


## Improving composition of ultrahigh energy cosmic rays with ground detector data

Blaž Bortolato<sup>1,2,\*</sup>, Jernej F. Kamenik<sup>1,2,†</sup> and Michele Tammaro<sup>1,‡</sup>

<sup>1</sup>*Jožef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia*

<sup>2</sup>*Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia*

 (Received 13 September 2023; accepted 19 January 2024; published 14 February 2024)

We show that the maximum shower depth ( $X_{\max}$ ) distributions of ultrahigh energy cosmic rays (UHECRs), as measured by fluorescence telescopes, can be augmented by building a mapping to observables collected by surface detectors. The resulting statistical improvement of such an augmented dataset depends in a universal way on the strength of the correlation exhibited by the mapping. Building upon the publicly available data on “golden hybrid” events from the Pierre Auger Observatory, we project possible improvements in the inferred composition of UHECRs for a range of possible mappings with varying correlation strengths.

DOI: [10.1103/PhysRevD.109.043023](https://doi.org/10.1103/PhysRevD.109.043023)

### I. INTRODUCTION

Cosmic rays (CRs) with ultrahigh energy  $E \geq 10^{18}$  eV are studied through the observation of the extensive air showers (EAS) that are generated in Earth’s atmosphere. The Pierre Auger Observatory [1] employs a “hybrid” detection method: Surface detectors (SDs) measure electrons and muons propagating to the surface, while fluorescent detectors (FDs) capture the isotropic light emitted by excited nitrogen and can track the development of the EAS. In particular, the former can measure the atmospheric depth at which the number of charged particles is at its maximum,  $X_{\max}$ , which is, in turn, a key ingredient to infer the mass composition of CRs.

While the hybrid measurement technique provides detailed insights for the EAS study, the run-time of FDs is largely limited by background light noise. This results in  $\sim 10\%$  of the total number of showers observed to be reconstructed as “hybrid,” i.e., with both FD and SD data, while the remaining part of the dataset is measured with the SD only. For example, in the Pierre Auger 2021 open data [2], there are 22731 SD measurements of EAS, which we refer to as nonhybrid (NH) showers, and 3156 “brass hybrid” (BH) events, that is, showers that have been

recorded simultaneously by the SD and the FD. Of these BH, 1602 are called “golden hybrids” (GHs), with independent SD and FD reconstructions. The low statistics of BH and especially GH events compared to NH is currently a dominant limiting factor in the determination of the CR mass composition.

A possible solution is to infer  $X_{\max}$  from ground data, exploiting GH events to build a map between SD and FD observables. Unfortunately, linear and simple nonlinear fits based solely on measured data, such as the  $\Delta$  variable proposed in Ref. [3], exhibit only relatively low levels of predictive power (which we quantify in Sec. III). Recently, an alternative has been proposed in terms of a deep neural network trained on simulated ground data for four different primaries to directly predict  $X_{\max}$  [4]. Unfortunately, simulations currently cannot faithfully reproduce the observed ground data of Pierre Auger events [5]. This exposes any approach directly relying on such simulations to potentially large systematic errors. In addition, the chosen composition of simulated events used to train the network can act as a confounding variable on the inferred  $X_{\max}$ . This can, in turn, lead to biased estimates of CR composition.

In this work, we address these drawbacks and propose a general method for exploiting any observed correlations between different datasets of a single process to maximize their combined statistical power. In particular, given one (smaller) dataset, containing measurements of a desired observable (i.e.,  $X_{\max}$ ), and another (bigger) dataset, which contains only measurements from which  $X_{\max}$  cannot be reconstructed directly but which do exhibit correlations with  $X_{\max}$ , we demonstrate how to effectively increase the

\*blaz.bortolato@ijs.si

†jernej.kamenik@cern.ch

‡michele.tammaro@ijs.si

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.*

statistics of the smaller dataset in determining (the distribution of)  $X_{\max}$ . The sole assumption is that both datasets are drawn from the same underlying distribution; i.e., the observed physical process is the same. This method is perfectly suited for the specific case of ultrahigh energy CRs, where we can exploit the correlation between SD and FD data to enhance the statistics of  $X_{\max}$ .

The remainder of the paper is structured as follows: After briefly reviewing possible maps in Sec. II, in Sec. III we show a rigorous way to combine the GH data with the inferred one, which leads to the enhancement of the dataset by an effective number of GH showers,  $N_{\text{eff}} = f \times N_{\text{GH}}$ . The latter number monotonically increases with the level of correlation given by the map. Finally, in Sec. IV, we employ this method to improve the results on the full mass composition obtained in Ref. [6].

