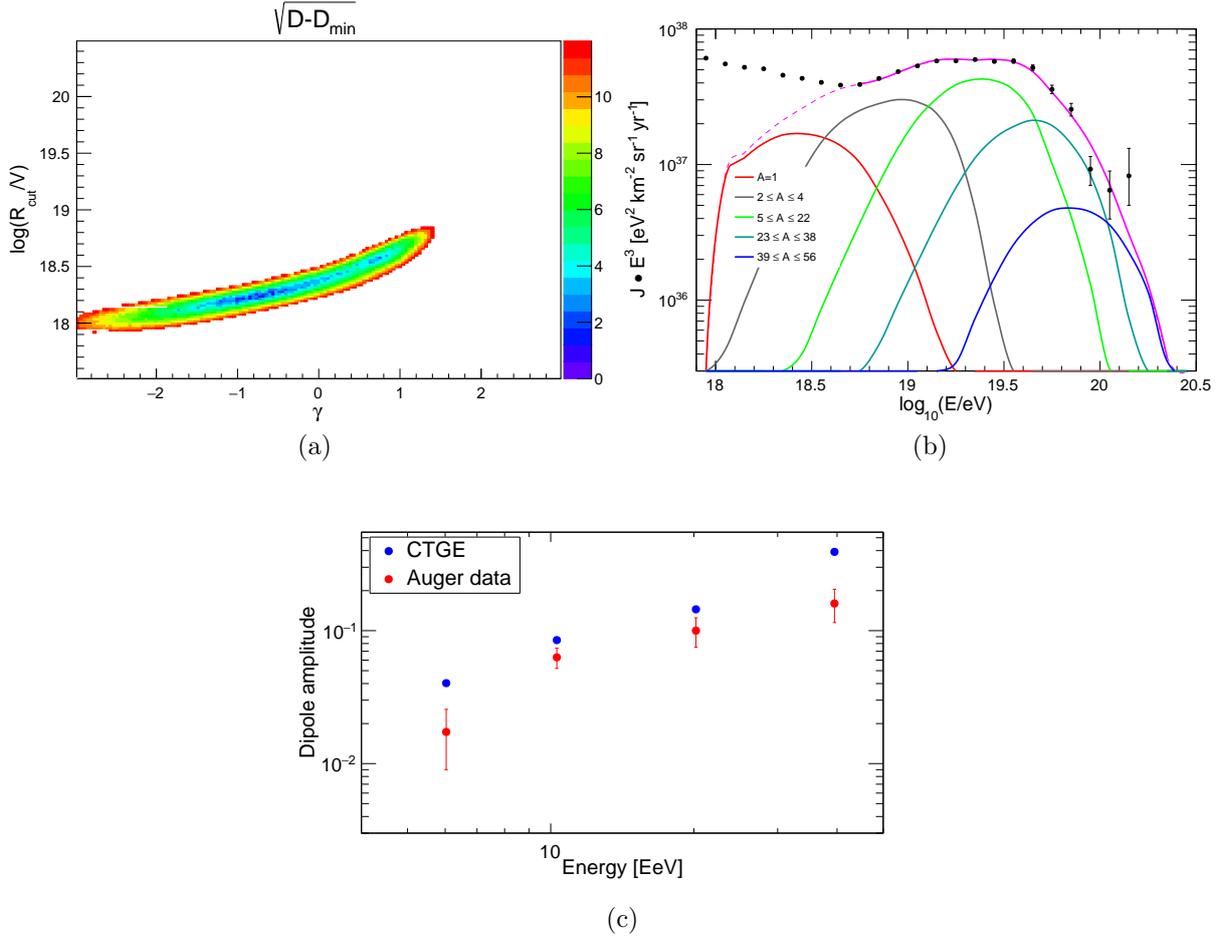


Figure B.5: (a) Minimum values of $\sqrt{D - D_{\min}}$ as a function of $(\gamma, \log_{10} R_{\text{cut}}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red) using $\sigma_0 = 20^\circ$ and considering source evolution $m = -3$

APPENDIX B. RESULTS IN THE BEST-FIT SCENARIO

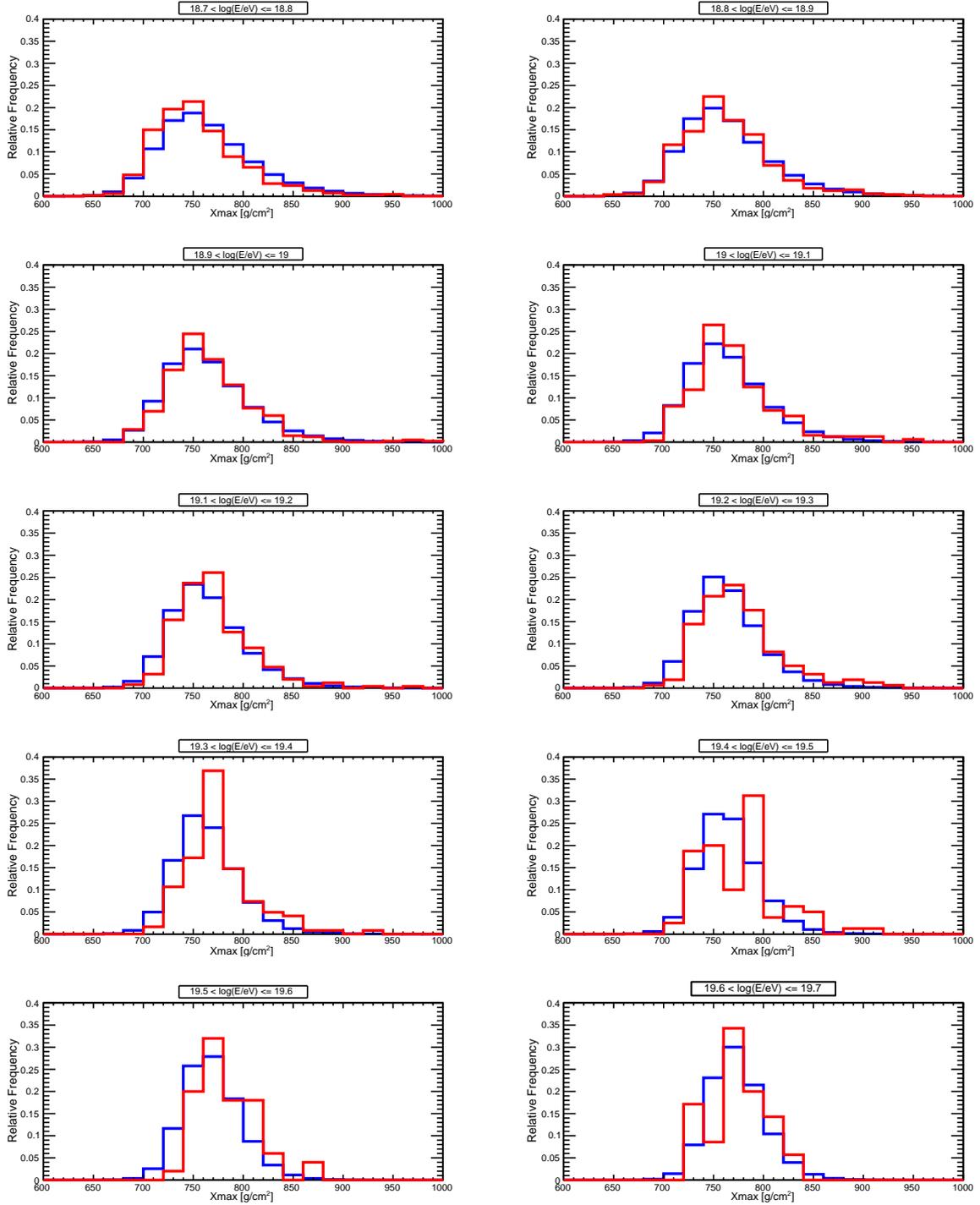


Figure B.6: X_{\max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset.

