



**HAL**  
open science

# Constraints on the population of ultra-high energy cosmic ray sources inferred from the data of the Pierre Auger Observatory

Sullivan Marafico

► **To cite this version:**

Sullivan Marafico. Constraints on the population of ultra-high energy cosmic ray sources inferred from the data of the Pierre Auger Observatory. *Cosmology and Extra-Galactic Astrophysics [astro-ph.CO]*. Université Paris-Saclay, 2022. English. NNT : 2022UPASP102 . tel-03985642

**HAL Id: tel-03985642**

**<https://theses.hal.science/tel-03985642>**

Submitted on 13 Feb 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Constraints on the population of ultra-high energy cosmic ray sources inferred from the data of the Pierre Auger Observatory

*Contraintes sur la population de sources des rayons cosmiques d'ultra-haute énergie à partir des données de l'observatoire Pierre Auger*

**Thèse de doctorat de l'université Paris-Saclay**

École doctorale n° 576, Particules Hadrons Énergie et Noyau :  
Instrumentation, Imagerie, Cosmos et Simulation (PHENIICS)  
Spécialité de doctorat : Astroparticules et Cosmologie  
Graduate School : Physique, Référent : Faculté des sciences d'Orsay

Thèse préparée dans l'unité de recherche **IJCLab (Université Paris-Saclay, CNRS)**, sous la direction d'**Olivier DELIGNY**, directeur de recherche, et le co-encadrement de **Jonathan BITEAU**, maître de conférence

**Thèse soutenue à Paris-Saclay, le 29 septembre 2022, par**

**Sullivan Marafico**

## Composition du Jury

Membres du jury avec voix délibérative

<b>Elias KHAN</b> Professeur, Université Paris Saclay	Président
<b>Corinne BERRAT</b> Directrice de recherche, LPSC (CNRS)	Rapporteur & Examinatrice
<b>Peter Tinyakov</b> Professeur, Université Libre de Bruxelles	Rapporteur & Examineur
<b>Martin Lemoine</b> Directeur de recherche, IAP (CNRS)	Examineur
<b>Véronique Van Elewyck</b> Maîtresse de conférences, Université de Paris	Examinatrice

université  
PARIS-SACLAY

**ÉCOLE DOCTORALE**

Particules, hadrons, énergie et noyau:  
instrumentation, imagerie, cosmos  
et simulation (PHENIICS)

# Contents

<b>Glossary</b>	<b>1</b>
<b>Résumé en français</b>	<b>3</b>
<b>Summary</b>	<b>9</b>
<b>1 Cosmic-ray introduction</b>	<b>15</b>
1.1 Discovery of cosmic rays . . . . .	15
1.1.1 An unexpected discovery . . . . .	15
1.1.2 Astroparticle spectrum . . . . .	17
1.2 Origin of Cosmic rays . . . . .	20
1.2.1 Acceleration mechanism . . . . .	21
1.2.2 Source candidates . . . . .	24
1.3 Propagation of UHECRs . . . . .	29
1.3.1 Cosmological framework and adiabatic losses . . . . .	29
1.3.2 The extragalactic medium is not empty! . . . . .	31
1.3.3 Attenuation length . . . . .	34
1.4 Extensive air showers . . . . .	36
1.4.1 Modelling . . . . .	37
1.4.2 Characteristics . . . . .	40
<b>2 The Pierre Auger Observatory and observables</b>	<b>43</b>
2.1 The Pierre Auger Observatory . . . . .	43
2.1.1 General view . . . . .	43
2.1.2 Other detectors . . . . .	48
2.2 Fluorescence detectors . . . . .	50
2.2.1 Fluorescence telescope: An extensive air shower camera . . . . .	50
2.2.2 Arrival direction reconstruction . . . . .	51
2.2.3 Energy and maximum depth of shower reconstruction . . . . .	52
2.3 Surface Detectors . . . . .	53
2.3.1 A water Cherenkov detector . . . . .	53
2.3.2 Arrival direction reconstruction . . . . .	54
2.3.3 Energy measurement . . . . .	55
2.4 In the Northern Hemisphere: Telescope Array . . . . .	58
<b>3 Arrival directions</b>	<b>61</b>



3.1	Reconstruction of flux map . . . . .	61
3.1.1	Method: $N/E_{\text{tot}}$ . . . . .	64
3.1.2	Method: $1/\omega$ . . . . .	66
3.1.3	Comparison of the two reconstruction methods . . . . .	67
3.2	Large scale anisotropy . . . . .	71
3.2.1	Rayleigh analysis in right ascension . . . . .	72
3.3	Smaller scales and composition-dependent anisotropy . . . . .	76
3.3.1	Intermediate scale anisotropy . . . . .	76
3.3.2	Indication of mass anisotropy . . . . .	79
<b>4</b>	<b>Study of the composition of UHECR</b>	<b>81</b>
4.1	From $X_{\text{max}}$ estimators to composition . . . . .	81
4.1.1	Fitting procedure . . . . .	82
4.1.2	Monte-Carlo simulations . . . . .	87
4.1.3	Results on data . . . . .	89
4.2	A biased approach? . . . . .	92
4.2.1	Test with fractions compatible with the boundaries . . . . .	92
4.2.2	Negative fractions? . . . . .	93
4.2.3	Test with parameters at boundaries . . . . .	95
4.3	Comparison between free and strictly positive fractions on data . . . . .	96
4.3.1	Bias vs resolution . . . . .	97
4.3.2	Discussion and conclusion . . . . .	98
<b>5</b>	<b>Combining spectrum and composition information</b>	<b>101</b>
5.1	Combined Fit methods . . . . .	102
5.1.1	Fraction and flux approach . . . . .	105
5.1.2	Emissivities . . . . .	107
5.1.3	Comparison between fraction and flux approaches . . . . .	107
5.1.4	Validation of the model . . . . .	108
5.1.5	Impact of the photodisintegration model . . . . .	109
5.1.6	Derivation of likelihood for the spectrum data . . . . .	110
5.1.7	Benchmark scenario . . . . .	111
5.2	Evolution term $S(z)$ . . . . .	115
5.2.1	Stellar mass density from Mpc to Gpc scales . . . . .	115
5.2.2	Star Formation Rate Density . . . . .	117
5.2.3	Results . . . . .	117
5.2.4	Discussion . . . . .	121
5.3	Across the ankle . . . . .	122
5.3.1	Proton spectrum and fit . . . . .	122
5.3.2	Results and discussion . . . . .	124

5.3.3	Escaped neutrons . . . . .	124
5.3.4	Results . . . . .	127
5.3.5	Systematic uncertainties . . . . .	128
5.3.6	Discussion . . . . .	129
5.4	Conclusion . . . . .	132
<b>6</b>	<b>An astrophysical model that describes the three observables</b>	<b>135</b>
6.1	From the Combined Fit to arrival directions . . . . .	135
6.1.1	Tensor approach . . . . .	135
6.1.2	Model . . . . .	138
6.1.3	Arrival direction map and comparison to data . . . . .	142
6.1.4	At the highest energy . . . . .	144
6.2	Impact of the cluster’s magnetic field . . . . .	146
6.2.1	Propagation inside the cluster . . . . .	147
6.2.2	Impact on the sky maps . . . . .	153
6.3	Below 1 Mpc? . . . . .	155
6.3.1	Impact of galaxies in the Local Group . . . . .	155
6.3.2	Transient events and time spread induced by magnetic envi- ronments . . . . .	156
6.3.3	Results and discussion . . . . .	160
6.4	Conclusion . . . . .	169
	<b>Conclusion</b>	<b>171</b>
	<b>Bibliography</b>	<b>175</b>
	<b>List of Figures</b>	<b>185</b>
	<b>List of Tables</b>	<b>195</b>
<b>A</b>	<b>Comparison of the two reconstruction methods using starburst study</b>	<b>197</b>
<b>B</b>	<b>Propagation in a turbulent magnetic field</b>	<b>201</b>
B.1	Starting from stochastic differential equation . . . . .	201
B.1.1	Taylor expansion . . . . .	203
B.1.2	Intermediate computation . . . . .	204
B.2	Time delay and duration of a burst . . . . .	205
B.2.1	Time delay . . . . .	205
B.2.2	Duration of a burst . . . . .	206
B.2.3	Deviation from the line of sight . . . . .	207
	<b>Remerciements</b>	<b>209</b>



# Glossary

**AGN:** Active Galactic Nuclei

**CC-SN:** Core Collapse Supernova

**CDAS:** Central Data Acquisition System

**CMB:** Cosmic Microwave Background

**DAQ:** Data Acquisition

**EBL:** Extragalactic Background Light

**FD:** Fluorescence Detector

**GRB:** Gamma-Ray Burst

**GZK limit:** Greisen Zatsepin Kuzmin limit

**ICRC:** International Cosmic Rays Conference

**LDF:** Lateral Distribution Function

**IGRB:** Long Gamma-Ray Burst

**LL-IGRB:** Low Luminosity Long Gamma-Ray Burst

**HL-IGRB:** High Luminosity Long Gamma-Ray Burst

**SD:** Surface Detector

**SFR:** Star Formation Rate

**SFRD:** Star Formation Rate Density

**SMD:** Star Mass Density

**SNR:** Supernova remnant

**TA:** Telescope Array

**TDE:** Tidal Disruption Event

**UHECR:** Ultra-High Energy Cosmic Ray

**WCD:** Water Cherenkov Detector



# Résumé en français

## Contexte

Les rayons cosmiques sont des noyaux d'atomes relativistes d'origine extraterrestre découverts au début du 20ème siècle. Les plus énergétiques d'entre eux, ceux dépassant les  $10^{17}$  eV, sont appelés rayons cosmiques d'ultra-haute énergie (RCUHE). Les origines des rayons cosmiques et en particulier des RCUHE restent un mystère. Les RCUHE qui arrivent sur Terre interagissent avec des noyaux d'oxygène ou d'azote présents dans l'atmosphère et produisent des gerbes atmosphériques. Ces gerbes atmosphériques sont détectées par l'observatoire Pierre Auger. C'est dans ce cadre que cette thèse fut écrite, dans le but de mieux comprendre leur origine à travers les mesures effectuées à l'Observatoire.

L'observatoire Pierre Auger est situé à Malargüe en Argentine, au pied de la cordillère des Andes. Il est constitué de deux types de détecteurs, des télescopes à fluorescence et des détecteurs de surface. Une gerbe atmosphérique est composée, notamment, de particules chargées de hautes énergies. Ces particules excitent les molécules de l'atmosphère qui, en se désexcitant, émettent de la lumière de fluorescence. Cette lumière est récoltée par les télescopes et permet de déduire des propriétés du rayon cosmique incident, telles que l'énergie, la direction d'arrivée et la profondeur maximum de la gerbe,  $X_{\max}$ , une longueur caractéristique qui peut être liée à la masse du rayon cosmique incident. Les télescopes n'étant fonctionnels que durant la nuit astronomique, ils sont complétés par des détecteurs de surface. Les détecteurs de surface sont des cuves d'eau à effet Tcherenkov. Il y en a 1,600 réparties sur  $3,000 \text{ km}^2$ , où elles sont espacées de 1,500 m chacune. Ces cuves détectent les particules chargées de la gerbe qui les atteignent. À partir de ces détections, la direction d'arrivée et l'énergie du RCUHE incident sont reconstruites.

En 2018, la collaboration Pierre Auger a publié l'observation d'un dipôle en ascension droite au-dessus de 8 EeV, permettant d'exclure l'hypothèse d'isotropie avec un niveau de confiance de  $5\sigma$  (The Pierre Auger Collaboration, 2018b). Ce dipôle pointe dans une direction à plus de  $100^\circ$  du centre galactique. De plus, il s'avère que la direction de l'anti-dipôle est compatible à  $1\sigma$  avec celle du vide local, une vaste région vide de l'univers adjacente à la nôtre (The Pierre Auger collaboration, 2022). Cette

découverte semble indiquer une origine extra-galactique des rayons cosmiques au-delà de 8 EeV. À plus haute énergie, une étude a récemment comparé les directions d'arrivée des rayons cosmiques avec des catalogues de sources extragalactiques (galaxies à flambée d'étoiles, noyaux actifs de galaxies, toutes les galaxies, etc.) et un fond isotrope. Au-dessus de 38 EeV, le catalogue de galaxies à flambée d'étoiles permet d'exclure l'hypothèse d'isotropie avec un niveau de confiance de  $4.2\sigma$  (The Pierre Auger collaboration, 2022).

Des études en composition (Bellido J. on behalf of the Pierre Auger Collaboration, 2018; The Pierre Auger Collaboration, 2014b) ont été menées par la collaboration Auger à partir des distributions en  $X_{\max}$ . L'étude revient à estimer les fractions de RCUHE semblables à des protons, de l'hélium, du CNO et de fer pour une gamme d'énergie donnée. La composition des RCUHE déduite par la collaboration Auger est une composition mixte, discréditant les modèles théoriques expliquant le spectre en utilisant des protons uniquement (Berezinsky, 2006). La composition est dominée par des protons aux alentours de l'EeV, puis la quantité de proton diminue, remplacée par de l'hélium, qui atteint son pique vers 10 EeV. Les plus hautes énergies semblent constituées majoritairement de noyaux de CNO, même si, vu les faibles statistiques, il est difficile de faire des analyses de composition au-delà de 40 EeV. Il faut cependant garder à l'esprit que les études en composition dépendent des modèles hadroniques utilisés, qui actuellement n'arrivent pas à reproduire les données dans toutes les gammes d'énergies balayées par l'observatoire.

Un modèle cosmologique tirant parti des études spectrales et en composition a été développé par la collaboration (The Pierre Auger Collaboration, 2017). Ce modèle suppose une origine extragalactique des rayons cosmiques au-dessus de 5 EeV. En partant d'une distribution de sources de RCUHE en fonction du redshift,  $S(z)$ , et d'un spectre de RCUHE injecté aux sources suivant une loi de puissance  $\Phi(E) \propto E^{-\gamma}$  avec une coupure ( $E_Z^{\max}$ ), des rayons cosmiques sont propagés dans l'univers et interagissent avec les photons des fonds diffus cosmologique et extragalactique. Pour limiter le nombre de paramètres ajustés, cinq nucléides caractéristiques sont injectés (H, He, N, Si, Fe). À chacun de ces nucléides est associé un paramètre qui traduit la quantité injectée. Une fois arrivés sur Terre, les rayons cosmiques sont comparés au spectre et aux distributions en  $X_{\max}$  (composition) et les paramètres optimaux sont déduits. Cette étude a permis de mettre en avant un indice spectral aux sources plus dur que celui attendu par les modèles d'accélération théorique de Fermi  $\gamma < 2$  (cf. 1.2.1).

## Les travaux issus de la thèse

En 2020, la collaboration a décidé de rendre public 10% de ses données à la communauté scientifique. Pour faciliter la prise en main de ces données, des notebooks python ont été réalisés reprenant les analyses majeures de la collaboration Auger. Dans le chapitre 3, une analyse simplifiée du dipôle décrit plus haut est faite et présentée. Cette analyse se retrouve dans le notebook d'anisotropie disponible sur le site de l'Open data (Marafico, Biteau, & Deligny, 2020). De plus, dans le chapitre 3, une étude a été menée pour comparer et caractériser en termes de biais et de variance deux méthodes de reconstruction utilisées dans la collaboration pour réaliser des cartes de flux à partir des données de l'observatoire Auger. Les résultats montrent que les deux méthodes semblent équivalentes en terme de biais et de variance.

L'étude en composition utilisée dans Bellido J. on behalf of the Pierre Auger Collaboration, 2018; The Pierre Auger Collaboration, 2014b a été reproduite en créant un code indépendant. Ce code pourra être utilisé dans le futur pour faire de nouvelles analyses de composition ou servir de code de vérification. Une étude a été réalisée pour caractériser le biais et l'écart-type à partir de données et de simulations. Il s'avère que les fractions déduites dans le cadre de cette analyse peuvent être biaisées dans le cas où le nombre d'évènements détectés est faible et que les fractions sont proches de zéro ou de un. Une approche autorisant les fractions négatives est proposée comme solution pour obtenir des fractions non biaisées. Cette approche permet de faire disparaître le biais, mais au prix fort : l'incertitude sur les fractions augmente et l'interprétation de fractions négatives devient plus subtile.

L'étude combinée de la composition et du spectre présentée au Chap. 5 est faite au-dessus de 5 EeV, en considérant des sources extragalactiques. Alors que l'étude faite par la collaboration utilise un scénario dans lequel les sources des RCUHE suivent une distribution plate en redshift  $S(z) = 1$ , ou une évolution cosmologique  $S(z) = (1 + z)^m$  avec  $m$  quelconque, le choix fait dans cette thèse est de tester deux scénarios. Dans le premier scénario, les sources suivent la densité de taux de formation d'étoiles (SFRD), dans le second, les sources suivent la densité de masse stellaire (SMD). Ces deux évolutions de sources prennent en compte l'évolution cosmique et locale. En effet, même si l'univers est homogène à grande échelle de distance, le SFRD et le SMD sont localement impactés par la distribution des galaxies se trouvant dans les feuillets, les filaments, les clusters et par la présence de vides cosmiques.



Les deux scénarios reproduisent similairement les données avec une déviance réduite proche de 2 (similaire à un  $\chi^2$  réduit). Cette déviance assez haute met en avant les difficultés des modèles hadroniques à reproduire les distributions en  $X_{\max}$ . Malgré le fait que les deux scénarios reproduisent les données avec une déviance similaire, le spectre injecté aux sources, quant à lui, diffère. Le scénario SMD est un scénario avec un spectre dur ( $\gamma \sim -0.4$ ) dominé par l'azote et l'hélium, avec des émissivités respectives de  $\bar{\mathcal{L}}_{\text{N}} = (12.7 \pm 1.4) \times 10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}$  et  $\bar{\mathcal{L}}_{\text{He}} = (22.1 \pm 2.0) \times 10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}$ . Les résultats du scénario SMD sont similaires à un scénario où la distribution des sources est plate car en première approche, pour les sources situées à distance inférieure à 1 Gpc, la densité de masse stellaire évolue peu. De l'autre côté, le scénario SFRD donne lieu à un spectre encore plus dur ( $\gamma \sim -1.8$ ) avec un spectre dominé par l'azote représentant 80% de l'émissivité totale,  $\bar{\mathcal{L}}_{\text{N}} = (191.0 \pm 2.3) \times 10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}$ . L'émissivité totale est six fois plus grande dans le scénario SFRD comparé au scénario SMD. Ceci est dû au fait que le taux de formation d'étoiles a atteint son pic il y a  $\sim 10$  Gyr et que bon nombre de noyaux perdent une partie importante de leur énergie sur de telles distances.

Jusqu'à maintenant, l'étude se basait sur le spectre et la composition au-delà de l'énergie de la cheville (5 EeV). Si l'accélération des noyaux suit un cycle de Peters (cf. 1.2.1), les protons observés en deçà de l'énergie de la cheville pourraient avoir la même origine que les noyaux détectés au-delà. Le spectre de protons en deçà de la cheville peut être reconstruit avec les fractions de protons calculés dans l'étude de composition,  $J_{\text{p}}(E) = f_{\text{p}}(E) \times J(E)$  où  $J(E)$  est le flux total. Ce flux de protons est comparé aux protons détectés en deçà de la cheville au-dessus de  $10^{17.8} \text{ eV}$  (seuil d'efficacité des télescopes à fluorescence). Alors que le scénario SFRD permet d'expliquer les protons en deçà de la cheville de manière raisonnable (déviance réduite de  $\sim 2$ ), le scénario SMD, tel qu'il a été formulé jusqu'ici, ne permet pas de reproduire les protons en deçà de la cheville (déviance réduite de  $\sim 5$ ). Pour ce faire, un nouveau paramètre est pris en compte, le paramètre d'indice spectral  $\gamma$  utilisé précédent est séparé en deux, un indice spectral pour les protons  $\gamma_{\text{p}}$  et un pour les noyaux  $\gamma_{\text{A}}$ . Comme montré dans Biehl et al., 2018, les scénarios d'accélération et de radiation pourraient produire des neutrons qui, n'étant pas chargés, s'échapperaient facilement de l'accélérateur. Une fois échappé, ces neutrons se désintègreraient en protons, dont le spectre pourrait être plus mou que celui des noyaux plus lourds. Ce nouveau paramètre introduit n'a pas d'impact dans le scénario SFRD, alors qu'il est crucial dans le scénario SMD (la déviance réduite retombe  $\sim 2$  avec  $\gamma_{\text{p}} \sim 3$ ). Cette étude montre que si les protons en deçà de la cheville ont la même origine que les noyaux au-delà, les deux scénarios sont valables. Dans le scénario SFRD, les

protons détectés en deçà de la cheville sont des protons secondaires créés en chemin par l'interaction de noyaux avec des photons. Dans le scénario SMD, les protons sont issus d'un mécanisme d'échappement aux sources qui diffère entre protons et noyaux.

Afin de prédire le ciel attendu dans les scénarios précédents, l'étude du Chap. 6 part d'un catalogue de près de 400,000 galaxies situées dans les 350 Mpc autour de la Terre. On peut associer à chacune des galaxies un flux basé sur les paramètres obtenus dans l'étude combinée (Chap. 5). Connaissant la position de chacune des galaxies, des cartes du ciel peuvent être produites et comparées aux données. Les deux scénarios (SMD et SFRD) donnent des résultats similaires en termes de cartes du ciel au premier ordre. Au-dessus de 40 EeV, les cartes sont dominées par l'amas de la Vierge. Cependant, l'amas de la Vierge n'est pas dominant dans les données en RCUHE. Le faible flux de cet amas peut être expliqué par le confinement des rayons cosmiques dû au champ magnétique présent dans l'amas. Un modèle simplifié est développé pour prendre en compte cet effet pour l'amas de la Vierge. L'aspect transitoire des sources de rayons cosmiques est aussi étudié. Si les sources sont transitoires, elles ne devraient être visibles depuis la Terre que durant un temps  $\Delta\tau$  correspondant à l'étalement dû aux champs magnétiques présents le long de la ligne de visée. L'étude montre que dans un scénario transitoire, si le taux de salves de RCUHE est trop élevé, les galaxies très proches tels les nuages de Magellan, la galaxie d'Andromède ou même la Voie lactée devraient être visibles dans les cartes de flux en RCUHE. Or, ce n'est pas le cas. Ce critère est utilisé pour mettre une limite supérieure à ce taux de salves. D'un autre côté, si le taux de salves est trop faible, les galaxies du Conseil des Géants ne devraient pas être visibles, or ces galaxies sont probablement responsables des indications d'anisotropies au-delà de 40 EeV (The Pierre Auger collaboration, 2022). À partir de ces observations, on est capable de mettre une limite inférieure au taux de salves. La présence du Conseil des Géants et la non-présence des galaxies proches nous permettent aussi de mettre des contraintes sur le champ magnétique issu du feuillet local qui doit être de l'ordre de quelques nG dans un scénario transitoire. Pour finir, les contraintes sur le taux de salves sont comparées à des candidats astrophysiques. La comparaison montre que les meilleurs candidats dans un scénario transitoire suivant le SFRD sont les sursauts gammas longs de faible luminosité et les événements de rupture par effet de marée dans le scénario SMD.



# Summary

## Context

Cosmic rays are relativistic atomic nuclei of extraterrestrial origin discovered in the early 20th century. The most energetic of them, those exceeding  $10^{17}$  eV, are called ultra-high energy cosmic rays (UHECR). The origins of cosmic rays, and in particular UHECRs, remain a mystery. UHECRs that arrive at Earth interact with oxygen or nitrogen nuclei in the atmosphere and produce extensive air showers. These atmospheric showers are detected by the Pierre Auger Observatory. It is within this framework that this thesis was written, with the aim of better understanding their origin through the measurements of the Pierre Auger Observatory.

The Pierre Auger Observatory is located in Malargüe, Argentina, a stone's throw from the Andes Mountains. The observatory consists of two types of detectors, fluorescence telescopes and surface detectors. An atmospheric shower is composed, in particular, of high-energy charged particles. These particles excite the molecules in the atmosphere, which then de-excite and emit fluorescent light. This light is collected by telescopes and allows us to deduce properties of the incident cosmic ray, such as the energy, the direction of arrival and the depth of shower maximum,  $X_{\max}$ , a characteristic length that can be related to the mass of the incident cosmic ray. As the telescopes are only operational during the astronomical night, they are supplemented by surface detectors. The surface detectors are Cherenkov water tanks. There are 1,600 spread over 3,000 km<sup>2</sup>, where they are spaced at 1,500 m each. These tanks detect the charged particles from the extensive air showers that reach them. From these detections, the direction of arrival and the energy of the incident UHECR are reconstructed.

In 2018, the Pierre Auger collaboration published the observation of a dipole in right ascension above 8 EeV, with a confidence level of  $5\sigma$  (The Pierre Auger Collaboration, 2018b). This dipole points in a direction more than  $100^\circ$  away from the galactic centre. Moreover, it turns out that the direction of the anti-dipole is at the  $1\sigma$  confidence level consistent with that of the local void, a vast empty region of the universe adjacent to our own (The Pierre Auger collaboration, 2022). This discovery seems to indicate an extra-galactic origin of cosmic rays beyond 8 EeV.

At higher energies, a recent study compared the arrival directions of cosmic rays with catalogues of extragalactic sources (starburst galaxies, active galactic nuclei, all galaxies, etc.) and an isotropic background. Above 38 EeV, the catalogue of starburst galaxies is preferred with a confidence level of  $4.2\sigma$  compared to isotropy.

Compositional studies (Bellido J. on behalf of the Pierre Auger Collaboration, 2018; The Pierre Auger Collaboration, 2014b) have been carried out by the Auger collaboration using  $X_{\max}$  distributions. The study amounts to estimating the fractions of proton-like, helium-like, CNO-like and iron-like UHECRs for a given energy range. The composition of the UHECRs deduced by the Auger collaboration is a mixed composition, discrediting theoretical models explaining the spectrum using protons only (Berezinsky, 2006). The composition is dominated by protons around EeV, then the amount of proton decreases, replaced by helium, which reaches its peak around 10 EeV. The highest energies seem to be constituted mainly by CNO nuclei, even if, given the low statistics, it is difficult to make compositional analyses beyond 40 EeV. It should be borne in mind, however, that compositional studies depend on the hadronic models used, which are currently unable to reproduce the data in all the energy ranges observed by the observatory.

A cosmological model based on spectral and compositional studies has been developed by the collaboration (The Pierre Auger Collaboration, 2017). This model assumes an extragalactic origin of cosmic rays above 5 EeV. Starting from a distribution of UHECR sources as a function of redshift,  $S(z)$ , and an UHECR spectrum injected at the sources following a power law  $\Phi(E) \propto E^{-\gamma}$  with a cutoff ( $E_{\frac{Z}{2}}^{\max}$ ), cosmic rays are propagated in the universe and interact with photons of the cosmic microwave background and of the extragalactic background light. To limit the number of parameters adjusted, five characteristic nuclides are injected (H, He, N, Si, Fe), each of these nuclides is associated with a parameter that reflects the quantity injected. Once on earth, the cosmic rays are compared with the spectrum and the  $X_{\max}$  distributions (composition) and the optimal parameters are deduced. This study has shown that the spectral index at the sources is harder than that expected by the theoretical Fermi acceleration models  $\gamma < 2$  (cf. 1.2.1).

## The work resulting from the thesis

In 2020, the collaboration decided to make 10% of its data public to the scientific community. To facilitate the handling of these data, python notebooks have been made containing the major analyses of the Auger collaboration. In chapter 3, a simplified analysis of the dipole described above is made and presented. This analysis can be found in the anisotropy notebook available on the Open Data website. In addition, in chapter 3, a study was carried out to compare and characterize in terms of bias and variance two reconstruction methods used in the collaboration to produce flux maps from the Auger Observatory data. The results show that the two methods seem equivalent in terms of bias and variance.

The compositional study used in Bellido J. on behalf of the Pierre Auger Collaboration, 2018; The Pierre Auger Collaboration, 2014b has been replicated by creating an independent code. This code can be used in the future for further compositional analysis or as a verification code. A study was carried out to characterize the bias and standard deviation from data and simulations. It was found that the fractions deduced in this analysis may be biased in the case where the number of events detected is small, and the fractions are close to zero or one. An approach allowing negative fractions is proposed as a solution to obtain unbiased fractions. This approach allows the bias to disappear, but at a high price: the uncertainty on the fractions increases and the interpretation of negative fractions becomes more subtle.

The combined compositional and spectral study presented in Chap. 5 is done above 5 EeV, considering extragalactic sources. While the study done by the collaboration uses a scenario in which the UHECR sources follow a flat redshift distribution  $S(z) = 1$ , or a cosmological evolution  $S(z) = (1 + z)^m$  with any  $m$ , the choice made in this thesis is to test two scenarios. In the first scenario, the sources follow the Star Formation Rate Density (SFRD), in the second, the sources follow the Stellar Mass Density (SMD). Both source evolutions take into account cosmic and local evolutions. Indeed, even if the universe is homogeneous at large distances, locally the SFRD and SMD are impacted by the distribution of galaxies in sheets, filaments, clusters and by the presence of cosmic voids.

Both scenarios reproduce the data similarly with a reduced deviance close to 2 (similar to a reduced  $\chi^2$ ). This rather high deviance highlights the difficulties of hadronic models to reproduce  $X_{\max}$  distributions. Despite the fact that both scenarios reproduce the data with a similar deviance, the spectrum injected at the sources differs. The SMD scenario is a hard spectrum scenario ( $\gamma \sim -0.4$ )

dominated by nitrogen and helium, with respective emissivities of  $\bar{\mathcal{L}}_{\text{N}} = (12.7 \pm 1.4) \times 10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}$  and  $\bar{\mathcal{L}}_{\text{He}} = (22.1 \pm 2.0) \times 10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}$ . The results of the SMD scenario are similar to a scenario where the distribution of sources is flat because, in the first approach, for sources located at distances smaller than 1 Gpc, the stellar mass density evolves little. On the other hand, the SFRD scenario gives rise to a harder spectrum ( $\gamma \sim -1.8$ ) with a spectrum dominated by nitrogen representing 80% of the total emissivity,  $\bar{\mathcal{L}}_{\text{N}} = (191.0 \pm 2.3) \times 10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}$ . The total emissivity is six times higher in the SFRD scenario compared to the SMD scenario. This is due to the fact that the star formation rate peaked  $\sim 10$  Gyr ago and many nuclei lose a significant amount of energy over such distances.

Until now, the study has been based on the spectrum and composition above the ankle (5 EeV). If the acceleration of the nuclei follows a Peters cycle (cf. 1.2.1), the protons observed below the ankle could have the same origin as the nuclei detected above. The proton spectrum below the ankle can be reconstructed with the proton fractions calculated in the composition study,  $J_{\text{p}}(E) = f_{\text{p}}(E) \times J(E)$  where  $J(E)$  is the total flux. This proton flux is compared to the protons detected under the ankle above  $10^{17.8}$  eV (the efficiency threshold of fluorescence telescopes). While the SFRD scenario allows explaining the protons below the ankle reasonably (deviance reduced by  $\sim 2$ ), the SMD scenario, as it has been formulated so far, does not allow reproducing the protons below the ankle (deviance reduced by  $\sim 5$ ). To do this, a new parameter is taken into account, the spectral index  $\gamma$  used previously is split into two, one spectral index for protons  $\gamma_{\text{p}}$  and one for nuclei  $\gamma_{\text{A}}$ . As shown in Biehl et al., 2018, the acceleration and radiation scenarios could produce neutrons which, being uncharged, would easily escape from the accelerator. Once escaped, these neutrons would decay into protons, whose spectrum could be softer than that of heavier nuclei. This newly introduced parameter has no impact in the SFRD scenario, whereas it is crucial in the SMD scenario (the reduced deviance drops to 2 with  $\gamma_{\text{p}} \sim 3$ ). This study shows that if the protons below the ankle have the same origin as the nuclei above, both scenarios are valid. In the SFRD scenario, the protons detected under the ankle are secondary protons created on the way by the interaction of nuclei with photons. In the SMD scenario, the protons originate from an escape mechanism from the sources that differs between protons and nuclei.

In order to predict the sky expected in the previous scenarios, the study of Chap. 6 starts from a catalogue of nearly 400,000 galaxies located in the 350 Mpc around the Earth. A flux can be associated with each galaxy based on the parameters obtained in the combined-fit study (Chap. 5). Knowing the position of each galaxy, sky maps can be produced and compared with the data. Both scenarios (SMD and SFRD) give similar results in terms of first-order sky maps. Above 40 EeV, the maps are

dominated by the Virgo cluster. However, the Virgo cluster is not dominant in the UHECR data. The low flux of this cluster can be explained by the confinement of cosmic rays due to the magnetic field present in the cluster. A simplified model is developed to take this effect into account for the Virgo cluster. The transient aspect of cosmic-ray sources is also studied. If the sources are transient, they should only be visible from Earth for a time  $\Delta\tau$  corresponding to the spread due to the magnetic fields along the line of sight. The study shows that in a transient scenario, if the UHECR burst rate is too high, very close galaxies such as the Magellanic Clouds, the Andromeda Galaxy or even the Milky Way should be visible in the UHECR flux maps, but this is not the case. This criterion is used to put an upper limit on the burst rate. On the other hand, if the burst rate is too low, the galaxies of the council of giants should not be visible, but these galaxies are probably responsible for the indications of anisotropies above 40 EeV. From these observations, we are able to put a lower limit on the burst rate. The presence of the council of giants and the non-presence of nearby galaxies also allows us to put constraints on the magnetic field from the local sheet which must be of the order of a few nG in a transient scenario. Finally, the constraints on the burst rate are compared with astrophysical candidates. The comparison shows that the best candidates in a transient scenario following the SFRD are long gamma-ray bursts with low luminosity and tidal disruption events in the SMD scenario.





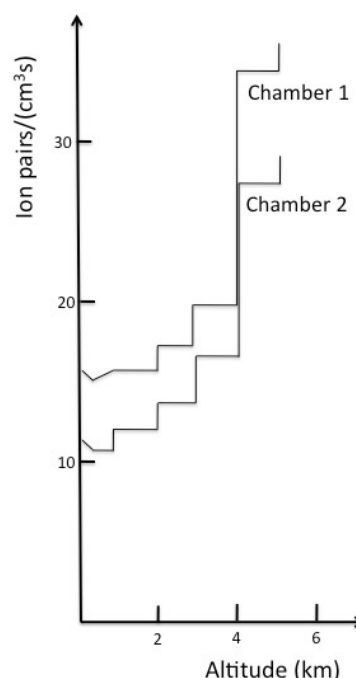
# Cosmic-ray introduction

## 1.1 Discovery of cosmic rays

### 1.1.1 An unexpected discovery

The discovery of cosmic rays (CRs) goes back to the beginning of the XIX century. At this time, physicists could measure the ionization rate of air using electrometers. It was believed that the ionization of air was coming from the radiation of radioactive elements discovered by Henri Becquerel in 1896. Theodor Wulf was the first one, in 1909, to design an experiment using an electrometer sealed in a hermetical container to study the dependence of the ionization rate on altitude. T. Wulf went to the Eiffel Tower and measured an ionization rate higher at its top than at ground level.

Two years later, in 1911, Domenico Pacini made a similar experiment measuring the ionization rate over the sea, over a lake and 3 meters below the surface of the water. He observed a significant decrease of the ionization rate underwater. He concluded that a specific part of the ionization rate should not come from the radioactivity of Earth. In 1912, Victor Hess went off to 5,300 m high in a hot air balloon flight (cf. Fig. 1.2). During the rise, he measured the ionization rate using different chambers (Hess, 1912). As shown in Fig. 1.1, the ionization increased significantly with altitude. He found that the ionization rate at 5,300 m was about four times the



**Fig. 1.1.:** Ionization rate measured by Victor Hess. Graph from Alessandro De Angelis, <https://commons.wikimedia.org/wiki/File:HessKol.jpg>.

one detected at ground level. It was the first proof that particles coming from the sky were ionizing the air. They decided to name them: Cosmic Rays.



**Fig. 1.2.:** Victor Hess making a balloon flight in order to measure the number of electroscope discharges for different heights.

New mysteries were raised after the experiments of Wulf, Pacini & Hess: What are those particles? Where are they coming from? The first question created a huge debate in the scientific community. Based on his experiments, Milikan proposed a theory where cosmic rays were high-energy photons produced in interstellar medium and propagated to Earth. The photons would transfer energy

to electrons via Compton effect when entering the atmosphere, thus it would explain the ionization rate increasing with altitude. Milikan's theory was not convincing to many scientists, especially to Arthur Compton, a former student of his.



**Fig. 1.3.:** New York Times article on the clash between Milikan and Compton about the nature of cosmic rays.

Compton was convinced that cosmic rays were charged particles. A debate began in the scientific community and went public ; articles were published in the famous New York Times about it (cf. Fig. 1.3). To prove that cosmic rays are charged particles, scientists were interested in measuring the cosmic-ray flux at different latitudes. Since charged particles would be deviated by Earth's magnetic field, one should expect the cosmic-ray flux to depend on latitude. In 1927, Jacob Clay, sailing from Java to the Netherlands, was the first one to find evidence of latitude dependence of cosmic-ray flux. At the same time, Compton went world-wide and collaborated with many scientists in order to measure cosmic-ray flux from Arctic to New Zealand. In 1932, he returned from his trip with measurements proving that Milikan was wrong. There is a clear latitude dependence, thus cosmic rays are charged particles.

### 1.1.2 Astroparticle spectrum

In the previous section, the question “What are those particles?” was not fully answered. In the '30s, Compton & others proved that the detected particles were charged. In the following decade, multiple investigations showed that cosmic rays were mainly protons, but all kinds of nuclei were also detected. Nowadays, all kind of high energy particles are detected in the sky. These particles refer to one domain of study: Astroparticle physics.

Figure 1.4 shows the spectrum of all kinds of detected astroparticles. The spectrum of astroparticles is given by  $\phi(E)$  with units  $[\phi(E)] = \text{eV}^{-1} \text{m}^{-2} \text{s}^{-1} \text{sr}^{-1}$ . Since the fluxes roughly follow a power law,  $\phi(E) \propto E^{-3}$ , it is common to represent the energy flux,  $E^2\phi(E)$  to enhance the spectral features. The second advantage to show it as energy flux is to ease the computation of the total energy in log-scale:  $\mathcal{E}_{\text{tot}} = \int E \phi(E) dE = \int E^2 \phi(E) d\log(E)$ .

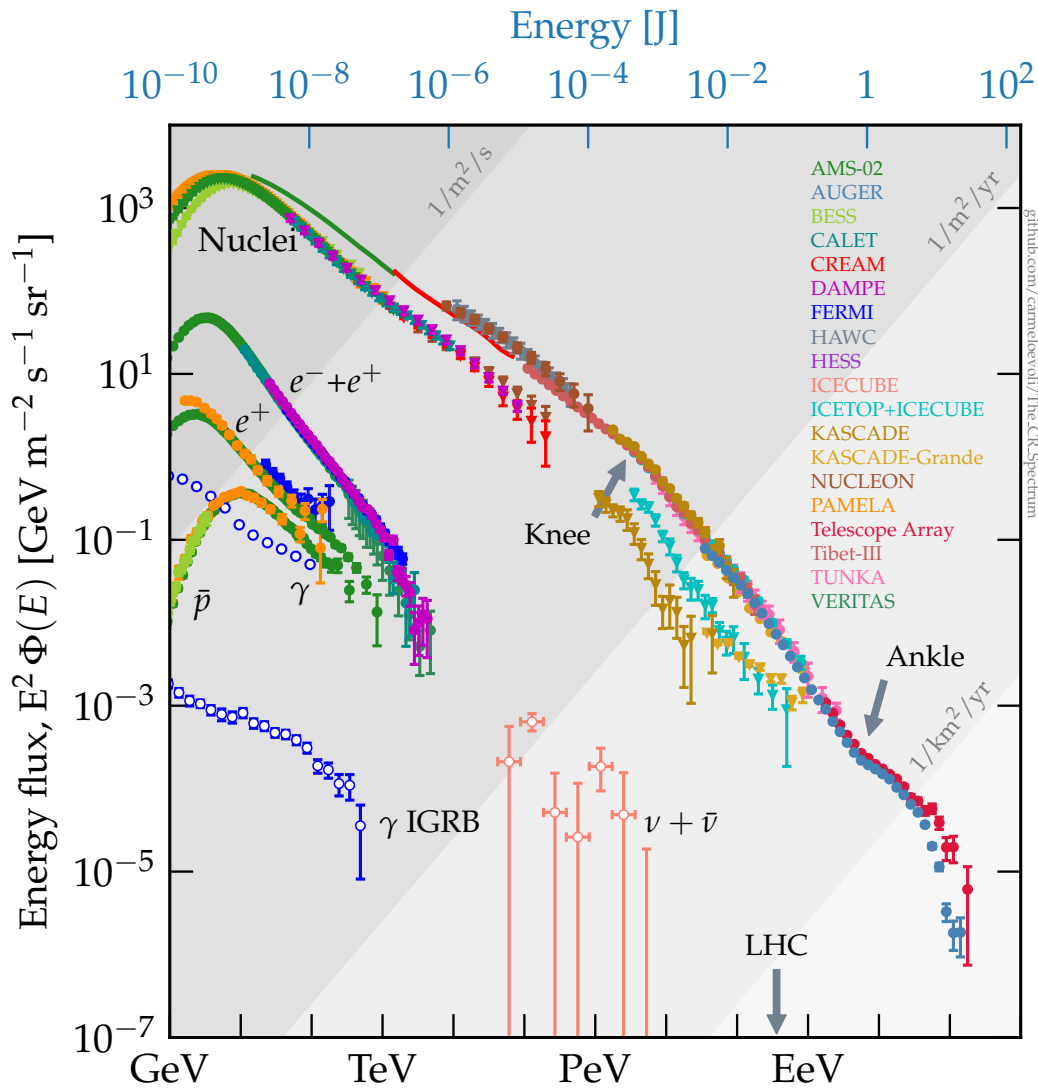


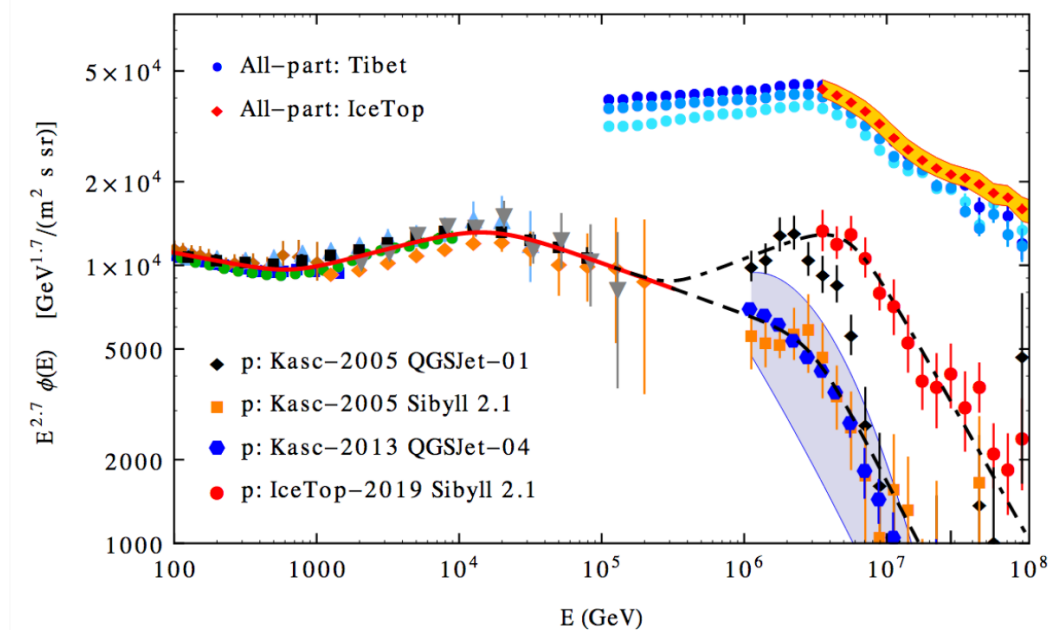
Fig. 1.4.: Astroparticle energy flux measured up to  $10^{21}$  eV and down to  $10^9$  eV. Adapted from [https://github.com/carmeloevoli/The\\_CR\\_Spectrum](https://github.com/carmeloevoli/The_CR_Spectrum)

In this PhD manuscript, the focus is put on Ultra High Energy Cosmic Rays (UHECR), which are cosmic rays above  $10^{17}$  eV. UHECRs are the witness of the most violent astrophysical phenomena. One can also expect that those violent astrophysical phenomena or/and the propagation of UHECRs have counterparts in neutrinos and in the isotropic gamma rays flux. On the other side, low-energy cosmic rays, positrons, electrons, anti-protons are, most of the time, treated in a separate way.

## Cosmic rays

The spectrum of cosmic rays shown in Fig. 1.4 goes from low-energies at  $10^9$  eV to  $10^{21}$  eV. The flux of cosmic rays spreads over 32 orders of magnitude. In terms of measurements, it can only be obtained by using very different observation methods. Several features of the spectrum have been well observed.

The first one, the steepening of the spectrum around 3 PeV, is the so-called knee. The corresponding composition has been measured by the KASCADE experiment (Antoni et al., 2005): the knee corresponds to a cut-off of light elements. One could also observe a second knee around 100 PeV (Apel et al., 2011). The second knee corresponds to a cut-off of heavy elements (Apel et al., 2013). As explained in the next subsection (cf. 1.2.1), this behaviour seems to look like the end of a mechanism of acceleration of cosmic rays. On top of the knee and the second knee, new features appear in the same energy range when considering the proton flux only. This is shown in Fig. 1.5.



**Fig. 1.5.:** Cosmic ray spectrum. In the top right corner, the flux of all particles is shown in blue circles (Tibet experiment) and red diamond (IceTop experiment). The others markers corresponds to the proton flux measured with different experiments. The red solid line is a fit using a two-break expression (cf. Lipari and Vernetto, 2020) that estimates the proton-flux below PeV energies. Above PeV energies, cosmic rays are measured through indirect measurement (cf. 1.4) which makes the deduced composition depend on the hadronic model. Therefore, different scenarios are shown in dashed black lines for the proton flux. Extracted from Lipari and Vernetto, 2020.

Figure 1.5 shows that the steepening of the proton flux changes as a function of energy. It becomes harder around  $\simeq 1$  TeV, then softer again around  $\simeq 10$  TeV. As said previously, a cut-off in proton is also observed at the knee. The systematics on the proton spectrum are not small enough to know whether the spectrum gets harder before the cut-off. Therefore, these features show that the simple description of cosmic-ray flux as a power-law is not valid. One way to explain those characteristics, would be that they correspond to the signature of a mechanism of acceleration or propagation. Many models have been put forward in order to reproduce this signature. However, another explanation would be to consider a variety of sources with different characteristics that sum-up to produce the global cosmic-ray spectrum (Lipari & Vernetto, 2020). In that scenario, the features below the knee correspond to the end of a mechanism of acceleration of a given source or group of sources, and the knee (and second knee) corresponds to the end of higher-energy emitters, for a given acceleration mechanism. It is believed that the second knee is a signature of the end of emission from sources within the Milky Way.

At 5 EeV, there is the ankle. For a long time, people expected the ankle to be the signature of the dip-model of Berezhinsky, where a single population of extragalactic protons explains the spectrum above 1 EeV (Berezhinsky, 2006). However, the mass composition has been revealed to be a mixed one at the Pierre Auger Observatory (cf. Chap 4), thus it is in conflict with the dip-model. Nowadays, the ankle is believed to correspond to the emergence of an extragalactic component (cf. Chap 3 & 4, and The Pierre Auger Collaboration, 2017). Between the second knee and the ankle, there is an unknown component: the component B as it was called by Hillas, 2005. The spectrum finishes with a two-step suppression. The first step, called the instep, corresponds to a steepening of the spectrum at  $\sim 10$  EeV. Finally, the cut-off at  $\sim 50$  EeV further reduces the flux observed up to  $\sim 100$  EeV (The Pierre Auger Collaboration, 2020c).

## 1.2 Origin of Cosmic rays

The origin of cosmic rays remains debated. However, multiple scenarios have been studied. In the following section, different mechanisms of acceleration and candidate sources of cosmic rays are discussed.

## 1.2.1 Acceleration mechanism

For a particle to gain energy, it must encounter an electric field. Since the electric field is reference frame dependent, Lorentz invariants are used to characterise an accelerating medium:  $\vec{E} \cdot \vec{B}$  and  $\vec{E}^2 - \frac{\vec{B}^2}{c^2}$ . From these Lorentz invariants, two kinds of acceleration can be discussed:

- Linear acceleration:  $\vec{E} \cdot \vec{B} \neq 0$  and  $\vec{E}^2 - \frac{\vec{B}^2}{c^2} > 0$ .
- Fermi acceleration:  $\vec{E} \cdot \vec{B} = 0$  and  $\vec{E}^2 - \frac{\vec{B}^2}{c^2} < 0$ .

The first one, also known as "one-shot" acceleration is hard to find in astrophysical environments. Ionized gases have conductivities similar to the conductivity of metal (Spitzer, 1962). Hence, if an electric field appears, it is shortly cancelled by the motion of free charges as in any conductors. Therefore, this kind of scenario is expected to only play a role in the case of low energy particle acceleration such as electron/positron in the environment of pulsars (magnetospheres) (Goldreich et al., 1969) or solar flares through magnetic reconnection (Shibata & Magara, 2011).

The second way of accelerating particles is described below in Fermi-type scenarios. In such a scenario, the particle propagates in the accelerator and sees a changing magnetic field. Although this scenario is not unanimous, it is nevertheless considered the most likely scenario for the acceleration of UHECRs.

### Second-order Fermi acceleration

A first model of acceleration in astrophysical environments has been developed by Enrico Fermi in 1949 (Fermi, 1949). Fermi proposed a stochastic model where the particles rebound on clouds in the interstellar medium. The clouds are considered as magnetic irregularities in the ambient magnetic field, moving at a random velocity  $V$ . Fermi showed that a particle gains energy after each collision with a cloud according to:

$$\left\langle \frac{\Delta E}{E} \right\rangle = \frac{8}{3} \left( \frac{V}{c} \right)^2. \quad (1.1)$$

If we consider that the mean free path between clouds is  $L$ , the time between collisions is given by  $\frac{L}{c \cos(\phi)}$  where  $\phi$  is the angle between the particle and the magnetic field direction. Averaging over  $\phi$ , the average time becomes  $\frac{2L}{c}$ . Then, the energy gain is given by:



$$\frac{dE}{dt} = \frac{4c}{3L} \left(\frac{V}{c}\right)^2 E = \alpha E. \quad (1.2)$$

Now, if a particle remains for a time  $\tau_{\text{esc}}$  in the accelerator, and neglecting the losses, using the transport equation with a continuous energy gain term (Longair, 2011), the flux of cosmic rays is given by:

$$\phi(E) \propto E^{-\left(1+\frac{1}{\alpha\tau_{\text{esc}}}\right)}. \quad (1.3)$$

This model is called the second-order Fermi acceleration, since the term  $\left\langle \frac{\Delta E}{E} \right\rangle$  is proportional to  $\left(\frac{V}{c}\right)^2$ . Such a scenario can explain a power-law observed in the spectrum. However, the second-order in  $\left(\frac{V}{c}\right)^2$  results in a very slow gain of energy per collision. Moreover, there is no particular reason that  $1 + \frac{1}{\alpha\tau_{\text{esc}}} \approx -2$ , as suggested by the spectrum (1.1.2).

### Diffusive shock acceleration

Diffusive shock acceleration (DSA) is a class of acceleration models that is one of the most promising theory to explain the cosmic-ray spectrum. In DSA models, cosmic rays cross a strong shock wave multiple times and gain energy each time they encounter the shock wave.

Let's define  $\beta$ , the energy gain after one collision,  $E = \beta E_0$ , and let's define  $p$ , the probability that the particle stays inside the accelerator after each collision. If we consider  $k$  collisions, one can deduce the number of particles in the accelerator  $N = N_0 p^k$ , and their energy  $E = E_0 \beta^k$ . Then, eliminating  $k$  yields:

$$\frac{\ln \frac{N}{N_0}}{\ln \frac{E}{E_0}} = \frac{\ln p}{\ln \beta}, \quad (1.4)$$

$$\implies \frac{N}{N_0} = \left(\frac{E}{E_0}\right)^{\frac{\ln p}{\ln \beta}}. \quad (1.5)$$

Here,  $N$  corresponds to the number of particles that reach at least the energy  $E$ , meaning  $N = N(\geq E)$ . The number of particles with energy between  $E$  and  $E + dE$  is then:

$$\phi(E) \propto E^{\frac{\ln p}{\ln \beta} - 1}. \quad (1.6)$$

When considering shock waves, one can show that the typical energy gain when a particle crosses the front back and forth is given by (cf. Longair, 2011):

$$\left\langle \frac{\Delta E}{E} \right\rangle = \frac{4}{3} \left( \frac{V}{c} \right), \quad (1.7)$$

which results in:

$$\beta = 1 + \frac{4V}{3c}. \quad (1.8)$$

On the other side, it can be proven (e.g. Bell, 1978) that  $p$  can be written as:

$$p = 1 - \frac{4V}{3c}. \quad (1.9)$$

The shocks are assumed to be non-relativistic, hence:

$$\frac{\ln p}{\ln \beta} = \frac{\ln 1 - \frac{4V}{3c}}{\ln 1 + \frac{4V}{3c}} \simeq \frac{-\frac{4V}{3c}}{\frac{4V}{3c}} \simeq -1. \quad (1.10)$$

Injecting eq. 1.10 into eq.1.6 one can get the expected spectrum:

$$\phi(E) \propto E^{-2}. \quad (1.11)$$

The DSA model naturally produces a spectrum  $\phi(E) \propto E^{-2}$  that roughly follows the observed spectrum in section 1.1.2. This model is also called the first-order Fermi acceleration model, since the energy gain after each shock is first-order in  $V$  (cf. eq. 1.7). This first order in  $V$  makes it very efficient compared to the second-order Fermi acceleration mechanism.

### Peters' cycle

Motivated by the knee feature in the spectrum (see 1.1.2), Peters proposed in 1961 (Peters, 1961) a scenario where cosmic rays are accelerated depending on their rigidity. Since the sources are supposed to accelerate cosmic rays using the electromagnetic force, one can expect them to do so up to a maximum rigidity  $R_{\max}$

in V. Then, the maximum energy depends on the injected nuclei, and it is given by:

$$E_{\max}(Z) = Ze \times R_{\max} = Z \times E_{\max}|_{\text{proton}} \quad (1.12)$$

Considering eq. 1.12, for one mechanism of acceleration, one could expect a progressive cut-off in energy: first a light element cut-off, then heavier element cut-offs. Indeed, this behaviour is observed in the spectrum at the knee (light elements cut-off) at 3 PeV, and at the second knee (heavy elements) at  $\sim 100$  PeV. If one considers that light elements are made of protons, and heavy elements of iron, the cut-off of iron nuclei is expected to be  $E_{\max}({}_{26}^{56}\text{Fe}) = 26 \times E_{\max}|_{\text{proton}} = 26 \times 3 \text{ PeV} \sim 100 \text{ PeV}$ , which roughly corresponds to the second knee energy.

## 1.2.2 Source candidates

The sources of cosmic rays are still unknown. However, after a century of research, the scientific community starts to have an idea of which could or could not be a source of cosmic rays.

The first condition on the sources of cosmic rays is expressed with the well-known Hillas condition (cf. Hillas, 1984). The Hillas condition is a necessary one for an astrophysical object to accelerate cosmic rays. It states that cosmic rays should remain spatially confined in the accelerator to be accelerated up to an energy  $E_{\max}$ . If we consider the gyration radius  $r_g$  of cosmic rays,  $r_g$  should be smaller than the size of the accelerator  $L$ . The gyration radius is given by  $r_g = E/Zc||\vec{B}||$ , where  $E$  is the energy,  $c$  the speed of light,  $Z$  is the charge of the nucleus, and  $\vec{B}$  the magnetic field. One can deduce that the energy of the cosmic rays should respect:

$$E \leq Zc||\vec{B}||L. \quad (1.13)$$

Equation 1.13 is the first argument of Hillas. One can refine this equation by computing the order of magnitude of the electric field in an accelerator:

$$\nabla \times E = \frac{\partial B}{\partial t} \implies \frac{E}{L} \sim \frac{B}{L/u}, \quad (1.14)$$

where  $u$  is a characteristic velocity which corresponds to the motion of the conductor in a one-shot acceleration, or to the scattering centres for the two types of Fermi accelerations. Then, the maximum energy is given by the electric field times the size of the accelerator:

$$E_{\max} = L \times Z \|\vec{E}\| = L \times ZuB, \quad (1.15)$$

$$E_{\max} = Z \left(\frac{u}{c}\right) \left(\frac{B}{1\mu\text{G}}\right) \left(\frac{L}{1\text{pc}}\right) \text{PeV}, \quad (1.16)$$

The second way of computing the Hillas condition is to consider Fermi acceleration and equating the time of acceleration with the diffusive escape time (Hillas, 1984).

### Low-energy cosmic rays

Cosmic rays with energies below the second knee are believed to be accelerated through DSA in strong shock waves. The strong shocks would be created by supernovae and propagate through the interstellar medium ahead of the supersonic shells of supernova remnants (SNRs). The first clue is found by computing the Hillas condition applied to SNRs, considering protons ( $Z = 1$ ) and typical value for  $u$ ,  $L$  and  $B$ , respectively 5000 km/s, 1 pc (from Lagage and Cesarsky, 1983) and  $1 \mu\text{G}$ . The maximum energy is found to be  $E_{\max} \simeq 20 \text{ TeV}$ , which is quite below the knee. This maximum energy can be increased if one considers the possible retroaction of the accelerated cosmic rays on the magnetic field ( $\vec{B}$  is amplified, as in Lucek and Bell, 2000). Hence, SNRs could match the maximum observed energy of protons around the knee.

The second clue comes from the energy budget of supernovae. Knowing the energy density of cosmic rays below the knee  $\rho = 10^{-12} \text{ erg cm}^{-3}$  estimated from the spectrum (Sec. 1.1.2), the Galactic confinement volume estimated with a cylinder  $V = \pi (20 \text{ kpc})^2 100 \text{ pc} \simeq 3 \times 10^{67} \text{ cm}^3$  and the cosmic-ray residence time  $\tau \sim 3 \times 10^7 \text{ yr}$  (Lipari, 2014), one can find the required power to fuel the Galactic energy density  $P = 3 \times 10^{40} \text{ erg s}^{-1}$ . It has to be compared with the energy input from all Galactic SNRs  $P_{\text{SNRs}} = 10^{42} \text{ erg s}^{-1}$ . Therefore, only a few percent of the energy budget of SNRs are needed to account for the energy density of cosmic rays up to the knee.

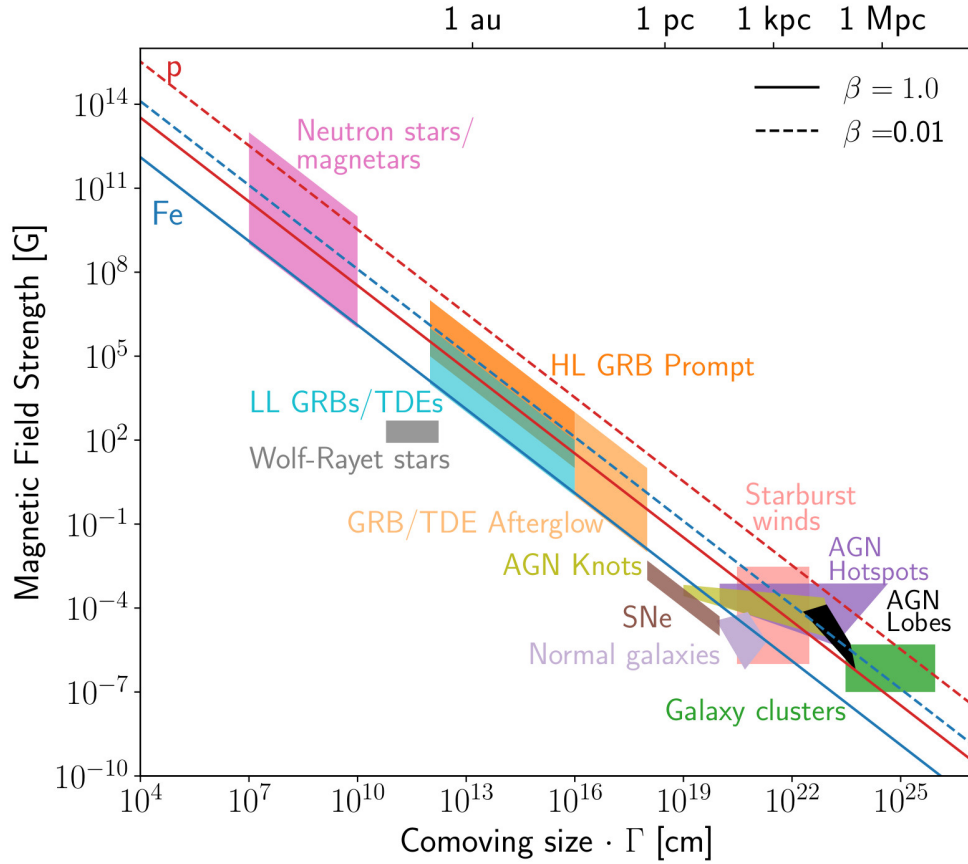
### UHECR sources

While there is no certainty about the sources of UHECRs, the scientific community has certainty about what is not a source of UHECRs. Following the discussion in Ptitsyna and Troitsky, 2010, one can see that an accelerator needs to satisfy several conditions in order to be considered as a potential candidate for UHECR source:

- Spatial: The particle should not leave the source before being accelerated (Hillas condition).
- Emissivities: The density of sources times their power should match the observed flux.
- Radiation losses: The energy lost through radiative processes should not overpass the energy gain.
- Interaction losses: The energy lost through interactions with other particles should not overpass the energy gain.
- Induced particles: Low energy cosmic rays, photons and neutrinos induced from the interactions should not outmatch the observed flux from a given source nor the diffuse background.

Here, the focus will be put on the spatial and emissivity conditions, because these conditions are less model dependent. The spatial condition is reached using the Hillas condition used before. For UHECRs, the Hillas condition can be written with a Lorentz factor  $\Gamma$  (as in eq. 1.17). The Lorentz factor comes from the fact that the acceleration region can be Lorentz boosted, which implies  $E_{\max} = E_{\max}^{\text{acc}} \Gamma$ . It can be the case for AGN jets ( $\Gamma \sim 10 - 50$ ) and for GRBs ( $\Gamma \sim 10 - 1000$ ) (Batista et al., 2019),

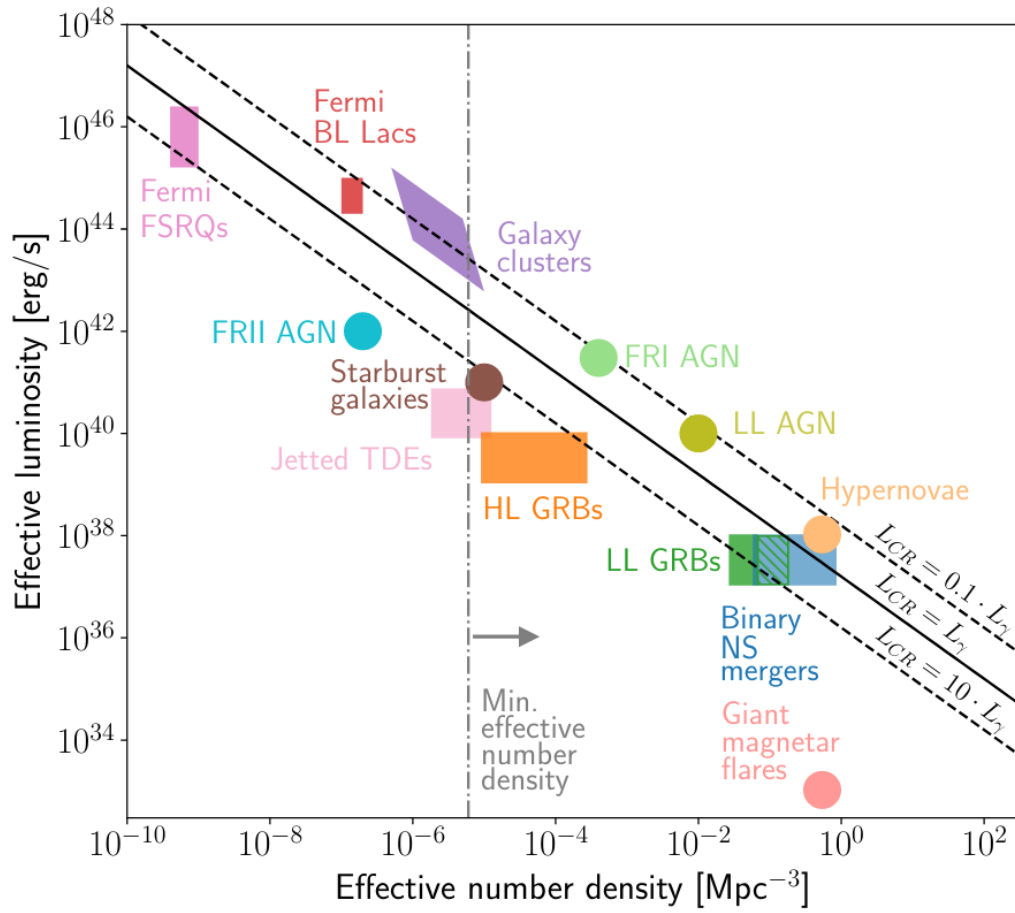
$$E_{\max} = Z\beta_{\text{sh}}BL\Gamma. \quad (1.17)$$



**Fig. 1.6.:** Hillas diagram. The abscissa shows the product  $L \times \Gamma$  which corresponds to the product of the comoving size of the engine times the Lorentz factor of the engine. Extracted from Batista et al., 2019.

Figure 1.6 is called Hillas diagram and helps to visualize the Hillas condition. The lines in Fig. 1.6 show the product  $BL\Gamma$  needed to confine iron (blue) or proton (red) with  $\beta_{\text{sh}} \simeq 1$  (solid) or  $\beta_{\text{sh}} \simeq 0.01$  (dashed) at  $10^{20}$  eV. On top of the lines, there are coloured areas corresponding to different astrophysical objects. One can see that the outflows of Supernovae, Normal galaxies and Wolf-Rayet stars do not have a sufficient  $BL\Gamma$  to confine nuclei long enough, for any  $\beta_{\text{sh}}$ . Hence, Supernovae, Normal galaxies and Wolf-Rayet stars do not fulfil the Hillas condition and can be excluded as sources of UHECR.

Another condition can be put on the energetics of the objects. In The Pierre Auger Collaboration, 2017, the emissivities  $\mathcal{L}$  in UHECR have been estimated above the ankle to be  $\mathcal{L} = 5 \times 10^{44} \text{ erg s}^{-1} \text{ yr}^{-1}$ .



**Fig. 1.7.:** Luminosity diagram. Effective luminosity of sources for transient sources<sup>1</sup> and average luminosity of sources as a function of source number density. The blue solid lines corresponds to the total luminosity of UHECR, estimated in The Pierre Auger Collaboration, 2017 which is  $\mathcal{L} = 5 \times 10^{44} \text{ erg Mpc}^{-3} \text{ yr}^{-1}$ . The orange dotted-dashed line gives the lower limit for the effective density of sources estimated from The Pierre Auger collaboration, 2013. Extracted from Batista et al., 2019.

In Fig. 1.7, the luminosity [erg/s] of different classes of sources are shown as a function of their effective number density [ $\text{Mpc}^{-3}$ ]. The luminosity of sources is deduced from various measurements in photons. The solid line corresponds to the estimated UHECR emissivity. If an astrophysical object is below that line, the photon luminosity is below the one observed in UHECR. However, it is not easy to compare the UHECR luminosity with that in photons. Hence, the dashed lines express two hypotheses, a correlation between the two of a factor 10 times bigger/smaller.

Looking at the anisotropy in UHECR sky, The Pierre Auger collaboration, 2013 estimated a minimum density of sources, using events above 70 EeV and an estimated

<sup>1</sup>Considering a characteristic time spread of  $3 \times 10^5 \text{ yr}$ .

spread due to magnetic deflections of  $30^\circ$ . This minimum density is shown as a dashed and dotted line in Fig. 1.7. Hence, gamma-ray emitting blazars (shown as Fermi FSRQs and BL Lacs), Active Galactic Nuclei of FR II type, galaxy clusters and giant magnetar flares are excluded as potential sources of UHECRs.

Hence, even if the sources are not known, multiple conditions help to exclude some candidates. In Chapter 4, we will see that on top of those minimum conditions, there are indications on the nature of the sources.

## 1.3 Propagation of UHECRs

After their acceleration in the sources, UHECRs travel through space, experiencing energy loss and/or photo-disintegration. These phenomena are explained in the following pages, starting with the cosmological framework in which the UHECRs evolve.

### 1.3.1 Cosmological framework and adiabatic losses

In 1915, Einstein published field equations that are the fundamental basis of General relativity (cf. eq. 1.18). The equation shows the link between the Einstein tensor  $G_{\mu\nu}$  expressing the curvature of the universe,  $g_{\mu\nu}$  expressing the metric tensor, and  $T_{\mu\nu}$  the stress-energy tensor that refers to matter itself:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}. \quad (1.18)$$

Here,  $\Lambda$  and  $G$  are respectively the cosmological and the gravitational constant. Starting from Einstein's equation (eq. 1.18), Friedman established in 1922 the equations that govern a homogeneous and isotropic universe. The first Friedman equation is obtained using the  $_{00}$  component of eq. 1.18:

$$\frac{\dot{a}(t)^2 + kc^2}{a(t)^2} = \frac{8\pi G\rho + \Lambda c^2}{3}, \quad (1.19)$$

where:

- the time  $t$  is taken as  $t = 0$  for nowadays and increases when one looks back in past,



- $a(t)$  is the scale factor, a dimensionless factor that shows how the distance between two celestial objects varies with time due to the expansion of the universe.  $a(t = 0) = 1$ ,  $a(t = t_{\text{big-bang}}) = 0$ , and any measured distance  $R_0$  at  $t = 0$  scales in time as:  $R(t) = R_0 a(t)$ ,
- $k$  is the current spatial curvature. If  $k > 0$ , the Universe is spatially closed and is a hypersphere, if  $k = 0$ , the Universe is spatially flat (preferred by the data), if  $k < 0$ , the Universe is spatially hyperbolic.

