

A Bayesian Framework for UHECR Source Association and Parameter Inference

Keito Watanabe,^{a,*} Anatoli Fedynitch,^{b,d} Francesca Capel^c and Hiroyuki Sagawa^d

^a*Institute for Astroparticle Physics, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany*

^b*Institute of Physics, Academia Sinica, No. 128, Sec. 2, Academia Rd, Nangang District, 11529 Taipei, Taiwan*

^c*Max Planck Institute for Physics, Boltzmannstr. 8, Garching 85748, 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, anatoli@gate.sinica.edu.tw, capel@mpp.mpg.de, hsagawa@icrr.u-tokyo.ac.jp

The identification of potential sources of ultra-high-energy cosmic rays (UHECRs) remains challenging due to magnetic deflections and propagation losses, which are particularly strong for nuclei. In previous iterations of this work, we proposed an approach for UHECR astronomy based on Bayesian inference through explicit modelling of propagation and magnetic deflection effects. The event-by-event mass information is expected to provide tighter constraints on these parameters and to help identify unknown sources. However, the measurements of the average mass through observations from the surface detectors at the Pierre Auger Observatory already indicate that the UHECR masses are well represented through its statistical average. In this contribution, we present our framework which uses energy and mass moments of $\ln A$ to infer the source parameters of UHECRs, including the mass composition at the source. We demonstrate the performance of our model using simulated datasets based on the Pierre Auger Observatory and Telescope Array Project. Our model can be readily applied to currently available data, and we discuss the implications of our results for UHECR source identification.

39th International Cosmic Ray Conference (ICRC2025)
15–24 July 2025
Geneva, Switzerland



*Speaker

1. Introduction

With more than 20 years of observations, the Pierre Auger Observatory (Auger) and Telescope Array Project (TA) have provided us with a wealth of data on ultra-high energy cosmic rays (UHECRs), and will provide us with even more data with the AugerPrime upgrade [1] and TA_{x4} experiment [2]. However, the association of UHECRs with their potential sources remains a challenge due to the limited statistics and the complex propagation effects in the intergalactic and Galactic medium. In particular, the mass composition of UHECRs has not been measured on an event-by-event basis, but rather only as a statistical average of all observed UHECRs. This limits the constraints placed on the rigidity-dependent deflections and mass-dependent propagation losses of each UHECR. Nevertheless, with sophisticated analyses techniques, the combination of currently available arrival directions, energies, and mass composition data can be used to constrain the properties of potential sources and the propagation effects of UHECRs [3].

In previous works, we developed a Bayesian hierarchical model capable of inferring ultra-high-energy cosmic ray (UHECR) source parameters, such as luminosity and spectral indices, while accounting for mass-dependent propagation effects using the nuclear propagation code `PrINce` [4–7]. However, earlier iterations of this model required assumptions about the arrival composition falling within a predefined mass group, limiting its direct applicability to observational data. In this work, we relax this assumption by explicitly incorporating the mean and variance of the logarithmic mass ($\ln A$) distribution measured at Earth. We demonstrate that combining these mass composition moments with individual event energies is sufficient to robustly infer source parameters, including the initial UHECR mass composition at the sources. We present detailed analyses of simulated datasets reflecting the exposures of the Pierre Auger Observatory and the Telescope Array, highlighting the method’s performance and applicability to currently available observational data.

2. Model

As with our previous work, the model used in this work is based on a Bayesian hierarchical framework, where we set priors on the source parameters, such as the luminosity and spectral indices, as well as the mass fractions of UHECRs at the source. The inference is performed using the probabilistic programming language `stan` [8], which performs Markov Chain Monte Carlo (MCMC) sampling through the Hamiltonian Monte Carlo (HMC) algorithm to directly obtain the posterior distributions of the model parameters.

