PROCEEDINGS OF SCIENCE

Approaching UHECR astronomy using mass-sensitive data from the Pierre Auger Observatory and the Telescope Array Project

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

[a] *Institute for Astroparticle Physics, Karlsruhe Institute of Technology (KIT), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany*

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

[c] *Max Planck Institute for Physics, Boltzmannstr. 8, 85748 Garching, Germany*

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

E-mail: keito.watanabe@kit.edu

The field of ultra-high-energy cosmic ray (UHECR) astronomy has been facing an ongoing challenge due to the unknown impact of magnetic deflections on the observed events. However, with the dawn of the era of mass-sensitive data, we would have the available information to perform UHECR astronomy. To do so, constructing a sophisticated analysis that can accurately reconstruct source parameters will be crucial. In this work, we construct a Bayesian hierarchical framework that utilises the spatial, energy, and mass composition information to infer the source properties such as luminosity or the spectral index as well as the strength of the extragalactic magnetic field while marginalising the uncertainties of the detector. We model the mass-dependent spatial deflections from the Galactic magnetic field (GMF) on an event-by-event basis using the latest available GMF model [1] and developed a novel method to determine the nuclear composition of each UHECR source by applying the propagation solver PriNCe [2] that incorporates the energy losses due to photo-nuclear interactions. We apply this new method on realistic simulated datasets with nuclear composition information observed in both Northern and Southern skies to demonstrate the method's capabilities of the Bayesian inference of source parameters. It is shown that leveraging the full three-dimensional mass-sensitive data while incorporating the most accurate physical models of UHECR acceleration and propagation allows a more unbiased reconstruction of source properties and extragalactic magnetic field strengths.

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

*Speaker

 ISSN 1824-8039 . Published by SISSA Medialab.

PoS(UHECR2024)017

1. Introduction

The origin of ultra-high-energy cosmic rays (UHECRs) continues to be a challenging question due to their mass composition. They primarily undergo rigidity-dependent deflections from the Galactic and extragalactic magnetic fields (GMF and EGMF respectively) and experience energy losses that transform their composition through propagation. While we currently lack individual mass-sensitive measurements of these UHECRs, with the recent achievements in utilising deep neural networks, such datasets will be available within the next few years. Furthermore, the multi-hybrid measurements from the AugerPrime upgrade and the increased exposure from the TAx4 experiment will be crucial in improving such datasets [3, 4].

To most effectively utilise the upcoming mass-sensitive data, developing a sophisticated analysis framework that encompasses our current understanding of UHECR propagation for source associations is crucial. In this work, we report on the progress on the development of a Bayesian hierarchical framework that not only performs source-UHECR associations but also infers source parameters such as the spectral index, luminosity, and EGMF strengths from rigidity-based data. Our approach is an extension on the framework from Capel & Mortlock 2018 [5], which also extends the works developed by Watson et al. 2010 and Soiaporn et al. 2013 [6, 7]. In our previous reports, we showcased our efforts to extend this model by including rigidity-based deflections from the GMF and EGMF [8] (W23a hereafter) and mass-dependent energy loss processes based on the current theory of UHECR propagation [9] (W23b hereafter). Here we showcase further improvements to our model and apply it to a realistic scenario based on simulated datasets from the Pierre Auger Observatory and the Telescope Array Project (PAO and TA hereafter). We verify our model by showcasing the reconstruction uncertainty for the Northern and Southern skies using such datasets, which highlight the importance of accurate mass reconstruction for source association analyses.

2. Model

The model that we utilise in this analysis framework is an improvement from our previous works [8, 9]. As before, we construct our likelihood assuming an approximate conservation of rigidity ($R_S \approx E/Z \approx R_E$) throughout propagation as the boost is approximately conserved for UHECR nuclei [10].

Propagation Loss Model: The energy loss model in our framework follows similarly from W23b, however, we instead utilise `PRiNCe` [2], a 1-D numerical propagation code that models UHECR transport from a uniform distribution of sources, to generate the propagation database that models the energy loss process for each rigidity R, source composition A_S, and source distance D. We assume a flat spectrum for all energies per nucleon E/A, given that the source emits only a single composition at each distance D. We then utilise `PRiNCe` to numerically compute the resulting energy spectrum for each arrival composition A_E at Earth. As the charge-to-mass ratio is approximately conserved, the derived spectrum can be transformed to describe the rigidity spectrum at Earth. Performing this over all source composition $A_S \in [2, 56]$ and distances $D \in [1, 120]$ Mpc and summing over arrival compositions that arrive within a particular mass group (labelled as MG) [11], we can generate a propagation matrix $\mathcal{M}^{\mathrm{MG}}_{S\to E}$ that encompasses the fraction of arrived particles