One can also express the scale factor in terms of redshift  $z$ :

$$a(t) = \frac{1}{1+z}, \quad (1.20)$$

and define the Hubble parameter as:

$$H(t) = \frac{\dot{a}(t)}{a(t)}. \quad (1.21)$$

Using equations 1.19, 1.20, 1.21, and considering a flat spatial curvature (in agreement with observations), one can find:

$$H(z) = H_0 \sqrt{\Omega_r(1+z)^4 + \Omega_m(1+z)^3 + \Omega_\Lambda} = -\frac{1}{1+z} \frac{dz}{dt}, \quad (1.22)$$

where  $\Omega_r$ ,  $\Omega_m$ ,  $\Omega_\Lambda$  correspond to the radiative density, matter density, and cosmological constant (or vacuum density), respectively, in critical density unit:  $\Omega \equiv \frac{\rho}{\rho_c} = \frac{3H_0^2 \rho}{8\pi G}$ .

If we consider an UHECR going through space with an energy  $E$ , one can associate to it its *de Broglie wavelength*,  $\lambda \simeq \frac{hc}{E}$  (because the particle is ultra-relativistic,  $E \gg mc^2$ ). The wavelength  $\lambda$  undergoes the expansion of the universe, which results in

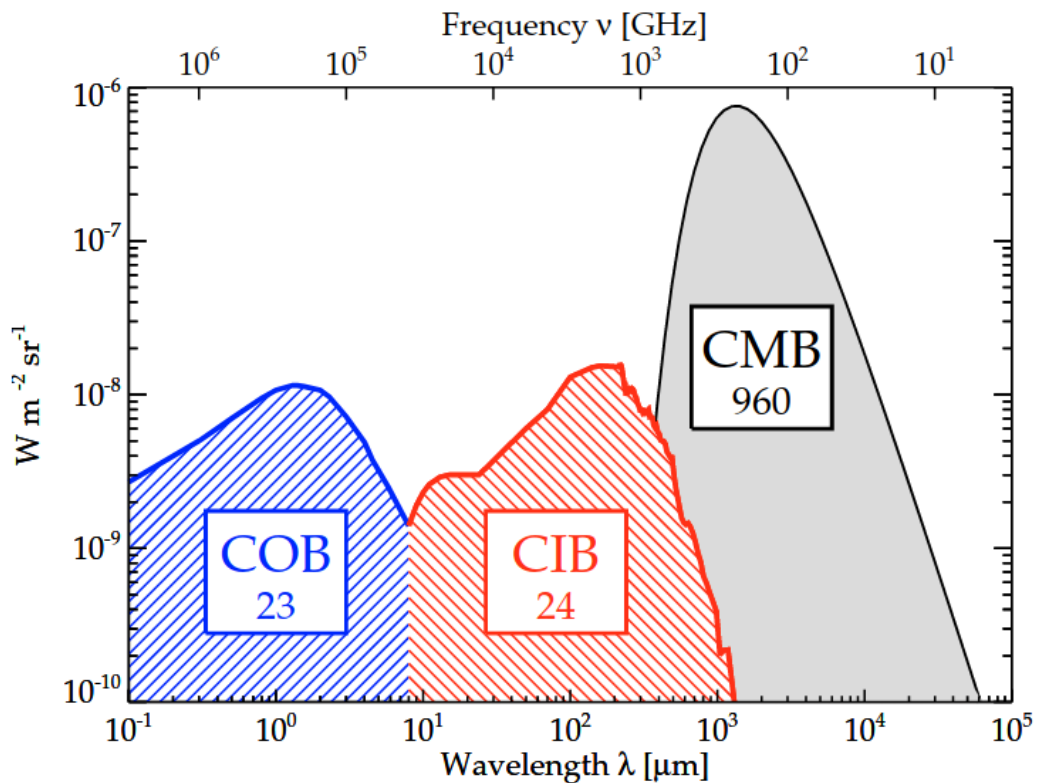
$$\frac{1}{E} \frac{dE}{dz} = \frac{\lambda}{hc(1+z)} \times \frac{hc}{\lambda} = \frac{1}{1+z}. \quad (1.23)$$

Now, defining the adiabatic losses as  $\frac{1}{E} \frac{dE}{dt} \Big|_{\text{ad}}$  and combining equations 1.22 and 1.23, one can find the adiabatic losses of a UHECR:

$$\frac{1}{E} \frac{dE}{dt} \Big|_{\text{ad}} = \frac{1}{E} \frac{dE}{dz} \frac{dz}{dt} = -H(z). \quad (1.24)$$

### 1.3.2 The extragalactic medium is not empty!

The extragalactic medium is not empty! While the presence of matter in the extragalactic medium is negligible for the propagation of UHECRs, the presence of photons has an impact. The first population of photons that interact with UHECRs are those of the Cosmic Microwave Background (CMB). CMB has been emitted  $\sim 380,000$  years after the Big Bang ( $z = 1100$ ) when protons bond with electrons to make hydrogen atom. This epoch is called the Recombination Era. The second population of photons of interest for the propagation of UHECRs are the photons of the Extragalactic Background Light (EBL). The EBL corresponds to the Ultraviolet/Optical/Infrared light emitted by all the stars, from the first ones up to now. The EBL can be split into two components, the Cosmic Optical Background (COB), which is also called the CUVOB because it is composed of optical and a small amount of UV. It comes from the direct emission of stars. The second component is the Cosmic Infrared Background (CIB). It comes from the re-emission of light from interstellar dust. The spectral of the EBL and the CMB are shown in Fig. 1.8.



**Fig. 1.8.:** Schematic spectral energy distribution of the CMB (grey) and the EBL (blue & red). The numbers in the boxes correspond to approximate brightness [ $\text{nW m}^{-2} \text{sr}^{-1}$ ]. Extracted from Dole et al., 2006.

UHECRs have a Lorentz factor  $\Gamma$  such that when they encounter a photon of the CMB/EBL, the energy of the photon in the rest frame of the nucleus is sufficient to damage it, in terms of mass (photo-disintegration) or/and in terms of energy (pair production or pion production). These phenomena are detailed below.

### Electron/positron pair production

Any nucleus that meets a photon with the energy above the  $E_\gamma = E_\gamma^{\min}$  can create an electron/positron pair. The electron/positron pair production follows from the reaction



The reaction can only occur if there is enough energy in the rest frame to create the electron/positron pair. The threshold can be found using the conservation of the norm of the four-momentum. It is done in eq. 1.26, by equating the norm of the four-momentum in the nucleus rest-frame before collision with the norm of the four-momentum after collision, where all particles are at rest (minimum energy):

$$\left| \begin{pmatrix} mc^2 \\ 0 \end{pmatrix} + \begin{pmatrix} E_\gamma^{\min} \\ E_\gamma^{\min} \end{pmatrix} \right|^2 = \left| \begin{pmatrix} mc^2 + 2m_e c^2 \\ 0 \end{pmatrix} \right|^2 \quad (1.26)$$

It results in a minimum energy of the photon in the rest frame of  $E_\gamma^{\min} = 2m_e c^2 (1 + \frac{m_e}{m}) \simeq 1 \text{ MeV}$ .

As shown in Berezinsky, 2006, creating a  $e^\pm$  pair results in a small energy loss of the order of  $f = 0.1\%$  when  $E_\gamma = E_\gamma^{\min}$ . This fraction  $f$  is monotonically decreasing with increasing  $E_\gamma$ . It can be interesting to compute the product fraction of energy loss per interaction  $f(E_\gamma)$  times the cross-section  $\sigma(E_\gamma)$ . For a proton, it is roughly constant for  $E_\gamma \geq 5 \text{ MeV}$ ,  $f(E_\gamma)\sigma(E_\gamma) \sim 0.5 \mu\text{bn}$  (Berezinsky, 2006).

### Pion production

Pion production can happen when a nucleon interacts with a photon. The pion production follows from the reactions

$$\begin{aligned}
p + \gamma &\rightarrow (\Delta^+) \rightarrow p + \pi^0, \\
p + \gamma &\rightarrow (\Delta^+) \rightarrow n + \pi^+.
\end{aligned}
\tag{1.27}$$

If the particle is a neutron, the reactions become:

$$\begin{aligned}
n + \gamma &\rightarrow (\Delta^0) \rightarrow n + \pi^0, \\
n + \gamma &\rightarrow (\Delta^0) \rightarrow p + \pi^-.
\end{aligned}
\tag{1.28}$$

Following eq. 1.26, one can find that the minimum energy for a photon in the nucleus rest frame to produce pions is given by:  $E_{\gamma\pi}^{\min} = m_{\pi}c^2(1 + \frac{m_{\pi}}{2m_p}) \simeq 145$  MeV. The product  $f(E_{\gamma})\sigma(E_{\gamma}) \sim 50 \mu\text{bn}$  is about two orders of magnitude above the product  $f(E_{\gamma})\sigma(E_{\gamma})$  found in the electron/positron pair production (Berezinsky, 2006). It means that above the energy threshold  $E_{\gamma\pi}^{\min}$ , for a given path, the nucleus loses  $\sim 100$  times more energy through pion production than electron/positron pair.

### Photo-erosion (photo-disintegration)

If the UHECR is a nucleus, all kinds of nuclear reactions can happen (eq. 1.29):

$${}^A_Z\text{X} + \gamma \rightarrow {}^{A-\alpha}_{Z-\beta}\text{Y} + (\alpha - \beta)\text{n} + \beta\text{p},
\tag{1.29}$$

where  $\alpha$  and  $\beta$  are positive integers, with  $\alpha - \beta \geq 0$ .

Those reactions can happen with very different phenomena, depending on the energy of the photon in the nucleus rest frame (Allard et al., 2006).

- $10 \text{ MeV} \lesssim E_{\gamma} \lesssim 20 \text{ MeV}$ : The disintegration process is dominated by the Giant Dipole Resonance (GDR). At these energies, the photon excites a collective nuclear mode. The protons and the neutrons can be described with one oscillating wave function, with the proton wave function in anti-phase compared to neutrons. It leads to the ejection of one or multiple nucleons (n, p, 2n, 2p, np,  $\alpha\dots$ ).

- $E_\gamma \approx 30 \text{ MeV}$ : The disintegration process is dominated by the quasi-deuteron (QD) process. At these intermediate energies, the photon wavelength is too small to excite collective motion, but too big to see the nucleus as individual nucleons. Hence, the photon will interact with a pair of correlated nucleons, resulting in the ejection of the pair as well as possibly more nucleons.
- $E_\gamma \gtrsim 150 \text{ MeV}$ : The disintegration process is dominated by the baryonic resonances (BR). At such high energies, the photon sees the nuclei as a bunch of individual nucleons. Hence, the reaction is the one explained in the previous section (cf. Pion production).

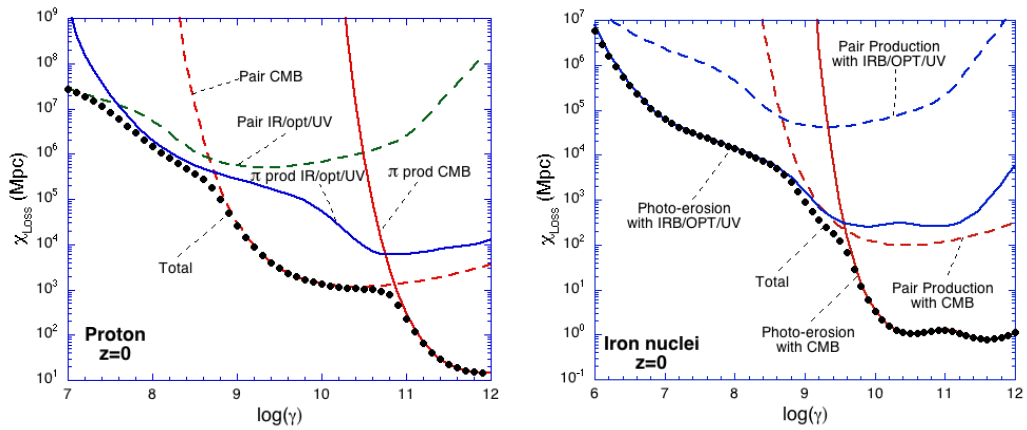
### 1.3.3 Attenuation length

The attenuation length is a characteristic length of interaction for UHECRs, which is defined in eq. 1.30 (Aloisio et al., 2013):

$$\chi_{\text{loss}}^{-1} = \frac{1}{E} \frac{dE}{dx}, \quad (1.30)$$

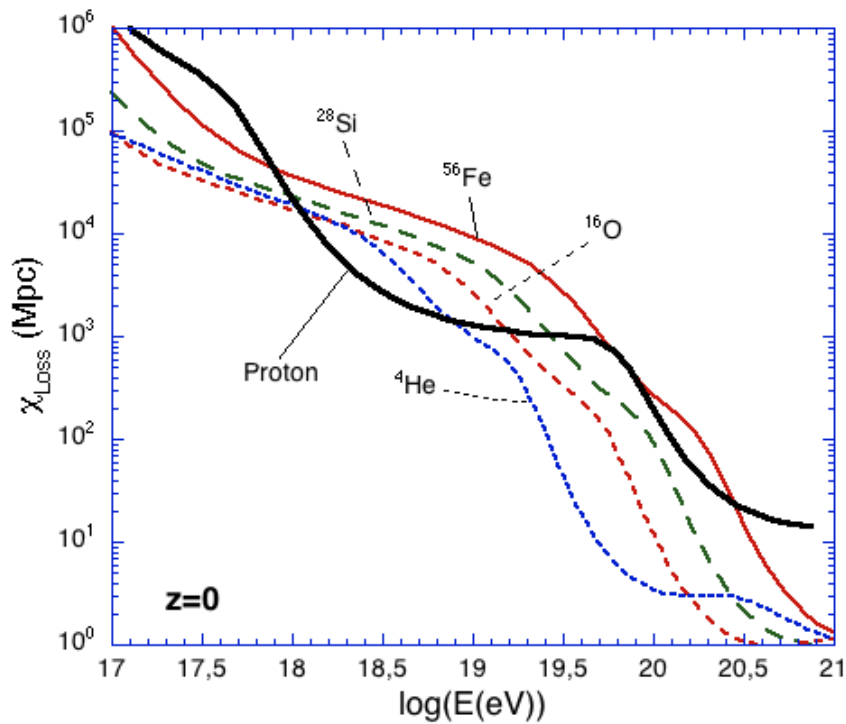
where  $x$  is the travelled distance. Figure 1.9 shows the attenuation length for a proton (left) and for an iron nuclei (right) as a function of the Lorentz factor  $\gamma$ . The losses are decomposed in electron/positron pair production on CMB/EBL photons and pion production ( $\pi$  prod) for protons, and in terms of photo-erosion process (GDR+QD+BR) and pair production for iron nuclei. It is important to note that those graphs have been done at a redshift  $z = 0$ , as if the nucleus evolves in a hypothetical world where the universe is static and stays as of now. When simulations are performed, not only the nucleus endures the adiabatic losses as explained before, but the CMB and EBL evolution as well. Indeed, both energy and density of photons fields are function of redshift and so do the resulting interactions. On the left-hand side of Fig. 1.9, one can see that the attenuation length drops around  $\log(\gamma) = 11$ , where  $\gamma$  is the UHECR's Lorentz factor. This drop corresponds to the well-known GZK cut-off (Greisen, 1966; Zatsepin & Kuz'min, 1966). The GZK cut-off is the analytical computation of the interaction of UHECR protons with CMB photons. Greisen, Zatsepin, and Kuzmin showed that the attenuation length is such that no proton from sources beyond  $\sim 10 \text{ Mpc}$  should be observed on Earth above this Lorentz factor.

Figure 1.10 shows the attenuation length of different nuclei as a function of energy. This figure sums-up the importance of studying the composition of UHECRs. Since



**Fig. 1.9.:** Attenuation length of protons (left) and iron nuclei (right) as a function of the Lorentz factor at  $z = 0$ . The contribution of pair production and pion production from the CMB and EBL (IR/Opt/UV) are separated for the protons. For iron, the production of pairs is separated from photo-erosion cases. Extracted from Allard et al., 2006.

different nuclei have different attenuation lengths, the composition of the observed UHECRs gives information about the maximum distance of sources.

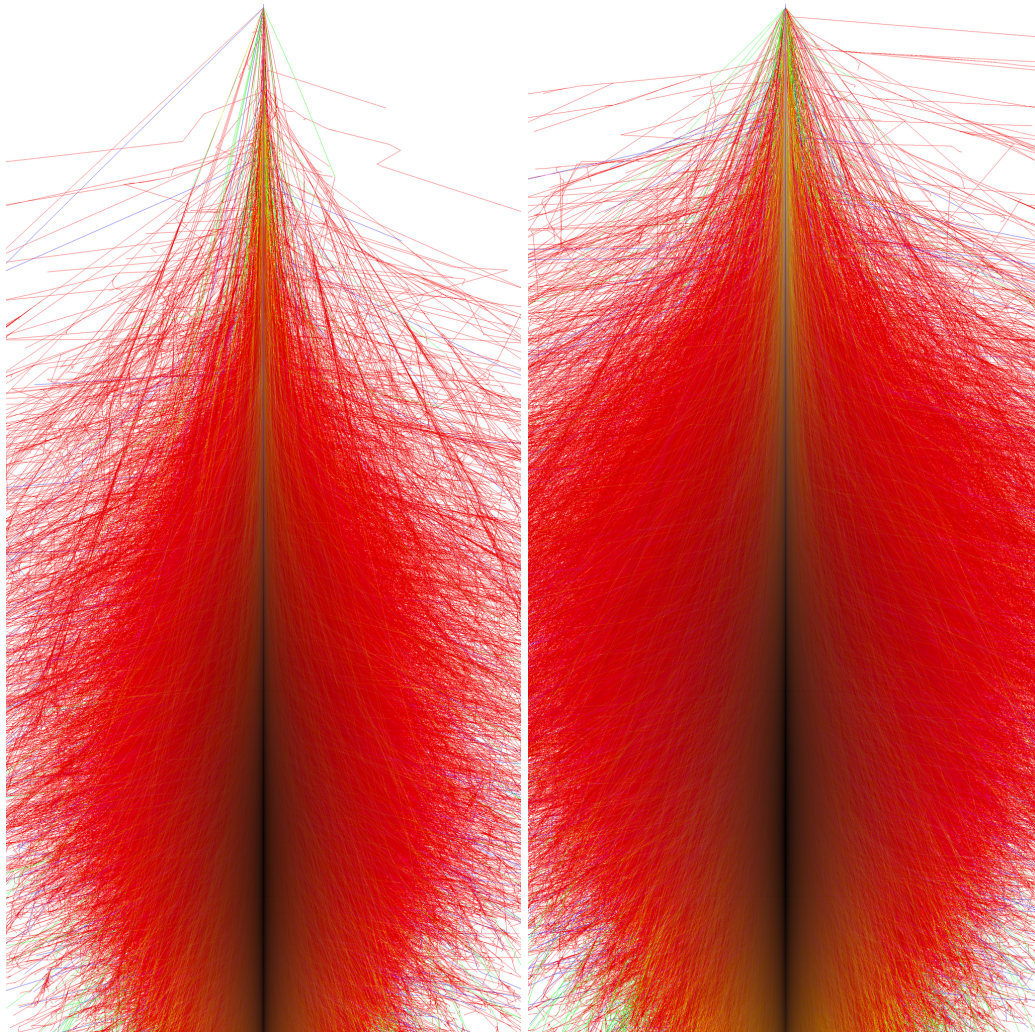


**Fig. 1.10.:** Attenuation length as a function of energy for different nuclei, at a redshift  $z = 0$ . Extracted from Allard et al., 2006.



## 1.4 Extensive air showers

After propagating in the universe, UHECRs arrive on Earth. When a highly relativistic particle enters the atmosphere, it collides with a nucleus. This collision results in a number of new high-energy particles. The newly-created particles also interact, creating again particles. This process continues until the particles do not have enough energy to create new particles. This cascade is called Extensive Air Shower (EAS). Two examples of simulated EAS are shown in Fig. 1.11.



**Fig. 1.11.:** CORSIKA simulation of extensive air showers of a proton (left) and iron (right) at  $10^{15}$  eV. The colours used are red for electromagnetic particles, green for muons and blue for hadronic particles. Extracted from <https://www.iap.kit.edu/corsika/>, Fabian Schmidt, University of Leeds, UK.

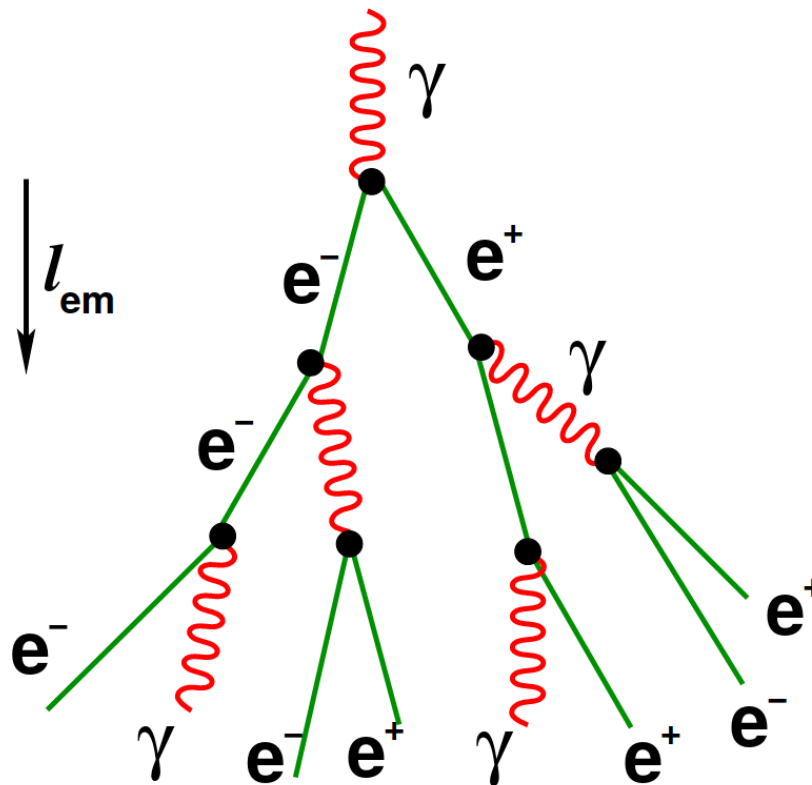
### 1.4.1 Modelling

#### Heitler model

The simplest model of EAS is the Heitler model, which describes electromagnetic showers initiated by a photon or an electron/positron. Two processes are possible:

- An electron/positron creates a photon through Bremsstrahlung effect.
- A photon's energy is converted into an electron-positron pair.

In the first case, the electron/positron experiences the electric field of an electron or a nucleus and gives part of its energy to a photon. In the second case, the photon needs to be in the vicinity of a nucleus for conserving the four-momentum of the reaction ; the nucleus receives some recoil. Each of these two particles will also split their energy in two, resulting in a cascade of particles. A schematic view of an electromagnetic cascade is shown in Fig. 1.12.

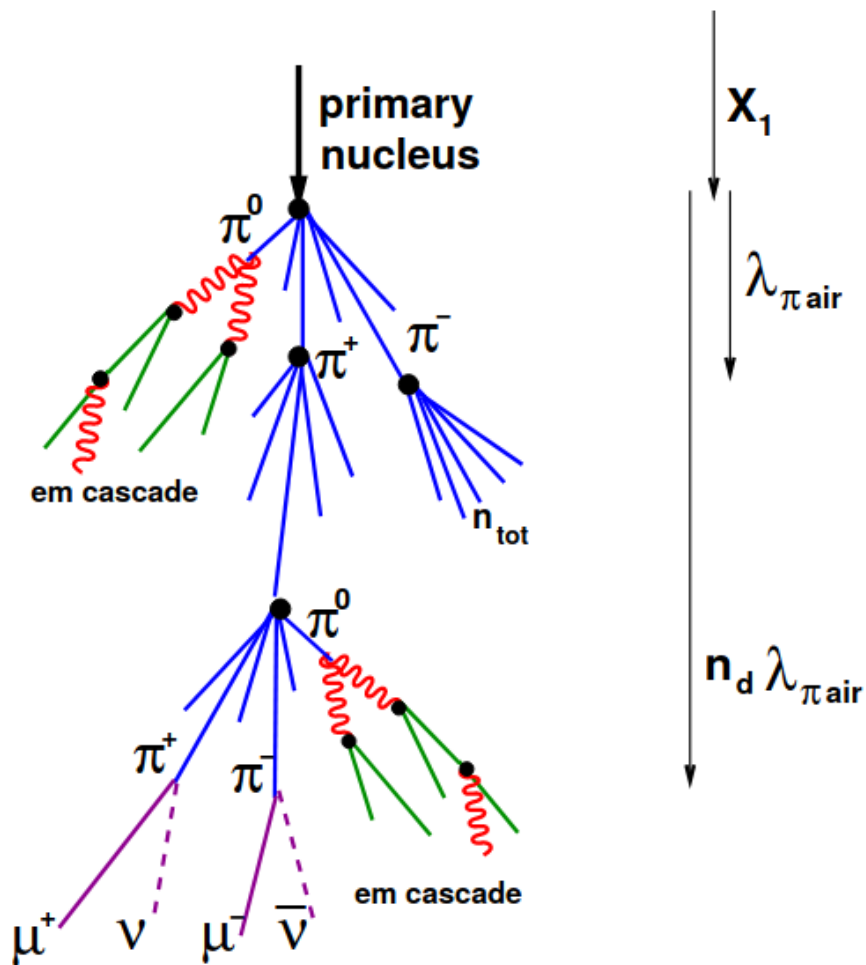


**Fig. 1.12.:** Schematic view of an electromagnetic cascade. The red particles are photons, while the green lines correspond to electron/positron. Extracted from Mollerach and Roulet, 2018.



The characteristic splitting length is given by  $l_{\text{em}} = X_0 \ln 2$ , where  $X_0$  is the radiation length (in air  $X_0 = 36.7 \text{ g/cm}^2$ ). After  $n$  splittings, the distance is given by  $x = n X_0 \ln 2$ , the number of particle is given by  $N = 2^n$  and the energy per particle is given by  $E = \frac{E_0}{N}$  where  $E_0$  is the energy of the initial particle. The cascade process stops when the secondary particles do not have enough energy to produce an electron/positron pair (for photons) or to produce a photon using Bremsstrahlung effect (for  $e^\pm$ ). This critical energy  $E_c$  is reached when the radiative energy loss becomes less than collisional energy loss (in air  $E_c = 85 \text{ MeV}$ ).

### Heitler model with hadronic interactions (Heitler-Matthews model)



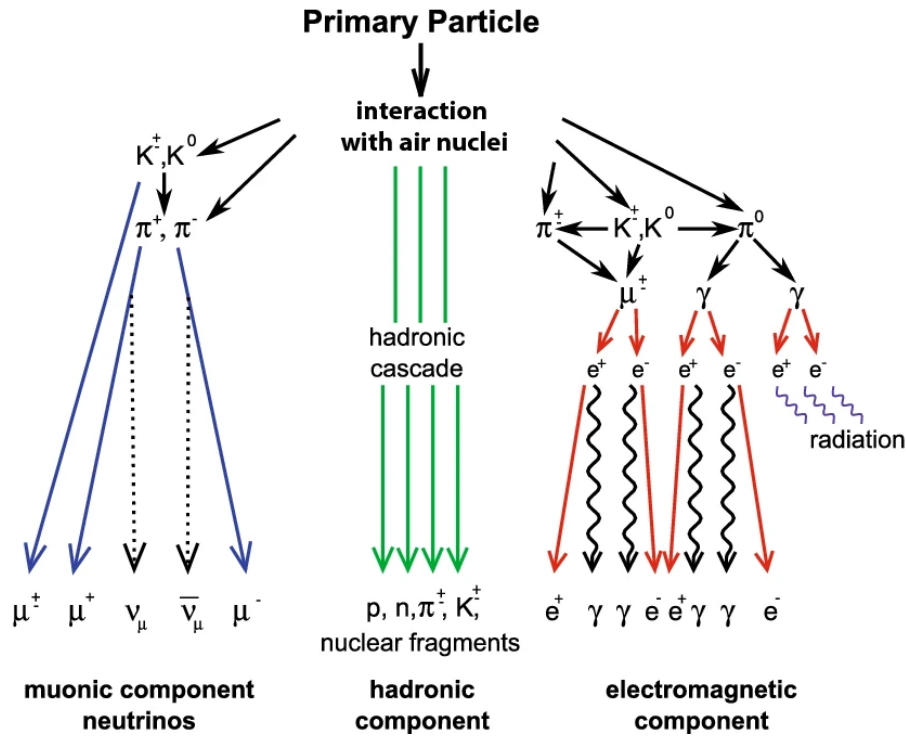
**Fig. 1.13.:** Schematic view of a hadronic cascade. Blue lines corresponds to hadronic particles, red corresponds to photons, green corresponds to electron/positron, and purple correspond to muons and neutrinos. Extracted from Mollerach and Roulet, 2018.

Hadronic air showers are modelled using a method similar to Heitler's.

The atmosphere is decomposed into layers of constant thickness  $\lambda_{\pi \text{ air}} = \lambda_I \ln 2$ , where  $\lambda_I$  is the interaction length of strongly interacting particles.  $\lambda_I$  can be considered constant over a large energy range (10–1000 GeV), and is given by  $\lambda_I = 120 \text{ g/cm}^2$  for pions (Matthews, 2005). After passing through one layer, hadrons produce  $N_{\text{ch}}$  charged pions ( $\pi^+$ ,  $\pi^-$ ) and  $\frac{N_{\text{ch}}}{2}$  of neutral pions ( $\pi^0$ ). Charged pions continue to interact when they pass through another layer. The operation is repeated until the energy of the  $\pi^\pm$  falls below a critical energy level  $E_{\pi_c}$ . They are, then, decaying into muons and associated neutrinos. The neutral pions instantly decay into two photons, initiating electromagnetic showers. It is important to note that the hadrons can also create kaons, which are strange pions; the process remains the same.

### **UHECR cascade**

An extensive air shower always starts with hadronic interactions, creating all kinds of hadrons (nuclear fragments, nucleons, pion and kaons). The neutral pions will decay into photons, creating electromagnetic showers, as explained before. The other particles nourish the hadronic shower until they reach their critical energy and decay into muons and neutrinos. It is important to note that, since muons are directly a tracer of the number of hadrons, the number of muons is linked to the mass of the incoming cosmic ray. Then, if the muons are energetic enough, they will reach the ground without interacting much. If the muons have low energies, they will decay into neutrinos and electrons, and initiate electromagnetic showers. At the end, the EAS can be described using three components: Hadronic components, electromagnetic component and a muonic/neutrinos component. The processes of EAS for an UHECR is summed-up in Fig. 1.14 .



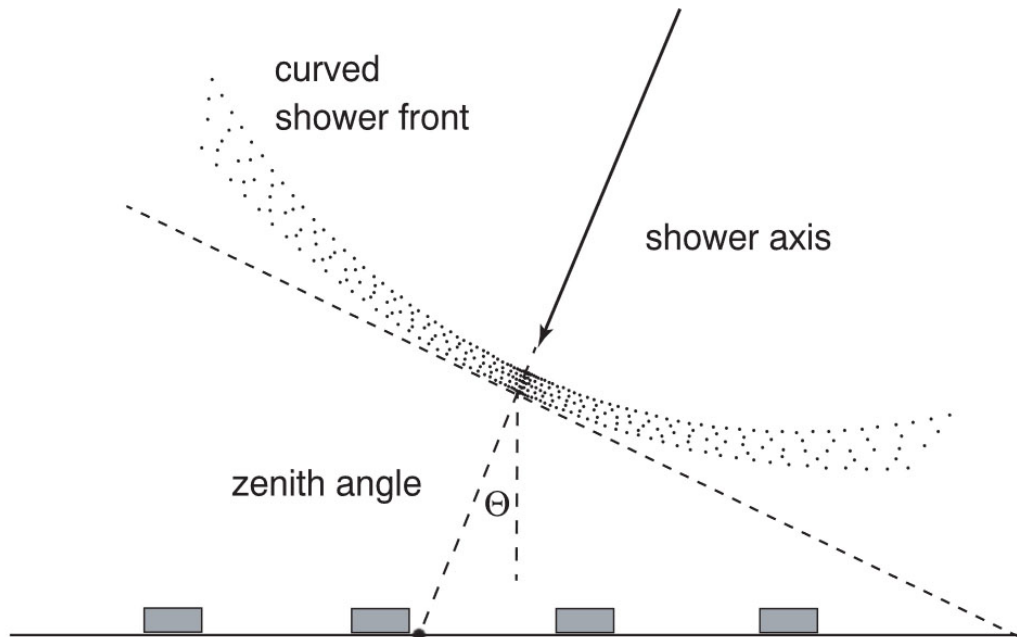
**Fig. 1.14.:** Schematic view of the composition of an extensive air shower. The shower can be explained using three main components. Green: hadronic component. Red: muonic component. Blue: electromagnetic component. Extracted from Haungs et al., 2018.

## 1.4.2 Characteristics

Since UHECRs are quite rare (about 1 per  $\text{km}^2$  per yr over the entire atmosphere around the ankle [cf. Fig. 1.4]), the only way for detecting them is to observe the EAS they initiate. This is an indirect measurement of cosmic rays. For an observer, the EAS is not easily split in three components. The observer sees a curved shower of all kinds of particles. It is represented in Fig. 1.15 and described with:

- The shower axis: it defines the arrival direction of UHECR;
- The core of the shower: it is the point where the shower axis meets the ground;
- The zenith angle: it corresponds to the angle between the vertical and the shower axis;
- The lateral distribution function: it is the projection of the number of particles onto the axis perpendicular to the shower axis;

- The longitudinal profile: it is the projection of the number of particles onto the shower axis.



**Fig. 1.15.:** Schematic view of an extensive air shower. Adapted from Fleck et al., 2021.

The lateral distribution function and the longitudinal profile are related to the energy of the incoming cosmic ray. To get an information on the mass of the incoming cosmic ray, one can look at the nature of the particles at ground level (number of muons). The other solution to get information on the mass is to look at the longitudinal profile.

The next chapter will focus on the detection of UHECRs. It will show how the lateral distribution and the longitudinal profile are reconstructed at the Pierre Auger Observatory, and how one can derive the energy, arrival direction and the composition of UHECRs.



# The Pierre Auger Observatory and observables

## 2.1 The Pierre Auger Observatory

Inaugurated in 2008, the Pierre Auger Observatory is the largest observatory to study UHECRs. It is located in Malargüe in the Mendoza Province, Argentina, near the Andes. Since UHECRs are rarely observed on Earth (cf. 1.1.2), the Pierre Auger Observatory covers a surface of  $3,000 \text{ km}^2$  (approximately corresponding to 30 times the surface of Paris).

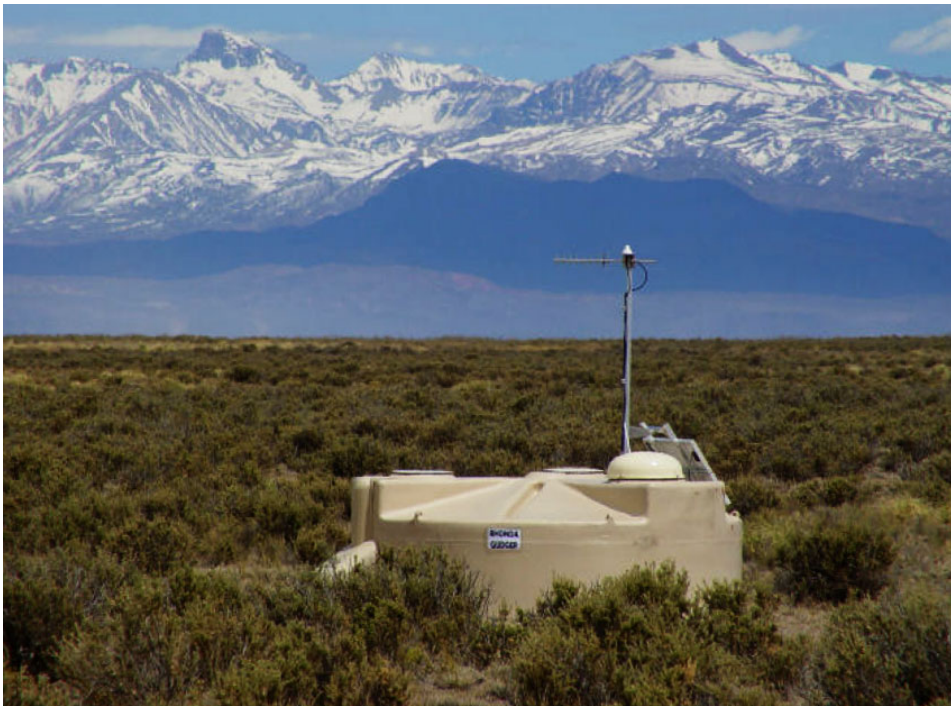
### 2.1.1 General view

The Pierre Auger Observatory is composed of two types of detectors: Fluorescence telescopes, also called fluorescence detectors (FD), and water tank detectors, also called water Cherenkov detectors (WCD) or simply surface detectors (SD). Images of a FD and a WCD are shown in Fig. 2.1 and Fig. 2.2, respectively. We shall see below that the two detectors are complementary in providing measurements of the longitudinal shower profile (with the FD) and the lateral shower profile (with the SD). The WCDs operate day and night, regardless of weather conditions, while the FDs, being telescopes, cannot operate during daytime, rainy weather, or full-moon nights. Therefore, the duty cycle of the FD, on the order of 15%, is relatively low compared to that of the SD (almost 100%). While the SD operates without requiring human control, the FD must be controlled. During observation nights (typically 3 weeks during the new Moon), shifts are organized to control the FD. These shifts can be done at the Observatory or remotely. During my first year of PhD, I had the chance to do a half-shift on site.

Figure 2.3 shows the map of the Pierre Auger Observatory. Each point represents a WCD while the black lines show the field of view of the FD. There are 1,600 WCDs, each of them spaced 1,500 m apart, while there are 4 FD sites, each composed of 6 fluorescence telescopes. As the cascade passes through the atmosphere, nitrogen and oxygen molecules get excited by the many ionizing electrons created along

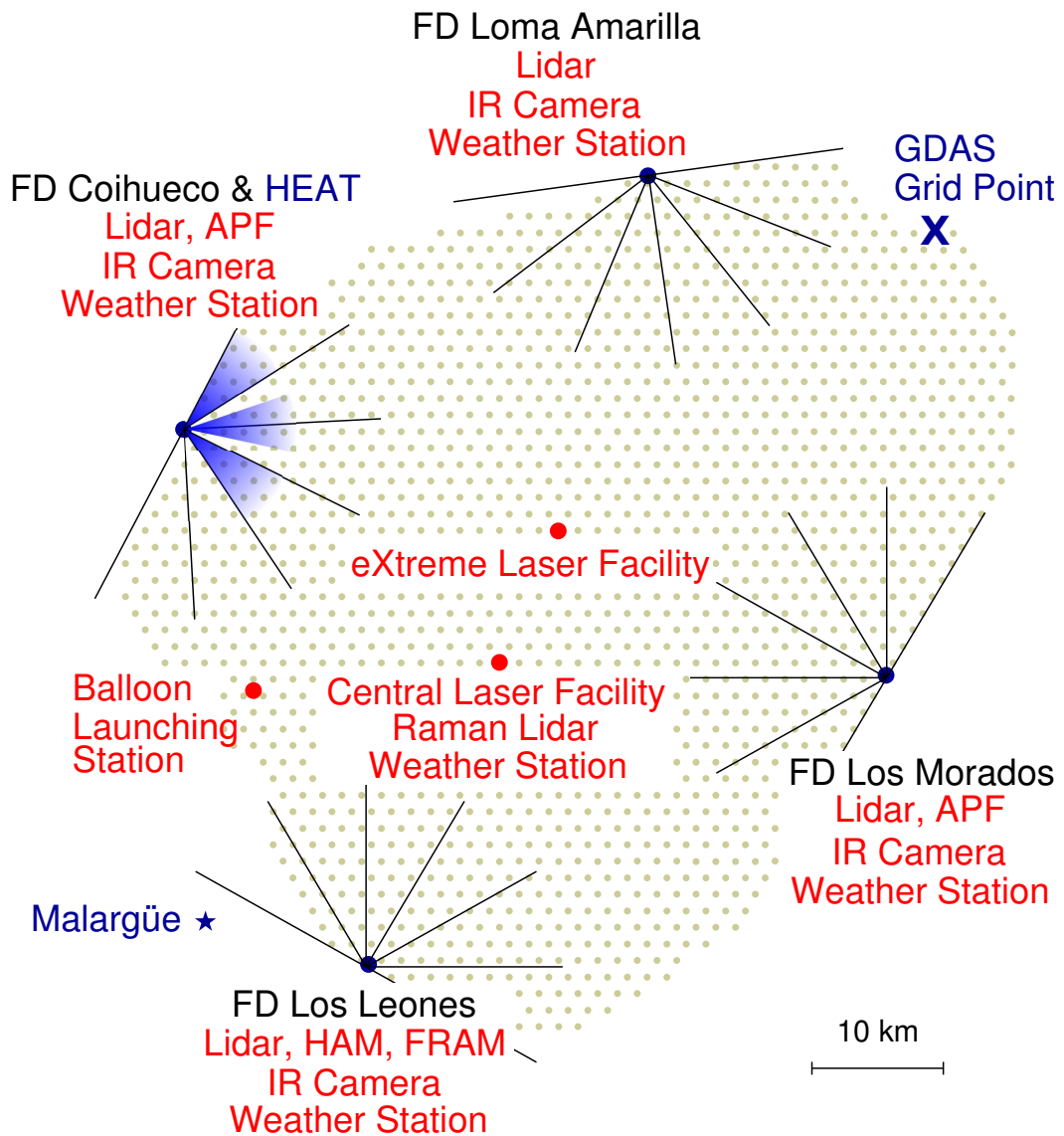


**Fig. 2.1.:** The Loma Amarilla site of fluorescence detector. Credit: Pierre Auger Collaboration.



**Fig. 2.2.:** One of the WCD of the surface detector of the Pierre Auger Observatory, with the Andes in the background. Credit: Pierre Auger Collaboration.





**Fig. 2.3.:** Map of Pierre Auger Observatory. Each dot represents an SD tank. The four FD sites are shown in black (Loma Amarilla, Los Morados, Los Leones, Colhueco). The black lines correspond to the viewing angle of each FD telescope. The low-energy FDs are shown in blue (HEAT). Equipment used for atmosphere monitoring is shown in red. The infilled array is not represented. Extracted from The Pierre Auger Collaboration, 2015b.

the EAS track. The ultraviolet fluorescence caused by the subsequent de-excitation of the molecules is detected by the FDs, made up of arrays of several hundreds of photomultiplier tubes that, thanks to a set of mirrors, each monitor a small portion of the sky. The isotropic emission enables observing the cascades side-on up to 30 or 40 km away on moonless nights and thus reconstructing the longitudinal profile of the EAS. In turn, measurement of the longitudinal profile allows inferring



both the energy of the EAS in a calorimetric way, without recourse to external information to calibrate the energy estimator, and the slant depth of maximum of shower development ( $X_{\max}$ ), which is a proxy, the best up to date, of the primary mass of the particles. By contrast, WCDs deployed on the ground, providing a harvest of data thanks to their quasi-permanent duty cycle, require assumptions about the primary mass and the hadronic processes that control the cascade development to infer the energies.

To derive an SD energy estimation that is free from assumptions about primary mass and hadronic physics, a hybrid approach is preferable. One can see that the field of view of the FD overlooks the SD array. This allows for measuring some events twice, once with the SD and once with the FD. Such events, called the golden-ones, are used to calibrate the SD energy estimator in terms of energy with the FD. This approach provides a calorimetric estimate of the energy for events recorded during periods when the FDs cannot be operated.

On top of the main SD array and FDs, there are enhancements and many facilities. Those enhancements are described in the next section. In terms of facilities, the main one is called the Central Data Acquisition System (CDAS), which is located in the city of Malargüe. Data acquisition is subtle. Each FD and WCD have their own triggers. They are called “local triggers”. Then, if a signal passes the first triggers, it is sent to CDAS through radio antennas. The CDAS then looks for coincidences with other WCDs and possibly triggers to record the event.

## Atmosphere monitoring

Since the Pierre Auger Observatory uses the atmosphere as a calorimeter, it is important to have precise measurements of it. The complete profile of the state of the atmosphere (temperature, humidity, pressure, wind) used to be measured with a balloon launched at the “Balloon Launching Station” (cf. Fig. 2.3). Nowadays, the state of the atmosphere is measured on the ground, every 5 minutes, using meteorological stations located on FD sites. The profile is computed from an atmospheric model called GDAS (Global Data Assimilation System). Aerosols, which are the dust in the atmosphere, are measured with different instruments. The eXtreme Laser Facility (XLF) and the Central Laser Facility (CLF) shoot lasers that are recorded by the FD and analyzed to reconstruct the vertical optical depth due to dust. Two other facilities are used to measure aerosol properties, the Aerosol Phase Function Monitor (AFM) and the Horizontal Attenuation Monitor (HAM), which emit light, also observed by the FD and processed.



**Fig. 2.4.:** Picture of a LIDAR during a FD shift.

Clouds are also monitored. The first way to monitor clouds is to use cloud cameras that are located at FD sites. These are IR cameras that have the same field of view as fluorescence telescopes. Knowing the position of the stars, the cloud cameras can compare the expected light from the stars and the measured light to find the position of the clouds. The second way to observe clouds is to use LIDAR. For the Pierre Auger Observatory, LIDARs are lasers mounted on a rotating telescope that scans the sky during a FD shift and records the re-emitted light. I was also able to visit a FD site with the associated LIDAR. A picture of a LIDAR in operation during a FD shift is shown in Fig. 2.4. As for dust, CLF and XLF are also used to detect clouds. Finally, FRAM (ph(F)otometric Robotic Atmospheric Monitor), an optical telescope that measures starlight, is also used to infer cloud information. A summary table of atmospheric monitoring is presented in Tab. 2.1.

**Tab. 2.1.:** Atmospheric measurements performed and the instruments that are used. Table extracted from The Pierre Auger Collaboration, 2015b.

Category	Variable	Frequency	Instrument(s)
State	At ground: Pressure, Temp., Wind, Humidity	5 min	Weather Stations
	Profile: Pressure, Temp., Humidity	3 hours	GDAS <sup>a</sup>
Aerosols	Vert. Optical Depth ( $z$ )	hourly	CLF, XLF + FD
	Phase Function	hourly	2 APF units
	Ångström Coefficient	hourly	FRAM (HAM)
Clouds	Presence in FD pixels	15 min	4 Cloud Cameras
	Behind FD sites	15 min	4 LIDAR stations
	Along select tracks	avg. 1/night	FRAM, LIDAR
	Above CLF/XLF	hourly	CLF, XLF + FD

<sup>a</sup>atmospheric model developed at the National Centers for Environmental Prediction, operated by NOAA; provided via READY - Real time Environmental Applications and Display sYstem.

## 2.1.2 Other detectors

### Infill and HEAT

High Elevation Auger Telescopes (HEAT) are FD located at Coihueco site. HEAT is an inclined FD with an elevation angle of  $\sim 30^\circ$ . Such an inclination allows for observing EAS that develop in high altitude but that cannot be observed with the conventional FD. This type of EAS corresponds to “low-energy” UHECRs ( $\lesssim \sim 10^{18}$  eV). Underneath the field of view of HEAT, there is the “Infill”, an array of infilled WCDs that are spaced 750 m apart (compared to the 1,500 m of the main array) over  $\sim 30$  km<sup>2</sup>. As with HEAT, the Infill was made to measure smaller EAS induced by “low-energy” UHECRs. At these targeted energies, the flux is high enough to have sufficient statistics with  $\sim 30$  km<sup>2</sup>.

### AugerPrime: Underground Muon Detector

Aimed at measuring the mass composition of UHECRs at the highest energies to discriminate between presently-viable scenarios explaining the current data, an upgrade of the Observatory, called AugerPrime or Pierre Auger Observatory Phase II, is under development at the time this thesis is written (early 2022). The first device implemented is the Underground Muon Detector (UMD). UMDs are scintillator detectors buried under the field. Burying the detectors allows them to be sensitive to the muon component of the EAS, using the ground as a shield against the

electromagnetic component. As shown in the previous chapter, the muon component is a function of the mass of the incoming UHECRs, so the UMD is sensitive to the composition of cosmic rays. The UMD has been deployed under the Infill array to detect low-energy UHECRs, between  $2 \times 10^{17}$  eV and  $2 \times 10^{18}$  eV.

The muon densities measured by UMD and published in The Pierre Auger Collaboration, 2020b were higher than those predicted by simulations using the latest hadronic model; this is the so-called “muon puzzle”. Investigations are still underway to understand this muon excess. The excess could come from an overproduction of charged pions that decay into muons. Such a measurement is crucial to understand EAS and constrain the hadronic models used to describe them.

### **AugerPrime: SSD, small PMT and radio upgrade**

In the same spirit of understanding the composition of UHECRs, plastic scintillators (SSDs) are installed above each WCD. Since the WCDs are 3D detectors, they primarily record the muon signal as it passes through the tank (the electromagnetic component is quickly absorbed and stops emitting light). On the other hand, SSDs are 2D detectors that also measure the electromagnetic and muon components. Thus, by comparing the WCD signal with the SSD signal, one can differentiate between muons and electrons/positrons. Therefore, as with UMD in the case of low-energy EAS, the combined SSD and WCD data will give information on the number of muons in the EAS, and thus on the mass of incoming UHECRs.

Since the SSDs are placed at the top of the WCDs, they are only sensitive to vertical showers. However, EAS are known to produce radio emissions in the 10-100 MHz range because they are made of charged particles undergoing acceleration in the atmosphere. As with SSDs, the combination of radio detection with WCD data allows for the measurement of the muon/ $e^\pm$  ratio. In order to better understand the composition of very inclined showers, radio antennas are installed on each WCD.

The final feature of the AugerPrime upgrade is a small PMT added to each WCD. The small PMT is less sensitive and will not be saturated if the shower core is close to an SD tank. Therefore, the small PMT improves the uncertainties of saturated signals.

## 2.2 Fluorescence detectors

### 2.2.1 Fluorescence telescope: An extensive air shower camera

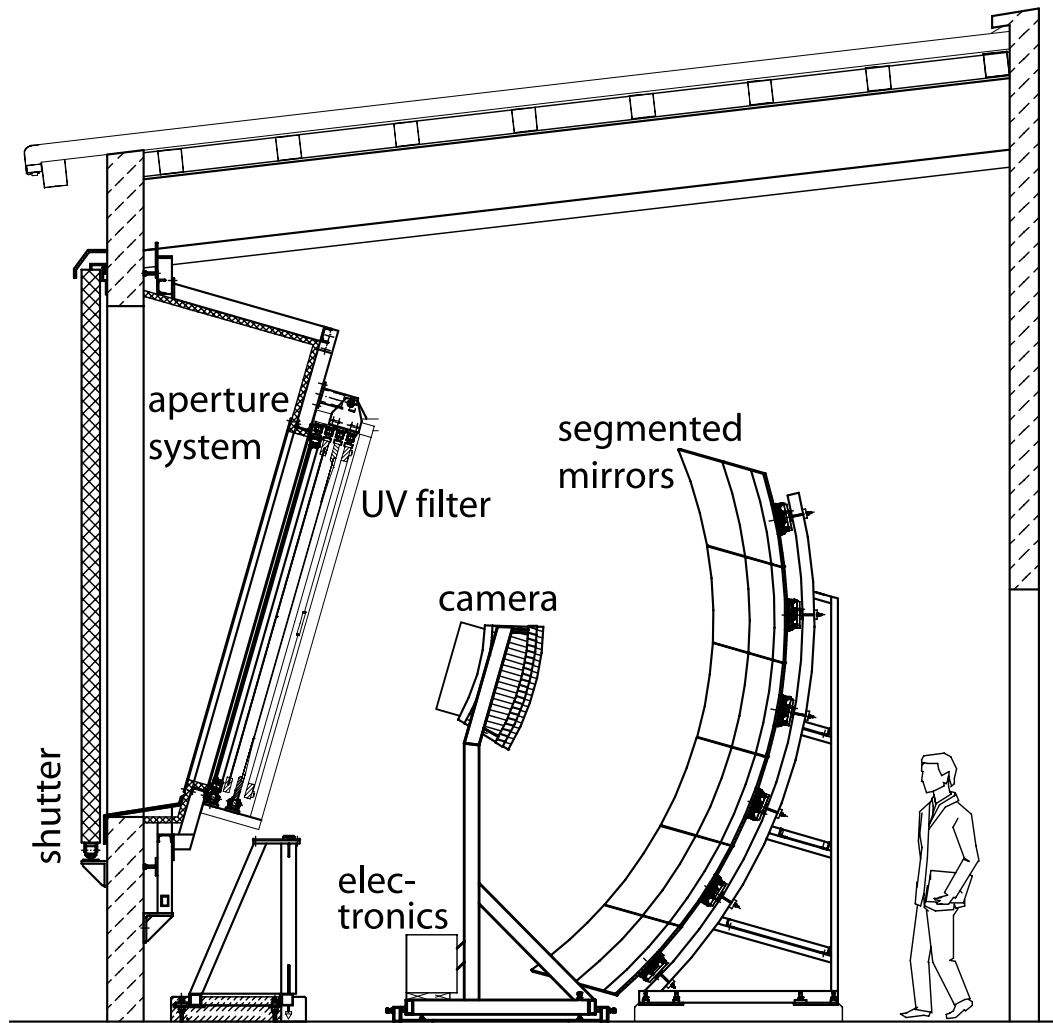
The FDs are telescopes that take pictures of EAS. Unlike other telescopes, the FD do not point directly to the sky: they point almost horizontally in the troposphere to see the light emitted by the EAS. Two kinds of light are emitted:

- Fluorescence light: Emitted by the excitation of atmospheric molecules by the charged particles of the EAS (same phenomenon as aurora borealis);
- Cherenkov light: Emitted by charged particles with a velocity higher than the phase velocity of light in the medium (here atmosphere).

The light can be detected directly and indirectly (by Rayleigh scattering on air or Mie scattering on aerosols). At high energies, the main contribution comes from fluorescence light while at low energies, it is dominated by Cherenkov light. For more details, see The Pierre Auger Collaboration, 2014a. FDs are fully effective above  $10^{19}$  eV (The Pierre Auger Collaboration, 2010).

A schematic view of a FD is shown in Fig. 2.5. The shutters are closed during the day or during rainy weather. The aperture system remains open during FD operation, but closes automatically if there is too much light to protect the cameras. Light passes through the filter, reflects off the segmented mirrors and hits the camera. The camera is composed of  $22 \times 20 = 440$  photomultiplier tubes (PMTs), each of them being a pixel. Each pixel covers a field of view of  $1.5^\circ$ . Each PMT signal is processed by the electronics.

The cameras must be calibrated to convert a number of photons (ADC) into energy. The main calibration is performed with a 2.5 m diameter light source, called a “drum”, placed at the telescope aperture. The drum provides uniform light to all pixels, which is used to calibrate the PMTs. In addition to the calibration, two relative calibrations are performed before and at the end of each observation night. There are different ways to perform the relative calibration, but it mainly consists in illuminating each PMT with a Xenon light source or a laser, directly or indirectly by illuminating the mirrors.



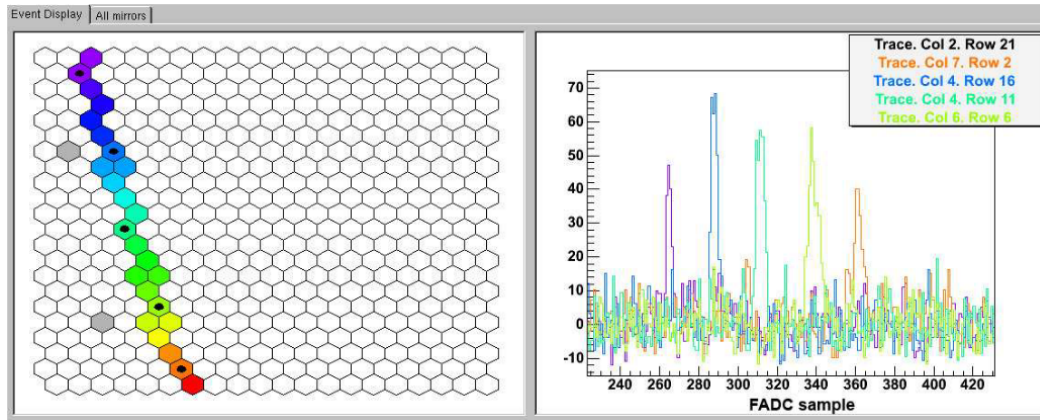
**Fig. 2.5.:** A schematic view of a FD compared to human size. Extracted from The Pierre Auger Collaboration, 2015b.

## 2.2.2 Arrival direction reconstruction

An example of event observed using the FD is shown in Fig. 2.6. Each hexagon corresponds to one of the 440 PMTs in the camera. The colored hexagons correspond to the PMTs that have been selected with a significant signal by the Data Acquisition (DAQ). The right-hand side of Fig. 2.6 shows the measured PMT signal for the corresponding dashed hexagon in FADC units (time equivalent). The detection time for each PMT gives the angle between the shower core and the zenith in the shower-detector plane. The direction of the trace in the camera (left part of Fig. 2.6) gives the angle between the shower core and the zenith in the plane perpendicular to the shower-detector plane. Therefore, by combining the two angles, one can determine the arrival direction of the incoming UHECR. Since the SD has a 100%



duty cycle, most of the energetic-enough events detected by the FD are also detected by the SD. The arrival direction obtained using both the SD and the FD is then reconstructed with a resolution of  $\sim 0.6^\circ$  (The Pierre Auger Collaboration, 2010).

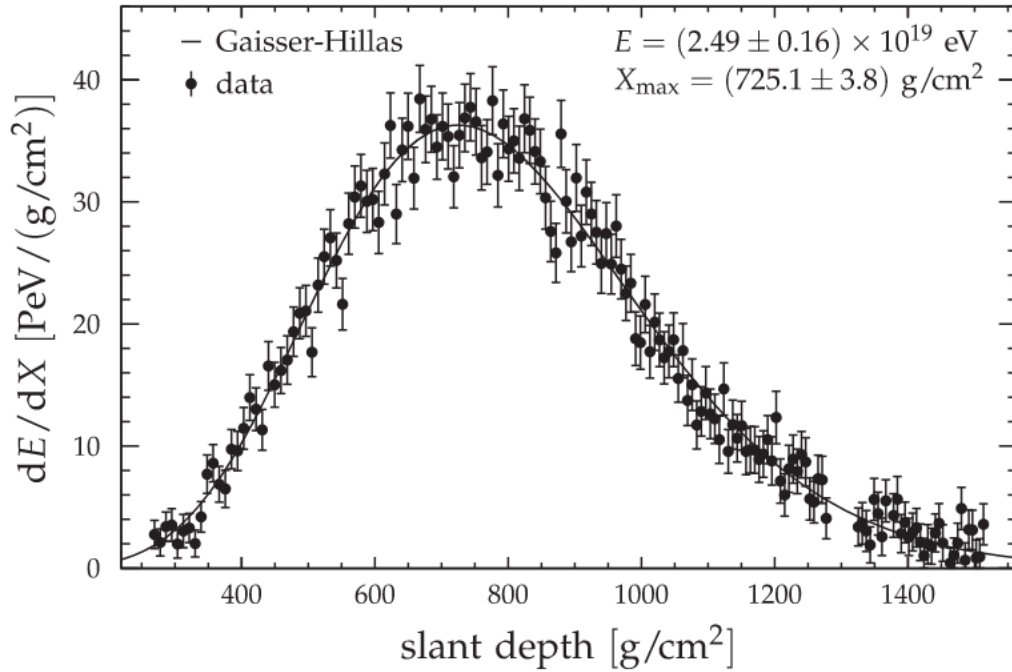


**Fig. 2.6.:** A FD event as observed using a telescope. Left: Each hexagon correspond to one PMT of the camera. The coloured hexagon corresponds to the PMT with a signal above a given threshold. The colours represent the time of detection of each PMT (purple is the earlier and red is the later). Right: The PMT signals of the dotted PMT are shown with the corresponding times. Extracted from The Pierre Auger Collaboration, 2010.

### 2.2.3 Energy and maximum depth of shower reconstruction

Knowing the geometry of the shower, the light collected at the aperture as a function of time can be translated into energy deposit as a function of slant depth, as shown in Fig. 2.7. The energy deposit function is fitted using the Gaisser-Hillas parametrization (Gaisser & Hillas, 1977). The calorimetric energy of a shower is estimated by integrating the Gaisser-Hillas function. Then, the estimated energy is corrected for the “invisible” energy of high-energy neutrinos and muons using. The systematic uncertainty on the energy for the FD is  $\sim 14\%$  (The Pierre Auger Collaboration, 2013).

The maximum depth of the shower  $X_{\max}$ , corresponds to the slant depth point where the Gaisser-Hillas function is maximized. It coincides with the depth where the number of particles is maximum.  $X_{\max}$  is directly related to the mass; however  $X_{\max}$  fluctuates between showers of same energy and primary mass. Therefore, it is not possible to infer the mass of a single incoming cosmic ray with sufficient accuracy. However, a given primary mass leads to a given  $X_{\max}$  distribution that differs from that of another primary mass, so one can get an estimate of the UHECR mass composition using distributions of  $X_{\max}$  values expected for a given scenario of



**Fig. 2.7.:** Energy deposit profile  $\frac{dE}{dx}$  as a function of the slant depth. The dots represent the measurements from a FD telescope while the solid line is a fit. Extracted from The Pierre Auger Collaboration, 2014b.

mass composition. Such a study is presented in chapter 4. The systematic uncertainty on  $X_{\max}$  is on the order of 10 g/cm<sup>2</sup> and depends slightly on the energy (The Pierre Auger Collaboration, 2014a).

## 2.3 Surface Detectors

### 2.3.1 A water Cherenkov detector

The SD observes charged particles of EAS that reach the ground. Each WCD has reflective walls and is filled with 12,000 liters of ultrapure water. Above the water, three 9" PMTs and associated electronics are powered by batteries that are recharged by solar panels. A complete schematic view of the detector is shown in Fig. 2.8.

When muons and electrons/positrons pass through the tank at a speed greater than the speed of light in water, they emit Cherenkov light. This light is reflected inside the tank and reaches the PMTs.



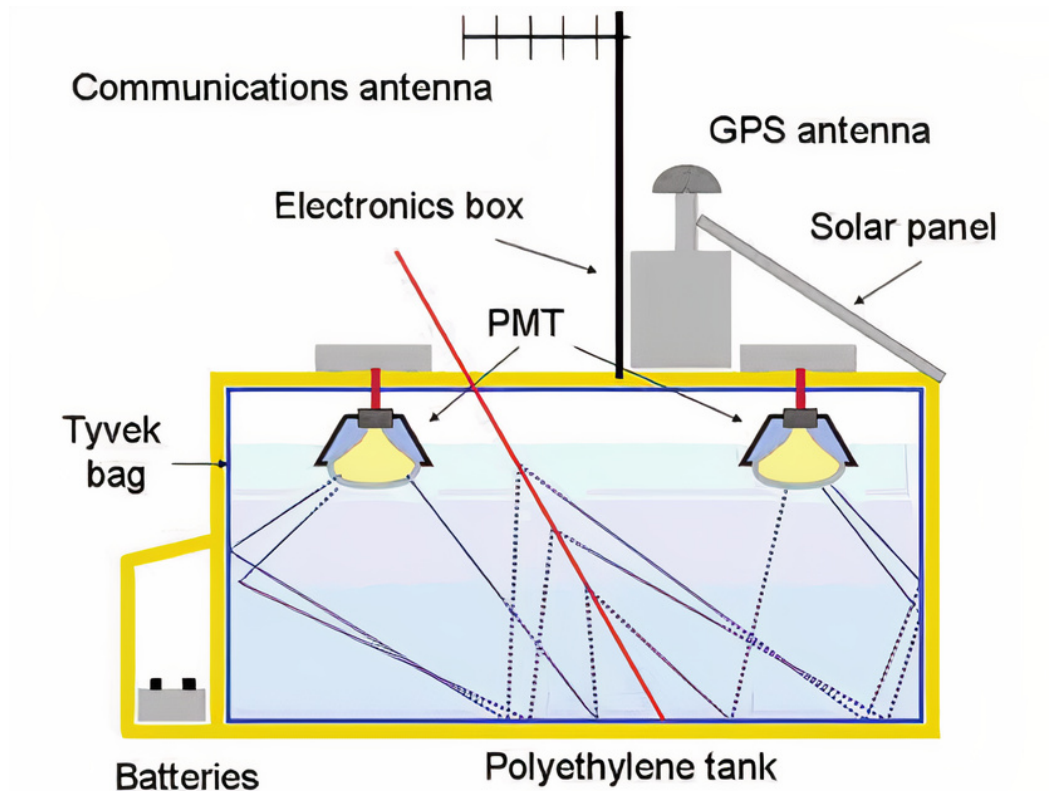


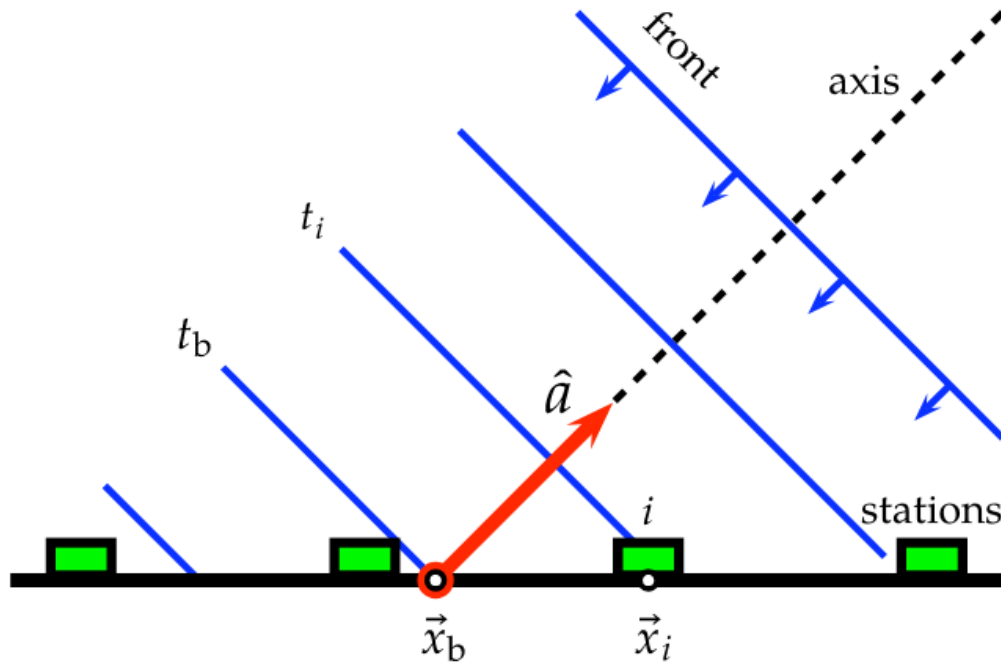
Fig. 2.8.: Schematic view of a SD tank. Extracted from Shellard, 2006.

The PMTs convert Cherenkov photons into electrical signals, which are then digitized into ADC counts. The ADC count is converted into a reference unit, the vertical equivalent muon (VEM). A VEM corresponds to the energy deposited by a one GeV muon passing vertically through the tank. The signal is then processed by the electronics and if it reaches the triggers, it is sent by radio antennas to CDAS. CDAS looks for a coincidence with other nearby SD tanks, and if so, the event is recorded. The SD array is fully efficient above  $\sim 3$  EeV.

### 2.3.2 Arrival direction reconstruction

Fig. 2.9 shows a schematic view of an EAS reaching the SDs (green square), considering the development of a plane-front air shower. The front of the EAS is perpendicular to the shower axis, described by the vector  $\hat{\mathbf{a}}$ . The core of the shower is represented by the point  $\vec{x}_b$ . Assuming that the front moves at the speed of light  $c$ . We can deduce that the shower reaches the  $\vec{x}$  position after a time  $t(\vec{x})$  according to eq. 2.1:

$$ct(\vec{x}) = ct_b - \hat{a}c(\vec{x} - \vec{x}_b) \quad (2.1)$$



**Fig. 2.9.:** Schematic view of an extensive air shower (blue). The SD detectors are shown in green. Extracted from The Pierre Auger Collaboration, 2020e.

By measuring the time  $t_i$  when a station  $i$  is triggered and knowing its position  $\vec{x}_i$ , we can deduce the core of the shower  $\vec{x}_b$ , and the arrival direction  $\hat{a}$ . Above 3 EeV, the angular resolution is less than  $1.4^\circ$  for vertical showers and decreases monotonically with zenith angle. For very inclined or very high-energy showers, the angular resolution can reach  $0.5^\circ$ . Such a behavior is expected because a horizontal EAS or a very high energy one will trigger many more stations (The Pierre Auger Collaboration, 2020e).

### 2.3.3 Energy measurement

The Lateral Distribution Function (LDF) of extensive air showers,  $f_{\text{LDF}}(r)$ , is a fundamental quantity for determining EAS sizes, which in turn play the role of a track-length integral for evaluating empirically the primary energy. Knowledge of the LDF is based on data-driven determinations of key parameters that govern functional shapes describing the falloff of the particle density with the distance  $r$ . Such functional shapes can be inspired from theoretical results on electromagnetic

cascades supplemented by experimental observations of the muon content of showers (commonly dubbed as NKG functions), or they can just be a fitting representation of data. For historical reasons, two software have been developed to reconstruct the EASs at the Pierre Auger Observatory. The first is named *Herald*, while the second one is named *Observer*.

*Herald* uses a log-log parabola function for LDF:

$$\ln f_{\text{LDF}}(r) = \beta\rho + \gamma\rho^2, \quad (2.2)$$

where  $\rho = \ln(r/r_{\text{opt}})$  and  $r_{\text{opt}}$  is a normalization constant.

