

Amplifying UHECR arrival direction information using mass estimators at the Pierre Auger Observatory

Lorenzo Apollonio^{a,b,*} for the Pierre Auger Collaboration

^aDipartimento di Fisica, Università degli Studi di Milano, Milano, Italy

^bIstituto Nazionale di Fisica Nucleare, Sezione Milano, Milano, Italy

Full author list: https://www.auger.org/archive/authors_2024_11.html

E-mail: spokesperson@auger.org

The origin of Ultra-High-Energy Cosmic Rays (UHECRs) is one of the biggest mysteries in modern astrophysics. Since UHECRs are deflected by Galactic and extragalactic magnetic fields, their arrival directions do not point to their sources. Previous analyses conducted on the arrival directions of high-energy events ($E \geq 32 \text{ EeV}$) recorded by the Surface Detector of the Pierre Auger Observatory have not shown significant anisotropies. The largest excess found in the first 19 years of data - at the 4.0σ level - is in the region around Centaurus A, and it is also the driving force of a correlation of UHECR arrival directions with a catalog of Starburst Galaxies, which is at the 3.8σ level. Since UHECRs are mostly nuclei, the lightest ones (least charged) are also the least deflected. While the mass of the events can be estimated better using the Fluorescence Detector of the Pierre Auger Observatory, the Surface Detector provides the necessary statistics needed for astrophysical studies. The introduction of novel mass-estimation techniques, such as machine learning models and an algorithm based on air-shower universality, will help identify high-rigidity events in the Surface Detector data of the Pierre Auger Observatory. With this work, we present how event-per-event mass estimators can help enhance the sensitivity in the search for anisotropies in the arrival directions of UHECRs at small and intermediate angular scales using simulations.

7th International Symposium on Ultra High Energy Cosmic Rays (UHECR2024)
17 – 21 November 2024
Malargüe, Mendoza, Argentina

*Speaker

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0). All rights for text and data mining, AI training, and similar technologies for commercial purposes, are reserved. ISSN 1824-8039. Published by SISSA Medialab.

<https://pos.sissa.it/>

1. Introduction

Ultra-High-Energy Cosmic Rays (UHECRs) with energies beyond 1 EeV are the highest energy particles observed by humankind. While being studied for over a century, the origin of UHECRs is still unknown [1]. Significant findings have been made studying the arrival directions of the high-quality data of the Pierre Auger Observatory [2], the largest cosmic-ray observatory in the world, located in the southern hemisphere in Malargüe, Argentina. Two independent detectors operate at the observatory: the Fluorescence Detector (FD) and the Surface Detector (SD) [3, 4]. The FD detects UHECR showers collecting the fluorescence light emitted in the atmosphere. As it can operate only on moonless nights with a clear atmosphere, its duty cycle is $\sim 15\%$ of the time. The SD records the UHECR air-shower particle footprints using an array of 1600 water-Cherenkov detectors at the ground, placed in a triangular grid with 1500 m spacing for its main array. In contrast to the FD, the SD duty cycle is $\simeq 100\%$.

Arrival-direction analyses on a large angular scale have shown the presence of a dipolar modulation of the UHECRs with energy larger than 8 EeV with a significance of 6.8σ [5]. Analyses on small and intermediate angular scales using high-energy cosmic rays (≥ 32 EeV) have yielded non-significant results [6, 7]. Nevertheless, searches for a signal around the position of Centaurus A have shown the presence of an excess with post-trial p -value of 3.0×10^{-5} (4.0σ one-sided). Additionally, a hint of correlation (post-trial p -value of 6.6×10^{-5} , equal to 3.8σ one-sided) with a catalog of Starburst Galaxies (SBGs) has been observed with a likelihood analysis. However, the correlation of UHECR arrival directions with the positions of SBGs is driven by the local excess in the Centaurus A region. Other analyses have shown no indications of anisotropy in the data; for example, an autocorrelation analysis has resulted in a post-trial p -value of 0.24.