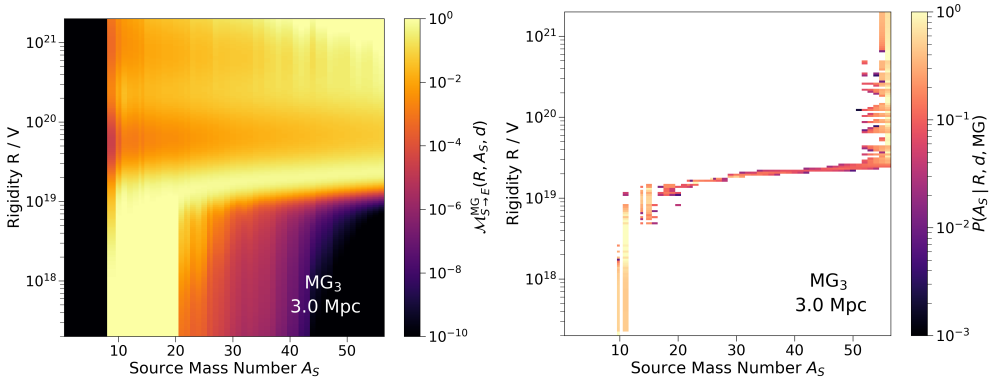

Figure 1: Left: Propagation matrix $\mathcal{M}^{\mathrm{MG}}_{S\to E}(R, A_S, D)$ for mass group 3 for a source placed at a distance of $D = 3$ Mpc that describes the fraction of arrived particles at each rigidity and source mass. Right: The source mass PDF $P(A_S \mid R, D; \mathrm{MG})$ describing the distribution of source masses that maximally produce UHECRs at each rigidity within MG$_3$ at a distance of 3 Mpc. The white region indicates values of (R, A_S) that are not efficient enough to produce UHECRs at MG$_3$.

within a particular mass group from each source with composition A_S and distance D:

$$\mathcal{M}^{\mathrm{MG}}_{S\to E}(R, A_S, D) = \sum_{A_E \in \mathrm{MG}} \frac{\Delta N_E(R, A_S, A_E, D)}{\Delta N_S(R, A_S, D)}. \tag{1}$$

The left panel of Figure 1 shows the propagation matrix for MG$_3$ (i.e. with arrival compositions $A_E \in [10, 20]$) at a source distance of 3 Mpc. From this matrix, we observe that at low rigidities, source masses within the mass group retain their identity at Earth, however, heavier source masses reach the given mass group with decreasing probability with mass as they experience stronger photo-pion interactions. There is also a sharp cutoff at around 3×10^{19} eV for all source masses in which the energy loss dominates, where the majority of the produced secondaries lie in MG$_1$ or MG$_2$ at Earth.

Using these propagation matrices, we then construct a cost function that minimises the fraction of arrived particles outside of the desired mass group, and maximises those that arrive within the mass group:

$$f(\mathcal{M}_{\mathrm{inj,MG}}) = \log_{10}\left[\sum_{A_E}\left(\frac{\Delta N_E(A_E \notin \mathrm{MG})\,\mathcal{M}_{\mathrm{inj,MG}}}{\Delta N_E(A_E \in \mathrm{MG})}\right)\right] \tag{2}$$

where we omit the dependence of R, A_S, and D here. Minimising this function allows us to derive the injection matrix $\mathcal{M}_{\mathrm{inj,MG}}$ that encodes the weights to ensure maximal production of UHECRs for this mass group. The source mass PDF $P(A_S \mid R, D; \mathrm{MG})$, defined by normalising $\mathcal{M}_{\mathrm{inj,MG}}$ over all source masses, now describes the distribution of source masses that can maximally produce UHECRs within the given mass group. The right panel of Figure 1 shows the resulting source mass PDF for a source at $D = 3$ Mpc. We observe that at low rigidities, only UHECRs with the same mass group at the source can be injected, but with higher rigidities the effects of propagation losses require the injection of elements with higher mass.

