Physica Scripta

PAPER • OPEN ACCESS

Calculation of rescaling factors and nuclear multiplication of muons in extensive air showers

To cite this article: Kevin Almeida Cheminant *et al* 2023 *Phys. Scr.* **98** 045301

IOP Publishing Phys. Scr. 98 (2023) 045301 https://doi.org/10.1088/1402-4896/acbca9

Physica Scripta

OPEN ACCESS

RECEIVED
3 November 2022

REVISED
13 January 2023

ACCEPTED FOR PUBLICATION
16 February 2023

PUBLISHED
3 March 2023

PAPER

Calculation of rescaling factors and nuclear multiplication of muons in extensive air showers

Kevin Almeida Cheminant[1], Dariusz Góra[1], Nataliia Borodai[1], Ralph Engel[2], Tanguy Pierog[2], Jan Pękala[1], Markus Roth[2], Michael Unger[2], Darko Veberič[2] and Henryk Wilczyński[1]

[1] Institute of Nuclear Physics PAS, Radzikowskiego 152, Krakow, Poland
[2] Karlsruhe Institute of Technology, Institute for Astroparticle Physics, Karlsruhe, Germany

E-mail: dariusz.gora@ifj.edu.pl

Keywords: rescaling, factors, nuclear, multiplication, muons

Abstract

Recent results obtained from leading cosmic ray experiments indicate that simulations using LHC-tuned hadronic interaction models underestimate the number of muons in extensive air showers compared to experimental data. This is the so-called muon deficit problem. Determination of the muon component in the air shower is crucial for inferring the mass of the primary particle, which is a key ingredient in the efforts to pinpoint the sources of ultra-high energy cosmic rays. In this paper, we present a new method to derive the muon signal in detectors, which uses the difference between the total reconstructed ('data') and simulated signals, and is in turn related to the muon signal which is roughly independent of the zenith angle, but depends on the mass of the primary cosmic ray. Such a method offers an opportunity not only to test/calibrate the hadronic interaction models, but also to derive the β exponent, which describes an increase of the number of muons in a shower as a function of the energy and mass of the primary cosmic ray. 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. We validate the method by using Monte-Carlo simulations for the EPOS-LHC and QGSJetII-04 hadronic interaction models, and show that this method allows us to recover the ratio of the muon signal between EPOS-LHC and QGSJetII-04 and the average β exponent for the studied system, within less than a few percent. This is a consequence of the good recovery of the muon signal for each primary included in the analysis.

1. Introduction

Discovered at the beginning of the 20th century by Victor F Hess, cosmic rays are protons and atomic nuclei that constantly bombard Earth's atmosphere. Before arriving to the surface of Earth they first interact with the nuclei of the atmosphere to produce cascades of secondary particles that may develop all the way to the ground. This physical phenomenon, also called extensive air shower (EAS), can be detected via multiple channels of observations, e.g. Cherenkov and fluorescence light, or radio emission, which can measure different physical quantities that can be used to determine the nature of the primary cosmic ray, its arrival direction, and its energy.

In the initial phase of the cascading process, the number of particles increases while the energy per particle drops and distinct components emerge, namely the hadronic, electromagnetic, and muonic components. Such growth carries on until a maximum is reached, at a traversed depth usually referred to as $X_{\max}$, as particles below a certain energy threshold are no longer capable of producing additional particles but decay instead as atmospheric absorption processes start taking over. As many as 10^6 to 10^9 secondary particles may reach the ground over an area that can extend up to several square kilometers.