## II. CORRELATING GROUND DETECTOR DATA WITH $X_{\max}$

For our purpose, the  $i$ th GH event in the dataset can be characterized by one FD measurement,  $x^{(i)} \equiv X_{\max}^{(i)}$ , and a set of  $m$  SD observables,  $\mathcal{Q}^{(i)} = \{Q_1^{(i)}, \dots, Q_m^{(i)}\}$ . We can then consider functions of SD observables  $y(\{\mathcal{Q}\})$  for which the GH set exhibits some correlation between  $x$  and  $y$ , as parametrized by

$$\rho(x, y) = \frac{\text{Cov}(x, y)}{\sigma(x)\sigma(y)}, \quad (1)$$

where  $\sigma(x)$ ,  $\sigma(y)$ , and  $\text{Cov}(x, y)$  are the standard deviations of  $x$  and  $y$  and their covariance, respectively.

Possibly the simplest example of  $y$  is the weighted average of  $Q_j$ :

$$y^{(i)} = \sum_{j=1}^m a_j Q_j^{(i)}, \quad (2)$$

where the weights  $a_j$  are fixed by maximizing  $\rho(x, y)^2$ . For simplicity, we assume that the Parton Distribution Function (PDF) of each variable is a normal distribution described by its measurement and the respective uncertainty. Namely, we have

$$P^{(i)}(x) \sim \mathcal{N}(x|x^{(i)}, \delta x^{(i)}), \quad (3)$$

and similarly for  $P^{(i)}(y)$ . Note that the fit has to be performed separately for each subset of GH events (i.e., in a particular CR energy bin or angular window) for which one wants to infer composition. This ensures an unbiased estimator, assuming the GH and NH events in the same subset are of the same composition and follow the same distribution of  $\mathcal{Q}$ .

As a proof of concept, we use the set of observables<sup>1</sup>

$$\mathcal{Q}^{(i)} = \{\theta, \langle T_s \rangle, \sigma(T_s), \max(T_s)\}^{(i)}, \quad (4)$$

with their respective uncertainties. Here,  $\theta$  is the azimuthal angle of the shower core, while  $T_s = (t_{\text{end}} - t_{\text{start}})_s$  is the duration of the signal in the  $s$ th Cherenkov station hit by the shower, defined as the difference between the final and initial time bin. We take the mean, standard deviation, and maximum values over the active stations for each event. For each shower, we generate multiple samples of the observables from their PDF, Eq. (3). In such a way, we can also compute the uncertainty associated with  $y^{(i)}$ . Uncertainties on the bins  $t_{\text{start}}$  and  $t_{\text{end}}$  are not provided in the open data; hence, we assume the uncertainties on both quantities to be  $\pm$  one bin. These samples are then utilized to maximize  $\rho^2$  through the ‘‘Nelder-Mead’’ (also known as ‘‘simplex-downhill’’) method (any similar optimization algorithm would work). Finally, we employ  $K$ -folding to check for overfitting. Following this procedure, we are able to achieve  $\rho = (30 \pm 4)\%$  correlation between  $x$  and  $y$  for GH in the energy bin [2, 5] EeV, with 455 GH showers. We chose this bin as the only one where GH and NH data overlap in the Pierre Auger open dataset.

In practice, more advanced approaches, such as non-linear fits or neural networks, may be employed to obtain stronger correlations between  $y$  and  $x$ . As two explicit examples of such more complicated observables, we mention the  $\Delta$  variable [3] and the deep neural network trained on simulated ground data [4]. In the same energy bin, the former yields  $\rho = 0.26 \pm 0.02$ ,<sup>2</sup> while the latter can reach  $\rho = 63\%$ . We stress that for our purpose the strength of an observable is measured solely in terms its correlation with  $x$  as established on a particular measured GH dataset  $\{(x, y)\}$ . In this way, we avoid potential systematic biases associated with direct inference of individual  $x^{(i)}$  from nonequivalent datasets or simulations.

## III. INFERENCE OF $X_{\max}$ DISTRIBUTION

Following the previous section, we can represent the GH dataset as a set of  $N$  pairs  $(x_j, y_j)$ , with  $j = 1, \dots, N$ , and we assume that there is some level of correlation between the two observables. The NH dataset is then represented by the  $M$  pairs  $(\hat{x}_k, \hat{y}_k)$ , with  $k = 1, \dots, M$ , where  $M \gg N$ .

<sup>1</sup>We have repeated this exercise with multiple combinations of available observables. The chosen set represents the one with the largest correlation achievable without overfitting of the data.

<sup>2</sup>Uncertainties due to limited statistics and above stated systematics in both cases have been estimated using bootstrapping [7].