As the magnetic deflection of cosmic rays scales with the inverse rigidity R (i.e. the ratio of the energy of the cosmic ray and its charge), we expect high-rigidity events to be less deflected and thus more spatially correlated with their sources. Identifying and selecting the high-rigidity events in the dataset could, therefore, help to amplify arrival direction information in the data, at the price of reducing the statistics. As arrival direction studies require the largest possible statistics, the advent of mass estimators able to reconstruct mass-dependent information with the SD of the Pierre Auger Observatory (e.g. [8–11]) will be crucial to possibly enhance the arrival-direction information. In this proceeding, we review the SD-based mass estimators that are currently being used and developed by the Pierre Auger Collaboration. We present analyses based on air-shower simulations to demonstrate how mass estimators can be utilized in the autocorrelation analyses and in the targeted search for a signal around Centaurus A.

2. Mass estimators

Extensive air showers from UHECRs exhibit several mass-dependent observables. The slanted atmospheric depth X_{\max} of the maximum particle content of the shower is one of them. The depth X_{\max} is dependent on the logarithmic energy per nucleus $\lg(E_0/A)$, with E_0 being the energy of the primary UHECR generating the shower, and A being the atomic mass number. Thus, at the same energy, a light (heavy) nucleus induces a cascade that reaches its maximum later (earlier) in the atmosphere, producing a deeper (shallower) X_{\max} . The depth of the maximum shower can be

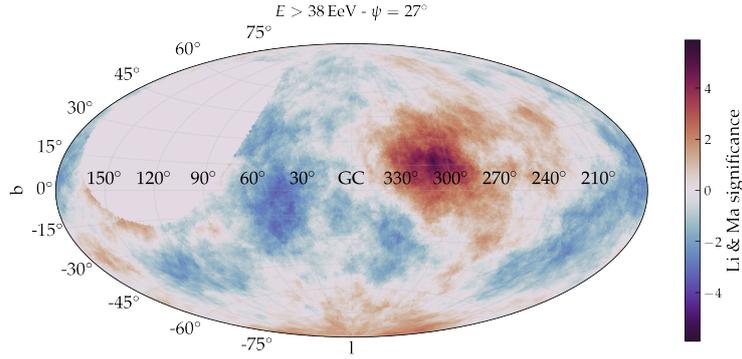


Figure 1: Local Li-Ma significance map within a top-hat window of $\psi = 27^\circ$ for cosmic rays at $E > 38 \text{ EeV}$ [6, 7].

estimated from the time-dependent signal traces in the water-Cherenkov detectors. This connection was already used to obtain mass-composition information with the surface detector of the Pierre Auger Observatory [12], but is now being refined using more precise mass estimators for arrival direction studies. Another promising mass-dependent observable is the relative number of muons in a shower, R_μ . The number of muons produced per nucleon is dependent both on the energy and the mass number, increasing as $\sim (E_0/20 \text{ GeV} \cdot 1/A)^\beta$ with $\beta \simeq 0.95$. Thus, at the same energy, light nuclei produce fewer muons in air showers than heavy nuclei [13].

Different mass estimators have been developed in the Pierre Auger Collaboration using surface detector information, particularly deep neural networks (DNN) and a likelihood fit based on air-shower universality [8–10] are being used to extend the physics reach of the existing data of the observatory. The universality model relies on the assumption that electromagnetic particles in air showers generated by different primary particles at different energies E_0 show the same lateral and longitudinal distributions as well as universal spectra. The model extends the particle content of air showers to hadronic particles and their decay products, while their relative contribution R_μ is left as a free parameter. In this way, universality can be restored for hadronic showers. The air shower universality fit can be used to estimate X_{max} and R_μ . However, for events reconstructed by the surface detector only, the energy is estimated by the shower footprint, and the estimate of R_μ is expected to be biased. Thus, so far R_μ can be estimated accurately only using *Golden Hybrid* showers, i.e. showers that are both independently reconstructable by the fluorescence and surface detectors, where the energy estimation with the fluorescence detector can be used as an additional input. The machine learning models are trained to map the input data from the surface detector to a desired mass-dependent observable, i.e. X_{max} and/or R_μ , using large simulation data bases. The AIXNET model is trained on shower simulations to reconstruct X_{max} from the footprint and detector time traces [8]; the resulting data set is finally calibrated to direct measurements, using the fluorescence detector data where available.