*Observer* uses a modified NKG function:

$$f_{\text{LDF}}(r) = \left(\frac{r}{r_{\text{opt}}}\right)^\beta \left(\frac{r + r_s}{r_{\text{opt}} + r_s}\right)^{(\beta+\gamma)}, \quad (2.3)$$

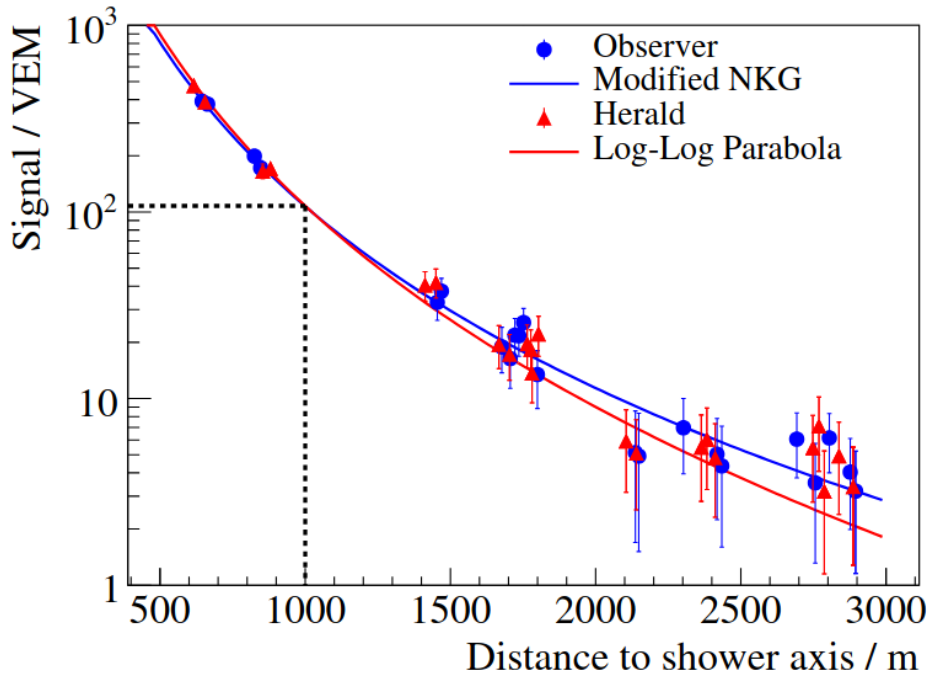
where  $r_s$  is a fixed at  $r_s = 700$  m.

The parameters  $\beta$  and  $\gamma$  govern the logarithmic slope that control the falloff with the distance to the core:  $\beta$ , which is the main logarithmic slope, and  $\gamma$ , which is the long-range correction to the logarithmic slope. Fitting both parameters on a set of EAS with same zenith angle and shower size results in reconstructed parameters significantly larger than the expected fluctuations, evidencing genuine fluctuations of LDF between showers of same size and inclination. This reflects that showers are sampled at the ground level at a different time after first interaction, due to shower-to-shower fluctuations. Therefore, mean parameters are extracted from subsets of events with an important number of trigger station.

Hence, the subsequent LDF originates from these fluctuations and must be regarded as a varying quantity on an event-by-event basis. However, the average dependence of the slope parameters are the only ones usually accessible from data of the SD array. This is because the spacing of the detectors is too large to allow for providing a sufficient lever arm for measuring the  $\beta$  values of the bulk of events triggering the SD. Only subsets of specific events can be used to infer average properties, and therefore to draw average LDFs. Subsequently, the signal deposit of an event is adjusted to the average LDF by scaling its normalization (or its size) in a fitting procedure, using a normalized LDF function:

$$S(r) = S(r_{\text{opt}}) f_{\text{LDF}}(r). \quad (2.4)$$

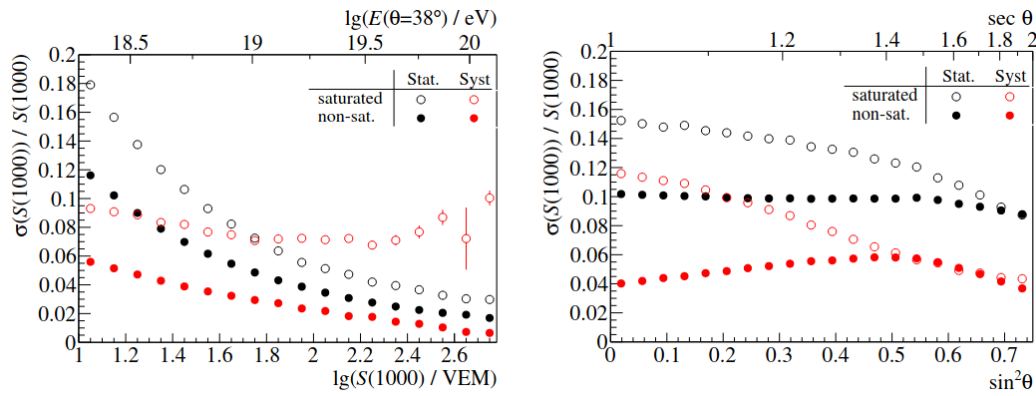
The size obtained in this way, namely  $S(r_{\text{opt}})$ , plays the role of a track-length integral. Such a track-length integral would be hazardous to estimate by integrating over some distance range the particle-density content of the LDF because of the unknown fluctuations from the average functional shape. The value of the “optimal distance,  $r_{\text{opt}}$ ”, which translates to the distance range that yields the most accurate interpolated signal, is determined by the presence of measurement points around it, and thus by the geometry of the SD array. In the case of the Auger SD array,  $r_{\text{opt}} = 1,000$  m (Newton et al., 2007).



**Fig. 2.10.:** Lateral distribution function of an SD event using two different analysis pipelines. The red dots are computed using Herald software, with an associated fit using a log-parabola function. The blue dots are computed using Observer software, with the associated fit using a NKG modified function. Extracted from The Pierre Auger Collaboration, 2020e.

The uncertainty of  $S(1000)$  is shown in Fig. 2.11. The uncertainty decreases with the energy of the incoming UHECR. This behavior is expected: the higher the incoming cosmic ray energy, the bigger the EAS. Therefore, more stations are triggered and the  $S(r)$  fit is better constrained.

The graph on the right-hand side of Fig. 2.11 shows that the uncertainties of the unsaturated events are approximately constant in terms of zenith angle up to  $60^\circ$ , and that the total uncertainty is better than to 15%.



**Fig. 2.11.:** Relative uncertainty of  $S(1000)$  as a function of  $\lg(S(1000)/\text{VEM}/\text{energy})$  (left) or as a function of zenith angle  $\sin^2 \theta$  (right) for saturated and non-saturated events. Extracted from The Pierre Auger Collaboration, 2020e.

For very inclined EASs, the electromagnetic component has been completely absorbed by the atmosphere and the only remaining particles are muons. Those muons traveled a long distance and, due to the Earth's magnetic field, created an asymmetric LDF. Therefore, another analysis is performed for the EASs with zenith angle larger than  $60^\circ$  to get an energy estimate. To do this, the analysis uses the muon distribution measured at ground level and corrects for the missing electromagnetic component. A review of the energy estimate for very inclined events is available in The Pierre Auger Collaboration, 2014c.

## 2.4 In the Northern Hemisphere: Telescope Array

In the field of UEHCR detection, a second experiment is often cited: Telescope Array. Telescope array is a complementary experiment to the Pierre Auger Observatory because it is located in Utah, United States, which is in the Northern Hemisphere, where the Auger experiment cannot see. Telescope Array (TA) was inaugurated in 2008. The principle of operation is the same as that of Auger. There is a FD that measures the cosmic ray energies, the arrival directions and  $X_{\max}$ , and a SD that measures the energy and the arrival directions. The main differences between the two observatories concern the SD. TA uses scintillators, which are fully efficient only for vertical showers with a zenith angle  $\theta \leq 45^\circ$  above 10 EeV (Telescope Array Collaboration, 2012) while Cherenkov water tanks are used at the Auger Observatory. The latter enable the measurement of very inclined showers up to  $\theta = 80^\circ$ . Therefore, the field of view of TA is reduced compared to that of the Auger

Observatory. Moreover, TA covers an area of  $760 \text{ km}^2$ , which is about four times smaller than that of the Auger Observatory.



## Arrival directions

*This chapter is partly adapted from the author's work described in the Auger open-data tutorial, the "The UHECR sky" analysis (Marafico, Biteau, & Deligny, 2020)*

### 3.1 Reconstruction of flux map

Reconstructing flux maps is one of the key tool to study UHECRs. Excess or deficit of flux in a region of the sky could be correlated to the distribution of sources of UHECRs or propagation effect in astrophysical structures (magnetic field of the cosmic web, of our galaxy). In this section, two methods to reconstruct flux maps are tested and compared, and their applications are shown.

In order to reconstruct the flux, one should compute the exposure of the Observatory. The exposure is the time-integrated effective collecting area of the Observatory times its field of view, in units of  $\text{km}^2 \text{ yr sr}$ . Knowing the total exposure  $E_{\text{tot}}$  and the number of detected events  $N$  in a given energy band, one can compute the flux in the energy band using  $\frac{N}{E_{\text{tot}}}$ . In the regime of full efficiency, and for uniform exposure to each direction of the sky, the total exposure can be computed using:

$$E_{\text{tot}} = S \times \Delta\Omega \times T, \quad (3.1)$$

where  $S$  is the surface of the detector in  $\text{km}^2$ ,  $\Delta\Omega$  is the field of view of the detector in  $\text{sr}$ , and  $T$  is the cumulated time of the detector running in  $\text{yr}$ . However, the exposure is not homogenous on the celestial sphere. As shown in Sommers, 2001, the relative exposure in a given direction of the sky can be computed geometrically. Since the Observatory is operating full-time, the exposure is constant in right ascension, and only depends on declination,  $\delta$ . Assuming the detector is fully efficient below a zenith angle  $\theta_m$ , and is located at a latitude  $\alpha_1$ , one can compute the directional exposure as:

$$\omega(\delta) = A (\cos(\alpha_1) \cos(\delta) \sin(\alpha_m) + \alpha_m \sin(\alpha_1) \sin(\delta)), \quad (3.2)$$

where  $A$  is normalization constant and  $\alpha_m$  is given by:

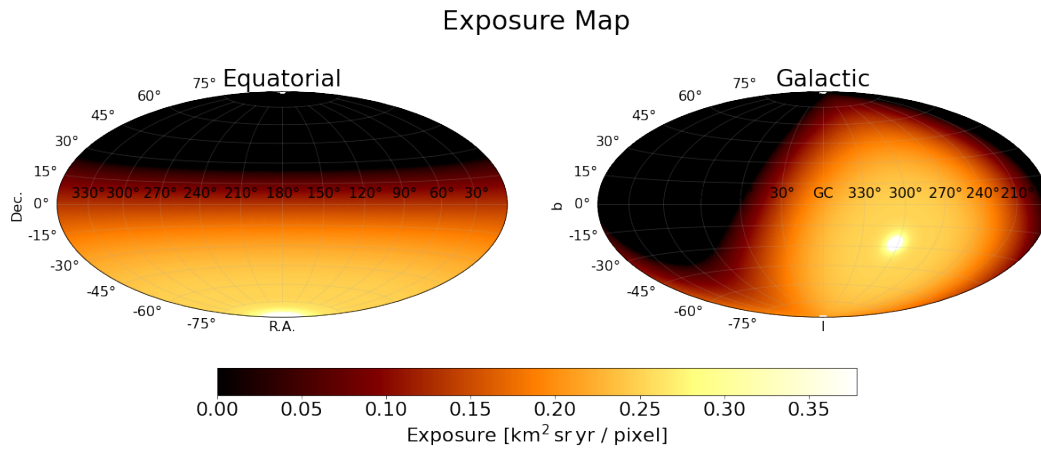


$$\alpha_m = \begin{cases} 0 & \text{if } \xi > 1, \\ \pi & \text{if } \xi < -1, \\ \cos^{-1}(\xi) & \text{otherwise,} \end{cases} \quad (3.3)$$

with

$$\xi = \frac{\cos(\theta_m) - \sin(\alpha_1) \sin(\delta)}{\cos(\alpha_1) \cos(\delta)}. \quad (3.4)$$

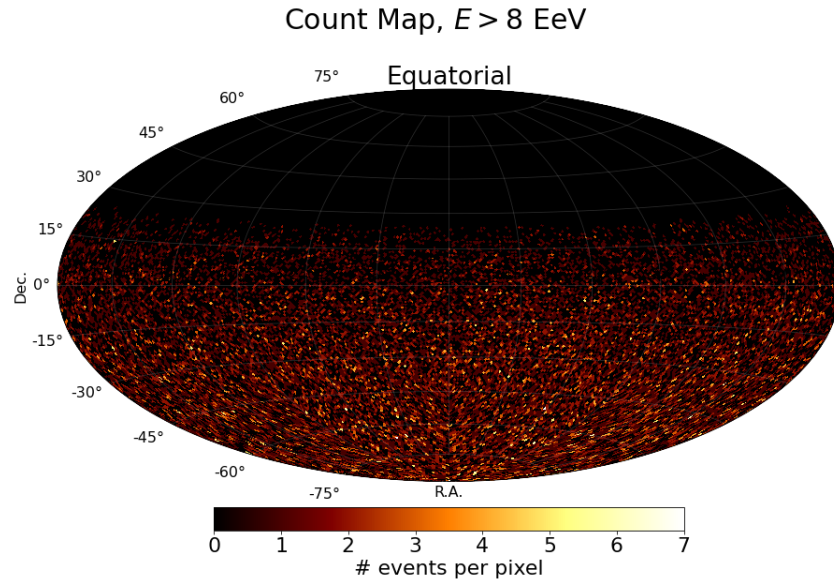
Knowing the exposure as a function of declination, one can represent it on a sky map using Healpy (Zonca et al., 2019). Healpy allows us to divide the celestial sphere into equal surface pixels. Then, the exposure is computed using Eq. 3.2 in each pixel and is normalized in order to get  $\sum_i \omega_i = E_{\text{tot}}$  where  $i$  is the index of each pixel. The total exposure is taken to be  $E_{\text{tot}} = 59,927.8 \text{ km}^2 \text{ yr sr}$ , which corresponds to the period from the first data taking (2004) up to December 2020 for vertical events. Two sky map representations of the exposure are shown in Fig. 3.1 considering vertical events only ( $\theta_m = 60^\circ$ ). The left hand-side corresponds to Equatorial coordinates. It is the “natural” view of the detector. Since the Pierre Auger Observatory is located in the Southern Hemisphere, most of its exposure covers negative declinations. On the right hand-side, the exposure is shown in Galactic coordinates, where the center of the map corresponds to the Galactic center and the Galactic longitude  $b = 0$  corresponds to the Galactic plane.



**Fig. 3.1.:** Exposure map in Equatorial (left) and Galactic coordinates (right).

The next step to compute the directional flux is to create a count map. To do so, vertical events above an energy  $E$  (here  $E = 8 \text{ EeV}$ ) are selected from the same period as the one taken for computing the exposure (from 2004 to December 2020).

An iteration is performed over all events. When an event is located inside a pixel  $i$ , the pixel value is increased by one (in practice, the implementation used here is strictly equivalent to that of a 2D histogram with uneven binning). At the end, a count map is created and it is shown in Fig. 3.2.



**Fig. 3.2.:** Count map showing the number of detected events in each pixel.

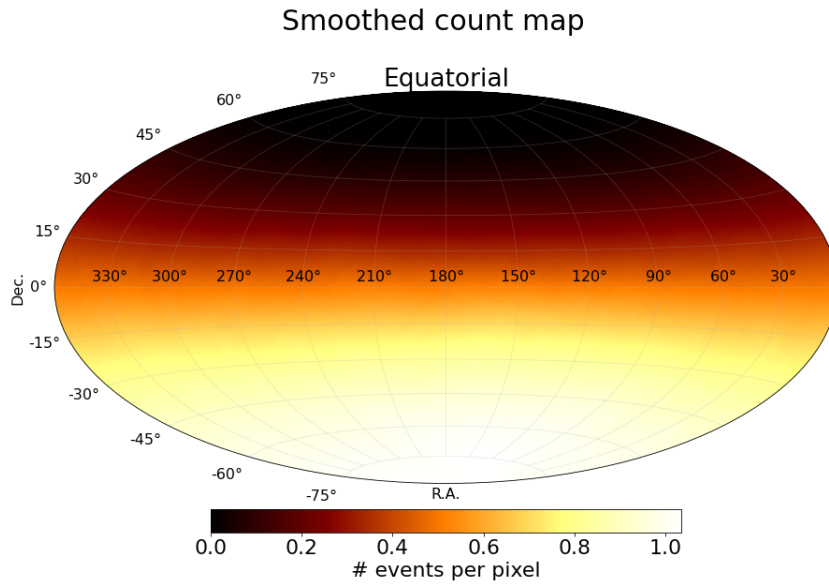
Fig. 3.1 and Fig. 3.2 give us the required building blocks to reconstruct a flux map above an energy  $E$ . However, the two maps depend on the pixel size, while one could hope to have a flux map which does not depend on the binning of the celestial sky. In the two next sections, two approaches to reconstruct flux sky maps are tested and compared.

### 3.1.1 Method: $N/E_{\text{tot}}$

The first method, called  $N/E_{\text{tot}}$ , starts by considering a smoothing function  $S(d)$  where  $d$  is the angular distance between two pixels on the 2D-Sphere, with  $[S(d)] = \frac{1}{\text{sr}}$ . Two smoothing functions are commonly used. The first one is a top-hat function. The second one is a von Mises–Fisher distribution, which corresponds to a Gaussian smoothing on a 2D-sphere. In the  $N/E_{\text{tot}}$  method, first, the count map is smoothed:

$$N_{\text{smoothed}}(\alpha, \delta) = \int_{\text{Sphere}} N(\theta, \phi) S(\alpha, \delta, \theta, \phi) d\Omega. \quad (3.5)$$

Figure 3.3 shows the smoothed count map using a top-hat function of  $45^\circ$  as a smoothing function.



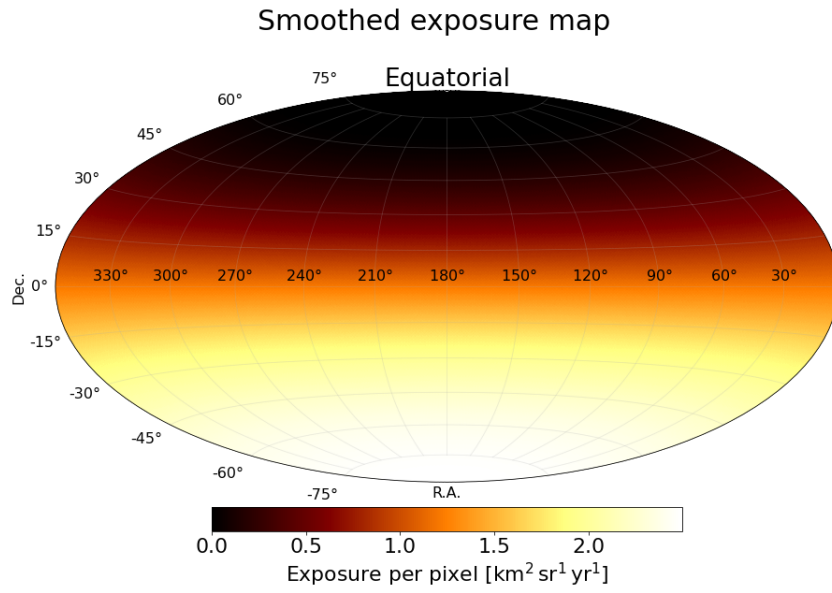
**Fig. 3.3.:** Smoothed count map  $N_{\text{smoothed}}(\alpha, \delta)$ .

Then, the exposure map is smoothed using the same smoothing function as before:

$$\omega_{\text{smoothed}}(\alpha, \delta) = \int_{\text{Sphere}} \omega(\theta, \phi) S(\alpha, \delta, \theta, \phi) d\Omega, \quad (3.6)$$

where  $\omega(\theta, \phi)$  is normalized to have  $\frac{1}{d\Omega} \int_{\text{Sphere}} \omega(\theta, \phi) d\Omega = E_{\text{tot}}$ . The smoothed exposure map is shown in Fig. 3.4.

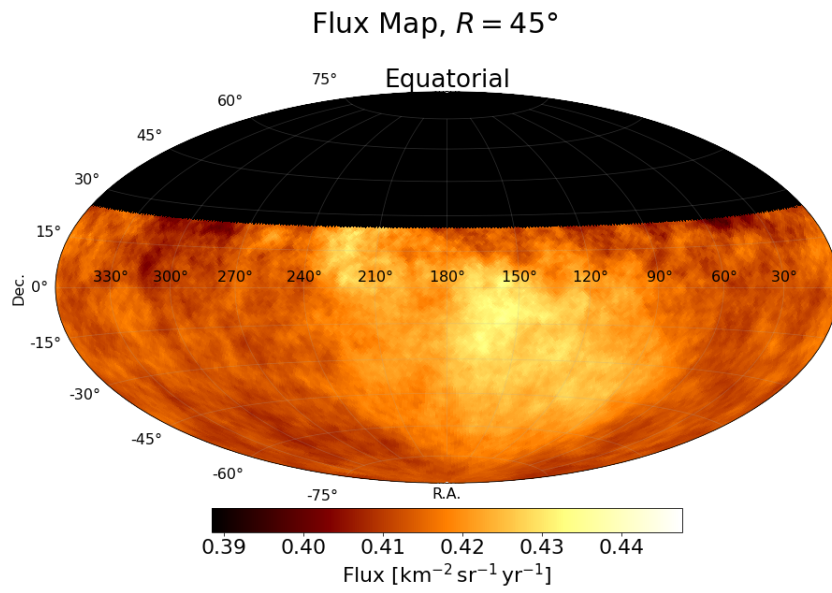
Finally, the smoothed flux map is computed by dividing the smoothed count map by the smoothed exposure map:



**Fig. 3.4.:** Smoothed exposure map,  $\omega_{\text{smoothed}}(\alpha, \delta)$ .

$$\Phi_{\text{smoothed}}(\alpha, \delta) = \frac{N_{\text{smoothed}}(\alpha, \delta)}{\omega_{\text{smoothed}}(\alpha, \delta)} \quad (3.7)$$

Such a method is binning independent. The resulting flux map is shown in Fig. 3.5.



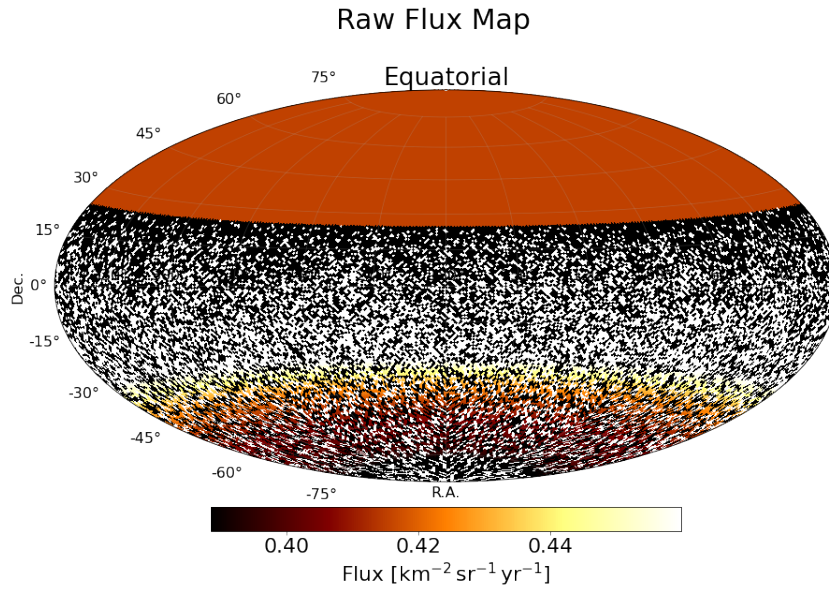
**Fig. 3.5.:** Flux map obtained using the  $N/E_{\text{tot}}$  method.

### 3.1.2 Method: $1/\omega$

The second method to reconstruct a flux sky map is the  $1/\omega$  method. It consists of computing a “raw” flux map, by dividing the count map by the exposure:

$$\Phi_{\text{raw}}(\alpha, \delta) = \frac{N(\alpha, \delta)}{\omega(\alpha, \delta)}. \quad (3.8)$$

Such a map is shown in Fig. 3.6. The area out of the field of view is filled with the average flux:  $\phi_{\text{raw}}^{\text{outside}} = \frac{\sum_i N_i}{E_{\text{tot}}}$ .



**Fig. 3.6.:** Raw flux map, where each pixel corresponds to the number of detected events in a single pixel divided by the exposure of the same pixel.

Then the smoothing is applied on the raw flux:

$$\Phi_{\text{smoothed}}(\alpha, \delta) = \int_{\text{Sphere}} \Phi_{\text{raw}}(\alpha, \delta) S(\alpha, \delta, \theta, \phi) d\Omega. \quad (3.9)$$

Here, the smoothing function is the same as before, a  $45^\circ$  radius top-hat function. The reconstructed flux map is shown in Fig. 3.7.

Comparing Fig. 3.5 and Fig. 3.7, one can see that the two results are broadly similar but differ near the border of the field of view. In the following section, the two methods are tested and compared using a benchmark scenario.

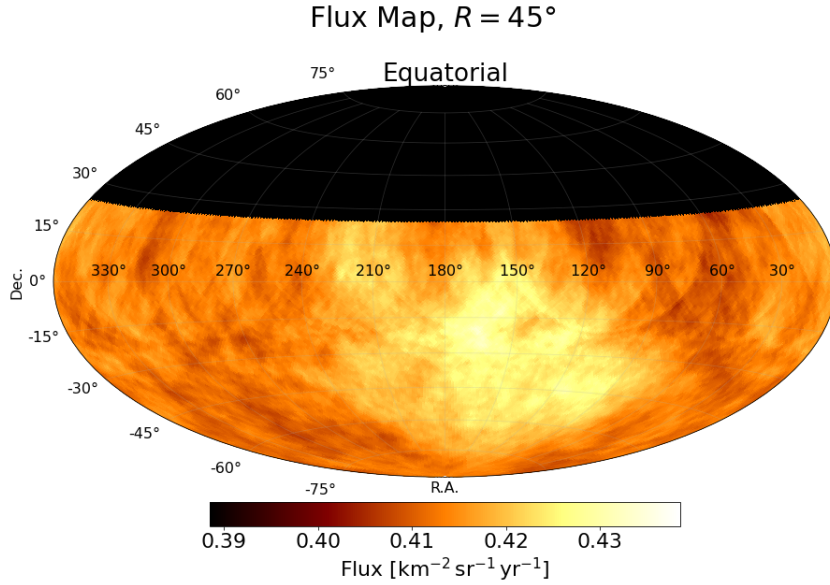
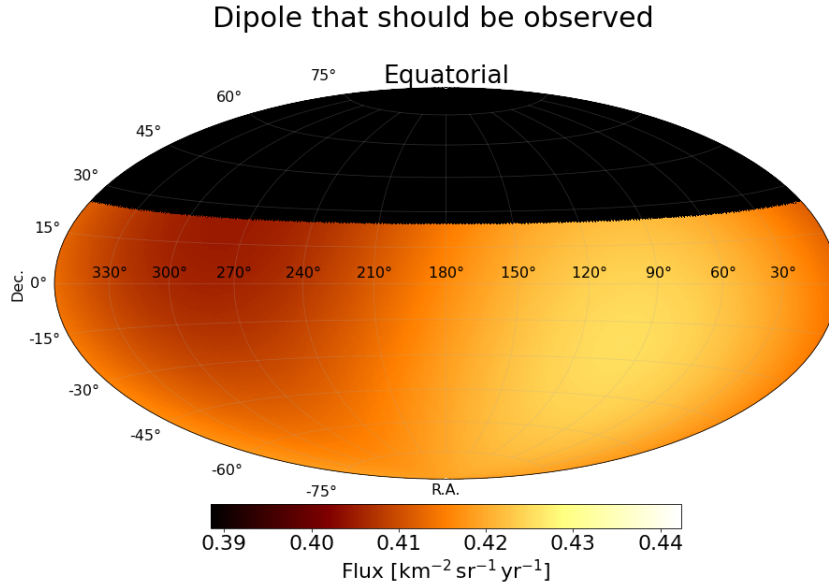


Fig. 3.7.: Flux map computed using  $1/\omega$  method.

### 3.1.3 Comparison of the two reconstruction methods

To test the two reconstruction methods, a benchmark scenario is constructed. The benchmark scenario is based on the dipole observed in The Pierre Auger Collaboration, 2018b (cf. 3.2), *i.e.* it is composed of a monopole together with a 6% dipole pointing towards right ascension  $\alpha = 100^\circ$  and declination  $\delta = -24^\circ$ . The map from the benchmark scenario is then multiplied by the exposure of the detector in order to get a hypothetical expected count map. From the hypothetical count map, 1000 random mock maps are created using an acceptance-rejection method. All the mocks have the same number of events  $\mathcal{N}$ .  $\mathcal{N}$  is taken to be the number of events detected from the start of the experiment up to December 2020 (date of the analysis). For each mock, the flux is reconstructed with the two methods using the same smoothing function as The Pierre Auger Collaboration, 2018b, which corresponds to a top-hat function with  $45^\circ$  radius. The expected dipole that should be reconstructed (modulo the randomness of mocks) is shown in Fig. 3.8. It corresponds to injected/observed dipole with a mask which takes into account the field of view of the detector.

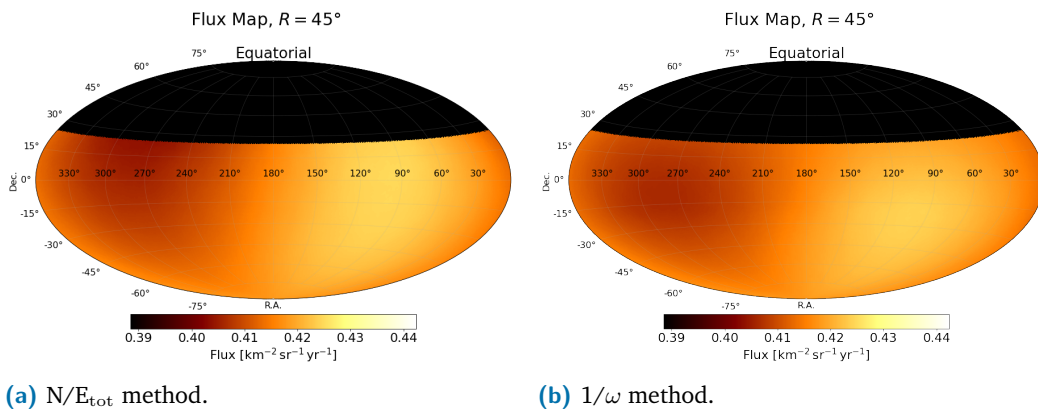
Figure 3.9 shows the average reconstructed map considering the two approaches. To enhance the differences, a bias map is built by subtracting and normalizing the expected dipole to the reconstructed ones:  $\text{Bias}(\alpha, \delta) = \frac{\phi^{\text{reconstructed}}(\alpha, \delta) - \phi^{\text{expected}}(\alpha, \delta)}{\phi^{\text{expected}}(\alpha, \delta)}$ . The bias map is set to zero in the masked area by definition.



**Fig. 3.8.:** Benchmark scenario,  $\phi^{\text{expected}}(\alpha, \delta)$ , with an applied mask corresponding to the field of view of the Pierre Auger Observatory.

The bias map is shown in Fig. 3.10. Looking at the bias, one can see that most differences appear at the border of the field of view. A check has been performed using a hypothetical full-sky field of view, and results in no bias and no differences between the two approaches. It confirms that the differences between the two approaches come from the limited aperture of the detector. Looking at Fig. 3.10, it seems that the  $N/E_{\text{tot}}$  method increases the contrast near the border while the  $1/\omega$  method reduces the contrast. To quantify the bias, the average bias is computed using:

$$\langle \delta \rangle = \sqrt{\frac{\sum_i \delta_i^2}{\mathcal{N}_{\text{pix}}}}, \quad (3.10)$$

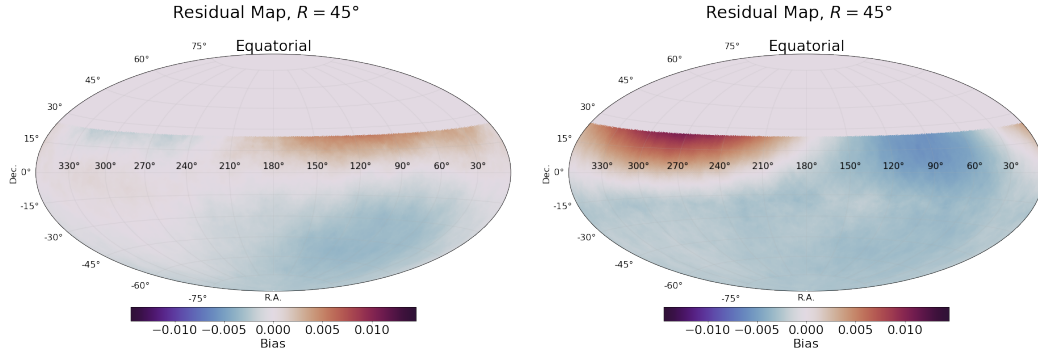


**Fig. 3.9.:** Average reconstructed dipole sky maps considering the two methods.



where  $\mathcal{N}_{\text{pix}}$  is the number of pixels and  $\delta_i$  is the bias in the  $i$ -pixel. It results in an average bias of:

$$\begin{cases} \langle \delta \rangle^{N/E_{\text{tot}}} = 0.2\% \\ \langle \delta \rangle^{1/\omega} = 0.3\% \end{cases} \quad (3.11)$$



(a)  $N/E_{\text{tot}}$  method.

(b)  $1/\omega$  method.

**Fig. 3.10.:** Bias map of the two reconstruction methods.

A standard deviation map is obtained by computing the standard deviation of the 1000 mock maps in each pixel. The standard deviation map is shown in Fig. 3.11. For the two methods, the variance is the same far away from the borders, however, the variance changes when being  $45^\circ$  away from the border. For the  $N/E_{\text{tot}}$  method, the standard deviation gradually increases when getting closer to the border, while for the  $1/\omega$  method, the standard deviation is separated in two areas: above and below  $45^\circ$  away from the border. As in the  $N/E_{\text{tot}}$  method, the area closer to the border has a bigger standard deviation than the area far from the border. As for the bias, the average standard deviation is computed using:

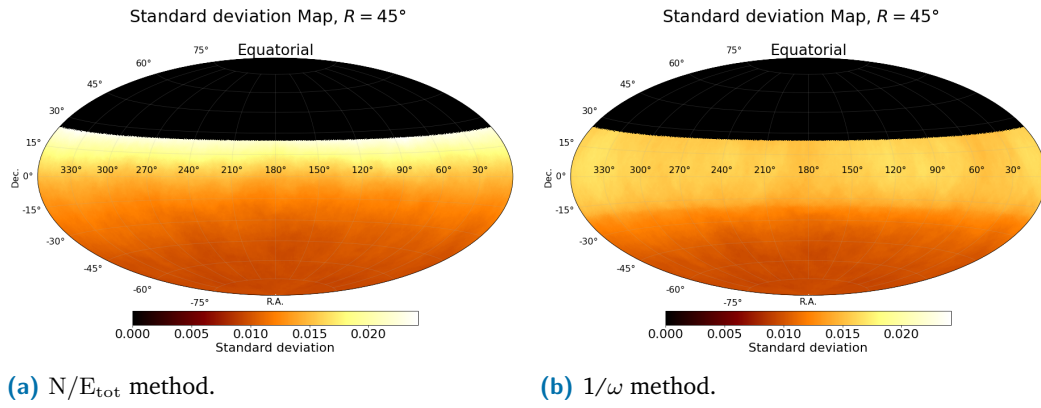
$$\langle \sigma \rangle = \sqrt{\frac{\sum_i \sigma_i^2}{\mathcal{N}_{\text{pix}} - 1}}. \quad (3.12)$$

It results in an average standard deviation of:

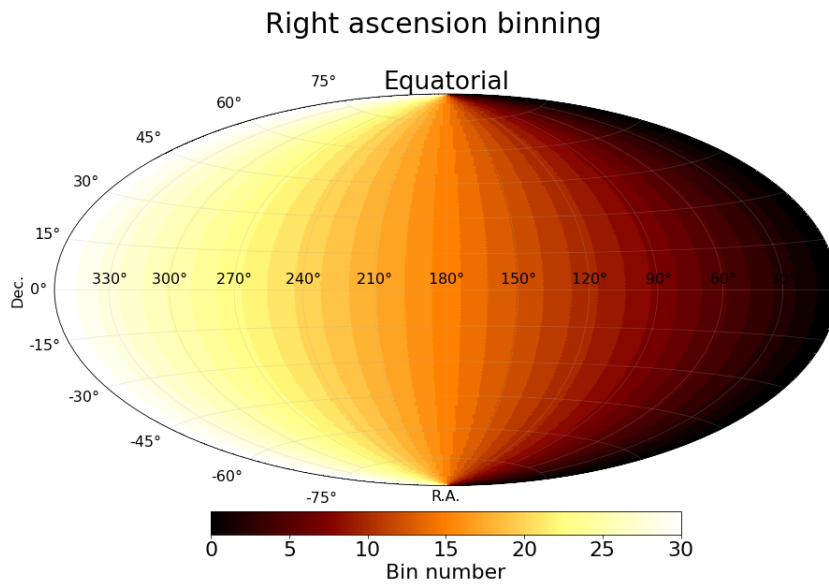
$$\begin{cases} \langle \sigma \rangle^{N/E_{\text{tot}}} = 1.4\% \\ \langle \sigma \rangle^{1/\omega} = 1.4\% \end{cases} \quad (3.13)$$

Besides, since the exposure does not depend on right ascension (cf. Eq. 3.2), an analysis in right ascension is performed (as it is done in The Pierre Auger Collaboration, 2018b). The average flux is computed in bin of right ascension. Each mock map is split in 30 bins of  $36^\circ$  as shown in Fig. 3.12





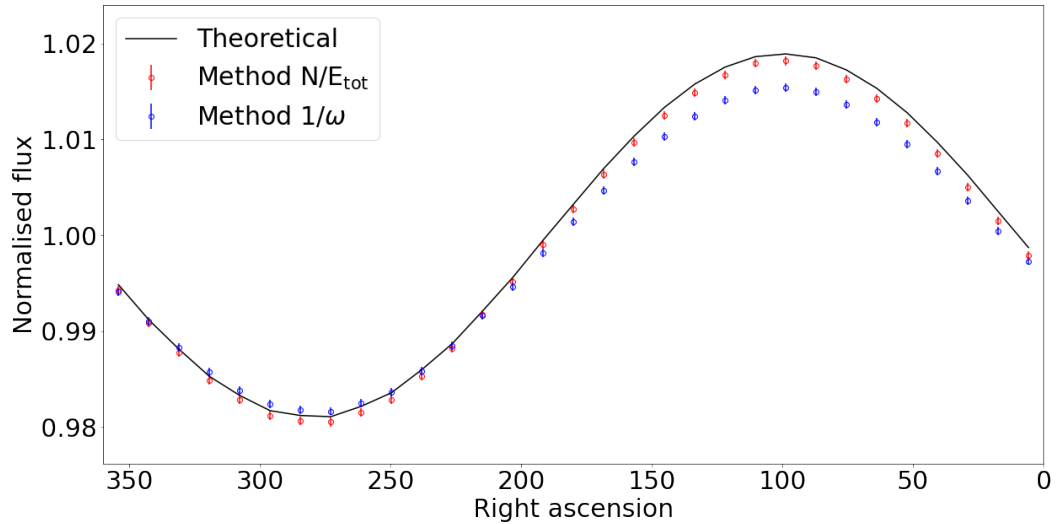
**Fig. 3.11.:** Standard deviation map considering the two reconstruction methods.



**Fig. 3.12.:** Bins used to reconstruct the flux in right ascension analysis.

The average flux reconstructed with the two methods and the average flux of the injected dipole as a function of right ascension are shown in Fig. 3.13. One can see that the  $N/E_{tot}$  method reproduces better the excess of flux around  $\alpha = 100^\circ$ , while the  $1/\omega$  method reproduces better the deficit of flux around  $\alpha = 280^\circ$ .

Comparing the average bias and average standard deviation in Eq. 3.11 & Eq. 3.13, one can see that the two approaches are equivalent in terms of bias and standard deviation. When looking at Fig. 3.13, one can see that the  $N/E_{tot}$  method reproduces better the excess, while the  $1/\omega$  method reproduces better the deficit. It looks like the  $N/E_{tot}$  method performs better on average, with mostly a small shift by  $\sim -0.1\%$ . An analysis has been performed using the catalogue found in 3.3.1 instead of the dipole. The results are similar and shown in Appendix A.



**Fig. 3.13.:** Flux analysis in right ascension. The dots correspond to the average flux in a bin of  $36^\circ$ .

## 3.2 Large scale anisotropy

Large-scale anisotropy can appear from a non-uniform spatial distribution of sources or if the cosmic ray propagation is diffusive in the case of a dominant source. Since 2017, the Pierre Auger Collaboration has reported a dipole in right ascension above 8 EeV. The dipole was found by means of a harmonic analysis on the observed arrival directions as a function of right ascension. The first-harmonic amplitude,  $r_\alpha = 0.06 \pm 0.01$ , allows the rejection of isotropy with  $6\sigma$  C.L., with a phase in the direction  $\alpha_0 = 98^\circ \pm 9^\circ$  (The Pierre Auger Collaboration, 2020a). A similar analysis was performed by the TA Collaboration, and results in  $r_\alpha = 0.03 \pm 0.02$  and  $\alpha_0 = 131^\circ \pm 33^\circ$  (Telescope Array Collaboration, 2020). The results of TA are compatible at  $1\sigma$  C.L. with the results of Auger. However, the small number of events makes the TA analysis rejecting the isotropy hypothesis with  $2\sigma$  C.L. only. A simplified analysis was developed by the author for the open data website of the Pierre Auger Observatory.<sup>1</sup> The simplified analysis is presented in the following section.

<sup>1</sup><https://opendata.auger.org/>

### 3.2.1 Rayleigh analysis in right ascension

A first-harmonic analysis consists of describing the flux as:

$$\Phi = \Phi_0 \times (1 + r_\alpha \cos(\alpha - \alpha_0)). \quad (3.14)$$

Such analysis is motivated by the exposure, which does not depend on the right ascension (cf. Eq. 3.2), and guarantees a robust analysis. Compared to the analysis reported in The Pierre Auger Collaboration, 2018b, the simplified analysis does not take into account the south-east tilt of the array and the small irregularities in the exposure of the array as a function of right ascension. Note as well that inclined events are not included here contrarily to The Pierre Auger Collaboration, 2018b. As in the previous section, the selected events are the vertical events detected up to December 2020, above an energy of 8 EeV.

In order to compute the amplitude,  $r_\alpha$ , and the phase,  $\alpha_0$ , of the first-harmonic (Eq. 3.14) a Rayleigh analysis can be performed. The Rayleigh analysis is done directly on the event basis, by computing:

$$\begin{cases} a_\alpha = \frac{2}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \cos \alpha_i \\ b_\alpha = \frac{2}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} \sin \alpha_i \end{cases}, \quad (3.15)$$

where  $\alpha_i$  is the right ascension of the  $i$ -th event and  $\mathcal{N}$  is the total number of events. Then, the amplitude and the phase can be computed using:

$$\begin{cases} r_\alpha = \sqrt{a_\alpha^2 + b_\alpha^2} \\ \alpha_0 = \arctan(b_\alpha/a_\alpha) \end{cases}. \quad (3.16)$$

From the analysis, one can compute the probability that a first-harmonic with an amplitude larger than  $r_\alpha$  appears in the data under the assumption of an isotropic scenario. Such a probability is given by:

$$P(r_\alpha) = \exp(-\mathcal{N} \frac{r_\alpha^2}{4}). \quad (3.17)$$

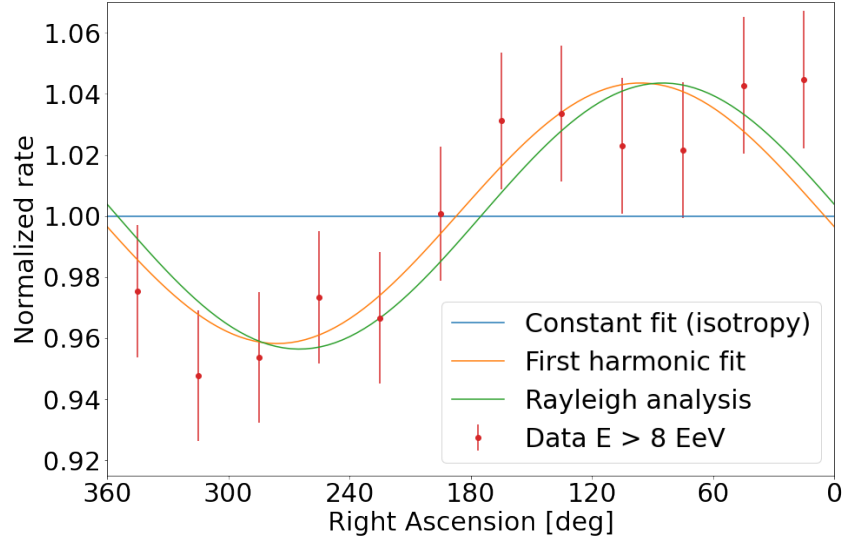
The uncertainties of the Fourier parameters,  $a_\alpha$  and  $b_\alpha$ , are analytical and given by

$$\sigma_{a_\alpha} = \sigma_{b_\alpha} = \sqrt{\frac{2}{\mathcal{N}}}, \quad (3.18)$$

Propagating uncertainty yields:

$$\begin{cases} \sigma_{r_\alpha} = \sqrt{\frac{2}{N}} \\ \sigma_{\alpha_0} = \frac{1}{r_\alpha} \sqrt{\frac{2}{N}} \end{cases} . \quad (3.19)$$

The results are shown as a green curve in Fig. 3.14.



**Fig. 3.14.:** Normalized flux of UHECRs as a function of right ascension. The green curve corresponds to the Rayleigh analysis, while the orange one corresponds to a first harmonic fit. A constant fit is shown in blue and refers to an isotropic scenario.

The points shown in Fig. 3.14 correspond to the normalized flux observed in a bin of right ascension. They are obtained by computing the number of events in a right ascension bin and normalizing to the expected average number of events per bin. The Rayleigh analysis is compared to a simple  $\chi^2$  test, where a cosine added a constant model are fitted to the data. The uncertainties of such a test are taken from the diagonal terms of the covariance matrix.

With Rayleigh statistics, the first harmonic is preferred at  $4.5\sigma$  compared to isotropy, with parameters:

$$\begin{cases} r_\alpha = 0.044 \pm 0.009 \\ \alpha_0 = 85^\circ \pm 12^\circ. \end{cases} \quad (3.20)$$

With  $\chi^2$  statistics, the first harmonic is preferred at  $4.4\sigma$  compared to isotropy, with parameters:

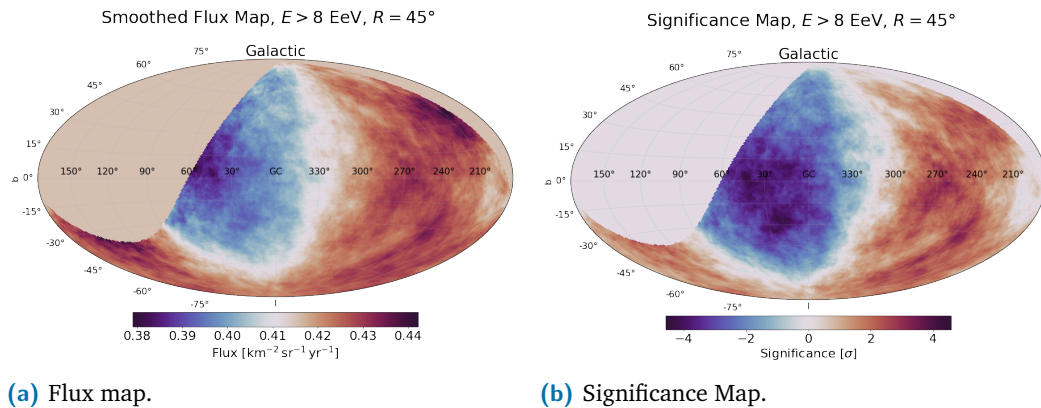
$$\begin{cases} \Phi_0 = 1.001 \pm 0.006 \\ r_\alpha = 0.043 \pm 0.008 \\ \alpha_0 = 96^\circ \pm 11^\circ. \end{cases} \quad (3.21)$$

The amplitude  $r_\alpha$  in right ascension has been studied as a function of energy. While the amplitude is not significant below  $E = 8 \text{ EeV}$ , it is found to increase with energy (The Pierre Auger Collaboration, 2020a). Besides, at EeV energies, the phase of the first harmonics is moving away from the Galactic center direction to the anti-center one (The Pierre Auger Collaboration, 2020a). Such a behaviour provides evidence that the pattern of UHECR arrival directions is not correlated to a Galactic origin but calls for an extragalactic one, as discussed in the following pages.

The flux sky map is also reconstructed using the  $N/E_{\text{tot}}$  method together with its significance map. The maps are shown in Fig. 3.15. The significance of excesses or deficits of events in a given region compared to the rest of the sky are computed using Li T.P. & Ma Y.Q., 1983. For a single region of interest, the significance is given by :

$$S = \sqrt{2} \times \sqrt{N_{\text{on}} \log \left[ \frac{1 + \alpha}{\alpha} \left( \frac{N_{\text{on}}}{N_{\text{on}} + N_{\text{off}}} \right) \right] + N_{\text{off}} \log \left[ (1 + \alpha) \left( \frac{N_{\text{off}}}{N_{\text{on}} + N_{\text{off}}} \right) \right]} \times \text{sign}(XS), \quad (3.22)$$

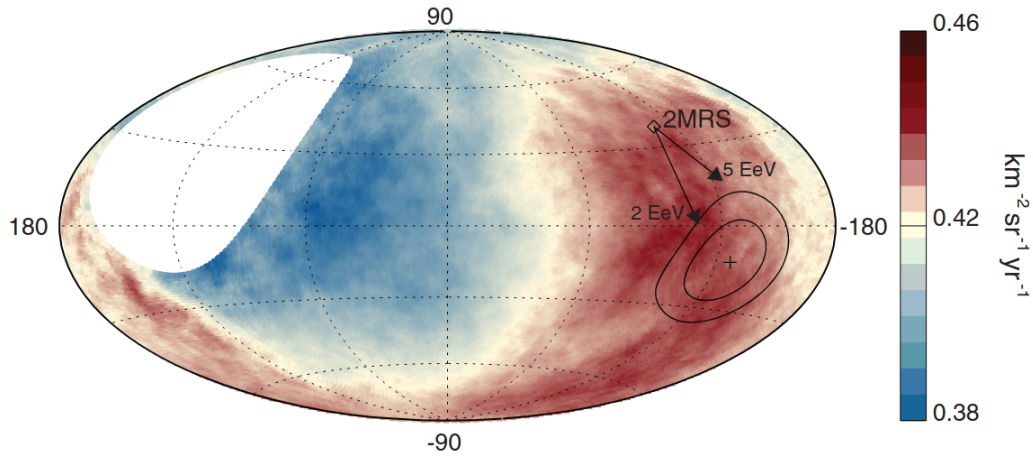
where  $N_{\text{on}}$  corresponds to the number of events counted in the region of interest (ON region), corresponding here to a disk of  $45^\circ$  radius.  $N_{\text{off}}$  corresponds to the number of background events counted in the rest of the sky (OFF region).  $\alpha$  is the



**Fig. 3.15.:** Reconstructed flux and significance map of events above 8 EeV.

ratio of the exposures of the ON and OFF regions, and  $XS = N_{\text{on}} - \alpha \times N_{\text{off}}$  is the number of events in excess in the ON region with respect to the background.

Figure 3.15 can be compared to Fig. 3.16, which is a sky map of the flux in Galactic coordinates observed above 8 EeV computed in The Pierre Auger Collaboration, 2018b. The main difference between this work and The Pierre Auger Collaboration, 2018b comes from the fact that the latter uses vertical and inclined events (events up to a zenith angle of  $80^\circ$ ) while this one uses vertical events only.



**Fig. 3.16.:** Auger UHECR flux in Galactic coordinates above 8 EeV smoothed with a  $45^\circ$ -radius top-hat function. The cross displays the observed dipole orientation, and the two contours stand for the 68% and 95% C.L. regions. The 2MRS dipole is indicated. Arrows indicate expected deflections of the 2MRS dipole considering the Galactic magnetic field of Jansson and Farrar, 2012, with protons at  $E = 5$  EeV or 2 EeV. Extracted from The Pierre Auger Collaboration, 2018b.

A spherical harmonic analysis in right ascension and declination allows the Auger Collaboration to reconstruct the 2D dipole on the sky, displayed in Fig. 3.16 as a cross. The 2D dipole points to  $115^\circ$  away from the Galactic center and  $55^\circ$  away from the direction of the 2MRS dipole, a near-infrared dipole that traces stellar mass within 250 Mpc. The arrows in Fig. 3.16 shows the expected dipole direction of UHECRs by accounting for the effect of the Galactic magnetic field if the sources are a fair sample of the 2MRS catalog. Each arrow corresponds to the average rigidity<sup>2</sup> obtained with different compositions  $R = 2$  EV (CNO dominated) and  $R = 5$  EV (Helium dominated). The two arrows are close to the  $2\sigma$  C.L. contour. Moreover, the anti-dipole direction is compatible at  $1\sigma$  C.L. with the Local-Void direction, a large empty intergalactic zone next to our own Local Group (The Pierre Auger collaboration, 2022). These arguments provide strong evidence for an extragalactic origin of UHECRs above 8 EeV. A similar analysis was performed using the Auger

<sup>2</sup> $R = E/Z$ , with  $E$  the energy and  $Z$  the charge of the particles

and TA data sets, and they are consistent with the Auger-only data set (Tinyakov P. on behalf of the Pierre Auger and Telescope Array collaborations, 2022).

## 3.3 Smaller scales and composition-dependent anisotropy

### 3.3.1 Intermediate scale anisotropy

By increasing energy at a fixed composition, we reduce the horizon of the UHECRs. Therefore, patterns from close sources can appear on smaller (intermediate) angular scales. In addition, increasing the energy can increase the rigidity  $R$  if the composition does not get heavier at the same pace. Therefore, if the average rigidity of the cosmic ray is increased, magnetic deflection is less.

Two types of studies have been performed at energy threshold,  $E_{\text{th}}$ , larger than 32 EeV. The first one consists in comparing the number of events observed above  $E_{\text{th}}$ , in a disk of radius  $\psi$ , with the number of events expected considering an isotropic scenario. A scan in threshold energy is performed. Such an analysis has been performed in several regions of interest (ROI):

- Scan over the full-sky,
- Supergalactic plane,
- Galactic plane,
- Galactic center,
- Centaurus region.

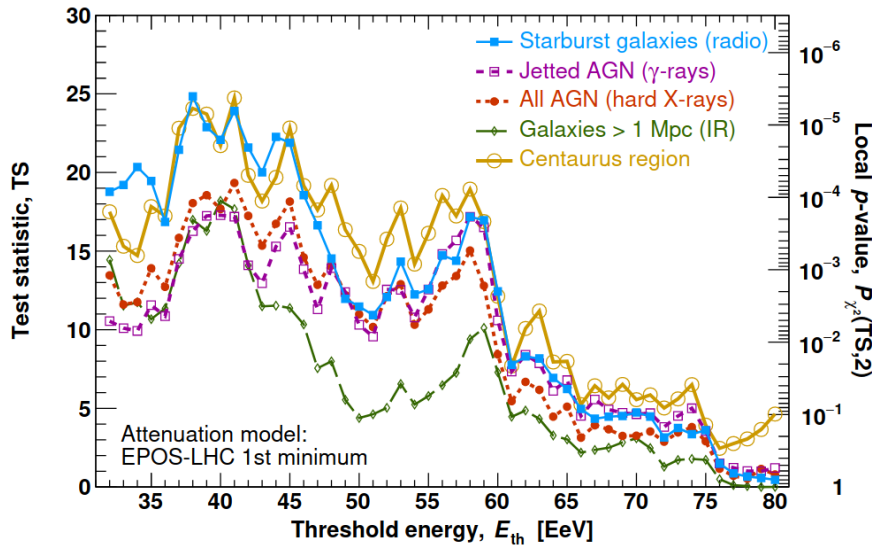
The last ROI of the list allows us to reject isotropy at  $4\sigma$  C.L. for  $E_{\text{th}} \sim 40$  EeV, while the other tests result in a significance smaller than  $2\sigma$  (The Pierre Auger collaboration, 2022). A search for autocorrelation of cosmic-ray directions has also been performed, yielding a significance smaller than  $1\sigma$ . The Centaurus region contains a major group of galaxies in the Council of Giants, with Centaurus A (radio galaxy), M83 (starburst galaxy), Circinus and NGC 4945 (galaxies with a Seyfert nucleus). The Centaurus region is thus composed of different types of galaxies and the host types cannot be distinguished from the UHECR point of view.

The second analysis is performed by comparing the arrival directions of UHECRs to catalogues of extragalactic sources. The chosen catalogues for the study are:

- Starburst galaxies, corresponding to galaxies with a high star formation rate;
- Jetted AGNs;
- All AGNs (both jetted and non jetted, i.e. Seyfert Galaxies);
- All galaxies beyond 1 Mpc.

The model corrects the flux of each galaxy according to the best-fit model prediction of the spectral and composition data (The Pierre Auger Collaboration, 2017). The flux of the galaxies is normalized to represent a fraction  $\alpha$  of the total flux. The flux is completed by an isotropic contribution normalized to  $(1 - \alpha)$ . The sky map model is smoothed on an angular scale  $\theta$  using a Fisher - von Mises function.  $\theta$  and  $\alpha$  are the two free parameters of the models. Finally, the data are compared to the model above  $E_{th}$  and a scan is performed on  $E_{th}$  by steps of 1 EeV.

The results of the scan are shown in Fig. 3.17. The result for the Centaurus region is also shown in yellow for comparison. In terms of structure, all curves seem to show a peak around  $\sim 40$  EeV and  $\sim 60$  EeV. Such a behavior could be explained in a rigidity-dependent scenario (Peters cycle) where the first peak would correspond

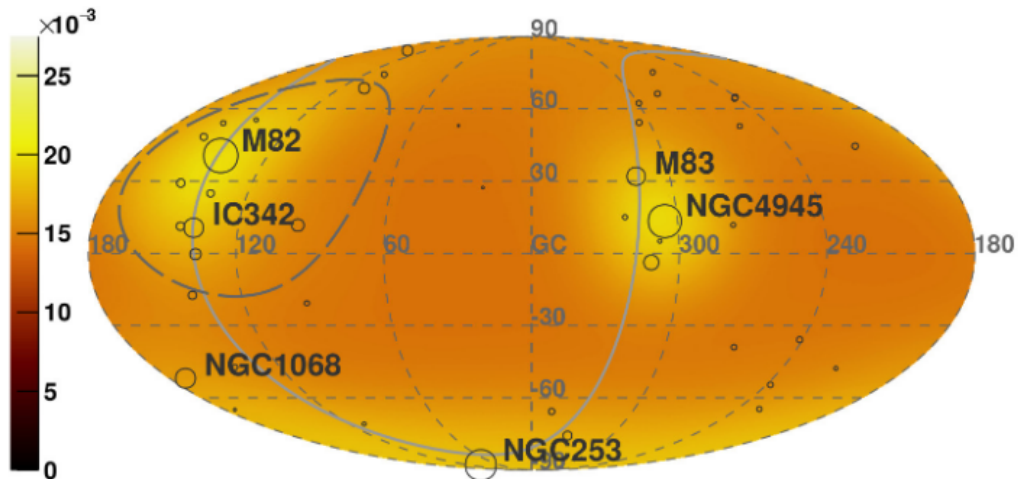


**Fig. 3.17.:** The test-statistic profile as a function of threshold energy for the catalogues studied and the Centaurus region. The local  $p$ -value shown in the right-hand side is penalized for signal fraction and angular scale scans (catalogues) and for the radius scans (Centaurus region). Extracted from Biteau J. on behalf of the Pierre Auger Collaboration, 2022.



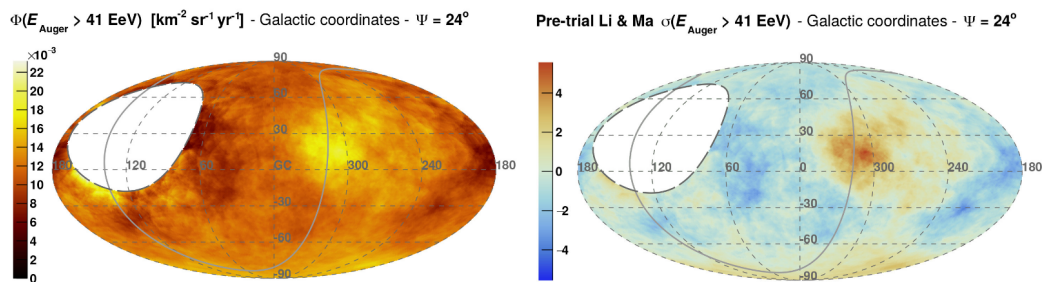
to a dominant abundance of CNO, the deficit around  $\sim 50$  EeV would correspond to a deficit of nuclei near fluorine in the accelerator, and the second peak would correspond to a dominant abundance of nuclei near neon. The catalog that provides the most significant signal is the starburst galaxy catalog, with a C.L. to reject isotropy similar to that of the Centaurus region ( $4\sigma$ ) above 38 EeV (Biteau J. on behalf of the Pierre Auger Collaboration, 2022).

**Starburst galaxies (radio) - expected  $\Phi(E_{\text{Auger}} > 38 \text{ EeV}) [\text{km}^{-2} \text{sr}^{-1} \text{yr}^{-1}]$**



**Fig. 3.18.:** The best-fit flux model from Starburst galaxies catalogues smoothed using a  $24^\circ$  radius top-hat function as for Fig. 3.19. Extracted from Biteau J. on behalf of the Pierre Auger Collaboration, 2022.

The pattern of the starburst-galaxy model is shown in Fig. 3.18. One can see that the main flux excess is in the Centaurus region coming from M83 and NGC4945, additional excesses could be expected from NGC253, and M82. The model can be



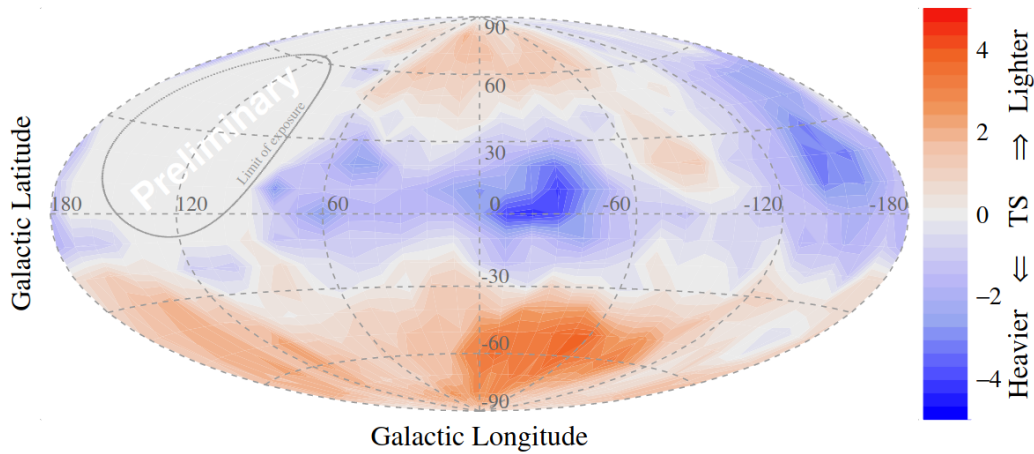
**Fig. 3.19.:** *Left:* Flux of cosmic rays observed above 41 EeV observed at the Pierre Auger Observatory smoothed with a top-hat function of radius  $24^\circ$ . *Right:* Associated Li-Ma pre-trial significance map. The solid gray line shows the super-galactic plane. Extracted from Biteau J. on behalf of the Pierre Auger Collaboration, 2022.

compared to the data (Fig. 3.19). One can see that the fluxes expected from M82 and IC342 are in the blind-spot of Auger. That is why the analysis was repeated by a joint working group of members of the Auger Collaboration and the TA Collaboration (Di Matteo A. on behalf of the Pierre Auger and Telescope Array collaborations, 2021). Using the two data sets, the significance of the starburst model to reject isotropy increases at  $4.2\sigma$  C.L. (Di Matteo A. on behalf of the Pierre Auger and Telescope Array collaborations, 2021).

On smaller scales, a search for magnetically-induced signatures at the event scale was performed by the Pierre Auger Collaboration without significant deviation from isotropy (The Pierre Auger Collaboration, 2020f).

### 3.3.2 Indication of mass anisotropy

In 2021, the Pierre Auger Collaboration published an indication of a mass-dependent anisotropy above the ankle energy at 5 EeV (Mayotte, E. on behalf of the Pierre Auger Collaboration, 2021). The sky map is divided into two regions, the on-plane region with Galactic latitude  $b \leq |b_{\text{split}}|$  and an off-plane region with Galactic latitude  $b \geq |b_{\text{split}}|$ . The two  $X_{\text{max}}$  distributions are compared using a 2-sample Anderson-Darling test with a scan performed on  $b_{\text{split}}$  by  $5^\circ$  step and  $E_{\text{th}}$  by logarithmic steps of  $\Delta \log_{10}(\frac{E}{\text{eV}}) = 0.1$ .



**Fig. 3.20.:** Mass-sensitive sky map for  $E \geq 10^{18.7}$  eV. Extracted from Mayotte, E. on behalf of the Pierre Auger Collaboration, 2021.

The most significant results are found at  $b_{\text{split}} = 30^\circ$  and  $E_{\text{th}} = 10^{18.7}$  eV  $\simeq 5$  EeV. The mean value of  $X_{\text{max}}$  is  $9.1 \pm 1.6$  g/cm<sup>2</sup> smaller in the on-plane region than in the off-plane region. The standard deviation of  $X_{\text{max}}$  is  $5.9 \pm 2.1$  g/cm<sup>2</sup> smaller

in the on-plane region than in the off-plane region. Such differences between the out-of-plane region and the in-plane region provide indication for a mass-dependent structure in arrival directions, with  $3.3\sigma$  C.L. after penalization for the trials and accounting for systematic uncertainties.

A map showing the test statistic (TS) is displayed in Fig. 3.20. The results suggest that the Galactic magnetic field may have an effect on the mass-dependent anisotropy. However, the different horizons of different nuclei could also explain the composition anisotropy. Therefore, further studies are needed to investigate the mass-dependent anisotropy in depth. The use of different  $X_{\max}$  reconstruction methods (air-shower universality (Bridgeman A. on behalf of the Pierre Auger Collaboration, 2018) or deep learning techniques (The Pierre Auger Collaboration, 2021a)), as well as the data of the upgraded Auger Observatory, could be crucial to study the composition sky map.

However, before obtaining a composition sky map, a study of the composition by itself is necessary. The next chapter shows how the composition of cosmic rays in a given energy range can be estimated from  $X_{\max}$  distributions.

# Study of the composition of UHECR

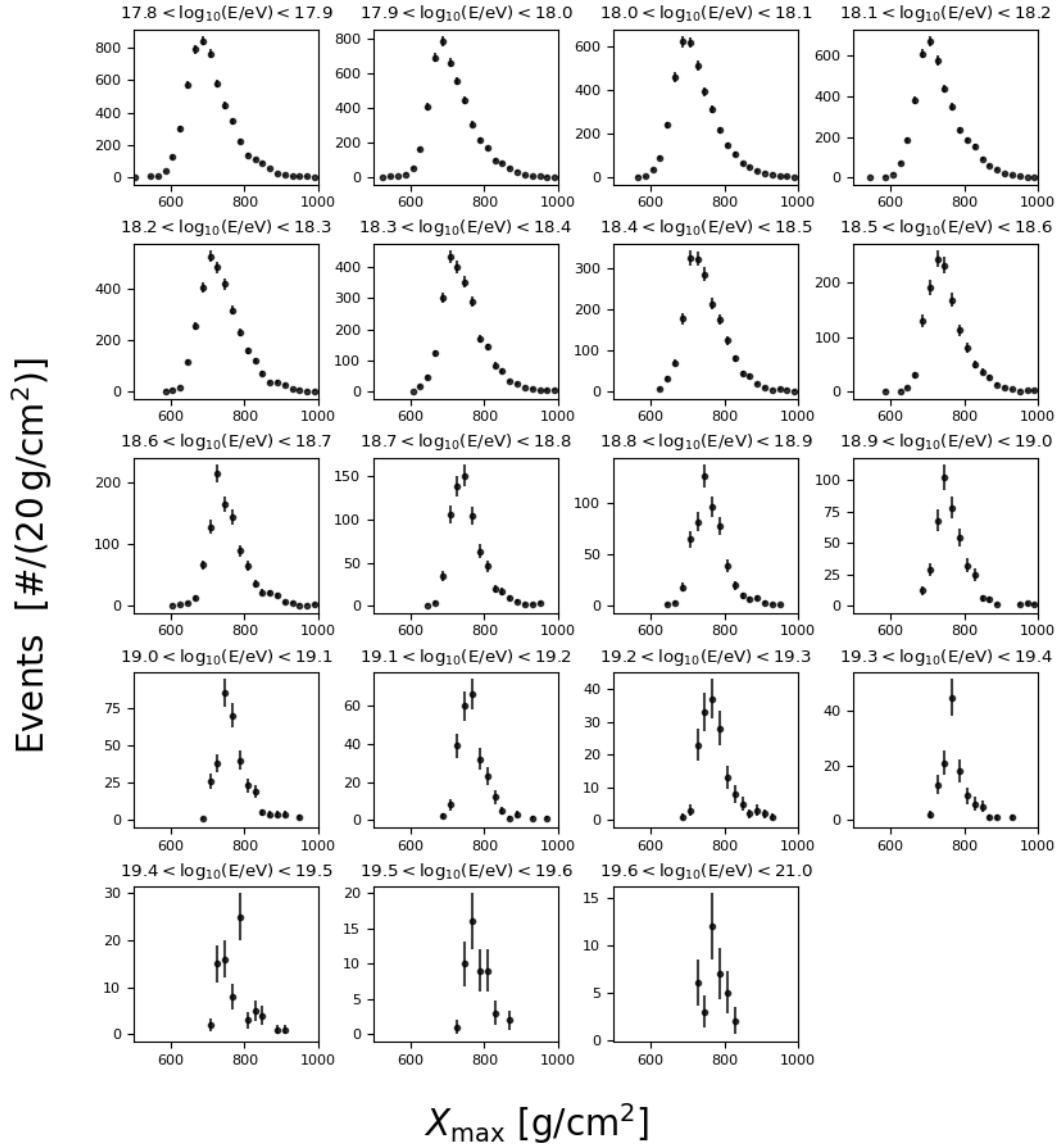
*This chapter is partly adapted from the author's work described in the Auger internal note, GAP2021 021 "Negative fractions in a frequentist analysis of composition derived from  $X_{\max}$  distributions", Marafico et al., 2021.*

## 4.1 From $X_{\max}$ estimators to composition

We have seen in the first chapter that the propagation of UHECRs depends on their mass (cf. Fig. 1.10). As an example, for a given travel distance, helium nuclei have a larger probability to photodisintegrate compared to iron nuclei. Therefore, constraining the composition of UHECRs can give us an information on the maximum distance of sources, as well as information about their environment and mechanism of acceleration.