In the latter set, the  $\hat{x}_k$  entries, which are missing in the original data, can be inferred by exploiting the map built through the  $\{(x, y)\}$  set correlation, that is,  $\hat{x} = \hat{x}(\hat{y})$ .

Let  $P(x, y)$  be the joint probability distribution of the  $\{(x, y)\}$  set. For simplicity, we start by assuming that such distribution follows a bivariate normal (BN) model, while we comment later for the general case. Each BN distribution is defined by the means  $\mu_x$  and  $\mu_y$ , the standard deviations  $\sigma_x$  and  $\sigma_y$ , and the correlation  $\rho$ . The conditional distribution  $P(x|y)$  is then simply given by the normal distribution with mean  $\mu_x + \rho \frac{\sigma_x}{\sigma_y}(y - \mu_y)$  and standard deviation  $\sigma_x \sqrt{1 + \rho^2}$ . Therefore, one can sample  $\hat{x} \sim P(x|y)$  following the formula

$$\hat{x}(\hat{y}) = \mu_x + \rho \frac{\sigma_x}{\sigma_y}(\hat{y}' - \mu_y) + \sigma_x \sqrt{1 + \rho^2} \epsilon, \quad (5)$$

where  $\hat{y}' = \hat{y} + \delta\hat{y}\epsilon$  accounts for systematic uncertainty of  $\hat{y}$  and  $\epsilon$  is a random number drawn from a normal distribution,  $\epsilon \sim \mathcal{N}(0, 1)$ . In this way, we obtain a set of inferred values,  $\{\hat{x}(\hat{y}_1), \dots, \hat{x}(\hat{y}_M)\}$ .

Because of the finite number ( $N$  and  $M$ ) of events, we need to account for statistical uncertainties in the inference via bootstrapping [7]. The latter method consists of inferring values  $\hat{x}$  for each  $\hat{y}$  multiple times, each time starting with a different set of pairs  $\{(\tilde{x}, \tilde{y})\}$ , called the bootstrapped sample. The procedure is repeated  $B$  times, where each time the bootstrapped sample is given by sampling  $N$  pairs  $(\tilde{x}_j, \tilde{y}_j)$  from the first set  $\{(x, y)\}$  with allowed repetitions.

Namely, at bootstrap step  $\ell$ , where  $\ell = 1, \dots, B$ , we have

$$\begin{aligned} \tilde{x}_{j,\ell} &= (x_j + \delta x_j \epsilon_{j,\ell}) \mathcal{O}_{j,\ell}, \\ \tilde{y}_{j,\ell} &= (y_j + \delta y_j \epsilon_{j,\ell}) \mathcal{O}_{j,\ell}. \end{aligned} \quad (6)$$

Here,  $\delta x_j$  and  $\delta y_j$  represent the systematic uncertainties on  $x_j$  and  $y_j$ , while  $\epsilon_{j,\ell} \sim \mathcal{N}(0, 1)$ . Finally, the  $N$ -dimensional vector  $\mathcal{O}_\ell$  is drawn from the multinomial distribution with  $N$  events and  $N$  classes which are all equally probable,  $\text{Mult}(N, p = (1/N, \dots, 1/N))$ . The values in this array,  $\mathcal{O}_{j,\ell}$ , count how many times each pair  $j$  is chosen in the  $\ell$ th bootstrapped sample. It, thus, satisfies the relation  $\sum_j \mathcal{O}_{j,\ell} = N$ . As a result of this procedure, we obtain at each step  $\ell$  the set of  $N$  couples

$$\mathcal{X}_\ell = \{(\tilde{x}, \tilde{y})\}_\ell. \quad (7)$$

We can now estimate the  $\ell$ th joint distribution  $P_\ell(\tilde{x}, \tilde{y})$  from the bootstrapped samples.

Similarly, we define

$$\tilde{y}_{k,\ell} = (\hat{y}_k + \delta\hat{y}_k \epsilon_{k,\ell}) \tilde{\mathcal{O}}_{k,\ell}, \quad (8)$$

where now the index  $k$  runs over the  $\hat{y}$  elements,  $k = 1, \dots, M$ , and  $\tilde{\mathcal{O}}_{k,\ell} \sim \text{Mult}(M, p = (1/M, \dots, 1/M))$ .

The  $\tilde{y}_{k,\ell}$  account for both statistical and systematic uncertainties.