3. Simulated Analysis for Autocorrelation Search

The model described in [9] is trained to estimate the depth of the shower maximum and the muon content of the shower, and combines them to a global mass-dependent proxy observable Y

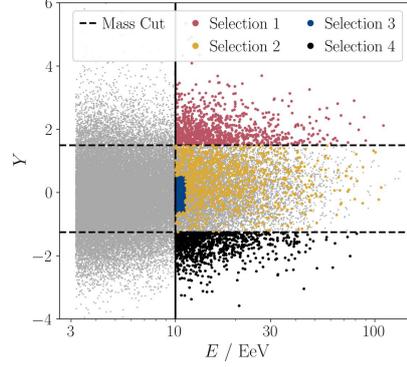


Figure 2: The estimated mass-proxy observable Y as a function of the primary energy for simulated air showers. Four regions are defined for later data selection. See the text for details.

(note the factor of -1 with respect to the definition in [9]),

$$Y = 0.5 \frac{X_{\max} - \langle X_{\max} \rangle}{\sigma(X_{\max})} - \frac{R_{\mu} - \langle R_{\mu} \rangle}{\sigma(R_{\mu})}. \quad (1)$$

A similar approach for a mass-dependent proxy-observable is described in [14]. For larger values of Y , we expect generally lighter masses of UHECRs; Y is intrinsically unit-less and depends pseudo-linearly on the logarithmic atomic mass $\ln A$. The performance of the network to study autocorrelation of arrival directions in the data of the Pierre Auger Observatory was tested using a simulation data base of realistic events. For the simulation data base, we use Offline simulations [15] of CORSIKA showers [16], generated with the EPOS-LHC model of hadronic interaction [17]. The arrival directions of the events were simulated considering uniformly distributed sources with source density ρ_{src} . Each source has the same emission spectrum defined by a power-law with a rigidity cut-off, as in [18]. The attenuation in the UHECR propagation is evaluated with CRPROP3 [19], and the galactic magnetic deflections are computed considering the Jansson-Farrar model [20]. The Pierre Auger Observatory exposure is taken into account when completing the simulations. The simulated X_{\max} are then mapped to the arrival direction of the events, matching the energy at Earth.

The mass-proxy observable Y is used to define various data selections, which will be subsets of the full event data to perform autocorrelation analyses. The distribution of reconstructed Y from simulated showers is depicted in Fig. 2. Selection 1 is a light and thus high-rigidity data set (it is defined by choosing the 2000 events with the largest Y above 10 EeV); Selection 2 is defined by constant estimated X_{\max} (events with distance in X_{\max} lower than 25 g/cm^2 from the average X_{\max} of the data set), and is thus approximately resulting in a constant mid-rigidity data set; Selection 3 is expected to be approximately constant Y in a narrow region of primary energy (between 10 EeV and 11.2 EeV), and thus composition, and is expected to be a pure composition data set; Selection 4 is a heavy and thus low-rigidity data set used as a cross-check in the analysis (it is defined by choosing the 2000 events with the lowest Y above 10 EeV). The four selections are defined to ensure no overlap between them. The evaluation of the simulation data sets is depicted in Fig. 3.

Each panel depicts a simulated source scenario. From left to right, the first panel considers $\rho_{\text{src}} = 10^{-2} \text{ Mpc}^{-3}$, the second panel considers $\rho_{\text{src}} = 5 \times 10^{-3} \text{ Mpc}^{-3}$, and the third panel considers

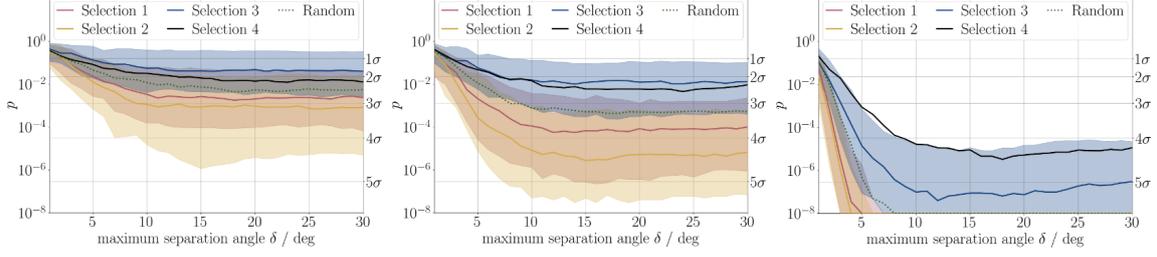


Figure 3: Significance of the autocorrelation signals for simulated data sets as a function of the separation angle of events. The three panels depict three different source scenarios. See the text for details.