2.1 Modelling UHECR Propagation with `PrINce`

To describe the mass-dependent propagation effects, we use the nuclear propagation code `PrINce` [7]. This code describes its nuclear propagation as a series of partial differential equations and numerically solves for the co-moving density of UHECRs $Y_i(E_i, z) = N_i(E_i, z)/(1+z)^3$ for each mass component i at a given energy E_i and redshift z . The propagation equation is given by:

$$\partial_z Y_i(E_i, z) = -\partial_E(b_{\text{ad}} Y_i) - \partial_E(b_{e^+e^-} Y_i) - \Gamma_i Y_i + \sum_j Q_{i \rightarrow j} Y_j + J_i, \quad (1)$$

which describes (in order from left to right) the adiabatic losses, pair production losses and photo-nuclear interactions (decay and re-injection). The last term $J_i = J_i(E_i, z, A_i^S)$ describes the injection from a distribution of homogeneously distributed and isotropically emitting sources, where A_i^S denotes the mass number of component i at the source. In this work, we describe the photo-nuclear and photo-hadronic cross sections with TALYS [9] and SOPHIA [10] respectively.

Source model: While the same propagation model is used to describe the UHECRs emitted from the source and from background sources, we model them separately. Given N_S sources, each source (labelled as k) with a given luminosity L_k at distance d_k (flux $F_k \propto L_k/4\pi d_k^2$) is treated as a point source that emits UHECRs with a single injection mass composition A_l^S isotropically without any background sources. This amounts to setting $J_i = 0$ in Eq. (1) and setting the initial condition of the co-moving density Y_i at $z_k = z(d_k)$ with the energy distribution of UHECRs for a single injected mass at the source (i.e. the source spectrum). The source spectrum follows that of [11]:

$$\mathcal{J}_{lk}^S(E_i, Z_l^S | \alpha_k, R_{\max}) \propto \left(\frac{E_i}{1 \text{ EeV}} \right)^{-\alpha_k} \times \begin{cases} 1 & : E_i < Z_l^S R_{\max} \\ \exp \left(1 - \frac{E_i}{Z_l^S R_{\max}} \right) & : E_i \geq Z_l^S R_{\max} \end{cases} \quad (2)$$

where Z_l^S denotes the charge number corresponding to the injected mass A_l^S . Each source is characterized by a spectral index α_k and a maximal rigidity R_{\max} , which is assumed to be identical for all injected elements. Although previous studies have indicated that maximal rigidity may vary between sources due to differences in their acceleration properties, we currently fix $R_{\max} = 1.7 \text{ EeV}$ for all sources following earlier analyses [7, 12]. The normalization constant, representing the number of injected particles with mass A_l^S per comoving volume, per unit energy, and per year, is determined within the forward model implemented in `stan`.

Background Model: Our background model assumes the continuous injection of UHECRs from sources homogeneously distributed from redshift $z_{\max} = 3$ down to a truncation redshift z_{trunc} . This is modeled by setting initial conditions $Y_i(E_i, z) = 0$ and defining the injection function J_{il} as:

$$J_{il}(E_i, z, Z_l^S | \alpha_0, R_{\max}) = \begin{cases} n_{\text{evol}}(z) \mathcal{J}_{lk}^S(E_i, Z_l^S | \alpha_0, R_{\max}), & z \geq z_{\text{trunc}} \\ 0, & z < z_{\text{trunc}}. \end{cases} \quad (3)$$

Analogous to the source model, the background model is described by a spectral index α_0 , a fixed maximal rigidity $R_{\max} = 1.7 \text{ EeV}$, mass fractions f_{l0}^A for each injected composition, and a total normalization flux F_0 . We assume that the source density $n_{\text{evol}}(z)$ evolves with redshift according to the star formation rate (SFR) parameterization from Ref. [13].

Using Eqs. (2) to (3), we generate a database of solutions by solving Eq. (1) on a grid of spectral indices and injection masses. In our analysis, we consider $N_{\text{inj}}^A = 5$ injected mass groups: proton ($A = 1$), helium ($A = 4$), nitrogen ($A = 14$), silicon ($A = 28$), and iron ($A = 56$). The resulting comoving density at Earth, $Y_{ilk}^{\text{Earth}} := Y_i(E_i, z = 0, Z_l^S | \alpha_k)$, is used to calculate the flux at Earth $\mathcal{J}_{lk}^{\text{Earth}}$, as well as the spectrum-weighted mean and variance of $\ln A$. These quantities are obtained by summing over all mass species at Earth:

$$\mathcal{J}_{lk}^{\text{Earth}} = \sum_i Y_{ilk}^{\text{Earth}}, \quad \langle \ln A \rangle_{lk} = \frac{\sum_i Y_{ilk}^{\text{Earth}} \ln A_i}{\sum_i Y_{ilk}^{\text{Earth}}}, \quad \text{Var}(\ln A)_{lk} = \frac{\sum_i Y_{ilk}^{\text{Earth}} (\ln A_i)^2}{\sum_i Y_{ilk}^{\text{Earth}}} - \langle \ln A \rangle_{lk}^2. \quad (4)$$