For each bootstrapped sample, we can now obtain a set of  $M$  pairs

$$\mathcal{X}_\ell^{\text{inf}} = \{(\hat{x}(\tilde{y}), \tilde{y})\}_\ell, \quad (9)$$

where the first component of the pair is inferred by plugging  $\tilde{y}_{k,\ell}$  in Eq. (5). Finally, we can combine the two sets to obtain

$$\mathcal{X}_\ell^{\text{comb}} = \mathcal{X}_\ell \cup \mathcal{X}_\ell^{\text{inf}} = \{(X, Y)\}_\ell. \quad (10)$$

At this point, one can use the larger combined dataset to perform the desired analysis. As we obtain the  $\hat{x}$  elements from the inference method described above, it is fair to ask what is the gain in terms of statistical power with the combined dataset. In order to estimate it, we can compute the variance of the distribution of the means,  $\text{Var}(\bar{x})$ . That is, we compute the mean value of  $\tilde{x}$  for each  $\mathcal{X}_\ell$  set,  $\bar{x}_\ell$ ; the resulting distribution defined by all  $\bar{x}_\ell$  is a Gaussian with variance  $\text{Var}(\bar{x})$ . The ratio between the variances obtained with the initial dataset  $\mathcal{X}_\ell$  and with the combined one,  $\text{Var}(\bar{X})$ , can provide a measure of the improvement in the statistical significance obtained with this method. Indeed, the quantity

$$f \equiv \frac{\text{Var}(\bar{x})}{\text{Var}(\bar{X})} \quad (11)$$

can be interpreted as the effective increase of events in the combined dataset, with respect to the original one,  $N_{\text{eff}} = f \times N$ . This interpretation follows from the fact that the variance of the mean distribution is given by the variance divided by the number of events. Namely, if  $f = 1$ , it means no additional information is contained in the inferred set  $\mathcal{X}^{\text{inf}}$ . On the contrary, if  $f > 1$ , the effective statistical power of  $\mathcal{X}^{\text{comb}}$  amounts to having  $f \times N$  total events. The functional dependence of  $f$  on the correlation  $\rho$ , for different values of  $M$  and  $N$ , is shown in Figs. 1 and 2. Points in these figures are obtained numerically, with the procedure described above, using  $B = 10^5$  bootstrapped samples. Furthermore, we assumed that systematic uncertainties are small. The latter translates to having  $\rho \sigma_x / \sigma_y \delta\hat{y} \ll \sigma_x \sqrt{1 + \rho^2}$  in Eq. (5). In this scenario, the ratio  $f$  increases monotonically with  $\rho$  and does not depend on the number of events  $M$  and  $N$  but only on their ratio; see Fig. 1. Larger  $M/N$  ratios allow then for larger values of  $f$  for the same level of correlation, as shown in Fig. 2.

In the limit of  $\rho \rightarrow 1$ ,  $f$  tends to the maximal value  $f = (M + 1)/N$ . In practice,  $\rho = 1$  cannot be obtained

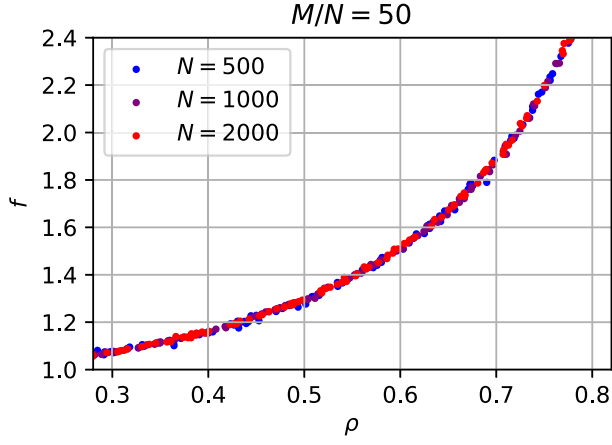


FIG. 1. Effective increase  $f$  as a function of correlation  $\rho$ , for three values of GH dataset size  $N$ , when fixing the ratio  $M/N = 50$ .

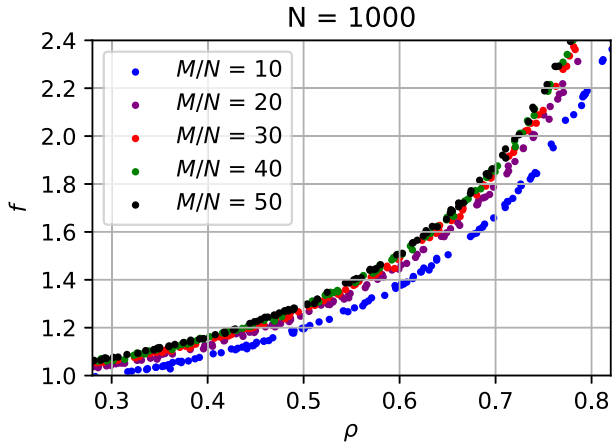


FIG. 2. The same as Fig. 1, with  $N = 1000$  and different values of the ratio  $M/N$ .

with a finite number of events  $N$  due to statistical uncertainties; thus, the maximum value of  $f$  is smaller.