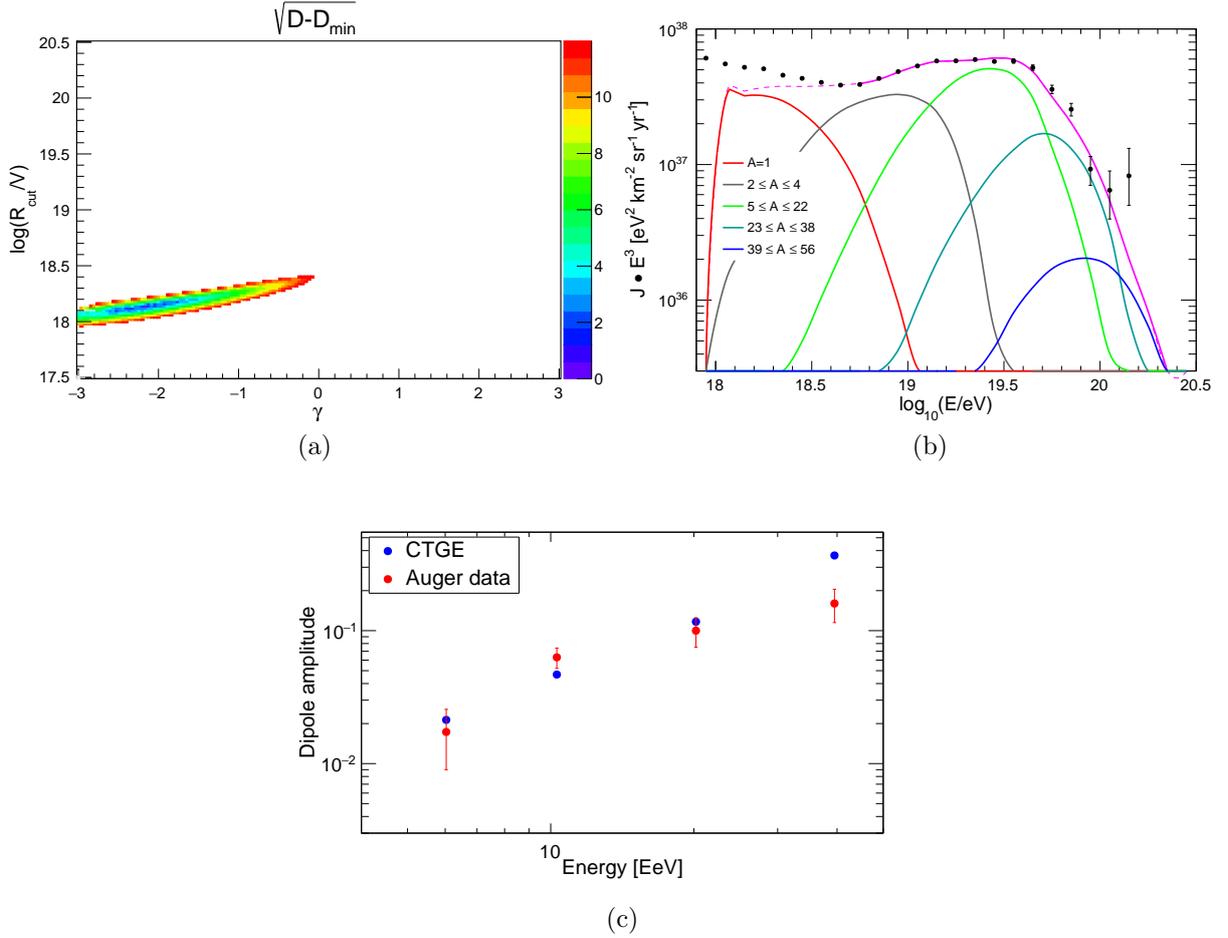


Figure B.7: (a) Minimum values of $\sqrt{D - D_{\min}}$ as a function of $(\gamma, \log_{10} R_{\text{cut}}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red) using $\sigma_0 = 20^\circ$ and considering source evolution $m = +3$

APPENDIX B. RESULTS IN THE BEST-FIT SCENARIO

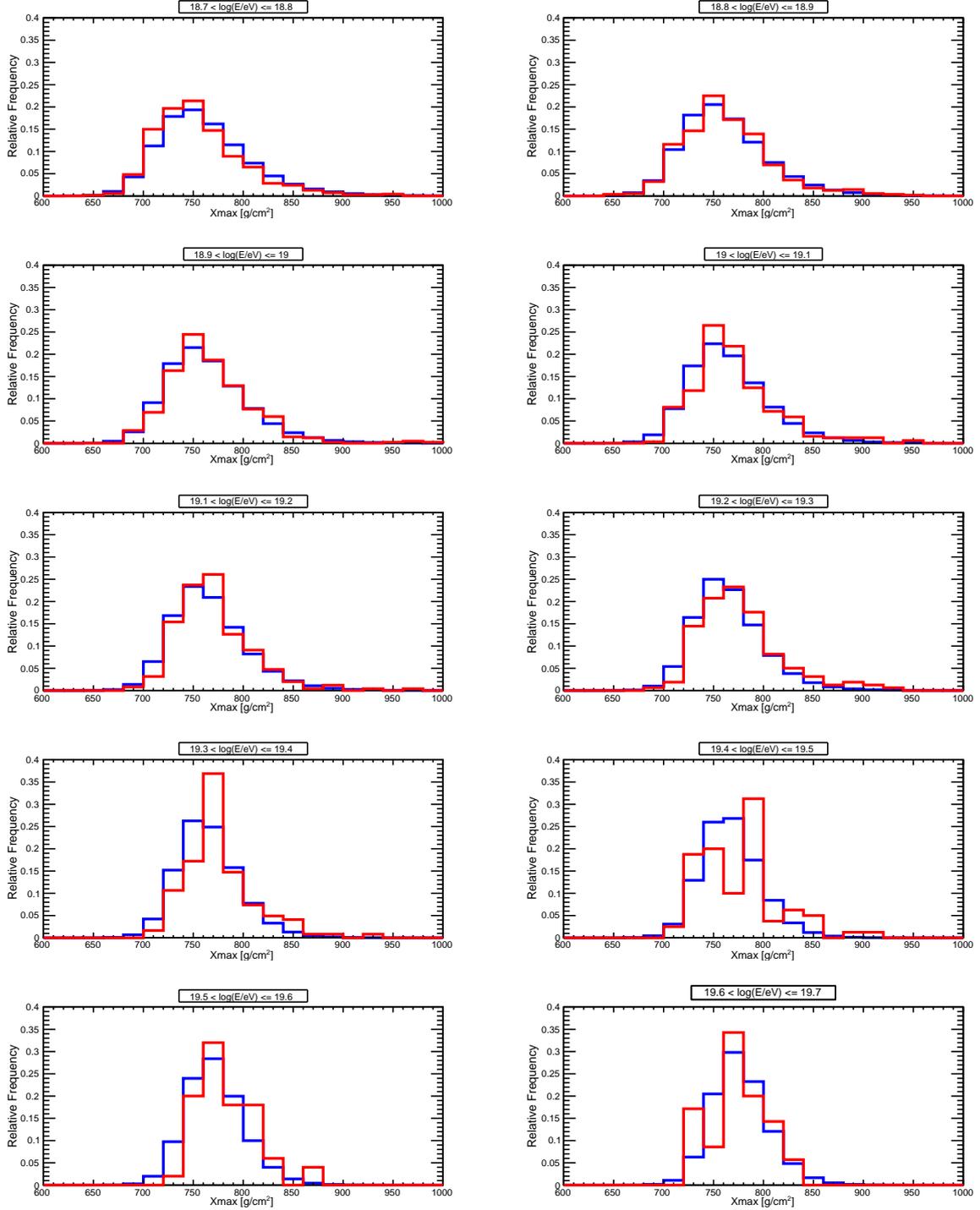


Figure B.8: X_{\max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset.

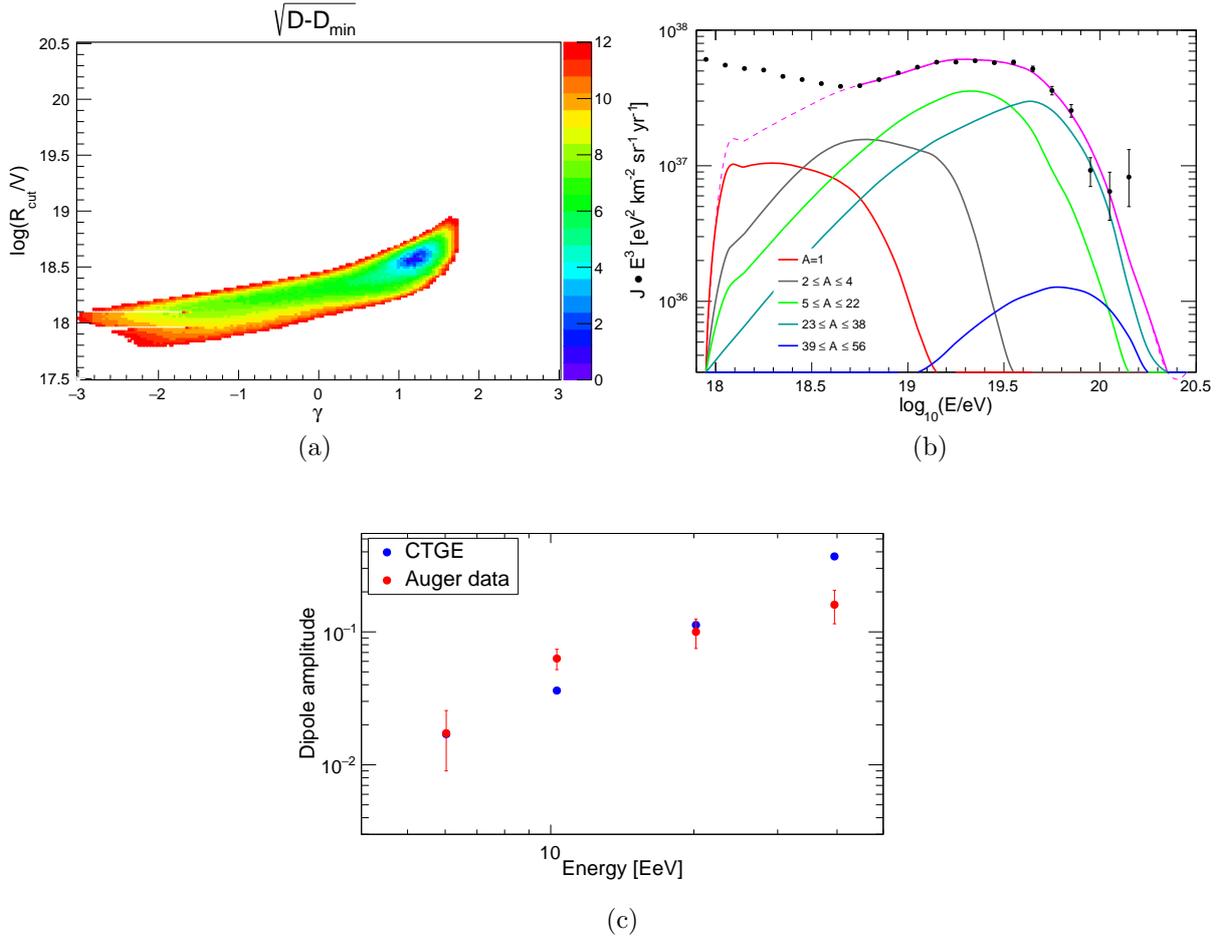


Figure B.9: (a) Minimum values of $\sqrt{D - D_{\min}}$ as a function of $(\gamma, \log_{10} R_{\text{cut}}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red).