To generate a corresponding rigidity spectrum that we can sample at Earth, we multiply the source mass PDF to a power-law (defined by spectral index α_s) with an exponential cutoff (with

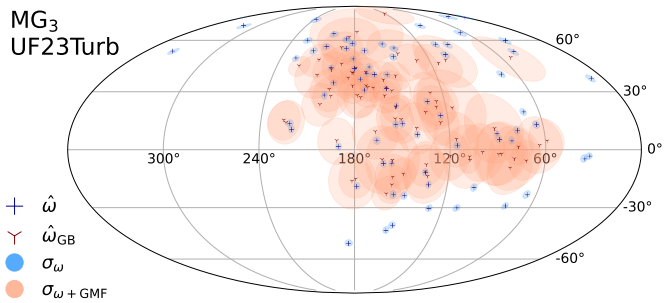

Figure 2: Skymap of mean deflected events (in red crosses) for the event sample from TA2014 (in blue crosses) in Galactic coordinates. The mean deflected events are generated using `CRPropa3`, assuming all arrival compositions lie within MG_3 and using the UF23Turb Galactic magnetic field model. The blue and red circles indicate the angular reconstruction uncertainty of the directions σ_ω and the angular scale accounting for GMF deflections $\sigma_{\omega+\mathrm{GMF}}$ respectively.

$R_{\rm cutoff} = 50\,\mathrm{EV}$) to define a rigidity spectrum at the source. Convolving this with the derived propagation matrix $\mathcal{M}^{\rm MG}_{S\to E}(R, A_S, D)$, we derive the rigidity spectrum at Earth which not only encompasses propagation losses but also takes into account the injection efficiency of UHECRs, under the assumption that all masses land within a certain mass group (see further details in W23b):

$$\frac{\mathrm{d}N_E}{\mathrm{d}R}(R, \alpha_s, D; \mathrm{MG}) \propto R^{-\alpha_s} \exp(-R/R_{\rm cutoff}) \sum_{A_S} \mathcal{M}^{\rm MG}_{S\to E}(R, A_S, D)\, Z_S^{1-\alpha_s}\, P(A_S\,|R, D, \mathrm{MG}) \quad (3)$$

The arrival spectrum is normalised to the luminosity L of each source, which we leave as a free parameter within our framework.

Deflection Model: The extragalactic deflections are modelled as with W23a and W23b where we assume a fully-turbulent field with a coherence length $D \gg l_c$, allowing us to describe the root-mean-square angular scale $\sigma_{\rm EGMF}$ as a function of the rigidity R, source distance D, and root-mean-square extragalactic magnetic field strength $B_{\rm EGMF}$. The deflection model utilises a von Mises-Fischer distribution [12] centered at the source direction $\hat{\omega}_S$ where the angular scale is converted to a deflection parameter $\kappa_{\rm EGMF} \approx 7552\,(\sigma_{\rm EGMF}/1\,\mathrm{deg})^{-2}$ [7].

The GMF deflections are modelled by backtracking events that are sampled around the true arrival direction $\hat{\omega}$ and energy $\hat{E}$ using the reported angular and energy reconstruction uncertainty. We convert the energy to a rigidity scale using the mean charge within the given mass group $\langle Z\rangle_{\rm MG}$. The events are then sampled using a von Mises-Fischer and Gaussian distribution respectively before backtracking them using `CRPropa3` through the GMF. The GMF is modelled through the averaged field from the 8 coherent models developed by Unger & Farrar 2023 [1] combined with the Planck re-tune of the random turbulent field with coherence length $l_c^{\rm GMF} = 60\,\mathrm{pc}$ (UF23Turb hereafter). We record the mean deflected direction at the Galactic boundary (GB) $\hat{\omega}_{\rm GB}$ and fit the deflection parameter $\kappa_{\rm GMF}$ using the deflected samples with `stan` [13], a Python-based package for high-dimensional statistical computation. Performing this for each mass group and dataset, we obtain a new dataset at the Galactic boundary which we use for our analysis framework.

Figure 2 shows a skymap of the deflected events from the public dataset from TA ([14], TA2014 hereafter) with an energy threshold of $E_{\rm th} = 57\,\mathrm{EeV}$. The blue and red circles indicate the angular reconstruction efficiency σ_ω and the angular scale for GMF deflections $\sigma_{\omega+\mathrm{GMF}}$ that corresponds

to "1 σ" in a normal distribution respectively. The events in this sample are all assumed to lie within MG_3, corresponding to a mean charge of $\langle Z \rangle_{\mathrm{MG}} = 7$. We observe that the majority of events are deflected towards the Galactic anti-center, with deflections as strong as $\sim 60\,\mathrm{deg}$ observed from several events. The deflection angular scale takes into account the strength of the deflection experienced by each event. Furthermore, as the backtracked events are constrained by the detector exposure, the (de-)magnification of events due to the GMF is implicitly taken into account.

