The muon deficit problem: a new method to calculate the muon rescaling factors and the Heitler-Matthews β exponent

Kevin Almeida Cheminant,[a] **Dariusz Góra,**[a] **Nataliia Borodai,**[a] **Ralph Engel,**[b] **Tanguy Pierog,**[b] **Jan Pękala,**[a,*] **Markus Roth,**[b] **Jarosław Stasielak,**[a] **Michael Unger,**[b] **Darko Veberič**[b] **and Henryk Wilczyński**[a]

[a] *Institute of Nuclear Physics PAS,*
Radzikowskiego 152, Kraków, Poland

[b] *Karlsruhe Institute of Technology, Institute for Astroparticle Physics,*
Karlsruhe, Germany

E-mail: jan.pekala@ifj.edu.pl

Simulations of extensive air showers using current hadronic interaction models predict too small numbers of muons compared to events observed in the air-shower experiments, which is known as the muon-deficit problem. In this work, we present a new method to calculate the factor by which the muon signal obtained via Monte-Carlo simulations must be rescaled to match the data, as well as the β exponent from the Heitler-Matthews model which governs the number of muons found in an extensive air shower as a function of the mass and the energy of the primary cosmic ray. This method uses the so-called z variable (difference between the total reconstructed and the simulated signals), which is connected to the muon signal and is roughly independent of the zenith angle, but depends on the mass of the primary cosmic ray. Using a mock dataset built from QGSJetII-04, we show that such a method allows us to reproduce the average muon signal from this dataset using Monte-Carlo events generated with the EPOS-LHC hadronic model, with accuracy better than 6%. As a consequence of the good recovery of the muon signal for each primary included in the analysis, also the β exponent can be obtained with accuracy of less than 1% for the studied system. Detailed simulations show a dependence of the β exponent on hadronic interaction properties, thus the determination of this parameter is important for understanding the muon deficit problem.

---

*Speaker

1. Introduction

In this work we examine the deficit of muons in simulations of extensive air showers – the muon number observed in data at energies around 10^{19} eV is 30% to 60% higher than in corresponding simulations [1]. The analyses of the observational data depend on simulations, therefore this deficit will systematically influence the results, e.g. studies of the composition of the primary cosmic rays based on observed muon numbers produce results shifted towards heavier elements as compared with analyses based on measurements of $X_{\max}$ (i.e. the atmospheric depth at which the air shower reaches its maximum size) [2].

To study the muon deficit we use a top-down (TD) method – its chain of simulations and reconstructions enable us to calculate signals in the fluorescence (FD) and surface detectors (SD) of the Pierre Auger Observatory [3] (Auger). For each observed shower, starting with a large number of simulated air showers with varying initial conditions, we select the one which has a longitudinal profile most similar to the profile of the observed shower (the reference profile). As a result of the simulation-reconstruction chain we get an event, with complete information about the distributions of the signals in the detectors (including information on the specific components that contribute to these signals) – these signals can then be compared with their reference counterparts. Since the results of the simulations depend on the properties of the hadronic interaction models that are included in the simulation software, by comparing the simulations with corresponding observational results we should be able to verify these models at energies exceeding those available in any man-made accelerators. We expect that we will gain new information, which will enable improvment of the interaction models, and in this way we are also able to reduce the discrepancy between the observations and simulations.
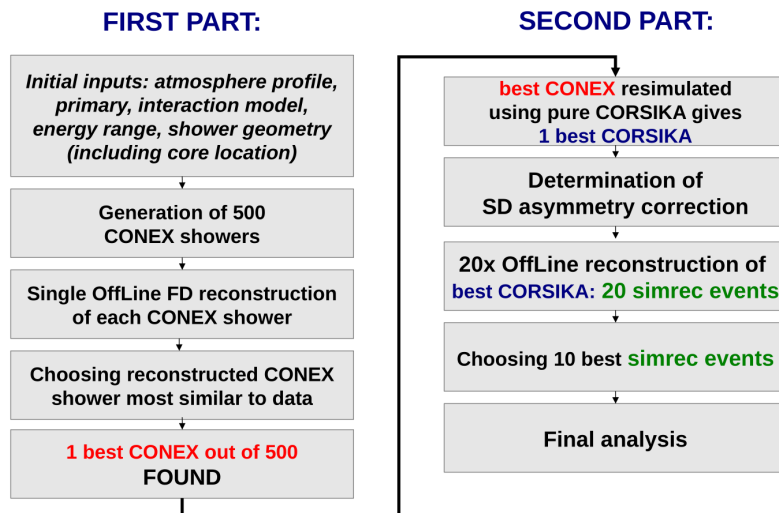