APPENDIX B. RESULTS IN THE BEST-FIT SCENARIO

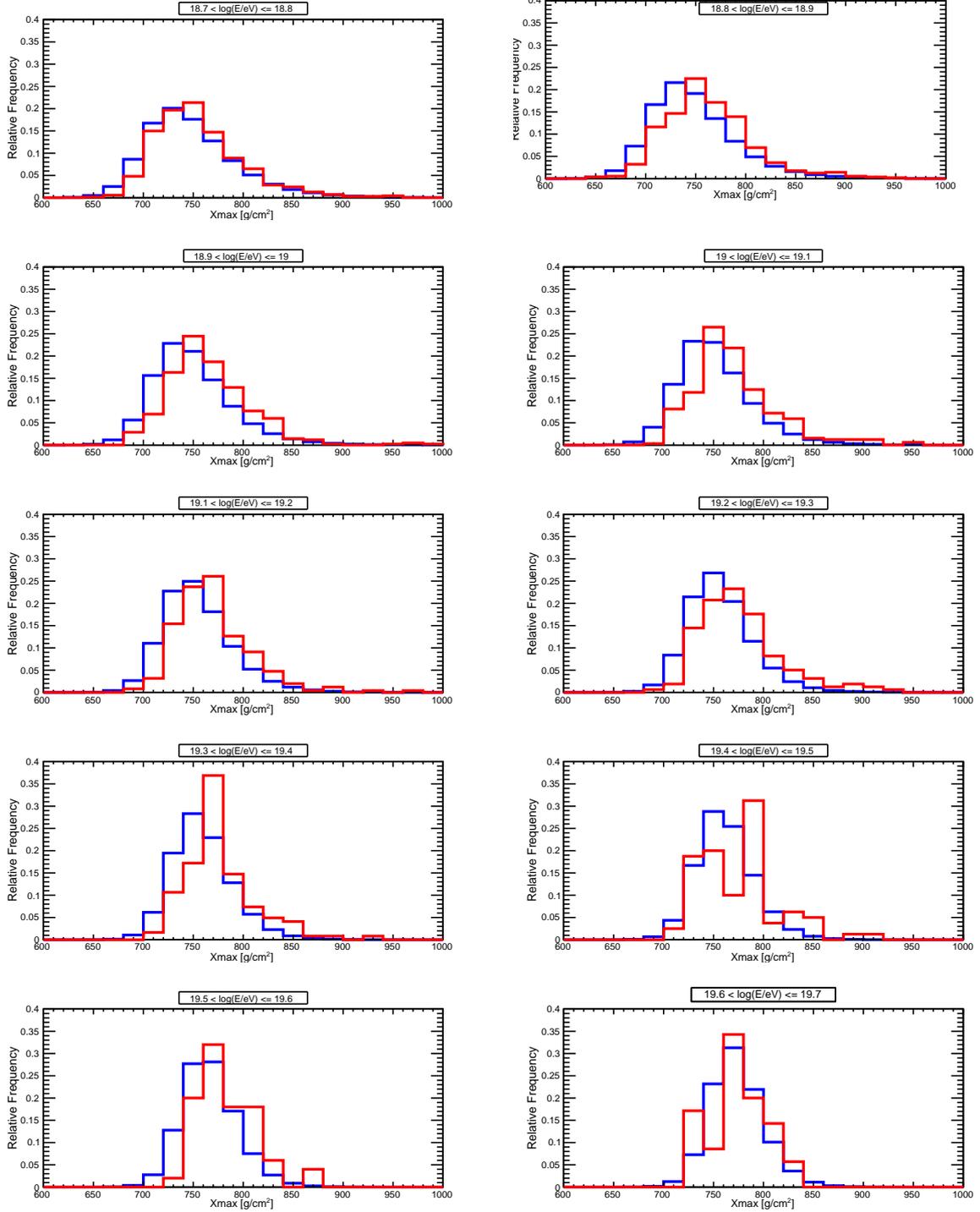


Figure B.10: X_{\max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset.

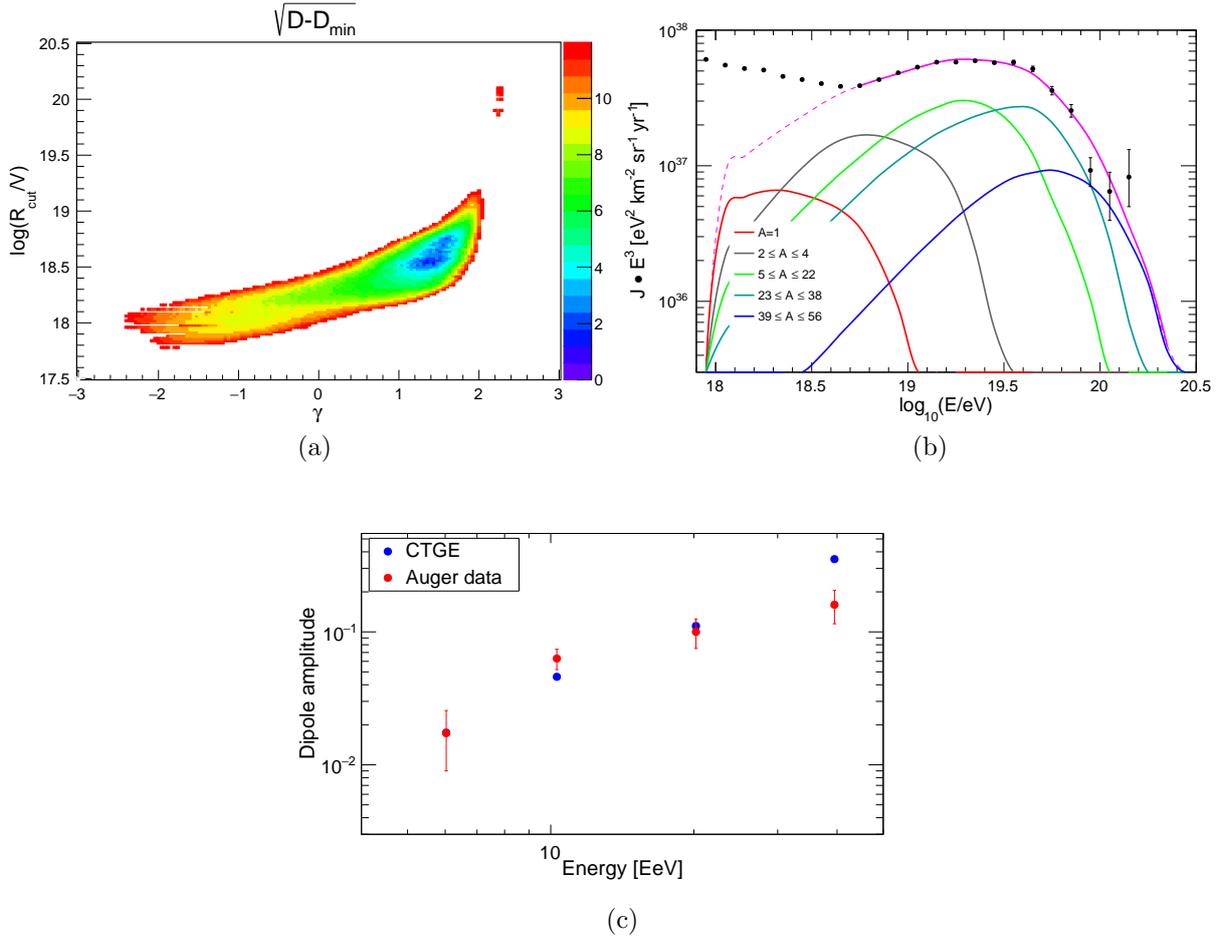


Figure B.11: (a) Minimum values of $\sqrt{D - D_{\min}}$ as a function of $(\gamma, \log_{10} R_{\text{cut}}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red).