This precomputed database is then used in the implementation of the forward model in `stan` to perform Bayesian inference.

2.2 Forward Modeling & Inference with `stan`

Energy & Mass Model: The total UHECR energy spectrum at Earth is obtained by summing contributions from each injected mass species A_l^S and all sources, including the diffuse background:

$$\mathcal{J}(E) = \sum_{k=0}^{N_S} \sum_{l=1}^{N_{\text{inj}}^A} f_{lk}^A F_k \mathcal{J}_{lk}^{\text{Earth}}(E) = \sum_{k=0}^{N_S} \sum_{l=1}^{N_{\text{inj}}^A} f_{lk}^A f_k^F F_{\text{tot}} \mathcal{J}_{lk}^{\text{Earth}}(E), \quad (5)$$

where the expected number of detected events N_k^{ex} for each source is determined by weighting the detector exposure ϵ_T with the integrated energy spectrum. Here, the spectrum at Earth is weighted by the source- and composition-dependent mass fractions f_{lk}^A and the source-dependent flux fractions f_k^F , multiplied by the total flux F_{tot} . These weighted spectra are used to sample individual UHECR energies E_{Earth} in our inference procedure.

The mean and variance of $\ln A$ at Earth are similarly computed by summing over the injected mass species contributions, now weighting by the expected number fraction of events from each source, $N_k^{\text{ex}}/N_{\text{tot}}^{\text{ex}}$, instead of the flux:

$$\langle \ln A \rangle(E_{\text{bin}}) = \sum_{k=0}^{N_S} \sum_{l=1}^{N_{\text{inj}}^A} f_{lk}^A \frac{N_k^{\text{ex}}}{N_{\text{tot}}^{\text{ex}}} \langle \ln A \rangle_{lk}, \quad \text{Var}(\ln A)(E_{\text{bin}}) = \sum_{k=0}^{N_S} \sum_{l=1}^{N_{\text{inj}}^A} f_{lk}^A \frac{N_k^{\text{ex}}}{N_{\text{tot}}^{\text{ex}}} \text{Var}(\ln A)_{lk}. \quad (6)$$

Since event-by-event composition information is currently unavailable, we adopt a binned approach, sampling the mean and variance of $\ln A$ within discrete energy bins E_{bin} separate from the individually sampled energies in Eq. (5). Thus, mass composition measurements act as global external constraints within our Bayesian inference framework.

Detector Model: UHECR measurements at observatories like Auger or TA are influenced by both statistical fluctuations and systematic uncertainties arising from calibration effects and hadronic model dependencies in the reconstruction process. To account for these effects, we model the detector response for each sampled energy E , and for each bin in $\langle \ln A \rangle$ and $\text{Var}(\ln A)$, as truncated Gaussian distributions centered around the true values, including global systematic shifts:

$$E_{\text{det}} \sim \text{TruncatedLogNormal}(\log E + \nu_{\log E}, \sigma_{\log E}, E_{\text{th}}, E_{\text{max}}), \quad (7)$$

$$\langle \ln A \rangle_{\text{det}}(E_{\text{bin}}) \sim \text{TruncatedNormal}(\langle \ln A \rangle(E_{\text{bin}}) + \nu_{\mu \ln A}, \sigma_{\mu \ln A}(E_{\text{bin}}), 0, \infty), \quad (8)$$

$$\text{Var}(\ln A)_{\text{det}}(E_{\text{bin}}) \sim \text{TruncatedNormal}(\text{Var}(\ln A)(E_{\text{bin}}) + \nu_{\text{Var} \ln A}, \sigma_{\text{Var} \ln A}(E_{\text{bin}}), -1, \infty), \quad (9)$$