$\rho_{\text{src}} = 10^{-3} \text{ Mpc}^{-3}$. In each panel, we show the evolution of the p -value of the autocorrelation analysis as a function of the separation angle considered. We show the results obtained for each selection and the outcomes expected for a random selection (green dotted line). The solid line is the median value obtained for 1000 simulations, while the shaded area shows the minimum and maximum p -values obtained in the relative selection. Fig. 3 shows that Selection 1 and Selection 2 are optimal, as they give in every source scenario a lower p -value than expected from a random selection. On the other hand, Selection 3 rarely gives a lower p -value than the random expectation, while Selection 4 gives the larger p -values as expected.

4. Simulated Analysis for Centaurus A Targeted Search

Using simulations, we test several scenarios to investigate the best strategy to be applied to enhance a possible high-rigidity excess around Centaurus A. We focus primarily on finding the optimal number of low-rigidity events that have to be discarded to maximize the significance. The simulations are completed defining the Centaurus A region as the circle in the sky with a radius equal to 27° centered around Centaurus A and considering the events with reconstructed energy $E \geq 38 \text{ EeV}$. The combination of the two parameters (radius and energy threshold) was chosen after the optimization that was completed on the dataset of events above 32 EeV [6, 7].

We investigate the hypothesis that a high-rigidity signal population is present in the data set and it is possibly more closely concentrated around Centaurus A, driving the excess that we observe in the data. We assume that it is real and not due to statistical fluctuations. In the arrival-direction data, we define a signal population S and background population B , with S being present only inside the Centaurus A region and B present both inside and outside. Since the considered mass can only be determined on events with a zenith angle smaller than 60° (vertical events), we randomly assign $N_{\text{ver}}^{\text{sig}}$ vertical events in the Centaurus A region as a signal, and associate them with a simulated light mass. We set all the other events of the data set as background and we assign them a heavy mass according to our simulation scenario. For each simulation scenario, we consider different Centaurus A signal fraction ($f_{\text{sig}}^{\text{Cen A}}$), which is the fraction of vertical events in the Centaurus A region marked as a signal.

We then complete the simulations by mapping showers reconstructed with the universality mass estimator to real events, using close matching values of the primary energy and the zenith angle. As in Section 3, for the simulation data base, we use Offline simulations [15] of CORSIKA showers [16], generated with the EPOS-LHC model of hadronic interactions [17]. The air-shower

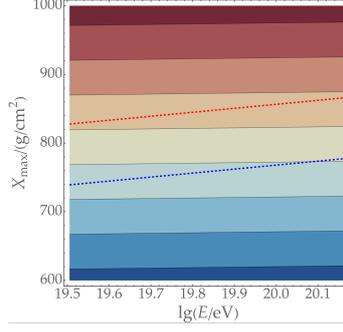


Figure 4: Two-dimensional sketch of contours with equal-estimated \tilde{R} . The colour-coded shades show larger (red) and lower (blue) values of $\ln \tilde{R}$ as a function of X_{\max} and $\lg(E_0/eV)$. The red and blue dashed lines depict the EPOS-LHC expectation values for X_{\max} from simulated proton and iron showers, respectively.

simulations are now completed assuming four primaries: proton (p), helium (He), oxygen (O) and iron (Fe). Two of the considered scenarios are as follows:

- (i) **helium scenario:** the signal composition is made of He particles, while the background composition follows the Auger mix composition;
- (ii) **mixed composition scenario:** in this scenario, we consider possible deflections and place lower-rigidity events further away from Centaurus A. We considered that the cosmic rays are deflected following a von Mises distribution ($p_{\text{VM}}(d, R)$), where d is the angular distance from Centaurus A and R is the rigidity of the event), which, for an event with atomic number Z_i and energy E , is defined by the concentration parameter

$$k_i = \frac{1}{\Theta_0^2} \left(\frac{E/Z_i e}{10 \text{ EV}} \right). \quad (2)$$

We impose $\Theta_0 = 16.8^\circ$. This value was taken from the outcome of the scenario considering Centaurus A as UHECR source in [18]. The probability for an event to be of the atomic species Z_i is given by

$$w_i = \frac{p_{\text{VM}}(d, E/Z_i e)}{\sum_j p_{\text{VM}}(d, E/Z_j e)}, \quad (3)$$

where $Z_j \in [1, 2, 8]$, i.e. we consider a mixed population of proton, helium and oxygen. As in the previous scenario, the background population follows the Auger mix.