APPENDIX B. RESULTS IN THE BEST-FIT SCENARIO

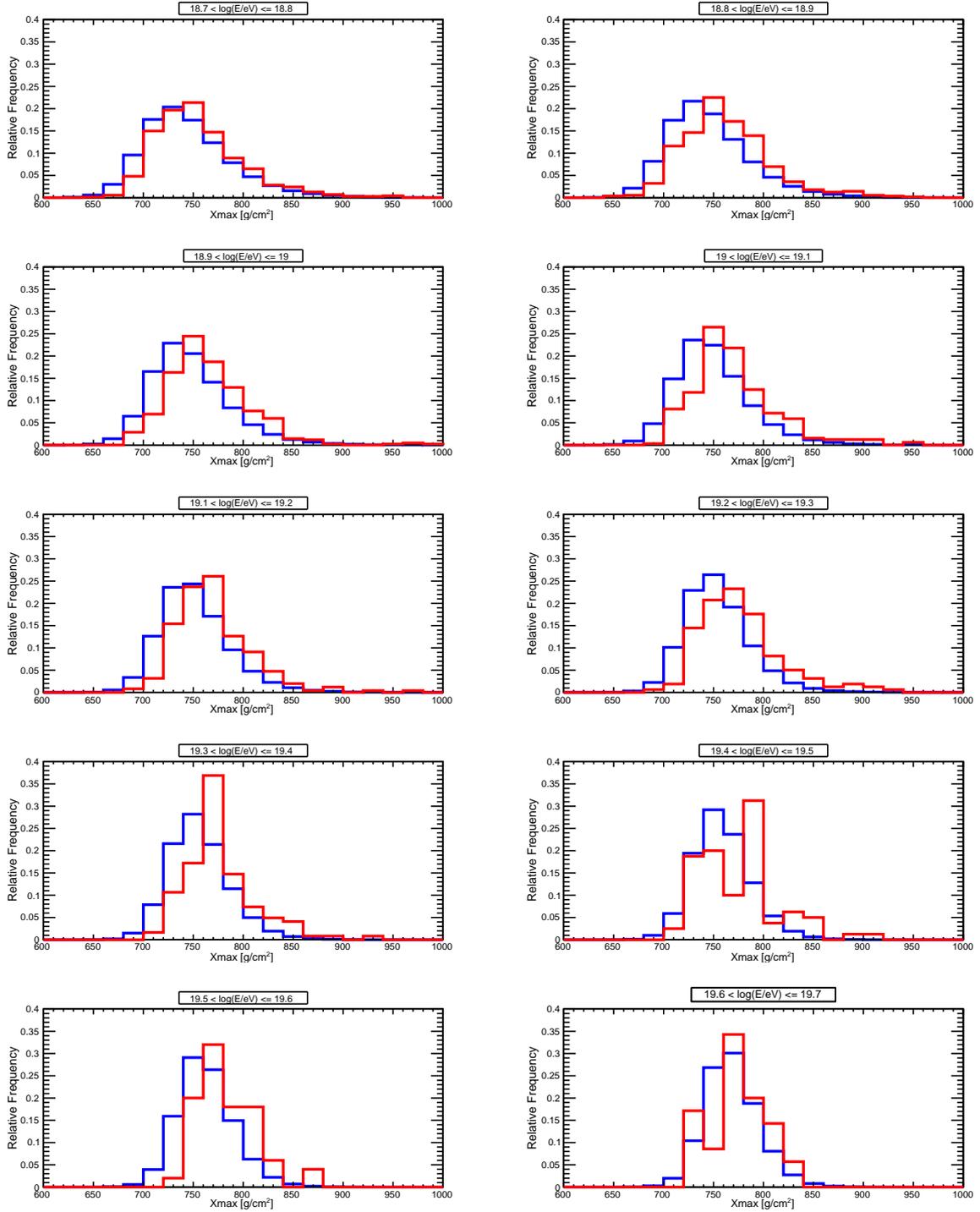


Figure B.12: X_{\max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset.

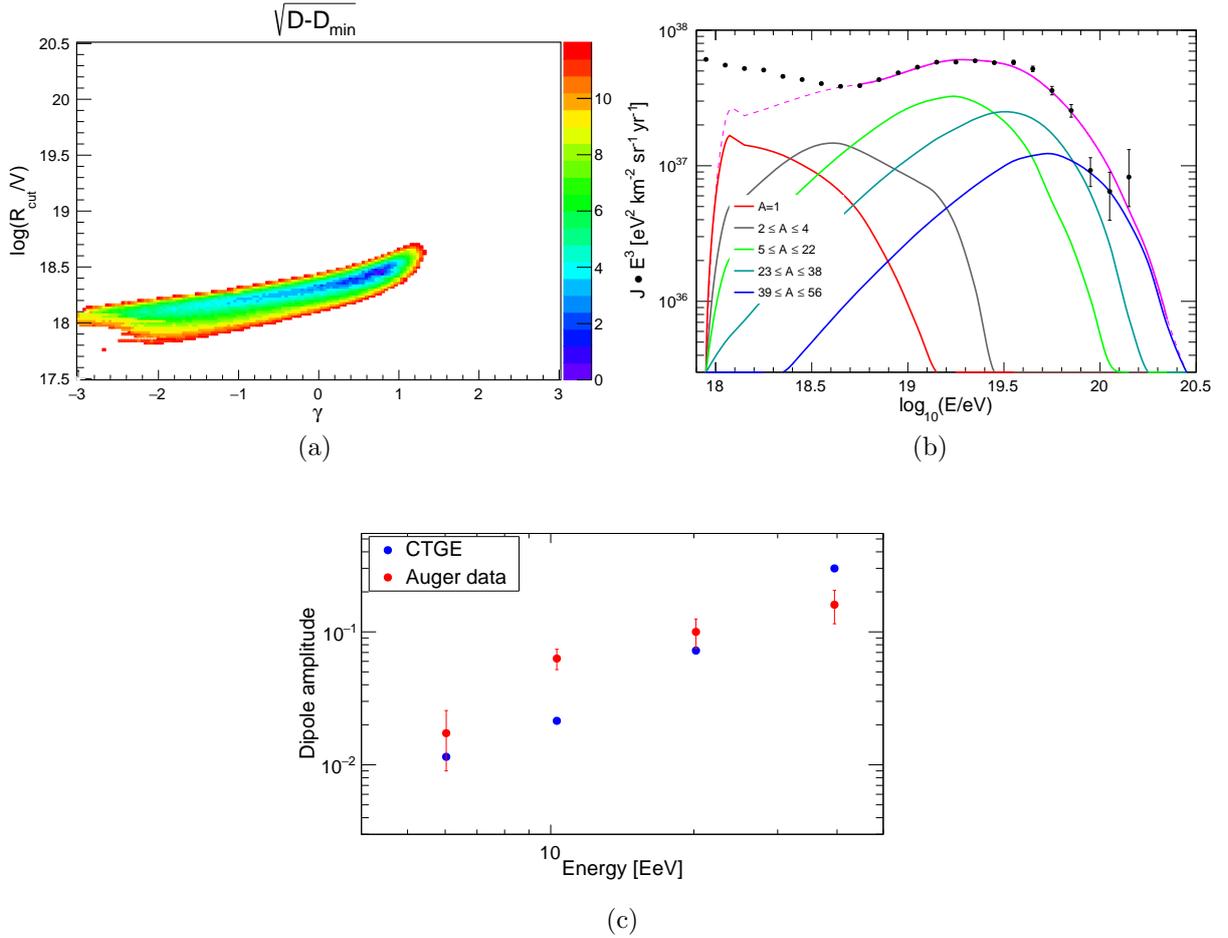


Figure B.13: (a) Minimum values of $\sqrt{D - D_{\min}}$ as a function of $(\gamma, \log_{10} R_{\text{cut}}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red).