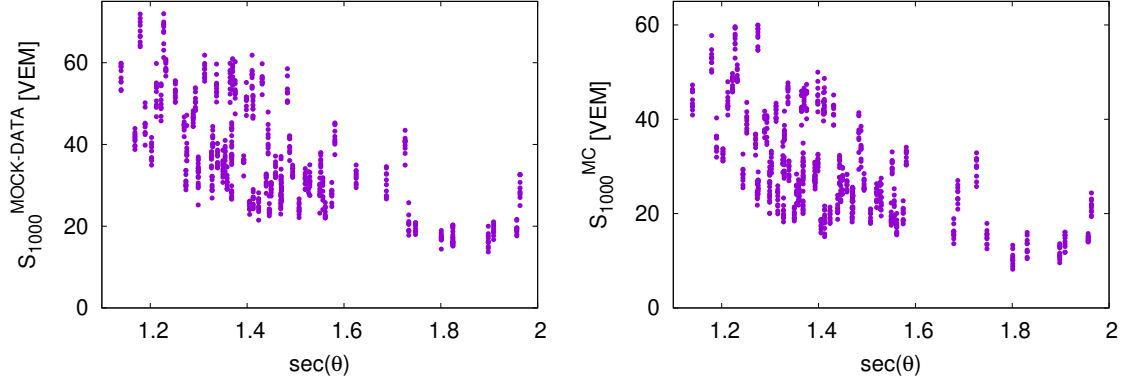**Figure 1:** A scheme of top-down simulation-reconstruction chain.

Figure 2: *Left*: the total MOCK-DATA signal S_{1000} as a function of zenith angle θ, at a reference distance of 1000 m from the shower core; air showers of 10^{19} eV energy with iron primaries and simulated with EPOS-LHC model. *Right*: the total MC signal, for proton air showers of 10^{19} eV simulated with QGSJetII-04.

2. Top-Down method

In this work we use a set of simulated air showers (MOCK-DATA) as a reference – by starting with showers with exactly known properties, and reconstructing them independently we are able to demonstrate the capabilities of the TD method. The TD method we use is an improved version of the simulation-reconstruction chain described in Ref. [1]. We have replaced the SENECA software [4], that was used in simulations, with CORSIKA [5], which was extensively developed and is currently more reliable and accurate. The details of our TD method are described in Refs. [6, 7]. Fig. 1 presents the scheme of the TD chain, that for each of the reference shower generates a set of simulated showers, from which the best one is selected (based on χ^2 calculated between reference and simulated profile). As a final result of this method we get 10 reconstructions of this shower (performed with the Offline software [8]) which are used in the following analysis. The TD calculations were done using two of the most up-to-date nuclear interaction models: we use 78 events for EPOS-LHC [9] and 89 for QGSJetII-04 [10]. Simulations with EPOS-LHC and iron primaries were chosen as the MOCK-DATA showers (Fig. 2) as they produce muon signals most similar to the observed ones at the energy of 10^{19} eV [11]. The QGSJetII-04 simulations that are used to reconstruct the muon signals of the MOCK-DATA set were repeated for four different types of primary cosmic rays: proton, helium, nitrogen and iron.

Almost the whole of the total signal at the ground comes from the electromagnetic and muonic components [1] – we will denote these contributions S_{EM} and S_μ, respectively. Other components of the shower [12] do not contribute significantly. Any discrepancy between corresponding observed and simulated values can be described by linear scaling factors: R_E for the electromagnetic part of an air shower, and R_μ for the muonic part. With these assumptions the ground signal at 1000 m from the axis (S_{1000}) for MOCK-DATA and simulated air shower j can be respectively written as:

