PROCEEDINGS OF SCIENCE

Overcoming Challenges in Finding Ultra-High-Energy Cosmic Ray Sources with a Bayesian Hierarchical Framework: Impact of the Galactic magnetic field and mass composition

Keito Watanabe,[a,*] **Anatoli Fedynitch,**[b,d] **Francesca Capel**[c] **and Hiroyuki Sagawa**[d]

[a] *Universität zu Köln, I. Physikalisches Institut, Zülpicher Str. 77, D-50937 Köln, Germany*

[b] *Institute of Physics, Academia Sinica, Taipei City, 11529, Taiwan*

[c] *Max Planck Institute for Physics, Föhringer Ring 6, 80805 Munich, Germany*

[d] *Institute for Cosmic Ray Research, the University of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa, Chiba 277-8582, Japan*

E-mail: watanabe@ph1.uni-koeln.de

The origins of ultra-high-energy cosmic rays (UHECRs) continue to elude us due to the intricate interplay between their acceleration and escape mechanisms, propagation physics, and detection limitations. Developing a comprehensive statistical model is further complicated by the numerous parameters and uncertainties involved. In this study, we present a statistical analysis of the connection between UHECRs and astrophysical sources using a Bayesian hierarchical framework introduced in Capel and Mortlock [1]. We expand on this framework by incorporating the effects of the Galactic magnetic field on UHECR propagation and introduce a novel method for inferring the nuclear composition at the source based on the observed composition at Earth. Our approach respects rigidity-dependent deflections and energy-loss horizons for each event, depending on its observed mass, energy, and arrival direction. We applied our approach to publicly available data from the Telescope Array. Since event-by-event composition-sensitive variables are not yet available, we assume an average observed composition. By providing a more accurate understanding of the nuclear composition at the source, our approach can improve our knowledge of the processes that generate cosmic rays. Our results offer new insights into the differences between Northern and Southern skies and enhance our ability to understand the astrophysical phenomena underlying UHECR production.

38th International Cosmic Ray Conference (ICRC2023)
26 July - 3 August, 2023
Nagoya, Japan

ICRC2023
38th International Cosmic Ray Conference
The Astroparticle Physics Conference

*Speaker

PoS(ICRC2023)479

1. Introduction

The association of UHECR events observed by the two large-scale observatories, Telescope Array (TA) and the Pierre Auger Observatory (PAO), with nearby sources is a question of great interest. Both observatories have been operating for over 15 years and have collected unprecedented amounts of UHECR data at the highest energies. Even as more data are collected, our understanding of the astrophysical objects responsible for UHECR acceleration remains incomplete. The complexity of UHECR phenomenology challenges the interpretation of the data and motivates the development of sophisticated analysis techniques to make the most of the information available to us.

In this work, we report on the progress in developing a new physical model for UHECR propagation and its integration in the statistical framework for source–UHECR association developed in Capel and Mortlock (CM19) [1] and based on pioneering works by Watson et al. and Soiaporn et al. [2, 3]. The framework is a Bayesian hierarchical model implemented in Stan [4] via a Python interface. Stan is a software package for high-performance statistical computation that contains an implementation of a Hamiltonian Markov Chain Monte Carlo that we make use of to sample from high-dimensional parameter spaces efficiently. The advantage of this approach is that it allows us to incorporate realistic physical models for the UHECR transport, the source properties and detector model, and relevant uncertainties associated with these models. We have previously reported on our efforts to include the Galactic magnetic field (GMF) lensing effects and non-proton source compositions into this framework in Watanabe et al. [5]. Here, we describe our recently improved treatment of the UHECR composition and its impact on possible source–UHECR associations.

2. Model for Intergalactic Propagation of UHECR nuclei

As UHECRs traverse the neighboring intergalactic medium (within tens of Mpc), they encounter two primary energy loss processes: a magnetic horizon at lower rigidities (EV and below) and photo-hadronic losses at extremely high energies, notably greater than 10 EeV (see, e.g., [6]). The interactions between the cosmic microwave and infrared backgrounds and UHECRs differ significantly between protons and nuclei. Protons interact through photo-hadronic excitation of a Δ^+ resonance or higher order processes, leading to kinetic energy loss due to recoil from the subsequent two-body decay of the Δ^+. For nuclei, the dominant threshold process is the Giant Dipole Resonance (GDR), where excitation energy dissipates through the evaporation of one or more nucleons or light fragments ($< \alpha$) [7].