where $\nu_{\log E}$, $\nu_{\mu \ln A}$, and $\nu_{\text{Var} \ln A}$ represent systematic uncertainties in the reconstructed energy, mean $\ln A$, and variance of $\ln A$, respectively. The corresponding statistical uncertainties, denoted by $\sigma_{\log E}$, $\sigma_{\mu \ln A}$, and $\sigma_{\text{Var} \ln A}$, are allowed to vary within each energy bin. The truncation ensures

physically meaningful results. Specifically, for the variance, a lower truncation bound of -1 is chosen to account for systematic shifts potentially leading to negative observed variance values.

Prior Model: We specify weakly informative priors for the model parameters as follows:

$$\alpha_k \sim \mathcal{N}(-1, 3), \quad f_k^A \sim \text{Dir}(\text{Vec}(1, N_{\text{inj}}^A)), \quad f^F \sim \text{Dir}(\text{Vec}(1, N_S+1)), \quad \log_{10} F_{\text{tot}} \sim \mathcal{N}(-1, 3), \quad (10)$$

where Dirichlet priors ensure that both the mass fractions f_k^A and flux fractions f^F sum to unity.

3. Results

To verify our model, we apply it to simulated datasets based on the detector configurations of Auger and TA. We assume a single point source (now labeled SRC) located at $d = 4$ Mpc, motivated by potential source candidates (Centaurus A, M82) for both hemispheres at similar distances [14, 15]. The continuous injection from background sources (labeled BG) is also truncated at the same distance. We set identical source parameters for both detectors, specifically choosing: $\alpha_{\text{SRC}} = -0.5$, $\alpha_{\text{BG}} = 0.5$, $L_{\text{SRC}} = 7 \times 10^{41} \text{ erg s}^{-1}$, $f_{\text{SRC}}^A = (0, 0.35, 0.5, 0.1, 0.05)$, and $f_{\text{BG}}^A = (0, 0.2, 0.4, 0.3, 0.1)$. The source association fraction is set to $f_{\text{assos}} = \sum_{k=1}^{N_S} N_k^{\text{ex}} / N_{\text{tot}}^{\text{ex}} = 0.1$, defining the relative contribution of the point source(s) to the overall observed data. Simulated datasets are generated using Eq. (5) to sample energies E_{Earth} for each UHECR, and Eq. (6) to sample the mean and variance of $\ln A$, utilizing energy bins from [16]. Detector responses are then applied to emulate observed energies and mass compositions, which we subsequently use in our inference model to reconstruct source parameters.

Detector	N_{ev}	$\epsilon_T / \text{km}^2 \text{ yr sr}$	$\sigma_{\log E}$	$\sigma_{\mu \ln A}$	$\sigma_{\text{Var} \ln A}$	$\nu_{\log E}$	$\nu_{\mu \ln A}$	$\nu_{\text{Var} \ln A}$
TA	680	30500	0.2	0.2	0.2	-0.05	0.3	0.5
Auger	2750	122000	0.1	0.1	0.1	0.05	0.3	0.5

Table 1: Detector parameters used in this work. Statistical uncertainties reflect differences in detector exposure ϵ_T , as indicated by the number of events N_{ev} . Systematic uncertainties are motivated by the energy rescaling from the joint Auger-TA energy spectrum working group and upper limits from current $\ln A$ measurements.

Table 1 summarizes the statistical and systematic parameters for each detector model. To ensure comparability, we use the detector exposure from the Auger Phase One dataset [17], and take 1/4 of this total exposure for the detector exposure of TA. Correspondingly, statistical uncertainties for TA are doubled. Systematic uncertainties are derived from energy rescaling efforts by the Auger-TA energy spectrum working group [18] and upper limits from [16].

Figure 1 presents reconstruction results from simulated TA and Auger datasets. The inferred source spectra and spectrum-weighted mass fractions of each source component are recovered well, demonstrating accurate propagation of information through the model. The reconstructed energy spectra and mass composition moments for both detectors align well with true values, although the TA results exhibit larger uncertainties due to lower exposure. Nonetheless, the true parameter values consistently lie within the inferred 2σ uncertainty intervals for all inferred source parameters, as shown in Figure 2.