Background Model: To differentiate the UHECRs originating from sources, we construct a separate model for events that cannot be associated with any sources within our framework. Here, all background events are assumed to originate at $D = 0\,\mathrm{kpc}$, and are described by a power law, i.e. $\mathrm{d}N_b/\mathrm{d}R \propto R^{-\alpha_b}$, where we separately define a spectral index α_b for background events. The events are normalised by a background flux F_0 and the directions are distributed uniformly, i.e. with $\kappa_{\mathrm{EGMF}} = 0$.

Statistical Model: Our likelihood function, which combines the deflection and propagation model, now takes the form $P(R, \hat{\omega}_{\mathrm{GB}} \mid \alpha_s, L, B_{\mathrm{EGMF}} l_c^{1/2}, \alpha_b, F_0)$, i.e. it is a joint distribution of the rigidity and the arrival direction at the Galactic boundary. The normalisation of the likelihood is performed similarly to W23b, however now encoding the GMF deflections within the calculation of the effective exposure. As in our previous works, we then use this likelihood into a Bayesian hierarchical framework. We choose weakly informative priors for our hyperparameters, motivated by constraints from UHECR phenomenology. Contrary to previous works, we choose to set a pessimistic upper limit of $1\,\mathrm{nG\,Mpc}^{1/2}$ for $B_{\mathrm{EGMF}} l_c^{1/2}$ as we do not include the spatial variation of the EGMF strength, which can vary from $\ll 1\mathrm{nG}$ to $> 100\,\mathrm{nG}$ strengths [15]. The joint posterior distribution for the hyperparameters are solved using `stan` [13].

3. Results

To showcase and verify our results, we apply our model to realistic simulated datasets based on TA2014 [14] and the Phase One dataset from PAO [16] (PAO2022 hereafter), assuming that the UHECRs are produced via the source M82 and Centaurus A with a source fraction of $f = 0.1$ and $f = 0.01$ respectively[1]. These sources were chosen as they are prime source candidates in the Northern and Southern skies respectively [17, 18]. We assume that both arrival compositions lie within MG_3 and assume the UF23Turb model for the Galactic magnetic field.

To generate the simulated dataset, we first sample the rigidity R and arrival directions ω_{GB} at the Galactic boundary from Equation (3) and the von Mises-Fischer distribution for EGMF deflections respectively, assuming that all events are produced from the given source (D, ω_S) and lie within a given mass group MG. These samples are passed through a lensing matrix (for a given GMF model) encoding the GMF deflections experienced from the Galactic boundary to Earth, and we then apply cuts to the deflected events using the detector exposure and reported detector uncertainty to generate the rigidity $\hat{R}$ and arrival direction $\hat{\omega}_E$ at Earth. Finally, we backtrack the events through the GMF using our GMF deflection model as described above to generate $\hat{\omega}_{\mathrm{GB}}$ and κ_{GMF} at the Galactic boundary.

[1] The source fraction is defined as the flux ratio of observed UHECRs from all sources within the catalogue compared to the total observed flux.

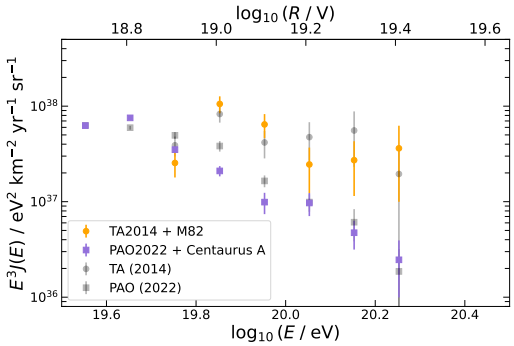

Figure 3: Energy spectrum of the realisation of the simulated dataset for TA2014 + M82 (orange circles) and PAO2022 + Centaurus A (purple squares) used in this work. We assume an arrival composition within MG_3 and histogram samples through equal logarithmically spaced energy bins. We also overlay the observed data from TA2014 [14] (black circles) and PAO2022 [16] (black squares), which are histogrammed similarly. The corresponding rigidity $R = E/\langle Z \rangle_{MG}$ is shown in the top axis.

Figure 3 shows the energy spectrum for the particular realisation for the two simulated datasets used in this work. The simulated datasets resemble that of the true event sample (overlaid in the figure), which further motivates our choice of the source fraction above.