APPENDIX B. RESULTS IN THE BEST-FIT SCENARIO

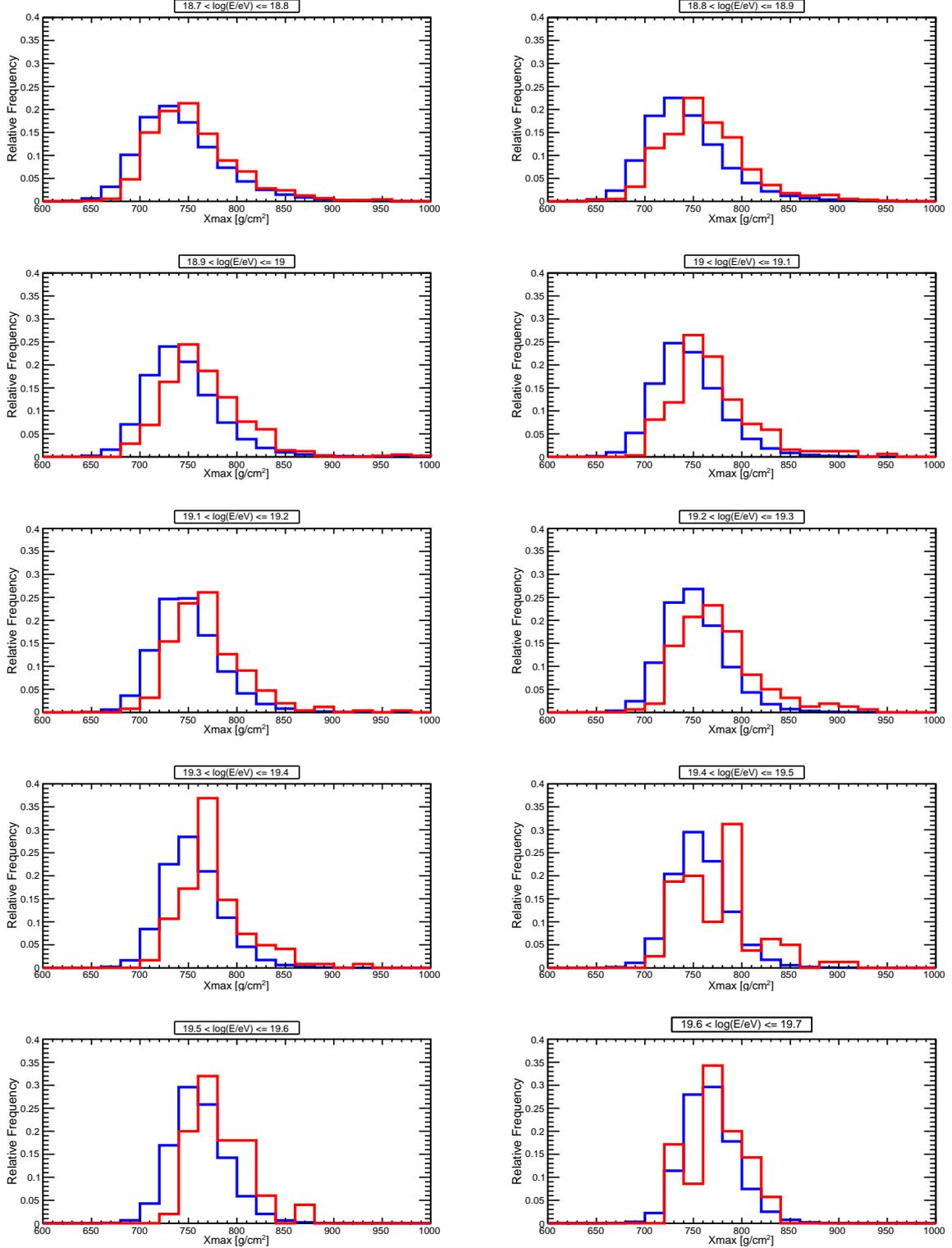


Figure B.14: X_{\max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset.

Beyond including different interaction models, here we will show the results for the best fit using the scenario CTGS, with $\sigma_0 = 10$. Figures B.15 and B.16 show all the results using Sibyll 2.3d from Table 8.6. Figures B.17 and B.18 show all the results using Sibyll

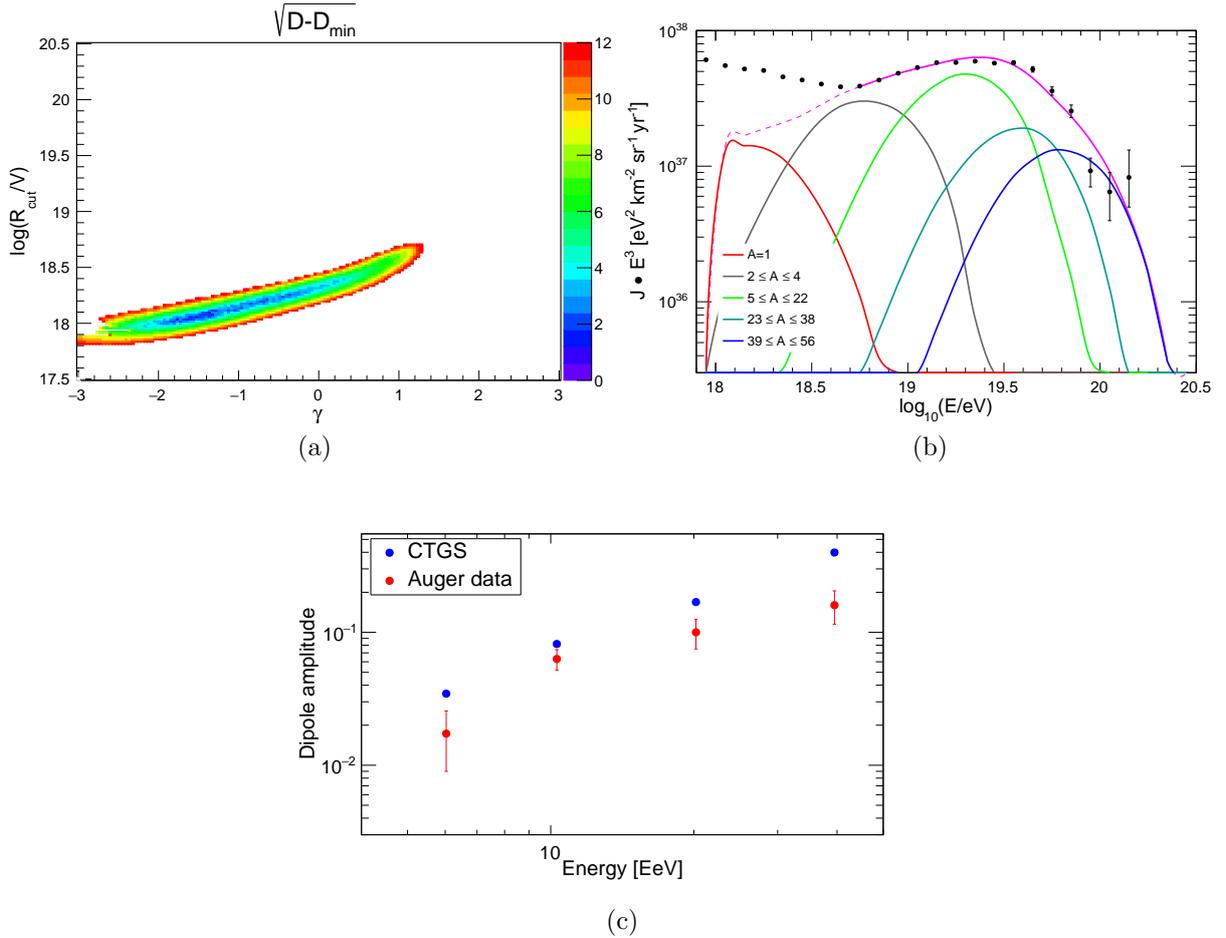


Figure B.15: (a) Minimum values of $\sqrt{D - D_{\min}}$ as a function of $(\gamma, \log_{10} R_{\text{cut}}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red).

2.3d considering a shift of δ in the X_{\max} and $\sigma_0 = 10$.

Figures B.19 and B.20 show all the results using Sibyll 2.3d, with $\sigma_0 = 20$, reported in Table 8.6.