Two important aspects arise from these interactions. Firstly, due to the relatively weak nuclear binding energies, the loss of boost ($\Gamma = E/m$) for nuclei is negligible. Secondly, most light stable-valley elements with mass $A \lesssim 56$ and charge Z have approximately the same number of neutrons and protons ($Z/A \sim 1/2$). These considerations lead to an **approximate conservation of rigidity** $R_S = E/Z \equiv R_E$ [1], implying that a nucleus and its nuclear disintegration products retain identical magnetic history during their journey from the source to Earth. We have verified the validity of this approximation using CRPropa3 [8] and use it throughout the work. This effect also suggests that nucleons ($Z/A = 1$) resulting from the disintegration of nuclei from a specific source may

[1]The subscripts refer to (S)ource and (E)earth throughout the text.

experience spatial decorrelation, making them indistinguishable from other background particles originating from higher redshifts.

In this study, we center our attention on energies > 32 EeV, where the PAO reports a mass composition primarily dominated by elements from the Nitrogen group, with minimal contamination from heavier elements [9]. As defined in the Global Spline Fit (GSF) [10], we categorize relative fractions of elements into the following mass groups (one unit wide in $\langle \ln A \rangle$): MG_1 for protons; MG_2 for $2 \leq A < 8$; MG_3 for $8 \leq A < 20$; MG_4 for $20 \leq A \leq 54$. We assume a composition comprised of a single nuclear mass group (MG_2–MG_4) across the entire sky.

The energy or mass loss model is derived from simulations using CRPropa3 by injecting N_S nuclei of mass A at a distance D_k on a grid of source rigidities R_S, which within our approximation are equal to rigidities R_E at Earth (R in the following). We select the subset of N_E events at Earth, which include at least one residual nucleus of arrival mass A_E within a given mass group MG_i and define the general composition weights such that

$$w_{\mathrm{C}}(R, D, A_E, A_S) = \frac{N_{\mathrm{E}}(R, D, A_E)}{N_{\mathrm{S}}(R, D, A_S)}. \tag{1}$$

Marginalizing over the mass groups then gives

$$w_{\mathrm{MG}}(R, D, A_S, \mathrm{MG}_i) = \sum_{A_E \in \mathrm{MG}_i} \frac{N_{\mathrm{E}}(R, D, A_E)}{N_{\mathrm{S}}(R, D, A_S)}, \quad A_S \geq \min \mathrm{MG}_i. \tag{2}$$

These weights simultaneously incorporate the effect of mass loss, equivalent to energy loss in the proton case, and total particle loss due to passing certain thresholds, such as the range of masses within a mass group, equivalent to arrival energy dropping below a threshold in the proton case.

Normalizing the weights in Equation 2 into a probability density function (composition PDF) gives

$$\mathrm{P}_{\mathrm{k}}(A_S \mid R, D_k, \mathrm{MG}_i) = \frac{w_{\mathrm{MG}}(R, D_k, A_S, \mathrm{MG}_i)}{\sum_{A_S=\min \mathrm{MG}_i}^{A_{\max}} w_{\mathrm{MG}}(R, D_k, A_S, \mathrm{MG}_i)}, \tag{3}$$

we can quantify the rigidity dependence of the mass composition at a source located at a distance D_k. Effectively, Equation 3 is a relative normalization for the spectrum of each element. As demonstrated in Figure 2, the average composition of the arrival spectrum is centered on the desired mass group. This composition choice minimizes the luminosity of the elements that don't contribute to the desired spectrum and maximizes that of elements actively delivering arrival elements within the desired MG. For harder spectral indices, photo-disintegration effects dominate, producing an over-abundance of protons. However, they are suppressed as they only remain with half the original rigidity, which in turn is suppressed due to the steepness of the spectrum.