In the general case, the joint distribution  $P(x, y)$  does not follow a BN distribution. One may then try to use kernel density estimation to obtain  $P(x, y)$ ; however, if the latter is computed inappropriately, we may introduce a bias in inferring  $\hat{x}$  values (variance-bias trade-off). For this reason, we propose to transform bootstrapped samples using a probability integral transformation in 2D [8], which has the advantage to preserve the correlation between  $x$  and  $y$ .

In the Auger open data, NH and GH events given in the open data in the energy interval  $E/E_{\text{eV}} \in [2.5, 10]$  amount to  $M \sim 15700$  and  $N \sim 300$ , respectively, thus with a ratio  $M/N \sim 50$ . In the case of maps between  $y$  and  $x \equiv X_{\text{max}}$  constructed with linear or nonlinear fits (see the previous section), the correlation is  $\rho \simeq 0.3$ , which corresponds to  $f \approx 1.06$ . On the other hand, the deep neural network

presented in Ref. [4] can reach  $\rho = 63\%$ , resulting in  $f = 1.4\text{--}1.5$  for  $M/N = 10\text{--}50$ .

#### IV. PROJECTIONS OF UHECR COMPOSITION FROM NONHYBRID EVENTS

Using the  $X_{\text{max}}$  distribution inferred from a combination of GH and NH events, we follow our previous analysis of the mass composition of ultrahigh energy cosmic rays (UHECRs) [6]. That is, we infer the posterior distribution of the composition  $w = (w_p, \dots, w_{\text{Fe}})$  using the moment decomposition of measured and simulated  $X_{\text{max}}$  distributions in a fixed energy bin. However, due to the small number of events in the energy bin  $E/E_{\text{eV}} \in [2.5, 5]$ , we use here the larger  $E/E_{\text{eV}} \in [0.65, 5]$  energy bin for projections. We use the same set of shower simulations and the same inference procedure of Ref. [6]; the improvements shown here come solely from considering the enhanced dataset,  $\mathcal{X}_{\ell}^{\text{comb}} \equiv \{X_{\ell,k} | k = 1, \dots, N + M\}$ .

At each bootstrap step  $\ell$ , we compute the moments of the  $X_{\text{max}}$  distribution as

$$z_{\ell,1} = \frac{1}{N+M} \sum_{k=1}^{N+M} X_{\ell,k}, \quad (12)$$

$$z_{\ell,n} = \frac{1}{N+M} \sum_{k=1}^{N+M} (X_{\ell,k} - z_{\ell,1})^n \quad (13)$$

for  $n = 2, 3, 4$ . We then calculate the mean of the distribution of moments  $\mu$  and the covariance matrix  $\Sigma$ . The effect of the enhanced dataset can be included here by considering the new covariance matrix  $\Sigma_{\text{eff}} = \Sigma/f$ .

In Fig. 3, we show the composition obtained using a mixture model of four primaries (p, He, N, and Fe), using

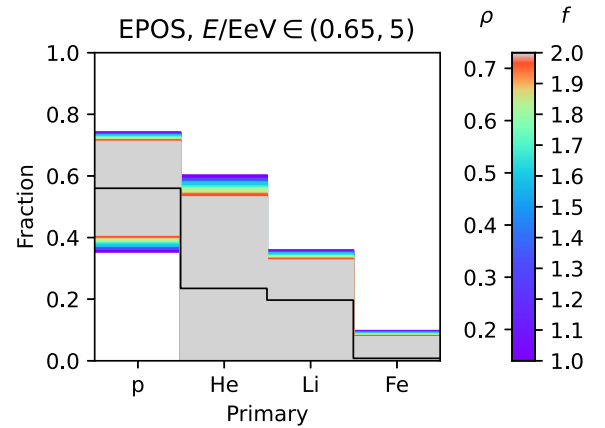


FIG. 3. Fraction of primaries p, He, Li, and Fe shown as  $2\sigma$  confidence intervals for various values of  $f$  and corresponding  $\rho$  for hadronic model EPOS. The black solid line represents the most probable composition.