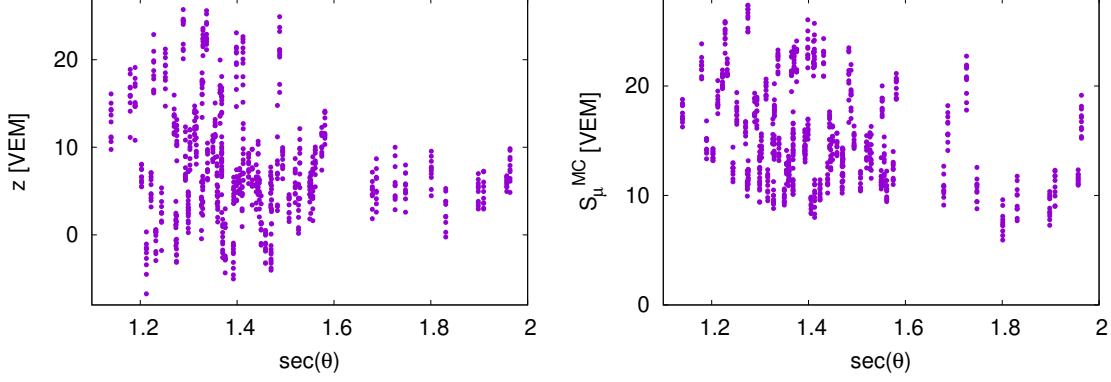$$S_{1000,j}^{MC} \equiv S_{EM,j}^{MC} + S_{\mu,j}^{MC}, \tag{1}$$

3

Figure 3: *Left*: the variable z. *Right*: S_μ for proton air showers of 10^{19} eV, simulated with QGSJetII-04.

$$S_{1000,j}^{\text{MOCK-DATA}}(R_E, R_\mu)_j \equiv S_{\text{EM},j}^{\text{MOCK-DATA}} + S_{\mu,j}^{\text{MOCK-DATA}},$$
$$S_{\text{EM},j}^{\text{MOCK-DATA}} = R_E\, S_{\text{EM},j}^{\text{MC}}, \tag{2}$$
$$S_{\mu,j}^{\text{MOCK-DATA}} = R_\mu\, S_{\mu,j}^{\text{MC}}.$$

The comparison of simulations and measurements performed in Ref. [1] has shown that the value of R_E is very close to one. In our TD method, when we select the simulated shower through comparison with the longitudinal profile of the MOCK-DATA shower, we increase our confidence in the accuracy of the simulation even more. Therefore in our analysis we assume $R_E = 1$.

Rather than using the ground signal directly, we use the difference between MOCK-DATA and the MC signal as the main observable:

$$z_j \equiv S_{1000,j}^{\text{MOCK-DATA}} - S_{1000,j}^{\text{MC}}.$$

This difference explicitly shows the discrepancy between observations and simulations. The variable z is roughly independent of the zenith angle – as is shown for example in Fig. 3 (left), the distribution of z only slightly depends on the zenith angle, as opposed to the S_{1000} signal shown in fig. 2. The largest advantage of using the variable z becomes apparent after comparing Eqs. (1) and (2). By simple subtraction, and assuming $R_E = 1$, we get

$$S_{\mu,i,j}^{\text{MC}} = \frac{z_{i,j}}{R_{\mu,i,j} - 1}. \tag{3}$$

We see that the simulated muon signal is proportional to the difference between observed and MC signal z, and the ratio depends on the muon scaling factor. The average signal difference, $\langle z_i \rangle$, depends on the primary type, and we have to account for this effect (additional index $i \in \{$p, He, N, Fe$\}$). Air showers initiated by heavier primary cosmic rays of the same energy contain larger number of muons, and correspondingly larger muon signals at the ground are expected [13], therefore muon scaling factors $R_{\mu,i}$ should be studied separately at different primary masses.

The average muon signal as a fraction $g_{\mu,i}(\theta)$ of the total signal at the ground has been calculated in previous analyses. This fraction depends on the shower zenith angle and the type of the primary cosmic ray, and only slightly on different hadronic interaction models [11]. For our

primary type i	model	$\langle S_{\mu,i}^{MC} \rangle$ [VEM]	st. dev. [VEM]	model	$\langle S_{\mu,i}^{MC} \rangle$ [VEM]	st. dev. [VEM]
p	EPOS-LHC	16.89 ± 0.31	5.1	QGSJetII-04	15.05 ± 0.3	4.3
He		18.74 ± 0.42	5.7		16.82 ± 0.4	4.7
N		20.67 ± 0.37	5.9		18.96 ± 0.4	5.1
Fe		23.09 ± 0.42	6.4		21.08 ± 0.4	5.7

Table 1: Results obtained using the TD method: the mean muon signal $S_{1000,\mu,i}$ and its standard deviation for EPOS-LHC and QGSJetII-04.

analysis the $g_{\mu,i}(\theta)$ was obtained from the analysis of simulations of muon signal in stations located at a distance of 1000 m from the shower core for showers with energies of $10^{18.5}$ to 10^{19} eV.