In our model, nuclei of all masses escape a steady-state source environment following a power law with index α_S in rigidity between $R_{\min} = 1\,\mathrm{EV}$ and $R_{\max} = 1\,\mathrm{ZV}$. However, as outlined above, each mass spectrum is weighted by the corresponding factor from the composition PDF

$$\frac{dN_k}{dRdt}(R, D_k, A_S, \alpha_s, \mathrm{MG}_i) = k_{\alpha_s,k} \mathrm{P}_{\mathrm{k}}(A_S \mid R, D_k, \mathrm{MG}_i) Z^{1-\alpha_s} R^{-\alpha_s} \exp(-R/R_{\mathrm{cutoff,k}}). \tag{4}$$

Since at very high rigidities, it is challenging to sample sufficient N_E in CRPropa3 and generate sufficiently smooth arrival spectra; we add an additional exponential cutoff, which is algorithmically

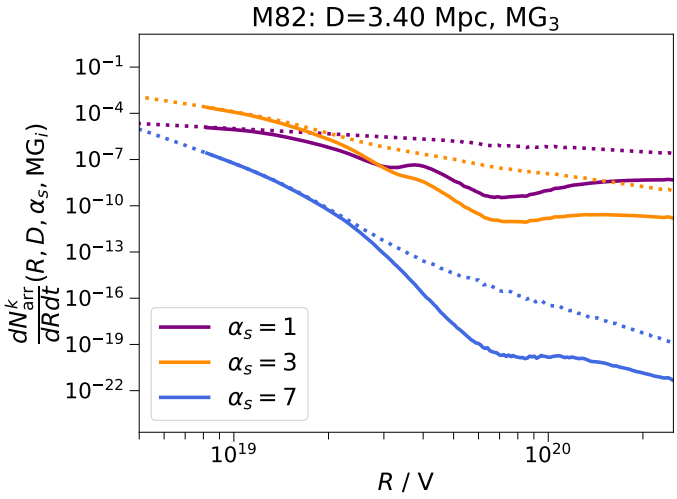

Figure 1: The rigidity spectrum at the source (dashed) and at Earth (solid) for different source spectral index α_s. The convolution with the composition weights takes propagation losses into account as seen by the dip in the spectrum at Earth.

chosen to lie above the cutoff seen in the arrival spectra. The normalization $k_{\alpha_s,k}$ is determined such that the luminosity L_k (at the source) is equal for all sources within a catalog.

Using the weights from Equation 1, we can calculate the rigidity spectrum at Earth using:

$$\frac{dN^k_{\rm arr}}{dRdt}(R, D_k, \alpha_s, {\rm MG}_i) = \sum_{A_S} w_{\rm MG}(R, D_k, A_S, {\rm MG}_i) \frac{dN_k}{dRdt}(R, D_k, A_S, \alpha_s, {\rm MG}_i). \tag{5}$$

Figure 1 illustrates the correspondence between arrival spectrum at Earth and the source spectrum. From the spectrum, we derive the expected quantities of interest, such as the mean rigidity, mean number of detected particles, mean deflection angle etc.

We model the background as a source located at $D = 0\,$Mpc, assuming an isotropic flux F_0 following a power law $dN_b/dRdt \propto R^{-\alpha_b}$ with a spectral index that can differ from the α_s, since it already incorporates any possible propagation effects.

The magnetic deflections from the extra-Galactic magnetic field (EGMF) for a UHECR is described analytically as derived in e.g. [11]. This formalism assumes that the EGMF is a Gaussian random field described by a Kolmogorov power spectrum, an RMS EGMF strength, $\bar{B}$, and a coherence length, l_c. The GMF deflections are modeled by backtracking each UHECR with a detected arrival direction $\hat{\omega}$ and rigidity $\hat{R}$ within a given mass group using CRPropa3, assuming a GMF model postulated by Janson & Farrar [12]. The details of this approach are given in [5].

As in CM19, we model the detected arrival directions $\hat{\omega}$ by using a von-Mises-Fischer (vMF) distribution [13] where ω is the reported "true" arrival direction, and κ_d is the concentration parameter, proportional to the inverse square of the angular reconstruction uncertainty σ_ω. The detector exposure is calculated following CM19, using the reported total exposure α_T. At this preliminary stage of work, the detected rigidities $\hat{R}$ are modeled by a normal distribution centered around the "true" arrival rigidity R with a width scaled by the energy reconstruction uncertainty f_E.