For each scenario, we impose three different values of $f_{\text{sig}}^{\text{Cen A}}$ (0.25, 0.5 and 0.75), and complete 1000 sky simulations.

The simulated events in each sky are ordered in a rigidity-proxy parameter to select a data set in which the signal is amplified. The rigidity order is obtained using the depth of shower maximum at 10^{19} eV (X_{\max}^{19}),

$$X_{\max}^{19} = X_{\max} - D \lg \left(\frac{E}{10^{19} \text{ eV}} \right), \quad (4)$$

where the energy dependence of X_{\max} is removed by subtracting the decadal elongation rate D . We impose $D = 58 \text{ g/cm}^2$, which is approximately equal in all hadronic interaction models. The

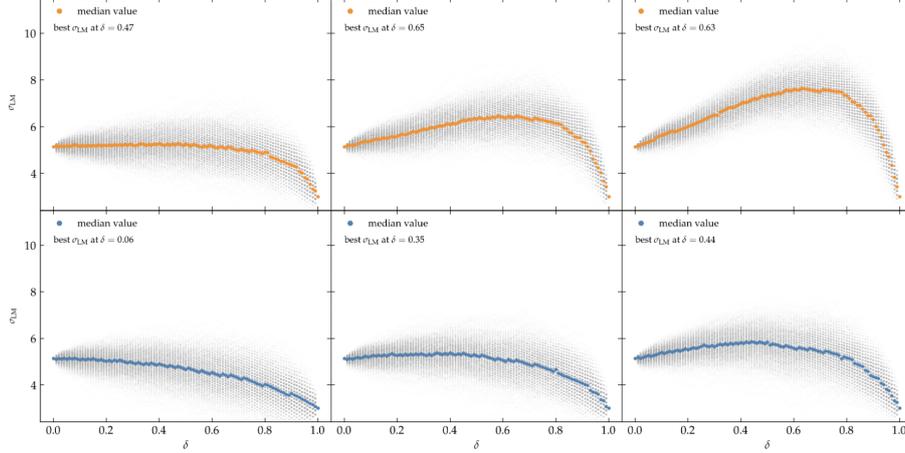


Figure 5: Evolution of the Li-Ma significance (σ_{LM}) of the Centaurus A region excess as a function of the fraction of the low-rigidity events discarded (δ) for the helium scenario (top plots) and the mixed composition scenario (bottom plots). The left plots consider $f_{\text{sig}}^{\text{Cen A}}$ equal to 0.25, the middle plots equal to 0.5 and the right plots equal to 0.75.

rigidity proxy parameter \tilde{R} is then defined as

$$\tilde{R} := E_0/\tilde{Z}, \quad \text{using} \quad (5)$$

$$1/\tilde{Z} := 2 \exp\left((X_{\text{max}}^{19} - x_{\text{ref}})/\lambda\right)/e, \quad (6)$$

with \tilde{Z} being a real number in units of the elementary charge e . We note here that \tilde{R} is not the absolute rigidity of the events, but it preserves the rigidity order of the dataset, given an ideal reconstruction. For Eq. (6), we make use of X_{max}^{19} , the (arbitrary) reference value $x_{\text{ref}} = 742 \text{ g/cm}^2$ (we selected the average value of X_{max}^{19} for the universality reconstructed dataset), and the increase in the average depth of the shower maximum per unit step in $\ln(A)$ based on the Heitler-Matthews model, $\lambda \simeq 22.3 \text{ g/cm}^2$, which is approximately the same for all hadronic interaction models. In Fig. 4, we show the two-dimensional sketch of equal \tilde{R} as a function of X_{max} and $\lg(E_0/\text{eV})$. The red (blue) dotted line is the X_{max} expectation value for proton (iron) in the EPos-LHC model, while the lines separating the coloured sectors are lines of equal \tilde{R} .