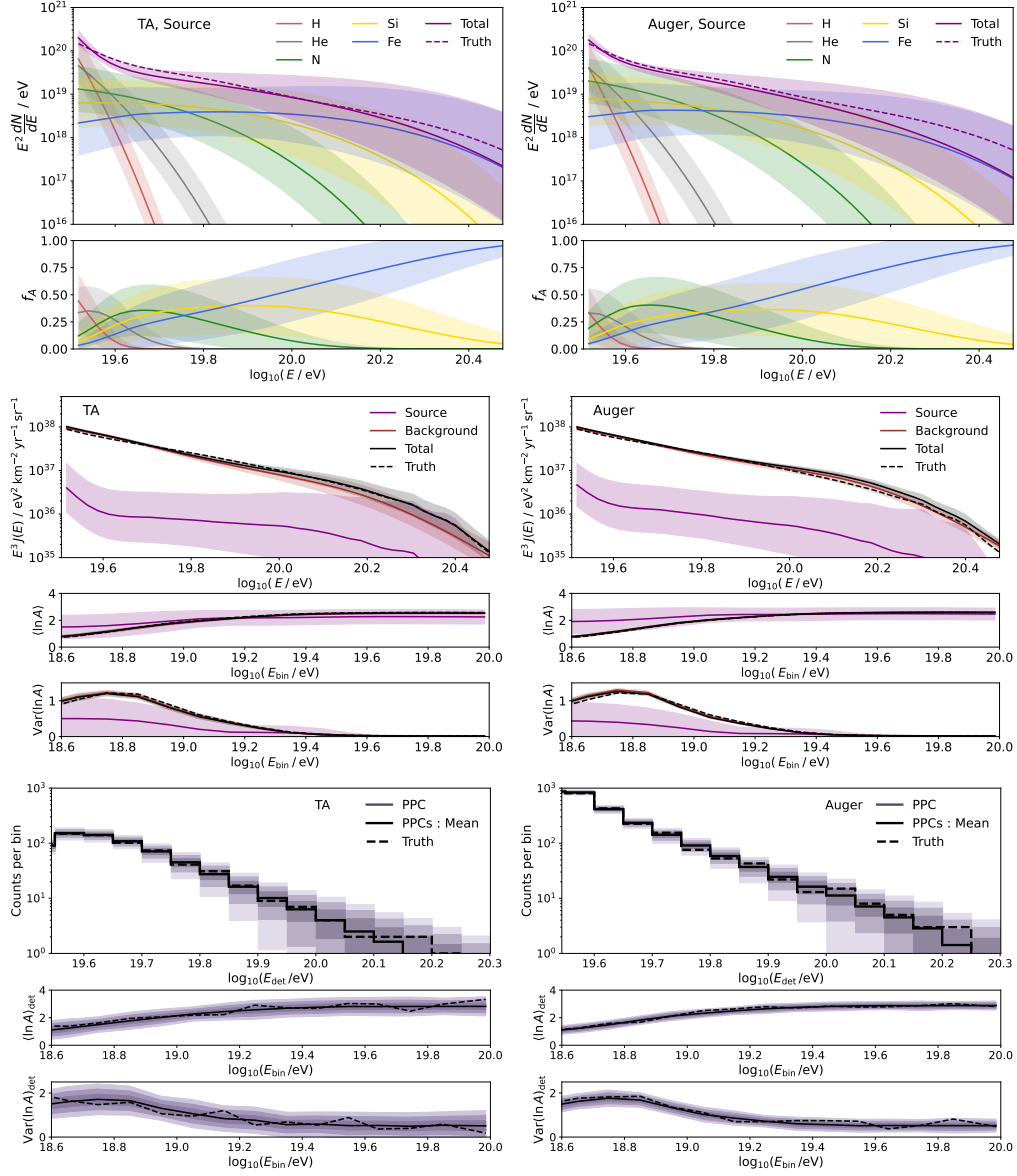


Figure 1: Reconstructed results from simulated datasets based on TA (left) and Auger (right). *Top:* Inferred source spectrum of the point source, weighted by inferred mass fractions for each injected composition. Insets show spectrum-weighted mass fractions at each energy. *Middle:* Inferred energy spectrum, weighted by the inferred flux contributions from the point and background sources. Insets display mean and variance of $\ln A$ per energy bin. *Bottom:* Posterior predictive checks (PPCs), demonstrating forward-modeled results. Shaded regions represent the 1, 2, and 3σ confidence intervals around the mean (black solid lines). True values are indicated by dashed lines.