As explained in the first chapter,  $X_{\max}$  fluctuates from shower to shower, and it cannot be used to determine the mass of a single event. However, for a given energy and a given mass, one can expect a given  $X_{\max}$  distribution. In this chapter, we will study  $X_{\max}$  distributions in order to get information on the mass of the incoming cosmic rays.

In order to do so, FD events are selected and distributed in terms of  $X_{\max}$  for a given energy bin. The energy bin is chosen to be of width  $\Delta \log_{10} = 0.1$  from  $10^{17.8}$  eV to  $10^{19.6}$  eV. Above  $10^{19.6}$  eV, there are too few events (e.g. 35 as of 2021) to have a bin of width  $\Delta \log_{10} = 0.1$ . Hence, a last bin is done with events with energy  $E \geq 10^{19.6}$  eV. The binning in  $X_{\max}$  is chosen to be of 20 g/cm<sup>2</sup>. The resulting distributions are shown in Fig. 4.1. The statistical uncertainties are expected to follow a Poisson distribution, and they are displayed following  $\sigma_{\text{stat}} = \sqrt{N}$  with  $N$  the number of events detected in a  $X_{\max}$  bin.



**Fig. 4.1.:**  $X_{\max}$  distribution for given energy bins from the ICRC 2019 dataset (dataset produced for the 37th International Cosmic Rays Conference).

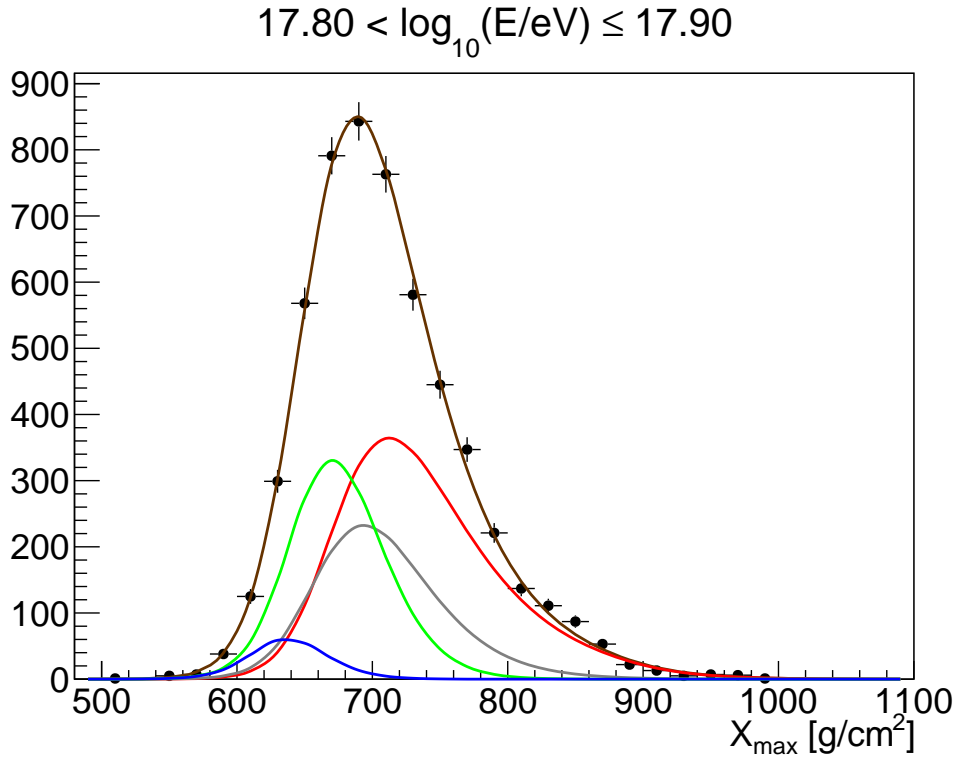
#### 4.1.1 Fitting procedure

The idea now is to reproduce the  $X_{\max}$  distributions using expected ones from simulations. Using distribution of all kinds of nuclei would result in an important number of free parameters. Hence, only a few selected representative masses are used to reproduce the data. Most of the time, the representative masses are chosen as follows:  ${}^1_1\text{H}$ ,  ${}^2_4\text{He}$ ,  ${}^7_{14}\text{N}$ ,  ${}^{26}_{54}\text{Fe}$ . Such a choice is motivated by the fact they are equally logarithmically spaced. Sometimes, an extra mass is used,  ${}^{14}_{28}\text{Si}$ . In the end,

the distribution is modelled as the superposition of simple mass species contributing with an amplitude  $\alpha_A$  :

$$G^{\text{model}}(X_{\text{max}}|E) = \sum_A \alpha_A G_A^{\text{model}}(X_{\text{max}}|E), \quad (4.1)$$

where  $A$  is the mass of each representative nucleus and  $G_A^{\text{model}}(X_{\text{max}}|E)$  is the expected  $X_{\text{max}}$  distribution for a nucleus of mass  $A$  and energy  $E$ . An example of model reproducing the data is shown in Fig. 4.2.



**Fig. 4.2.:** Example of an  $X_{\text{max}}$  distribution modelled with four different masses. The red curve corresponds to  ${}^1_1\text{H}$ , grey to  ${}^4_2\text{He}$ , green to  ${}^{14}_7\text{N}$ , blue to  ${}^{54}_{26}\text{Fe}$ ; the sum is shown in brown.

To construct the distribution of representative masses  $G_A^{\text{model}}(X_{\text{max}}|E)$ , EAS have been simulated using CONEX (Pierog et al., 2006) for a given energy bin and for a given mass (Sergio Petrera, 2020). A hadronic model needs to be associated to CONEX in order to describe the hadronic interactions in EAS. Currently, there are three main hadronic models available on the market: EPOS-LHC (Pierog et al., 2015), Sibyll2.3d (Riehn et al., 2020), and QGS-JETII-04 (Ostapchenko, 2011). The simulated distributions are modelled with a parametrized function. Several functions have been tested: log-normal distribution, exponentially modified Gaussian distribution, generalized Gumbel distribution... As shown in Arbeletche

and de Souza, 2020, the generalized Gumbel distribution reproduces better the  $X_{\max}$  distribution. Therefore, generalized Gumbel distribution is used in this study, as shown in eq. 4.2:

$$g(u) = \frac{1}{\sigma} \frac{\lambda^\lambda}{\Gamma(\lambda)} \exp(-\lambda u - \lambda \exp(-u)) \text{ with } u = \frac{X_{\max} - X_0}{\sigma}. \quad (4.2)$$

The parameters  $\lambda$ ,  $\sigma$ ,  $X_0$  depend on the mass and the energy of the nucleus and differ for each hadronic model. We use as a baseline the parameterizations obtained with EPOS-LHC (Sergio Petrera, 2020). The function is then discretized to be compared to histograms of Fig. 4.1:

$$g(u) \rightarrow g_{jxA}, \quad (4.3)$$

where  $j$  stands for the energy bin,  $A$  is the mass of the nucleus, and  $x$  is the  $X_{\max}$  bin.

In our case, the distributions shown in Fig. 4.1 are impacted by the resolution and the acceptance of the detector. Hence, the model to be compared to the data is convolved with the acceptance and the resolution:

$$G_{jxA}^{\text{model}} = \sum_{x'} g_{jx'A} \text{Acc}_j(x') \text{Res}_j(x - x'), \quad (4.4)$$

where  $\text{Acc}_j(x')$  is the acceptance of the detector and  $\text{Res}_j(x - x')$  is the resolution of the detector. Finally, the probability to observe an event in a bin  $x$  of  $X_{\max}$  and in a bin  $j$  of energy, for a given mass  $A$ , is given by  $G_{jxA}^{\text{model}}$ .

## Likelihood estimator: General point

In order to fit the model to the data, we use a likelihood procedure. When matching the observed  $X_{\max}$  distributions, two different likelihood estimators are possible. The first one involves a multinomial approach where the number of expected events in the model is equal to the number of detected events:

$$n_{\text{expected}} = n_{\text{detected}}. \quad (4.5)$$

The second one involves a Poisson approach, where the number of expected events in the model is not fixed. The next two pages describe the two approaches.

## Likelihood estimator: Multinomial approach

In the multinomial approach, the model is normalized to verify the condition in eq. 4.5:

$$G_{jx}^{\text{model}} = \frac{n_{\text{detected}}}{n_{\text{expected}}} \sum_A \alpha_A G_{jxA}^{\text{model}} \quad (4.6)$$

where  $\alpha_A$  are the parameters of the model, each one being associated to a representative mass. The number of expected events,  $n_{\text{expected}}$ , is given by the integral of the model:  $n_{\text{expected}} = \sum_x \sum_A \alpha_A G_{jxA}^{\text{model}}$ . Normalizing using  $\frac{1}{n_{\text{expected}}}$  as done in Eq. 4.6 is equivalent to have  $n-1$   $\alpha_A$  parameters and compute the last one with  $\alpha_{\text{last}} = \frac{n_{\text{detected}} - \sum_x \sum_A^{n-1} \alpha_A G_{jxA}^{\text{model}}}{\sum_x G_{jxA_{\text{last}}}^{\text{model}}}$ . Hence, the approach has three effective parameters when four components are considered.

The likelihood associated to the multinomial approach is given by:

$$\mathcal{L}_j = n_j^{\text{data}}! \prod_x \frac{1}{k_{jx}^{\text{data}}!} (G_{jx}^{\text{model}})^{k_{jx}^{\text{data}}}, \quad (4.7)$$

where  $n_j^{\text{data}}$  is the number of observed events in an energy bin  $j$  and  $k_{jx}^{\text{data}}$  is the number of observed events in a bin  $x$ , at an energy  $j$ .

From the likelihood, one can define the deviance, a goodness-of-fit estimator. It is computed as:

$$D_j = -2 \log \left( \frac{\mathcal{L}_j}{\mathcal{L}_j^{\text{Sat}}} \right), \quad (4.8)$$



where  $\mathcal{L}_j^{\text{Sat}}$  is the likelihood of the saturated model which describes perfectly the data. Hence, the deviance is given by:

$$D_j = -2 \sum_x k_{jx}^{\text{data}} \times (\log(G_{jx}^{\text{model}}) - \log(k_{jx}^{\text{data}})) \quad (4.9)$$

### Likelihood estimator: Poisson approach

In the Poisson approach, the number of expected events is not fixed. Hence, four parameters are needed and the model is given by:

$$G_{jx}^{\text{model}} = \sum_A \alpha_A G_{jxA}^{\text{model}} \quad (4.10)$$

The likelihood associated to the Poisson approach is given by:

$$\mathcal{L}_j = \prod_x \frac{e^{-G_{jx}^{\text{model}}}}{k_{jx}^{\text{data}}!} (G_{jx}^{\text{model}})^{k_{jx}^{\text{data}}}, \quad (4.11)$$

Following eq.4.8, the associated deviance is:

$$D_j = -2 \sum_x k_{jx}^{\text{data}} \times (\log(G_{jx}^{\text{model}}) - \log(k_{jx}^{\text{data}})) + (k_{jx}^{\text{data}} - G_{jx}^{\text{model}}) \quad (4.12)$$

For both approaches, the results are shown in terms of fractions instead of  $\alpha_A$ , where the fractions are computed as:

$$f_A = \frac{\alpha_A}{\sum_A \alpha_A} \quad (4.13)$$

We checked that using equation 4.13 is strictly equivalent to the equation 4.14 which corresponds to the true fractions observed. The fractions are not “acceptance dependent”.

$$f_A = \frac{\sum_x \alpha_A \sum_{x'} g_{jx'A} \text{Res}_j(x - x')}{\sum_x \sum_A \alpha_A \sum_{x'} g_{jx'A} \text{Res}_j(x - x')} \quad (4.14)$$

## $p_{\text{value}}$ : $\chi^2$ and C-statistics

From the deviance, one can estimate a  $p_{\text{value}}$ . If one considers  $\chi^2$  statistics, the  $p_{\text{value}}$  is given by:

$$p_{\text{value}} = 1 - \mathcal{P}_{\text{ndf}}^{\chi^2}(x < D), \quad (4.15)$$

where  $\text{ndf}$  denotes the number of degrees of freedom (number of points – number of parameters), and  $\mathcal{P}_{\text{ndf}}^{\chi^2}(x < D)$  is the cumulative distribution function of  $\chi^2$  distribution of parameter  $\text{ndf}$ . It corresponds to the probability that a correct model gives a better deviance than  $D$ . Hence, the  $p_{\text{value}}$  denotes the probability that a correct model yields a deviance at least as extreme as  $D$ .

A simple  $\chi^2$   $p_{\text{value}}$  does not take into account empty  $X_{\text{max}}$  bins. However, an empty bin (i.e. not measuring something) contains information. Therefore, under the assumption that simulated deviances are Gaussian distributed, one can use C-statistics<sup>1</sup> as shown in Bonamente, 2019. C-statistics takes into account empty bins to compute the  $p_{\text{value}}$ . The  $p_{\text{value}}$  is then given by:

$$p_{\text{value}} = 1 - \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{D - \langle D \rangle - n_{\text{par}}}{\sqrt{2 \times (\sigma_D^2 - 2 \times n_{\text{par}})}} \right) \right] \quad (4.16)$$

where  $\langle D \rangle$  is the expected deviance and  $n_{\text{par}}$  is the number of representative masses,

$$\langle D \rangle = \sum_x D_x = \sum_x \sum_{n=0}^{\infty} 2 \left( n \log \left( \frac{n}{G_{jx}^{\text{model}}} \right) + G_{jx}^{\text{model}} - n \right) \frac{\exp(-G_{jx}^{\text{model}}) (G_{jx}^{\text{model}})^n}{n!} \quad (4.17)$$

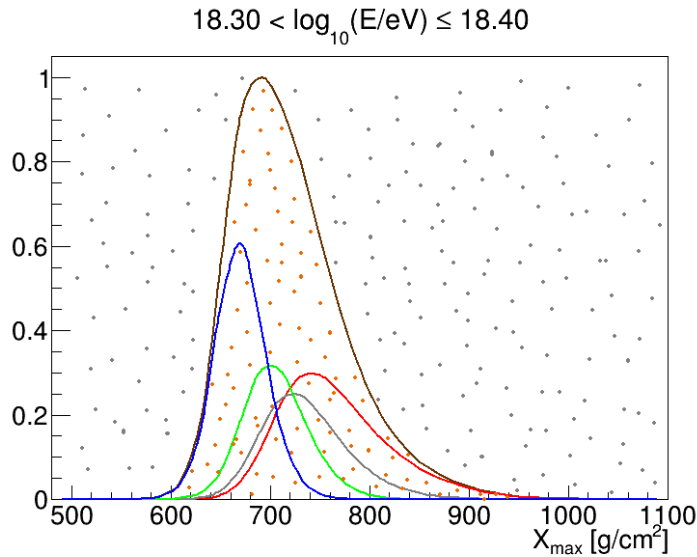
where  $\sigma_D$  is the expected rms:  $\sigma_D = \sqrt{\langle D^2 \rangle - \langle D \rangle^2}$ .

### 4.1.2 Monte-Carlo simulations

To evaluate the performance of the multinomial and Poisson approaches, one can run mock/Monte-Carlo (MC) simulations to test the model. Here, 1000 MC simulations are run from a benchmark scenario. Both likelihoods are employed to match each of the 1000 MC simulations and to compare the reconstructed parameters to the injected benchmark scenario.

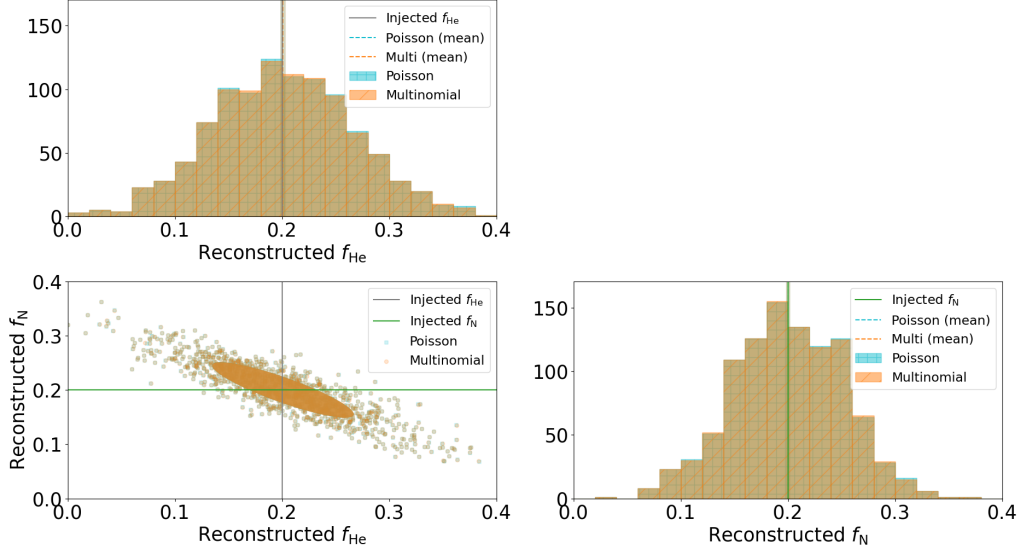
<sup>1</sup>Proposed by Armando Di Matteo for  $X_{\text{max}}$  distributions.

For the first test, the benchmark scenario is run from a composition with  $f_{\text{H}} = 0.3$ ,  $f_{\text{He}} = 0.2$ ,  $f_{\text{N}} = 0.2$ ,  $f_{\text{Fe}} = 0.3$ , in the energy bin  $18.3 \leq \log_{10}(\text{E}/\text{eV}) < 18.4$ . Each MC simulation contains  $N$  events, where  $N$  is taken randomly for each simulation from a Poisson distribution with parameter  $\lambda = 3000$ . The mocks are built using the acceptance rejection method starting from the benchmark scenario. An example is shown in Fig. 4.3. Points are drawn at random in the plane shown. If the points are below the model, they are kept (orange dots), if they are above, they are thrown away (grey). When there are  $N$  orange points, a histogram is built and a new draw is made. The 1000 MC simulated distributions are modelled using the multinomial and Poisson approaches. Figure 4.4 shows the results obtained for the two different approaches. In Fig. 4.4, each dot represents the best-fit fraction for Helium and Nitrogen for a given MC simulation.



**Fig. 4.3.:** Distribution used in the benchmark scenario ( $f_{\text{H}} = 0.3$ ,  $f_{\text{He}} = 0.2$ ,  $f_{\text{N}} = 0.2$ ,  $f_{\text{Fe}} = 0.3$ , in the energy bin  $18.3 \leq \log_{10}(\text{E}/\text{eV}) < 18.4$ ). Points represent the acceptance-rejection method: Orange for kept points, grey for rejected points.

Figure 4.4 shows that on average, for both reconstructions, the 1000 MC simulations reproduce the injected composition. In this test, both approaches give the same outcome irrespective of the number of injected events (tested with  $N$  ranging from 30 to 3000). However, we notice up to 5 outliers are obtained out of 1000 MC simulations for the multinomial approach. Hence, during the rest of the study, we only show the results coming from the Poisson approach, which appears more robust.

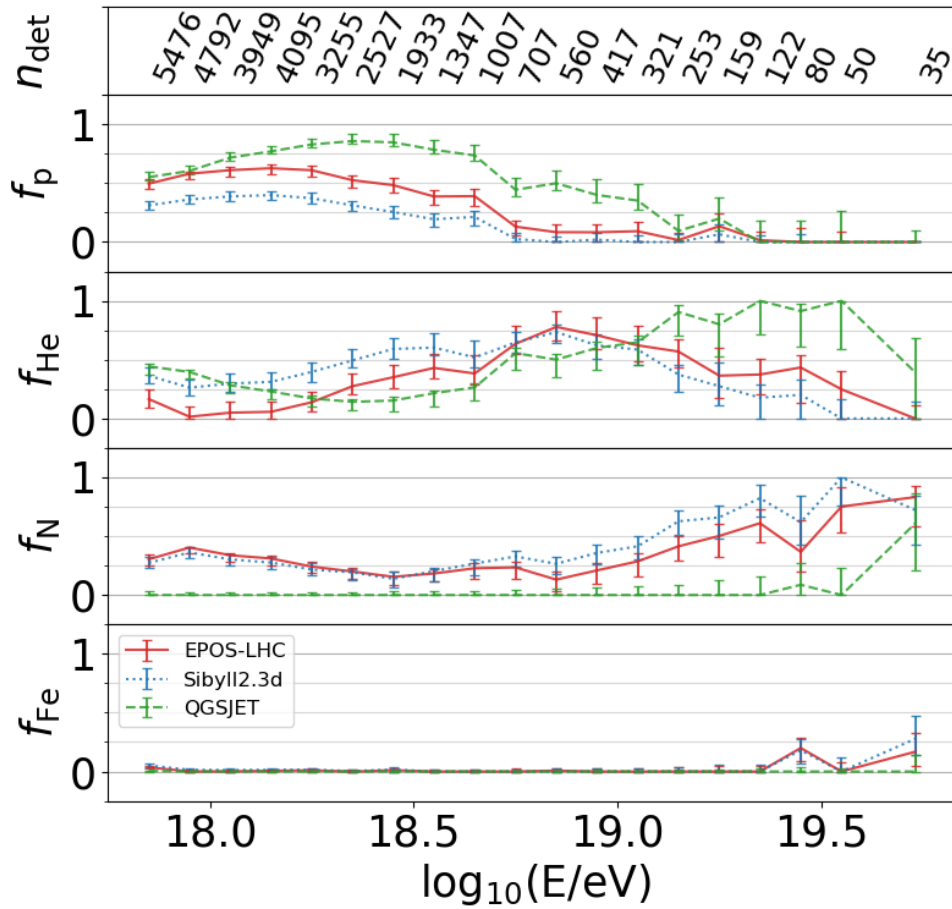


**Fig. 4.4.:** Multinomial (red) & Poisson (blue) reconstructions of 1000 MC simulations, with an injected composition  $f_H = 0.3$ ,  $f_{He} = 0.2$ ,  $f_N = 0.2$ ,  $f_{Fe} = 0.3$  in the energy bin  $18.3 \leq \log_{10}(E/\text{eV}) < 18.4$ , with  $N = \text{Poisson}(3000)$ . The grey line corresponds to the fraction of Helium injected, the green line corresponds to the fraction of Nitrogen injected.

### 4.1.3 Results on data

Figure 4.5 shows the result of the fit applied to data from ICRC 2019 (Yushkov A. on behalf of the Pierre Auger Collaboration, 2021). The uncertainties are computed using MC simulations. 1000 MC simulations are run from the best-fit scenario of each energy bin. As for the previous test, each MC simulation contains  $N$  events, with  $N$  following a Poisson distribution with parameter  $\lambda$  equal to  $n_{\text{det}}$ , the number of detected events in the energy bin. Each MC is fitted 1000 times, and the 68% confidence interval of the reconstructed fractions of the mocks gives the uncertainty. The three main hadronic models are tested. Fig. 4.5 shows that the three hadronic models result in a composition getting heavier as a function of energy. Such a behaviour can be expected from Peters' cycle (cf. 1.2.1).

EPOS-LHC gives a composition dominated by proton (hydrogen) below  $10^{18.2}$  eV. At this point, the He fraction starts to increase and peaks around  $10^{18.7}$  eV before decreasing. The nitrogen fraction, which traces the composition in terms of Carbon, Nitrogen, Oxygen seems to slightly decrease from  $10^{17.8}$  eV up to  $10^{18.7}$  eV, then increases up to the last energy bin. Sibyll2.3d hadronic models results in similar trends, with a composition going from proton to nitrogen when increasing the energy of interest. On the other side, QGSJET hadronic model results in a light composition



**Fig. 4.5.:** Best-fit parameters for each energy bin using the ICRC 2019 dataset (Yushkov A. on behalf of the Pierre Auger Collaboration, 2021) considering different hadronic models: EPOS-LHC (solid red), Sibyll2.3d (dotted blue), QGS-JETII-04 (dashed green). The lines are here to guide the eye.

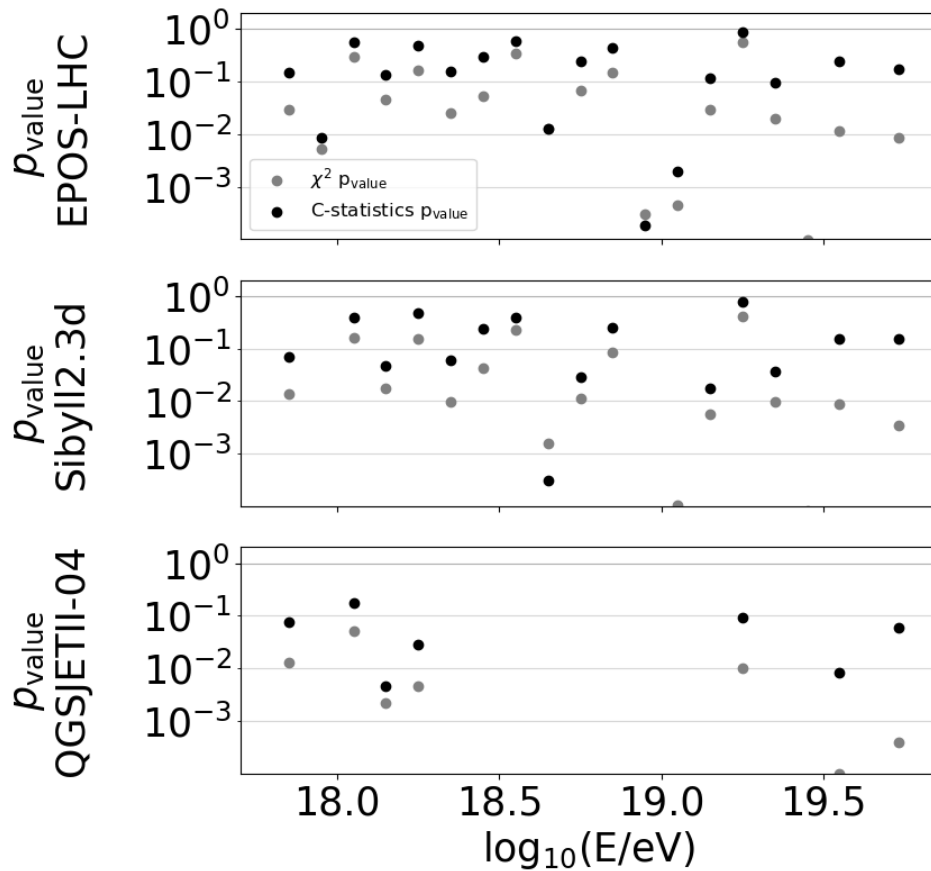
made exclusively of proton and helium. It is important to note that iron is not significantly detected for any hadronic interactions model.

The associated  $p_{value}$  computed using  $\chi^2$  statistics and C-statistics are shown in Fig. 4.6. One can see that the  $p_{value}$  using C-statistics are better than the ones found with  $\chi^2$  statistics. This comes from the fact that C-statistics consider empty bins, which provides a better assessment of the agreement between the model and the data. It is important to note that Sibyll and especially EPOS-LHC reproduce rather well the observed data for most of the energy bins (good  $p_{value}$  overall). However, QGS-JETII-04 does not successfully reproduce the observed  $X_{max}$  distributions. Hence, QGS-JETII-04 cannot be used to interpret Auger data. Given the number of

test bins,  $n_E$ , one can quantitatively estimate the probability to obtain a  $p_{\text{value}}$  at least as bad as observed:  $1 - p_{\text{post}} = (1 - p_{\text{worst}})^{n_E}$ .

- EPOS-LHC:  $p_{\text{post}} = 6.3 \times 10^{-4}$
- Sibyll2.3D:  $p_{\text{post}} = 1.8 \times 10^{-6}$
- QGS-JET22-04:  $p_{\text{post}} = 6.3 \times 10^{-15}$

As before, the two first hadronic interaction models are better than QGS-JET. However, overall, none of the hadronic interaction models is perfect, and it shows how difficult it is to reproduce  $X_{\text{max}}$  distributions.



**Fig. 4.6.:**  $p_{\text{value}}$  obtained with different hadronic models. Grey dot show the  $p_{\text{value}}$  obtained from  $\chi^2$  statistics. Black dot show the  $p_{\text{value}}$  obtained from C-statistics.

A composition analysis has been performed by the Telescope Array collaboration (Hanlon W. on behalf of the Telescope Array collaboration, 2021). Telescope Array uses QGS-JET model, which results in a composition mostly made of light nuclei (proton and helium). However, when using representative masses as done above,

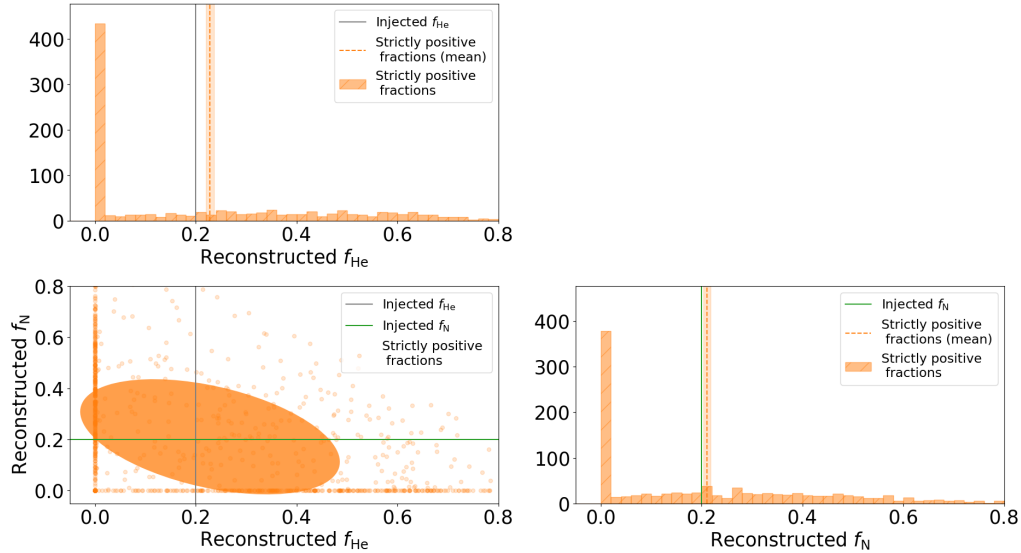
they found evidence of mixed composition, including light nuclei, medium nuclei (nitrogen) and heavy elements (iron) (Hanlon W. on behalf of the Telescope Array collaboration, 2021).

## 4.2 A biased approach?

In the previous section, we described a composition study similar to that developed in Bellido J. on behalf of the Pierre Auger Collaboration, 2018. As shown in Fig. 4.4, when the parameters are well-chosen, the results are not biased. However, biases can appear when getting closer to the boundaries of the parameter space. In the following section, the bias is studied.

### 4.2.1 Test with fractions compatible with the boundaries

A test is performed with the same scenario injected in Fig. 4.4 ( $f_H = 0.3$ ,  $f_{He} = 0.2$ ,  $f_N = 0.2$ ,  $f_{Fe} = 0.3$ , in the energy bin  $18.3 \leq \log_{10}(E/eV) < 18.4$ ), but with a smaller number of detected events. For each MC simulation, the number of events  $N$  is given by  $N = \text{Poisson}(30)$ . The results are shown in Fig. 4.7.

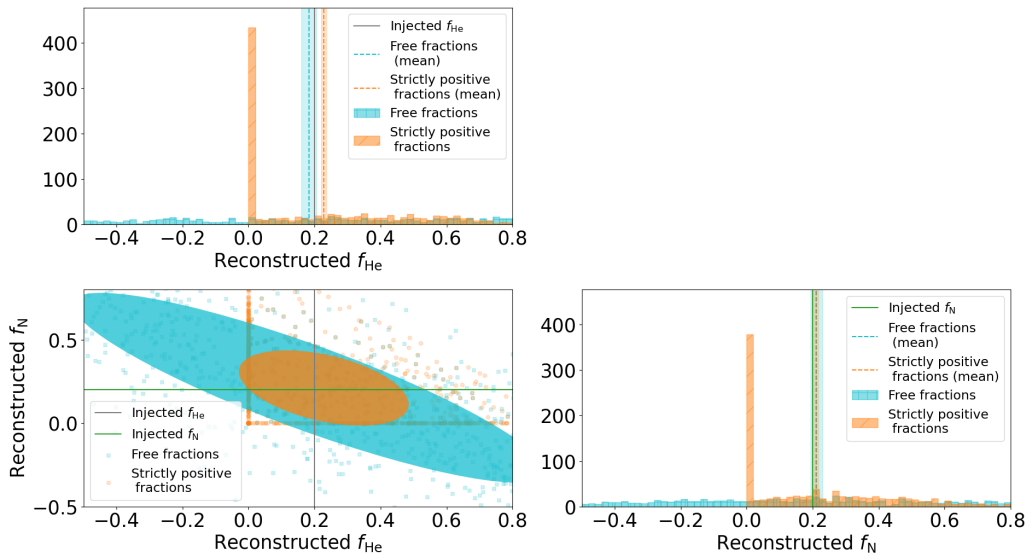


**Fig. 4.7.:** Poisson reconstruction of 1000 MC simulations, with an injected composition  $f_H = 0.3$ ,  $f_{He} = 0.2$ ,  $f_N = 0.2$ ,  $f_{Fe} = 0.3$  in the energy bin  $18.3 \leq \log_{10}(E/eV) < 18.4$ , with  $N = \text{Poisson}(30)$ . The grey line corresponds to the fraction of Helium injected, the green line corresponds to the fraction of Nitrogen injected. Dashed lines correspond to the average reconstructed parameters with its  $1\sigma$  band.

Fig. 4.7 shows that the fit does not reproduce at  $1\sigma$  C.L. the injected composition. This comes from the asymmetry between the left-hand side and the right-hand side of the distribution of detected fractions. Lowering the number of detected events increases the variance of the distribution. When the distribution reaches the region  $f_A < 0$ , the fit is "stuck" at  $f_A = 0$ . Allowing for negative fractions removes this feature.

## 4.2.2 Negative fractions?

To allow for a symmetric behaviour between the left hand-side & the right hand-side of the distribution of detected fractions, negative fractions are allowed. However, to be meaningful, the model should only predict a positive number of events per  $X_{\max}$  bin (it is particularly true for the deviance which is not defined for a negative value of the parameters). Hence, the absolute value of the model is considered in eq. 4.6 & 4.10 to regularize the likelihood,  $G_{jx}^{\text{model}} \rightarrow |G_{jx}^{\text{model}}|$ . The results of the test are shown in Fig. 4.8.



**Fig. 4.8.:** Poisson reconstruction of 1000 MC simulations, with an injected composition  $f_H = 0.3$ ,  $f_{\text{He}} = 0.2$ ,  $f_N = 0.2$ ,  $f_{\text{Fe}} = 0.3$  in the energy bin  $18.3 \leq \log_{10}(E/\text{eV}) < 18.4$ , with  $N = \text{Poisson}(100)$ . Negative fractions are allowed. The grey line corresponds to the fraction of Helium injected, the green line corresponds to the fraction of Nitrogen injected. Dashed lines correspond to the average reconstructed parameters with its  $1\sigma$  band.

Figure 4.8 shows that considering negative fractions allows us to reproduce on average the injected composition, in a scenario with a few detected events. The



negative fractions remove the bias created by the condition  $f \leq 0$  (orange results in Fig. 4.8). However, the variance increases compared to considering strictly positive fractions.

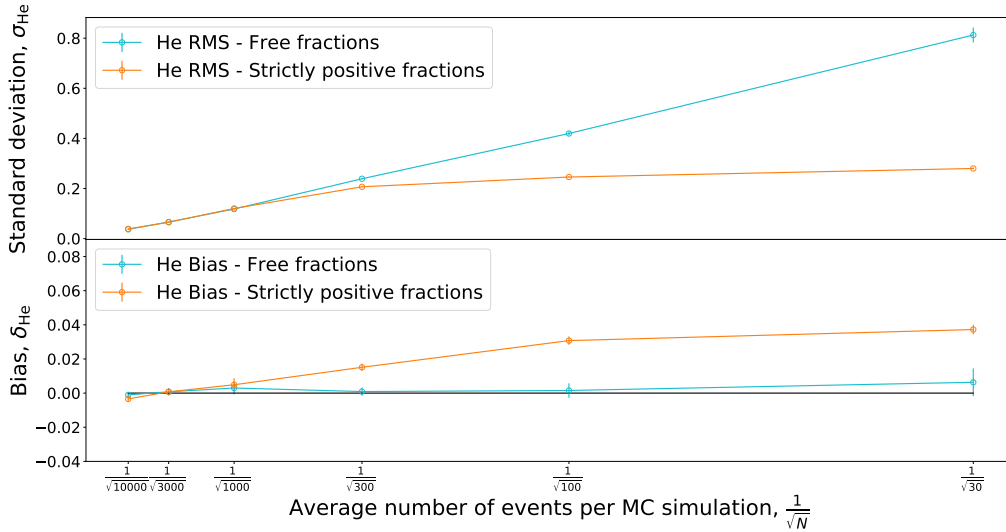
To study the impact of the number of events on the bias and the variance, a scan is performed for the same fractions, in the same energy bin over different  $N$  for strictly positive and free fractions ( $f_H = 0.3$ ,  $f_{He} = 0.2$ ,  $f_N = 0.2$ ,  $f_{Fe} = 0.3$ ,  $18.3 \leq \log_{10}(E/eV) < 18.4$ ). The bias is computed using

$$\delta_A = \langle f_A^{\text{reconstructed}} \rangle - f_A^{\text{injected}}. \quad (4.18)$$

The standard deviation is computed for the free fraction approach. For the strictly positive fractions, due to the strong asymmetry in the distribution, an effective standard deviation is computed following:

$$\sigma_A^{\text{effective}} = \frac{f_A^{\text{reconstructed}}|_{84\% \text{ quantile}} - f_A^{\text{reconstructed}}|_{16\% \text{ quantile}}}{2} \quad (4.19)$$

The (effective) standard deviations and the biases of the two methods are shown in Fig. 4.9.



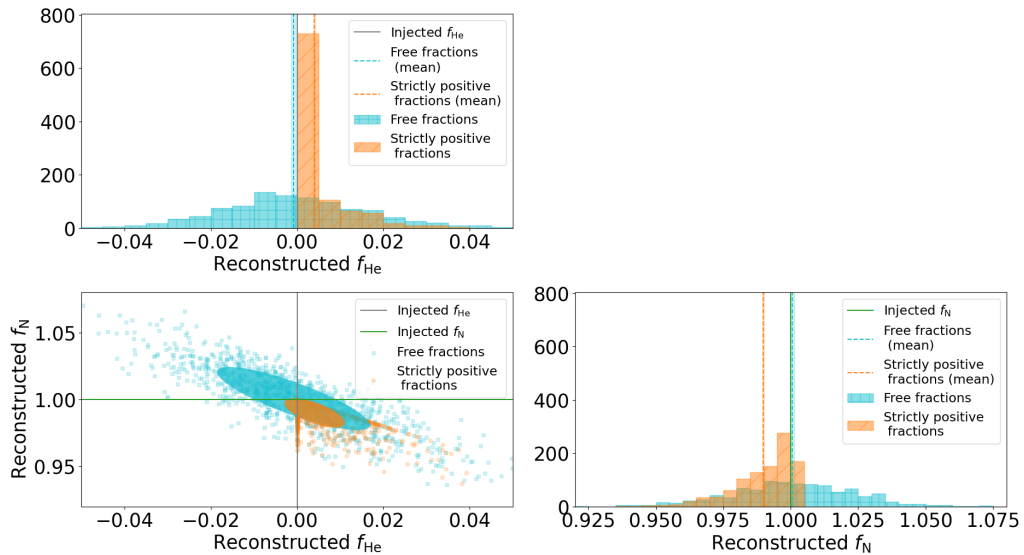
**Fig. 4.9.:** Standard deviation (top) and bias (bottom) of Helium fraction for 1000 MC simulations, with an injected composition  $f_H = 0.3$ ,  $f_{He} = 0.2$ ,  $f_N = 0.2$ ,  $f_{Fe} = 0.3$  in the energy bin  $18.3 \leq \log_{10}(E/eV) < 18.4$ , with  $N$  varying from 30 to 10,000. The lines are here to guide the eye.

Figure 4.9 shows that the free fractions allow us to remove the bias of the model when reducing the number of events detected. On the other side, the standard

deviation of both the strictly positive fractions and the free fraction approach are proportional to  $\frac{1}{\sqrt{N}}$  for  $N \geq 1000$  as expected in a Poisson distribution. However, when  $N \leq 300$ , for this set of parameters, the standard deviation is not proportional to  $\frac{1}{\sqrt{N}}$  for the strictly positive fractions approach, whereas it remains proportional to  $\frac{1}{\sqrt{N}}$  for the free fractions approach. Such a behaviour may be due to the information got from the boundaries. When the distribution of reconstructed fractions touches the boundaries of the parameter space, the fractions are stuck at the boundaries and therefore the spread of reconstructed fractions is reduced, leading to an effective standard deviation smaller than in the free fraction approach.

### 4.2.3 Test with parameters at boundaries

The next test consists in testing a benchmark scenario at boundaries. Here, 1000 MC simulations are run starting from a pure Nitrogen scenario,  $f_H = 0$ ,  $f_{He} = 0$ ,  $f_N = 1$ ,  $f_{Fe} = 0$  for each MC simulation, with  $N = \text{Poisson}(3000)$  in the energy bin  $18.8 \leq \log_{10}(E/\text{eV}) < 18.9$ . Here, even if the number of events is important, the positive-fraction approach cannot reproduce on average the injected composition (see orange in Fig 4.10).

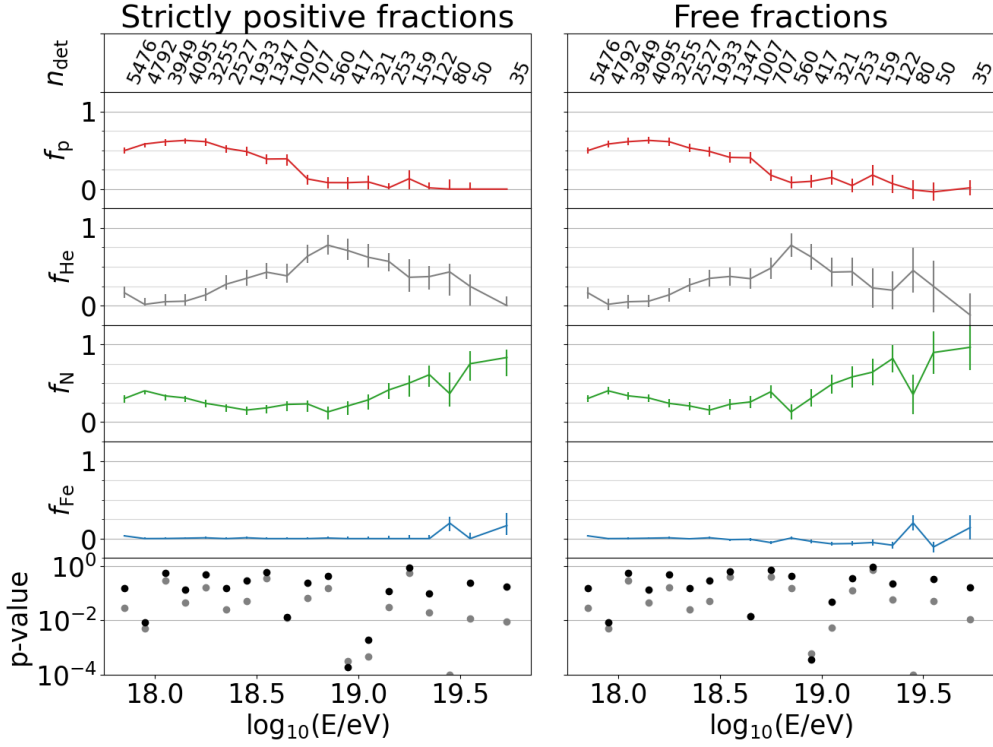


**Fig. 4.10.:** Poisson approach with 1000 MC simulations, with an injected composition  $f_H = 0$ ,  $f_{He} = 0$ ,  $f_N = 1$ ,  $f_{Fe} = 0$  in the energy bin  $18.8 \leq \log_{10}(E/\text{eV}) < 18.9$ , with  $N = \text{Poisson}(3000)$ . Negative fractions allowed.

As shown in Fig. 4.10, using negative fractions reduces the bias observed with strictly positive fractions, but results in a larger variance of the reconstructed fractions as before (see blue in Fig. 4.10).

### 4.3 Comparison between free and strictly positive fractions on data

Figure 4.11 shows the result using free fractions on the data compared to using strictly positive fractions. As before, the errors are estimated through the confidence interval obtained with the reconstruction of 1000 MC simulations based on the best-fit scenario. Applied to the data, the two approaches lead to quite similar results. Most differences appear in the last energy bins (i.e. the bins with the fewest events). As expected, the fractions of iron, which were zero in the strictly positive fractions case, oscillate around zero when using free fractions. Regarding Fig. 4.9, the errors are bigger when considering free fractions. The differences in terms of bias and resolution between the two approaches are discussed in the following pages.



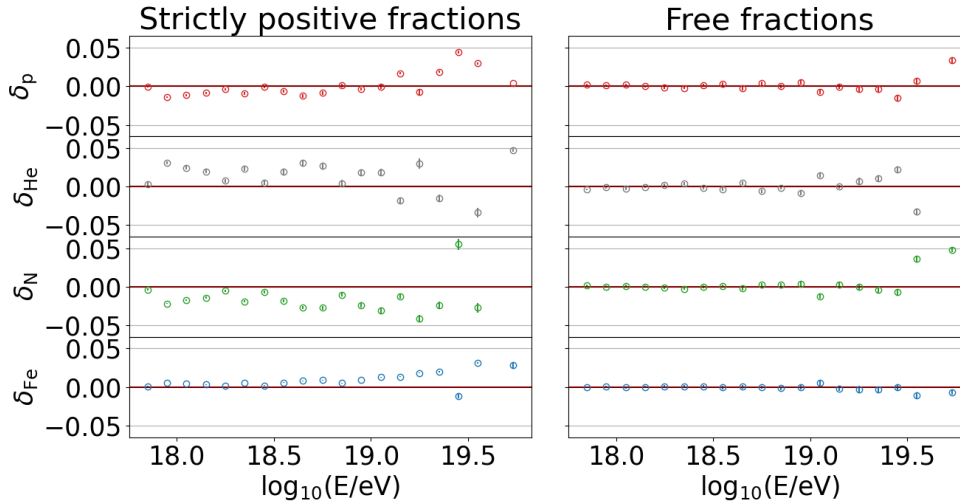
**Fig. 4.11.:** Fractions obtained from the best-fit parameters using a Poisson approach and EPOS-LHC as hadronic-interaction model for the data from ICRC 2019 (Yushkov A. on behalf of the Pierre Auger Collaboration, 2021). The last panel shows the  $p_{\text{value}}$ . The black dots correspond to the C-statistics  $p_{\text{value}}$ , and can be compared to a more classical  $\chi^2$  statistics (grey dots).

### 4.3.1 Bias vs resolution

To compare the approach employing strictly positive fractions and that allowing for negative ones, a bias estimator is computed using the mock simulations as done before in Eq. 4.18, where  $f_A^{\text{injected}}$  is the reconstructed fraction found in Fig. 4.11 for each approach. A positive  $\delta_A$  shows that on average, the parameter is overestimated, whereas a negative  $\delta_A$  shows that on average, the parameter is underestimated. The bias estimator  $\delta_A$  is shown in Fig. 4.12.

To compare the resolution obtained with the two methods, Fig. 4.13 shows the (effective) standard deviations obtained from MC simulations for each fraction and each energy bin for (strictly positive) free fractions.

Figure 4.12 shows that considering positive fractions only (left-hand side of the figure) creates a bias of a few percents in most energy bins, whereas allowing for negative fractions reproduces most of the data with a bias smaller than 1.5% for



**Fig. 4.12.:** Bias estimator for scenarios with and without negative fractions allowed.

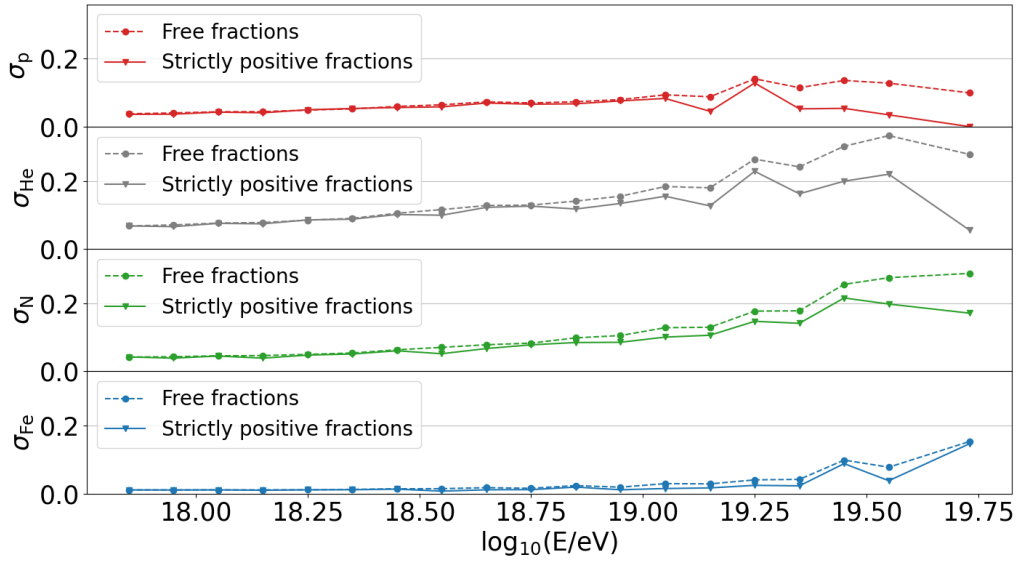
$\log_{10}(E/eV) < 19.5$  (right hand-side), i.e. for more than 50 events per bin. On the other side, Figure 4.13 shows that the resolution is at best the same for both methods, and if not, the resolution is better when only positive fractions are considered.

To sum-up the differences between the two approaches, the difference  $\Delta f_A$  between the fractions obtained with the strictly positive fraction and the free fraction approaches are shown in Fig. 4.14. The dark-coloured band shows for each energy bin the maximum bias from free and strictly positive fractions, obtained in Fig. 4.12. The clear coloured band shows the absolute value of the maximum resolution from free and strictly positive fractions, obtained in Fig. 4.13.

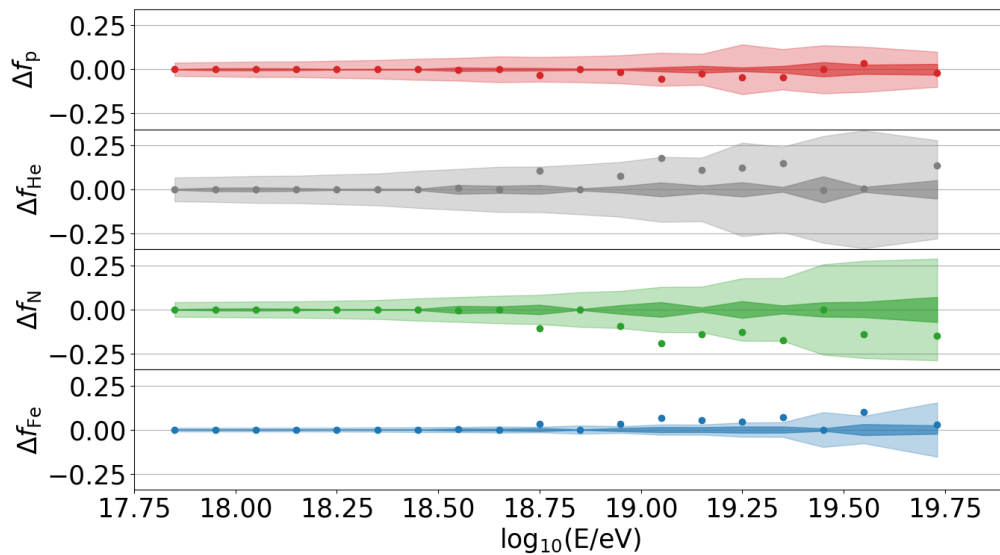
Figure 4.14 shows that the difference between the fractions reconstructed with the two methods is larger than the expected bias. Such a behaviour comes from the fact that the free fractions approach opens a new phase-space region where parameters have non-physical meaning. Therefore, the differences between the two best-fit parameters can be bigger than the expected bias.

### 4.3.2 Discussion and conclusion

In all tested situations, the Poisson & multinomial approaches give similar outputs, except for a few outliers in the multinomial approach. Figure 4.4 shows that the fit manages to reproduce, on average, the parameters injected in a situation with a high number of detected events (here 3000) and with injected elements far from the boundaries (here  $f_H = 0.3$ ,  $f_{He} = 0.2$ ,  $f_N = 0.2$ ,  $f_{Fe} = 0.3$ ). However, the fit



**Fig. 4.13.:** Standard deviation of each fraction for free (dashed lines) and strictly positive fractions (solid lines). The lines are here to guide the eye.



**Fig. 4.14.:** Differences between the fractions obtained from the strictly positive fraction approach and the free fraction approach in Fig. 4.11. The dark-coloured band shows, for each energy bin, the maximum bias from the two methods. The light coloured band shows the maximum resolution of the two methods, obtained in Fig. 4.13.

considering strictly positive fractions does not reproduce on average the simulated values when considering a small amount of events (see Fig. 4.7), or when one or more fractions are close to the boundaries (see Fig. 4.10). In fact, the closest

fractions are to zero, the more events are needed to avoid the bias coming from the asymmetries at  $f_A \geq 0$  or  $f_A \leq 1$ . To correct for this bias, the solution evaluated is to consider negative fractions. This feature allows us to reproduce, on average, the simulated true value (see Fig. 4.8, 4.9 & 4.10).

Two choices are possible. Either one considers positive fractions, which results in biased parameters, or one considers free fractions, which does not result in biased parameters, but the resolution is worse compared to positive fractions. On top of the worsen resolution, the free fractions approach can result in unphysical parameters (negative fractions has no physical meaning). Therefore, free fractions is a heavy price solution to pay to have an unbiased approach.

In the next chapter, together with the spectrum, the fitting procedure of the  $X_{\max}$  distribution is used in order to construct a cosmological model of a population of UHECR sources and to constrain their distribution as well as the injected spectrum of sources.

## Combining spectrum and composition information

*This chapter is partly adapted from the author’s work described in the Auger internal note, GAP2020 035 “Stellar mass density as evolution of sources for the Combined Fit”, Marafico, Biteau, Deligny, and Luce, 2020 and in the published paper Q. Luce, S. Marafico et al., “Observational constraints on cosmic-ray escape from UHE accelerators” 2022 (The Astrophysical Journal).*

### Introduction

In the previous chapter, the  $X_{\max}$  distributions are fitted in order to infer the mass composition of UHECRs observed on Earth. The work presented in this chapter aims at reproducing both the  $X_{\max}$  distributions and the energy spectrum, starting from a cosmic scenario where sources of UHECRs inject nuclei. The nuclei are then propagated in the universe and interact with background photons. When the particles reach Earth, predictions for the  $X_{\max}$  distributions and the energy spectrum can be made and compared to observation through likelihood procedures. From the comparison, a best-fit scenario of the sources is reconstructed. In The Pierre Auger Collaboration, 2017, such an analysis is called “combined fit”; we use this terming in the following.

The journey starts at the sources. The first step consists in considering the number of injected particles,  $n(E_g, A_g)$ , of mass  $A_g$  and of energy within  $[E_g; E_g + dE_g]$  per unit covolume and unit energy,

$$n(E_g, A_g)dE_g = \int_{t_{\min}}^{t_0} dt q_{A_g}(E_g) S(t) dE_g, \quad (5.1)$$

where  $q_{A_g}(E_g)$  is the production rate of a nucleus of mass  $A_g$  per unit covolume and per unit energy, with an energy  $E_g$  at the source. The evolution term,  $S(t)$ , describes the cumulative intensity of UHECR sources at a given cosmic time,  $t$ . The



second step consists in considering the propagation of the nuclei, that is, solving the transport equation to retrieve the expected UHECR energy spectrum,

$$J(E) \propto c \int_{z_{\min}}^{z_{\max}} dz \left| \frac{dt}{dz} \right| \sum_{A_g} q_{A_g}(E_g) S(z) \left| \frac{dE_g}{dE} \right|. \quad (5.2)$$

The term  $dE_g/dE$  describes the energy losses of injected nuclei due to adiabatic expansion and to interactions with photons from the cosmic microwave background (CMB) and from the extragalactic background light (EBL). The integration is performed over lookback time, the role of which is played by redshift using

$$\left( \frac{dt}{dz} \right)^{-1} = H_0(1+z) \sqrt{\Omega_M(1+z)^3 + \Omega_\Lambda}, \quad (5.3)$$

where  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$  is the Hubble constant,  $\Omega_M = 0.3$  is the dark-matter energy density and  $\Omega_\Lambda = 1 - \Omega_M$  is the dark-energy density in a flat  $\Lambda$ CDM cosmology. The resulting spectrum is then used for the comparisons with the measured energy spectrum and the  $X_{\max}$  distributions as a function of energy. The complete method is described below.

## 5.1 Combined Fit methods

The combined fit follows the procedure of The Pierre Auger Collaboration, [2017](#). SimProp (Aloisio et al., [2017](#)) is used to simulate the 1D propagation of nuclei from  $z = 2.5$  to  $z = 0$ . The interactions with the CMB and the EBL, described by Gilmore et al., [2012](#) for the latter, are taken into account. The photodisintegration processes are described by using Puget et al., [1976](#) (PSB) or Koning et al., [2007](#) (TALYS). Events are simulated with a uniform distribution in redshift and a weight is applied to each event so that the resulting weighted distribution follows  $S(z)$ . For practical reasons, the simulation is divided in seven shells in redshift, with boundaries at  $z = \{0, 0.01, 0.05, 0.10, 0.20, 0.50, 2.50\}$ . As in the previous chapter, representative masses are injected to keep the number of free parameters manageable. Hence, in each shell and for each representative injected mass ( $\text{H}^1$ ,  $\text{He}^4$ ,  $\text{N}^{14}$ ,  $\text{Si}^{26}$ ,  $\text{Fe}^{56}$ ), 500,000 nuclei are simulated and weighted by:

$$w(z) = S(z) \times \Delta z_i / \Delta z_0 \times \left( \frac{dt}{dz} \right)^{-1}, \quad (5.4)$$

where  $\Delta z_i$  is the width of the shell  $i$ .

Events are stored in a 4D histogram,  $N_{ijkl}$ , as a function of

$$\begin{cases} i & \Leftrightarrow E_g, \\ j & \Leftrightarrow E, \\ k & \Leftrightarrow A_g, \\ l & \Leftrightarrow A. \end{cases}$$

To model the number of UHECRs  $\mathcal{N}_{jl}$  detected on Earth with an energy  $j$  and a mass  $l$ , the production rate,  $q_{A_g}(E_g)$ , is integrated by summing over all injected energies,  $i$ , and mass,  $k$ , accounting for a spectral weight,  $\omega_{ik}$ ,

$$\mathcal{N}_{jl} = \sum_i \sum_k \omega_{ik} N_{ijkl} \Delta E_i. \quad (5.5)$$

The spectral weights are customarily defined using a broken exponential cut-off power-law to model the injection process:

$$\omega_{ik} = \frac{dN_{A_g}}{dE_g}(E_g) = f_k J_0 \left( \frac{E_i}{E_{\text{ref}}} \right)^{-\gamma} \times \begin{cases} 1 & \text{if } E_i < Z \times E_{\text{max}}^{\text{p}}, \\ \exp\left(1 - \frac{E_i}{Z \times E_{\text{max}}^{\text{p}}}\right) & \text{otherwise,} \end{cases} \quad (5.6)$$

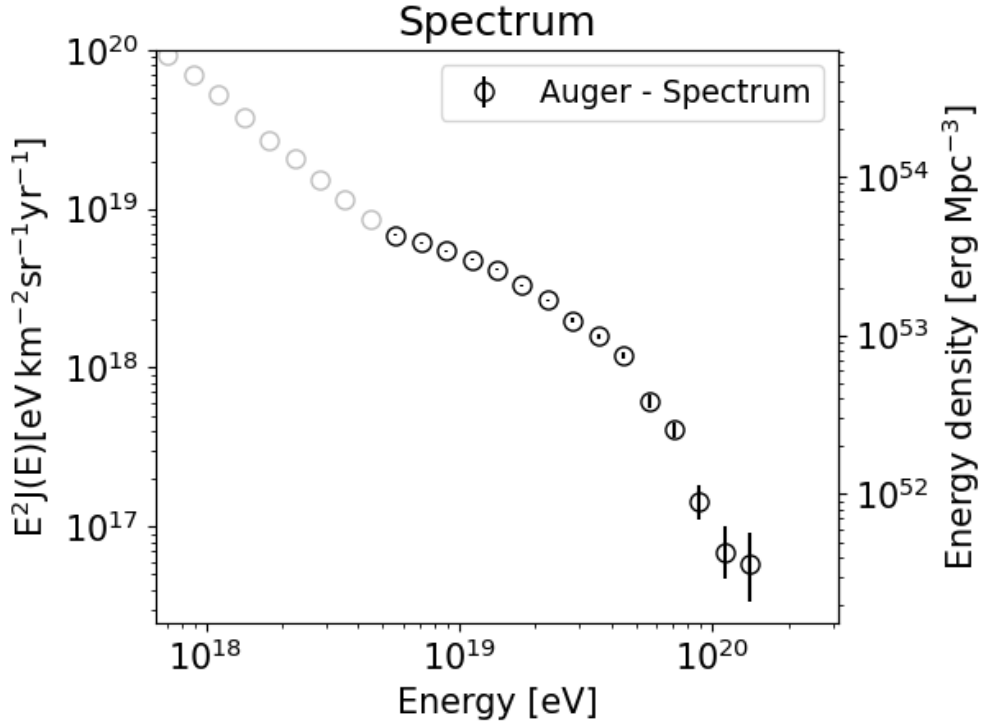
where  $f_k$  is the fraction of injected element of mass  $k$ ,  $J_0$  is a normalization,  $E_{\text{ref}}$  is an arbitrary pivot energy,  $E_{\text{max}}^{\text{p}}$  is the cutoff energy of proton, and  $\gamma$  is the spectral index at escape from the sources.

The UHECR spectrum model is constructed as  $J_j^{\text{model}} = \sum_l \mathcal{N}_{jl} / \Delta E_j$ . The spectral model can then be compared to the observed spectrum, which is shown in Fig. 5.1 as obtained at the Pierre Auger Observatory (The Pierre Auger Collaboration, 2021b). Note that what is actually shown is the energy flux instead of the flux, as it allows us to show enhanced features of the flux as well as reading directly the measured energy density,  $4\pi E^2 J(E)/c$ , (right axis of the same plot).

The energy threshold for the combined fit, namely  $E \geq 10^{18.7}$  eV, is chosen under the assumption that the extragalactic sources considered in the combined fit contribute 100% of the flux above the ankle energy. The comparison between data and expectations in the spectrum sector is carried out using the following likelihood function:

$$L_J = \prod_j \frac{1}{\sigma_j^{\text{data}} \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{J_j^{\text{data}} - J_j^{\text{model}}}{\sigma_j^{\text{data}}}\right)^2\right). \quad (5.7)$$

The minimization as a function of  $J_0$  can be performed analytically, solving for  $\partial \ln L_J / \partial J_0 = 0$ .



**Fig. 5.1.:** Energy flux measured at the Pierre Auger Observatory. Point above the ankle energy are shown in black

As for the composition sector, the fitting procedure follows from that of Chap. 3. For each detected nucleus, the  $X_{\max}$  distribution is modelled with generalized Gumbel distributions (cf. Eq. 4.2) using the parameters of Sergio Petrera, 2020, considering EPOS-LHC (Pierog et al., 2015) as hadronic interaction model as a baseline. The resolution and the acceptance of the detector are taken into account in the comparison of the generalized Gumbel distributions with the data. The probability to observe an event in a bin  $x$  of  $X_{\max}$  and in a bin of energy  $j$  is

$$G_{jx}^{\text{model}} = \frac{1}{\sum_l \mathcal{N}_{jl}} \times \sum_l \mathcal{N}_{jl} \sum_x g_{ljx} \text{Acc}_j(x) \text{Res}_j(X_{\max}^{\text{det}} - X_{\max}), \quad (5.8)$$

where  $\text{Acc}_j(x)$  is the acceptance of the detector and  $\text{Res}_j(X_{\max}^{\text{det}} - X_{\max})$  is the resolution of the detector in the  $j$ -th energy bin. Here, since all the bins are fitted simultaneously, it is not possible to have a ‘‘Poisson’’ approach as it was the case in 4.1.1. Therefore, a multinomial approach is used instead, and the overall model is normalized to the data, so that:

$$\sum_x G_{jx}^{\text{model}} = n_j^{\text{data}}, \quad (5.9)$$

where  $n_j^{\text{data}}$  is the number of observed events in an energy bin  $j$ . The likelihood associated to composition is then the product of multinomial distributions:

$$L_X = \prod_j n_j^{\text{data}}! \prod_x \frac{1}{k_{jx}^{\text{data}}!} (G_{jx}^{\text{model}})^{k_{jx}^{\text{data}}}, \quad (5.10)$$

where  $k_{jx}^{\text{data}}$  is the number of observed events in a bin  $x$ , at an energy  $j$ . As before, the comparison is carried out only for  $X_{\text{max}}$  distributions above  $E \geq 10^{18.7}$  eV.

The goodness of fit is given by the deviance  $D_{\text{tot}}$ , which is a generalized  $\chi^2$ .  $D_{\text{tot}}$  is defined as the sum of the deviance of the spectrum,  $D_J = -2 \ln L_J/L_J^{\text{sat}}$ , and of the deviance of the composition,  $D_X = -2 \ln L_X/L_X^{\text{sat}}$ , with  $L_{J/X}^{\text{sat}}$  the likelihood of the saturated model that perfectly describes the data. Therefore, the deviance of the spectrum is given by:

$$D_J = \sum_j \left( \frac{J_j^{\text{data}} - J_j^{\text{model}}}{\sigma_j^{\text{data}}} \right)^2, \quad (5.11)$$

and the deviance of the composition is given by:

$$D_X = \sum_j D_{Xj} = -2 \sum_j \sum_x k_{jx}^{\text{data}} \times (\log(G_{jx}^{\text{model}}) - \log(k_{jx}^{\text{data}})). \quad (5.12)$$

The number of degrees of freedom reported in this work accounts only for the non-empty bins in the data.

As shown in Chap. 4, the observed  $X_{\text{max}}$  distributions are not correctly reproduced by the hadronic interaction models used to model the extensive air showers. Because the hadronic interactions models do not manage to reproduce the  $X_{\text{max}}$  distributions when using  $4 \times 10 = 40$  parameters (cf. Chap. 4), the composition deviance  $D_X$  of the Combined Fit using 7 parameters only cannot be smaller than a value  $D_{\text{tot}}^{\text{best}}$  corresponding to the scenario of Chap. 4 ( $4 \times 10 = 40$  parameters).  $D_{\text{tot}}^{\text{best}}$  is estimated by summing the deviance found in each energy bin of Chap. 4. From Fig. 4.5,  $D_{\text{tot}}^{\text{best}} = 221.6$ .

### 5.1.1 Fraction and flux approach

#### Fraction approach

The combined fit used in The Pierre Auger Collaboration, 2017 is based on a fraction approach. It uses seven free parameters:

- $\gamma$ : one spectral index at escape from the sources,
- $E_{\max}^{\text{p}}$ : one proton maximum energy, which scales as  $E_{\max}^Z = Z \times E_{\max}^{\text{p}}$  for a nucleus of charge  $Z$ ,
- $f_k$ : four fractions of elements ( $\text{H}^1$ ,  $\text{He}^4$ ,  $\text{N}^{14}$ ,  $\text{Si}^{26}$ ), the last one being given by  $f_{\text{Fe}} = 1 - \sum f_k$ ,
- $J_0$ : one spectral normalization at fixed reference energy  $E_{\text{ref}}$  chosen below the cut-off  $E_{\max}^{\text{p}}$ .

## Flux approach

Using fractions  $f_k$ , with  $\sum_k f_k = 1$ , implies a strong correlation between the parameters, because the last fraction is computed as  $f_{\text{Fe}} = 1 - \sum_{k \neq \text{Fe}} f_k$ . Alternatively, we consider here five parameters  $J_k$  that are equivalent to the parameters  $f_k \times J_0$ . This allows us to get rid of the constraint  $\sum f_k = 1$ , which is a hindrance for the minimizer. Equation 5.6 is then changed into

$$\omega_{ik} = \frac{dN_{A_g}}{dE_g}(E_g) = J_k \left( \frac{E_i}{E_{\text{ref}}} \right)^{-\gamma} \times \begin{cases} 1 & \text{if } E_i < Z \times E_{\max}^{\text{p}}, \\ \exp\left(1 - \frac{E_i}{Z \times E_{\max}^{\text{p}}}\right) & \text{otherwise.} \end{cases} \quad (5.13)$$

In this approach, termed as flux approach, there are also seven free parameters:

- $\gamma$ : one spectral index at escape from the sources,
- $E_{\max}^{\text{p}}$ : one proton maximum energy, which scales as  $E_{\max}^Z = Z \times E_{\max}^{\text{p}}$  for a nucleus of charge  $Z$ ,
- $J_k$ : five injection fluxes at fixed reference energy  $E_{\text{ref}}$ .