In order to describe how EAS are formed in the atmosphere, simple toy models, such as the one described by Heitler and Matthews [1], have been developed and are capable of providing accurate predictions of some of the main quantities that characterize air showers without the need for high-performance computing. Although simplistic, the Heitler-Matthews model is powerful enough to allow the discrimination of EAS produced by

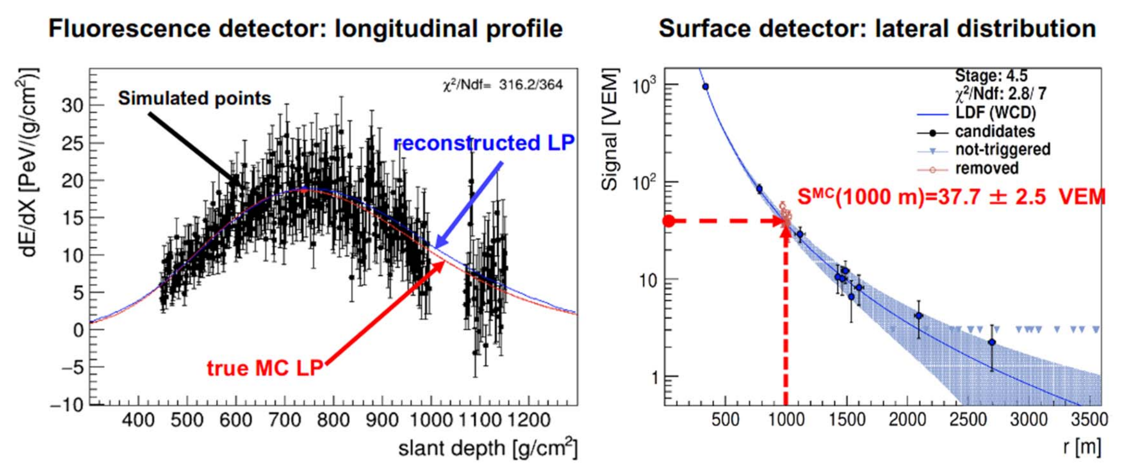

Figure 1. Example of TD—simulated event. **Left**: The measured longitudinal profile (LP) of an illustrative air shower with the matching simulated showers, using QGSJetII-04 for proton with the energy 10^{19} eV. The longitudinal profile for true simulated event (true MC LP) and reconstructed LP are also shown. **Right**: The simulated ground signals for the same event. The signal at 1000 m (S_{1000}) in VEM is also shown. 1 VEM corresponds to the most-likely signal deposited by a muon that traverses vertically the center of the SD station.

protons/nuclei and photons. The number of muons N_μ^A in EAS initiated by a nucleus with mass number A can be related to the number of muons $N_\mu^{\rm p}$ produced in a shower initiated by a proton with the same energy through $N_\mu^A = N_\mu^{\rm p}\, A^{1-\beta}$, where $1 - \beta \simeq 0.1$. Muons in EAS have also large decay lengths and small radiative energy losses and are produced at different stages of the shower development. Therefore, muons can reach surface and underground detector arrays while keeping relevant information about the hadronic cascade.

In recent years, our understanding of the nature of cosmic rays has significantly improved thanks to experiments spread out all over the world and using different methods of detections such as gamma-ray telescopes (H.E.S.S. [2], MAGIC [3],VERITAS [4], HAWC [5], and others) and cosmic-ray observatories (Telescope Array [6], Pierre Auger Observatory [7]). As of today, the energy spectrum of cosmic rays has been measured from a few GeV (giga electron-Volts, 10^9eV) up to 100 EeV (100 exa electron-Volts, 10^{20} eV), well beyond the energy accessible in terrestrial particle accelerators, and falls rapidly as the energy increases. The general consensus is that cosmic rays below 10^{17} to 10^{18} eV are of Galactic origin, most likely from supernovae, while particles above this energy range have their origin in extra-galactic sources, with active galactic nuclei and starburst galaxies being the most plausible candidates.