In Fig. 5, we show the evolution of the significance of the Centaurus A excess, evaluated using the Li-Ma significance (σ_{LM}) [21], as a function of the fraction of low-rigidity events discarded (δ_{rem}). The top plots show the results for the helium scenario, while the bottom plots show the results for the mixed composition scenario. The left plots show the results obtained for $f_{\text{sig}}^{\text{Cen A}} = 0.25$, the middle plots for $f_{\text{sig}}^{\text{Cen A}} = 0.5$, and the right plots for $f_{\text{sig}}^{\text{Cen A}} = 0.75$. Fig. 5 shows that, using mass-estimators, we can enhance the significance of an excess formed by high-rigidity events present in the Centaurus A region. Furthermore, the significance enhancement is larger when we have a larger mass difference between the signal population and the background population, and the significance increases with the signal fraction, as expected. Finally, Fig. 5 shows that if we are able to enhance a signal produced by high-rigidity events, we expect to find the maximum significance, discarding a substantial portion of the low-rigidity events present in the data set ($\delta \sim 0.4 - 0.6$). We note,

however, that these results are obtained in the specific assumption that a high-rigidity signal is present only in the Centaurus A region; thus results on data may vary significantly.

5. Conclusions and outlook

As the UHECR magnetic deflections scale inversely to the rigidity of the events, the introduction of mass estimators can lead to the identification of the least deflected particles. However, as arrival-direction studies require the maximum possible statistics, using estimators that can be applied to events detected with SD is needed. With this proceeding, we completed simulations to study the performances in arrival direction studies of the mass estimators introduced by the Pierre Auger Collaboration, focusing on the autocorrelation analysis and the targeted analysis around Centaurus A. In Section 3, we considered a transformer-based neural network and defined four possible selections to enhance the significance of the autocorrelation analysis with the data above 10 EeV. Two selections, a light-related one (Selection 1) and an X_{\max} -based one (Selection 2) resulted to be very promising, as they lead, on the same data sets, to a better significance than the one obtained from random selections. The simulations performed in Section 4 show how the significance of the excess around Centaurus A evolves as a function of the fraction of low-rigidity events discarded, assuming that high-rigidity particles are more concentrated around it. Furthermore, they show that, if a high-rigidity signal is present in the region, we expect to be able to enhance it by discarding a moderate fraction of the events in the data set.

References

- [1] R. Alves Batista et al., *Front. Astron. Space Sci.* **6** (2019) 23 [1903.06714].
- [2] PIERRE AUGER Collaboration, *Nucl. Instrum. Meth. A* **798** (2015) 172 [1502.01323].
- [3] PIERRE AUGER Collaboration, *Nucl. Instrum. Meth. A* **620** (2010) 227 [0907.4282].
- [4] PIERRE AUGER Collaboration, *Nucl. Instrum. Meth. A* **586** (2008) 409 [0712.2832].
- [5] PIERRE AUGER Collaboration, *Astrophys. J.* **976** (2024) 48 [2408.05292].
- [6] PIERRE AUGER Collaboration, *Astrophys. J.* **935** (2022) 170 [2206.13492].
- [7] PIERRE AUGER Collaboration, *PoS ICRC2023* (2023) 252.
- [8] PIERRE AUGER Collaboration, *JINST* **16** (2021) P07019 [2101.02946].
- [9] PIERRE AUGER Collaboration, *PoS ICRC2023* (2023) 371.
- [10] M. Stadelmaier, R. Engel, M. Roth, D. Schmidt et al., *Phys. Rev. D* **110** (2024) 023030 [2405.03494].
- [11] PIERRE AUGER Collaboration, *PoS UHECR2024* (2024) 080.
- [12] PIERRE AUGER Collaboration, *Phys. Rev. D* **96** (2017) 122003 [1710.07249].
- [13] J. Matthews, *Astropart. Phys.* **22** (2005) 387.
- [14] M. Stadelmaier, *Astrophys. J.* **946** (2023) 100.
- [15] S. Argiro, S.L.C. Barroso, J. Gonzalez, L. Nellen et al., *Nucl. Instrum. Meth. A* **580** (2007) 1485 [0707.1652].
- [16] D. Heck, J. Knapp, J. Capdevielle, G. Schatz et al., *Report FZKA* **6019** (1998) .
- [17] T. Pierog, I. Karpenko, J.M. Katzy, E. Yatsenko et al., *Phys. Rev. C* **92** (2015) 034906 [1306.0121].
- [18] PIERRE AUGER Collaboration, *JCAP* **2024** (2024) 022.
- [19] R. Alves Batista, A. Dundovic, M. Erdmann, K.-H. Kampert et al., *JCAP* **05** (2016) 038 [1603.07142].
- [20] R. Jansson and G.R. Farrar, *Astrophys. J.* **757** (2012) 14 [1204.3662].
- [21] T.P. Li and Y.Q. Ma, *Astrophys. J.* **272** (1983) 317.