Fractions can still be computed for comparison with other studies using  $f_k = \frac{J_k}{\sum_k J_k}$ .

## 5.1.2 Emissivities

Instead of showing the parameters in terms of  $f_k$  or  $J_k$ , which does not encode the amount of energy injected at the sources per species, we use the emissivity, which is referred to as the bolometric energy production rate. The differential energy production rate per comoving volume unit of the sources, which is directly related to their differential luminosity, is then  $\ell_j(E_g, z) = E_g^2 q_{A_g}(E_g) S(z)$ , where  $S(z)$  describes the redshift evolution of the UHECR luminosity density. The quantity  $\ell_j(E_g, z)$  is hereafter called differential energy production rate for convenience. On the other hand, the bolometric energy production rate per comoving volume unit at redshift  $z$  is obtained as:

$$\mathcal{L}_A = S(z) \int_{E_{\min}}^{\infty} E_g dE_g q_{A_g}(E_g, A_g), \quad (5.14)$$

in units of  $\text{erg Mpc}^{-3} \text{yr}^{-1}$ . Its average value in a volume spanning  $z_{\min} - z_{\max}$  is computed as

$$\bar{\mathcal{L}}_A = \frac{1}{\int_{z_{\min}}^{z_{\max}} dz \left| \frac{dt}{dz} \right|} \int_{z_{\min}}^{z_{\max}} dz \left| \frac{dt}{dz} \right| S(z) \int_{E_{\min}}^{\infty} E_g dE_g q_{A_g}(E_g, A_g). \quad (5.15)$$

The energy production rate is computed above an energy threshold, taken here at  $E_{\min} = 10^{17.8} \text{eV}$ , in order to be comparable to the work across the ankle shown at the end of this chapter (cf. 5.3).

The associated uncertainties are propagated, using :

$$\delta \bar{\mathcal{L}}_k = \sqrt{\sum_{a=1}^n \left( \frac{\partial \bar{\mathcal{L}}_k}{\partial a} \right)^2 (\delta a)^2 + 2 \sum_{a=1}^{n-1} \sum_{b=i+1}^n \frac{\partial \bar{\mathcal{L}}_k}{\partial a} \frac{\partial \bar{\mathcal{L}}_k}{\partial b} \text{Cov}(a, b)}. \quad (5.16)$$

## 5.1.3 Comparison between fraction and flux approaches

110 fits are performed using fixed  $\gamma$  and  $E_{\max}^p$  to compare the fraction approach with the flux one. The flux approach is 40% faster than the fraction one (an average of 815 flux iterations compared to 1338 fraction iterations). Furthermore, while the fraction approach yields 10% of outliers for which the combined fit does not converge, the flux approach yields no outlier.

In The Pierre Auger Collaboration, 2017,  $\gamma$  and  $E_{\max}^p$  were fixed on a grid to allow the combined fit to converge. Such an approach requires hundreds of tasks on the

Lyon computing center and between 24 and 48 hours of computing to achieve it. The flux approach makes the minimization of the deviance more robust, allowing us to let the parameter  $\gamma$  free together with all other parameters except  $E_{\text{max}}^{\text{p}}$ , over which a scan is performed. The combined fit converges in 30 min on a common computer using a single core. Hence, we use the flux approach for this analysis.

### 5.1.4 Validation of the model

A benchmark scenario is run with a flat distribution of sources, EPOS-LHC as hadronic model and PSB as photodisintegration model. This choice is motivated from the benchmark scenario of The Pierre Auger Collaboration, 2017. In order to validate the code, a comparison is done with the work of Condorelli, 2021, which uses the latest parameterization of generalized Gumbel distributions as well as the unfolded spectrum, while The Pierre Auger Collaboration, 2017 used outdated parameterization of generalized Gumbel distributions and the raw spectrum.

## Results

The results are shown in terms of fractions and emissivity in Table 5.1 using the ICRC 2015 and 2021 datasets.

	ICRC 2015 this work	ICRC 2015 Condorelli, 2021	ICRC 2021 this work
$f_{\text{H}}$ [%]	$0.0 \pm 4.7$	0.0	$66.8 \pm 6.2$
$f_{\text{He}}$ [%]	$81.2 \pm 4.5$	81.5	$32.8 \pm 1.0$
$f_{\text{N}}$ [%]	$16.4 \pm 1.4$	16.1	$0.4 \pm 0.0$
$f_{\text{Si}}$ [%]	$2.4 \pm 0.3$	2.4	$0.0 \pm 0.0$
$f_{\text{Fe}}$ [%]	$0.0 \pm 0.0$	0.0	$0.0 \pm 0.0$
$\mathcal{L}_{\text{H}}$ [ $10^{43}$ erg Mpc $^{-3}$ ]	$0.0 \pm 0.0$		$2.9 \pm 0.3$
$\mathcal{L}_{\text{He}}$ [ $10^{43}$ erg Mpc $^{-3}$ ]	$16.7 \pm 1.6$		$22.7 \pm 1.0$
$\mathcal{L}_{\text{N}}$ [ $10^{43}$ erg Mpc $^{-3}$ ]	$21.4 \pm 1.3$		$18.5 \pm 0.9$
$\mathcal{L}_{\text{Si}}$ [ $10^{43}$ erg Mpc $^{-3}$ ]	$8.1 \pm 1.2$		$6.0 \pm 1.7$
$\mathcal{L}_{\text{Fe}}$ [ $10^{43}$ erg Mpc $^{-3}$ ]	$0.0 \pm 0.0$		$1.8 \pm 1.9$
$D_{\text{tot}}/\text{ndf}$	193.1/118	195.7/118	309.5/129
$D_{\text{J}}/N_{\text{J}}$	18.4/15	21.1/15	27.6/15
$(D_{\text{X}})/N_{\text{X}}$	174.7/110	174.6/110	281.9/121
$\log_{10} E_{\text{max}}^{\text{p}}/\text{eV}$	$18.59 \pm 0.01$	$18.59 \pm 0.03$	$18.24 \pm 0.01$
$\gamma$	$0.69 \pm 0.02$	$0.67 \pm 0.12$	$-1.17 \pm 0.01$

**Tab. 5.1.:** Comparison between this work and that of Condorelli, 2021 using ICRC 2015 and 2021 datasets.

The first two columns of 5.1 show results compatible within  $1\sigma$  C.L. From 2015 to 2021, the total deviance increased by  $\approx 115$  units. Therefore, the tension between the model and the data increases with larger statistics. A significant contribution comes from the new way of binning the highest detected energies (one bin used for  $19.5 < \log_{10} E/\text{eV} < 20.0$  in 2015, contributing to the deviance with  $\sim 5$  units; two bins used in 2021, contributing to the deviance with  $\sim 30$  units).

It is interesting to note that while the fraction of nitrogen seems negligible when using the ICRC 2021 dataset ( $f_N = 0.4\%$ ), its emissivity is not ( $\tilde{\mathcal{L}}_N = 18.5 \times 10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}$ ). Such a behaviour is due to the fact that fractions represent the proportion of injected elements at a low reference energy. The resulting fractions do not scale with the amount of injected energy. This is why the results are shown in terms of emissivity rather than fractions in this thesis.

### 5.1.5 Impact of the photodisintegration model

The results presented above make use of PSB (Puget et al., 1976; Stecker & Salamon, 1999) as photodisintegration model. PSB is a simple model that takes into account one- and two-nucleon channels together with multiple nucleon channels. Other kinds of processes are neglected. On the other side, TALYS (Koning et al., 2007) is a much more sophisticated and complete model that describes, in addition to the previous channels, channels of deuterons, tritium,  $^3\text{He}$ , and  $^4\text{He}$  nuclei. To fully describe the propagation and take into account He channels, the fit is run with TALYS and the results are shown in Tab. 5.2.

	PSB	TALYS
$\tilde{\mathcal{L}}_H [10^{43} \text{ erg Mpc}^{-3}]$	$2.9 \pm 0.3$	$2.0 \pm 0.4$
$\tilde{\mathcal{L}}_{\text{He}} [10^{43} \text{ erg Mpc}^{-3}]$	$22.7 \pm 1.0$	$12.8 \pm 1.3$
$\tilde{\mathcal{L}}_N [10^{43} \text{ erg Mpc}^{-3}]$	$18.5 \pm 0.9$	$29.8 \pm 0.8$
$\tilde{\mathcal{L}}_{\text{Si}} [10^{43} \text{ erg Mpc}^{-3}]$	$6.0 \pm 1.7$	$7.2 \pm 1.3$
$\tilde{\mathcal{L}}_{\text{Fe}} [10^{43} \text{ erg Mpc}^{-3}]$	$1.8 \pm 1.9$	$1.4 \pm 1.4$
$D_{\text{tot}}/\text{ndf}$	309.5/129	292.9/129
$D_J/N_J$	27.6/15	27.1/15
$(D_X)/N_X$	281.9/121	268.8/121
$\log_{10} E_{\text{max}}^{\text{p}}/\text{eV}$	$18.24 \pm 0.01$	$18.27 \pm 0.01$
$\gamma$	$-1.17 \pm 0.01$	$-0.97 \pm 0.1$

**Tab. 5.2.:** Comparison between PSB and TALYS as photodisintegration models.

Overall, the TALYS photodisintegration model improves the deviance by  $\approx 20$  units. The best-fit parameters using TALYS are close to the ones from PSB. However, the



emissivity of Helium is twice smaller with TALYS than with PSB. This is due to the presence of He nuclei in the channels of TALYS. Therefore, He nuclei are produced during the travel of heavier nuclei while it is not the case with PSB. Hence, the fit increases the emissivity of Nitrogen to compensate the non-creation of He during the journey of nuclei when using TALYS.

### 5.1.6 Derivation of likelihood for the spectrum data

The likelihood function used in Eq. 5.7 does not account for the bin-to-bin correlations inherited from the unfolding procedure used to infer the energy spectrum at the Pierre Auger Observatory, which corrects for the probability that an event with an energy  $E$  corresponding to a bin  $j$  has been reconstructed in a bin  $k$  with  $k \neq j$ . In order to account for the associated correlation matrix, the model is changed to predict the expected number of events in each energy bin as  $\mu_j^{\text{exp}}(\mathbf{s}) = \sum_l \mathcal{N}_{jl}(\mathbf{s})$ . The set of free parameters, denoted by  $\mathbf{s}$ , can then be fitted to match the observed  $\mu_j$  quantities (“unfolded number of events”), obtained by applying the invert migration matrix  $R$  provided in The Pierre Auger Collaboration, 2020d to the observed number of events. The fit is carried out by means of the likelihood function

$$L_J(\mathbf{s}) = \frac{1}{(2\pi)^{N_{\text{bin}}/2} |M(\mathbf{s})|^{1/2}} \exp \left( -\frac{1}{2} \sum_{j_1, j_2}^{N_{\text{bin}}} (\mu_{j_1} - \mu_{j_1}^{\text{exp}}(\mathbf{s})) M_{j_1 j_2}^{-1}(\mathbf{s}) (\mu_{j_2} - \mu_{j_2}^{\text{exp}}(\mathbf{s})) \right), \quad (5.17)$$

with  $M_{ij}(\mathbf{s}) = R_{ik}^{-1} V_{kl}(\mathbf{s}) (R_{lj}^{-1})^T$  the covariance matrix in the unfolded space, obtained by applying the standard transformation rules to the covariance matrix in the smeared space, namely  $V_{kl}(\mathbf{s}) = R_{k\ell} \mu_{\ell}^{\text{exp}}(\mathbf{s}) \delta_{k\ell}$ . The matrix  $M$  takes into account the bin-to-bin correlations arising from the unfolding procedure. Note that, in the absence, currently, of a published combination procedure of the inclined and vertical spectrum, only the so-called vertical spectrum is used in this work. The study is pursued by using the vertical spectrum and the spectrum likelihood defined in eq. 5.17.

### 5.1.7 Benchmark scenario

A benchmark scenario is run, using:

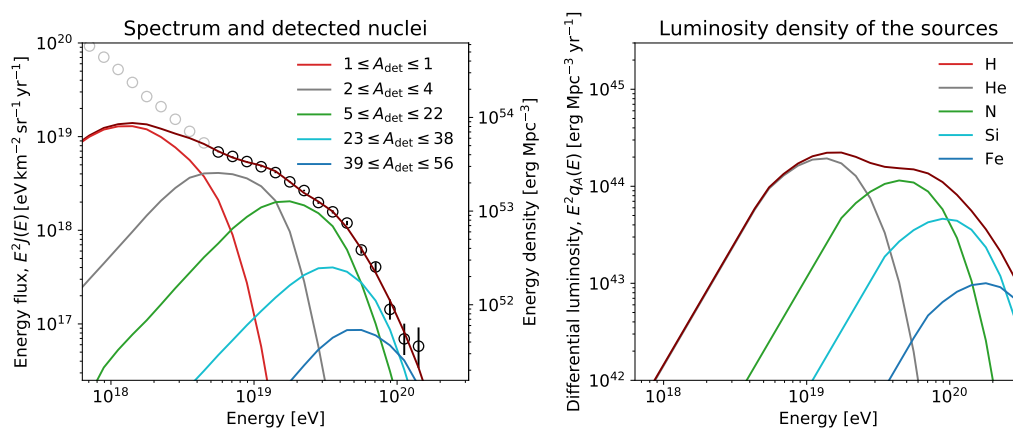
- ICRC 2021 dataset (The Pierre Auger Collaboration, 2021b; Yushkov A. on behalf of the Pierre Auger Collaboration, 2021);
- TALYS photodisintegration model (Koning et al., 2007);
- EPOS-LHC as hadronic interaction model (Pierog et al., 2015);
- Gilmore as EBL model (Gilmore et al., 2012);
- $S(z) = 1$ , a flat evolution of sources;
- Spectrum likelihood with correlation matrix.

The best fit parameters are shown in Tab. 5.3.

	Benchmark scenario
$\mathcal{L}_H [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$2.0 \pm 0.4$
$\mathcal{L}_{\text{He}} [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$13.7 \pm 1.4$
$\mathcal{L}_N [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$30.0 \pm 0.4$
$\mathcal{L}_{\text{Si}} [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$6.8 \pm 1.2$
$\mathcal{L}_{\text{Fe}} [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$1.5 \pm 1.0$
$D_{\text{tot}}/\text{ndf}$	288.6/129
$D_J/N_J$	25.1/15
$(D_X)/N_X$	263.5/121
$\log_{10} E_{\text{max}}^{\text{p}}/\text{eV}$	$18.25 \pm 0.01$
$\gamma$	$-1.18 \pm 0.01$

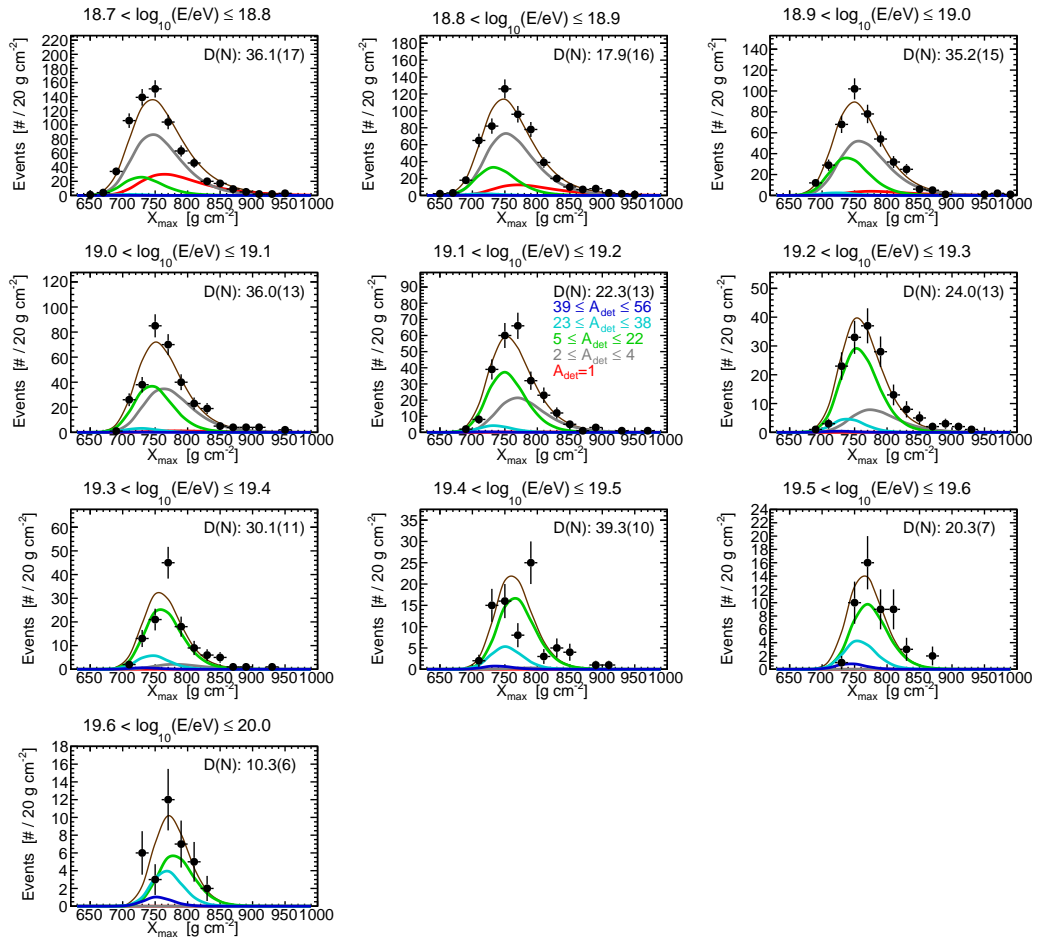
**Tab. 5.3.:** Benchmark scenario using a flat evolution of sources.

The results shown in Tab. 5.3 are very close to the last column of Tab. 5.2. The only difference in the inputs is the use of the correlation matrix in the likelihood. Fig. 5.2 shows the observed spectrum and the best-fit scenario of the model. Since the representative nuclei can be photodisintegrated during the journey, all kinds of nuclei can be detected on Earth. Therefore, to show the contribution of different types of nuclei, the flux is split into groups: proton  $A = 1$  (red), Helium-like  $2 \leq A \leq 4$  (grey), CNO-like  $5 \leq A \leq 22$  (green), Silicon-like  $23 \leq A \leq 38$  (cyan), Iron-like  $39 \leq A \leq 56$  (blue). On the right hand-side of Fig. 5.2, the (differential) luminosity density of the sources,  $E_g^2 q_{A_g}(E_g)$ , is shown as a function of energy. Note that this is an astrophysical quantity of interest that can be compared to that of non-thermal astrophysical sites, *e.g.* in the radio and/or gamma-ray bands.



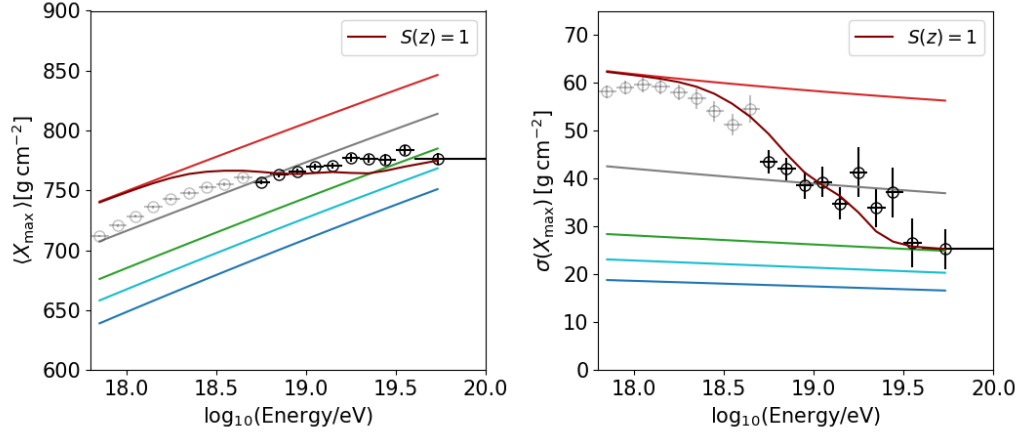
**Fig. 5.2.:** *Left:* Detected spectrum with detected nuclei. Colors represent the detected nuclei, red for  $A = 1$ , grey for  $2 \leq A \leq 4$ , green for  $5 \leq A \leq 22$ , cyan  $23 \leq A \leq 38$ , blue for  $39 \leq A \leq 56$ , the sum is shown in brown. *Right:* Luminosity density of the sources as a function of energy, red for H, grey for He, green for N, cyan Si, blue for Fe, the sum is shown in brown.

The  $X_{\max}$  distributions are shown in Fig. 5.3. The nuclei are arranged in the same groups as before. The deviance and the number of points associated to each energy bin are shown in top-right corners of each graph.



**Fig. 5.3.:**  $X_{\max}$  distributions with best-fit model. The model is shown in brown.

An economical way of seeing how the model manages to reproduce the  $X_{\max}$  distributions, without showing the full distributions, is to compare the two first  $X_{\max}$  moments from the data and from the model. Fig 5.4 does so, by showing the measured  $X_{\max}$  moments unfolded from acceptance and resolution effects (Yushkov A. on behalf of the Pierre Auger Collaboration, 2021) compared to the moments computed from the best-fit scenario using non-convolved Gumbel generalized function (before taking acceptance and resolution of the detector). Although the two first moments are interesting to describe the  $X_{\max}$  distributions, one should keep in mind that two different mass compositions can result in the same average and standard deviation values (The Pierre Auger Collaboration, 2014b).



**Fig. 5.4.:**  $X_{\max}$  distribution moments for each energy bin. The brown curve represents the best-fit scenario. Other curves represent the expected values for a pure composition, red for H, grey for He, green for N, cyan for Si, blue for Fe.

The largest tension between the model and the data appears to be driven by the evolution of the average maximum of shower depth,  $\langle X_{\max} \rangle$ , with energy.

## Discussion

The value of  $E_{\max}^{\text{p}}$  which determines the cut-off of nuclear species at  $E_{\max}^{\text{Z}} = Z \times E_{\max}^{\text{p}}$ , implies that the suppression of the spectrum is due to the combination of the cut-off energy at the sources for the heavier nuclei and the propagation effects (GZK limit). The spectral index,  $\gamma$ , is in turn determined by the increase of the average mass with energy (cf. Chap 4), which is almost monoelemental, to reproduce the  $X_{\max}$  distributions as well as possible. The solution provided by the fit therefore consists in imposing a hard index for nuclei so that the contribution of each element mixes as little as possible: high-energy suppression imposed by the cut-off beyond  $E_{\max}^{\text{Z}}$  and low-energy suppression via the hard index  $\gamma$ . Regarding the deviance, one can see that the spectrum is roughly reproduced by the model,  $D_J = 25.1$  for  $N_J = 15$  points for the spectrum. On the other hand, the  $X_{\max}$  distribution is less well reproduced,  $D_X = 263.5$  for  $N_J = 121$  points when considering all  $X_{\max}$  distributions. The large value of the deviance on  $X_{\max}$  reflects the difficulty for hadronic-interaction generators to reproduce the data, especially the average  $X_{\max}$  as shown in Fig. 5.4.

## 5.2 Evolution term $S(z)$

Instead of a slow source evolution on Gpc scales based on simplified analytical terms such as  $S(z) = (1 + z)^m$ , e.g. with  $m = \{-3; 0; 3\}$ , widely used in the literature, we test hereafter two alternative evolution terms that account for the over-density of matter surrounding the Milky Way:

- Stellar Mass Density (SMD),  $\rho_*(z)$  [ $M_\odot \text{ Mpc}^{-3}$ ]
- Star Formation Rate Density (SFRD),  $\dot{\rho}_*(z)$  [ $M_\odot \text{ Mpc}^{-3} \text{ yr}^{-1}$ ]

The two evolutions trace different astrophysical objects and scenarios. For instance, SMD is a good tracer of intermediate luminosity AGN, while SFRD is a good tracer of long gamma ray bursts. More details about the differences between the two terms will be shown in Chap. 6.

Accounting for the over-density of matter surrounding the Milky Way is motivated by the  $4\sigma$  preference for the “starburst” model over isotropy at energies larger than 40 EeV (cf. 3.3.1, The Pierre Auger Collaboration, 2018a; The Pierre Auger collaboration, 2022). The brightest objects in this sky model are located in the Council of Giants at  $\approx 4 \text{ Mpc}$  (McCall, 2014), as for instance Centaurus A. We aim here at testing to which extent such an over-density of sources could be in line with the observed UHECR spectrum and composition.

### 5.2.1 Stellar mass density from Mpc to Gpc scales

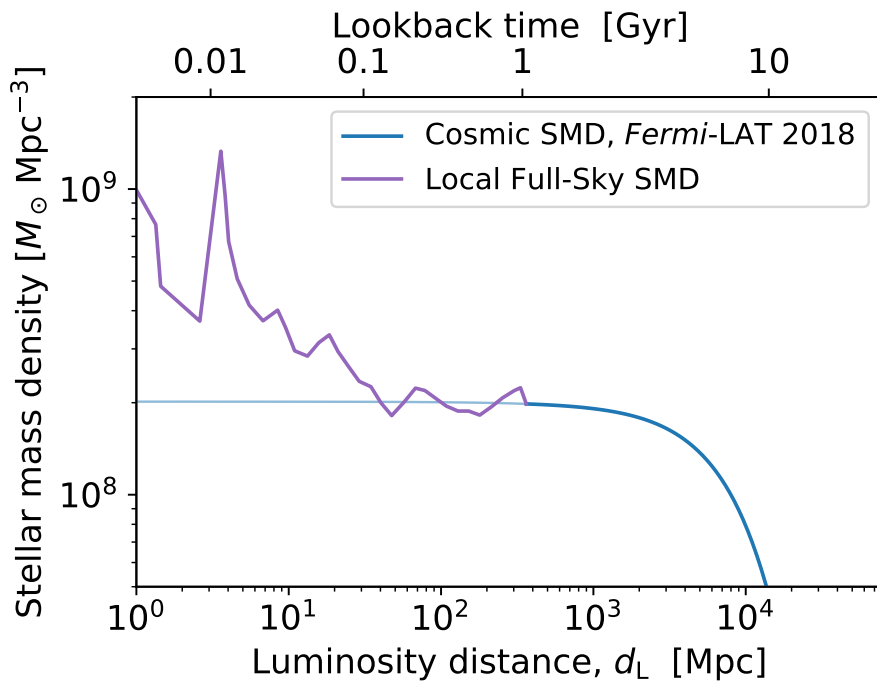
The cosmic evolution of the SMD,  $\rho_*^{\text{cosmo}}$ , can be computed by integrating over time the Cosmic Star Formation History (CSFH). CSFH is the cosmic evolution of the star formation rate as a function of redshift. Here, CSFH is based on the parameterization of Koushan et al., 2021 of the latest Fermi-LAT results. In this way,  $\rho_*^{\text{cosmo}}$  is computed as:

$$\rho_*^{\text{cosmo}}(z) = (1 - R) \int_z^\infty \text{CSFH}(z') \left| \frac{dt}{dz'} \right| dz', \quad (5.18)$$

where  $R = 0.27$  is the “return fraction”, which corresponds to the proportion of the mass of each generation of stars that is put back into the interstellar and intergalactic medium. However, as shown in Madau and Dickinson, 2014, computing SMD from CSFH results in an overestimation of the SMD. Therefore,  $\rho_*^{\text{cosmo}}(z)$  is re-normalized in order to have its average result between  $z = 0.02$  and  $z = 0.08$  equal to measured SMD from Driver et al., 2018 in the same redshift range. On smaller scales, the

density of stellar matter,  $\rho_*^{\text{local}}(z)$ , has been characterized within 350 Mpc based on near-infrared surveys (Biteau, 2021).

Figure 5.5 shows the evolution of sources in a SMD scenario (purple for the local SMD and blue for the cosmic one) ranging from luminosity distances of 1 Mpc up to 100 Gpc. The sharp feature at 1 Mpc corresponds to the influence of the Local Group, dominated in mass by Andromeda at  $\approx 750$  kpc. The peak at  $D = 4$  Mpc corresponds to the Council of Giants. The second peak at  $D \approx 15$  Mpc corresponds to the Virgo cluster (Shaya et al., 2017). Since the weights are important at small distances, a sufficiently large number of events needs to be simulated in the first redshift shell to avoid fluctuations in the reconstructed quantities. Thus, 5,000,000 events are simulated for each injected mass within  $0.00 \leq z \leq 0.01$ , instead of 500,000 for more distant shells.

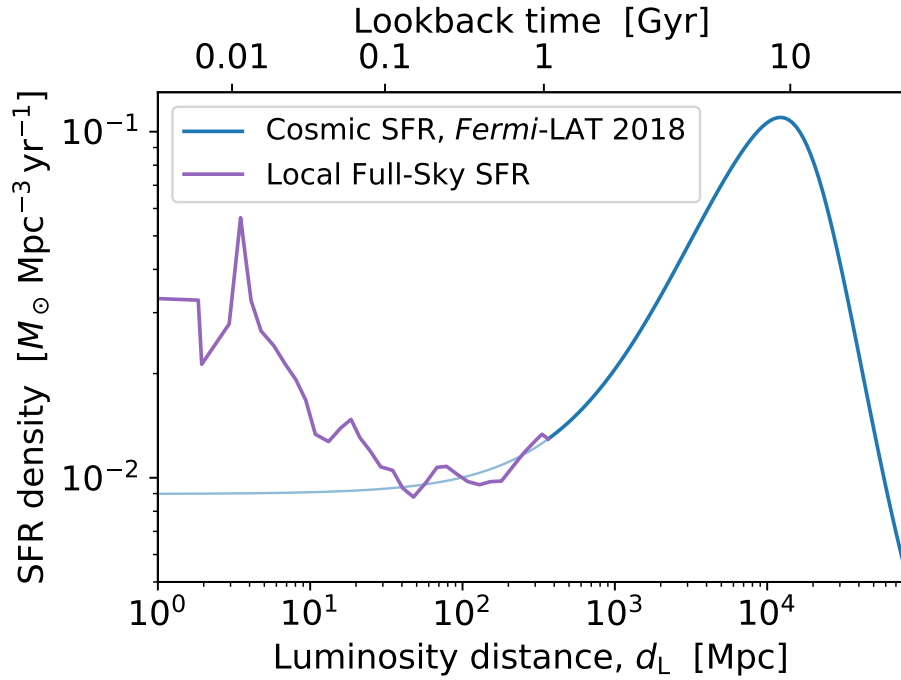


**Fig. 5.5.:** Stellar mass density as a function of luminosity distance. The local measured SMD is shown in purple, while the cosmic SMD is shown in blue.

A minimum distance,  $D_{\text{min}}$ , beyond which nuclei are injected, is introduced. Setting  $D_{\text{min}} = 0$  would correspond to injection within the Milky Way, with a diverging stellar mass density. Even for  $D_{\text{min}} = 50$  kpc, one could expect the UHECR sky to be dominated by the Magellanic clouds if the UHECR flux was simply traced by stellar mass. Therefore,  $D_{\text{min}} = 1$  Mpc is chosen. In this way, the Local Group is excluded. A more detail study of the sources below 1 Mpc is done in Chap. 6.

## 5.2.2 Star Formation Rate Density

As for the SMD, the local SFRD has been characterized in the first 350 Mpc in Biteau, 2021 from  $H_\alpha$ , near-infrared and morphology measurements in local galaxies. Together with the CSFH used to compute SMD, the SFR density is reconstructed and shown in Fig. 5.6. As before, the distance of injection is taken to be  $D_{\min} = 1$  Mpc to exclude the Local Group and avoid the divergence of the SFR density at small distances due to the Local Group and the Milky Way.



**Fig. 5.6.:** Star formation rate density as a function of luminosity distance. The local measured SFRD is shown in purple, while the CSFH is shown in blue.

As before, features appear at  $D = 4$  Mpc and  $D \approx 15$  Mpc, which correspond to the Council of Giants and Virgo cluster respectively. These peaks appear for both the SMD and the SFR models because they are quite correlated: the more there are galaxies, the larger stellar mass and SFR.

## 5.2.3 Results

Tab. 5.4 shows the combined fit results for a scenario with a flat source evolution,  $S(z) = 1$ , the SMD scenario, and the SFRD scenario, with  $D_{\min} = 1$  Mpc. The deviance improves by  $\sim 15$  units when considering the SMD and  $\sim 7$  units when considering the SFRD.

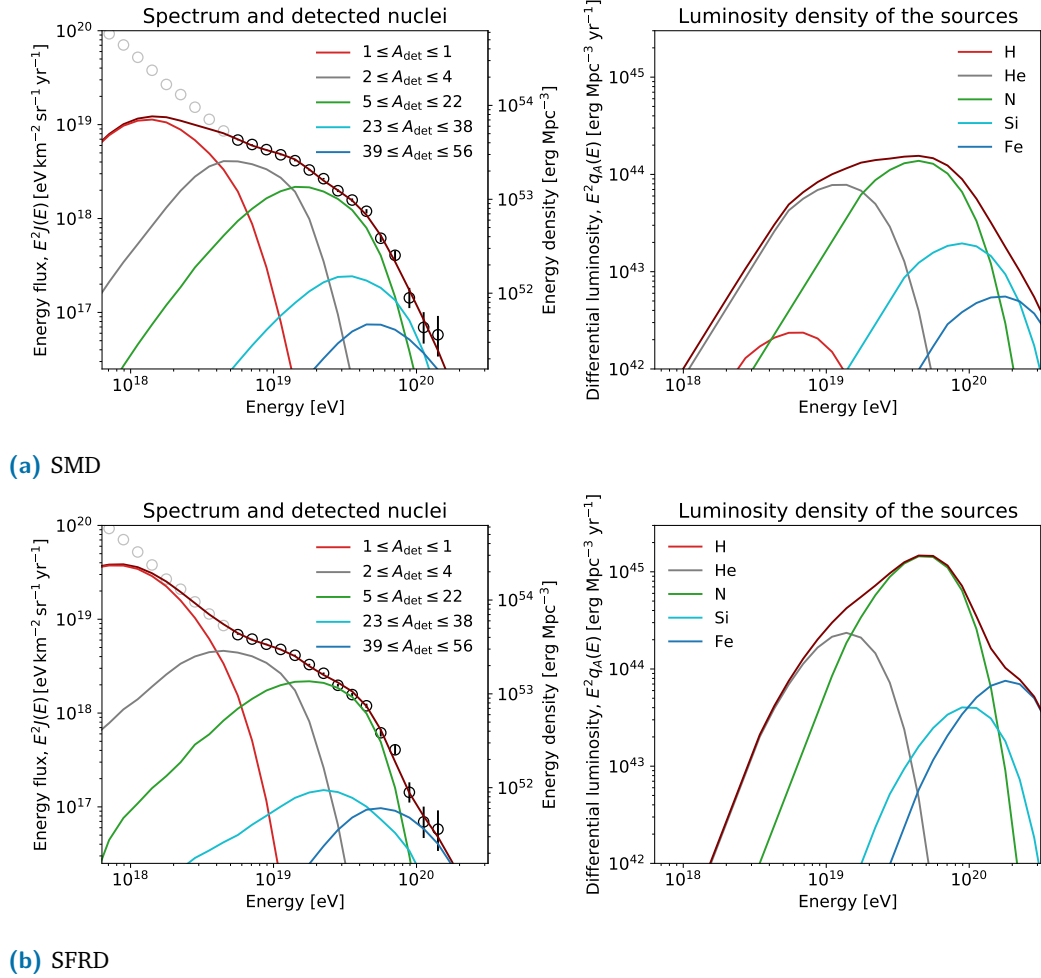


	$S(z) = 1$ PSB	$S(z) = \rho_*(z)$ $D_{\min} = 1 \text{ Mpc}$ SMD	$S(z) = \dot{\rho}_*(z)$ $D_{\min} = 1 \text{ Mpc}$ SFRD
$\tilde{\mathcal{L}}_{\text{H}} [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$2.0 \pm 0.4$	$0.4 \pm 0.8$	$0.0 \pm 0.0$
$\tilde{\mathcal{L}}_{\text{He}} [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$13.7 \pm 1.4$	$12.7 \pm 1.4$	$30.6 \pm 1.5$
$\tilde{\mathcal{L}}_{\text{N}} [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$30.0 \pm 0.4$	$22.1 \pm 2.0$	$191.0 \pm 2.3$
$\tilde{\mathcal{L}}_{\text{Si}} [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$6.8 \pm 1.2$	$3.1 \pm 1.6$	$5.3 \pm 11.3$
$\tilde{\mathcal{L}}_{\text{Fe}} [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$1.5 \pm 1.0$	$0.9 \pm 0.6$	$9.8 \pm 0.9$
$D_{\text{tot}}/\text{ndf}$	288.6/129	274.5/129	281.5/129
$D_J/N_J$	25.1/15	25.7/15	37.0/15
$(D_X)/N_X$	263.5/121	248.7/121	244.5/121
$\log_{10} E_{\text{max}}^{\text{P}}/\text{eV}$	$18.25 \pm 0.01$	$18.37 \pm 0.05$	$18.21 \pm 0.01$
$\gamma$	$-1.18 \pm 0.01$	$-0.39 \pm 0.01$	$-1.85 \pm 0.01$

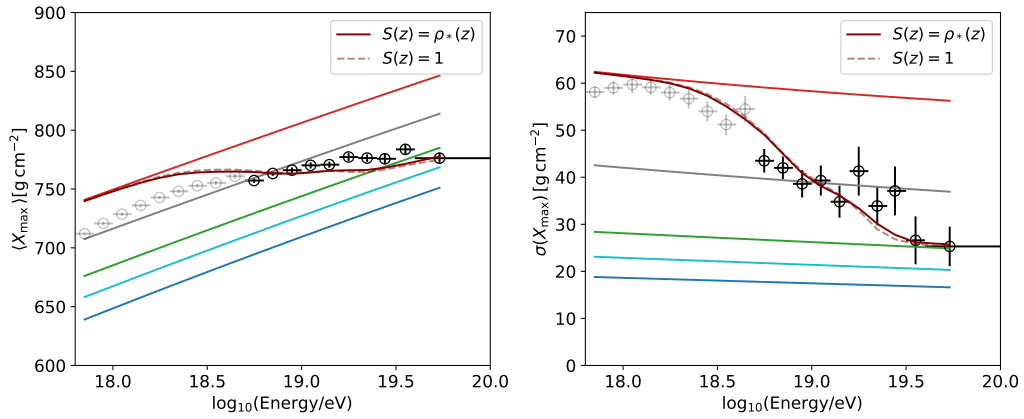
**Tab. 5.4.:** Results obtained with a luminosity evolution term that follows the SMD or SFRD (middle SMD, right SFRD) compared to those obtained with a flat distribution of sources (left). The ICRC 2021 dataset is used.

The detected spectrum and luminosity density of the sources for the two evolutions of sources are shown in Fig. 5.7. On the other side, Fig. 5.8 shows the two first  $X_{\text{max}}$  moments for the two evolution of sources. The SMD evolution results in a softer spectrum with  $\gamma \sim -0.4$ ; the SFRD evolution of sources results in a very hard spectrum  $\gamma \sim -1.8$ . Compared to the SMD scenario, the SFRD one results in much more secondary protons produced at low energy. Regarding the luminosities, the SFRD scenario is dominated by CNO and the total emissivity is an order of magnitude above the one expected considering the SMD scenario.

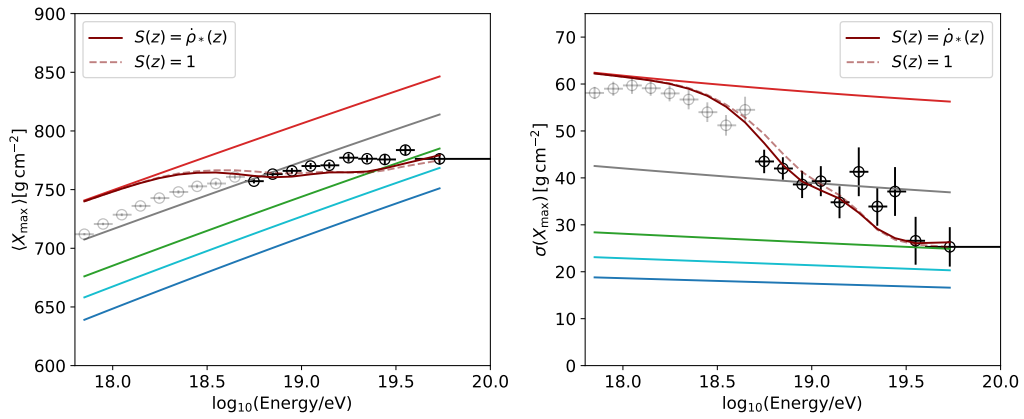
Fig. 5.8 shows that the average  $X_{\text{max}}$  is not well reproduced by the two evolution of sources, as it is the case when considering a flat scenario.



**Fig. 5.7.:** *Left:* Observed spectrum with detected nuclei. Colors represent the detected nuclei, red for  $A = 1$ , grey for  $2 \leq A \leq 4$ , green for  $5 \leq A \leq 22$ , cyan  $23 \leq A \leq 38$ , blue for  $39 \leq A \leq 56$ , the sum is shown in brown. *Right:* Luminosity density of the sources as a function of energy, red for H, grey for He, green for N, cyan Si, blue for Fe, the sum is shown in brown.



(a) SMD



(b) SFRD

**Fig. 5.8.:**  $X_{\max}$  moments for each energy bin. The brown curve represents the best-fit scenario. The dashed brown curve represents the best-fit configuration for a flat distribution of sources. Coloured lines represent the expected values for a pure composition, red for H, grey for He, green for N, cyan for Si, blue for Fe.

## 5.2.4 Discussion

With increased statistics, there are more and more pieces of evidence that the widely-used flat distribution of sources is not an optimal input for the combined fit to reproduce the spectrum and composition data ( $D_{\text{tot}} = 288.6$ ). In order to better reproduce these data, two models of source evolution have been tested, SMD and SFRD. Considering the SMD as the evolution term in the combined fit leads to an improved deviance of  $\approx 15$  units, while considering SFRD as the evolution term leads to an improved deviance of  $\approx 7$  units. These results are obtained by considering local SMD/SFRD to describe the first 350 Mpc with a minimal distance of injection of 1 Mpc. This minimal distance of injection excludes the Local Group as a host of UHECR accelerators. The total emissivity, which estimates the energy budget needed to reproduce the spectrum, is smaller for the SMD ( $\bar{\mathcal{L}}_{\text{tot}} = (3.9 \pm 0.3) \times 10^{44}$ ) than for a flat-source scenario ( $\bar{\mathcal{L}}_{\text{tot}} = (5.4 \pm 0.4) \times 10^{44}$ ). This is due to the excess of stellar mass in the first 100 Mpc and the deficit of sources beyond  $\sim 10$  Gpc. Due to a distribution of sources that decreases with distance, the impact of the propagation is reduced. Therefore, the nuclei mix less and the injected spectrum can be softer ( $\gamma \sim -0.4$ ) than that inferred with the flat distribution ( $\gamma \sim -1.2$ ). To have a similar cut-off in energy with a softer spectral index,  $\log_{10} E_{\text{max}}^p$  is increased (from 18.25 to 18.37). On the other side, the SFRD, with a similar deviance, results in a very different scenario at the sources. Many secondary protons are expected at low energy; this is expected from the propagation and the SFRD that peaks at 10 Gpc. In the SFRD scenario, the ankle feature is created by secondary protons produced during the journey of UHECRs. In the SMD case, the proton flux is not big enough to reproduce the ankle. Therefore, the origin of the ankle feature can more easily be created by introducing a second type of extragalactic sources with a significantly steeper spectral index (Aloisio et al., 2014; Guido, E. on behalf of the Pierre Auger Collaboration, 2021; Mollerach & Roulet, 2020). In order to keep the  $X_{\text{max}}$  distribution as monoelemental as possible with a larger source density at 10 Gpc in the SFRD scenario, the injection spectrum is steeper than that with a flat evolution of sources ( $\gamma \sim -1.8$ ). Such an injection spectrum is not expected by most of the theoretical approaches. The second counterpart of having a distribution of sources peaking around 10 Gpc is the increased emissivity. The total emissivity is increased by a factor  $\sim 6.0$  compared to that of a flat distribution of sources. About 80% of the emissivity comes from CNO nuclei, the remaining emissivity comes from helium. It means that in the SFRD scenario, the production of secondaries *en route* plays a major role to reproduce the  $X_{\text{max}}$  distributions and to reproduce the ankle feature. A scenario where the evolution of sources follow the cosmic behaviour of SMD and SFRD has also been tested (without account for the local environment).

The cosmic SMD evolution of sources results in parameters and deviance very close to the flat evolution of source, which is not surprising, regarding the flatness of the cosmic SMD. Using the cosmic SFRD brings similar results as using SFRD with the local environment. The differences lie in the emissivities with  $\sim -25\%$  of CNO and  $\sim +50\%$  of helium comparing cosmic with cosmic+local SFRD, together with a total deviance of 273 which is better than the cosmic+local one. The fact that a “deficit” of sources nearby compared to the cosmic+local SFRD scenario, reproduces better the data could be a sign that sources of UHECRs are transient. This hypothesis is studied in the last chapter of this thesis.

## 5.3 Across the ankle

The analysis presented above starts at the ankle energy. This choice is motivated by the widely-used assumption that UHECRs with energies above the ankle feature are of extra-galactic origin, an assumption that has been strengthened by the large-scale anisotropy findings (cf. 3.2). Below the ankle, an additional component of unknown origin, the component B (Hillas, 2005), is believed to be responsible for making up the all-particle energy spectrum up to the ankle feature. If the component B behaves as a Peters’ cycle (cf. 1.2.1), one can expect the component ending with heavy nuclei; therefore no proton from the component B would be expected below the ankle. Hence, the protons observed below the ankle should be produced by the “above-the-ankle” component. Interestingly, as shown in Fig. 5.7, the two approaches tested before result in a significant number of protons expected below the ankle energy. The idea, then, is to compare the protons expected from the combined fit to the proton spectrum below the ankle.

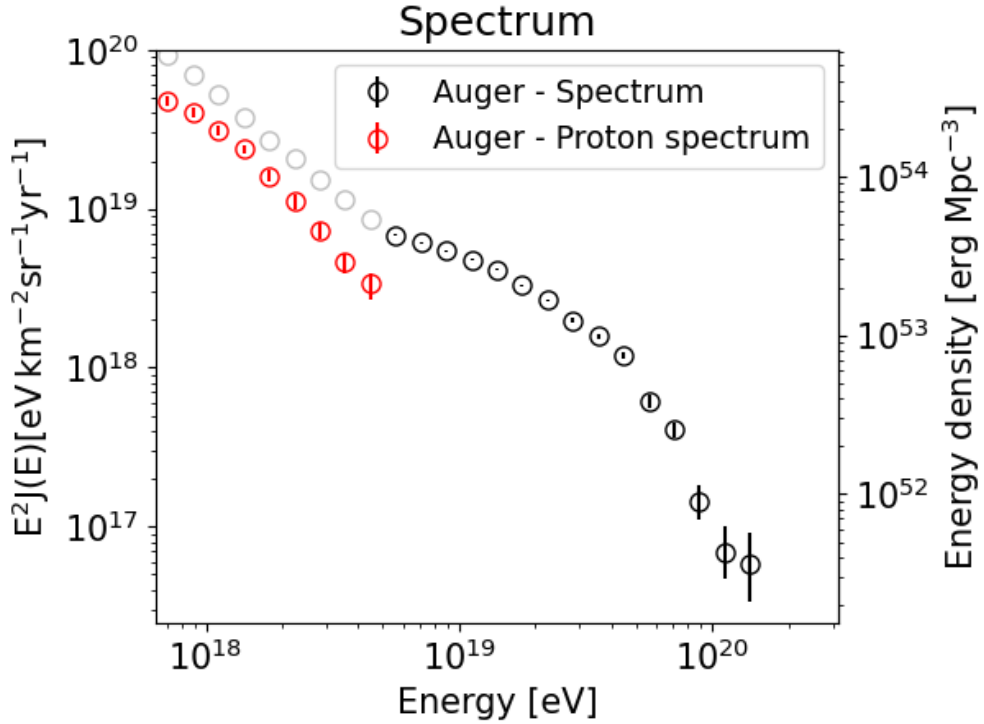
### 5.3.1 Proton spectrum and fit

#### Proton spectrum

The proton spectrum is reconstructed below the ankle energy using:

$$J^{\text{proton}}(E) = f_p \times J(E), \quad (5.19)$$

where  $f_p$  is the fraction of protons estimated in each energy bin using strictly positive fractions (cf Chap. 4).  $J(E)$  is the total flux measured at the Auger Observatory. The uncertainties are dominated by the fraction uncertainties; hence the uncertainties in



**Fig. 5.9.:** Energy spectrum measured at the Pierre Auger Observatory and its associated proton component. Both the red and black points are modelled in the following.

the spectrum are computed using the  $1\sigma$  confidence interval of the fractions. The resulting spectrum is shown in Fig. 5.9

### Fitting procedure

As before, the full spectrum and the  $X_{\max}$  distributions are modelled above the ankle energy. Below the ankle, on the other hand, only the detected proton flux is compared to the predicted proton flux. The proton flux is modelled from  $10^{17.8}$  eV (threshold of the FD) up to the ankle energy  $10^{18.7}$  using a Gaussian likelihood, as shown in Eq.5.7. To explore energies below  $10^{17.8}$  eV, one should take into account the HEAT dataset; this is not done in this analysis. Since the covariance matrix of the unfolded spectrum is not available at low energy, the spectrum likelihood is here taken to be the diagonal Gaussian likelihood shown in Eq. 5.7. On top of the two deviances used before, a new deviance is computed, the deviance associated to the proton spectrum  $D_{J_p}$ .

### 5.3.2 Results and discussion

The fit is re-run including the extra comparison with protons below the ankle energy. Since the model goes down to  $10^{17.8}$  eV, the emissivities are computed from  $E_{\min} = 10^{17.8}$  eV (Eq. 5.14). The results are shown in Tab. 5.5.

	$S(z) = \rho_*(z)$ SMD	$S(z) = \dot{\rho}_*(z)$ SFRD
$\bar{\mathcal{L}}_{\text{H}} [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$4.7 \pm 0.6$	$0.0 \pm 0.0$
$\bar{\mathcal{L}}_{\text{He}} [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$11.0 \pm 1.2$	$23.1 \pm 3.6$
$\bar{\mathcal{L}}_{\text{N}} [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$21.0 \pm 0.9$	$177.0 \pm 11.5$
$\bar{\mathcal{L}}_{\text{Si}} [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$3.0 \pm 1.0$	$17.2 \pm 9.6$
$\bar{\mathcal{L}}_{\text{Fe}} [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$0.9 \pm 1.3$	$4.9 \pm 0.7$
$D_{\text{tot}}/\text{ndf}$	698.7/138	323.8/138
$D_J/N_J$	25.5/15	32.6/15
$D_{J_p}/N_{J_p}$	413.6/9	30.9/9
$(D_X)/N_X$	259.7/121	260.3/121
$\log_{10} E_{\text{max}}/\text{eV}$	$18.18 \pm 0.03$	$18.20 \pm 0.01$
$\gamma$	$-1.53 \pm 0.02$	$-1.67 \pm 0.02$

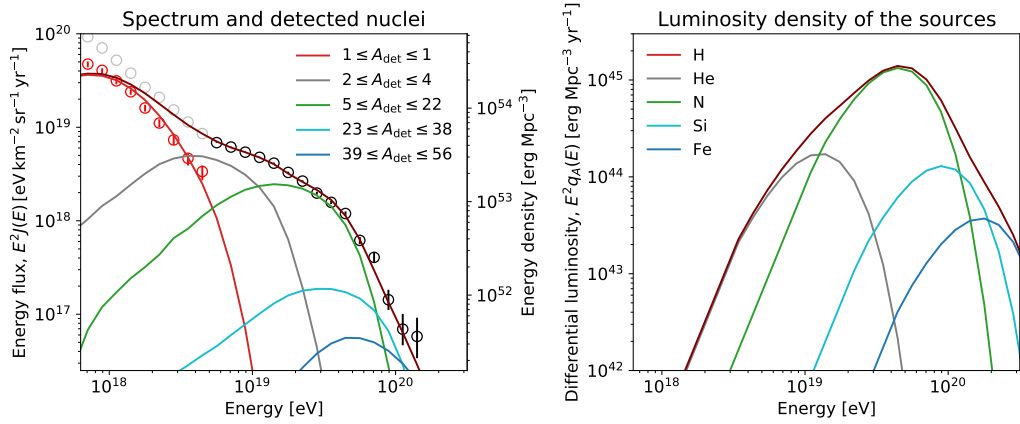
**Tab. 5.5.:** Results obtained with a luminosity evolution term that follows the SMD (left) and SFRD (right), considering the comparison between the observed proton flux across the ankle with the proton flux of the model. The ICRC 2021 dataset is used.

While the total deviance remains similar to the previous case for the SFRD case (+16%), the deviance when using SMD (+150%) shows that such an evolution of sources cannot explain the protons below the ankle ( $D_{\text{tot}}/\text{ndf} = 698.7/138$ ). The best-fit spectrum and the associated differential luminosity of the SFRD case are shown in Fig. 5.10.

Fig. 5.10 shows that the secondary protons reproduce roughly the proton below the ankle. It is interesting to note that in the SFRD scenario, no protons are injected at the sources and all the protons are produced *en route*. As already commented, this is due to the peak of SFRD at a far distance of  $\sim 10$  Gpc.

### 5.3.3 Escaped neutrons

The SFRD scenario gives an explanation for the low energy protons, but it also presents a crucial problem: the interpretation of the value inferred for the spectral index  $\gamma$ , which is given by  $\gamma \sim -1.7$ .



**Fig. 5.10.:** SFRD scenario. *Left:* Detected spectrum with detected nuclei. Colors represent the detected nuclei, red for  $A = 1$ , grey for  $2 \leq A \leq 4$ , green for  $5 \leq A \leq 22$ , cyan  $23 \leq A \leq 38$ , blue for  $39 \leq A \leq 56$ , the sum is shown in brown. *Right:* Luminosity density of the sources as a function of energy, red for H, grey for He, green for N, cyan Si, blue for Fe, the sum is shown in brown.

If the inferred  $\gamma$  value describes the acceleration mechanism, such a hard spectral index would favour accelerators with large-scale non-zero electric fields, which, due to the high electrical conductivity of most of astrophysical plasmas, can be found in special environments only that are not well known nor well constrained.

An example is a spinning black hole threaded by magnetic fields generated in surrounding accretion flows (Levinson, 2000). For such environment, the injection rate of hadrons in the gaps remains poorly constrained (e.g. Levinson, 2000), as well as the sizes of the gaps, which are crucial to determine the maximum acceleration energy and which rely at present on inputs from kinetic simulations of black hole magnetospheres (Crinquand et al., 2021; Levinson & Cerutti, 2018). Jets with tangled magnetic fields can also give rise to a net local electric field due to the interchange of field lines of opposite polarity. However, simulations show that acceleration in such magnetic reconnection regions can only be effective if the magnetic energy density greatly exceeds that of the plasma in the injection region (large magnetization), in which case the particles quickly escape from the acceleration regions and cannot reach ultra-high energies (Kagan et al., 2016; Petropoulou & Sironi, 2018). Although, these acceleration mechanisms result in a hard spectrum ( $\gamma \sim 1$  for magnetic reconnection (Guo et al., 2015)), such mechanisms cannot reproduce a hard spectrum of  $\gamma \sim -1.5$  as seen in the data.

On the other side, first-order Fermi shock acceleration is more constrained and could achieve the acceleration of UHECRs. However, as shown in 1.2.1, the typical spectral index for this mechanism is soft with  $\gamma \sim 2$ , which is even more significantly



different from the hard values favoured from the data. Therefore, if first-order Fermi processes are responsible for UHECR acceleration, the interplay between the escape from the sources and the acceleration mechanisms must be taken into account to correctly apprehend the emission spectrum (e.g. Fujita et al., 2009), in particular when escape is influenced by in-source interactions (Biehl et al., 2018; Boncioli et al., 2019; Globus et al., 2015; Supanitsky et al., 2018; Unger et al., 2015; Zhang et al., 2018).

The cosmic-ray luminosity of the source candidates is governed, among other things, by the levels of radiation and magnetic-field densities. While gaining energy, UHECRs can interact with this radiation and possibly escape from the magnetized zone. Thus, the energy spectrum of the ejected particles as well as the amount of ejected nuclei may differ strongly from those injected into the electromagnetic field. This has two important consequences: a) the ejected spectrum of the charged nuclei can be much harder than that injected, due to the escape mechanism and in particular the behaviour of the nuclei-photon cross section at high energies; b) the interactions can produce secondary neutrons of energy  $E_n = E/A$ . These neutrons can escape freely from the magnetic confinement zones, with an ejection spectrum much softer than that of nuclei, to decay into protons on their way to the Earth after an average travel distance of  $9.1 \text{ kpc} \times (E_n/1 \text{ EeV})$ .

## Model

Detailed studies of the nuclear cascade developing in sources with different levels of radiation densities from photodissociation, such as that presented in Biehl et al., 2018, show that a softer flux of secondary neutrons is expected from in-source interactions. To mimic this behaviour, the  $q_{A_g}(E_g)$  in Eq. 5.1 is split into two terms:  $q_p(E_g)$  which describes the escaped neutron, and  $q_{A_g}(E_g)$  which describes the escaped nuclei as before. The first term results in a broken exponential cut-off power-law with an index  $\gamma_p$ :

$$\omega_{i0} = \frac{dN_A}{dE_g}(E_g) = J_p \left( \frac{E_i}{E_{\text{ref}}} \right)^{-\gamma_p} \times \begin{cases} 1 & \text{if } E_i < Z \times E_{\text{max}}^p, \\ \exp\left(1 - \frac{E_i}{Z \times E_{\text{max}}^p}\right) & \text{otherwise.} \end{cases} \quad (5.20)$$

The second term is also modelled with a broken exponential cut-off power-law with a different index  $\gamma_A$ :

$$\omega_{ik} = \frac{dN_A}{dE_g}(E_g) = J_k \left( \frac{E_i}{E_{\text{ref}}} \right)^{-\gamma_A} \times \begin{cases} 1 & \text{if } E_i < Z \times E_{\text{max}}^p, \\ \exp \left( 1 - \frac{E_i}{Z \times E_{\text{max}}^p} \right) & \text{otherwise.} \end{cases} \quad (5.21)$$

The combined fit then has one extra parameter, the proton index  $\gamma_p$ . Such a model of the injected spectrum is a simplified approach aimed at parametrizing effectively the results of Biehl et al., 2018. It is important to note that the approach remains valid only in the range of energy of interest considered here.

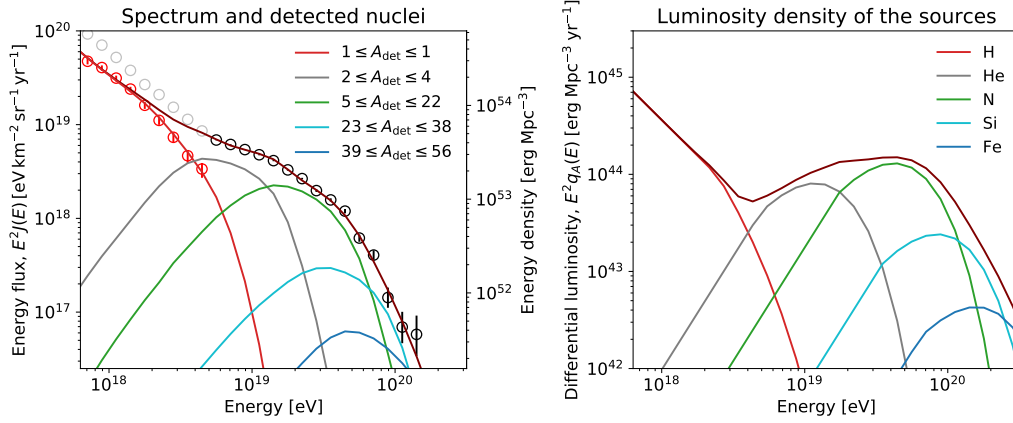
### 5.3.4 Results

The fit is run with the extra parameters ( $\gamma \rightarrow \gamma_p$  &  $\gamma_A$ ). The results are shown in Tab. 5.6.

	$S(z) = \rho_*(z)$ SMD	$S(z) = \dot{\rho}_*(z)$ SFRD
$\mathcal{L}_H [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$40.5 \pm 2.0$	$42.1 \pm 33.5$
$\mathcal{L}_{\text{He}} [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$13.0 \pm 1.2$	$16.1 \pm 3.3$
$\mathcal{L}_N [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$20.9 \pm 1.4$	$183.2 \pm 5.6$
$\mathcal{L}_{\text{Si}} [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$3.9 \pm 1.3$	$13.7 \pm 1.7$
$\mathcal{L}_{\text{Fe}} [10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}]$	$0.7 \pm 0.4$	$6.2 \pm 0.7$
$D_{\text{tot}}/\text{ndf}$	290.3/137	315.6/137
$D_J/N_J$	32.2/15	36.8/15
$D_{J_p}/N_{J_p}$	5.4/9	12.4/9
$(D_X)/N_X$	252.5/121	266.5/121
$\log_{10} E_{\text{max}}^p/\text{eV}$	$18.36 \pm 0.01$	$18.23 \pm 0.01$
$\gamma$	$-0.36 \pm 0.02$	$-1.43 \pm 0.01$
$\gamma_p$	$3.48 \pm 0.14$	$6.7 \pm 1.8$

**Tab. 5.6.:** Results obtained with a luminosity evolution term that follows the SMD (left) and SFRD (right), considering the proton across the ankle and two spectral indices  $\gamma_p$  for proton and  $\gamma_A$  for the other nuclei. The ICRC 2021 dataset is used.

The new model does not significantly improve the goodness of fit when considering SFRD. In that case, the emissivity of protons is compatible with 0 at  $\sim 1.3\sigma$  C.L. with an underconstrained  $\gamma_p = 6.7 \pm 1.8$ . On the other side, using  $\gamma_p$  improves the goodness of fit when considering SMD ( $D_{\text{tot}}/\text{ndf} = 698.7/138 \rightarrow D_{\text{tot}}/\text{ndf} = 290.3/137$ ): the reduced deviance is  $\sim 2.1$ , which is the same value as that obtained without accounting for protons below the ankle. The spectrum and the differential luminosity density are shown in Fig. 5.11.



**Fig. 5.11.:** SMD scenario with  $\gamma_p$  &  $\gamma_A$ . *Left:* Detected spectrum with detected nuclei. Colors represent the detected nuclei, red for  $A = 1$ , grey for  $2 \leq A \leq 4$ , green for  $5 \leq A \leq 22$ , cyan  $23 \leq A \leq 38$ , blue for  $39 \leq A \leq 56$ , the sum is shown in brown. *Right:* Luminosity density of the sources as a function of energy, red for H, grey for He, green for N, cyan Si, blue for Fe, the sum is shown in brown.

Fig. 5.11 allows one to clearly see the soft index of the proton component that contrasts with the hard index for the other nuclei.

### 5.3.5 Systematic uncertainties

#### Hadronic model

The analysis has been run changing the hadronic model from EPOS-LHC to Sibyll2.3d. The deviance of the SFRD scenario with one spectral index  $\gamma = \gamma_p = \gamma_A$  is 10 units better, while the SMD scenario with different spectral index  $\gamma_p \neq \gamma_A$  is 10 units worse. Sibyll tends to soften the injected spectrum of the two approaches:  $\gamma = -1.67 \rightarrow \gamma = -1.1$  for SFRD and  $(\gamma_p = 3.48; \gamma_A = -0.36) \rightarrow (\gamma_p = 3.91; \gamma_A = 0.01)$ .

#### EBL model

The range of  $E_{\text{max}}^Z$  makes the role of the EBL more important than that of the CMB to control the energy losses by photodissociation of the nuclei. Consequently, the uncertainties affecting both the far-infrared intensity of the EBL and the partial cross sections of channels in which protons are ejected can alter the results. Considering the EBL model of Dominguez et al., 2011 instead of Gilmore et al., 2012 results in an increased far-infrared density of photons, which enhances the intensity of secondary protons from photodissociations *en route*. Most notably, for the SMD scenario, this

enhancement is compensated in the fit to the data through an energy production rate of CNO elements that is increased by  $\sim 10\%$  while that of protons is decreased by  $\sim 10\%$ . Also, the spectral index of protons is increased by  $+1$  for a total deviance increases of 10 units. While one could expect the SFRD to better reproduce the low energy protons with an increase of secondaries, it is not the case. Using the EBL model of Dominguez et al., 2011 cannot reproduce the low energy protons because the number of secondaries is such that the secondaries protons predicted are higher than the total flux of UHECRs below the ankle.

### Photodisintegration cross-sections

The use of PSB (Puget et al., 1976; Stecker & Salamon, 1999) instead of TALYS (Koning et al., 2007) as photodisintegration model is tested as well. For the two approaches, the best-fit parameters are similar except for an increase of helium emissivity (a factor  $\times 2$  for SMD and  $\times 4$  for SFRD) and a decrease of CNO emissivity (a factor 0.5 for SMD and SFRD) for the two approaches. As explained in 5.1.5, this is due to a more complete description of interaction of TALYS, which takes into account the creation of  $\alpha$  particles and therefore produces helium during the propagation, while PSB does not.

### $X_{\max}$ distributions

The  $X_{\max}$  distributions are shifted to  $\pm 1\sigma$  together with the proton fractions used in Eq. 5.19. For the SMD scenario, the deviance is worsened ( $D_{\text{tot}} = 423.5$  and  $D_{\text{tot}} = 628.3$ , for respectively  $+1\sigma$  and  $-1\sigma$ ). For the SFRD case, the deviance is worsened when shifting at  $+1\sigma$  the  $X_{\max}$  data ( $D_{\text{tot}} = 628.3$ ) and the deviance is  $\sim 40$  units better when considering a negative shift of  $-1\sigma$  ( $D_{\text{tot}} = 285.1$ ). Considering this shift increases the proportion of CNO at the sources, which now represent 85% of the emissivities. The spectrum is also softened with an index  $\gamma = -1.11 \pm 0.01$

## 5.3.6 Discussion

The results presented in the previous sections direct the interpretation of the origin of protons below the ankle energy under the assumption of an extragalactic origin. While a SFRD scenario results in protons produced *en route*, a SMD scenario needs a proton component at the source with a spectral index  $\gamma_p \neq \gamma_A$ . In this picture,

the component of protons is of extragalactic origin well below the ankle energy and exponentially suppressed above it, while heavier nuclei steadily take over to the highest energies through a rigidity-dependent maximum-energy scenario.

### SMD scenario

For the SMD scenario, the soft spectral index for the proton population allows populating the region below the ankle, while the hard index of nuclei does not populate this region. Interestingly, such a behaviour qualitatively fits with scenarios of in-source interactions in which copious fluxes of neutrons, which are produced while accelerated charged particles interact with the bath of photons permeating the sources, escape freely the electromagnetic fields.

The exploratory study presented above calls for further investigation of specific scenarios; yet some general statements can be drawn on some properties of the source environments and acceleration/escape mechanisms. The simultaneous requirements of a soft spectral index for protons and of the presence of intermediate and heavy nuclei at the highest energies imply, expressed in a simple way, interactions in the source environments that produce hard ejection spectra for nuclei below the maximum rigidity and generate secondary neutrons at lower energies. At the same time, the interactions should not be too numerous to avoid the destruction of the nuclei at the highest energies. These requirements favour source environments that are optically thin to nuclei-photon interactions, or optically thick to nuclei-photon interactions but simultaneously optically thin to proton-photon interactions (Biehl et al., 2018). The electromagnetic luminosity of the sources, governed by the level of radiation density, should thus be relatively low or intermediate.

The energy production rate inferred for protons accounts for those produced by the escaping neutrons, inheriting hence an energy of the order of  $E/A$  of the nuclei energy  $E$ . The maximum energy that this population of protons can reach is then  $E_{\max}^Z/A \sim E_{\max}^P/2$ . It should be noted that this is not reflected in the rigidity-dependent acceleration scheme model by Equation 5.20. The extreme case in which all ejected protons would be from photodissociation by-products can be tested by replacing  $E_{\max}^Z = Z \times E_{\max}^P$  by  $E_{\max}^Z = \frac{A}{2} \times E_{\max}^P$ . It is interesting that the main changes in the results concern the spectral index of protons, which becomes  $\gamma_p \sim 2.7$ . The deviance is also a bit smaller ( $D_{\text{tot}} = 290.3 \rightarrow D_{\text{tot}} = 283.7$ ). Further characterization of the balance between the population of accelerated protons from the initial abundance in the source environment and that of secondaries from nuclear cascades calls for modeling of specific sources and environments. Such models would

in particular help determining the neutrino flux expected from in-source interactions and be compared to the cosmogenic neutrino limit observed by Auger (The Pierre Auger Collaboration, 2015a) and IceCube (IceCube Collaboration, 2019).

## SFRD

If UHECRs are produced in regions with high star formation rate, most of them would interact during their journey and produce low-energy protons. Due to losses, the emissivity at the engine is much more important in the SMD scenario. It is important to note that the EBL model of Dominguez et al., 2011, which is brighter in the far-infrared domain compared to Gilmore et al., 2012, is incompatible with the SFRD scenario due to the numerous secondaries produced on. Therefore, if the SFRD scenario was confirmed, it would invalidate the EBL model found in Dominguez et al., 2011. Alternatively, if the EBL model found in Dominguez et al., 2011 turns out to be correct, it would thus invalidate SFRD as a tracer of the sources of UHECRs. Although the SFRD can correctly reproduce the data (especially when considering a  $-1\sigma_{\text{syst}}$  shift of the  $X_{\text{max}}$  scale), the spectral index remains hard as it is always smaller than  $-1$ . As for the SMD scenario, a hard spectrum favours sources with photon-rich environments in order to shape an accelerated spectrum expected in a Fermi scenario from  $\gamma = 2 \rightarrow \gamma = -1$ . Although dense photon fields can harden the spectrum,  $\gamma = -1$  seems at the limit of the theoretical models. Moreover, such photon fields should not be too optically thick to avoid destroying the CNO nuclei needed to reproduce the data.