Simulations of EAS using current hadronic interaction models predict fewer muons than observed in real events, which is known as the muon deficit problem [8]. As an example, data from the Pierre Auger Observatory indicate that the muon number predicted by the LHC-tuned models, such as EPOS-LHC [9] and QGSJetII-04 [10], is 30% to 60% lower than what is observed in showers with an energy of 10^{19}eV [11]. The muon excess over predictions seen by the Pierre Auger Collaboration is/was also seen in several other experiments like HiRes/MIA [12], NEVOD-DECOR [13], SUGAR array [14], Telescope Array [15]. However, experiments like KASCADE-Grande [16] and EAS-MSU [17] reported no discrepancy in the muon number around 10^{17}eV. In [18], after cross-calibration of the energy scales, the observed muon densities were scaled by using the so-called z-scale and compared to expectations from different hadronic models, also for data from IceCube [19] and AMIGA [20]. While such densities were found to be consistent with simulations up to 10^{16}eV, at higher energies the muon deficit increases in several experiments [18]. Since data interpretation relies on simulations, the muon deficit problem has deep implications: the data suggest a much heavier composition of cosmic rays based on muons only than the composition derived from $X_{\max}$ measurements [21].

To study the muon-number problem, a top-down (TD) reconstruction method was proposed by the Pierre Auger Collaboration [11] for the so-called hybrid events (events seen simultaneously by the array of particle detectors (SD) and by the fluorescence detector (FD)). The main aim of the TD reconstruction is to predict signals in the FD and SD on a simulation basis. In the TD method, one finds a simulated shower, which has a distribution of electromagnetic component along the shower axis (longitudinal profile, reconstructed LP) most similar to the observed longitudinal profile of the shower (i.e. reference profile (true MC LP)—see left panel of figure 1). The reference longitudinal profile is linked to the electromagnetic component of the shower, so the method relies on the fact that this component is accurately simulated.

As an output, the TD method provides a reconstructed event, in which the signal in the SD is determined using Monte-Carlo (MC) simulations. The simulated SD signals in the output shower, which depend on the interaction model, may then be compared with the data/initial shower. The SD signal includes the contribution

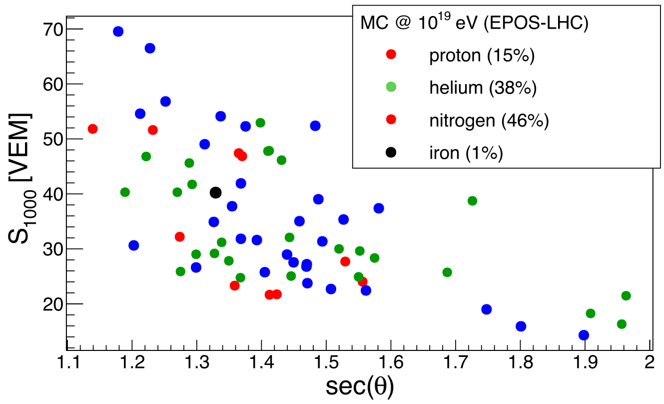

Figure 2. Total signal S_{1000} of the mock dataset shown with the proportion of primaries discussed in the text.

of muons, which are tracers of properties of the hadronic interactions. A comparison of the simulated SD signal with the corresponding signal in the data shower provides an opportunity to check the correctness of the lateral distribution of the simulated showers (see right panel of figure 1). Since the lateral distribution is sensitive to the hadronic interaction models used, an analysis of this distribution provides an opportunity to investigate indirectly the interaction models at energies much above the maximum energies provided by terrestrial accelerators. It is therefore expected that the TD method should allow us to calibrate the interaction models, and to reduce the discrepancy between the data and simulations.

The TD analysis performed in this work is similar to the one presented in [22, 23] and is based on the analysis found in [11]. In this work the TD chain includes a simulation of the SD response for the CORSIKA [24] simulated event[3]—the reference shower. The Pierre Auger Observatory response for the reference shower is simulated in the hybrid mode—the event is seen by SD and FD—using the Offline software [25] which provides 10 detector simulations for comparison of the station signals with the reference MC event.