APPENDIX B. RESULTS IN THE BEST-FIT SCENARIO

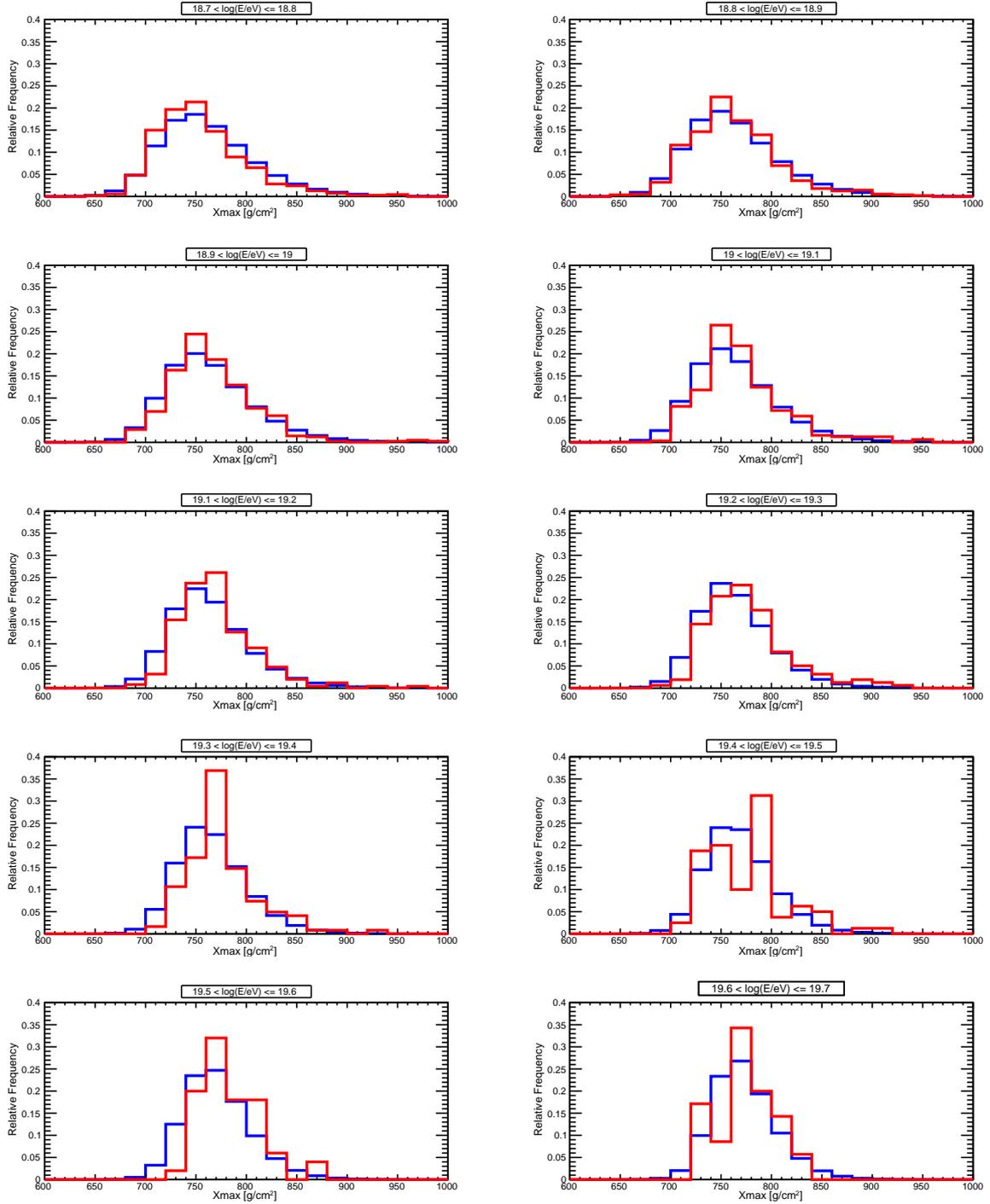


Figure B.16: X_{\max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset.

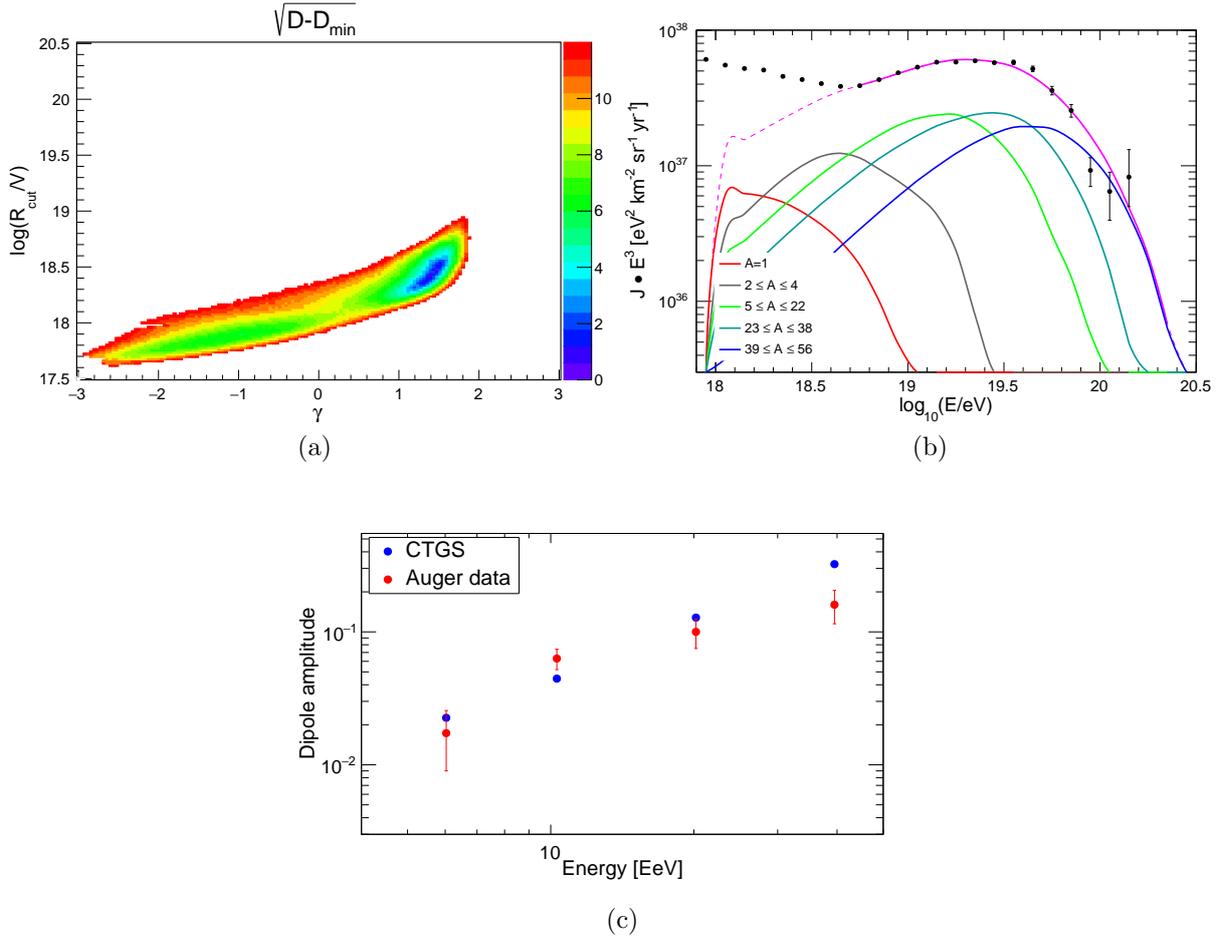


Figure B.17: (a) Minimum values of $\sqrt{D - D_{\min}}$ as a function of $(\gamma, \log_{10} R_{\text{cut}}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red).

APPENDIX B. RESULTS IN THE BEST-FIT SCENARIO

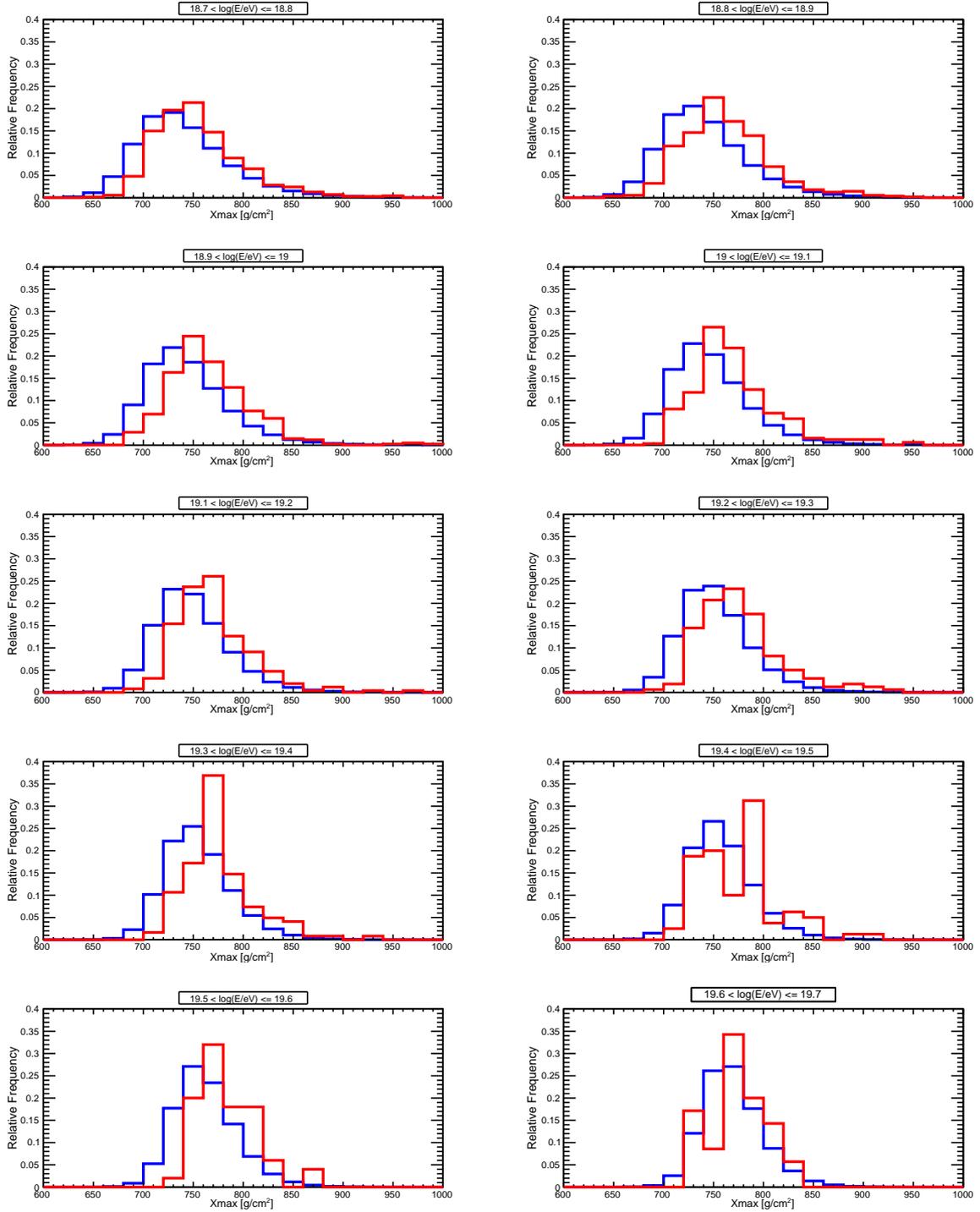


Figure B.18: X_{\max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset.