We know the values of $z_{i,j}$ and the S_{1000}, and using the parameterization $g_{\mu,i}(\theta)$, we can get the muon scaling factor even for an individual air shower j,

$$R_{\mu,i,j} = 1 + \frac{z_{i,j}}{g_{\mu,i}(\theta)\, S_{1000,i,j}^{MC}}, \tag{4}$$

where $S_{\mu,i,j}^{MC} \equiv g_{\mu,i}(\theta)\, S_{1000,i,j}^{MC}$.

3. Muon rescaling factors

The average reconstructed muon signals $\langle S_{\mu,i}^{MC} \rangle$ are listed in Table 1. For all primaries the QGSJetII-04 model predicts smaller muon signals than simulations with the EPOS-LHC model – that was expected based on the results from Ref. [14]. The difference between both models is in the number of muons, but also in the energy spectrum of the muons (e.g. EPOS-LHC predicts more muons with lower energies) – this leads to a difference of the muon signal at the ground between simulations done with both models. From Table 1, we can calculate the average ratio of the muon signals as $r = S_{1000,\mu,i}^{MC\text{-}EPOS} / S_{1000,\mu,i}^{MC\text{-}QGSJet.} \simeq 1.11 \pm 0.04$. The average value of individual factors $R_{\mu,i}$ is almost independent of the zenith angle and decreases with growing primary mass.

Table 2 presents values of the total muon signal $S_\mu^{MOCK\text{-}DATA} \equiv \langle R_{\mu,i} \rangle \langle S_{1000,\mu,i}^{MC} \rangle$. $S_\mu^{MC\text{-}True}$ is known from the initial MOCK-DATA simulations, and $S_\mu^{MOCK\text{-}DATA}$ is the corresponding estimate obtained in the TD procedure. The reconstructed signal is by only 2 to 4% different from the muon signal from the MOCK-DATA set ($S_\mu^{MC\text{-}True} = 23.09$ VEM) – this demonstrates the accuracy of the presented method.

4. Calculating the β exponent

Observations of air showers, and also simulations, demonstrated that the number of muons N_μ grows almost linearly with the shower energy E, and it also increases with a small power of the primary mass A. These relations can be reproduced in the framework of the Heitler-Matthews model of hadronic air showers [13]. This model predicts

$$N_\mu = A \left(\frac{E/A}{\epsilon_c^\pi} \right)^\beta,$$

primary type i	$\langle R_{\mu,i} \rangle$	$\sigma_{\langle R_{\mu,i} \rangle}$	$\langle S^{\text{MC}}_{\mu,i} \rangle$ [VEM]	$\langle R_{\mu,i} \rangle \langle S^{\text{MC}}_{\mu,i} \rangle$ [VEM]	k [%]
p	1.57 ± 0.01	0.27	15.05	23.63	2.3
He	1.42 ± 0.01	0.26	16.82	23.88	3.4
N	1.26 ± 0.01	0.21	18.96	23.89	3.5
Fe	1.14 ± 0.01	0.18	21.08	24.00	3.9

Table 2: The mean values of the muon rescaling parameters $R_{\mu,i}$ calculated from Eq. (4), and their standard deviations $\sigma_{\langle R_{\mu,i} \rangle}$, mean values of the muon ground signal $S^{\text{MC}}_{1000,\mu,i}$ from simulations with QGSJetII-04 model, reconstructed muon ground signal predicted for the MOCK-DATA set, and the ratio $k \equiv (\langle R_{\mu,i} \rangle \langle S^{\text{MC}}_{1000,\mu} \rangle - S^{\text{MC-True}}_{\mu})/S^{\text{MC-True}}_{\mu}$.

where $\beta \approx 0.9$. For any fixed energy it describes how the muon number depends on the primary mass $N_\mu = N_{\mu,\text{p}} A^{1-\beta}$ (the $N_{\mu,\text{p}}$ is the number of muons of a proton shower; ϵ^π_c is the critical energy at which pions decay into muons).