In Figure 4, we show the source association skymap of the simulated dataset at Earth and at the Galactic boundary. The individual source-UHECR association is indicated by the intensity of the black lines. We observe that at the Galactic boundary, the associations are clustered near their respective source. However, the associations appear non-trivial at Earth due to the deflected events. This highlights the importance of GMF deflections towards correlation studies utilising arrival directions. Figure 5 shows the joint posterior distribution for the inferred parameters α_s, $B_{\rm EGMF} l_C^{1/2}$, and L using the same datasets. We show that we can accurately reconstruct the true values for both Northern and Southern skies as the true values lie within 20% of the posterior maximum.

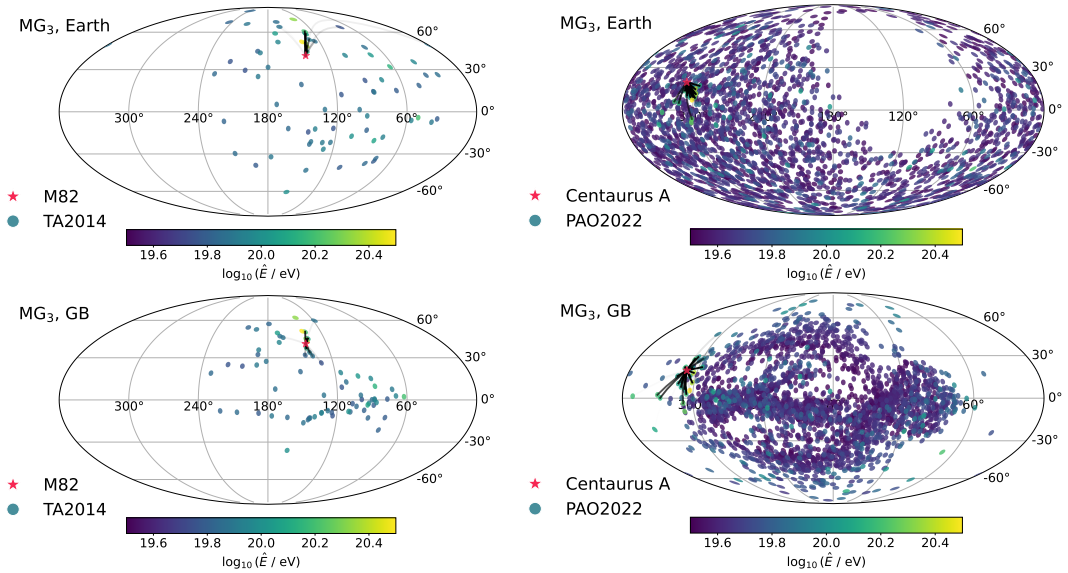

Figure 4: The association skymaps of the simulated dataset following that of TA2014 + M82 (left) and PAO2022 + Centaurus A (right) at Earth (top) and at the Galactic boundary (bottom). We assume that they lie within MG_3 and are deflected using the UF23Turb GMF model. The colorbar indicates the energy of each simulated UHECR, and the individual association probability between the sources at each UHECR is indicated by the solid lines, with darker lines indicating a stronger association.

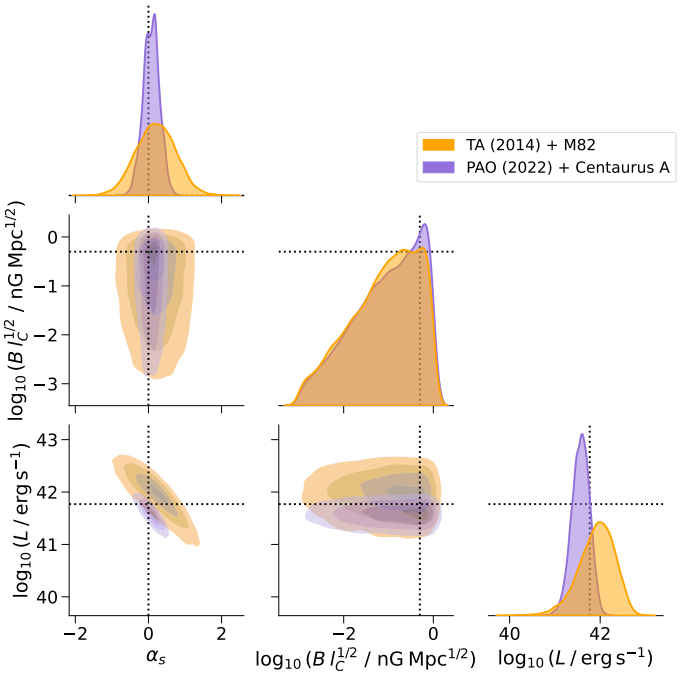

Figure 5: The joint posterior distribution for the source parameters (α_s, $B_{\rm EGMF}l_C^{1/2}$, and L) using the simulated datasets used in this work. The dashed lines indicate the true values used to generate both simulated datasets. The contours indicate values that are 10%, 30%, 70%, and 90% away from the posterior maximum.