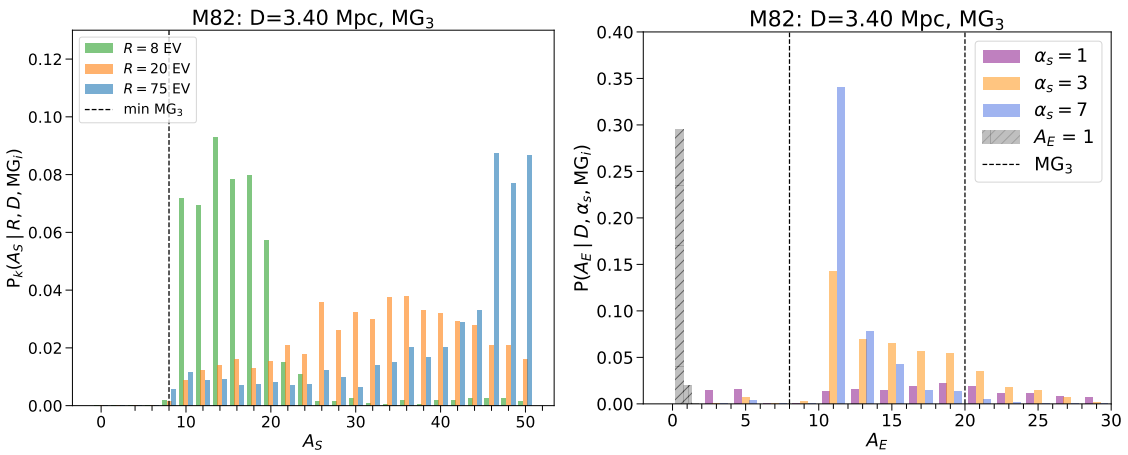

Figure 2: The left panel shows the composition PDF $\mathrm{P_k}(A_S \mid R, 3.4\,\mathrm{Mpc}, \mathrm{MG_3})$ of M82 for three rigidity values and $\mathrm{MG_3}$. At higher R, the mean mass at the source shifts to heavier elements to compensate for the more excessive mass loss. The right panel demonstrates the corresponding mass composition of the spectrum at Earth calculated using Equation 5 integrated over R. As desired, the arrival composition is well confined within the mass group. The contribution from protons are excluded as they do not contribute due to their overall suppression.

The true rigidities for a given mass group are evaluated by dividing the reported arrival energies E with the mean charge $\langle Z \rangle$ determined by the GSF model [10].

To appropriately normalize the likelihood, the number of expected events must be determined. This is performed similarly to CM19. However, we consider the energy losses and source spectrum under the assumptions of our new approach. In particular, the number of expected events can be determined by summing over all sources N_s as such

$$N_{\mathrm{ex}} = N_{\mathrm{b}} + \sum_{k=1}^{N_s} \frac{Q^k_{\mathrm{earth}}(L, \alpha_s, D_k, \mathrm{MG}_i)}{4\pi D_k^2} \epsilon_k(R^k_{\mathrm{ex}}, D_k, \bar{B}, \alpha_s), \tag{6}$$

where $N_b = F_0\,\alpha_T/4\pi$ is the number of expected events from the background, Q^k_{earth} is the number of expected UHECRs per unit time, and ϵ_k is the effective exposure factor for each source, calculated using CRPropa3 using the expected rigidity at earth R^k_{ex}.

The UHECR rate Q^k_{earth} and expected rigidity R^k_{ex} at Earth can be determined by evaluating the zeroth and first moment of the arrival spectrum, respectively

$$Q^k_{\mathrm{earth}} = \int_{R_{\mathrm{min}}}^{R_{\mathrm{max}}} dR\, \frac{dN^k_{\mathrm{arr}}}{dRdt}; \qquad R^k_{\mathrm{ex}} = \frac{1}{Q^k_{\mathrm{earth}}} \int_{R_{\mathrm{min}}}^{R_{\mathrm{max}}} dR\, R\, \frac{dN^k_{\mathrm{arr}}}{dRdt}. \tag{7}$$

Here we set the minimum rigidity at Earth to the detector-dependent threshold rigidity $R_{\mathrm{th}} = E_{\mathrm{th}}/\langle Z \rangle$, where E_{th} is the threshold energy, and set $R_{\mathrm{max}} = 100\,\mathrm{EV}$. Through Q^k_{earth}, we appropriately account for the source spectrum and energy loss effects for the expected number of events.