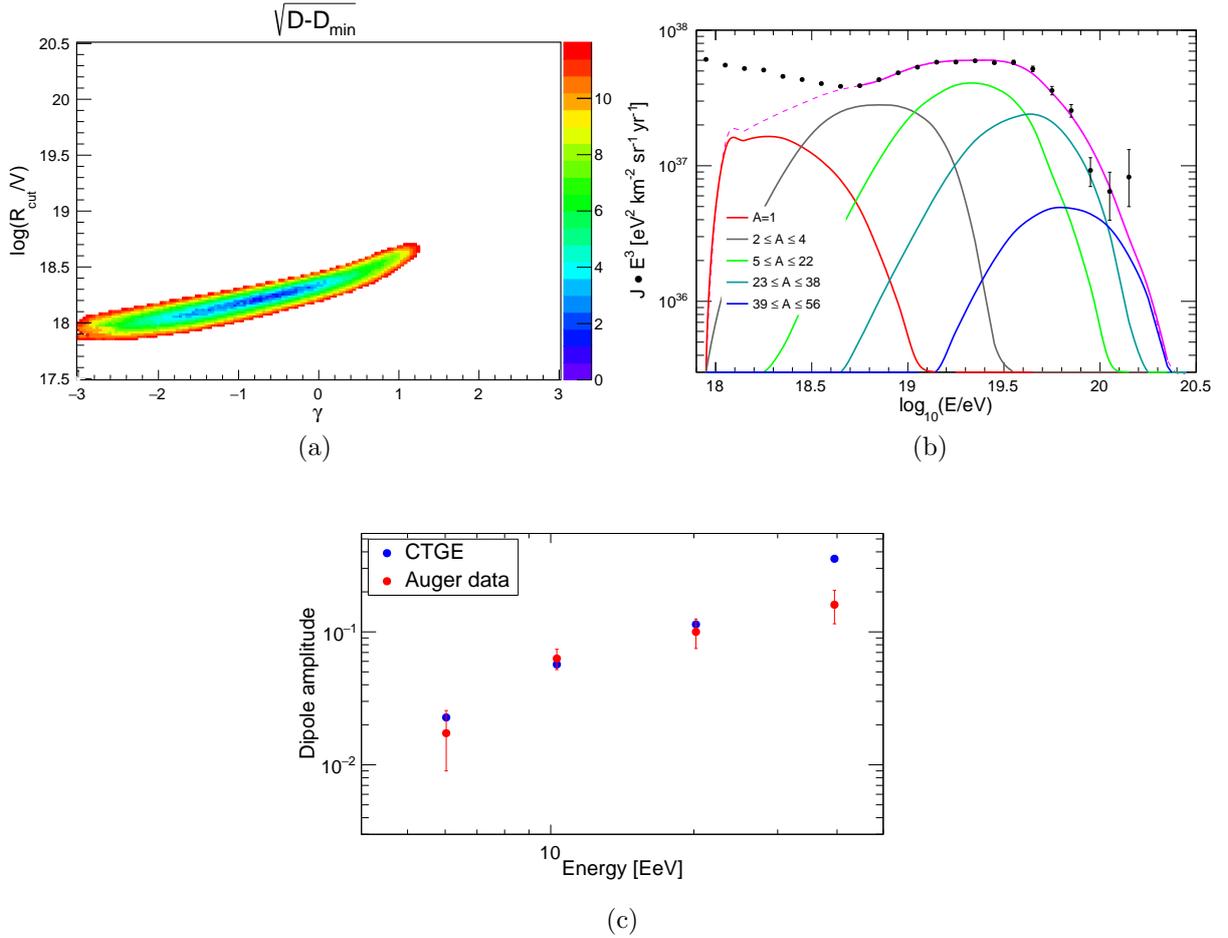


Figure B.19: (a) Minimum values of $\sqrt{D - D_{\min}}$ as a function of $(\gamma, \log_{10} R_{\text{cut}}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red).

APPENDIX B. RESULTS IN THE BEST-FIT SCENARIO

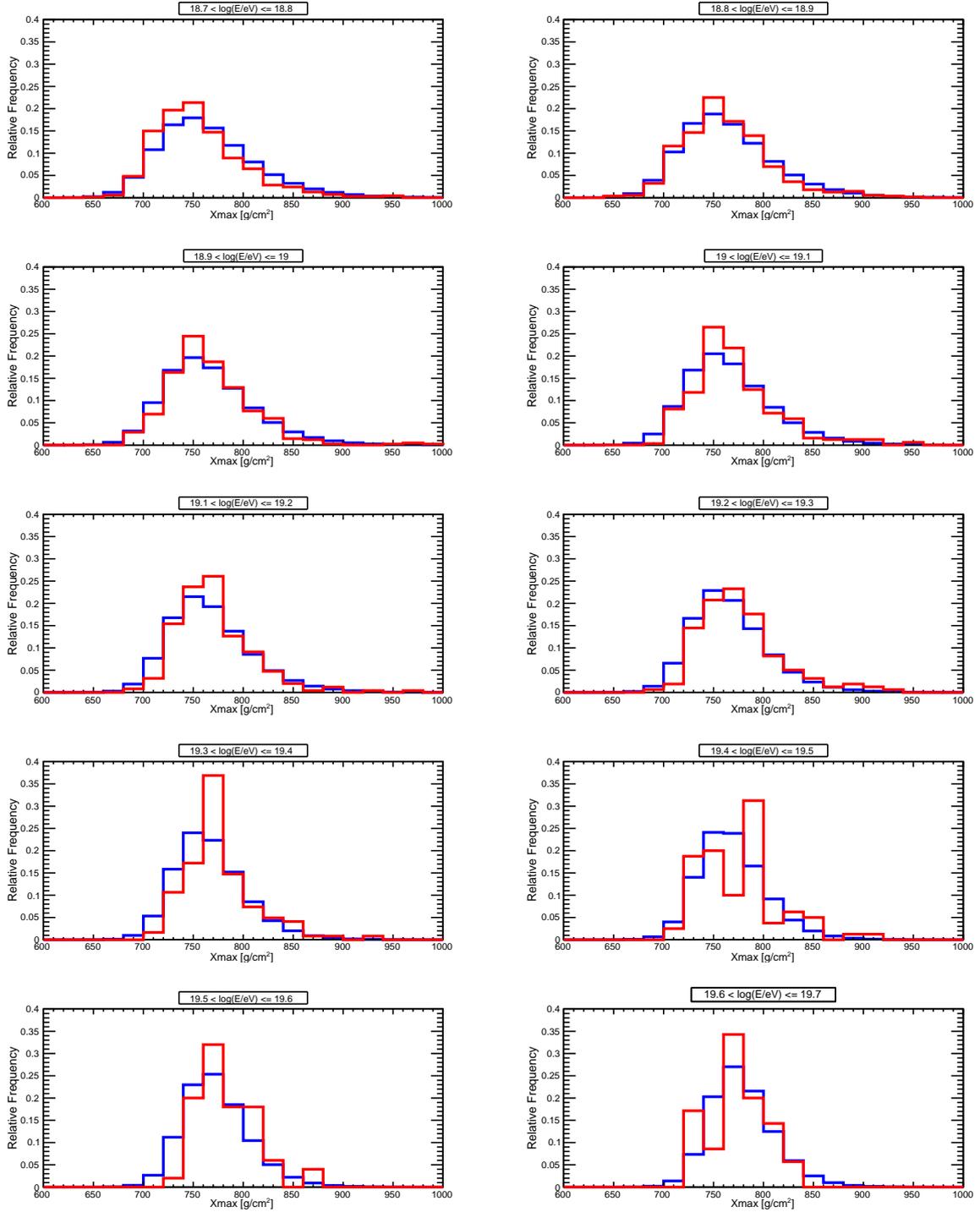


Figure B.20: X_{\max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset.

Figures B.21 and B.22 show all the results for the best fit using the scenario CTGS, with $\sigma_0 = 20$, and uncertainties on the X_{\max} , reported in Table 8.6.