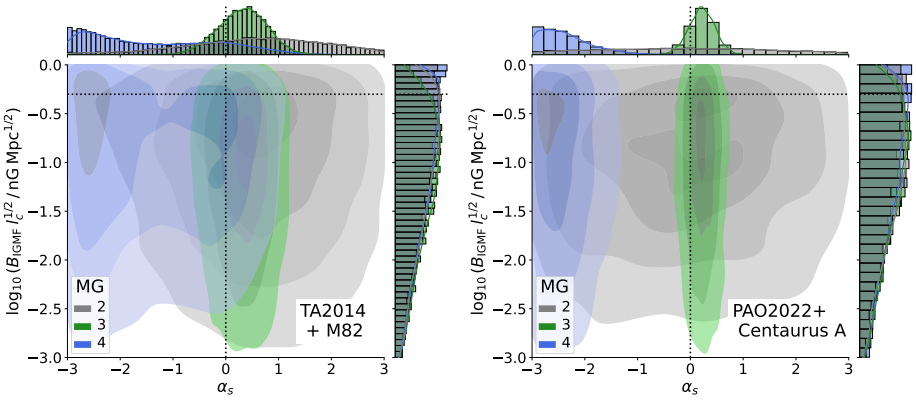

Figure 6: The joint posterior distribution for α_s and $B_{\rm EGMF}l_c^{1/2}$ for TA2014 + M82 (left) and PAO2022 + Centaurus A (right). We use the same simulated dataset in this work (i.e. assuming $\rm MG_3$), but assuming mass groups of 2 (gray), 3 (green), and 4 (blue) when performing the inference. The dashed lines indicate the true values used to generate the simulated dataset.

To verify the accuracy of the mass reconstruction, we also performed a "mass misidentification" study where we inferred the parameters by assuming mass groups $\rm MG_2$ - $\rm MG_4$ for reconstruction using the simulated datasets used in this work (i.e. simulating with $\rm MG_3$ for both datasets). Figure 6 shows the joint posterior density for α_s and $B_{\rm EGMF}l_c^{1/2}$ for both datasets, which shows that we can most accurately reconstruct the true parameters only by assuming the correct mass group when

performing the inference.

4. Conclusion

In this work, we show the progress on the development of a Bayesian analysis framework for the inference of source parameters and associations that encompasses our latest physical understanding of UHECR propagation. The energy and mass are described through a rigidity spectrum that includes the injection efficiency and propagation loss for an assumed arrival composition based on mass groups. We include rigidity-dependent deflections through a von-Mises Fisher distribution and utilise deflected events at the Galactic boundary as our event sample. We tested our framework on realistic simulated datasets for both Northern and Southern skies, showing an accurate reconstruction of the truths that lie within 20% of the posterior maxima. We also show that an accurate mass measurement is crucial in accurately inferring source parameters. We hope to perform further validation tests such as comparing the reconstruction efficiency with different GMF models.

References

[1] M. Unger and G.R. Farrar, *Astrophys. J.* **970** (2024) 95 [2311.12120].

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

[3] Telescope Array collaboration, *Nucl. Instrum. Meth. A* **1019** (2021) 165726 [2103.01086].

[4] Pierre Auger collaboration, *EPJ Web Conf.* **210** (2019) 06002 [1905.04472].

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

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

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

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

[9] K. Watanabe et al., *PoS* **ICRC2023** (2023) 479.

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

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

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

[13] S.D. Team, *Stan modeling language users guide and reference manual*, 2024.

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

[15] F. Vazza et al., *Mon. Not. Roy. Astron. Soc.* **439** (2014) 2662 [1401.4454].

[16] Pierre Auger collaboration, *Astrophys. J.* **935** (2022) 170 [2206.13492].

[17] H. Yüksel et al., *Astrophys. J.* **758** (2012) 16 [1203.3197].

[18] Telescope Array collaboration, *Astrophys. J. Lett.* **867** (2018) L27 [1809.01573].