3. Statistical Model

As in CM19, we implement the physical model described above into a Bayesian hierarchical framework. Our improved model for UHECR nuclei allows us to extend the idea of our statistical

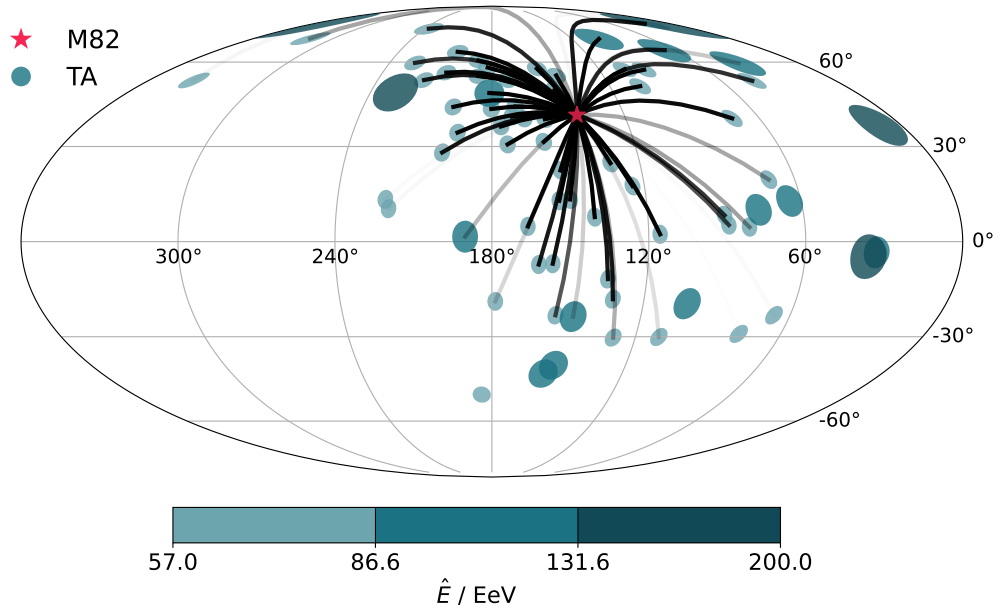

Figure 3: The association skymap of the starburst galaxy M82 with the public events with $E > 57$ EeV from TA [16], assuming a mass group of 2 (MG_2). The solid lines indicate source–UHECR association probabilities, with darker lines corresponding to stronger associations.

analysis fitting a "bubble" of possible origins for each UHECR event to non-proton UHECR compositions. Our likelihood function has the same form but is now implemented in terms of $\hat{R}$ and the corresponding rigidity spectrum $\mathrm{P}(R \mid \alpha_s, \alpha_b, D_k)$ instead of energy.

We choose weakly informative priors for the hyperparameters, or highest-level parameters, L, α_s, α_b, F_0, and $\bar{B}$, motivated by UHECR phenomenology and to avoid any unphysical values. In particular, we choose normal priors for the product of the EGMF strength $\bar{B}$ with the coherence length $l_c^{1/2}$, taking $\mu_{Bl_c^{1/2}} = 2\,\mathrm{nG\,Mpc}^{1/2}$, $\sigma_{Bl_c^{1/2}} = 5\,\mathrm{nG\,Mpc}^{1/2}$ and an upper limit of $10\,\mathrm{nG\,Mpc}^{1/2}$ based on current observational constraints [14, 15]. For other hyperparameters, we select appropriate upper and lower bounds and choose wide normal priors to reflect the lack of knowledge of such parameters. We have $\mathrm{P}(\alpha_s) \sim \mathcal{N}(5,3)$, $\mathrm{P}(\alpha_b) \sim \mathcal{N}(5,3)$, $\mathrm{P}(L) \sim \mathcal{N}(0,50)$ and $\mathrm{P}(F_0) \sim \mathcal{N}(0,10)$.