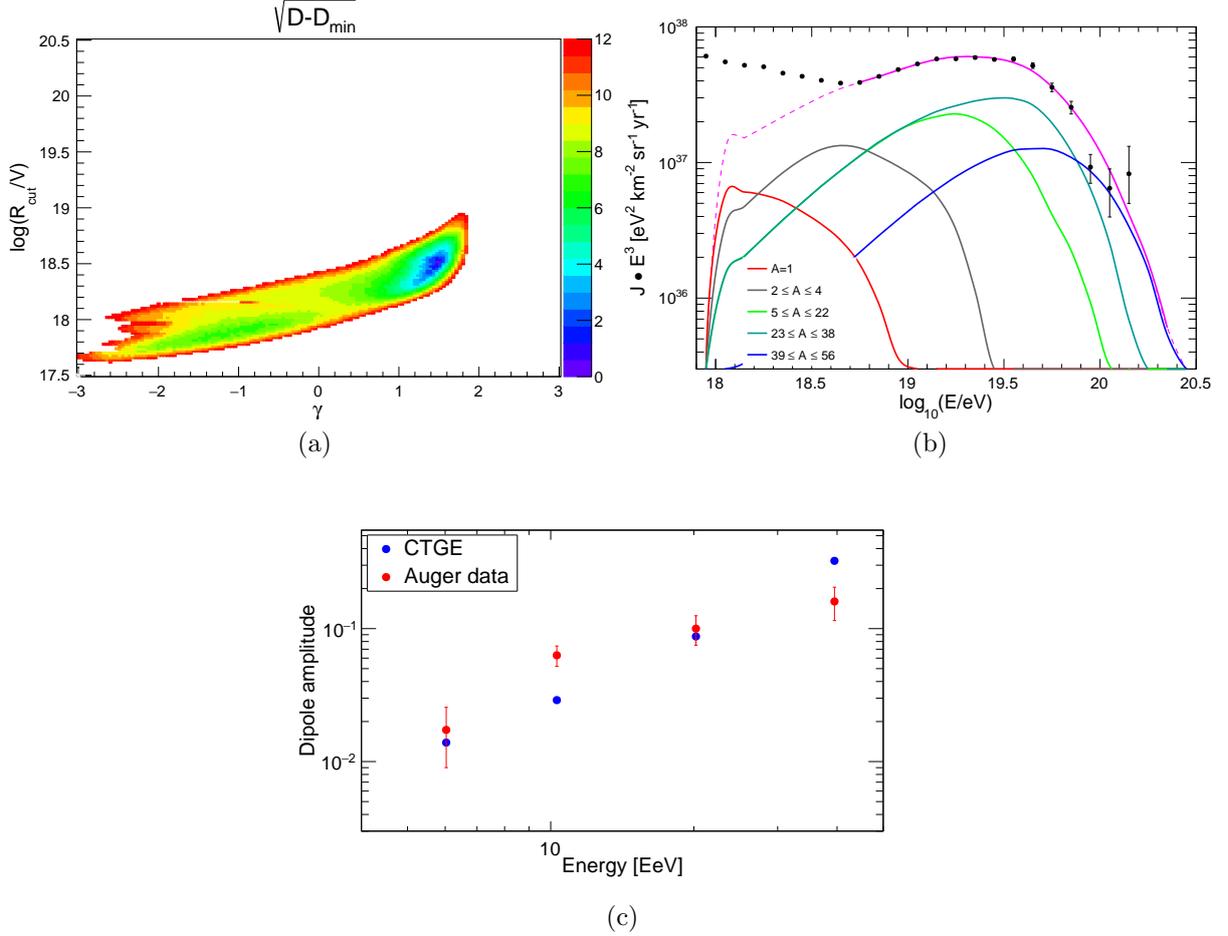


Figure B.21: (a) Minimum values of $\sqrt{D - D_{\min}}$ as a function of $(\gamma, \log_{10} R_{\text{cut}}/V)$ obtained using the package Minuit. (b) Energy spectrum of UHECRs (multiplied by E^3) for the best fit. The panel shows the abundances at the sources and the labels, the contribution of arriving particles within the mass number ranges. ICRC 2019 dataset. (c) Dipole amplitudes as a function of energy expected from the model (in blue) and measured by the Pierre Auger Collaboration (in red).

APPENDIX B. RESULTS IN THE BEST-FIT SCENARIO

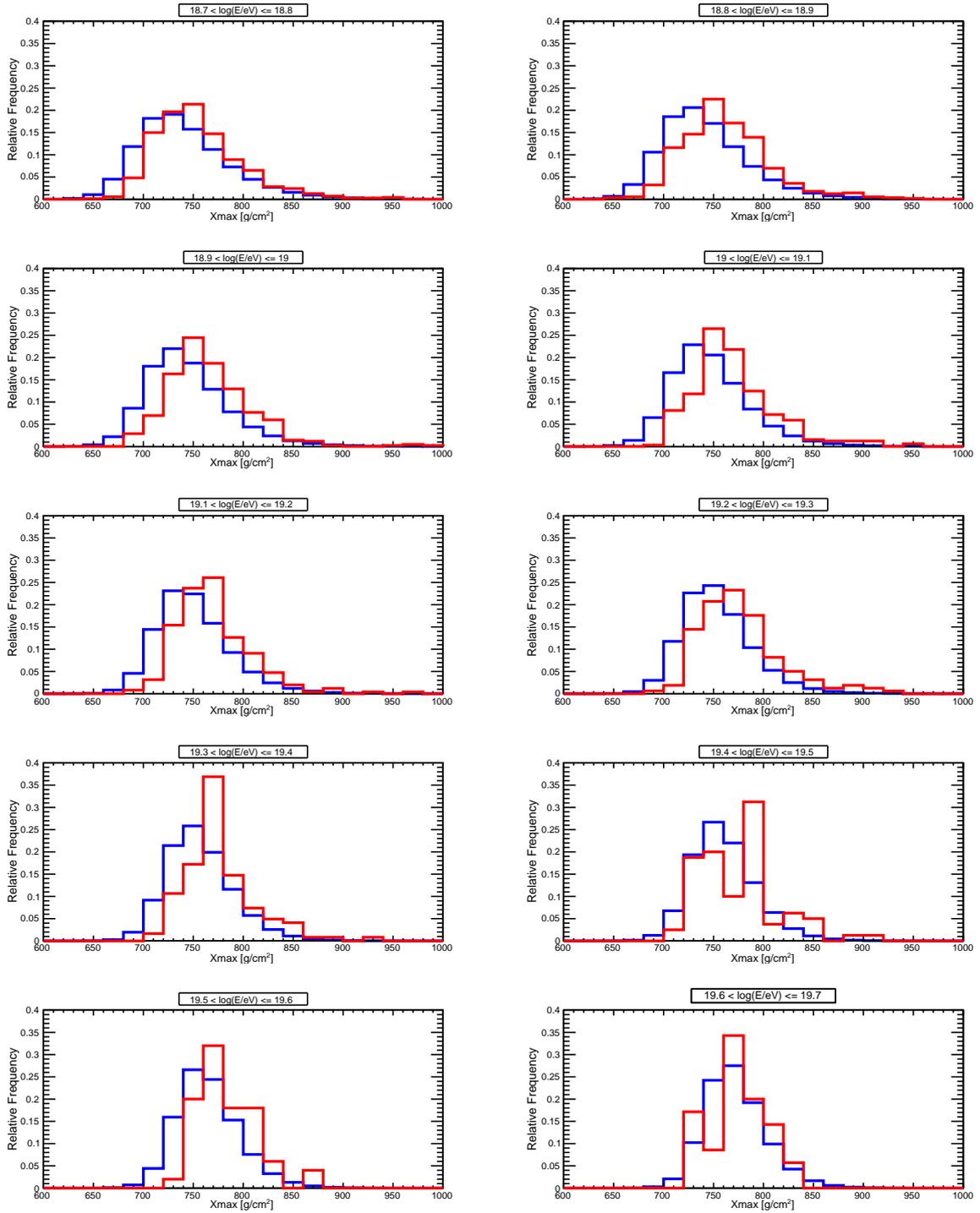


Figure B.22: X_{\max} distributions from the model for the best fit (blue histogram) and data (red histogram) for different energy bins. ICRC 2019 dataset.

B.1 Running the CODE

The range of the scan and the increment steps of both parameters can be set via an input file.

B.1.1 The input file

- `GammaLow = -3`. Sets the minimum value for the scan in γ . If `GammaStep == 0` the code runs in *full optimization* mode and the initial value of γ passed to the MINUIT is $(\text{GammaHigh} + \text{GammaLow})/2$.
- `GammaHigh = 3`. Sets the maximum value for the scan in γ .
- `GammaStep = 0.02` Sets the step for the scan in γ . Toggles *full optimization* mode if set to 0.
- `logRcutLow = 19.0` Sets the minimum value for the scan in $\log_{10} R_{\text{cut}}/V$. If `logRcutStep == 0` the code runs in *full optimization* mode and the initial value of $\log_{10} R_{\text{cut}}/V$ passed to the MINUIT is $(\text{logRcutHigh} + \text{logRcutLow})/2$.
- `logRcutHigh = 22`. Sets the maximum value for the scan in $\log_{10} R_{\text{cut}}/V$.
- `logRcutStep = 0.02` Sets the step for the scan in $\log_{10} R_{\text{cut}}/V$. Toggles *full optimization* mode if set to 0.
- `SpecData = ICRC2019` Switches the spectrum data. Possible inputs are ICRC2015, ICRC2019 and PRL2020.
- `SpecType = Unfolded`. Sets the spectrum type.
- `oldXmaxPar = 0` Toggles the *old* X_{max} parametrization.
- `XmaxPath = files_ICRC2019/` Specifies the path containing the X_{max} data files: `acceptance.txt`, `resolution.txt`, `xmaxHistograms.txt` and `xmaxMoments.txt`