Simulations have shown that β depends on various properties of hadronic interaction (e.g. multiplicity, charge ratio, baryon anti-baryon pair production [15]). Therefore, estimating the β exponent would be helpful in constraining the parameters of hadronic interactions and improving the accuracy of models.

We can assume that the average muon signal S_μ is proportional to muon number N_μ. With this we can calculate the logarithm of the muon number $N_{\mu,i}$ for primary i and iron ($A = 56$; we choose it as a reference), and we get: $\beta_i = 1 - \frac{\ln\langle S^{\text{MC}}_{\mu,\text{Fe}} \rangle - \ln\langle S^{\text{MC}}_{\mu,i} \rangle}{\ln 56 - \ln A_i}$. The β exponent calculated using the muon signals listed in the Table 1 is ~0.92 – this is close to the values reported in Ref [16]: $\beta = 0.927$ for EPOS-LHC and $\beta = 0.925$ for QGSJetII-04, which is another validation of our method. We can also use the reconstructed muon signals for any primary i, $S^{\text{MOCK-DATA}}_{\mu,i} \equiv \langle R_{\mu,i} \rangle \langle S^{\text{MC}}_{1000,\mu,i} \rangle$, to calculate the β exponent. In this case the β_i exponent is given by

$$\beta_i = 1 - \frac{\ln\langle S^{\text{MOCK-DATA}}_{\mu,\text{Fe}} \rangle - \ln\langle S^{\text{MOCK-DATA}}_{\mu,i} \rangle}{\ln 56 - \ln A_i} = 1 - \frac{\ln\langle (R_{\mu,\text{Fe}} \rangle \langle S^{\text{MC}}_{\mu,\text{Fe}} \rangle) - \ln(\langle R_{\mu,i} \rangle \langle S^{\text{MC}}_{\mu,i} \rangle)}{\ln 56 - \ln A_i}.$$

To calculate the β_i exponent for a set of air showers with a predefined composition we perform the following operations. First we generate a new MOCK-DATA set, shown in Fig. 4 (left). This MOCK-DATA set is constructed so that it contains a sample of events with fractions of primary types (f_i) that reflect the composition observed by Auger at 10^{19} eV ($f_\text{p} \simeq 15\%$, $f_\text{He} \simeq 38\%$, $f_\text{N} \simeq 46\%$, $f_\text{Fe} \simeq 1\%$ [17]). Also the distribution of the zenith angles of the events was selected so that it follows the distribution of the TD simulations. The average muon signal for all primaries is also similar (within ±0.5 VEM) to our previous results with the EPOS-LHC model. Using this new MOCK-DATA set we calculate for each of the events the z^{mix} variable, $z^{\text{mix}}_j \equiv S^{\text{MOCK-DATA}}_{1000,j} - \sum_{i\in\{\text{p,He,N,Fe}\}} f_i S^{\text{MC}}_{1000,i,j}$. The distribution of the z^{mix} is shown in Fig. 4 (right). To this z^{mix} histogram we fit the Gaussian function given by

$$P(A, \sigma, \mathbf{R}^{\text{fit}}_\mu) = B \exp(-(z^{\text{mix}} - \langle z^{\text{mix}} \rangle)^2 / 2\sigma^2), \tag{5}$$