## General remarks

Although they provide elements to answer several questions through a rather soft acceleration spectrum traced by protons ( $\gamma_p \gtrsim 2.5$ ) for the SMD or secondaries protons for the SFRD, the results obtained only partially address the origin of the ankle. If protons of extragalactic origin contribute to the sub-ankle component, other elements are needed to make up the all-particle spectrum that, beyond its impressive regularity in the energy region between the second knee near 0.1 EeV and the ankle near 5 EeV, hides beneath a complex intertwining of different astrophysical phenomena. Keeping in mind the reliance of the interpretations of  $X_{\text{max}}$  data on the validity of the hadronic interaction models, this intertwining is evidenced by the composition study in Bellido J. on behalf of the Pierre Auger Collaboration, 2018 as a function of energy that can be described broadly as follows. On the one hand, a steep fall-off of the Fe component is observed well below the ankle energy. This is along the

lines of the scenario for the bulk of Galactic cosmic rays characterized by a rigidity-dependent maximum acceleration energy,  $E_{\text{max, Gal.}}^Z \sim 3Z$  PeV, for particles with charge  $Z$  to explain the knee structures. On the other hand, all hadronic-interaction models indicate the presence of CNO nuclei between the second-knee and ankle energies. This extra-component, which, in the overall scenario explored in this thesis, contributes to shape the ankle feature, raises questions. Such intermediate-mass elements could be, for example, fuelled by extragalactic sources different from those producing the bulk of UHECRs above the ankle energy (Aloisio et al., 2014). Or, they could correspond to a second Galactic component, as first suggested in Hillas, 2005, for example one resulting from explosions of Wolf-Rayet stars (Thoudam et al., 2016).

## 5.4 Conclusion

A cosmological model called combined fit has been developed in order to constrain the distribution of sources together with their injected spectrum using the spectrum and  $X_{\text{max}}$  data beyond the ankle feature. The deviance  $D_{\text{tot}}$ , an estimation of the goodness of fit, is used to compare the model to the data. The deviance is composed of two terms, a composition deviance  $D_X$  and a spectrum deviance  $D_J$ . Seven parameters are used, one spectral index  $\gamma$ , one maximum proton energy  $E_{\text{max}}^p$ , five injections fluxes  $J_k$  for five representative masses ( $\text{H}^1$ ,  $\text{He}^4$ ,  $\text{N}^{14}$ ,  $\text{Si}^{26}$ ,  $\text{Fe}^{56}$ ). The benchmark scenario is a scenario that considers a flat evolution of sources ( $D_{\text{tot}}/\text{ndf} = 288.6/129$ ). With a reduced deviance of  $\sim 2.2$ , a flat evolution of sources does not reproduce the data. However, it is shown in Chap. 4 that this is mainly due to the incapacity of the hadronic interaction models to reproduce the observed  $X_{\text{max}}$  distributions, with a best composition deviance of  $D_X^{\text{best}} = 221.6$  when using 40 parameters to describe 10 energy bins, with 121 points.

From this benchmark scenario, two evolutions of sources are tested for the first time: one where UHECR sources follow the SMD and one where UHECR sources follow the SFRD. Above the ankle, the two scenarios are similar with a deviance of  $D_{\text{tot}}/\text{ndf} = 274.5/129$  for SMD, and  $D_{\text{tot}}/\text{ndf} = 281.5/129$  for SFRD. The SMD is composed of an overdensity at small distances that decreases on cosmic scales. The SMD scenario results in a hard spectrum  $\gamma \sim -0.4$  with an injected spectrum dominated by CNO and helium:  $\bar{\mathcal{L}}_{\text{N}} = (12.7 \pm 1.4) \times 10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}$  and  $\bar{\mathcal{L}}_{\text{He}} = (22.1 \pm 2.0) \times 10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}$ . The SFRD scenario is also described by an overdensity at small distances but increases from 100 Mpc to 10 Gpc before decreasing. The SFRD scenario is particularly dominated by CNO, which represents

80% of its total emissivity  $\bar{\mathcal{L}}_N = (191.0 \pm 2.3) \times 10^{43} \text{ erg Mpc}^{-3} \text{ yr}^{-1}$ . The total emissivity of SFRD is  $\sim 6$  times bigger than that of the SMD, due to density of sources at 10 Gpc. The second consequence of far-away sources is the very hard spectrum found, which is given by  $\gamma \sim -1.8$ .

From this classical combined fit above the ankle energy, an extension is done below the ankle by comparing the predicted flux of protons to the proton spectrum, following the work of Luce, 2019. Such a comparison is valid if the protons below the ankle have the same origin as the heavier component above the ankle. While the SFRD scenario roughly reproduces the spectrum ( $D_{\text{tot}}/\text{ndf} = 323.8/138$ ), the SMD scenario cannot reproduce the low energy proton component ( $D_{\text{tot}}/\text{ndf} = 698.7/138$ ). The SFRD scenario manages to reproduce the low-energy proton component and the ankle feature using the many secondary protons produced *en route*. On the other side, the SMD scenario, in which numerous sources are nearby, produces a few secondary protons that are not sufficient to explain the protons below the ankle and the ankle feature. However, as shown in Biehl et al., 2018, in-source interactions could produce neutrons with a soft index which easily escape and turn into protons. To mimic this study, the spectral index  $\gamma$  is split into two index  $\gamma_p$  for neutrons that turn into protons and  $\gamma_A$  for other nuclei. Using this new parameter does not change much the SFRD scenario, but allows the SMD scenario to reproduce the protons below the ankle ( $D_{\text{tot}}/\text{ndf} = 290.3/138$ ). The nuclei spectral index  $\gamma_A \sim -0.4$  is similar to the combined fit above the ankle. On the flip side, the proton spectral index  $\gamma_p \sim 3.5$  is soft.

The two scenarios remain possible in the systematics. To further analyse these hypotheses, the next chapter focuses on the impact of such scenarios on the arrival directions compared to data.





# An astrophysical model that describes the three observables

*This chapter is partly adapted from the author’s work described in Biteau, Marafico et al. “Cosmographic model of the astroparticle skies”, 2022 (ICRC2022) and Marafico et al., 2022 (in prep.).*

## 6.1 From the Combined Fit to arrival directions

In the previous chapter, a tool called Combined Fit has been used to constrain the spectrum at the sources, based on different hypotheses. In the first hypothesis, sources of UHECRs follow the Stellar Mass Density (SMD), while in the second hypothesis, sources of UHECRs follow the Star Formation Rate Density (SFRD). The characterisation of the spectrum at the sources is performed through a likelihood procedure that compares the model to the composition and spectrum data of the Pierre Auger Observatory. In this chapter, starting from those two scenarios, arrival-direction sky maps are derived and can be compared to the recent observation of Auger and TA.

### 6.1.1 Tensor approach

Compared to the previous chapter, the Combined Fit is redefined in order to use relevant parameters for this study. The “tensor approach” shown below allows us to run the code in a much lighter, quicker way and enables arrival direction studies.

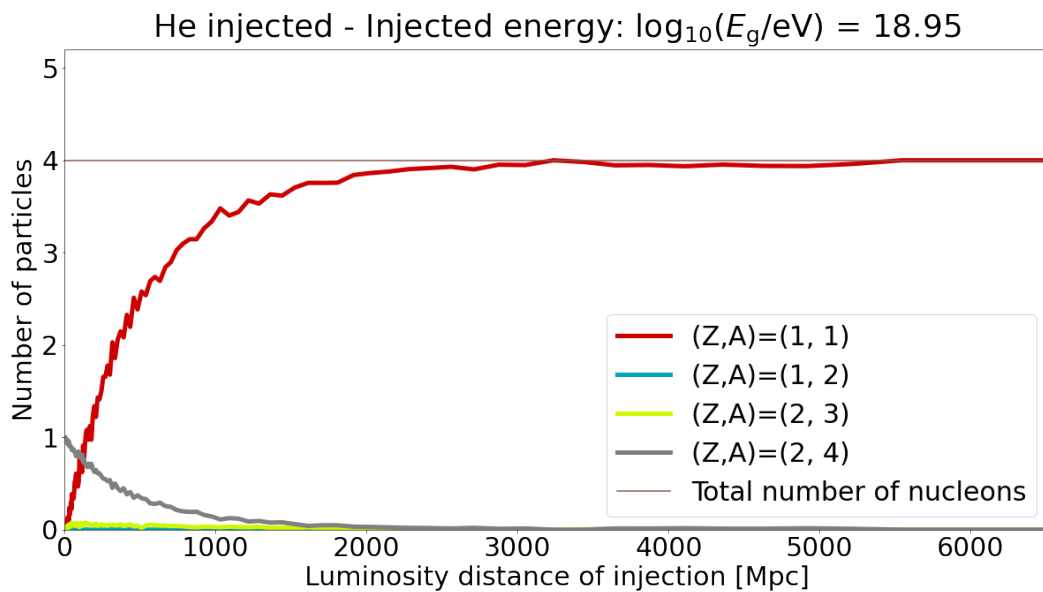
In order to create a more efficient Combined Fit, all the events generated with SimProp are stored in a normalized 5D tensor:  $T(E, A|E_g, A_g, z)$ .

The tensor gives the number of expected events of energy  $E$  and mass  $A$ , knowing the original nucleus was injected at a redshift  $z$ , with a mass  $A_g$  and energy  $E_g$ .

The first advantage is that a tensor is much quicker to load, compared to the numerous number of events needed in Chap 5. The second advantage is that the tensor can be produced using different inputs (EBL model, photodisintegration model, software used for propagating, etc.) and can be shared inside the community.

## Visualization tool

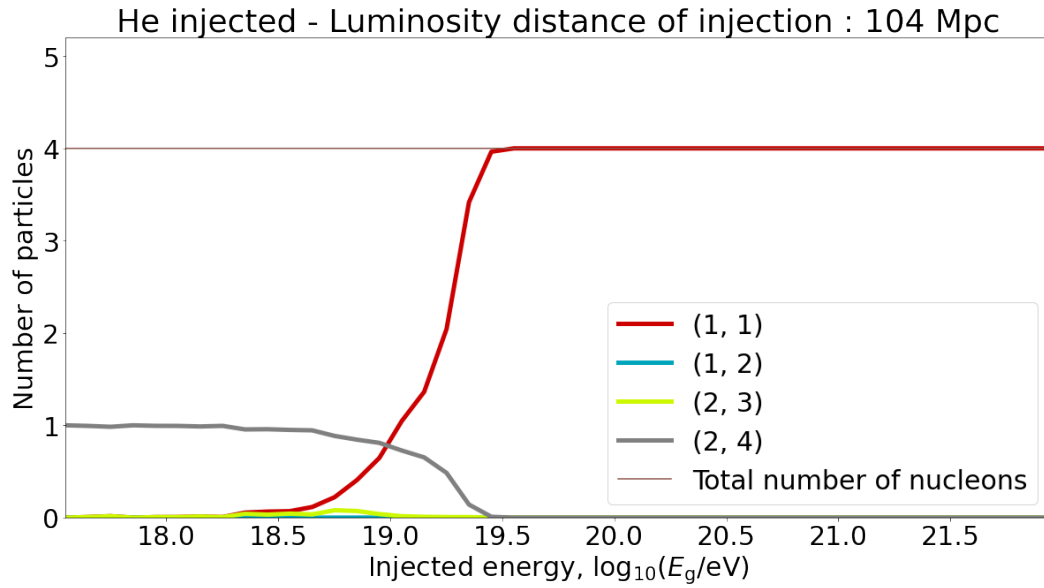
To check that the tensor is nicely defined and to better understand the effect of the propagation on nuclei, visualization tools have been made. The visualization presented here are done with a helium nucleus ( $Z = 2, A = 4$ ) injected, to avoid showing too many secondaries. The first visualisation is shown in Fig. 6.1.



**Fig. 6.1.:** Average number of nuclei detected at energies above  $10^{17}$  eV as a function of distance of injection for He injection at  $E = 10^{18.95}$  eV.

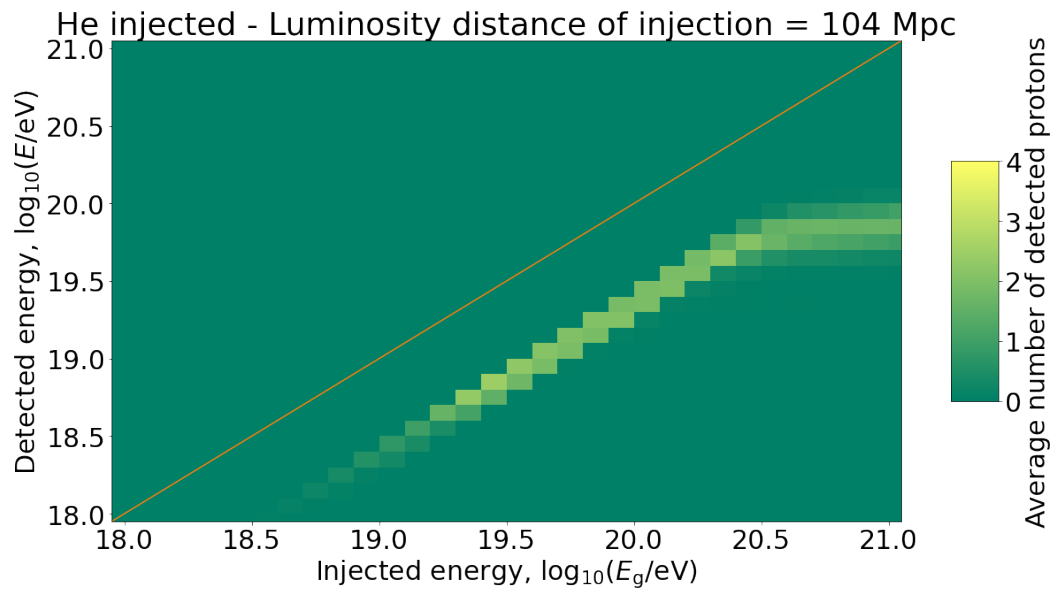
Fig. 6.1 shows the impact of the distance of injection on the secondaries detected on Earth when a helium nucleus is injected with an energy  $10^{18.95}$  eV. As expected, a helium injected at a luminosity distance close to 0 results in an average of one helium detected at Earth. When the distance of injection increases, the average number of helium nuclei drops and the average number of protons detected increases up to four, which corresponds to the split of the four nucleons of helium. One can note that apart from protons, the other channels are negligible.

Figure 6.2 shows similar results by fixing the distance of injection and varying the injected energy. Figure 6.2 shows that for an injected energy below  $10^{18.5}$  eV, a



**Fig. 6.2.:** Average number of detected nuclei at energies above  $10^{17}$  eV as a function of injected energy for He nuclei emitted at 104 Mpc.

distance of travel of  $\sim 100$  Mpc is too small to have a helium nucleus to be photodisintegrated. Above  $10^{19.5}$  eV, the helium nucleus has been fully photodisintegrated into four protons during its travel. The last plot, Fig. 6.3 shows the average number and energy of protons detected from a helium nucleus injected at  $\sim 100$  Mpc.



**Fig. 6.3.:** Average number of protons detected per bin for He injection at 104 Mpc, as a function of injected and detected energies.

Figure 6.3 shows that, at low energy (below  $10^{18.5}$  eV), an helium nucleus that travels  $\sim 100$  Mpc has not been photodisintegrated, so no protons are detected. Above  $10^{18.9}$  eV, the nucleus has been split into four protons with energy around  $E_{\text{inj}}/4$ . At the highest energies, (above  $10^{20.5}$  eV), the nucleus is split into four protons with a detected energy of  $\sim 5 \times 10^{19}$  eV corresponding to the GZK limit for UHECR protons.

## 6.1.2 Model

In this section, a similar model of the Combined Fit (shown in Sec. 5.1) is redefined in order to do arrival direction studies. As before in Sec. 5.1, the journey starts from a generation term,  $q_{A_g}(E_g, t)$ , which gives the number of nuclei of mass  $A$  created per units of energy, per second, and per covolume,  $[q_{A_g}(E_g, t)] = \text{eV}^{-1} \text{s}^{-1} \text{Mpc}^{-3}$ . From this generation term, one can compute the number of UHECRs per unit of covolume at a given energy  $E_g$  and at a given time  $t_0$ :

$$N(E_g, t_0) = \sum_{A_g} \int_0^{t_0} dt q_{A_g}(E_g, t) dE_g \quad (6.1)$$

Knowing the particles go to the speed of light, the left-hand side of Eq. 6.1 can be transformed into a flux (Eq. 6.2),

$$J(E_g, t_0) dE_g = \sum_{A_g} \frac{c}{4\pi} \int_0^{t_0} dt q_{A_g}(E_g, t) dE_g. \quad (6.2)$$

However, since the particles lose energy when propagating through EBL & CMB, the detected energy  $E$  is not equal to the initial energy  $E_g$ . Hence, Eq. 6.2 becomes Eq. 6.3, where the term  $\frac{dE_g}{dE}$  is a dimensionless function that describes the energy losses endured by the UHECR:

$$J(E, t_0) = \frac{c}{4\pi} \sum_{A_g} \int_0^{t_0} dt q_{A_g}(E_g, t) \frac{dE_g}{dE}. \quad (6.3)$$

The generation term can be decomposed into two distinct terms. A term  $\frac{dN}{dE_g}(E_g)$  which describes the number of generated particles per unit of energy per source, and a term  $S(t)$  that describes the number of sources per covolume unit and per second (Eq. 6.4),

$$q_{A_g}(E_g, t) = \frac{dN_{A_g}}{dE_g}(E_g) \times S(t). \quad (6.4)$$

The term  $\frac{dN_{A_g}}{dE_g}$  is described as a function of energy as shown in Eq. 6.5,

$$E_g \frac{dN_{A_g}}{dE_g}(E_g) = \mathcal{E}_{A_g} F(E_g) \text{ with } \int_0^\infty F(E_g) dE_g = 1. \quad (6.5)$$

$F(E_g)$  is the distribution of generated energy at the sources. It is a PDF. Thus,  $\mathcal{E}_{A_g}$  is total energy of UHECRs of mass  $A$  emitted per source. Here, we make the choice of a basic power-law function with an exponential cut-off for the  $F(E_g)$  term, it is shown in Eqs. 6.6 & 6.7:

$$F(E_g) = \frac{E_g f(E_g)}{\int_{E_{\min}}^\infty E_g f(E_g) dE_g}, \quad (6.6)$$

where  $E_{\min} = 10^{17.8}$  eV as done in the previous chapter when computing emissivities, and  $f(E_g)$  is given by:

$$f(E_g) = E^{-\gamma} \times \begin{cases} 1 & E \leq Z \times E_{\max}^p \\ e^{1 - \frac{E_g}{Z \times E_{\max}^p}} & E > Z \times E_{\max}^p \end{cases}, \quad (6.7)$$

where  $\gamma$ ,  $E_{\max}^p$  are parameters of the model and  $Z$  is the atomic number of the generated element.

As before, two evolution of sources are considered:

$$S(t) = k \times s(t) = k \times \begin{cases} \rho_*(t) \\ \dot{\rho}_*(t) \end{cases}, \quad (6.8)$$

where  $[k] = M_\odot^{-1}$  for the SFRD case and  $[k] = M_\odot^{-1} \text{s}^{-1}$  for the SMD case.

Taking Eq. 6.3 & 6.4, injecting equations 6.5 & 6.8, and transforming the integral from time  $t$  to redshift  $z$ , one can get Eq. 6.9:

$$J(E, z_0) = \frac{c}{4\pi} \sum_{A_g} \int_0^\infty dz \left| \frac{dt}{dz} \right| \mathcal{E}_{A_g} \frac{F(E_g)}{E_g} k \times s(z) \frac{dE_g}{dE}. \quad (6.9)$$

In equation 6.9, the term  $\frac{dE_g}{dE}$  is described using a 5D tensor that gives for a generated species  $A$ , at energy  $E_g$ , at a redshift  $z$ , the number of detected particles of mass  $A_{\text{det}}$ , and energy  $E$ , at redshift  $z = 0$ . The 5D propagation tensor is generated from

SimProp simulations (Aloisio et al., 2017). In SimProp, UHECR are propagated through the CMB and EBL (Gilmore et al., 2012), the photodisintegration cross-sections are taken from Koning et al., 2007 (TALYS).

To keep information about the detected nuclei, one can write equation 6.9 as:

$$J_A(E, z_0) = \frac{c}{4\pi} \sum_{A_g} \sum_{E_g} \int_0^\infty dz \left| \frac{dt}{dz} \right| \mathcal{E}_{A_g} \frac{F(E_g)}{E_g} k \times s(z) T(E, A|E_g, A_g, z) \frac{\Delta E_g}{\Delta E} \quad (6.10)$$

where  $J_A(E, z_0)$  is the flux of UHECR of detected mass  $A$  and  $T(E, A|E_g, A_g, z)$  is the 5D propagation tensor.

### List of input parameters

The parameters in Eq. 6.9 & 6.10 are  $\mathcal{E}_{A_g}$ ,  $k$ ,  $\gamma$ , and  $E_{\max}^p$ . Five representative masses are injected, in order to reproduce the spectrum ( $E_H$ ,  $E_{He}$ ,  $E_N$ ,  $E_{Si}$ ,  $E_{Fe}$ ). However,  $\mathcal{E}_{A_g}$  and  $k$  are degenerate, hence the effective parameters are:

$\mathcal{E}_{A_g} \times k$	Injected energy per stellar mass for SFRD and per stellar mass and per year for SMD scenario [5 parameters]
$\gamma$	spectral index at the sources [1 parameter]
$E_{\max}^p$	cut-off rigidity [1 parameter]

One can note that the average luminosity computed in Chap. 5,  $\bar{\mathcal{L}}_A$ , can be transformed to  $\mathcal{E}_{A_g} \times k$ , using:

$$\mathcal{E}_{A_g} \times k = \bar{\mathcal{L}}_A \frac{\int \frac{dt}{dz} dz}{\int s(z) \frac{dt}{dz} dz} \quad (6.11)$$

### Arrival direction reconstruction

Now that the model has been refined, the idea is to take the catalogue of Biteau, 2021, which contains nearly 400,000 galaxies in the first 350 Mpc ( $z = 0.08$ ) around the Milky Way, and associate to each of them a UHECR flux. However, the model described above considers a continuous evolution of sources  $s(z)$ . One needs to convert this model into a model that provides the flux for a given galaxy.

In order to do so, Eq. 6.10 is converted into Eq. 6.12 to compute the contribution of an individual galaxy:

$$\mathcal{N}(E \geq E_{\text{th}})_{\text{Gal}} = \frac{1}{d_L^2} \frac{1}{4\pi} \sum_A \sum_{A_g} \sum_{E_g} \mathcal{E}_{A_g} \frac{F(E_g)}{E_g} k s_{\text{Gal}} T(E, A|E_g, A_g, z) \Delta E_g, \quad (6.12)$$

where  $s_{\text{Gal}}$  is the tracer considered (i.e. stellar mass of the galaxy in a SMD scenario and star formation rate of the galaxy in a SFRD scenario), and  $d_L$  is the luminosity distance of the galaxy. As shown in Chap. 3, sky maps are done above a fixed energy threshold and provided in units  $[\mathcal{N}] = \# \text{ km}^{-2} \text{ yr}^{-1} \text{ sr}^{-1}$

The background is computed following Eq. 6.10 from a redshift of  $z = 0.08$  up to  $z = 2.50$ :

$$\begin{aligned} \mathcal{N}(E \geq E_{\text{th}})_{\text{background}}|_{0.08 \rightarrow 2.50} &= \frac{c}{4\pi} \sum_{E=E_{\text{th}}}^{\infty} \sum_A \sum_{A_g} \sum_{E_g} \int_{0.08}^{\infty} dz \left| \frac{dt}{dz} \right| \mathcal{E}_{A_g} \frac{F(E_g)}{E_g} k \\ &\quad \times s(z) T(E, A|E_g, A_g, z) \Delta E_g \end{aligned} \quad (6.13)$$

The contribution of all galaxies within 350 Mpc is normalized so that:

$$\sum_{\text{Gal}} \mathcal{N}(E \geq E_{\text{th}})_{\text{Gal}} = \mathcal{N}(E \geq E_{\text{th}})|_{0.00 \rightarrow 2.50} - \mathcal{N}(E \geq E_{\text{th}})|_{0.08 \rightarrow 2.50} \quad (6.14)$$

Then, the flux in a given pixel  $i$  is given by:

$$\mathcal{N}_i(E \geq E_{\text{th}}) = \mathcal{N}(E \geq E_{\text{th}})|_{0.08 \rightarrow 2.50} + \sum_{\text{Gal}} \mathcal{N}_i(E \geq E_{\text{th}})_{\text{Gal}}. \quad (6.15)$$

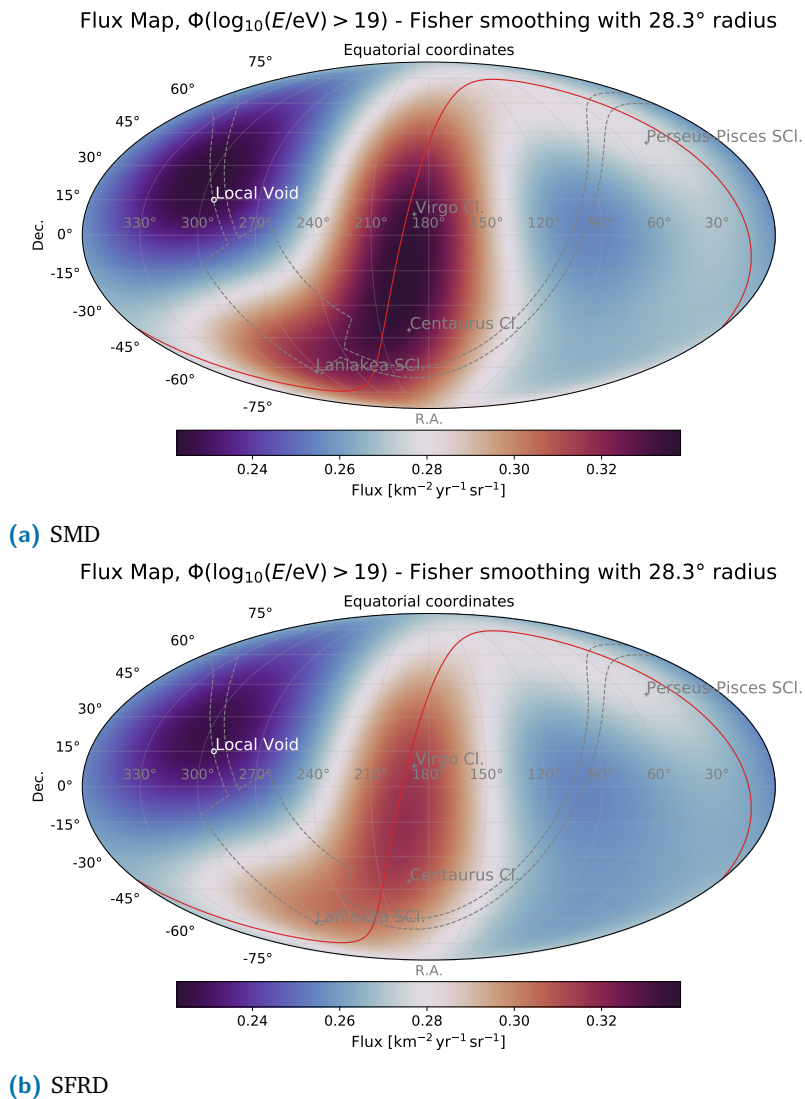
The flux is then smoothed following the procedure shown in Eq. 3.9 and using a von Mises–Fisher distribution. Compared to a top-hat smoothing, a von Mises-Fisher is continuous over the full-sky and results in a smoother outcome. As shown in The Pierre Auger collaboration, 2022, the radius  $R$  of smoothing used with a von Mises–Fisher distribution can be transposed to an equivalent radius  $\theta$  of a top-hat smoothing using the relation:

$$\theta = 1.59 R. \quad (6.16)$$

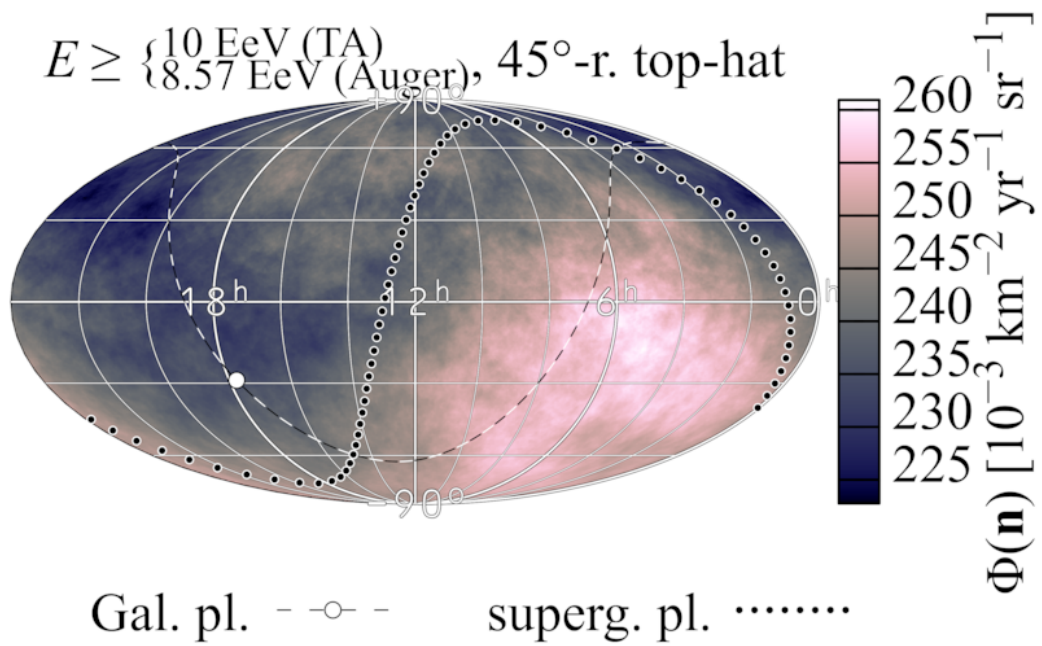


### 6.1.3 Arrival direction map and comparison to data

Sky maps are produced above a threshold of  $E_{\text{th}} = 10^{18.9} \text{ eV} = 8 \text{ EeV}$  for the two scenarios. This threshold is motivated by the observed dipole at the same energy. The resulting maps using SMD and SFRD scenarios are shown in Fig. 6.4. The maps are shown in equatorial coordinates in order to be compared the full-sky data. The maps are smoothed with a radius of  $R = 28.3^\circ$  using a Fisher distribution in order to match the  $45^\circ$  top-hat radius used to display the data. The full-sky data using Auger and Telescope Array observations are shown for comparison in Fig. 6.5.



**Fig. 6.4.:** Sky maps of the expected flux above  $E_{\text{th}} = 10^{18.9} \text{ eV}$  using the SMD and SFRD scenarios. The grey lines shows the zone of avoidance along the Galactic plane. The red line corresponds to the super galactic plane.



**Fig. 6.5.:** UHECR full sky flux seen from the Pierre Auger Observatory and Telescope Array above 8.57 EeV (Auger) and 10 EeV (TA). The map is shown in equatorial coordinate, the galactic center is shown as a white dot. Extracted from Tinyakov P. on behalf of the Pierre Auger and Telescope Array collaborations, 2022.

The data are taken at two different energy thresholds (8.57 EeV for Auger, and 10 EeV for TA) in order to take into account the energy shift of 16% observed between the two observatories. Note that TA is not fully efficient below 10 EeV, which corresponds to 8.57 EeV for Auger, slightly above 8 EeV considered above.

## Discussion

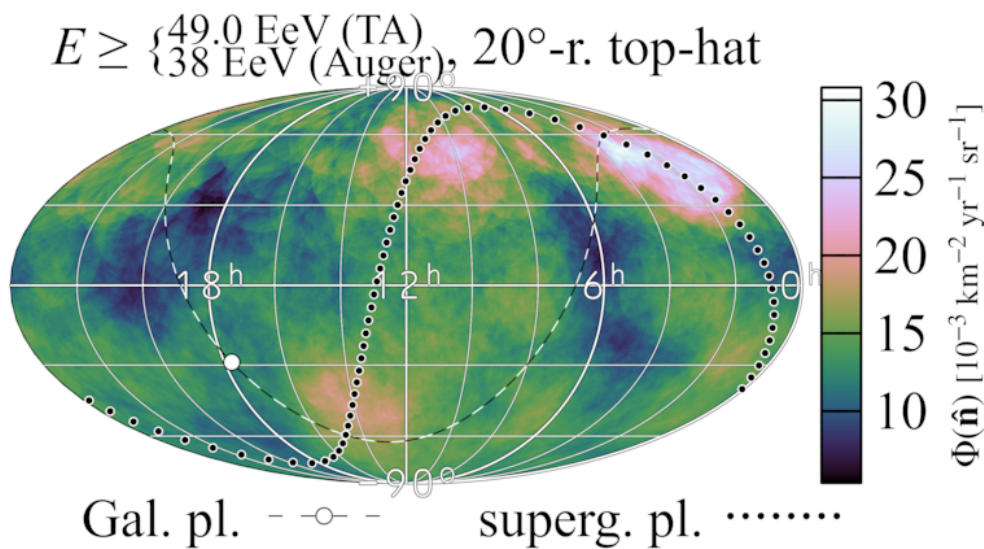
Despite the two different scenarios, the results are very similar in terms of structures in arrival direction. The main differences between the two are the contribution of the background (above 350 Mpc) compared to the foreground (below 350 Mpc, described by the catalogue of galaxies). The contribution of the background represents  $\sim 52\%$  of the total contribution in the SFRD scenario, while the contribution of the background represents  $\sim 38\%$  in the SMD scenario. This explains the larger contrast for the SMD scenario compared to the SFRD one. This contrast (i.e. the value between the minimum and maximum flux), compared to the data, could be a key to constrain the evolution of source. For the two scenarios the lowest flux comes, as expected, from the Local Void. This behaviour is also seen on the data

as the anti-dipole position is compatible with the local-void at  $1\sigma$  C.L. (The Pierre Auger collaboration, 2022). On the other side, the brightest region in the sky comes from the supergalactic plane, with major contribution in the directions of Laniakea (incl. the Centaurus Cluster), the Virgo Cluster and the Perseus-Pisces supercluster. When looking at the data (cf. Fig. 6.5), the brightest region is quite far from the supergalactic plane. This may be due to the galactic magnetic field, which deviated the UHECR as suggested in Ding et al., 2021; The Pierre Auger Collaboration, 2018b. It could also come from the presence of clusters or superclusters (such as Virgo, Shapley, Laniakea, Perseus Pisces... ) which may not contribute as much as presented in Fig. 6.4 due to confinement in the magnetized environment of clusters (cf. Sec. 6.2)

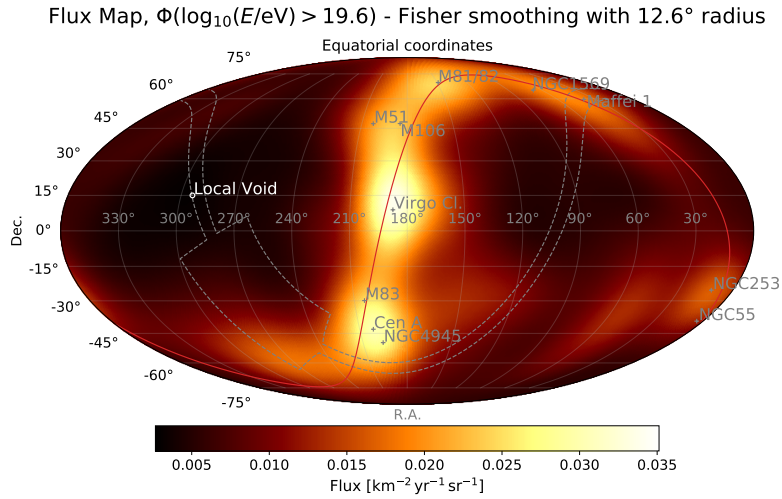
### 6.1.4 At the highest energy

As shown in Chap. 3, a correlation of  $4\sigma$  C.L. is found between a catalogue of starburst galaxies and the directions of arrival of UHECR above  $\sim 40$  EeV (The Pierre Auger collaboration, 2022). For comparison, sky maps are computed above an energy of  $10^{19.6}$  eV  $\sim 40$  EeV and shown in Fig. 6.7.

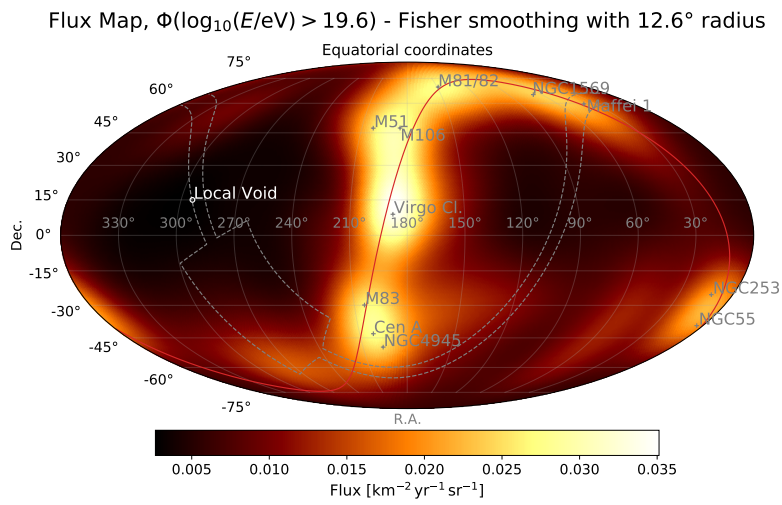
The full-sky data which uses Auger et TA data are shown in Fig. 6.6



**Fig. 6.6.:** UHECR full sky flux seen from the Pierre Auger Observatory and Telescope Array above 38 EeV (Auger) and 49 EeV (TA). The map is shown in equatorial coordinate, the galactic center is shown as a white dot. Extracted from Di Matteo A. on behalf of the Pierre Auger and Telescope Array collaborations, 2021 .



(a) SMD



(b) SFRD

**Fig. 6.7.:** Sky maps of the expected flux above  $E_{\text{th}} = 10^{18.9}$  eV using the SMD and SFRD scenarios.

## Discussion

As before, the differences between the SFRD and SMD scenario are qualitatively small. The major contributions are in the directions of:

- Virgo cluster at  $d = 15 - 20$  Mpc;
- The Centaurus group (Cen A, M83, NGC 4945, at  $d = 4 - 6$  Mpc);
- M51/M106: Spiral galaxies at  $d = 7 - 9$  Mpc;
- M81/M82: Spiral and interacting galaxies at  $d = 3 - 4$  Mpc;

- NGC1569: Interacting galaxy at  $d \sim 3$  Mpc (The Pierre Auger collaboration, 2022);
- Maffei 1: Lenticular galaxy at  $d \sim 4$  Mpc.

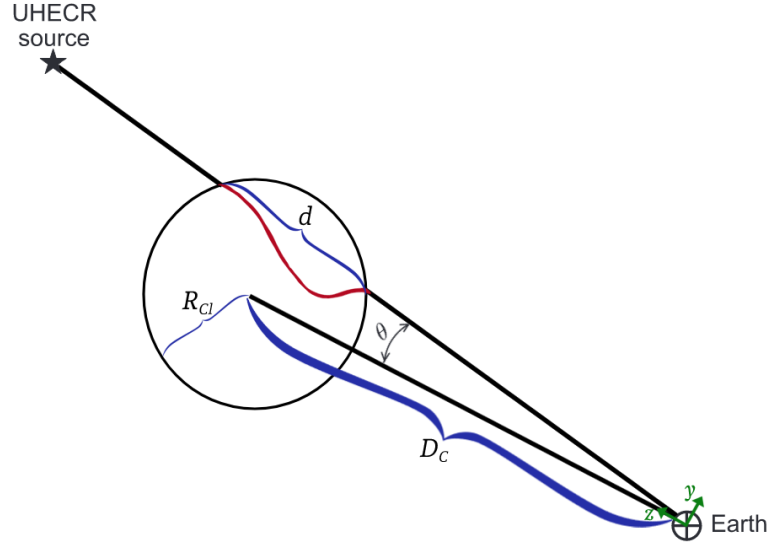
The Virgo cluster is the dominant source of UHECRs in this model. The Virgo cluster is the nearest cluster, located at  $\sim 16.5$  Mpc and has a mass of  $12 \times 10^{15} M_{\odot}$  within a radius of 2.2 Mpc from its center (Fouque et al., 2001). Virgo also contains a star forming region (Gerhard et al., 2002). This starforming region is not located at the center of a galaxy but at the border of a galaxy, and possibly inside the Virgo intracluster space area. This region could be created by energetic jet outflows from the active galaxy NGC 4388 (Gerhard et al., 2002). Because this star forming region is not strictly speaking a starburst galaxy, Virgo is not described by the starburst model used in The Pierre Auger collaboration, 2022. The important mass and the star forming region of Virgo makes it the brightest spot in the SFRD and SMD scenarios. However, when looking at the data in Fig. 6.6 there is no hot-spot observed in the direction of Virgo. Multiple hypotheses have been made in order to explain why this cluster does not appear in the UHECR data. Galactic magnetic fields (GMF) could deflect UHECRs from Virgo (Eichmann et al., 2022) and so could ExtraGalactic Magnetic Fields (EGMF) (de Oliveira & de Souza, 2021). However, one should keep in mind that GMF and EGMF are not well known, especially the EGMF whose amplitude is poorly constrained despite lower and upper bounds (Alves Batista & Saveliev, 2021). In de Oliveira and de Souza, 2021, its direction is evaluated through magnetohydrodynamics simulations, which results in large differences from one simulation to another (Hackstein et al., 2018). The following section suggests another scenario which accounts for the cluster's magnetic field (cf. Sec.6.2). Apart from Virgo, most of the sky map is dominated by nearby star forming galaxies which are present in the study of The Pierre Auger collaboration, 2022.

## 6.2 Impact of the cluster's magnetic field

As said above, the Virgo cluster is not seen in the data while it is the dominant source in the model. In this section, we evaluate whether Virgo is not seen in the data because of the cluster's magnetic field. If the cluster's magnetic field is high enough, UHECR could be trapped long enough to be photo-disintegrated before escaping.

Clusters' magnetic fields are not well known apart from that of the Coma cluster which has been well characterised (Bonafede et al., 2010). In this section, Virgo is

treated as if it behaved as the Coma cluster. The magnetic field of Virgo is modelled within a sphere of radius  $R_{Cl}$ , located at a distance  $D_C$  to the Earth. As shown in Fig. 6.8, an UHECR that is emitted by a source behind the cluster needs to cross a length  $d$  that depends on the angular separation between the centre of the cluster and the source,  $\theta$ .



**Fig. 6.8.:** Schematic of a UHECR that traverses a cluster magnetic field .

The distance  $d$  can be computed by determining the position of the two points that belong to the circle and the line of sight:

$$\begin{cases} y^2 + (z - D_c)^2 = R_{Cl}^2 \\ y = z \tan \theta \end{cases} . \quad (6.17)$$

From this distance  $d$ , we will try to evaluate the time spent inside the cluster by an UHECR.

### 6.2.1 Propagation inside the cluster

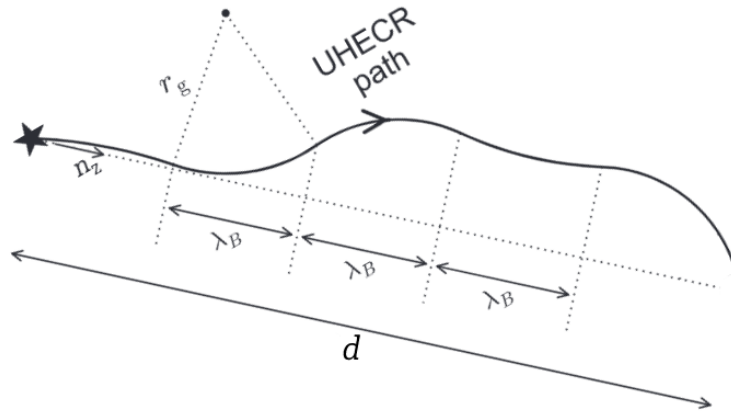
In this section, the time delay induced by the propagation in a homogeneous isotropic turbulent magnetic field is evaluated. One can define a unit vector  $\hat{n} \equiv (n_1, n_2, n_3)$  along the direction of flight. In this referential, a test particle can be described using a vector  $\mathbf{r} = (x, y, z)$ .



A charged particle in a given magnetic field  $B$  follows a radius of gyration  $r_g(E)$  that is given by:

$$r_g(E) = \frac{E}{ZecB}, \quad (6.18)$$

where  $Ze$  is the charge of the particle,  $c$  the speed of light,  $B$  the magnetic field traversed by the particle, and  $E$  its energy. Since the magnetic field of the cluster is turbulent (Bonafede et al., 2010), one can associate to it a coherence length  $\lambda_B$ . The coherence length is the typical length to see a change in direction of the magnetic field. Fig. 6.9 shows the propagation of a particle of gyroradius  $r_g(E)$  in a homogeneous isotropic turbulent magnetic field of coherence length  $\lambda_B$ .



**Fig. 6.9.:** Typical path of a UHECR in a homogeneous isotropic turbulent magnetic field of coherence length  $\lambda_B$  at a distance  $d$ , adapted from Bray and Scaife, 2018.

The magnetic field inside the cluster is modeled with a constant amplitude in successive cells of size  $\lambda_B$ . The coherence length  $\lambda_B = 10$  kpc is taken following the average value found in Coma cluster (Bonafede et al., 2010). The value of the magnetic field  $B$  is taken to be the average value of the magnetic field of the Coma cluster from 0 to  $R_{Cl}$ , which corresponds to  $B = 2.8 \mu\text{G}$  and  $B = 1.9 \mu\text{G}$  for  $R_{Cl} = 1$  Mpc and  $R_{Cl} = 2$  Mpc respectively (Bonafede et al., 2010).

In order to estimate the time delay in such environment, we closely follow, here, the procedure developed in Achterberg et al., 1999 which solves the propagation of UHECRs in a turbulent magnetic field using diffusive mechanism and assuming small-angle scattering.

## From magnetic field to a diffusion coefficient

The equations of motion for an ultra-relativistic particle with charge  $Ze$  and energy  $E$  in a quasi-static magnetic field can be written as

$$\frac{d\mathbf{x}}{ds} = \hat{\mathbf{n}} \quad , \quad \frac{d\hat{\mathbf{n}}}{ds} = \left( \frac{Ze|\mathbf{B}|}{E} \right) \hat{\mathbf{n}} \times \hat{\mathbf{b}} . \quad (6.19)$$

Here  $s \equiv ct$  is the path length of the UHECR and  $\hat{\mathbf{b}} \equiv (b_1, b_2, b_3)$  is the unit vector along the magnetic field. Assuming a uniform and constant magnetic field  $B_r$  with a random (isotropic) direction distribution gives:

$$|\mathbf{B}| = B_r \quad , \quad \langle b_i \rangle = 0 \quad , \quad \langle b_i b_j \rangle = \frac{1}{3} \delta_{ij} . \quad (6.20)$$

This is equivalent to cells that are randomly oriented and have a constant magnetic field intensity. Then, Eq. 6.19 can be written as:

$$\frac{dn_i}{ds} = \frac{\epsilon_{ijk} n_j b_k}{r_g(E)} , \quad (6.21)$$

where  $\epsilon_{ijk}$  the Levi-Civita symbol. The Einstein summation are used here. As shown in Achterberg et al., 1999, scattering of UHECRs can be described using a diffusion coefficient given by

$$\mathcal{D}_{ij} \equiv \frac{\langle \Delta n_i \Delta n_j \rangle}{2\Delta s} , \quad (6.22)$$

which follows the Kubo-Taylor formula (Taylor, 1922):

$$\mathcal{D}_{ij} = \frac{1}{r_g^2(E)} \int_0^\infty ds \epsilon_{ikl} n_k(0) \epsilon_{jqr} n_q(s) \langle b_l(0) b_r(\mathbf{x}(s)) \rangle . \quad (6.23)$$

Since small angle scattering is assumed, the direction of flight remains roughly the same  $n_k(0)n_q(s) \approx n_k(0)n_q(0)$ ,

$$\mathcal{D}_{ij} = \frac{1}{r_g^2(E)} \epsilon_{ikl} n_k(0) \epsilon_{jqr} n_q(0) \int_0^\infty ds \langle b_l(0) b_r(\mathbf{x}(s)) \rangle . \quad (6.24)$$

Since the magnetic field has a coherence length of  $\lambda_B$ , one can compute the integral to obtain the diffusion tensor:

$$\mathcal{D}_{ij} = \frac{1}{r_g^2(E)} \epsilon_{ikl} n_k(0) \epsilon_{jqr} n_q(0) \frac{\lambda_B \delta_{lr}}{3} . \quad (6.25)$$

Eq. 6.25 can be rewritten as

$$\mathcal{D}_{ij} = \mathcal{D}_0 (\delta_{ij} - n_i n_j) , \quad (6.26)$$



with  $\mathcal{D}_0$  the diffusion coefficient,

$$\mathcal{D}_0 = \frac{\lambda_B}{3 r_g^2(E)}. \quad (6.27)$$

The diffusion coefficient of Eq. 6.26 matches the one found in Achterberg et al., 1999.

### Stochastic process

As shown in Achterberg et al., 1999, the propagation can be model using stochastic differential equations with the diffusion coefficient  $\mathcal{D}_0$  that has just been computed:

$$dn_i = -2\mathcal{D}_0 n_i ds + \sqrt{2\mathcal{D}_0} P_{ij}(\hat{\mathbf{n}}) dW_j, \quad (6.28)$$

where  $P_{ij}(\hat{\mathbf{n}}) \equiv \delta_{ij} - n_i n_j$  is the projection tensor onto the plane perpendicular to  $\hat{\mathbf{n}}$ . The quantity  $d\mathbf{W} \equiv (dW_1, dW_2, dW_3)$  is a three-component Wiener process, which mathematically describes a continuous-time stochastic process.

Eq. 6.28 can be solved to retrieve the average position  $\langle ||\vec{r}|| \rangle$  induced by the diffusion inside the magnetic field at a time  $s = ct$ . From the average position, one can obtain the time delay. The full computation is a bit tedious and shown in Appendix B. The final result is shown in Eq. 6.29 and matches the result of Achterberg et al., 1999:

$$c t_{\text{del}} = \frac{\lambda_B}{9} \frac{(ZecB)^2}{E^2} d^2 \quad (6.29)$$

$$t_{\text{del}} \simeq 3.1 \times 10^3 \left( \frac{B}{10 \text{ nG}} \right)^2 \left( \frac{R}{10 \text{ EV}} \right)^{-2} \left( \frac{d}{1 \text{ Mpc}} \right)^2 \left( \frac{\lambda_B}{10 \text{ kpc}} \right) \text{ yrs.} \quad (6.30)$$

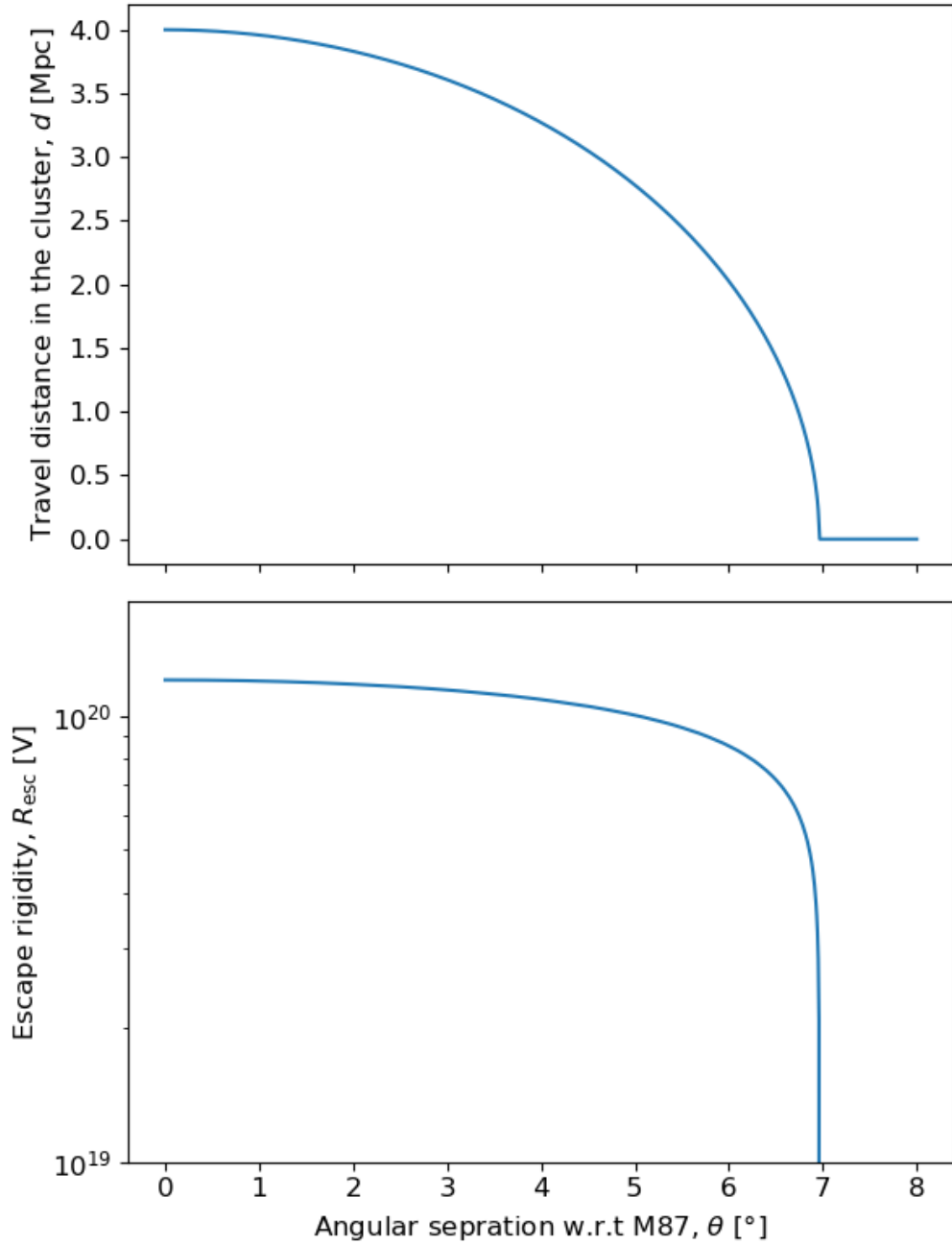
Now that we have the time delay  $t_{\text{del}}$  corresponding to the scattering of UHECRs in the Virgo cluster, we need to take a condition that states if an UHECR escapes or not the cluster. This condition is taken to be:

$$t_{\text{del}} = t_{\text{bal}} = \frac{d}{c}. \quad (6.31)$$

From Eq. 6.31 one can compute an escape rigidity  $R_{\text{esc}}$  which corresponds to the minimum rigidity needed for a UHECR to cross the cluster without entering a strong diffusion regime.

From Eq. 6.17 and Eq. 6.31, one can compute the distance  $d$  traveled from a source behind the cluster of Virgo and the associated  $R_{\text{esc}}$ . The results are shown in Fig. 6.10

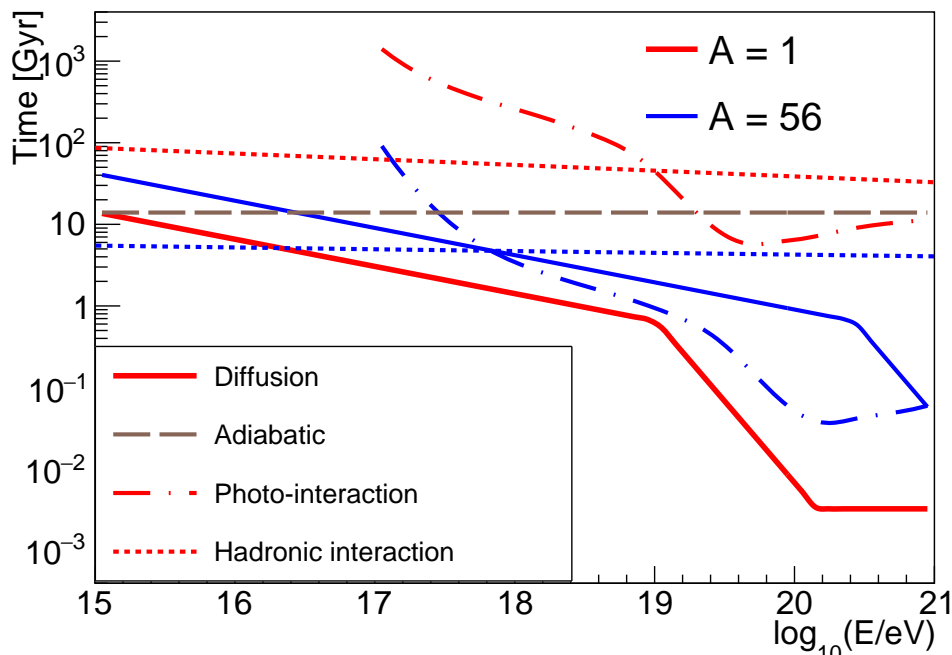
for a cluster of  $R_{Cl} = 2 \text{ Mpc}$  at a distance of  $D_C = 16.5 \text{ Mpc}$  corresponding to the distance of M87, considered as the center of Virgo (Mei et al., 2007).



**Fig. 6.10.:** *Top:* Distance  $d$  traveled by an UHECR emitted by a source located behind the cluster with an angular distance  $\theta$  with respect to the center of the cluster. *Bottom:* Corresponding escape rigidity  $R_{esc}$ .

## Validity of the model

The approach is done only as precursor analysis. A full analysis would need to determine whether UHECRs remain confined long enough in order to photodissociate or lose their energy. The analysis shall also determine whether the counterpart gamma-rays and neutrinos overpass the current observational constraints. Such a study is ongoing in our group. In order to do so, magnetic diffusion inside the cluster is taken into account inside SimProp (Aloisio et al., 2017). Fig. 6.11 shows the associated timescales of magnetic diffusion for iron and proton compared to different interaction timescale, considering a cluster of  $R_{Cl} = 1$  Mpc, and a magnetic Kolmogorov spectrum, assuming  $B_{RMS} = 1 \mu\text{G}$ , and  $l_c = 10$  kpc. The magnetic field follows a radial distribution characteristic of the Virgo cluster. The hadronic interactions are considered assuming a distribution of hadronic matter composed of proton and helium which are function of the radial distance. The photon field is only coming from EBL/CMB and is not source dependant.

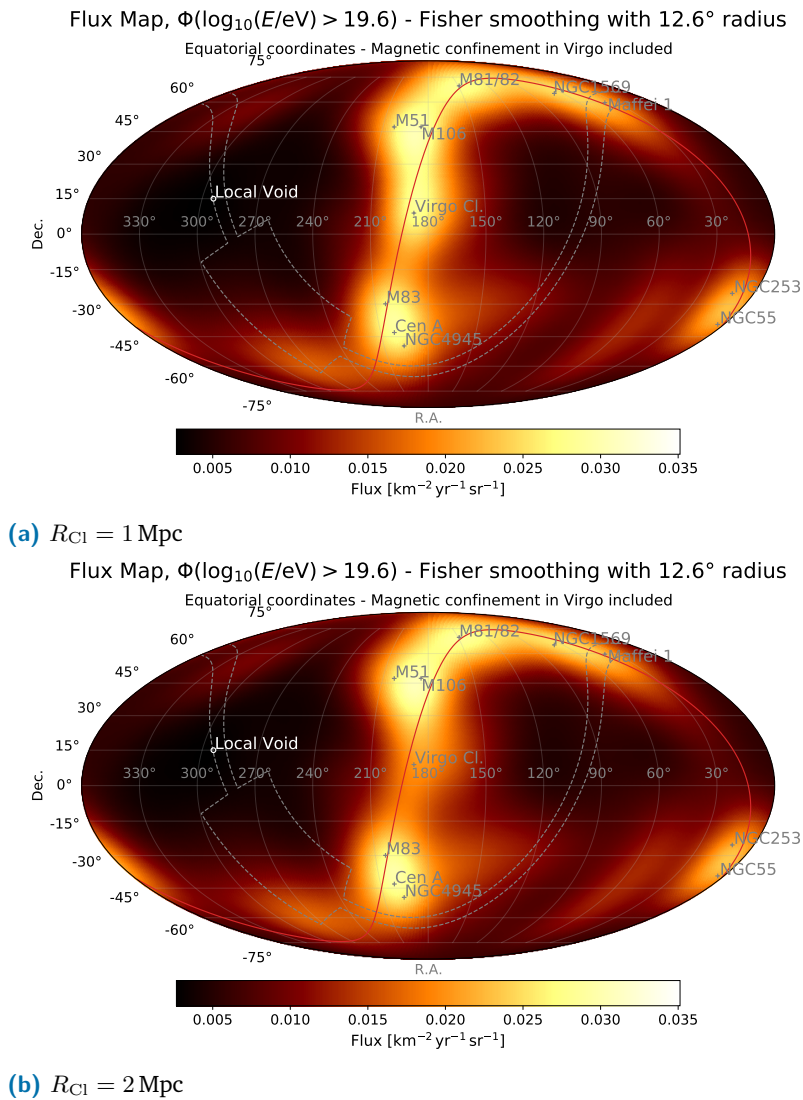


**Fig. 6.11.:** Comparison between diffusion timescale and different interaction timescales for a source at the centre of cluster. From Condorelli A. (private communication).

Fig. 6.11 shows that the diffusion time for protons is always smaller than the interaction time. Therefore, protons can escape the cluster. On the other side, for heavy nuclei, such as iron with an energy above 1 EeV, the photo-interaction time is smaller than the diffusion time and the nucleus can be photodisintegrated before it

reaches the border of the cluster. As a consequence, nuclei remain trapped inside the cluster. Therefore, the approach used here, although it is very simple, mimics such behaviour for the Virgo cluster. However, a more precise analysis using the work done in Fig. 6.11 is ongoing, that will not only consider the Virgo cluster, but also all kinds of clusters or superclusters that may impact the UHECR flux (Shapley, Perseus Pisces, Laniakea...).

## 6.2.2 Impact on the sky maps

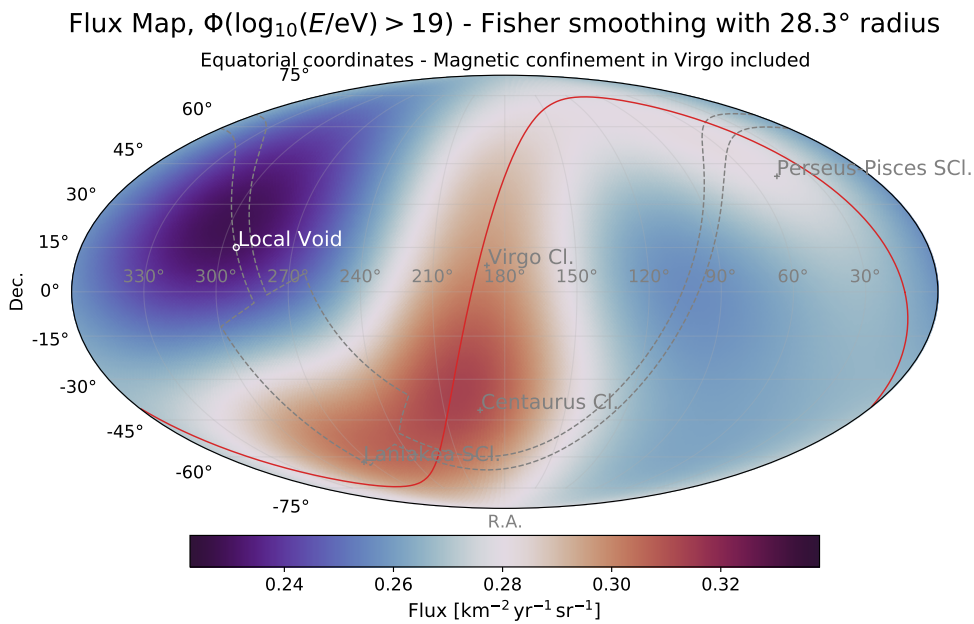


**Fig. 6.12.:** Sky map of the expected flux above  $E_{th} = 10^{19.6}$  eV considering Virgo's magnetic field.

Since the SFRD and SMD scenarios are quite close, from now on, the results are shown only in terms of SFRD. The results considering the impact of Virgo’s magnetic field are shown in Fig. 6.12 considering  $R_{\text{Cl}} = 1 \text{ Mpc}$  and  $R_{\text{Cl}} = 2 \text{ Mpc}$  above  $\sim 40 \text{ EeV}$ .

As shown in Fig. 6.12, considering a cluster of radius  $R_{\text{Cl}} = 1 \text{ Mpc}$  results in little difference between the analysis without taking into account the cluster magnetic field. However, considering a cluster of  $R_{\text{Cl}} = 2 \text{ Mpc}$  highly reduces the flux in the direction of Virgo. Such a radius is similar to the radius used in Mei et al., 2007, where the mass is estimated taking the galaxies within a radius of  $R_{\text{Cl}} \simeq 2.2 \text{ Mpc}$  around M87.

The same sky map is computed above  $\sim 8 \text{ EeV}$  to be compared to Fig. 6.5. The resulting map is shown in Fig. 6.13. The impact of the magnetic field of the Virgo cluster is negligible for the map above  $8 \text{ EeV}$ . This is mainly due to the fact that the angular scale used to smooth the map is larger. The remaining structures are the Laniakea and Perseus-Pisces super clusters, which host magnetized structures that will be modeled in the more advanced work ongoing in the group.



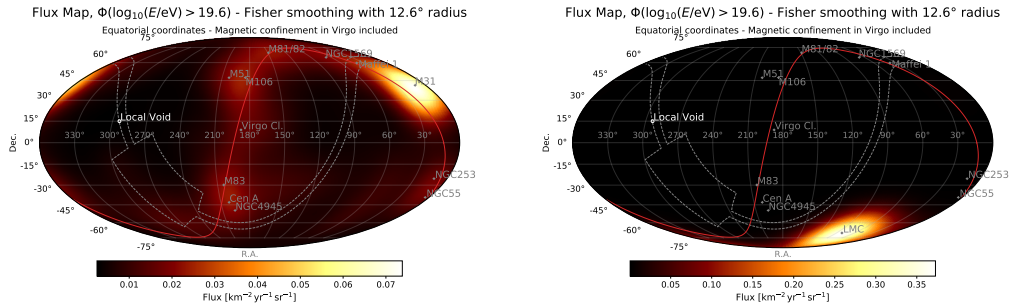
**Fig. 6.13.:** Sky map of the expected flux above  $E_{\text{th}} = 10^{18.9} \text{ eV}$  considering Virgo’s magnetic field spreading over a radial extent  $R_{\text{Cl}}$ .

## 6.3 Below 1 Mpc?

Up to this point, a minimum distance of injection, taken as  $D_{\min} = 1$  Mpc, has been chosen without any justification apart from the divergence of the SMD and SFRD at small distance. In the followings pages, this choice is justified and its implication on UHECR source candidates are explored.

### 6.3.1 Impact of galaxies in the Local Group

In order to study the impact of galaxies below 1 Mpc, two minimum distances of injection  $D_{\min}$  are tested. Fig. 6.14 shows the sky maps expected for two minimum distances of injection  $D_{\min} = 0.5$  Mpc and  $D_{\min} = 0.0$  Mpc, in the SFRD scenario, knowing that the SMD scenario is qualitatively similar. As shown in Fig. 6.14, if the minimum distance is taken to be  $D_{\min} = 0.5$  Mpc, the sky map is dominated by M31, also known as the Andromeda Galaxy, which is located at  $\sim 750$  kpc. If the minimum distance is reduced to zero and excluding the Milky Way, the sky map is highly dominated by the Magellanic clouds.



**Fig. 6.14.:** Sky maps of the expected flux above  $E_{\text{th}} = 10^{19.6}$  eV considering two minimum distances of injection  $D_{\min} = 0.5$  Mpc and  $D_{\min} = 0.0$  Mpc.

Another test can be performed by including the Milky Way inside the model. To simplify the approach, the Milky Way is placed as a point source at  $D = 8$  kpc which is the distance between the Sun and the centre of the Galaxy. When doing so, the sky map is dominated by the contribution of the Galaxy located in the galactic center. The following section describes a scenario that could explain why very nearby sources are not seen in arrival directions. From now on,  $D_{\min}$  is kept at zero.

### 6.3.2 Transient events and time spread induced by magnetic environments

In this section, source candidates arising from exploding star in particular core-collapse supernovae, are tested. To first order, the rate of those objects follows the birthrate of stars (SFRD), therefore, the SFRD scenario is considered here.

#### Introduction

The hypothesis presented above works for steady sources, which produce UHECR continuously. From now on, transient scenarios are considered. A transient scenario produces UHECR through bursts of duration  $\delta t$ . If there is no intervening magnetic field, each burst is observable over a duration,  $\delta t$ . However, due to the presence of magnetic fields, the bursts are spread in time and an observer would be able to see them during a period of time  $\Delta\tau$  larger than the initial  $\delta t$ . If the time  $\Delta\tau$  is small, the number of bursts per unit time must be important in order to see the sources as continuous stream in UHECR data. In this section, we compute the time  $\Delta\tau$  and look at the consequences of a transient scenario on the UHECR sky.