Here we also try to reproduce as accurately as possible the real data from the Pierre Auger Observatory by creating a mock data set of mixed composition from MC simulations obtained with the EPOS-LHC hadronic model, at 10^{19} eV. MC simulations at the same energy produced using the QGSJETII-04 model are then used to try to recover the muon signal from this mock data set, by calculating the muon scaling factors (relative to EPOS-LHC) for the primaries considered in this dataset.

In this paper, we present a validation test of the method for determining muon scaling factors by analysing reconstructions of a simulated hybrid shower, where mock-data showers are used as reference events. This method is based on the z-variable, which is the difference between the initially simulated and the reconstructed total signal at the detectors, 1000 meters away from the shower axis, and which is related to the muon signal. This variable is approximately independent of the zenith angle, but depends on the mass of the primary cosmic ray. We show that we can recover the ratio of the muon signal between EPOS-LHC and QGSJetII-04, on average, within less than 6%, and the average β exponent which governs the number of muons in simulated air showers [1], within less than 1%, which is a consequence of the good recovery of the muon signal for each primary.

2. Preparation of the mock dataset and QGSJetII-04 Monte-Carlo simulations

The mock data set of EPOS-LHC is built based on simulated CORSIKA events. CORSIKA simulations are performed for four potential primaries: proton, helium, nitrogen, and iron. If we restrict ourselves to shower energies $10^{18.8} < E < 10^{19.2}$ eV to zenith angles θ below 60°, we get 68 reference events (typical number of high-energy, high-quality events observed over several years by the Pierre Auger Observatory in a narrow energy interval). The total signal S_{1000} for each reference event as a function of sec θ is shown in figure 2. As expected, due to the attenuation of EAS in the atmosphere, we see that the signal depends on the zenith angle.

To calculate the muon signal for each MC event, we use the universal parametrization of muon fraction in the signal $g_{\mu,i}(\theta)$ presented in [23, 26]. This parametrization gives the muon signal for a primary i and a zenith angle θ such that $S_{\mu,i}^{\rm mock}(\theta) = g_{\mu,i}(\theta)\, S_{1000,i}^{\rm mock}(\theta)$.

In order to create the mock dataset of mixed composition, we consider the fractions of primaries f_i for proton, helium, nitrogen, and iron as measured by the Pierre Auger Observatory at 10^{19} eV from EPOS-LHC, as presented in [27]. These fractions are roughly estimated to be around 15%, 38%, 46%, and 1% for proton,

[3] We use one of the latest version of CORSIKA, i.e. version 7560.

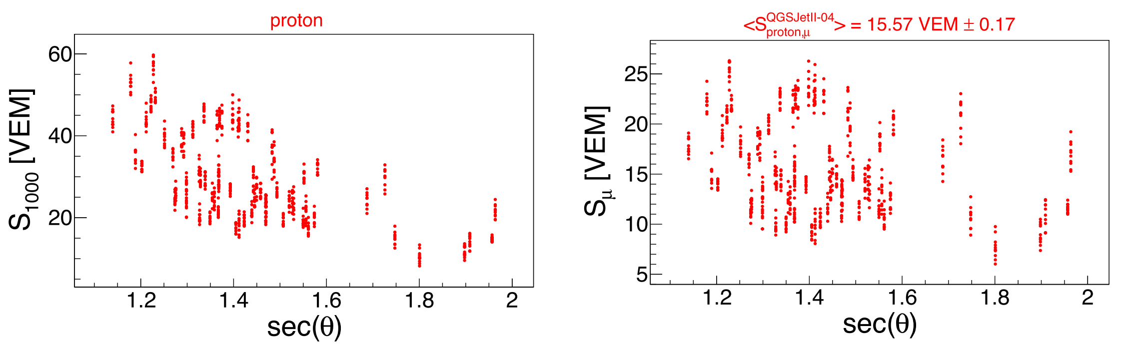

Figure 3. Left: Total signal S_{1000} in events simulated with QGSJetII-04 for proton primary. The average total signal for proton is (29.66 ± 0.41) VEM for the studied range of zenith angle, i.e. less than $60°$. **Right:** Muon signal in events simulated with QGSJetII-04 for proton primary. The average muon signal for proton is (15.45 ± 0.17) VEM.