$$\langle z^{\text{mix}} \rangle = \sum_{i\in\{\text{p,He,N,Fe}\}} f_i \langle S^{\text{MC}}_{\mu,i} \rangle (R^{\text{fit}}_{\mu,i} - 1). \tag{6}$$
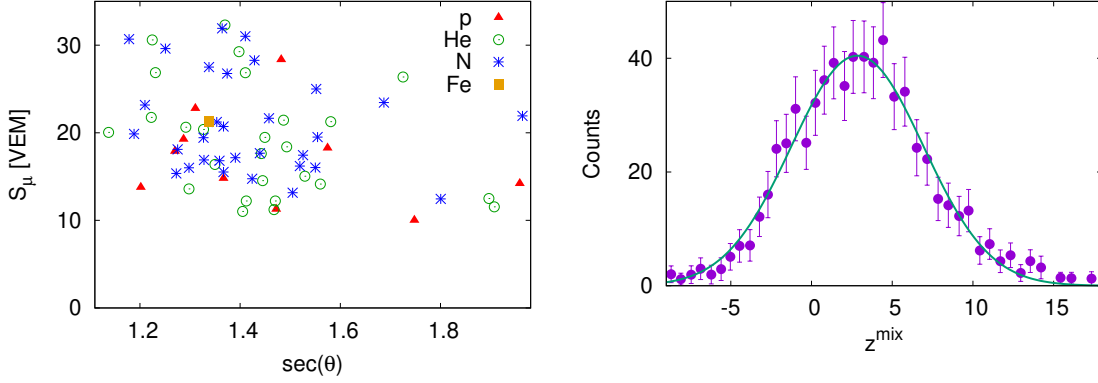
Figure 4: *Left*: the $S_{\mu,i}$ as a function of zenith angle. The distribution of events (68 in total) follows the composition reported by Auger for EPOS-LHC [17]. *Right*: the z^{mix} distribution for MOCK-DATA set. The green line shows an example of a Gaussian fit given by Eq. (5).

The fitting parameters are: the amplitude B, the standard deviation σ and the rescaling parameters $\mathbf{R}_\mu^{\mathrm{fit}} = \{R_{\mu,\mathrm{p}}^{\mathrm{fit}}, R_{\mu,\mathrm{He}}^{\mathrm{fit}}, R_{\mu,\mathrm{N}}^{\mathrm{fit}}, R_{\mu,\mathrm{Fe}}^{\mathrm{fit}}\}$. According to Eq. (3) the average of the muon signal is proportional to the z^{mix}, and $R_{\mu,i}^{\mathrm{fit}} \langle S_{\mu,i}^{\mathrm{MC}} \rangle$ are the contributions from the corresponding primaries i to the total muon signal.

To find the most likely solutions we have fitted the Gaussian function to z^{mix} histogram using the TMinuit routine from the ROOT package. We have done many minimizations with different initial values of the muon scaling parameters – we performed a scan for $1 \leq R_{\mu,i} \leq 2$ with a step of 0.005. Some cuts were applied to select the correct solutions. One condition was $R_{\mu,\mathrm{p}} \langle S_{\mu,\mathrm{p}}^{\mathrm{MC}} \rangle < R_{\mu,\mathrm{He}} \langle S_{\mu,\mathrm{He}}^{\mathrm{MC}} \rangle < R_{\mu,\mathrm{N}} \langle S_{\mu,\mathrm{N}}^{\mathrm{MC}} \rangle < R_{\mu,\mathrm{Fe}} \langle S_{\mu,\mathrm{Fe}}^{\mathrm{MC}} \rangle$ – from the physics of air showers we know that muon numbers should increase with the mass of the primary. Based on the Heitler-Matthews model it is also expected, that logarithm of the muon signal should increase linearly with logarithm of the primary mass, therefore corresponding linearity conditions were introduced:

$$\left| \frac{\ln(R_{\mu,\mathrm{p}} \langle S_{\mu,\mathrm{p}}^{\mathrm{MC}} \rangle) - \ln(R_{\mu,\mathrm{He}} \langle S_{\mu,\mathrm{He}}^{\mathrm{MC}} \rangle)}{\ln(A_{\mathrm{p}}) - \ln(A_{\mathrm{He}})} - \frac{\ln(R_{\mu,\mathrm{He}} \langle S_{\mu,\mathrm{He}}^{\mathrm{MC}} \rangle) - \ln(R_{\mu,\mathrm{N}} \langle S_{\mu,\mathrm{N}}^{\mathrm{MC}} \rangle)}{\ln(A_{\mathrm{He}}) - \ln(A_{\mathrm{N}})} \right| < \epsilon,$$

$$\left| \frac{\ln(R_{\mu,\mathrm{He}} \langle S_{\mu,\mathrm{He}}^{\mathrm{MC}} \rangle) - \ln(R_{\mu,\mathrm{N}} \langle S_{\mu,\mathrm{N}}^{\mathrm{MC}} \rangle)}{\ln(A_{\mathrm{He}}) - \ln(A_{N})} - \frac{\ln(R_{\mu,\mathrm{N}} \langle S_{\mu,\mathrm{N}}^{\mathrm{MC}} \rangle) - \ln(R_{\mu,\mathrm{Fe}} \langle S_{\mu,\mathrm{Fe}}^{\mathrm{MC}} \rangle)}{\ln(A_{\mathrm{N}}) - \ln(A_{\mathrm{Fe}})} \right| < \epsilon.$$