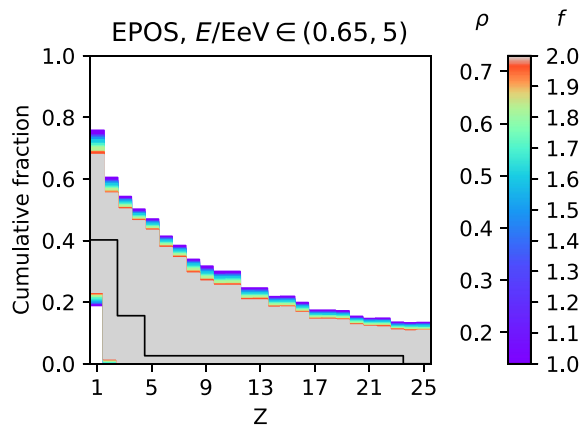


FIG. 4. Fraction of primaries shown as  $2\sigma$  confidence intervals for various values of  $f$  and corresponding  $\rho$  for hadronic model EPOS. The black solid line represents the most probable composition.

the EPOS<sup>3</sup> hadronic model. The black line indicates the best fit, while the bars represent the allowed fractions at 95% C.L. Different colors show the effect of including inferred values of  $X_{\max}$ , for different values of  $f$ : the gray band obtained for  $f = 2$ , that is, by doubling the size of the GH dataset, down to the purple one, where  $f = 1$ .

Similarly, in Fig. 4, we show the result for the full cumulative composition; that is, for each primary  $Z_0$ , we show the allowed fraction of heavier primaries,  $Z > Z_0$ . We observe again an improvement of the gray region with respect to the purple.

We provide the same results for the hadronic models Sibyll, QGSJet01, and QGSJetII-04 in the Appendix.

## V. CONCLUSIONS

We have developed a novel method to improve the statistical power of GH events by including  $X_{\max}$  values of UHECRs inferred from SD observables. The effective number of events obtained is quantified by the variance ratio  $f$ , which strongly depends on the level of correlation

<sup>3</sup>Energy conserving quantum mechanical approach, based on Partons, parton ladders, strings, Off-shell remnants and Splitting of parton ladders.

between ground and fluorescent data. While simple linear or nonlinear fits can achieve at most 30% correlation levels, resulting in rather modest values of  $f$ , more advanced approaches can reach above 60% correlation [4], leading to  $f \gtrsim 1.5$ .

We have projected the CR composition inference with such enhanced datasets, as a function of the ratio  $f$ . The resulting potential improvements in the composition C.L. for the Auger open data are shown in Figs. 3 and 4 and in the Appendix. Although the latter results seem meager, this study shows that, as a proof of concept, these improvements are possible and can be exploited by the experimental collaboration on their full dataset. Furthermore, it underlines the potential of probing deeper the correlation of SD and FD data, where the advancement in artificial intelligence tools can lead to interesting and powerful results.<sup>4</sup>

Finally, we point out that the procedure to enhance the  $X_{\max}$  dataset by SD observables is very general and based on statistics arguments only. That is, any analysis based on correlated observables can make use of this method to augment a dataset containing measurements of both observables with a (preferably much larger) dataset of measurements of only a single one.<sup>5</sup>

## ACKNOWLEDGMENTS

We are grateful to Gregor Šega for helping in reviewing the method outlined in this work. The authors acknowledge the financial support from the Slovenian Research Agency (Grant No. J1-3013 and research core funding No. P1-0035).

<sup>4</sup>We note that unbiased statistical improvement via our method relies purely on the correlation observed in the actual data. In the case one is able to construct and measure observables for which there is firm theoretical knowledge of their correlation with  $X_{\max}$ , then one can infer the latter directly (e.g., via a capable enough approximator like a neural network, as done in Ref. [4]).

<sup>5</sup>For the Pierre Auger Collaboration, this could be advantageous, i.e., in the exploration of observables to discriminate between hadronic models, a direction we leave for future work.

## APPENDIX: UHECR COMPOSITION FROM NONHYBRID EVENTS

Below, we present the projections of compositions for all hadronic models considered: EPOS (as in the main text), Sybill, QGSJetII-04, and QGSJet01 in the energy bin  $E/E_{\text{eV}} \in [0.65, 5]$ . In the case of four primaries, projections are shown in Fig. 5, while for 26 primaries are given in Fig. 6.