To compute the duration of a burst  $\Delta\tau$ , one can take Eq. 6.28 and compute the standard deviation of the position at a time  $s = ct$ :

$$\Delta\tau = \sqrt{\langle r^2 \rangle - \langle r \rangle^2} \quad (6.32)$$

As before, the full computation is shown in Appendix B and final results give:

$$c\Delta\tau = \frac{\sqrt{2}\lambda_B (ZecB)^2}{9 E^2} D^2 . \quad (6.33)$$

One can see that the duration of the burst is a factor  $\sqrt{2}$  larger than the time delay (Eq. 6.29).

$$c\Delta\tau \simeq 4.4 \times 10^3 \left( \frac{B}{10 \text{ nG}} \right)^2 \left( \frac{R}{10 \text{ EV}} \right)^{-2} \left( \frac{D}{1 \text{ Mpc}} \right)^2 \left( \frac{\lambda_B}{10 \text{ kpc}} \right) \text{ yrs.} \quad (6.34)$$

One can also compute the R.M.S deviation angle,  $\theta_{\text{RMS}} = \sqrt{\langle \theta^2 \rangle}$ .

$$\theta_{\text{RMS}} = \frac{2 (ZecB)}{3 E} \sqrt{\lambda_B D} \quad (6.35)$$

Finally,

$$\theta_{\text{RMS}} \simeq 3.4^\circ \left( \frac{B}{10 \text{ nG}} \right) \left( \frac{R}{10 \text{ EV}} \right) \left( \frac{D}{1 \text{ Mpc}} \right)^{0.5} \left( \frac{\lambda_{\text{B}}}{10 \text{ kpc}} \right)^{0.5}. \quad (6.36)$$

## Magnetic field

Now that the duration of the burst is known, the relevant question is: Which magnetic field impacts most significantly the propagation of UHECRs? We can list at least three relevant magnetic fields that could play a role on the propagation of UHECRs:

- the Galactic magnetic field [G]:  $B_{\text{G}} = 1 \mu\text{G}$  (Jansson and Farrar, 2012),  $\lambda_{\text{B}_{\text{G}}} = 100 \text{ pc}$  (Jansson and Farrar, 2012),  $L_{\text{max}} = 10 \text{ kpc}$  (size of the galaxy);
- the magnetic field in the Local Sheet [LS]:  $B_{\text{LS}} = [0.5, 25] \text{ nG}$  (cf. below),  $\lambda_{\text{B}_{\text{LS}}} = 10 \text{ kpc}$ ,  $L_{\text{max}} = 1 \text{ Mpc}$  (cf. below);
- the extragalactic magnetic field [EG]:  $B_{\text{EG}} = 0.1 \text{ pG}$  (Bray and Scaife, 2018),  $\lambda_{\text{B}_{\text{EG}}} = 1 \text{ Mpc}$  (Bray and Scaife, 2018),  $L_{\text{EG}}^{\text{max}} = \infty$ .

The extragalactic magnetic-field strength is evaluated between  $O(0.1 \text{ nG})$  and  $O(0.1 \text{ fG})$  (Bray & Scaife, 2018; Vazza et al., 2017). Here, we consider an intermediate value  $B_{\text{EF}} = 0.1 \text{ pG}$ . The upper-limit of  $O(0.1 \text{ nG})$  can be found by reversing an equation similar to Eq. 6.35 with a given the  $\theta_{\text{RMS}}$  seen on the data using the dipole, or intermediate-scale analysis and a similar upper limit is obtained by measuring the polarization of the CMB. The values used for the Galactic magnetic field are taken from one of the most up-to-date model (Jansson & Farrar, 2012). As said before, the only magnetic field robustly constrained in large-scale structures is the magnetic field of Coma. However, the Milky Way is not inside a cluster but inside a sheet, close to its center (McCall, 2014). MHD simulations try to evaluate these magnetic fields (Donnert et al., 2018). Note that those simulations are highly seed-dependent and struggle to reproduce cluster magnetic fields at the  $\mu\text{G}$  level. Therefore, in order to estimate the local magnetic field, we consider a typical size  $L_{\text{max}} = 1 \text{ Mpc}$ , which roughly corresponds to the radius of the Local Group. If one observed a correlation between UHECR arrival directions and a source at a distance  $D$  within an angle  $\theta_{\text{RMS}}$ , Eq. 6.35 can be inverted to obtain:

$$B\sqrt{\lambda_{\text{B}}} \leq \frac{3}{2} \frac{E}{Zec} \frac{1}{\sqrt{D}} \theta_{\text{RMS}}. \quad (6.37)$$



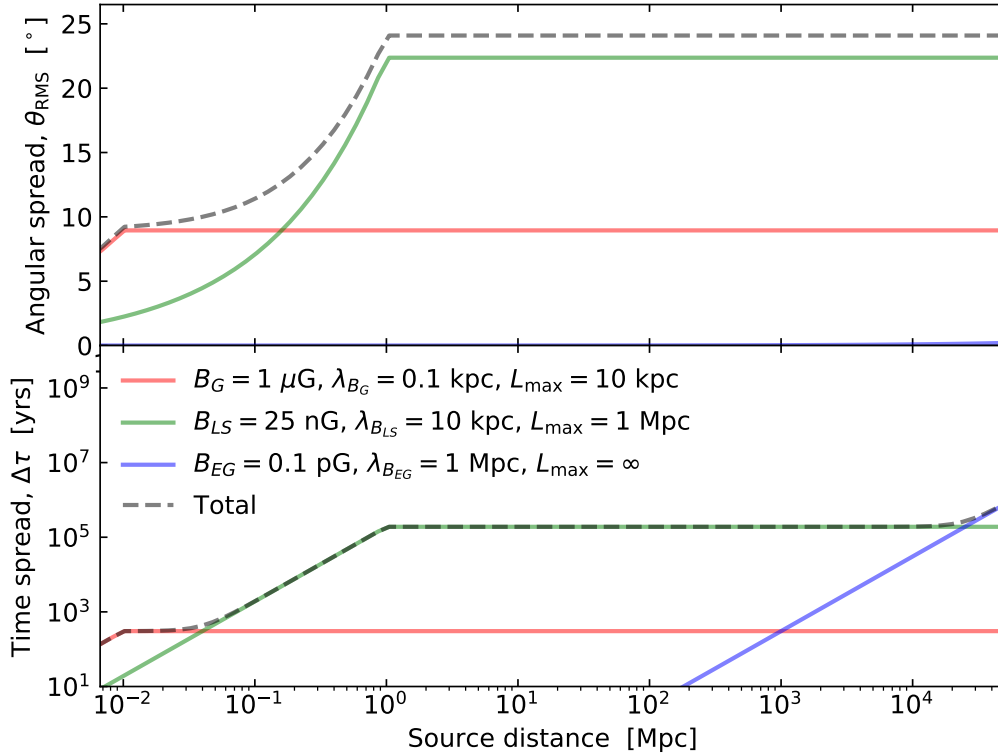
As shown in Chap. 3, The Pierre Auger collaboration, 2022 finds evidence for a match between starburst galaxies and UHECRs within search radius of  $\theta \in [11^\circ, 23^\circ]$ . This search angle is obtained with data above 38 EeV, and looking at the composition study done in Chap. 4, one can deduce the average charge of particles above 38 EeV,  $Z \sim 10$  (EPOS-LHC) and  $Z \sim 12$  (Sibyll2.3d). Since all the sources considered in The Pierre Auger collaboration, 2022 are above  $L_{\max} = 1$  Mpc, the length of the considered magnetic field, one can obtain:

$$B \left( \frac{\lambda_{B_{LS}}}{10 \text{ kpc}} \right)^{0.5} \lesssim 25 \times \left( \frac{Z}{10} \right)^{-1} \text{ nG.} \quad (6.38)$$

Therefore, we are able to constrain the product  $B \left( \frac{\lambda_B}{10 \text{ kpc}} \right)^{0.5}$  which is required and compute the time spread (Eq. 6.34). Three values are tested for the product  $B \left( \frac{\lambda_B}{10 \text{ kpc}} \right)^{0.5}$ :

- $B_{LS} = 0.5$  nG, and  $\lambda_{B_{LS}} = 10$  kpc. In this scenario, the magnetic field of the Local Sheet is negligible, since its impact is smaller than the impact of Galactic magnetic field. This scenario means that the pattern evidenced above  $\sim 40$  EeV is due to the Galactic magnetic field only.
- $B_{LS} = 10$  nG, and  $\lambda_{B_{LS}} = 10$  kpc. In this scenario, the magnetic field of the Local Sheet has a significant impact on the UHECR sources within 10 Mpc. This scenario means that the pattern evidenced above  $\sim 40$  EeV is equally due to the Galactic magnetic field and Local Sheet magnetic field.
- $B_{LS} = 25$  nG, and  $\lambda_{B_{LS}} = 10$  kpc. In this scenario, the magnetic field of the Local Sheet is at its upper limit, and the deflection of UHECR from the sources below 100 Mpc are dominated by this magnetic field. This scenario means that the pattern evidenced above  $\sim 40$  EeV is due to the local-sheet magnetic field only.

The angular spread and time spread induced by the three magnetic fields (Galactic, Local Sheet, extragalactic) are shown in Fig. 6.15 for a rigidity of  $\sim 4$  EV, which is close to the maximum rigidity inferred in the previous chapter.



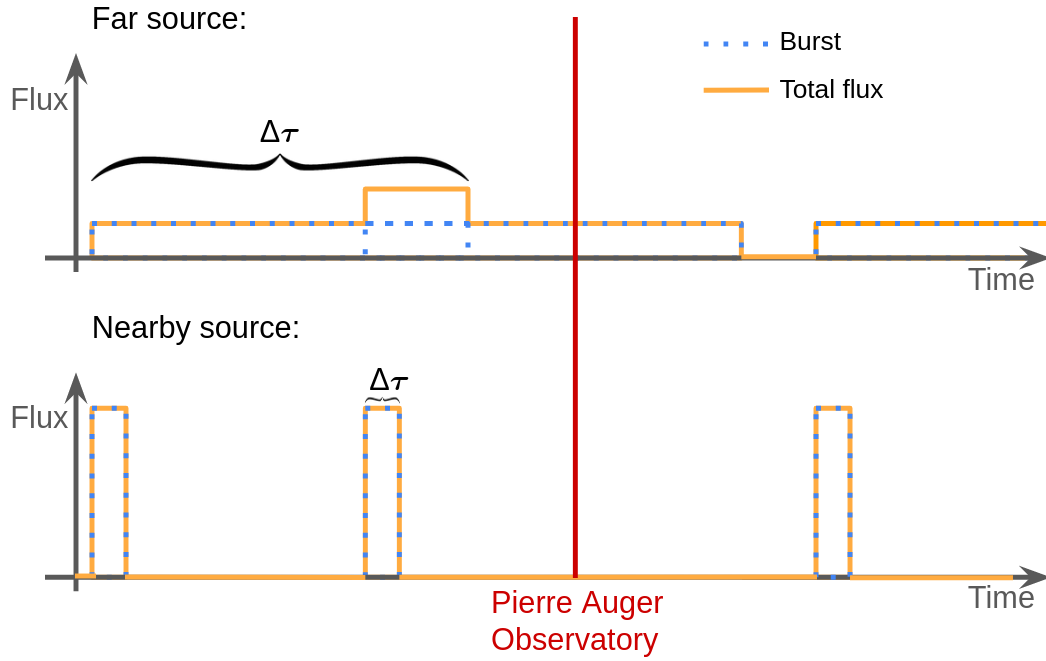
**Fig. 6.15.:** Angular and time spread as a function of the distance of the source for a rigidity of  $R \sim 4 \text{ EV}$ .

### Transient scenario

Due to  $\Delta\tau$  increasing with distance from Earth (as shown in Fig. 6.15), an UHECR burst will last longer for a far source compared to a nearby source. Fig. 6.16 shows two examples of bursts for a far and nearby source. The duration of a burst is given by the blue dots, the total flux is given in orange. Now, the Pierre Auger observatory has been running for roughly 15 years. 15 years is very small compared to the order of  $\sim 10^5$  years seen in Fig. 6.15. This duration is shown as a vertical red line in the figure.

Looking at Fig. 6.16, one can see that the probability to see a nearby source in a transient scenario is small compared to a far source. In this section, we will try to quantify the expected sky in such scenario.

Starting from Eq. 6.12, one can see that the term  $k \times \text{SFR}(z_{\text{Gal}})$  gives the number of bursts per year for a given source  $[\# \text{ yr}^{-1}]$ . Then, for a given rigidity,  $E/Z$ , each burst is viewable from Earth during a time  $\Delta\tau(z_{\text{Gal}}, E)$ . Since the time of the observation



**Fig. 6.16.:** Example of the impact of the magnetic spread in time in a transient scenario where UHECR are produced in bursts.

is small compared to  $\Delta\tau(t)$ , the expected number of bursts seen by an observer follows a Poisson distribution of parameter  $\lambda(z_{\text{Gal}}, E/Z)$ :

$$\lambda_{\text{Gal}}(E/Z) = k \text{SFR}_{\text{Gal}} \Delta\tau(z_{\text{Gal}}, E/Z). \quad (6.39)$$

Then Eq. 6.12 is multiplied and divided by  $\Delta\tau(z_{\text{Gal}})$  to make the term  $\lambda_{\text{Gal}}$  appear,

$$\mathcal{N}(E \geq E_{\text{th}})_{\text{Gal}} = \frac{1}{d_L^2} \frac{1}{4\pi} \sum_A \sum_{A_g} \sum_{E_g} \mathcal{E}_{A_g} \frac{F(E_g)}{E_g} \frac{\lambda_{\text{Gal}}}{\Delta\tau(z_{\text{Gal}})} T(E, A|E_g, A_g, z) \Delta E_g. \quad (6.40)$$

Then, for each galaxy,  $\lambda_{\text{Gal}}$  is replaced by a randomly drawn value from a Poisson distribution of parameter  $\lambda_{\text{Gal}}$ ,  $\lambda_{\text{Gal}} \rightarrow \lambda_{\text{Gal}}^{\text{MC}}$ . This is done for each galaxy and we obtain a mock map that represents one realisation of the sky in that particular scenario.

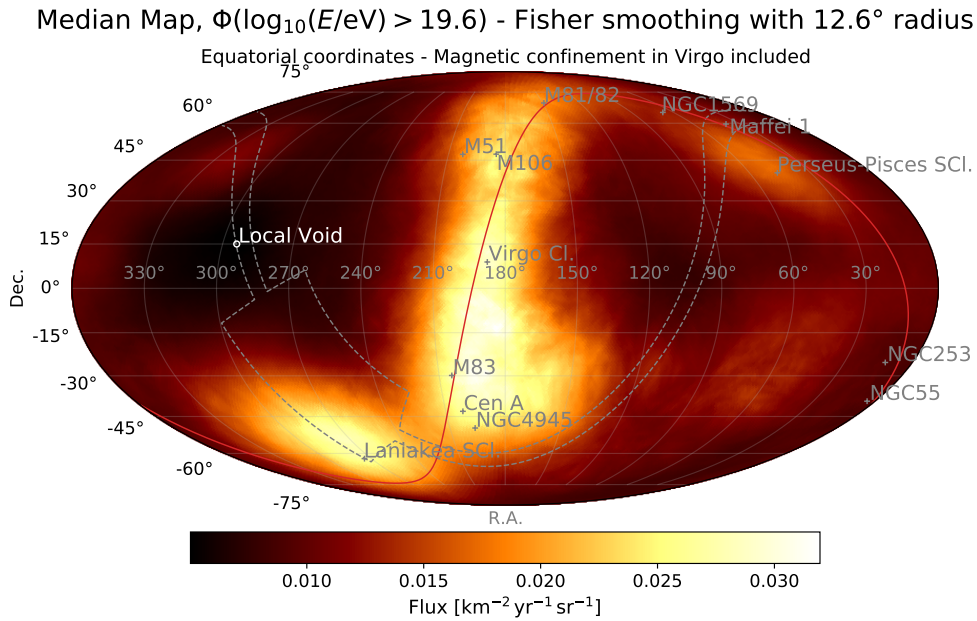
### 6.3.3 Results and discussion

A total of 100 mock maps are created for a given  $k$ . Taking the average of the mock would result in a scenario with steady sources, since the average of a Poisson

distribution is the parameter  $\lambda_{\text{Gal}}$ . Therefore, for each pixel, we take the median value of the 100 distributions.

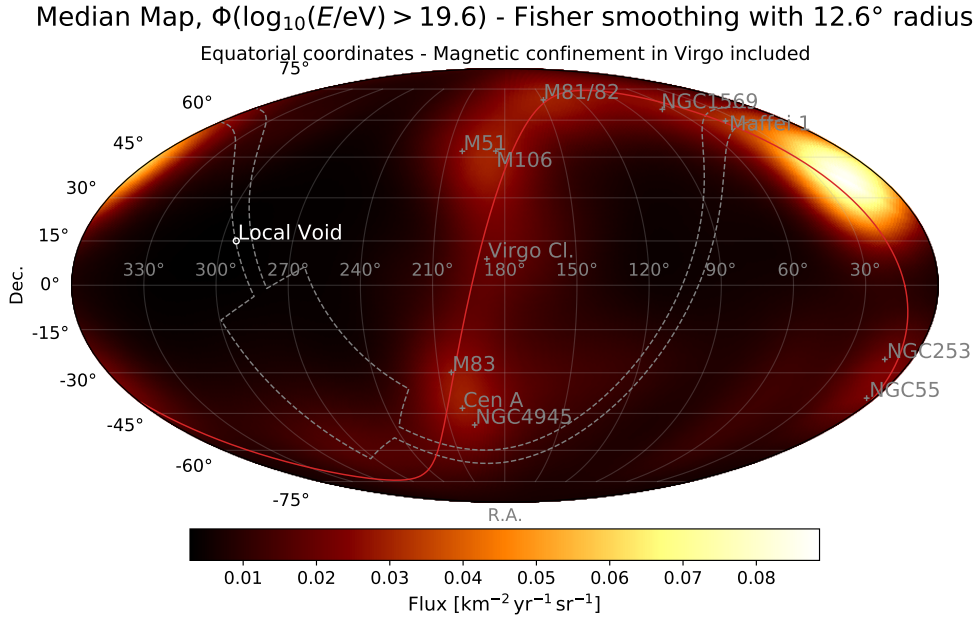
**Case**  $B_{\text{LS}} = 25 \text{ nG}$

We assume a magnetic field of  $B_{\text{LS}} = 25 \text{ nG}$ , which corresponds to the highest value possible for the Local Sheet. A first map is obtained with  $k = 10^{-7} M_{\odot}^{-1}$  and the result is shown in Fig. 6.17.



**Fig. 6.17.:** Sky map of the expected flux above  $E_{\text{th}} = 10^{19.6} \text{ eV}$  considering  $k = 10^{-7} M_{\odot}^{-1}$ .

Fig. 6.17 shows that the low value of  $k = 10^{-7} M_{\odot}^{-1}$  acts as a filter and removes the contribution of very nearby galaxy (Milky Way, Magellanic clouds, Andromeda) as expected from Fig. 6.16. Furthermore, the low value of  $k$  also reduces the contribution of M83/82, M51/106, Maffei 1, NGC 253/55 and results in a significant contribution in the direction of the Virgo Cluster and south of the Galactic Center. This second zone corresponds to part of Laniakea, the supercluster to which the Milky Way belongs. This sky map does not look like the data seen by Auger and TA shown in Fig. 6.6. Therefore, using the value  $k = 10^{-7} M_{\odot}^{-1}$  is too low and turns off regions of the sky which probably contribute to the UHECR flux. Another test is done with a higher value,  $k = 10^{-4} M_{\odot}^{-1}$  and the result is shown in Fig. 6.18. Fig. 6.18 shows that,  $k = 10^{-4} M_{\odot}^{-1}$  is so high that the main contribution would come from Andromeda galaxy which is also not seen in the data (Fig. 6.6). In such scenario, the value of  $k$  is between those two extreme values, and that is what we



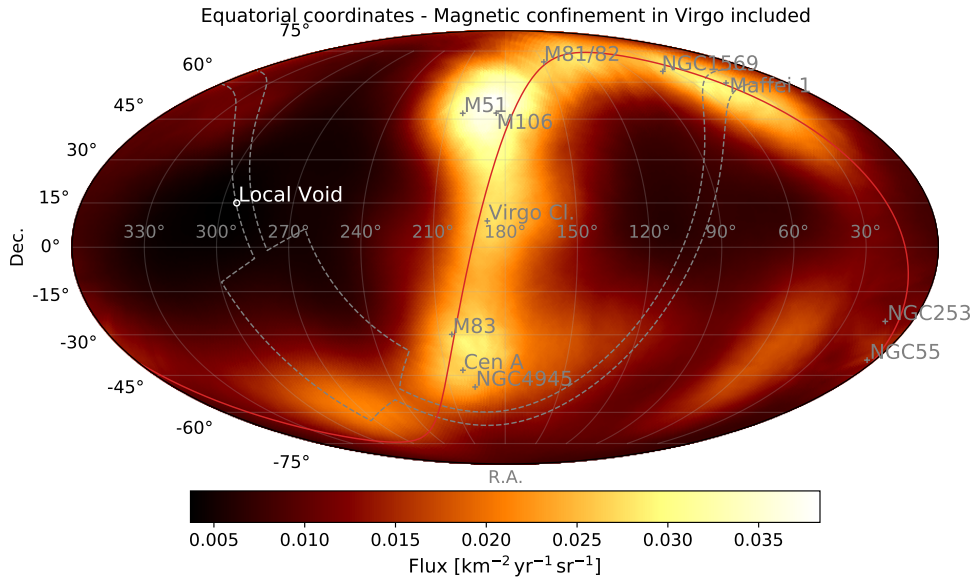
**Fig. 6.18.:** Sky map of the expected flux above  $E_{\text{th}} = 10^{19.6}$  eV considering  $k = 10^{-4} M_{\odot}^{-1}$ .

are planning to find. In order to find the suitable value of  $k$ , a scan is performed over  $k = [1 \times 10^{-7}, 3 \times 10^{-7}, 1 \times 10^{-6}, 3 \times 10^{-6}, 1 \times 10^{-5}, 3 \times 10^{-5}, 1 \times 10^{-4}, 3 \times 10^{-4}] M_{\odot}^{-1}$ . This choice is motivated by the equal logarithmic space between each of those values. From this scan, we can deduce a value  $k_{\text{min}}$  which corresponds to the minimum value needed to see the Council of Giants (McCall, 2014), composed of Centaurus A, NGC4945, M83, NGC253, among other galaxies. On the other side, one can also deduce a value  $k_{\text{max}}$  which correspond to the value where local galaxies such as Andromeda galaxy, the Magellanic clouds or the Milky Way would completely dominate the UHECR sky as shown in Fig. 6.18.

The minimum value found is  $k_{\text{min}} = 3 \times 10^{-6} M_{\odot}^{-1}$ , where there is a major contribution of M51/M106, from the Centaurus region and from Maffei 1. The tested value below  $k_{\text{min}}$  (i.e.  $k = 1 \times 10^{-6} M_{\odot}^{-1}$ ) results in a contribution dominated by M51/M106 together with Virgo, and the Laniakea region is as bright as the Centaurus region. Such sky map is disfavoured due to the difference with the data. On the other-side, the maximum value found is  $k_{\text{max}} = 1 \times 10^{-5} M_{\odot}^{-1}$ . Above, this value (i.e.  $k = 3 \times 10^{-5} M_{\odot}^{-1}$ ), the sky map is fully dominated by Andromeda. The two “acceptable” maps regarding the data are shown in Fig. 6.19 for  $k = k_{\text{min}}$ , and Fig. 6.20  $k = k_{\text{max}}$ .

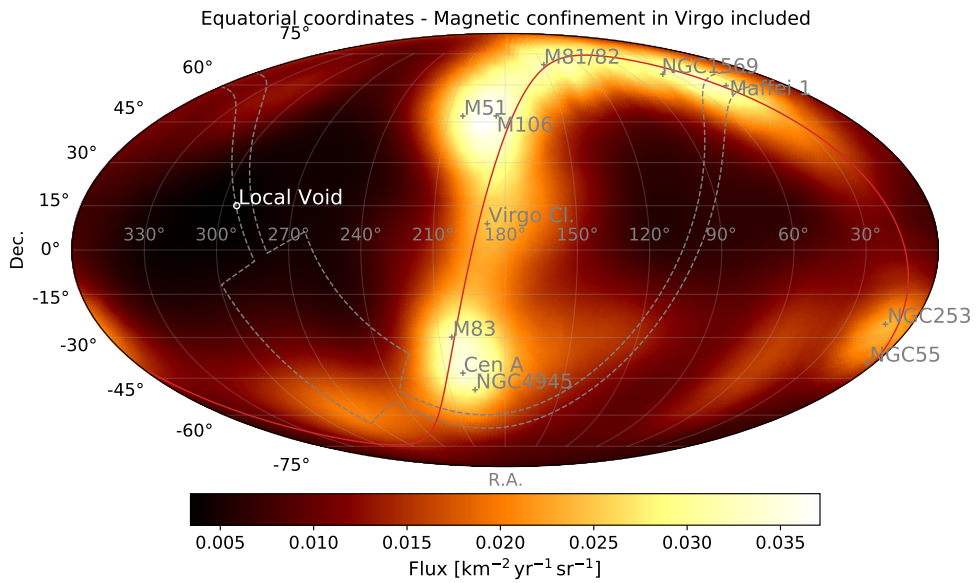
Fig. 6.19 and Fig. 6.20, are rather similar. The major differences come from the impact of M81/82 and NGC253/NGC55. It shows us that, considering such value of  $k$ , those galaxies may or may not have been bursting in the recent past.

Median Map,  $\Phi(\log_{10}(E/eV) > 19.6)$  - Fisher smoothing with  $12.6^\circ$  radius



**Fig. 6.19.:** Sky map of the expected flux above  $E_{\text{th}} = 10^{19.6}$  eV considering  $k = 3 \times 10^{-6} M_{\odot}^{-1}$ .

Median Map,  $\Phi(\log_{10}(E/eV) > 19.6)$  - Fisher smoothing with  $12.6^\circ$  radius



**Fig. 6.20.:** Sky map of the expected flux above  $E_{\text{th}} = 10^{19.6}$  eV considering  $k = 10^{-5} M_{\odot}^{-1}$ .

### Case $B_{\text{LS}} = 10$ nG

From now on, the Local Sheet magnetic field is reduced to 10 nG. The analysis is re-run in the same condition as before. Once again, one can find new limits for  $k$ .

The reduced value of  $B_{LS}$ , results in bigger limits, with  $k_{\min} = 3 \times 10^{-6} M_{\odot}^{-1}$  and  $k_{\max} = 1 \times 10^{-5} M_{\odot}^{-1}$ . The maps observed are similar to the ones shown before. Here, lowering the magnetic field only shift a bit the value of  $k_{\min}$  and  $k_{\max}$ .

### Case $B_{LS} = 0.5$ nG

When considering  $B_{LS} = 0.5$  nG, it is not possible to define a  $k_{\min}$  and a  $k_{\max}$ . Indeed, since only the Galactic magnetic field impacts the propagation, UHECRs coming from very nearby source (Magellanic cloud, Andromeda galaxy...) or from the Council of Giants experience the same magnetic time spread. Therefore, there are two possibilities. Either the sources have a low burst-rate and nearby sources are not viewable in the UHECR sky, so that the sky map is dominated by far away sources (no Council of Giants), or the burst-rate is high enough and the UHECR sky is dominated by the Magellanic clouds and Andromeda. It means that if the sources are transients that follow the evolution of stars, and the extragalactic magnetic field is small enough, the Local Sheet magnetic field needs to be larger than 0.5 nG in order to explain the UHECR arrival direction data. Such value matches the results from MHD simulation, despite the large uncertainty from one simulation to another ( $0.1 \text{ nG} < B_{LS}^{\text{MHD}} < 100 \text{ nG}$  from Vazza et al., 2017). Moreover, the value of  $B_{LS} \sim 10$  nG is characteristic of primordial origin, suggesting that the magnetic field were present at the early age of the universe with weak strength and got amplified in astrophysical structures.

### Constraint on $k$ in a SMD scenario

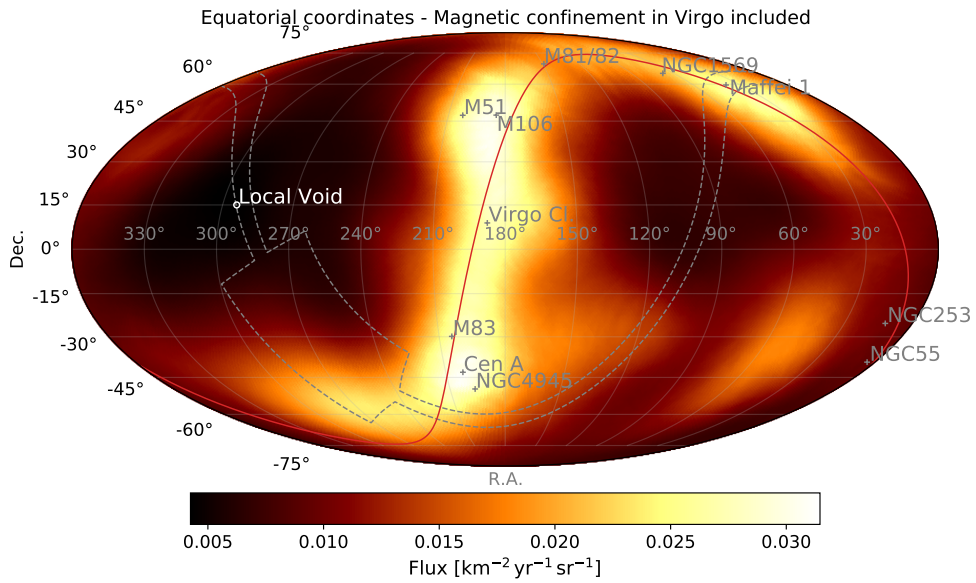
The analysis is done in the same way for the SMD scenario. A total of 100 mock maps are created for a given  $k$ . Then, for each pixel, we take the median value of the 100 distributions. Note that since we are in the SMD scenario here, the units of  $k$  are now  $[k] = M_{\odot}^{-1} \text{ yr}^{-1}$ . The minimum  $k_{\min}$  is taken from  $B_{LS} = 25$  nG, which correspond to the maximum magnetic value possible for the local sheet. The minimum is found to be  $k_{\min} = 10^{-16} M_{\odot}^{-1} \text{ yr}^{-1}$  and the map is shown in Fig. 6.21.

The maximum  $k_{\max}$  is taken from  $B_{LS} = 10$  nG. The maximum is found to be  $k_{\max} = 3 \times 10^{-15} M_{\odot}^{-1} \text{ yr}^{-1}$  and the map is shown in Fig. 6.22.

One can conclude that in a SMD scenario  $k \in [10^{-16}, 3 \times 10^{-15}] M_{\odot}^{-1} \text{ yr}^{-1}$ .

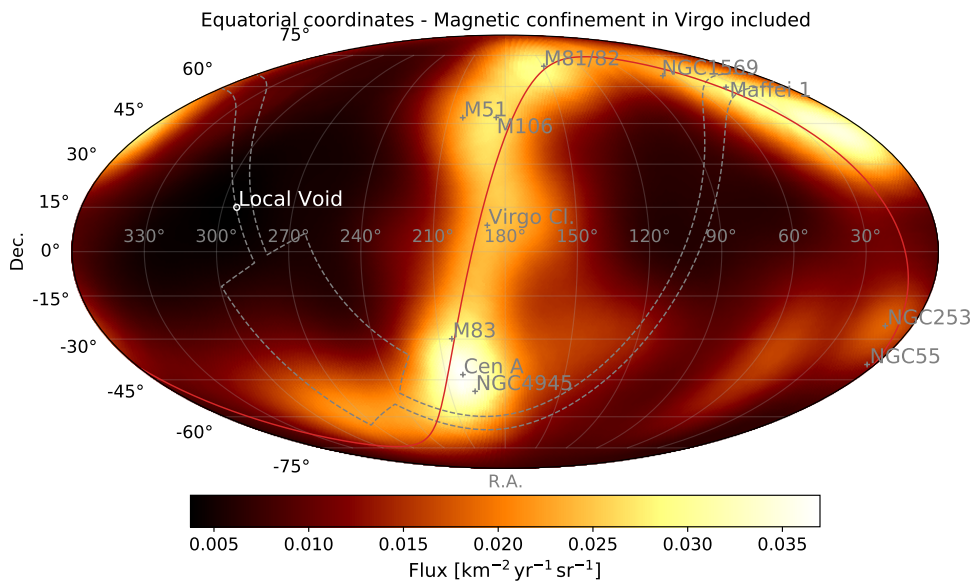


Median Map,  $\Phi(\log_{10}(E/\text{eV}) > 19.6)$  - Fisher smoothing with  $12.6^\circ$  radius



**Fig. 6.21.:** Sky map of the expected flux above  $E_{\text{th}} = 10^{19.6} \text{ eV}$  considering  $k = 10^{-16} M_{\odot}^{-1} \text{ yr}^{-1}$  and  $B_{\text{LS}} = 25 \text{ nG}$ .

Median Map,  $\Phi(\log_{10}(E/\text{eV}) > 19.6)$  - Fisher smoothing with  $12.6^\circ$  radius

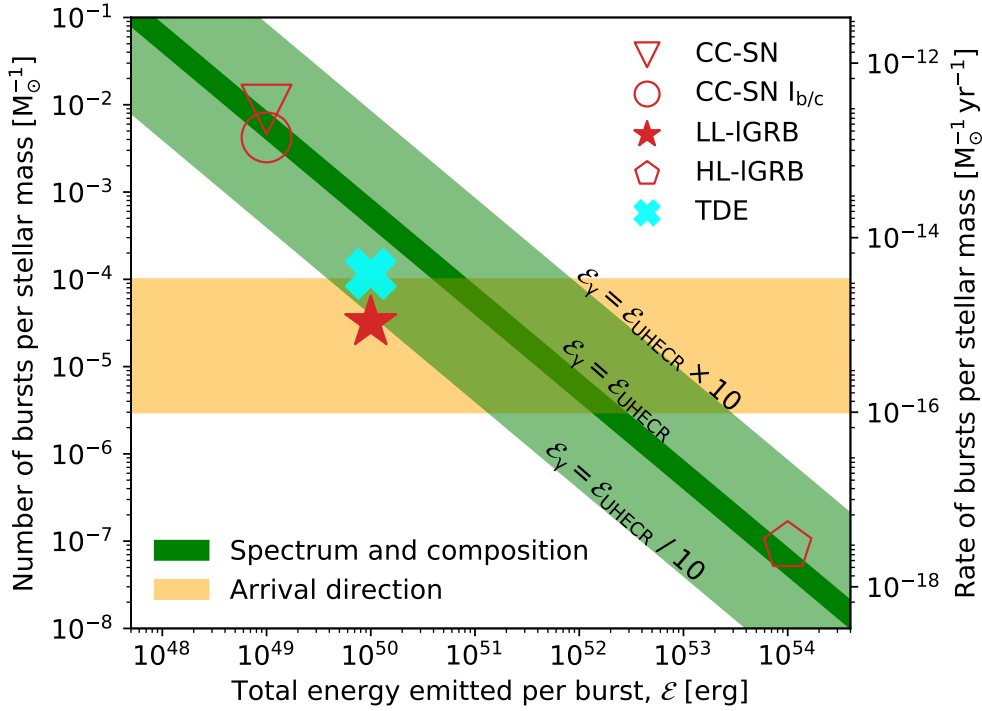


**Fig. 6.22.:** Sky map of the expected flux above  $E_{\text{th}} = 10^{19.6} \text{ eV}$  considering  $k = 3 \times 10^{-15} M_{\odot}^{-1} \text{ yr}^{-1}$  and  $B_{\text{LS}} = 10 \text{ nG}$ .



## Constraining the sources

Taking  $k_{\min}$  from  $B_{LS} = 25$  nG and  $k_{\max}$  from  $B_{LS} = 10$  nG together with the best-fit values of  $\mathcal{E}_{\text{tot}} \times k$ , one can put constraints on the sources of UHECRs. Those constraints are shown as orange and green bands in Fig. 6.23.



**Fig. 6.23.:** Constraints on  $k$  and on  $\mathcal{E}_{\text{tot}}$  in a SFRD scenario (left axis) and SMD scenario (right axis). Source candidates are shown as markers (red for those who follow the SFRD and blue for those who follow the SMD).

The right axis of Fig. 6.23 is obtained by performing the same analysis in the SMD scenario and by computing the factor  $A = \frac{k_{\min}^{\text{SMD}}}{k_{\min}^{\text{SFRD}}}$ . The green band is obtained using  $\mathcal{E}_{\text{tot}} \times k$  from the SMD scenario for the upper band and from the SFRD scenario for the minimum.

Multiple candidates are shown in Fig. 6.23:

- Core Collapse Supernova (CC-SN);
- CC-SN of type I<sub>b/c</sub> ;
- Low luminosity Long Gamma Ray Burst (LL-IGRB);
- High luminosity Long Gamma Ray Burst (HL-IGRB);

- Tidal Disruption Event (TDE).

The first one, CC-SN; regroups supernovae from the deaths of massive stars (above  $\sim 8 M_{\odot}$ ), which at first order, follows the SFRD. CC-SN of type  $I_{b/c}$  is a sub-group of CC-SN which is hydrogen-poor and is believed to be progenitors of long gamma ray bursts (LGRB). Then, from LGRB, two distinct populations of GRBs have been found, the low-luminosity and high luminosity ones (Liang et al., 2007). The isotropic photon emission of CC-SN and CC-SN  $I_{b/c}$ , is taken to be  $10^{53}$  erg (Smartt, 2009), with a local rate given by Madau and Dickinson, 2014 for the former, and Liu et al., 2011 for the latter. For LL-IGRB, we take an energy of  $10^{50}$  erg as it is done in Liu et al., 2011, which corresponds to the isotropic gamma-ray energy found in nearby LL-IGRB (GRB 060218 and GRB 100316D). The local rate is given in Liu et al., 2011 as well. For HL-IGRB, the energy is taken to be of  $10^{54}$  erg, it corresponds to multiplying the average luminosity of  $10^{52}$  erg/s found in Liu et al., 2011 with a characteristic time  $t_{90} = 100$  s, the local rate is also found in Liu et al., 2011.

On the other side, constraining the sources in a SMD scenario is more difficult. This is due to the fact that while SMD is a good tracer of intermediate luminosity AGN, the period of activity or non-activity of AGN is not well known. However, TDEs could follow the SMD. TDEs are transient astronomical phenomena that occur when a star approaches a black hole and is torn apart by the tidal forces of the black hole. The event rate of TDE and the energy is shown in Fig. 6.23. TDE releases around  $10^{50}$  erg in electromagnetic radiation (Lodato & Rossi, 2011), the local rate is taken from Khabibullin and Sazonov, 2014; van Velzen, 2018.

Regarding Fig. 6.23, the best candidates are LL-IGRB for the SFRD scenario. Studies have been done in favour or not of LL-IGRB as sources of UHECRs (Samuelsson et al., 2020; Zhang et al., 2018). For instance, Zhang et al., 2018 argues that LL-IGRB are better sources than HL-IGRB for accelerating nuclei (as seen in the data). Indeed, due to the intense photon field in the HL-IGRB jet, the nuclei are easily photodisintegrated in the HL-IGRB case compared to the LL-IGRB case. Moreover, the resulting injection spectrum expected in LL-IGRB is dominated by CNO and Si, without proton and iron, which is close to what the data shown in Chap. 5 (80% of CNO emissivity for the SFRD case). On the other side, Samuelsson et al., 2020 shows that taking the properties of GRB 060218 and comparing electron synchrotron radiation with observations excludes LL-IGRB as GRB 060218 to be the main sources of UHECRs. They disfavour such sources considering the phases of the bursts. The afterglow phase is excluded considering mildly relativistic outflows, and showing that the limited available energy in LL-IGRB is not sufficient. The prompt phase is excluded assuming that the prompt energy budget is comparable to the

afterglow kinetic energy. Although, one can highlight the fact that LL-IGRB are still poorly known and that the argument of Samuelsson et al., 2020 is based on a single GRB. One limitation of the current understanding is due to the low number of detected LL-IGRB due to their low emissivity, which can barely be accessed by current observatories. Therefore, a better understanding of LL-IGRB is key to determine if they are able to explain the UHECR data.

In the SMD scenario, TDE seems to be a good candidate for accelerating UHECRs. Two kinds of TDEs are detected, jetted TDEs and non-jetted TDEs. For now, a few jetted TDEs have been detected. Therefore, due to the low statistics, we do not consider them as a distinct population of sources. Multiple models have attempted to accelerate UHECRs using jetted-TDE (Alves Batista & Silk, 2017; Farrar, 2012; Guépin et al., 2018; Zhang et al., 2017). Zhang et al., 2017 shows that UHECR nuclei would survive external, reverse and forward shocks, while it is harder for them to survive internal shock apart from internal shock from low-luminosity TDE jets. They also show that TDE initiated by oxygen-neon-magnesium white dwarfs could explain the spectrum and the  $X_{\max}$  moments, although those kinds of white dwarfs may be too rare in order to explain the observed flux of UHECRs. Starting from the spectral energy distributions from two known TDEs and considering a thin accretion disk model, Farrar, 2012 manage to estimate the bolometric energy of the accretion disk. They show that the bolometric energy of the accretion disk is sufficient to accelerate low- $Z$  nuclei. Guépin et al., 2018 consider a simplified model where TDEs are split into two stages. One of short duration with a bright luminosity and a second one, long-lasting with medium luminosity. They show that their best-fit model manages to reproduce the spectrum and the  $X_{\max}$  moments from the Pierre Auger Observatory. While, the previous models always consider supermassive black holes, Alves Batista and Silk, 2017 consider the tidal disruption of a white dwarf with an intermediate-mass black hole. They show that such events naturally give an intermediate composition for UHECRs as observed and they suggest that globular clusters and dwarf galaxies could be interesting sites for accelerating UHECRs. Therefore, TDE is an interesting candidate for accelerating UHECRs.

The study presented here considers a transient scenario with sources that follow to first order the SFRD and SMD. A similar study could be done with a refined distribution of sources or considering other kind of transients that have not been studied here (AGN flares for instance). Moreover, we wish to highlight the fact that the approach used here is a qualitative one. In the continuity of this thesis, a more quantitative work is in progress in order to better constrain  $k$  through a statistical analysis against arrival-direction data. Finally, this study does not allow to decide between the SMD and SFRD scenarios. However, several aspects can be

explored to distinguish between the two approaches. If taking into account the magnetic fields from superclusters and clusters allows the dipole to be reproduced, then comparing the amplitude from the model with the measured amplitude could allow the two scenarios to be distinguished. On the other hand, Fig. 6.11 shows that protons escape from clusters more easily than nuclei. In the SMD scenario, protons detected at energies close to the ankle are produced nearby. In the SFRD scenario, the protons are secondary protons from nuclei injected further away. It would be interesting to look at the number of protons as a function of redshift  $z$  and to determine which clusters/superclusters they cross. If the number of protons produced behind the clusters/superclusters is larger in the SFRD scenario compared to the SMD scenario, one would expect to detect more protons than nuclei in the direction of the clusters/superclusters. Indeed, the clusters/superclusters could act as a filter, reducing the proportion of heavy nuclei and consequently increasing the fraction of protons in their direction. If this turns out to be the case, then this approach could explain the structures observed in the  $X_{\max}$  sky maps above the ankle, which could originate from clusters/superclusters located at some hundreds of Mpc (Mayotte, E. on behalf of the Pierre Auger Collaboration, 2021). Considering that the protons are not produced at the same redshift, the production of neutrinos associated to the propagation of protons should also be different in the two scenarios. Therefore, looking at the neutrinos could be another way to distinguish the two approaches.

## 6.4 Conclusion

The model described in this chapter manages to produce arrival-direction sky maps that can be compared to the data. The two approaches tested in Chap. 5 (i.e. SFRD and SMD scenarios) result in similar outcomes apart from a larger contrast for the SMD scenario due to the smaller flux from distant sources. The model sky map obtained in the SFRD and SMD scenario is dominated by starforming galaxies (as in The Pierre Auger collaboration, 2022). The Virgo cluster is the dominant contribution in those scenarios, although it does not appear in the data. The non presence of Virgo could be due to the cluster's magnetic field that traps UHECRs. A simplified model is explored to take into account such trapping. Finally, the transientness of sources of UHECRs is taken into account. Considering local magnetic fields, transient sources could explain why no contributions are seen from local galaxies as the Magellanic clouds or the Milky Way. From the presence of the Council of Giants (i.e. sources between 1 and 10 Mpc) and the non presence

of very nearby sources in the arrival directions data, we are able to constrain the burst rate from  $k_{\min} = 3 \times 10^{-6} M_{\odot}^{-1}$  to  $k_{\max} = 10^{-4} M_{\odot}^{-1}$  in the SFRD scenario and  $k_{\min} = 10^{-16} M_{\odot}^{-1} \text{yr}^{-1}$  to  $k_{\max} = 3 \times 10^{-15} M_{\odot}^{-1} \text{yr}^{-1}$  in a SMD scenario, depending on the magnetic field considered. This scenario also assumes that the magnetic field associated to the Local Sheet is at least of the orders of few nG. Together with the constraints put on the emissivities from Chap. 5, we are able to construct a diagram comparing the sources that follow the SFRD (i.e. exploding stars) or the SMD. While, CC-SN and CC-SN  $I_{b/c}$  reach the required emissivities in gammas, they are too frequent to explain the non-presence of Local Group galaxies. On the other side, HL-IGRB are not frequent enough to explain the presence of the Council of Giants. Finally, LL-IGRB match the two requirements, the burst-rate and the emissivity. We also show that, when considering a SMD scenario, TDEs also match the two requirements as well. LL-IGRB and TDEs represent the best candidate up to date as UHECR sources in this transient scenario analysis.

# Conclusion

The origin of UHECRs is a difficult question to tackle, as it lies at the crossroads of many fields of physics. Our knowledge of hadronic models, EBL, nucleus-photon cross-sections, biases on reconstruction are limitations in interpreting the astrophysics of the sources. In this thesis, some of these problems have been studied, including flux map reconstructions and a compositional study, but the majority of the work has been focused on the study of UHECR sources. Using the Combined Fit, a cosmic model that is compared to the compositional and spectral data, we have been able to constrain the injected scenario at the sources. Assuming that the sources follow a Peters cycle (i.e. the composition becomes heavier with increasing energy), we can assume that the protons seen below the ankle are from the same sources as the nuclei above. Taking these protons into account in the Combined Fit results in two possibilities. If the evolution of the sources follows the star formation rate density (SFRD), then the protons detected below the ankle are secondary protons produced *en route* from the interaction of nuclei with the EBL and CMB. If the evolution of the sources follows the stellar mass density, then it is necessary to have a different spectral index for protons at the sources ( $\gamma_p \sim 3$ ). This index is very different from the preferred hard index for data above the ankle ( $\gamma_N < 0$ ). Such a soft index for the protons could come from neutrons escaping from the sources on kpc scales. Indeed, as neutrons are neutral, if they are produced, they could easily escape from the sources and then decay into protons. Moreover, by crossing this astrophysical model with a catalogue of nearly 400,000 galaxies, we are able to produce maps of the sky expected by the models. The maps produced are dominated by the Virgo cluster, while no major excess is seen in the data in that direction. Apart from the Virgo cluster, the proposed model is consistent with the starburst model that was used by the Auger collaboration to announce the evidence for anisotropy above 40 EeV. The hypothesis proposed here to explain the non-presence of Virgo in the data is the confinement of UHECRs in the intracluster medium. A first approach is considered to study the Virgo cluster, by assuming that the time delay induced by a cluster magnetic field should not overcome the ballistic time in order for UHECR to escape. At the same time, a much more robust study is developed within the group, which confirms the findings presented here. From these flux maps, a study of the transient aspect of the sources was carried out. In fact, the effect of magnetic fields (galactic and associated with the local sheet) shows that, in the case of transient

events, nearby sources should not contribute to the UHECR flux. Not observing the very close galaxies (Magellanic clouds, Andromeda galaxy), while observing galaxies in the council of giants (Centaurus region, NGC4945), allows us to put constraints on the burst rate, relative to the SFRD,  $k \in [3 \times 10^{-6}; 10^{-4}] M_{\odot}^{-1}$  or SMD,  $k \in [10^{-16}; 3 \times 10^{-15}] M_{\odot}^{-1} \text{ yr}^{-1}$ . Together with the emissivities, the burst rate allows constraining the candidate sources if the sources are transients. This study seems to show that low luminosity long gamma ray bursts and tidal disruption events are the best candidates to match these conditions.

The analysis presented in this thesis should benefit from the many experimental advances in the coming years. TA has launched the TA $\times$ 4 upgrade, which consists in multiplying by four the area covered by TA for the most energetic EAS, which will thus have a similar sensitivity to that of Auger for the most energetic cosmic rays. This upgrade will allow one to better constrain UHECRs in the Northern Hemisphere. The Auger observatory is currently developing AugerPrime, the upgrade including, among others, scintillators on top of the water tanks (see Chap. 2). This upgrade will allow a better understanding of the composition through the measurement of the muon number, which is today a limit in the understanding of the composition of UHECRs. In a more distant and less certain future, the Giant Radio Array for Neutrino Detection (GRAND) would have the sensitivity to observe cosmogenic neutrinos. GRAND would better constrain the flux of these neutrinos and thus the source scenarios. For the study of UHECRs, projects such as the Extreme Universe Space Observatory (EUSO) and the Probe of Extreme Multi-Messenger Astrophysics (POEMMA) are in the prototyping phase. The objective of EUSO and POEMMA is to observe the fluorescence light emitted by atmospheric showers in the manner of FDs from space. EUSO and POEMMA would be the next generation of experiments measuring UHECRs.

The hadronic models used for the composition should benefit from the LHC Run 3 which foresees p-O and O-O collisions around 2024. As these collisions were not realized in the previous runs, they should constrain the hadronic models in the region where they differ, i.e. nuclear collisions. A better understanding of galaxy clusters would allow to better constrain their density profiles and the associated magnetic field. e-ROSITA, an X-ray satellite observing up to 10 keV launched in July 2019, seeks in particular to better understand clusters and their formation. Its future results could help us to better understand the impact of UHECR propagation in clusters. The Square Kilometer Array (SKA) is a giant array of radio telescopes covering 1 km<sup>2</sup>, for which construction of Phase 1 started in 2020. SKA will be of crucial importance to better understand the surrounding magnetic fields through the measurement of Faraday rotation.

Consequently and for all these reasons, the following years look promising, and we may be on the verge of discovering the sources of UHECRs.





# Bibliography

- Achterberg, A., Gallant, Y. A., Norman, C. A., & Melrose, D. B. (1999). Intergalactic Propagation of UHE Cosmic Rays. *19th Texas Symposium on Relativistic Astrophysics and Cosmology* (cit. on pp. 148–150, 201, 203, 207).
- Allard, D., Ave, M., Busca, N., Malkan, M. A., et al. (2006). Cosmogenic neutrinos from the propagation of ultrahigh energy nuclei. *Journal of Cosmology and Astroparticle Physics*, 2006(09) (cit. on pp. 33, 35).
- Aloisio, R., Berezhinsky, V., & Blasi, P. (2014). Ultra high energy cosmic rays: implications of Auger data for source spectra and chemical composition. *Journal of Cosmology and Astroparticle Physics*, 10, 020 (cit. on pp. 121, 132).
- Aloisio, R., Berezhinsky, V., & Grigorieva, S. (2013). Analytic calculations of the spectra of ultra-high energy cosmic ray nuclei. I. The case of CMB radiation. *Astroparticle Physics*, 41 (cit. on p. 34).
- Aloisio, R., Boncioli, D., Matteo, A. d., Grillo, A. F., et al. (2017). SimProp v2r4: Monte Carlo simulation code for UHECR propagation. *2017(11)* (cit. on pp. 102, 140, 152).
- Alves Batista, R., & Saveliev, A. (2021). The Gamma-Ray Window to Intergalactic Magnetism. *Universe*, 7(7), 223 (cit. on p. 146).
- Alves Batista, R., & Silk, J. (2017). Ultrahigh-energy cosmic rays from tidally-ignited white dwarfs. *Physical Review D*, 96(10), Article 103003, 103003 (cit. on p. 168).
- Antoni, T., Apel, W. D., Badea, A. F., Bekk, K., et al. (2005). KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems. *Astroparticle Physics*, 24(1-2) (cit. on p. 19).
- Apel, W. D., Arteaga-Velázquez, J. C., Bekk, K., Bertaina, M., et al. (2013). KASCADE-Grande measurements of energy spectra for elemental groups of cosmic rays. *Astroparticle Physics*, 47 (cit. on p. 19).
- Apel, W. D., Arteaga-Velázquez, J. C., Bekk, K., Bertaina, M., et al. (2011). Kneelike structure in the spectrum of the heavy component of cosmic rays observed with KASCADE-grande. *Physical Review Letters*, 107(17) (cit. on p. 19).
- Arbeletche, L. B., & de Souza, V. (2020). On the parametrization of the distributions of depth of shower maximum of ultra-high energy extensive air showers. *Astroparticle Physics*, 116 (cit. on p. 83).
- Batista, R. A., Biteau, J., Bustamante, M., Dolag, K., et al. (2019). Open Questions in Cosmic-Ray Research at Ultrahigh Energies. *Frontiers in Astronomy and Space Sciences*, 6 (cit. on pp. 26–28).
- Bell, A. R. (1978). The acceleration of cosmic rays in shock fronts - I. *Monthly Notices of the Royal Astronomical Society*, 182, 147–156 (cit. on p. 23).
- Bellido J. on behalf of the Pierre Auger Collaboration (2018). Depth of maximum of air-shower profiles at the Pierre Auger Observatory: Measurements above  $10^{17.2}$  eV and Composition Implications. *Proceeding of Science, 35th International Cosmic Ray Conference*, 301 (cit. on pp. 4, 5, 10, 11, 92, 131).

- Berezinsky, V. (2006). Dip in UHECR Spectrum as Signature of Proton Interactions with CMB (A. I. Studenikin, Ed.), 87 (cit. on pp. 4, 10, 20, 32, 33).
- Biehl, D., Boncioli, D., Fedynitch, A., & Winter, W. (2018). Cosmic-Ray and Neutrino Emission from Gamma-Ray Bursts with a Nuclear Cascade. *Astronomy and Astrophysics*, 611, A101 (cit. on pp. 6, 12, 126, 127, 130, 133).
- Biteau, J., Marafico, S., Kerfis, Y., & Deligny, O. (2022). Cosmographic model of the astroparticle skies. *Proceeding of Science, 37th International Cosmic Ray Conference*, Article 1012 (cit. on p. 135).
- Biteau, J. (2021). Stellar Mass and Star Formation Rate within a Billion Light-years. *The Astrophysical Journal Supplement Series*, 256(1) (cit. on pp. 116, 117, 140).
- Biteau J. on behalf of the Pierre Auger Collaboration (2022). The ultra-high-energy cosmic-ray sky above 32 EeV viewed from the Pierre Auger Observatory. 395 (cit. on pp. 77, 78).
- Bonafede, A., Feretti, L., Murgia, M., Govoni, F., et al. (2010). The Coma cluster magnetic field from Faraday rotation measures. *Astronomy & Astrophysics*, 513(5) (cit. on pp. 146, 148).
- Bonamente, M. (2019). Distribution of the C statistic with applications to the sample mean of Poisson data. *Journal of Applied Statistics*, 47(11) (cit. on p. 87).
- Boncioli, D., Biehl, D., & Winter, W. (2019). On the common origin of cosmic rays across the ankle and diffuse neutrinos at the highest energies from low-luminosity Gamma-Ray Bursts. *The Astrophysical Journal*, 872(1), 110 (cit. on p. 126).
- Bray, J. D., & Scaife, A. M. M. (2018). An Upper Limit on the Strength of the Extragalactic Magnetic Field from Ultra-high-energy Cosmic-Ray Anisotropy. *The Astrophysical Journal*, 861(1) (cit. on pp. 148, 157).
- Bridgeman A. on behalf of the Pierre Auger Collaboration (2018). Shower universality reconstruction of data from the Pierre Auger Observatory and validations with hadronic interaction models. *Proceeding of Science, 36th International Cosmic Ray Conference*, 301 (cit. on p. 80).
- Condorelli, A. (2021). *Study of cosmic-ray interactions in candidate sources. Applications to the interpretation of the energy spectrum and mass composition data of the Pierre Auger Observatory* (Doctoral dissertation). (Cit. on p. 108).
- Crinquand, B., Cerutti, B., Dubus, G., Parfrey, K., & Philippov, A. (2021). Synthetic gamma-ray light curves of Kerr black hole magnetospheric activity from particle-in-cell simulations. *Astronomy and Astrophysics*, 650, A163 (cit. on p. 125).
- de Oliveira, C., & de Souza, V. (2021). Magnetically Induced Anisotropies in the Arrival Directions of Ultra-high-energy Cosmic Rays from Nearby Radio Galaxies. *The Astrophysical Journal*, 925(1) (cit. on p. 146).
- Di Matteo A. on behalf of the Pierre Auger and Telescope Array collaborations (2021). UHECR arrival directions in the latest data from the original Auger and TA surface detectors and nearby galaxies. *Proceeding of Science, 37th International Cosmic Ray Conference*, 304 (cit. on pp. 79, 144).
- Ding, C., Globus, N., & Farrar, G. R. (2021). The Imprint of Large-scale Structure on the Ultrahigh-energy Cosmic-Ray Sky. *The Astrophysical Journal, Letters*, 913(1), Article L13, L13 (cit. on p. 144).
- Dole, H., Lagache, G., Puget, J.-L., Caputi, K. I., et al. (2006). The cosmic infrared background resolved by Spitzer. Contributions of mid-infrared galaxies to the far-infrared background. *Astronomy and Astrophysics*, 451(2) (cit. on p. 31).

- Dominguez, A., Primack, J. R., Rosario, D. J., Prada, F., et al. (2011). Extragalactic background light inferred from AEGIS galaxy-SED-type fractions. *Monthly Notices of the Royal Astronomical Society*, 410(4), 2556–2578 (cit. on pp. 128, 129, 131).
- Donnert, J., Vazza, F., Brüggén, M., Zuhone, J., et al. (2018). Magnetic Field Amplification in Galaxy Clusters and Its Simulation. *Space Science Reviews 2018 214:8*, 214(8) (cit. on p. 157).
- Driver, S. P., Andrews, S. K., Da Cunha, E., et al. (2018). GAMA/G10-COSMOS/3D-HST: the  $0 < z < 5$  cosmic star formation history, stellar-mass, and dust-mass densities. *Monthly Notices of the Royal Astronomical Society*, 475(3) (cit. on p. 115).
- Eichmann, B., Kachelrieß, M., & Oikonomou, F. (2022). Explaining the UHECR spectrum, composition and large-scale anisotropies with radio galaxies. *Journal of Cosmology and Astroparticle Physics*, 2022(7), Article 006, 006 (cit. on p. 146).
- Farrar, G. R. (2012). Tidal disruption flares as the source of ultra-high energy cosmic rays. *European Physical Journal Web of Conferences*, 39, Article 07005, 07005 (cit. on p. 168).
- Fermi, E. (1949). On the Origin of the Cosmic Radiation. *Physical Review*, 75(8) (cit. on p. 21).
- Fleck, I., Titov, M., Grupen, C., & Buvat, I. (2021). *Handbook of Particle Detection and Imaging*. (Cit. on p. 41).
- Fouque, P., Solanes, J. M., Sanchis, T., & Balkowski, C. (2001). Structure, mass and distance of the Virgo cluster from a Tolman-Bondi model. *Astronomy and Astrophysics*, 375(3) (cit. on p. 146).
- Fujita, Y., Ohira, Y., Tanaka, S. J., & Takahara, F. (2009). Molecular Clouds as a Probe of Cosmic-Ray Acceleration in a Supernova Remnant. *The Astrophysical Journal, Letters*, 707, L179–L183 (cit. on p. 126).
- Gaisser, T. K., & Hillas, A. M. (1977). Reliability of the Method of Constant Intensity Cuts for Reconstructing the Average Development of Vertical Showers. *15th International Cosmic Ray Conference*, 8 (cit. on p. 52).
- Gerhard, O., Arnaboldi, M., Freeman, K. C., & Okamura, S. (2002). Isolated Star Formation: A Compact HII Region in the Virgo Cluster. *The Astrophysical Journal, Letters*, 580(2) (cit. on p. 146).
- Gilmore, R. C., Somerville, R. S., Primack, J. R., & Dominguez, A. (2012). Semi-analytic modelling of the extragalactic background light and consequences for extragalactic gamma-ray spectra. *Monthly Notices of the Royal Astronomical Society*, 422(4) (cit. on pp. 102, 111, 128, 131, 140).
- Globus, N., Allard, D., & Parizot, E. (2015). A complete model of the cosmic ray spectrum and composition across the Galactic to extragalactic transition. *Physical Review D*, 92(2), 021302 (cit. on p. 126).
- Goldreich, P., Julian, W. H., Goldreich, P., & Julian, W. H. (1969). Pulsar Electrodynamics. *The Astrophysical Journal*, 157 (cit. on p. 21).
- Greisen, K. (1966). End to the Cosmic-Ray Spectrum? *Physical Review Letters*, vol. 16, Issue 17, pp. 748-750, 16(17) (cit. on p. 34).
- Guépin, C., Kotera, K., Barausse, E., Fang, K., & Murase, K. (2018). Ultra-high-energy cosmic rays and neutrinos from tidal disruptions by massive black holes. *Astronomy and Astrophysics*, 616, Article A179, A179 (cit. on p. 168).
- Guido, E. on behalf of the Pierre Auger Collaboration (2021). Combined fit of the energy spectrum and mass composition across the ankle with the data measured at the

- Pierre Auger Observatory. *Proceeding of Science, 37th International Cosmic Ray Conference, ICRC2021*, 311 (cit. on p. 121).
- Guo, F., Liu, Y. H., Daughton, W., & Li, H. (2015). Particle acceleration and plasma dynamics during magnetic reconnection in the magnetically dominated regime. *The Astrophysical Journal*, 806(2) (cit. on p. 125).
- Hackstein, S., Vazza, F., Brüggem, M., Sorce, J. G., & Gottlöber, S. (2018). Simulations of ultra-high energy cosmic rays in the local Universe and the origin of cosmic magnetic fields. *Monthly Notices of the Royal Astronomical Society*, 475(2) (cit. on p. 146).
- Hanlon W. on behalf of the Telescope Array collaboration (2021). Telescope Array 10 Year Composition. *Proceeding of Science, 37th International Cosmic Ray Conference*, 358 (cit. on pp. 91, 92).
- Haungs, A., Kang, D., Schoo, S., Wochele, D., et al. (2018). The KASCADE Cosmic-ray Data Centre KADC: granting open access to astroparticle physics research data. *European Physical Journal C*, 78 (cit. on p. 40).
- Hess, V. F. (1912). Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten. *Phys.Z.*, 13 (cit. on p. 15).
- Hillas, A. M. (1984). The Origin of Ultra-High-Energy Cosmic Rays. *Annual Review of Astronomy and Astrophysics*, 22(1) (cit. on pp. 24, 25).
- Hillas, A. M. (2005). Topical review: Can diffusive shock acceleration in supernova remnants account for high-energy galactic cosmic rays? *Journal of Physics G: Nuclear and Particle Physics*, 31, R95–R131 (cit. on pp. 20, 122, 132).
- IceCube Collaboration (2019). Cosmic ray spectrum and composition from PeV to EeV using 3 years of data from IceTop and IceCube. *Physical Review D*, 100(8), 082002 (cit. on p. 131).
- Jansson, R., & Farrar, G. R. (2012). A New Model of the Galactic Magnetic Field. *Astrophys.J.*, 757(1) (cit. on pp. 75, 157).
- Kagan, D., Nakar, E., & Piran, T. (2016). The effect of cooling on particle trajectories and acceleration in relativistic magnetic reconnection. *The Astrophysical Journal*, 833(2), 155 (cit. on p. 125).
- Khabibullin, I., & Sazonov, S. (2014). Stellar tidal disruption candidates found by cross-correlating the ROSAT Bright Source Catalogue and XMM-Newton observations. *Monthly Notices of the Royal Astronomical Society*, 444(2), 1041–1053 (cit. on p. 167).
- Koning, A. J., Hilaire, S., & Duijvestijn, M. C. (2007). TALYS-1.0. *International Conference on Nuclear Data for Science and Technology* (cit. on pp. 102, 109, 111, 129, 140).
- Koushan, S., Driver, S. P., Bellstedt, S., Davies, L. J., et al. (2021). GAMA/DEVILS: constraining the cosmic star formation history from improved measurements of the 0.3–2.2  $\mu\text{m}$  extragalactic background light. *Monthly Notices of the Royal Astronomical Society*, 503(2), 2033–2052 (cit. on p. 115).
- Lagage, P. O., & Cesarsky, C. J. (1983). The maximum energy of cosmic rays accelerated by supernova shocks. *Astronomy and Astrophysics*, 125 (cit. on p. 25).
- Levinson, A. (2000). Particle acceleration and curvature TeV emission by rotating, supermassive black holes. *Physical Review Letters*, 85, 912–915 (cit. on p. 125).
- Levinson, A., & Cerutti, B. (2018). Particle-in-cell simulations of pair discharges in a starved magnetosphere of a Kerr black hole. *Astronomy and Astrophysics*, 616, A184 (cit. on p. 125).
- Li T.P. & Ma Y.Q. (1983). Analysis methods for results in gamma-ray astronomy. *The Astrophysical Journal*, 272 (cit. on p. 74).

- Liang, E., Zhang, B., Virgili, F., & Dai, Z. G. (2007). Low-Luminosity Gamma-Ray Bursts as a Unique Population: Luminosity Function, Local Rate, and Beaming Factor. *The Astrophysical Journal*, 662(2), 1111–1118 (cit. on p. 167).
- Lipari, P. (2014). The lifetime of cosmic rays in the Milky Way. *arXiv:1407.5223* (cit. on p. 25).
- Lipari, P., & Vernetto, S. (2020). The shape of the cosmic ray proton spectrum. *Astroparticle Physics*, 120 (cit. on pp. 19, 20).
- Liu, R. Y., Wang, X. Y., & Dai, Z. G. (2011). Nearby low-luminosity gamma-ray bursts as the sources of ultra-high-energy cosmic rays revisited. *Monthly Notices of the Royal Astronomical Society*, 418(2) (cit. on p. 167).
- Lodato, G., & Rossi, E. M. (2011). Multiband light curves of tidal disruption events. *Monthly Notices of the Royal Astronomical Society*, 410(1), 359–367 (cit. on p. 167).
- Longair, M. S. (2011). *High Energy Astrophysics*. Cambridge University Press. (Cit. on pp. 22, 23).
- Luce, Q. (2019). *Etude des rayons cosmiques d'ultra-haute énergie avec l'Observatoire Pierre Auger : de l'extraction du signal à l'interprétation du spectre en énergie* (Doctoral dissertation). (Cit. on p. 133).
- Luce, Q., Marafico, S., Biteau, J., Condorelli, A., & Deligny, O. (2022). Observational constraints on cosmic-ray escape from uhe accelerators. *The Astrophysical Journal* (cit. on p. 101).
- Lucek, S. G., & Bell, A. R. (2000). Non-linear amplification of a magnetic field driven by cosmic ray streaming. *Monthly Notices of the Royal Astronomical Society*, 314(1) (cit. on p. 25).
- Madau, P., & Dickinson, M. (2014). Cosmic Star Formation History. *Annual Review of Astron and Astrophys*, 52, 415–486 (cit. on pp. 115, 167).
- Marafico, S., Biteau, J., & Deligny, O. (2021). Negative fractions in a frequentist analysis of composition derived from  $X_{\max}$  distributions. *Internal Auger note: GAP2021 021* (cit. on p. 81).
- Marafico, S., Biteau, J., & Deligny, O. (2020). The UHECR sky, <https://opendata.auger.org/analysis.php> (cit. on pp. 5, 61).
- Marafico, S., Biteau, J., Deligny, O., & Luce, Q. (2020). Stellar mass density as evolution of sources for the combined fit. *Internal Auger note: GAP2020 035* (cit. on p. 101).
- Matthews, J. (2005). A Heitler model of extensive air showers. *Astroparticle Physics*, 22(5) (cit. on p. 39).
- Mayotte, E. on behalf of the Pierre Auger Collaboration (2021). Indication of a mass-dependent anisotropy above 10<sup>18.7</sup> eV in the hybrid data of the Pierre Auger Observatory. *Proceeding of Science, 37th International Cosmic Ray Conference, ICRC2021 - 321* (cit. on pp. 79, 169).
- McCall, M. L. (2014). A council of giants. *Monthly Notices of the Royal Astronomical Society*, 440(1) (cit. on pp. 115, 157, 162).
- Mei, S., Blakeslee, J. P., Côté, P., Tonry, J. L., et al. (2007). The ACS Virgo Cluster Survey. XIII. SBF Distance Catalog and the Three-dimensional Structure of the Virgo Cluster. *The Astrophysical Journal*, 655(1), 144–162 (cit. on pp. 151, 154).
- Mollerach, S., & Roulet, E. (2018). Progress in high-energy cosmic ray physics. *Progress in Particle and Nuclear Physics*, 98, 85–118 (cit. on pp. 37, 38).
- Mollerach, S., & Roulet, E. (2020). Extragalactic cosmic rays diffusing from two populations of sources. *Physical Review D*, 101(10), 103024 (cit. on p. 121).



- Newton, D., Knapp, J., & Watson, A. A. (2007). The optimum distance at which to determine the size of a giant air shower. *Astroparticle Physics*, 26(6) (cit. on p. 57).
- Ostapchenko, S. (2011). Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 83(1) (cit. on p. 83).
- Peters, B. (1961). Primary cosmic radiation and extensive air showers. *Il Nuovo Cimento*, 22(4), 800–819 (cit. on p. 23).
- Petropoulou, M., & Sironi, L. (2018). The steady growth of the high-energy spectral cut-off in relativistic magnetic reconnection. *Monthly Notices of the Royal Astronomical Society*, 481(4), 5687–5701 (cit. on p. 125).
- Pierog, T., Karpenko, I., Katzy, J. M., Yatsenko, E., & Werner, K. (2015). EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider. *Physical Review C - Nuclear Physics*, 92(3) (cit. on pp. 83, 104, 111).
- Pierog, T., Alekseeva, M. K., Bergmann, T., Chernatkin, V., et al. (2006). First results of fast one-dimensional hybrid simulation of EAS using conex. *Nuclear Physics B - Proceedings Supplements*, 151(1) (cit. on p. 83).
- Ptitsyna, K. V., & Troitsky, S. V. (2010). Physical conditions in potential accelerators of ultra-high-energy cosmic rays: updated Hillas plot and radiation-loss constraints. *Physics-Uspexhi*, 53(7) (cit. on p. 25).
- Puget, J. L., Stecker, F. W., & Bredekamp, J. H. (1976). Photonuclear Interactions of Ultrahigh-Energy Cosmic Rays and their Astrophysical Consequences. *The Astrophysical Journal*, 205, 638–654 (cit. on pp. 102, 109, 129).
- Riehn, F., Engel, R., Fedynitch, A., Gaisser, T. K., & Stanev, T. (2020). Hadronic interaction model sibyll 2.3d and extensive air showers. *Physical Review D*, 102(6) (cit. on p. 83).
- Samuelsson, F., Bégué, D., Ryde, F., Pe'er, A., & Murase, K. (2020). Constraining Low-luminosity Gamma-Ray Bursts as Ultra-high-energy Cosmic Ray Sources Using GRB 060218 as a Proxy. *The Astrophysical Journal*, 902(2) (cit. on pp. 167, 168).
- Sergio Petrer (2020). *Internal Auger note: GAP2020\_058 Xmax parameterizations for post-LHC hadronic models* (tech. rep.). (Cit. on pp. 83, 84, 104).
- Shaya, E. J., Tully, R. B., Hoffman, Y., & Pomarède, D. (2017). Action Dynamics of the Local Supercluster. *The Astrophysical Journal*, 850(2) (cit. on p. 116).
- Shellard, R. C. (2006). First results from the Pierre Auger Observatory. *Brazilian Journal of Physics*, 36(4 A) (cit. on p. 54).
- Shibata, K., & Magara, T. (2011). Solar Flares: Magnetohydrodynamic Processes. *Living Reviews in Solar Physics* 2011 8:1, 8(1) (cit. on p. 21).
- Smartt, S. J. (2009). Progenitors of core-collapse supernovae. *Annual Review of Astronomy and Astrophysics*, 47(1) (cit. on p. 167).
- Sommers, P. (2001). Cosmic ray anisotropy analysis with a full-sky observatory. *Astroparticle Physics*, 14(4) (cit. on p. 61).
- Spitzer, L. (1962). Physics of fully ionized gases. *American Journal of Physics* (cit. on p. 21).
- Stecker, F. W., & Salamon, M. H. (1999). Photodisintegration of ultrahigh-energy cosmic rays: A New determination. *The Astrophysical Journal*, 512, 521–526 (cit. on pp. 109, 129).
- Supanitsky, A. D., Cobos, A., & Etchegoyen, A. (2018). Origin of the light cosmic ray component below the ankle. *Physical Review D*, 98(10), 103016 (cit. on p. 126).
- Taylor, G. I. (1922). Diffusion by Continuous Movements. *Proceedings of the London Mathematical Society*, s2-20(1) (cit. on p. 149).