4. Results

While we leave a more detailed discussion for a forthcoming publication, we highlight here one of the results, as shown in Figure 3 and Figure 4. The starburst galaxy M82 is located at about 3.4 Mpc within TA's exposure in the northern sky. It has been proposed as a possible origin of the TA Hotspot [16, 17]. Using the full model, we test whether M82 can contribute a sufficient number of events, given the observed average composition within MG_2 by TA [18]. We find that M82 can contribute about 80% of the events in TA's sample used for anisotropy measurements [16]. The fitted spectral indices are compatible with those obtained in UHECR fits that use a uniform distribution of sources, such as Ref. [19], that prefer hard ($\alpha_S \sim 1$) spectral indices. In particular, all lower energy events within the hotspot range can be strongly associated with M82, yet this source cannot easily explain the most energetic events.

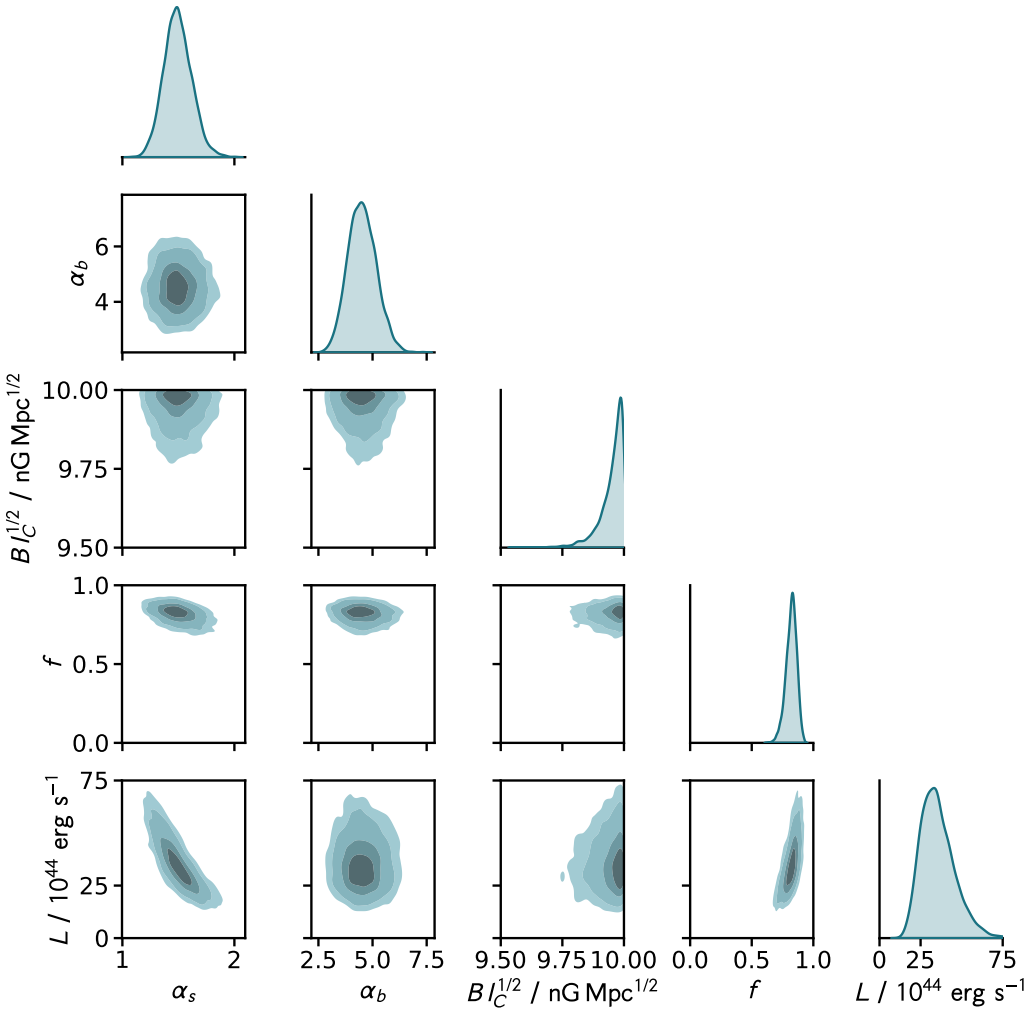

Figure 4: The joint marginalized posterior distributions for L, α_s, α_b, $\bar{B}\sqrt{l_c}$, and the source fraction f, using the configurations as in Figure 3 (M82 + TA + MG_2).