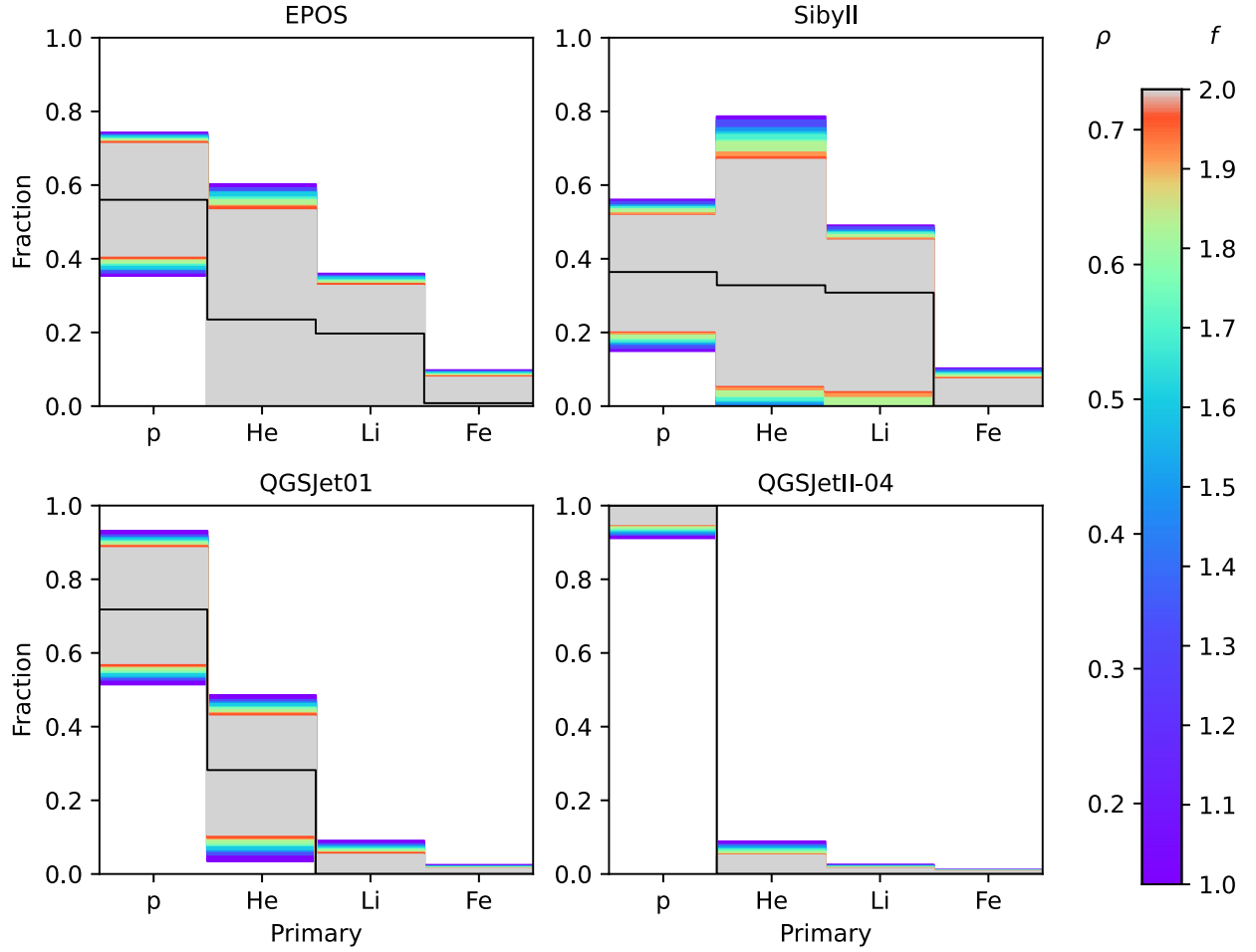


FIG. 5. Fraction of primaries p, He, Li, and Fe shown as  $2\sigma$  confidence intervals for various values of  $f$  and corresponding  $\rho$  for hadronic models EPOS, Sibyll, QGSJetII-04, and QGSJet01 in energy interval  $E/E_{\text{eV}} \in [0.65, 5]$ . The black solid line represents the most probable composition.