- Telescope Array Collaboration (2020). Search for Large-scale Anisotropy on Arrival Directions of Ultra-high-energy Cosmic Rays Observed with the Telescope Array Experiment. *The Astrophysical Journal Letters*, 898(2) (cit. on p. 71).
- Telescope Array Collaboration (2012). The surface detector array of the telescope array experiment. *Nuclear Instruments and Methods in Physics Research A*, 689 (cit. on p. 58).
- The Pierre Auger Collaboration (2018a). An Indication of Anisotropy in Arrival Directions of Ultra-high-energy Cosmic Rays through Comparison to the Flux Pattern of Extragalactic Gamma-Ray Sources. *The Astrophysical Journal*, 853(2) (cit. on pp. 115, 197).
- The Pierre Auger Collaboration (2017). Combined fit of spectrum and composition data as measured by the Pierre Auger Observatory. *Journal of Cosmology and Astroparticle Physics*, 2017(04) (cit. on pp. 4, 10, 20, 27, 28, 77, 101, 102, 105, 107, 108).
- The Pierre Auger Collaboration (2020a). Cosmic-Ray Anisotropies in Right Ascension Measured by the Pierre Auger Observatory. *The Astrophysical Journal*, 891(2) (cit. on pp. 71, 74).
- The Pierre Auger Collaboration (2021a). Deep learning based reconstruction of the shower maximum  $X_{\max}$  using the water Cherenkov detectors of the Pierre Auger Observatory. *Journal of Instrumentation*, 16(07) (cit. on p. 80).
- The Pierre Auger Collaboration (2014a). Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above  $10^{17.8}$  eV. *Physical Review D*, 90(12) (cit. on pp. 50, 53).
- The Pierre Auger Collaboration (2014b). Depth of maximum of air-shower profiles at the Pierre Auger Observatory. II. Composition implications. *Physical Review D*, 90(12), 122006 (cit. on pp. 4, 5, 10, 11, 53, 113).
- The Pierre Auger Collaboration (2020b). Direct measurement of the muonic content of extensive air showers between  $2 \times 10^{17}$  eV and  $2 \times 10^{18}$  eV at the Pierre Auger Observatory. *The European Physical Journal C* 2020 80:8, 80(8) (cit. on p. 49).
- The Pierre Auger Collaboration (2020c). Features of the energy spectrum of cosmic rays above  $2.5 \times 10^{18}$  eV using the Pierre Auger Observatory. *Physical Review Letters*, 125(12) (cit. on p. 20).
- The Pierre Auger Collaboration (2015a). Improved limit to the diffuse flux of ultrahigh energy neutrinos from the Pierre Auger observatory. *Physical Review D - Particles, Fields, Gravitation and Cosmology*, 91(9) (cit. on p. 131).
- The Pierre Auger Collaboration (2018b). Large-scale Cosmic-Ray Anisotropies above 4 EeV Measured by the Pierre Auger Observatory. *The Astrophysical Journal*, 868(1) (cit. on pp. 3, 9, 67, 69, 72, 75, 144, 197).
- The Pierre Auger Collaboration (2020d). Measurement of the cosmic-ray energy spectrum above  $2.5 \times 10^{18}$  eV using the Pierre Auger Observatory. *Physical Review D*, 102(6) (cit. on p. 110).
- The Pierre Auger Collaboration (2020e). Reconstruction of events recorded with the surface detector of the Pierre Auger Observatory. *Journal of Instrumentation*, 15(10) (cit. on pp. 55, 57, 58).
- The Pierre Auger Collaboration (2014c). Reconstruction of inclined air showers detected with the Pierre Auger Observatory. *Journal of Cosmology and Astroparticle Physics*, 2014(08) (cit. on p. 58).



- The Pierre Auger Collaboration (2020f). Search for magnetically-induced signatures in the arrival directions of ultra-high-energy cosmic rays measured at the Pierre Auger Observatory. *Journal of Cosmology and Astroparticle Physics*, 2020(6) (cit. on p. 79).
- The Pierre Auger Collaboration (2021b). The energy spectrum of cosmic rays beyond the turn-down around  $10^{17}$  eV as measured with the surface detector of the Pierre Auger Observatory. *European Physical Journal C*, 81(11) (cit. on pp. 103, 111).
- The Pierre Auger Collaboration (2010). The fluorescence detector of the Pierre Auger Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 620(2-3) (cit. on pp. 50, 52).
- The Pierre Auger Collaboration (2015b). The Pierre Auger Cosmic Ray Observatory. *Nuclear Instruments and Methods in Physics Research A*, 798 (cit. on pp. 45, 48, 51).
- The Pierre Auger Collaboration (2013). The Pierre Auger Observatory: Contributions to the 33rd International Cosmic Ray Conference (ICRC 2013) (cit. on p. 52).
- The Pierre Auger collaboration (2022). Arrival Directions of Cosmic Rays above 32 EeV from Phase One of the Pierre Auger Observatory. *The Astrophysical Journal, Supplement* (cit. on pp. 3, 4, 7, 9, 75, 76, 115, 141, 144, 146, 158, 169, 197).
- The Pierre Auger collaboration (2013). Bounds on the density of sources of ultra-high energy cosmic rays from the Pierre Auger Observatory. *Journal of Cosmology and Astroparticle Physics*, 2013(05) (cit. on p. 28).
- Thoudam, S., Rachen, J. P., van Vliet, A., Achterberg, A., et al. (2016). Cosmic-ray energy spectrum and composition up to the ankle: the case for a second Galactic component. *Astronomy and Astrophysics*, 595, A33 (cit. on p. 132).
- Tinyakov P. on behalf of the Pierre Auger and Telescope Array collaborations (2022). The UHECR dipole and quadrupole in the latest data from the original Auger and TA surface detectors. *Proceeding of Science, 37th International Cosmic Ray Conference*, 395 (cit. on pp. 76, 143).
- Unger, M., Farrar, G. R., & Anchordoqui, L. A. (2015). Origin of the ankle in the ultrahigh energy cosmic ray spectrum, and of the extragalactic protons below it. *Physical Review D*, 92(12), 123001 (cit. on p. 126).
- van Velzen, S. (2018). On the Mass and Luminosity Functions of Tidal Disruption Flares: Rate Suppression due to Black Hole Event Horizons. *The Astrophysical Journal*, 852(2), Article 72, 72 (cit. on p. 167).
- Vazza, F., Brüggem, M., Gheller, C., Hackstein, S., et al. (2017). Simulations of extragalactic magnetic fields and of their observables. *Classical and Quantum Gravity*, 34(23) (cit. on pp. 157, 164).
- Yushkov A. on behalf of the Pierre Auger Collaboration (2021). Mass Composition of Cosmic Rays with Energies above  $10^{17.2}$  eV from the Hybrid Data of the Pierre Auger Observatory. *Proceeding of Science, 37th International Cosmic Ray Conference*, 358 (cit. on pp. 89, 90, 97, 111, 113, 189, 190).
- Zatsepin, G. T., & Kuz'min, V. A. (1966). Upper Limit of the Spectrum of Cosmic Rays. *Soviet Journal of Experimental and Theoretical Physics Letters*, 4, 78 (cit. on p. 34).
- Zhang, B. T., Murase, K., Kimura, S. S., Horiuchi, S., & Mészáros, P. (2018). Low-luminosity gamma-ray bursts as the sources of ultrahigh-energy cosmic ray nuclei. *Physical Review D*, 97(8) (cit. on pp. 126, 167).
- Zhang, B. T., Murase, K., Oikonomou, F., & Li, Z. (2017). High-energy cosmic ray nuclei from tidal disruption events: Origin, survival, and implications. *Physical Review D*, 96(6), Article 063007, 063007 (cit. on p. 168).

Zonca, A., Singer, L. P., Lenz, D., et al. (2019). Healpy: Equal area pixelization and spherical harmonics transforms for data on the sphere in Python. *Journal of Open Source Software*, 4(35) (cit. on p. 62).



# List of Figures

1.1	Ionization rate measured by Victor Hess. Graph from Alessandro De Angelis, <a href="https://commons.wikimedia.org/wiki/File:HessKol.jpg">https://commons.wikimedia.org/wiki/File:HessKol.jpg</a> . . . .	15
1.2	Victor Hess making a balloon flight in order to measure the number of electroscopes discharges for different heights. . . . .	16
1.3	New York Times article on the clash between Milikan and Compton about the nature of cosmic rays. . . . .	17
1.4	Astroparticle energy flux measured up to $10^{21}$ eV and down to $10^9$ eV. Adapted from <a href="https://github.com/carmeloevoli/The_CR_Spectrum">https://github.com/carmeloevoli/The_CR_Spectrum</a> . .	18
1.5	Cosmic ray spectrum. In the top right corner, the flux of all particles is shown in blue circles (Tibet experiment) and red diamond (IceTop experiment). The others markers corresponds to the proton flux measured with different experiments. The red solid line is a fit using a two-break expression (cf. Lipari and Vernetto, 2020) that estimates the proton-flux below PeV energies. Above PeV energies, cosmic rays are measured through indirect measurement (cf. 1.4) which makes the deduced composition depend on the hadronic model. Therefore, different scenarios are shown in dashed black lines for the proton flux. Extracted from Lipari and Vernetto, 2020. . . . .	19
1.6	Hillas diagram. The abscissa shows the product $L \times \Gamma$ which corresponds to the product of the comoving size of the engine times the Lorentz factor of the engine. Extracted from Batista et al., 2019. . . . .	27
1.7	Luminosity diagram. . . . .	28
1.8	Schematic spectral energy distribution of the CMB (grey) and the EBL (blue & red). The numbers in the boxes correspond to approximate brightness [ $\text{nW m}^{-2} \text{sr}^{-1}$ ]. Extracted from Dole et al., 2006. . . . .	31
1.9	Attenuation length of protons (left) and iron nuclei (right) as a function of the Lorentz factor at $z = 0$ . The contribution of pair production and pion production from the CMB and EBL (IR/Opt/UV) are separated for the protons. For iron, the production of pairs is separated from photo-erosion cases. Extracted from Allard et al., 2006. . . . .	35
1.10	Attenuation length as a function of energy for different nuclei, at a redshift $z = 0$ . Extracted from Allard et al., 2006. . . . .	35

1.11	CORSIKA simulation of extensive air showers of a proton (left) and iron (right) at $10^{15}$ eV. The colours used are red for electromagnetic particles, green for muons and blue for hadronic particles. Extracted from <a href="https://www.iap.kit.edu/corsika/">https://www.iap.kit.edu/corsika/</a> , Fabian Schmidt, University of Leeds, UK. . . . .	36
1.12	Schematic view of an electromagnetic cascade. The red particles are photons, while the green lines correspond to electron/positron. Extracted from Mollerach and Roulet, 2018. . . . .	37
1.13	Schematic view of a hadronic cascade. Blue lines corresponds to hadronic particles, red corresponds to photons, green corresponds to electron/positron, and purple correspond to muons and neutrinos. Extracted from Mollerach and Roulet, 2018. . . . .	38
1.14	Schematic view of the composition of an extensive air shower. The shower can be explained using three mains components. Green: hadronic component. Red: muonic component. Blue: electromagnetic component. Extracted from Haungs et al., 2018. . . . .	40
1.15	Schematic view of an extensive air shower. Adapted from Fleck et al., 2021. . . . .	41
2.1	The Loma Amarilla site of fluorescence detector. Credit: Pierre Auger Collaboration. . . . .	44
2.2	One of the WCD of the surface detector of the Pierre Auger Observatory, with the Andes in the background. Credit: Pierre Auger Collaboration. . . . .	44
2.3	Map of Pierre Auger Observatory. Each dot represents an SD tank. The four FD sites are shown in black (Loma Amarilla, Los Morados, Los Leones, Colhueco). The blacks lines correspond to the viewing angle of each FD telescope. The low-energy FDs are shown in blue (HEAT). Equipment used for atmosphere monitoring is shown in red. The infilled array is not represented. Extracted from The Pierre Auger Collaboration, 2015b. . . . .	45
2.4	Picture of a LIDAR during a FD shift. . . . .	47
2.5	A schematic view of a FD compared to human size. Extracted from The Pierre Auger Collaboration, 2015b. . . . .	51

2.6	A FD event as observed using a telescope. Left: Each hexagon correspond to one PMT of the camera. The coloured hexagon corresponds to the PMT with a signal above a given threshold. The colours represent the time of detection of each PMT (purple is the earlier and red is the later). Right: The PMT signals of the dotted PMT are shown with the corresponding times. Extracted from The Pierre Auger Collaboration, 2010. . . . .	52
2.7	Energy deposit profile $\frac{dE}{dx}$ as a function of the slant depth. The dots represent the measurements from a FD telescope while the solid line is a fit. Extracted from The Pierre Auger Collaboration, 2014b. . . . .	53
2.8	Schematic view of a SD tank. Extracted from Shellard, 2006. . . . .	54
2.9	Schematic view of a extensive air shower (blue). The SD detectors are shown in green. Extracted from The Pierre Auger Collaboration, 2020e. . . . .	55
2.10	Lateral distribution function of an SD event using two different analysis pipelines. The red dots are computed using Herald software, with an associated fit using a log-parabola function. The blue dots are computed using Observer software, with the associated fit using a NKG modified function. Extracted from The Pierre Auger Collaboration, 2020e. . . . .	57
2.11	Relative uncertainty of $S(1000)$ as a function of $\lg(S(1000)/VEM/\text{energy})$ (left) or as a function of zenith angle $\sin^2 \theta$ (right) for saturated and non-saturated events. Extracted from The Pierre Auger Collaboration, 2020e. . . . .	58
3.1	Exposure map in Equatorial (left) and Galactic coordinates (right). . . . .	62
3.2	Count map showing the number of detected events in each pixel. . . . .	63
3.3	Smoothed count map $N_{\text{smoothed}}(\alpha, \delta)$ . . . . .	64
3.4	Smoothed exposure map, $\omega_{\text{smoothed}}(\alpha, \delta)$ . . . . .	65
3.5	Flux map obtained using the $N/E_{\text{tot}}$ method. . . . .	65
3.6	Raw flux map, where each pixel corresponds to the number of detected events in a single pixel divided by the exposure of the same pixel. . . . .	66
3.7	Flux map computed using $1/\omega$ method. . . . .	67
3.8	Benchmark scenario, $\phi^{\text{expected}}(\alpha, \delta)$ , with an applied mask corresponding to the field of view of the Pierre Auger Observatory. . . . .	68
3.9	Average reconstructed dipole sky maps considering the two methods. . . . .	68
3.10	Bias map of the two reconstruction methods. . . . .	69
3.11	Standard deviation map considering the two reconstruction methods. . . . .	70
3.12	Bins used to reconstruct the flux in right ascension analysis. . . . .	70
3.13	Flux analysis in right ascension. The dots correspond to the average flux in a bin of $36^\circ$ . . . . .	71

3.14	Normalized flux of UHECRs as a function of right ascension. The green curve corresponds to the Rayleigh analysis, while the orange one corresponds to a first harmonic fit. A constant fit is shown in blue and refers to an isotropic scenario. . . . .	73
3.15	Reconstructed flux and significance map of events above 8 EeV. . . . .	74
3.16	Auger UHECR flux in Galactic coordinates above 8 EeV smoothed with a $45^\circ$ -radius top-hat function. The cross displays the observed dipole orientation, and the two contours stand for the 68% and 95% C.L. regions. The 2MRS dipole is indicated. Arrows indicate expected deflections of the 2MRS dipole considering the Galactic magnetic field of Jansson and Farrar, 2012, with protons at $E = 5$ EeV or 2 EeV. Extracted from The Pierre Auger Collaboration, 2018b. . . . .	75
3.17	The test-statistic profile as a function of threshold energy for the catalogues studied and the Centaurus region. The local $p$ -value shown in the right-hand side is penalized for signal fraction and angular scale scans (catalogues) and for the radius scans (Centaurus region). Extracted from Biteau J. on behalf of the Pierre Auger Collaboration, 2022. . . . .	77
3.18	The best-fit flux model from Starburst galaxies catalogues smoothed using a $24^\circ$ radius top-hat function as for Fig. 3.19. Extracted from Biteau J. on behalf of the Pierre Auger Collaboration, 2022. . . . .	78
3.19	<i>Left:</i> Flux of cosmic rays observed above 41 EeV observed at the Pierre Auger Observatory smoothed with a top-hat function of radius $24^\circ$ . <i>Right:</i> Associated Li-Ma pre-trial significance map. The solid gray line shows the super-galactic plane. Extracted from Biteau J. on behalf of the Pierre Auger Collaboration, 2022. . . . .	78
3.20	Mass-sensitive sky map for $E \geq 10^{18.7}$ eV. Extracted from Mayotte, E. on behalf of the Pierre Auger Collaboration, 2021. . . . .	79
4.1	$X_{\max}$ distribution for given energy bins from the ICRC 2019 dataset (dataset produced for the 37th International Cosmic Rays Conference). . . . .	82
4.2	Example of an $X_{\max}$ distribution modelled with four different masses. The red curve corresponds to $^1_1\text{H}$ , grey to $^2_4\text{He}$ , green to $^7_{14}\text{N}$ , blue to $^{26}_{54}\text{Fe}$ ; the sum is shown in brown. . . . .	83
4.3	Distribution used in the benchmark scenario ( $f_{\text{H}} = 0.3$ , $f_{\text{He}} = 0.2$ , $f_{\text{N}} = 0.2$ , $f_{\text{Fe}} = 0.3$ , in the energy bin $18.3 \leq \log_{10}(E/\text{eV}) < 18.4$ ). Points represent the acceptance-rejection method: Orange for kept points, grey for rejected points. . . . .	88

4.4	Multinomial (red) & Poisson (blue) reconstructions of 1000 MC simulations, with an injected composition $f_H = 0.3$ , $f_{He} = 0.2$ , $f_N = 0.2$ , $f_{Fe} = 0.3$ in the energy bin $18.3 \leq \log_{10}(E/eV) < 18.4$ , with $N = \text{Poisson}(3000)$ . The grey line corresponds to the fraction of Helium injected, the green line corresponds to the fraction of Nitrogen injected.	89
4.5	Best-fit parameters for each energy bin using the ICRC 2019 dataset (Yushkov A. on behalf of the Pierre Auger Collaboration, 2021) considering different hadronic models: EPOS-LHC (solid red), Sibyll2.3d (dotted blue), QGS-JETII-04 (dashed green). The lines are here to guide the eye.	90
4.6	$p_{\text{value}}$ obtained with different hadronic models. Grey dot show the $p_{\text{value}}$ obtained from $\chi^2$ statistics. Black dot show the $p_{\text{value}}$ obtained from C-statistics.	91
4.7	Poisson reconstruction of 1000 MC simulations, with an injected composition $f_H = 0.3$ , $f_{He} = 0.2$ , $f_N = 0.2$ , $f_{Fe} = 0.3$ in the energy bin $18.3 \leq \log_{10}(E/eV) < 18.4$ , with $N = \text{Poisson}(30)$ . The grey line corresponds to the fraction of Helium injected, the green line corresponds to the fraction of Nitrogen injected. Dashed lines correspond to the average reconstructed parameters with its $1\sigma$ band.	92
4.8	Poisson reconstruction of 1000 MC simulations, with an injected composition $f_H = 0.3$ , $f_{He} = 0.2$ , $f_N = 0.2$ , $f_{Fe} = 0.3$ in the energy bin $18.3 \leq \log_{10}(E/eV) < 18.4$ , with $N = \text{Poisson}(100)$ . Negative fractions are allowed. The grey line corresponds to the fraction of Helium injected, the green line corresponds to the fraction of Nitrogen injected. Dashed lines correspond to the average reconstructed parameters with its $1\sigma$ band.	93
4.9	Standard deviation (top) and bias (bottom) of Helium fraction for 1000 MC simulations, with an injected composition $f_H = 0.3$ , $f_{He} = 0.2$ , $f_N = 0.2$ , $f_{Fe} = 0.3$ in the energy bin $18.3 \leq \log_{10}(E/eV) < 18.4$ , with $N$ varying from 30 to 10,000. The lines are here to guide the eye.	94
4.10	Poisson approach with 1000 MC simulations, with an injected composition $f_H = 0$ , $f_{He} = 0$ , $f_N = 1$ , $f_{Fe} = 0$ in the energy bin $18.8 \leq \log_{10}(E/eV) < 18.9$ , with $N = \text{Poisson}(3000)$ . Negative fractions allowed.	95



4.11	Fractions obtained from the best-fit parameters using a Poisson approach and EPOS-LHC as hadronic-interaction model for the data from ICRC 2019 (Yushkov A. on behalf of the Pierre Auger Collaboration, 2021). The last panel shows the $p_{\text{value}}$ . The black dots correspond to the C-statistics $p_{\text{value}}$ , and can be compared to a more classical $\chi^2$ statistics (grey dots). . . . .	97
4.12	Bias estimator for scenarios with and without negative fractions allowed. . . . .	98
4.13	Standard deviation of each fraction for free (dashed lines) and strictly positive fractions (solid lines). The lines are here to guide the eye. . .	99
4.14	Differences between the fractions obtained from the strictly positive fraction approach and the free fraction approach in Fig. 4.11. The dark-coloured band shows, for each energy bin, the maximum bias from the two methods. The light coloured band shows the maximum resolution of the two methods, obtained in Fig. 4.13. . . . .	99
5.1	Energy flux measured at the Pierre Auger Observatory. Point above the ankle energy are shown in black . . . . .	104
5.2	<i>Left:</i> Detected spectrum with detected nuclei. Colors represent the detected nuclei, red for $A = 1$ , grey for $2 \leq A \leq 4$ , green for $5 \leq A \leq 22$ , cyan $23 \leq A \leq 38$ , blue for $39 \leq A \leq 56$ , the sum is shown in brown. <i>Right:</i> Luminosity density of the sources as a function of energy, red for H, grey for He, green for N, cyan Si, blue for Fe, the sum is shown in brown. . . . .	112
5.3	$X_{\text{max}}$ distributions with best-fit model. The model is shown in brown. .	113
5.4	$X_{\text{max}}$ distribution moments for each energy bin. The brown curve represents the best-fit scenario. Other curves represent the expected values for a pure composition, red for H, grey for He, green for N, cyan for Si, blue for Fe. . . . .	114
5.5	Stellar mass density as a function of luminosity distance. The local measured SMD is shown in purple, while the cosmic SMD is shown in blue. . . . .	116
5.6	Star formation rate density as a function of luminosity distance. The local measured SFRD is shown in purple, while the CSFH is shown in blue. . . . .	117

5.7	<i>Left</i> : Observed spectrum with detected nuclei. Colors represent the detected nuclei, red for $A = 1$ , grey for $2 \leq A \leq 4$ , green for $5 \leq A \leq 22$ , cyan $23 \leq A \leq 38$ , blue for $39 \leq A \leq 56$ , the sum is shown in brown. <i>Right</i> : Luminosity density of the sources as a function of energy, red for H, grey for He, green for N, cyan Si, blue for Fe, the sum is shown in brown. . . . .	119
5.8	$X_{\max}$ moments for each energy bin. The brown curve represents the best-fit scenario. The dashed brown curve represents the best-fit configuration for a flat distribution of sources. Coloured lines represent the expected values for a pure composition, red for H, grey for He, green for N, cyan for Si, blue for Fe. . . . .	120
5.9	Energy spectrum measured at the Pierre Auger Observatory and its associated proton component. Both the red and black points are modelled in the following. . . . .	123
5.10	SFRD scenario. <i>Left</i> : Detected spectrum with detected nuclei. Colors represent the detected nuclei, red for $A = 1$ , grey for $2 \leq A \leq 4$ , green for $5 \leq A \leq 22$ , cyan $23 \leq A \leq 38$ , blue for $39 \leq A \leq 56$ , the sum is shown in brown. <i>Right</i> : Luminosity density of the sources as a function of energy, red for H, grey for He, green for N, cyan Si, blue for Fe, the sum is shown in brown. . . . .	125
5.11	SMD scenario with $\gamma_p$ & $\gamma_A$ . <i>Left</i> : Detected spectrum with detected nuclei. Colors represent the detected nuclei, red for $A = 1$ , grey for $2 \leq A \leq 4$ , green for $5 \leq A \leq 22$ , cyan $23 \leq A \leq 38$ , blue for $39 \leq A \leq 56$ , the sum is shown in brown. <i>Right</i> : Luminosity density of the sources as a function of energy, red for H, grey for He, green for N, cyan Si, blue for Fe, the sum is shown in brown. . . . .	128
6.1	Average number of nuclei detected at energies above $10^{17}$ eV as a function of distance of injection for He injection at $E = 10^{18.95}$ eV. . . .	136
6.2	Average number of detected nuclei at energies above $10^{17}$ eV as a function of injected energy for He nuclei emitted at 104 Mpc. . . . .	137
6.3	Average number of protons detected per bin for He injection at 104 Mpc, as a function of injected and detected energies. . . . .	137
6.4	Sky maps of the expected flux above $E_{\text{th}} = 10^{18.9}$ eV using the SMD and SFRD scenarios. The grey lines shows the zone of avoidance along the Galactic plane. The red line corresponds to the super galactic plane.	142

6.5	UHECR full sky flux seen from the Pierre Auger Observatory and Telescope Array above 8.57 EeV (Auger) and 10 EeV (TA). The map is shown in equatorial coordinate, the galactic center is shown as a white dot. Extracted from Tinyakov P. on behalf of the Pierre Auger and Telescope Array collaborations, <a href="#">2022</a> . . . . .	143
6.6	UHECR full sky flux seen from the Pierre Auger Observatory and Telescope Array above 38 EeV (Auger) and 49 EeV (TA). The map is shown in equatorial coordinate, the galactic center is shown as a white dot. Extracted from Di Matteo A. on behalf of the Pierre Auger and Telescope Array collaborations, <a href="#">2021</a> . . . . .	144
6.7	Sky maps of the expected flux above $E_{\text{th}} = 10^{18.9}$ eV using the SMD and SFRD scenarios. . . . .	145
6.8	Schematic of a UHECR that traverses a cluster magnetic field . . . . .	147
6.9	Typical path of a UHECR in a homogeneous isotropic turbulent magnetic field of coherence length $\lambda_B$ at a distance $d$ , adapted from Bray and Scaife, <a href="#">2018</a> . . . . .	148
6.10	<i>Top</i> : Distance $d$ traveled by an UHECR emitted by a source located behind the cluster with an angular distance $\theta$ with respect to the center of the cluster. <i>Bottom</i> : Corresponding escape rigidity $R_{\text{esc}}$ . . . . .	151
6.11	Comparison between diffusion timescale and different interaction timescales for a source at the centre of cluster. From Condorelli A. (private communication). . . . .	152
6.12	Sky map of the expected flux above $E_{\text{th}} = 10^{19.6}$ eV considering Virgo's magnetic field. . . . .	153
6.13	Sky map of the expected flux above $E_{\text{th}} = 10^{18.9}$ eV considering Virgo's magnetic field spreading over a radial extent $R_{\text{Cl}}$ . . . . .	154
6.14	Sky maps of the expected flux above $E_{\text{th}} = 10^{19.6}$ eV considering two minimum distances of injection $D_{\text{min}} = 0.5$ Mpc and $D_{\text{min}} = 0$ . Mpc. . . . .	155
6.15	Angular and time spread as a function of the distance of the source for a rigidity of $R \sim 4$ EV. . . . .	159
6.16	Example of the impact of the magnetic spread in time in a transient scenario where UHECR are produced in bursts. . . . .	160
6.17	Sky map of the expected flux above $E_{\text{th}} = 10^{19.6}$ eV considering $k = 10^{-7} M_{\odot}^{-1}$ . . . . .	161
6.18	Sky map of the expected flux above $E_{\text{th}} = 10^{19.6}$ eV considering $k = 10^{-4} M_{\odot}^{-1}$ . . . . .	162
6.19	Sky map of the expected flux above $E_{\text{th}} = 10^{19.6}$ eV considering $k = 3 \times 10^{-6} M_{\odot}^{-1}$ . . . . .	163

6.20	Sky map of the expected flux above $E_{\text{th}} = 10^{19.6}$ eV considering $k = 10^{-5} M_{\odot}^{-1}$ . . . . .	163
6.21	Sky map of the expected flux above $E_{\text{th}} = 10^{19.6}$ eV considering $k = 10^{-16} M_{\odot}^{-1} \text{ yr}^{-1}$ and $B_{\text{LS}} = 25$ nG . . . . .	165
6.22	Sky map of the expected flux above $E_{\text{th}} = 10^{19.6}$ eV considering $k = 3 \times 10^{-15} M_{\odot}^{-1} \text{ yr}^{-1}$ and $B_{\text{LS}} = 10$ nG . . . . .	165
6.23	Constraints on $k$ and on $\mathcal{E}_{\text{tot}}$ in a SFRD scenario (left axis) and SMD scenario (right axis). Source candidates are shown as markers (red for those who follow the SFRD and blue for those who follow the SMD). . . . .	166
A.1	Benchmark scenario, $\phi^{\text{expected}}(\alpha, \delta)$ , with an applied mask corresponding to the field of view of the Pierre Auger Observatory. . . . .	198
A.2	Average reconstructed sky maps considering the two methods. . . . .	198
A.3	Bias map of the two reconstruction methods . . . . .	199
A.4	Standard deviation map considering the two reconstruction methods. . . . .	199



# List of Tables

2.1	Atmospheric measurements performed and the instruments that are used. Table extracted from The Pierre Auger Collaboration, 2015b. . .	48
5.1	Comparison between this work and that of Condorelli, 2021 using ICRC 2015 and 2021 datasets. . . . .	108
5.2	Comparison between PSB and TALYS as photodisintegration models. .	109
5.3	Benchmark scenario using a flat evolution of sources. . . . .	111
5.4	Results obtained with a luminosity evolution term that follows the SMD or SFRD (middle SMD, right SFRD) compared to those obtained with a flat distribution of sources (left). The ICRC 2021 dataset is used. . . .	118
5.5	Results obtained with a luminosity evolution term that follows the SMD (left) and SFRD (right), considering the comparison between the observed proton flux across the ankle with the proton flux of the model. The ICRC 2021 dataset is used. . . . .	124
5.6	Results obtained with a luminosity evolution term that follows the SMD (left) and SFRD (right), considering the proton across the ankle and two spectral indices $\gamma_p$ for proton and $\gamma_A$ for the other nuclei. The ICRC 2021 dataset is used. . . . .	127

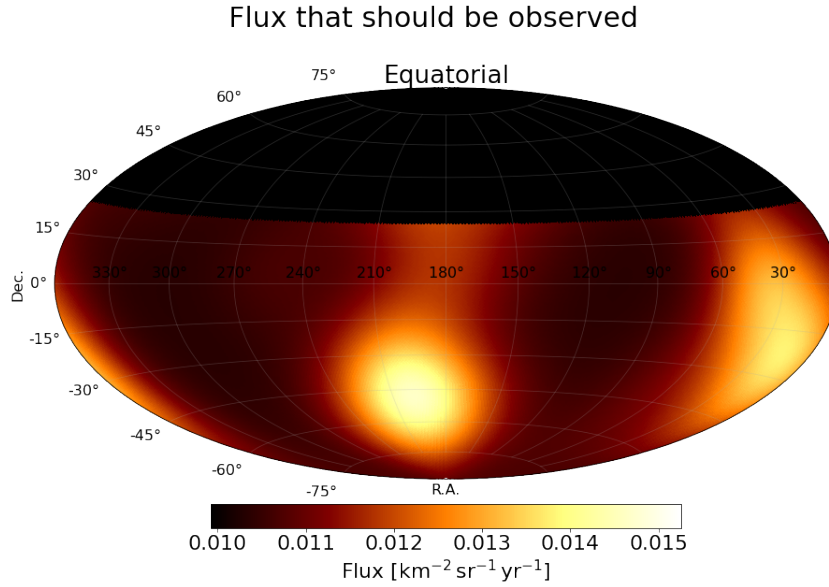


## Comparison of the two reconstruction methods using starburst study

In Sec. 3.1.3, two reconstruction methods are tested and compared, starting from a benchmark scenario based on the dipole observed in The Pierre Auger Collaboration, 2018b (cf. 3.2). In this section, the same analysis is re-run based on the evidence of anisotropy observed above 40 EeV (The Pierre Auger collaboration, 2022). To do so, the benchmark scenario is built from the catalogue of starburst galaxies used in The Pierre Auger Collaboration, 2018a using 90% of isotropy and 10% of flux coming from the catalogue. As in Chap. 4, the map from the benchmark scenario is then multiplied by the exposure of the detector in order to get a hypothetical expected count map. From the hypothetical count map, 1000 random mock maps are created using an acceptance-rejection method. All the mocks have the same number of events  $\mathcal{N}$ .  $\mathcal{N}$  is taken to be the number of events detected from the start of the experiment up to December 2020 (date of the analysis). For each mock, the flux is reconstructed with the two methods using the same smoothing function as in The Pierre Auger Collaboration, 2018a, which corresponds to a Von-Fisher function with  $13^\circ$  radius. The expected pattern that should be reconstructed (modulo the randomness of mocks) is shown in Fig. A.1. It corresponds to the injected/observed pattern with a mask which takes into account the field of view of the detector.

Figure A.2 shows the average reconstructed map considering the two approaches. To enhance the differences, a bias map is built by subtracting and normalizing the expected map to the reconstructed ones:  $\text{Bias}(\alpha, \delta) = \frac{\phi^{\text{reconstructed}}(\alpha, \delta) - \phi^{\text{expected}}(\alpha, \delta)}{\phi^{\text{expected}}(\alpha, \delta)}$ . The bias map is set to zero in the masked area by definition.

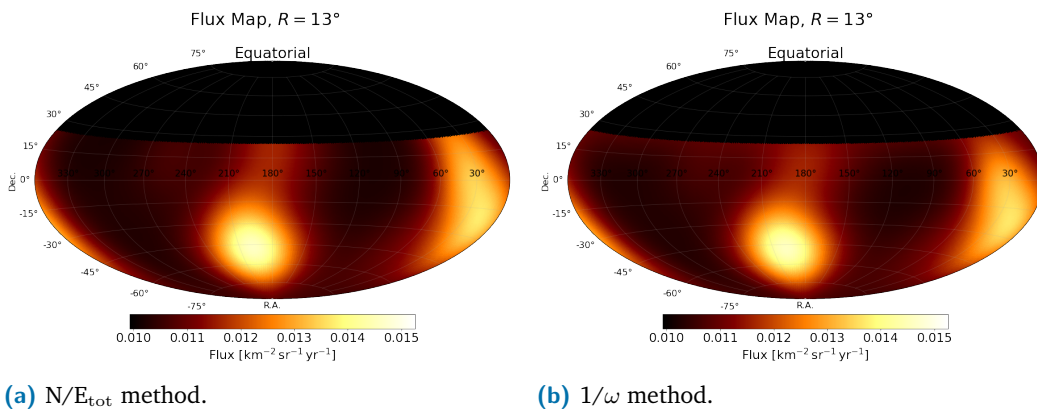




**Fig. A.1.:** Benchmark scenario,  $\phi^{\text{expected}}(\alpha, \delta)$ , with an applied mask corresponding to the field of view of the Pierre Auger Observatory.

The bias map is shown in Fig. A.3. Looking at the bias, one can see that, as before, most differences appear at the border of the field of view. Looking at Fig. A.3, it seems that the  $N/E_{\text{tot}}$  method increases the contrast near the border while the  $1/\omega$  method reduces the contrast, this behaviour was also seen in the dipole analysis (cf. Chap. 3). To quantify the bias, the average bias is computed using:

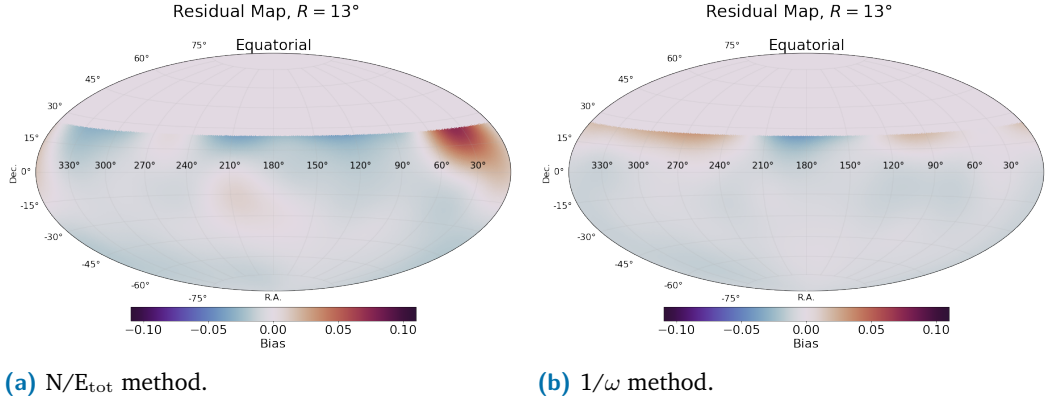
$$\langle \delta \rangle = \sqrt{\frac{\sum_i \delta_i^2}{\mathcal{N}_{\text{pix}}}}, \quad (\text{A.1})$$



**Fig. A.2.:** Average reconstructed sky maps considering the two methods.

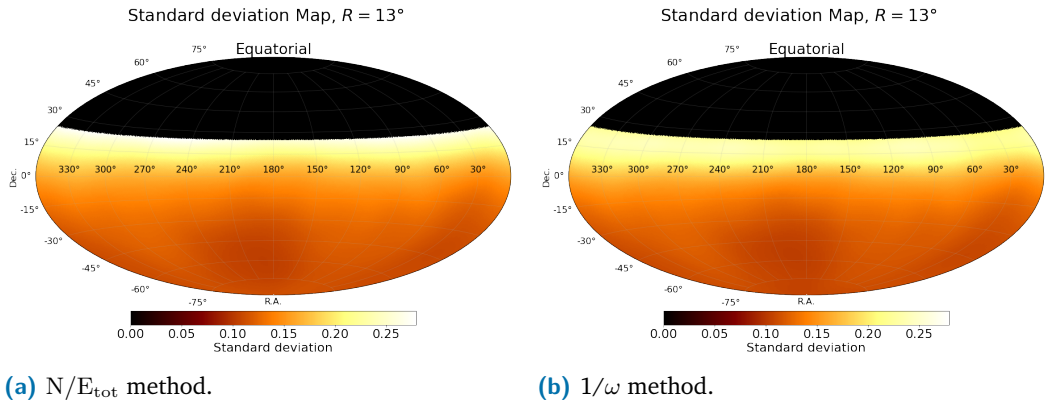
where  $\mathcal{N}_{\text{pix}}$  is the number of pixels and  $\delta_i$  is the bias in the  $i$ -pixel. It results in an average bias of:

$$\begin{cases} \langle \delta \rangle^{N/E_{\text{tot}}} = 1.2\% \\ \langle \delta \rangle^{1/\omega} = 1.0\% \end{cases} \quad (\text{A.2})$$



**Fig. A.3.:** Bias map of the two reconstruction methods .

A standard deviation map is obtained by computing the standard deviation of the 1000 mock maps in each pixel. The standard deviation map is shown in Fig. A.4. For the two methods, the variance is the same far away from the borders, however, the variance changes when being  $13^\circ$  away of the border. For the  $N/E_{\text{tot}}$  method, the standard deviation gradually increases when getting closer to the border, while for the  $1/\omega$  method, the standard deviation is separated in two areas: above and below  $13^\circ$  away from the border. As in the  $N/E_{\text{tot}}$  method, the area closer to the border has a bigger standard deviation than the area far from the border.



**Fig. A.4.:** Standard deviation map considering the two reconstruction methods.

As for the bias, the average standard deviation is computed using:

$$\langle \sigma \rangle = \sqrt{\frac{\sum_i \sigma_i^2}{\mathcal{N}_{\text{pix}} - 1}}. \quad (\text{A.3})$$

It results in an average standard deviation of:

$$\begin{cases} \langle \sigma \rangle^{N/E_{\text{tot}}} = 16.0\% \\ \langle \sigma \rangle^{1/\omega} = 16.0\% \end{cases} \quad (\text{A.4})$$

Comparing the average bias and average standard deviation in Eq. A.2 & Eq. A.4, one can see that the two approaches are equivalent in terms of bias and standard deviation.

# Propagation in a turbulent magnetic field

## B.1 Starting from stochastic differential equation

As shown in Achterberg et al., 1999, the propagation can be modeled using stochastic differential equations with the diffusion coefficient  $\mathcal{D}_0$ , that has been computed in Chap. 6:

$$dn_i = -2\mathcal{D}_0 n_i ds + \sqrt{2\mathcal{D}_0} P_{ij}(\hat{\mathbf{n}}) dW_j , \quad (\text{B.1})$$

where  $P_{ij}(\hat{\mathbf{n}}) \equiv \delta_{ij} - n_i n_j$  is the projection tensor onto the plane perpendicular to  $\hat{\mathbf{n}}$ . The tensor satisfies  $\mathbf{P} = \mathbf{P}^\dagger$  and  $P_{ij}P_{jk} = P_{ik}$ . The quantity  $d\mathbf{W} \equiv (dW_1, dW_2, dW_3)$  is a three-component Wiener process satisfying:

$$\langle dW_i \rangle = 0 \quad , \quad \langle dW_i dW_j \rangle = ds \delta_{ij} . \quad (\text{B.2})$$

The symbol  $\langle \cdot \rangle$  represents the average ensemble. From Eq. B.2, one can estimate:  $n_i(s)$ :

$$\frac{d \langle n_i \rangle}{ds} = -2\mathcal{D}_0 \langle n_i \rangle \implies \langle n_i \rangle (s) = \langle n_i \rangle_0 e^{-2\mathcal{D}_0 s} . \quad (\text{B.3})$$

The subscript 0 denotes the initial value at  $s = 0$ . In the following pages, the Itô rules is applied for two stochastic integrands  $X$  and  $Y$

$$d(X \cdot Y) = dX \cdot Y + X \cdot dY + dX \cdot dY , \quad (\text{B.4})$$

Keeping the term  $ds$  at first order and assuming

$$dW_i ds = 0 , \quad (\text{B.5})$$

$$dW_i dW_j = ds \delta_{ij} ,$$

One can find:

$$d(n_i n_j) = 2\mathcal{D}_0 (\delta_{ij} - 3 n_i n_j) ds + (n_i P_{jk} dW_k + n_j P_{ik} dW_k) .$$

Taking the ensemble average results in:

$$\frac{d \langle n_i n_j \rangle}{ds} = 2\mathcal{D}_0 \delta_{ij} - 6\mathcal{D}_0 \langle n_i n_j \rangle . \quad (\text{B.6})$$

Integrating along the UHECR path gives:

$$\langle n_i n_j \rangle = \langle n_i n_j \rangle_0 e^{-6\mathcal{D}_0 s} + \frac{\delta_{ij}}{3} (1 - e^{-6\mathcal{D}_0 s}) . \quad (\text{B.7})$$

The next step consists of computing the distance along an axis:

$$x_i(s) = \int_0^s ds' n_i(s') . \quad (\text{B.8})$$

Taking the average, using equation (B.3), one can find:

$$\langle x_i \rangle (s) = \frac{n_{i0}}{2\mathcal{D}_0} (1 - e^{-2\mathcal{D}_0 s}) . \quad (\text{B.9})$$

From Eq. B.4 and Eq. 6.19, one can find

$$d(x_i n_j) = (n_i n_j - 2\mathcal{D}_0 x_i n_j) ds + \sqrt{2\mathcal{D}_0} x_i P_{jk} dW_k . \quad (\text{B.10})$$

Taking the average yields a differential equation for  $\langle x_i n_j \rangle$ :

$$\frac{d \langle x_i n_j \rangle}{ds} = \langle n_i n_j \rangle - 2\mathcal{D}_0 \langle x_i n_j \rangle . \quad (\text{B.11})$$

Integrating the equation using Eq. B.7 one finds:

$$\begin{aligned} \langle x_i n_j \rangle (s) &= \frac{\langle n_i n_j \rangle_0 - \frac{1}{3} \delta_{ij}}{4\mathcal{D}_0} e^{-2\mathcal{D}_0 s} (1 - e^{-4\mathcal{D}_0 s}) \\ &+ \frac{\delta_{ij}}{6\mathcal{D}_0} (1 - e^{-2\mathcal{D}_0 s}) . \end{aligned} \quad (\text{B.12})$$

Note that this expression is symmetric in  $i$  and  $j$ . Similarly, one can derive an equation for  $\langle x_i x_j \rangle$ :

$$\frac{d \langle x_i x_j \rangle}{ds} = 2 \langle x_i n_j \rangle , \quad (\text{B.13})$$

where we used the above symmetry. Direct integration yields:

$$\begin{aligned}
\langle x_i x_j \rangle (s) &= \frac{\langle n_i n_j \rangle_0 - \frac{1}{3} \delta_{ij}}{4\mathcal{D}_0^2} \times \\
&\times \left[ 1 - e^{-2\mathcal{D}_0 s} - \frac{1}{3} \left( 1 - e^{-6\mathcal{D}_0 s} \right) \right] \\
&+ \frac{\delta_{ij}}{3\mathcal{D}_0} \left[ s - \frac{1}{2\mathcal{D}_0} \left( 1 - e^{-2\mathcal{D}_0 s} \right) \right].
\end{aligned} \tag{B.14}$$

Eq. B.9 and Eq. B.14 match the results of Achterberg et al., 1999.

### B.1.1 Taylor expansion

In order to derive the duration of a burst in the approximation of small-angle, we perform a Taylor expansion of the terms computed above, assuming  $\mathcal{D}_0 s \ll 1$ . We first do a Taylor expansion of the term  $e^{-2\mathcal{D}_0 s}$  and  $e^{-6\mathcal{D}_0 s}$ . The first Taylor expansion gives:

$$e^{-2\mathcal{D}_0 s} = 1 - 2\mathcal{D}_0 s + 2\mathcal{D}_0^2 s^2 - \frac{4}{3}\mathcal{D}_0^3 s^3 + \frac{2}{3}\mathcal{D}_0^4 s^4 \tag{B.15}$$

The second one gives:

$$e^{-6\mathcal{D}_0 s} = 1 - 6\mathcal{D}_0 s + 18\mathcal{D}_0^2 s^2 - 36\mathcal{D}_0^3 s^3 + 54\mathcal{D}_0^4 s^4 \tag{B.16}$$

Injecting Eq. B.15 into Eq. B.9, leads to:

$$\langle x_i \rangle (s) = \frac{n_{i0}}{2\mathcal{D}_0} \left( +2\mathcal{D}_0 s - 2\mathcal{D}_0^2 s^2 + \frac{4}{3}\mathcal{D}_0^3 s^3 - \frac{2}{3}\mathcal{D}_0^4 s^4 \right). \tag{B.17}$$

$$\langle x_i \rangle (s) = n_{i0} \left( +s - \mathcal{D}_0 s^2 + \frac{2}{3}\mathcal{D}_0^2 s^3 - \frac{1}{3}\mathcal{D}_0^3 s^4 \right). \tag{B.18}$$

Injecting Eq. B.15 and Eq. B.16 into Eq. B.14, leads to:

$$\begin{aligned}
\langle x_i x_i \rangle (s) &= \frac{\langle n_i n_j \rangle_0 - \frac{1}{3}}{4\mathcal{D}_0^2} \\
&\times \left[ 1 - (1 - 2\mathcal{D}_0 s + 2\mathcal{D}_0^2 s^2 - \frac{4}{3}\mathcal{D}_0^3 s^3 + \frac{2}{3}\mathcal{D}_0^4 s^4) \right. \\
&- \left. \frac{1}{3} \left( 1 - (1 - 6\mathcal{D}_0 s + 18\mathcal{D}_0^2 s^2 - 36\mathcal{D}_0^3 s^3 + 54\mathcal{D}_0^4 s^4) \right) \right] \\
&+ \frac{1}{3\mathcal{D}_0} \left[ s - \frac{1}{2\mathcal{D}_0} \left( 1 - (1 - 2\mathcal{D}_0 s + 2\mathcal{D}_0^2 s^2 - \frac{4}{3}\mathcal{D}_0^3 s^3 + \frac{2}{3}\mathcal{D}_0^4 s^4) \right) \right] \\
&\hspace{15em} \text{(B.19)} \\
&= \frac{\langle n_i n_j \rangle_0 - \frac{1}{3}}{4\mathcal{D}_0^2} \times \left[ +4\mathcal{D}_0^2 s^2 - \frac{32}{3}\mathcal{D}_0^3 s^3 + \frac{52}{3}\mathcal{D}_0^4 s^4 \right] \\
&+ \frac{1}{3\mathcal{D}_0} \left[ s - \frac{1}{2\mathcal{D}_0} \left( 1 - (1 - 2\mathcal{D}_0 s + 2\mathcal{D}_0^2 s^2 - \frac{4}{3}\mathcal{D}_0^3 s^3 + \frac{2}{3}\mathcal{D}_0^4 s^4) \right) \right] \\
&= \left( \langle n_i n_j \rangle_0 - \frac{1}{3} \right) \times \left[ s^2 - \frac{8}{3}\mathcal{D}_0 s^3 + \frac{13}{3}\mathcal{D}_0^2 s^4 \right] + \left( +\frac{1}{3}s^2 - \frac{2}{9}\mathcal{D}_0 s^3 + \frac{1}{9}\mathcal{D}_0^2 s^4 \right) .
\end{aligned}$$

### B.1.2 Intermediate computation

In order to compute the time delay and the duration of a burst, one needs to compute the average  $\langle x_i \rangle$  and the root-mean-square  $\langle x_i^2 \rangle$ , using Eq. B.18 and Eq. B.19. As shown in Fig. 6.9, the  $z$ -axis is taken to be the initial axis of the particle.

$$\begin{aligned}
\langle x \rangle = \langle y \rangle = 0 \quad , \quad \langle z \rangle = s - \mathcal{D}_0 s^2 + \frac{2}{3}\mathcal{D}_0^2 s^3 \quad , \\
\hspace{15em} \text{(B.20)} \\
\langle x^2 \rangle = \langle y^2 \rangle = \frac{2}{3}\mathcal{D}_0 s^3 - \frac{4}{3}\mathcal{D}_0^2 s^4 \quad , \quad \langle z^2 \rangle = s^2 - 2\mathcal{D}_0 s^3 + 3\mathcal{D}_0^2 s^4 .
\end{aligned}$$

## B.2 Time delay and duration of a burst

### B.2.1 Time delay

The propagation time  $\tau$  is given by the average value of  $r$

$$\langle r \rangle = \left\langle \sqrt{x^2 + y^2 + z^2} \right\rangle \quad (\text{B.21})$$

Since we are doing the small-angle approximation, one can use :  $|x|, |y| \ll |z|$  and do a Taylor expansion:

$$\langle r \rangle = \left\langle z + \frac{x^2 + y^2}{2z} \right\rangle \quad (\text{B.22})$$

We approximate  $\left\langle \frac{x^2 + y^2}{2z} \right\rangle$  as  $\frac{\langle x^2 + y^2 \rangle}{2\langle z \rangle}$ . A check through MC simulation has been performed to validate this approximation. The approximation remains valid at the  $10^{-4}$  level for  $\mathcal{D}_0 s \simeq 0.01$ . Injecting Eq. B.20 into Eq. B.23 gives:

$$\langle r \rangle = s - \mathcal{D}_0 s^2 + \frac{2}{3} \mathcal{D}_0^2 s^3 + \frac{\frac{2}{3} \mathcal{D}_0 s^3 - \frac{4}{3} \mathcal{D}_0^2 s^4}{s - \mathcal{D}_0 s^2} \quad (\text{B.23})$$

$$\langle r \rangle = s - \mathcal{D}_0 s^2 + \frac{2}{3} \mathcal{D}_0^2 s^3 + \frac{1}{s} \left( \frac{2}{3} \mathcal{D}_0 s^3 - \frac{4}{3} \mathcal{D}_0^2 s^4 \right) (1 + \mathcal{D}_0 s) \quad (\text{B.24})$$

$$\langle r \rangle = s - \mathcal{D}_0 s^2 + \frac{2}{3} \mathcal{D}_0^2 s^3 + \frac{2}{3} \mathcal{D}_0 s^2 + \frac{2}{3} \mathcal{D}_0^2 s^3 - \frac{4}{3} \mathcal{D}_0^2 s^3 \quad (\text{B.25})$$

$$\langle r \rangle = s - \frac{1}{3} \mathcal{D}_0 s^2 + o(\mathcal{D}_0^3 s^4) \quad (\text{B.26})$$

The difference  $\Delta \langle r \rangle$  between the average distance  $\langle r \rangle$  traveled by a charged particle and a photon  $\langle r_{\text{photon}} \rangle = s$  gives the delay time  $t_{\text{del}}$ :

$$c t_{\text{del}} = \Delta \langle r \rangle = s - \langle r \rangle = \frac{1}{3} \mathcal{D}_0 s^2 \quad (\text{B.27})$$

Injecting Eq. 6.27 into Eq. B.27 and taking  $s$  to be the distance of source  $D$  gives:

$$c t_{\text{del}} = \frac{\lambda_B (ZecB)^2}{9 E^2} D^2 \quad (\text{B.28})$$

$$t_{\text{del}} \simeq 3.1 \times 10^3 \left( \frac{B}{10 \text{ nG}} \right)^2 \left( \frac{R}{10 \text{ EV}} \right)^{-2} \left( \frac{D}{1 \text{ Mpc}} \right)^2 \left( \frac{\lambda_B}{10 \text{ kpc}} \right) \text{ yrs.} \quad (\text{B.29})$$



## B.2.2 Duration of a burst

The duration of a burst  $\Delta\tau$  is given by the standard deviation of  $r(s)$ :

$$c\Delta\tau = \sqrt{\langle r^2 \rangle - \langle r \rangle^2}. \quad (\text{B.30})$$

The first term  $\langle r^2 \rangle$  is computed following:

$$\langle r^2 \rangle = \langle x^2 \rangle + \langle y^2 \rangle + \langle z^2 \rangle, \quad (\text{B.31})$$

$$\langle r^2 \rangle = \frac{4}{3}\mathcal{D}_0 s^3 - \frac{8}{3}\mathcal{D}_0^2 s^4 + s^2 - 2\mathcal{D}_0 s^3 + 3\mathcal{D}_0^2 s^4, \quad (\text{B.32})$$

$$\langle r^2 \rangle = s^2 - \frac{2}{3}\mathcal{D}_0 s^3 + \frac{1}{3}\mathcal{D}_0^2 s^4 + o(\mathcal{D}_0^3 s^5). \quad (\text{B.33})$$

The second term is also computed:

$$\langle r \rangle^2 = \left( s - \frac{1}{3}\mathcal{D}_0 s^2 \right)^2, \quad (\text{B.34})$$

$$\langle r \rangle^2 = s^2 - \frac{2}{3}\mathcal{D}_0 s^3 + \frac{1}{9}\mathcal{D}_0^2 s^4 + o(\mathcal{D}_0^3 s^5). \quad (\text{B.35})$$

Injecting Eq. B.32 and Eq. B.35 into Eq. B.30 gives the duration of a burst:

$$c\Delta\tau = \sqrt{\frac{2}{9}\mathcal{D}_0^2 s^4}, \quad (\text{B.36})$$

$$c\Delta\tau = \frac{\sqrt{2}}{3}\mathcal{D}_0 s^2. \quad (\text{B.37})$$

Replacing  $\mathcal{D}_0$  gives:

$$c\Delta\tau = \frac{\sqrt{2}\lambda_B (ZecB)^2}{9 E^2} D^2. \quad (\text{B.38})$$

One can see that the duration of the burst is a factor  $\sqrt{2}$  bigger than the delay time (Eq. B.28).

$$c\Delta\tau \simeq 4.4 \times 10^3 \left( \frac{B}{10 \text{ nG}} \right)^2 \left( \frac{R}{10 \text{ EV}} \right)^{-2} \left( \frac{D}{1 \text{ Mpc}} \right)^2 \left( \frac{\lambda_B}{10 \text{ kpc}} \right) \text{ yrs.} \quad (\text{B.39})$$

### B.2.3 Deviation from the line of sight

The angle  $\theta$  between the line of sight and the position of the cosmic ray is given by the trigonometric relation:

$$\cos \theta = \frac{z}{r} = \frac{z}{\sqrt{x^2 + y^2 + z^2}} \quad (\text{B.40})$$

Assuming a small deviation angle  $\theta$ , one can do Taylor expansion of the first term of Eq. B.40 giving  $\cos \theta = 1 - \frac{1}{2}\theta^2$ , and the second term giving  $\frac{z}{r} = 1 - \frac{x^2+y^2}{2z^2}$ . Equating the two terms gives:

$$\langle \theta^2 \rangle = \left\langle \frac{x^2 + y^2}{z^2} \right\rangle \quad (\text{B.41})$$

As before, we approximate  $\left\langle \frac{x^2+y^2}{z^2} \right\rangle$  as  $\frac{\langle x^2+y^2 \rangle}{\langle z^2 \rangle}$ . A check through MC simulation has also been performed to validate this approximation, which shows that the approximation remains valid at the  $10^{-4}$  level for  $\mathcal{D}_0 s \simeq 0.01$  as before. Injecting Eq. B.20 into Eq. B.40, gives

$$\theta_{\text{RMS}} = \sqrt{\langle \theta^2 \rangle} = \sqrt{\frac{4}{3} \mathcal{D}_0 s}, \quad (\text{B.42})$$

which matches the results from Achterberg et al., 1999. Then, replacing  $\mathcal{D}_0$  gives:

$$\theta_{\text{RMS}} = \frac{2}{3} \frac{(ZecB)}{E} \sqrt{\lambda_B D} \quad (\text{B.43})$$

In the end,

$$\theta_{\text{RMS}} \simeq 3.4^\circ \left( \frac{B}{10 \text{ nG}} \right) \left( \frac{R}{10 \text{ EV}} \right) \left( \frac{D}{1 \text{ Mpc}} \right)^{0.5} \left( \frac{\lambda_B}{10 \text{ kpc}} \right)^{0.5} \text{ yrs}. \quad (\text{B.44})$$



# Remerciements

Avant d'entrer dans des remerciements détaillés, je souhaiterais exprimer ma gratitude pour toutes les personnes que je ne vais pas citer et qui, de près ou de loin, ont participé au bon déroulement de cette thèse. En particulier celles qui ne le savent même pas !

Ceci fait, commençons ! Avant toute chose, je tiens à remercier les membres du jury qui se sont rendus disponibles et ont pris de leur temps pour lire ma thèse et écouter ma soutenance.

Une thèse ne peut bien se dérouler que dans un cadre agréable. J'ai eu la chance lors de ces trois dernières années d'être dans un groupe où je me sentais à l'aise et à ma place et où j'ai pu rencontrer des gens avenants et chaleureux. Je pense aux thésards et postdocs, Quentin, Ziwei, Pooja, Pierpaolo, et en particulier à Lucas et Antonio avec qui j'ai passé de très agréables moments pendant cette laborieuse année de rédaction. Je pense aussi aux permanents, Tiina, Piera et Isabelle, merci pour votre bienveillance.

Je n'oublie pas les excellents enseignants que j'ai pu avoir lors de mon parcours scolaire. Je pense à Stéfane qui m'a donné le goût des maths en terminale et qui a été d'un rôle crucial dans le choix de mon orientation. Arrivé à la fac, j'ai aussi fait la rencontre de KV, mon responsable de L2. L'université a besoin de prof comme toi. Je te remercie pour ta bienveillance et pour l'écoute dont tu fais preuve chaque jour pour aider les jeunes étudiants. Je pense aussi à Claire, que j'ai plus appris à connaître en tant que collègue qu'en tant que prof. Claire fait partie de ces profs qui permettent à tant de jeunes d'avoir un point d'ancrage dans ces années si particulières que sont les études !

Au quotidien, j'ai passé beaucoup de temps avec le groupe de physique théorique. C'est un peu mon groupe de cœur... J'y ai été intégré grâce à Florian et Thomas et je m'y suis senti rapidement accepté. Thomas, merci pour nos grandes discussions sur l'histoire, la géopolitique et la politique, c'est toujours un plaisir de parler avec toi, tu es un puits de connaissances. Yann, commence à faire tes économies pour le hameau, je m'occupe du reste ;) . Pour David, merci d'avoir partagé avec moi tes connaissances sur le monde pour me rendre moins bête, et surtout merci pour tes excellents gâteaux !

Évidemment, dans une thèse, il y a toujours un thésard. . . Mais aussi des directeurs de thèses ! Jonathan, à l'heure où j'écris ces mots, je te connais depuis 5 ans. . . Aujourd'hui, je ne regrette pas de ne pas avoir respecté mon emploi de temps de M1 pour aller suivre mon TD de particules et noyaux avec toi. Le hasard fait bien les choses, car l'année suivante, j'ai choisi, sans le savoir, le projet de travaux de laboratoire que tu encadrais. Merci pour ces 5 années à me guider dans ma voie de physicien ! Merci d'avoir su me rassurer dans les moments stressants et de m'avoir mis un coup de pied au \*\*\* quand j'en avais besoin. Olivier, je t'ai rencontré il y a moins de temps, mais je suis ravi d'avoir passé ces 3 dernières années à tes côtés. Nos debriefs footballistiques vont me manquer, mais je retiendrai surtout les bons moments à critiquer les collègues. Merci d'avoir toujours été disponible et de m'avoir régulièrement accordé ta confiance.

Cette thèse n'aurait pas été réussie sans le soutien constant et infaillible de mes amis. Merci à Maxime, Keke et Dipsy de me carry au quotidien sur WoW, Counter Strike ou dans la vraie vie. Je pense aussi à Florian avec qui j'ai partagé la quasi-totalité de mes études supérieures. La première fois que je t'ai rencontré, tu me saoulais à être trop fort, puis les jours ont passé et maintenant, je raconte à tout le monde à quel point tu es un génie ! Je suis ravi de t'avoir rencontré et d'avoir pu franchir ces étapes à tes côtés. Désormais, un nouveau chemin se dessine, ça va faire bizarre de ne plus travailler avec toi, mais je sais que tu ne seras jamais trop loin. Marie, je t'ai rencontré, il n'y a pas si longtemps, tu m'as beaucoup aidé durant cette thèse. En particulier la première année, quand nous étions à la maison de la Corée ensemble. Merci pour ton écoute, merci d'avoir été là. Et enfin, Dimitri et Lionel, mes amis de (très) longue date. Depuis petit, vous m'aidez à grandir et à me développer, et vous faites partie intégrante de ma vie. Vous m'avez aidé à m'épanouir et je ne vous remercierai jamais assez d'être resté aussi longtemps près de moi.

Roxane, merci de m'avoir soutenu quand je craquais. Merci d'avoir été patiente et à l'écoute. Ton rôle est si particulier, car tu es la seule à l'avoir presque vécu de l'intérieur. Désolé pour ces journées pleines de stress que je t'ai partagé et merci de ne pas m'en avoir tenu rigueur.

Côté famille, j'ai la chance d'avoir été épaulé par une super équipe de frères et sœurs, Jess, Fred, Dono et Sissi. Merci de vous être toujours rendu disponibles lors de mes retours à la maison. Ces moments étaient si importants pour moi et ont agi comme une bouffée de bien-être avant de retrouver ma stressante vie parisienne.

Pour finir, je remercie mes parents qui m'ont donné tout l'amour et la stabilité nécessaire pour m'épanouir professionnellement. Je ne pourrais jamais vous re-

mercier à la hauteur de tout ce que vous avez fait pour moi. Sur ces mots, je pense qu'il est temps de fermer la dernière page de ce livre et de passer à la suite :) .

Prenez soin de vous,

Sully.



**Titre :** Contraintes sur la population de sources des rayons cosmiques d'ultra-haute énergie à partir des données de l'observatoire Pierre Auger

**Mots clés :** Rayons cosmiques, directions d'arrivées, Auger, transitoire

**Résumé :** Cette thèse porte sur les rayons cosmiques d'ultra haute énergie (RCUHE). L'origine de ces noyaux d'atomes, découverts il y a plus d'un siècle et accélérés jusqu'à des centaines d'EeV, reste un mystère. L'observatoire Pierre Auger, situé en Argentine, a été construit pour les étudier et tenter de percer ce mystère. En 2017, l'observatoire a permis de mettre en évidence leur origine extragalactique avant, l'année suivante, de fournir des indications de corrélation entre les directions d'arrivée des RCUHE au-delà de  $\simeq 40$  EeV et un catalogue de galaxies à haut taux de formation d'étoiles. L'étude présentée dans cette thèse s'inscrit dans la continuité de ces résultats. Premièrement, une étude a été réalisée sur les méthodes de reconstruction du flux en termes de direction d'arrivée. Une seconde étude, portant sur la composition des RCUHE à partir des profondeurs de maximum de gerbe ( $X_{\max}$ ), s'est concentrée sur l'estimation du biais et des incertitudes des méthodes utilisées. Enfin, une étude phénoménologique visant à contraindre la production de RCUHE à leurs sources à partir des données de composition et de spectre en énergie a été conduite. Deux modèles d'évolution des sources ont été évalués : un scénario d'évolution des sources qui suit la densité de masse stellaire (DMS), et un scénario où l'évolution des sources suit la densité de taux de formation d'étoiles (DTFE). Ces deux mo-

dèles prennent en compte l'évolution cosmique et l'évolution locale, laquelle est propre à notre environnement extragalactique proche. Le modèle est comparé au spectre en énergie et aux distributions en  $X_{\max}$  pour des énergies au-delà de la "cheville" (caractéristique spectrale à  $\simeq 5$  EeV) et au spectre de proton (calculé à partir de l'étude en composition) en deçà de la cheville. Cette étude montre que si les protons en deçà de la cheville sont issus de la même population de sources que celle au-delà, un échappement différent entre les protons et les autres noyaux est nécessaire pour expliquer les données dans le cas du scénario DMS, tandis que pour le scénario DTFE, les protons sous la cheville sont la conséquence de produits secondaires issus de la propagation des RCUHE dans l'univers. Dans le dernier chapitre de la thèse, ces modèles cosmiques ont été associés à un catalogue de près de 400,000 galaxies dans le but de déduire les cartes du ciel attendues. Ces modèles de ciel sont comparés aux données reconstruites à l'observatoire Pierre Auger et par l'expérience Telescope Array. Pour finir, l'aspect transitoire des sources de RCUHE a été caractérisé, en prenant en compte l'effet de divers champs magnétiques (Galactique, feuillet local et extragalactique). Cette étude a permis de contraindre le taux d'évènements produisant des RCUHE, ainsi que le champ magnétique imprégnant le feuillet local.



**Title :** Constraints on the population of ultra high energy cosmic ray sources inferred from the data of the Pierre Auger Observatory

**Keywords :** UHECR, sources, Auger, transient, arrival directions

**Abstract :** This thesis focuses on ultra high energy cosmic rays (UHECRs). The origin of these atomic nuclei, discovered over a century ago and accelerated to hundreds of EeV, remains a mystery. The Pierre Auger Observatory, located in Argentina, was built to study them and try to unravel this mystery. In 2017, the observatory showed the first evidence of an extragalactic origin of UHECRs. The following year, the observatory found indications of correlation between the arrival directions of UHECRs above  $\simeq 40$  EeV and a catalogue of star-forming galaxies. The study presented in this thesis is a continuation of the results from the Auger Observatory. Firstly, a study was carried out on the methods for reconstructing the flux as a function of arrival direction. A second study, concerning the estimation of the bias and the uncertainties of the composition of UHECRs from the distribution of the depth of shower maximum ( $X_{\max}$ ), was carried out. Finally, a phenomenological study, based on a tool developed by the collaboration (the Combined Fit) to constrain the production of UHECRs at their sources from composition and spectrum data, was used. Two source evolution models were considered, a source evolution that follows the stellar mass density (SMD) and a scenario where the source evolution follows

the star formation rate density (SFRD). These two evolution models take into account the cosmic and local evolution, which is specific to our immediate extragalactic environment. The model is compared to the all-particle spectrum and  $X_{\max}$  distributions above the ankle and to the proton spectrum (calculated from the composition study) below the ankle ( $\simeq 5$  EeV). This study shows that if the protons are from the same source population below the ankle and above it, a different escape between the protons and the other nuclei is needed to explain the data in the case of the SMD scenario, while for the SFRD scenario, the protons below the ankle are the consequence of secondary products from the propagation of UHECRs in the universe. In the last chapter of the thesis, these cosmic models have been associated with a catalogue of nearly 400,000 galaxies in order to derive sky maps expected from the model. These sky maps are compared with data reconstructed by the Pierre Auger Observatory and the Telescope Array experiment. Finally, the transient aspect of UHECR sources associated with the effect of various magnetic fields (Galactic, Local Sheet, and extragalactic) has been explored. This study constrains the rate of events producing UHECRs, as well as the magnetic field permeating the Local Sheet.