We also perform posterior predictive checks (PPCs) by generating realizations of the simulated datasets using 100 posterior samples. This verifies the accurate reconstruction of detected energies within 1σ confidence and the mass moments of $\ln A$ within 3σ , reflecting the reduced binning resolution employed in the current model. We also note that the inferred parameter contours in Figure 2 are not exclusively dominated by statistical uncertainty; rather, they reflect limitations due

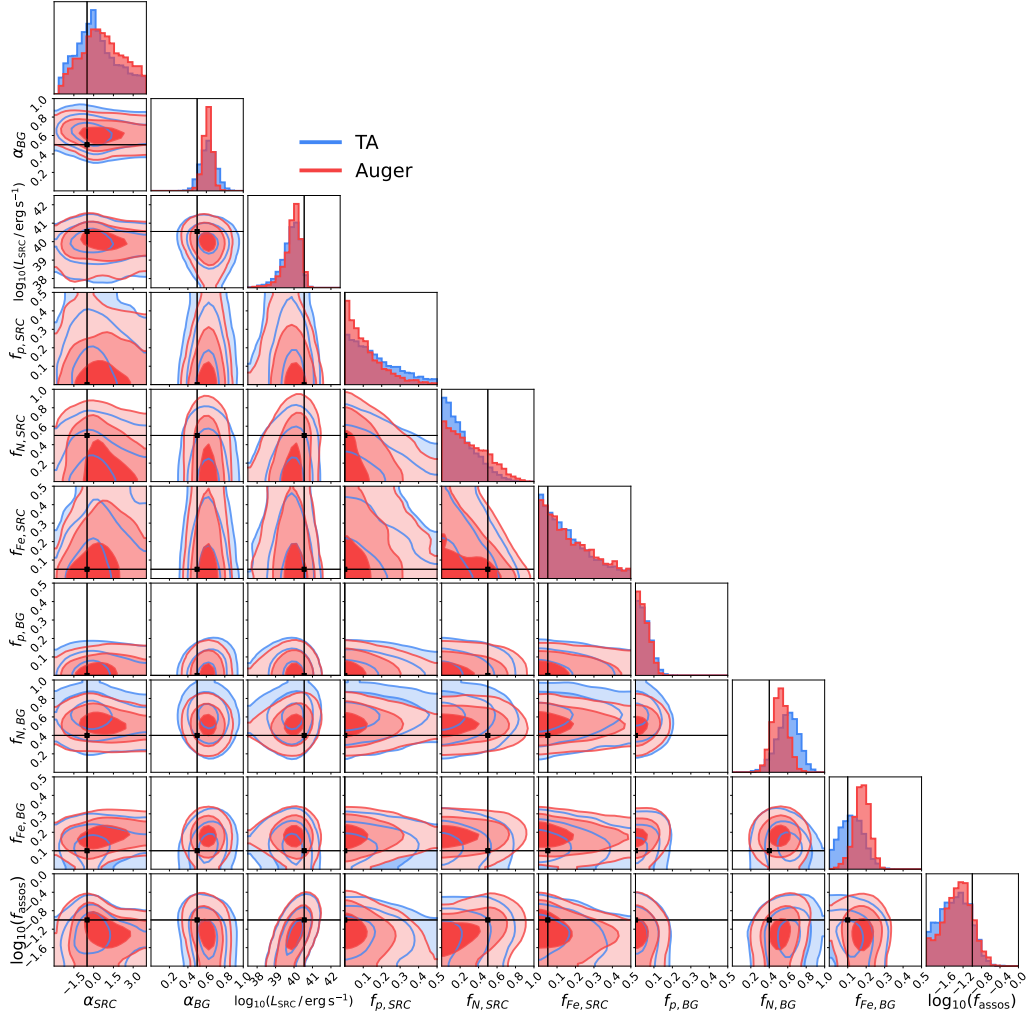


Figure 2: Corner plot of joint posterior distributions for key source parameters: spectral indices for point and background sources, source luminosity, proton, nitrogen, and iron mass fractions for point and background sources, and source association fraction. Results from TA (blue) and Auger (red) are shown, with contours representing 1, 2, and 3 σ intervals. True simulation values are indicated by black lines and dots. Figure generated using `corner.py` [19].