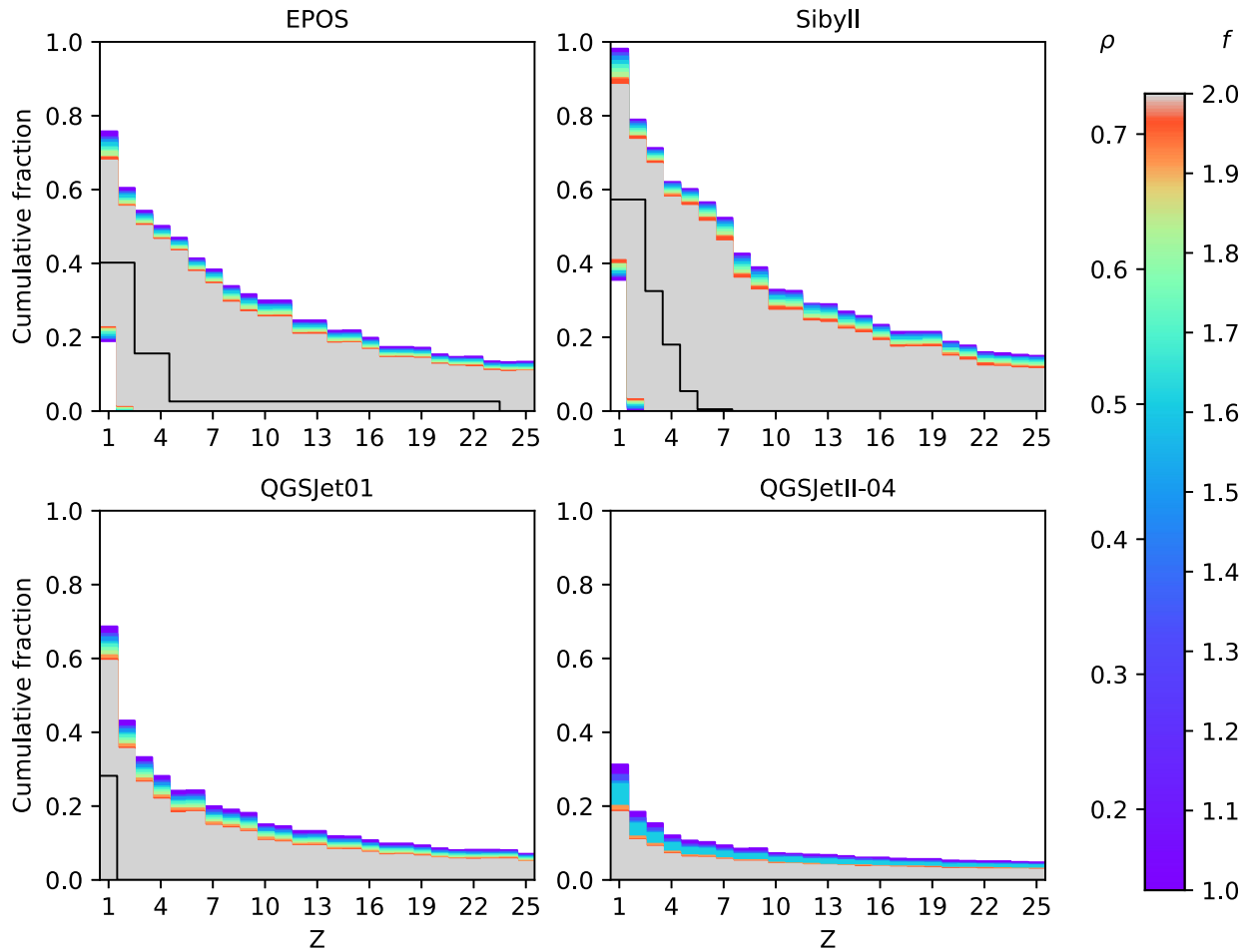


FIG. 6. Fraction of primaries shown as  $2\sigma$  confidence intervals for various values of  $f$  and corresponding  $\rho$  for hadronic models EPOS, Sibyll, QGSJetII-04, and QGSJet01 in energy interval  $E/E_{\text{eV}} \in [0.65, 5]$ . The black solid line represents the most probable composition.

- 
- [1] A. Aab *et al.* (Pierre Auger Collaboration), *Nucl. Instrum. Methods Phys. Res., Sect. A* **798**, 172 (2015).  
 [2] The Pierre Auger Collaboration, Pierre Auger Observatory 2021 open data, [10.5281/zenodo.4487613](https://zenodo.org/record/4487613) (2021).  
 [3] A. Aab *et al.* (Pierre Auger Collaboration), *Phys. Rev. D* **96**, 122003 (2017).  
 [4] A. Aab *et al.* (Pierre Auger Collaboration), *J. Instrum.* **16**, P07019 (2021).  
 [5] L. Cazon (EAS-MSU IceCube, KASCADE Grande, NEVOD-DECOR, Pierre Auger, SUGAR, Telescope Array, and Yakutsk EAS Array Collaborations), *Proc. Sci. ICRC2019* (2020) 214 [arXiv:2001.07508].  
 [6] B. Bortolato, J. F. Kamenik, and M. Tamaro, *Phys. Rev. D* **108**, 022004 (2023).  
 [7] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*, Monographs on Statistics and Applied Probability No. 57 (Chapman & Hall/CRC, Boca Raton, FL, 1993).  
 [8] F. N. David and N. L. Johnson, *Biometrika* **35**, 182 (1948).