However, a high luminosity $L \gtrsim 10^{45}$ erg s^{-1} is required to explain the observed source fraction. Additionally, we see that the EGMF tends to larger values $\sim 10nG$, within the physically motivated prior that we consider [14, 15]. Furthermore, as described in [5], our model for the GMF deflections neglects the directionality of the lensing effects, implying a conservative assumption on the GMF modelling. As shown in Figure 1, the maximal rigidities are currently set to a fixed value of $R_{\rm max} = 1$ ZV. A more realistic model would allow to fit for $R_{\rm max}$ to explain the absence of very high energies spatially correlated with the M82's direction.

5. Summary

We implement a model for the propagation of UHECR nuclei that is simple and fast enough to be implemented in Stan for efficient Bayesian fits of the UHECR observations, but still captures the key underlying physics. Our approach relies on the approximation of constant rigidity during propagation of nuclei, and the calculation of the relevant source composition PDFs with CRPropa3.

We demonstrate the application of our UHECR propagation model within a hierarchical Bayesian framework to study M82 as a possible source of the TA hotspot and find that ~80% of the events can be explained for the assumption of an MG_2 composition upon arrival. In a forthcoming publication, we apply our method to explore the implications for different source catalogs and the publicly available data for the PAO, which bring further constraints.

Acknowledgements A.F. performed this work as a JSPS International Research Fellow (JSPS KAKENHI Grant Number 19F19750) and hosted K.W. during his research visit at ICRR. The authors acknowledge the computing resources provided by the Academia Sinica Grid Computing Center (ASGC), supported by the Institute of Physics at Academia Sinica.

References

[1] F. Capel and D.J. Mortlock, *Mon. Not. Roy. Astron. Soc.* **484** (2019) 2324 [1811.06464].

[2] L.J. Watson et al., *Mon. Not. Roy. Astron. Soc.* **418** (2011) 206 [1010.0911].

[3] K. Soiaporn et al., *Ann. Appl. Stat.* **7** (2013) 1249 [1206.4569].

[4] Stan Development Team, "Stan Modeling Language Users Guide and Reference Manual." Online, 2022.

[5] K. Watanabe et al., *EPJ Web Conf.* **283** (2023) 03009.

[6] D. Allard, *Astropart. Phys.* **39-40** (2012) 33 [1111.3290].

[7] D. Boncioli et al., *Sci. Rep.* **7** (2017) 4882 [1607.07989].

[8] R. Alves Batista et al., *JCAP* **09** (2022) 035 [2208.00107].

[9] Pierre Auger collaboration, *PoS* **ICRC2021** (2021) 311.

[10] H.P. Dembinski et al., *PoS* **ICRC2017** (2018) 533 [1711.11432].

[11] D. Harari et al., *Journal of High Energy Physics* (2002) 045 .

[12] R. Jansson and G.R. Farrar, *Astrophys. J.* **757** (2012) 14 [1204.3662].

[13] R. Fisher, *Proceedings of the Royal Society of London Series A* **217** (1953) 295.

[14] R. Durrer and A. Neronov, *Astron. Astrophys. Rev.* **21** (2013) 62 [1303.7121].

[15] J.D. Bray and A.M.M. Scaife, *Astrophys. J.* **861** (2018) 3 [1805.07995].

[16] Telescope Array collaboration, *Astrophys. J. Lett.* **790** (2014) L21 [1404.5890].

[17] H.-N. He et al., *Phys. Rev. D* **93** (2016) 043011 [1411.5273].

[18] Telescope Array collaboration, *PoS* **ICRC2021** (2021) 300.

[19] J. Heinze et al., *Astrophys. J.* **873** (2019) 88 [1901.03338].