to insufficiently detailed features in the spectral and average composition data alone. Therefore, incorporating spatial information and, ultimately, event-by-event mass measurements is expected to significantly improve constraints on source parameters.

4. Conclusion

In this work, we presented a Bayesian hierarchical framework that integrates individual event energies and statistically averaged, binned mass composition data to infer the properties of UHECR sources, including luminosity, spectral indices, and injected mass fractions. The nuclear propagation code `PrINCE` is employed to model the observed energy spectrum and the mean and variance of $\ln A$ at Earth, clearly distinguishing between discrete injections from point sources

and continuous injections from a homogeneous background. Our forward model also accounts for detector responses, incorporating both statistical and systematic uncertainties. Through application to simulated datasets representative of the Auger and TA observatories, we demonstrated that our framework reliably reconstructs the true source parameters at each stage of the modeling process.

Previously, we showed that rigidity-dependent deflections from Galactic and extragalactic magnetic fields can be incorporated alongside energy and mass composition models [4–6]. Our future plans involve integrating this deflection model into the current framework, enabling a comprehensive analysis that utilizes all available observational information from existing UHECR observatories. We aim to apply this extended approach to publicly available data from Auger and TA, eventually expanding the analysis to include catalogs of potential point-source candidates such as starburst galaxies.

Acknowledgements We acknowledge support from Academia Sinica under Grant No. AS-GCS-113-M04. AF acknowledges additional support from the National Science and Technology Council under Grant No. 113-2112-M-001-060-MY3. Computational resources were provided by the Academia Sinica Grid Computing Center (ASGC) with support from the Institute of Physics, Academia Sinica.

References

- [1] PIERRE AUGER collaboration, *EPJ Web Conf.* **210** (2019) 06002 [1905.04472].
- [2] TELESCOPE ARRAY collaboration, *Nucl. Instrum. Meth. A* **1019** (2021) 165726 [2103.01086].
- [3] PIERRE AUGER collaboration, *JCAP* **01** (2024) 022 [2305.16693].
- [4] K. Watanabe et al., *EPJ Web Conf.* **283** (2023) 03009.
- [5] K. Watanabe et al., *PoS ICRC2023* (2023) 479.
- [6] K. Watanabe et al., *PoS UHECR2024* (2025) 017.
- [7] J. Heinze et al., *Astrophys. J.* **873** (2019) 88 [1901.03338].
- [8] S.D. Team, *Stan modeling language users guide and reference manual*, 2024.
- [9] A.J. Koning et al., *AIP Conf. Proc.* **769** (2005) 1154.
- [10] A. Mücke et al., *Comput. Phys. Commun.* **124** (2000) 290 [astro-ph/9903478].
- [11] PIERRE AUGER collaboration, *JCAP* **04** (2017) 038 [1612.07155].
- [12] D. Ehlert et al., *Phys. Rev. D* **107** (2023) 103045 [2207.10691].
- [13] H. Yüksel et al., *Astrophys. J. Lett.* **683** (2008) L5 [0804.4008].
- [14] TELESCOPE ARRAY collaboration, *Astrophys. J. Lett.* **867** (2018) L27 [1809.01573].
- [15] H. Yüksel et al., *Astrophys. J.* **758** (2012) 16 [1203.3197].
- [16] PIERRE AUGER collaboration, *Phys. Rev. D* **111** (2025) 022003 [2406.06319].
- [17] PIERRE AUGER collaboration, *Astrophys. J.* **935** (2022) 170 [2206.13492].
- [18] TELESCOPE ARRAY, PIERRE AUGER collaboration, *PoS ICRC2021* (2021) 337.
- [19] D. Foreman-Mackey, *The Journal of Open Source Software* **1** (2016) 24.