The results are shown in Fig. 5. We see that this method can reproduce the ratio of the muon signals of simulations using EPOS-LHC and QGSJetII-04 within ~5%: the ratio for MC-true is r = 1.11 ± 0.04 and the average reconstructed ratio (left part of Fig. 5) is 1.16 ± 0.02. The difference is caused by the fact that the signal for the MOCK-DATA set is not exactly equal to the one for EPOS-LHC (Table 1). We also recover the β parameter (average $\simeq$ 0.92) for the studied set.

5. Summary

This work presents a new method that can be used to determine the muon scaling factors, and also to estimate the β exponent. This method has been tested with results from TD reconstructions of a set of air showers with the energy about 10^{19} eV. The method allows us to calculate the average muon signal for any set of air showers. Knowing this component is crucial for all analyses of air shower observations, and indirectly can also be used to improve hadronic interaction models.
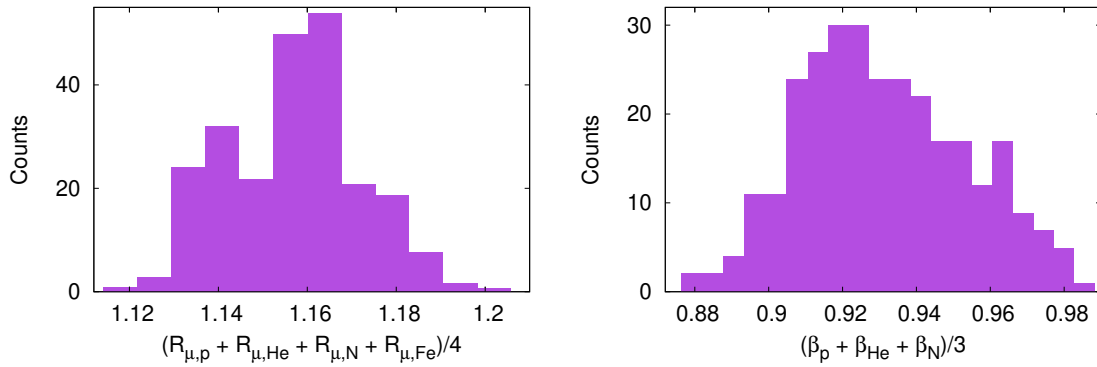
Figure 5: The average muon rescaling factor (left) and the average β_i exponent distribution (right) for events passing the selection cuts are shown.

Acknowledgments

References

[1] A. Aab et al., Phys. Rev. Lett. 117 (2016) 192001.

[2] A. Aab et al., Phys. Rev. D 90 (2014) 122005.

[3] A. Aab et al., Nucl. Instrum. Meth. A 798 (2015) 172.

[4] H.J. Drescher and G. Farrar, Phys. Rev. D 67 (2003) 116001; J.Allen et al., [arXiv:0708.2892].

[5] D. Heck et al., Report FZKA 6019, Forschungszentrum Karlsruhe (1998).

[6] Cz. Porowski, PhD thesis, IFJ PAN repository: https://rifj.ifj.edu.pl/handle/item/289

[7] D. Góra et al., PoS(ICRC2021) (2021) 207.

[8] S. Argirò et al., Nucl. Instrum. and Meth. in Phys. A 580 (2007) 1485.

[9] T. Pierog et al., Phys. Rev. C 92 (2015) 034906.

[10] S. Ostapchenko et al., Phys. Rev. D 83 (2011) 014018.

[11] B. Kegl et al., ICRC-2013 [arXiv:1307.5059].

[12] M. Ave, R. Engel, M. Roth, and A. Schulz, Astropart. Phys. 87 (2017) 23.

[13] J. Matthews, Astropart. Phys. 22 (2005) 387.

[14] T. Pierog, Eur. Phys. J. Web Conf. 52 (2013) 03001.

[15] R. Ulrich et al., PRD 83 (2011) 054026.

[16] L. Cazon et al., Phys. Lett. B 784 (2018) 68.

[17] J. Bellido et al., PoS(ICRC2017) (2017) 506.