helium, nitrogen, and iron, respectively. In this work 68 reference CORSIKA events were selected from EPOS-LHC simulations, taking into account these primary fractions, which roughly translates to 10 proton events, 26 helium events, 31 nitrogen events, and 1 iron event. The average muon signal for the mock dataset is 17.30 ± 0.25, 19.03 ± 0.30, 21.12 ± 0.38, and (23.42 ± 0.25) VEM for proton, helium, nitrogen, and iron primaries, respectively.

In this TD analysis, we consider air shower simulations obtained with QGSJetII-04 model as our MC sample. It is used to reconstruct the muon signal found in the mock dataset described in the previous section. 10 MC showers are associated to each shower from the mock dataset with a given zenith angle, for each primary. As an example, the distributions of $S^{\mathrm{MC}}_{1000,ij}$ and $S^{\mathrm{MC}}_{\mu,ij} = g_{\mu,i}(\theta) S^{\mathrm{MC}}_{1000,ij}$ as a function of $\sec\theta$ for QGSJetII-04 simulations of proton is shown in figure 3 (left) and figure 3 (right), respectively. In this case, the average muon signal for each primary is 15.57 ± 0.17, 17.25 ± 0.19, 19.37 ± 0.20, and (21.62 ± 0.23) VEM. One can observe that the average muon signal for each primary is larger for EPOS-LHC simulations than for QGSJetII-04. Thus the mean ratio averaged over the four primaries studied $r^{\mathrm{MC}}_{\mathrm{true}} = \langle \langle S^{\mathrm{EPOS}}_{\mu} \rangle / \langle S^{\mathrm{QGS}}_{\mu} \rangle \rangle$ is approximately equal to 1.10 ± 0.04. It is worth noting that in this way, i.e. using EPOS-LHC as mock data set and MC simulation from QGSJetII-04, we also mimic the muon problem seen in the real data.

3. Muon scaling factor

The observed SD signal of ultra-high energy air showers is significantly larger than predicted by hadronic models tuned to fit the accelerator data [11]. Such a disagreement can be corrected for by introducing linear scaling factors, for the electromagnetic part, R_{EM}, and the hadronic/muonic part, R_μ. Following this approach for a single shower j, the simulated ground signal at 1000 m from QGSJetII-04 MC and the mock dataset can be written as

$$S^{\mathrm{MC}}_{1000,j} \equiv S^{\mathrm{MC}}_{\mathrm{EM},j} + S^{\mathrm{MC}}_{\mu,j}, \tag{1}$$

$$S^{\mathrm{mock}}_{1000,j}(R_{\mathrm{EM}}, R_\mu) \equiv S^{\mathrm{mock}}_{\mathrm{EM},j} + S^{\mathrm{mock}}_{\mu,j} = R_{\mathrm{EM}}\, S^{\mathrm{MC}}_{\mathrm{EM},j} + R_{\mu,j}\, R^{\alpha}_{\mathrm{EM}}\, S^{\mathrm{MC}}_{\mu,j}. \tag{2}$$

In above equation (2) we have used the fact that some of the electromagnetic particles produced by muons in decay or radiation processes, as well as by low-energy π^0s, can be attributed to the electromagnetic signal by introducing an additional factor R^{α}_{EM}; but the muons that result from photoproduction are assigned to the electromagnetic signal, S_{EM}. As shown in [11], no rescaling is needed for the electromagnetic part, where the most likely solution is $R_{\mathrm{EM}} = 1$. Furthermore, in the TD method the reference longitudinal profile is related to the electromagnetic component of the shower, so the method ensures that this part is accurately simulated. Hence the assumption $R_{\mathrm{EM}} = 1$ used in this analysis.

In this work, we use the difference between the mock dataset and the MC ground signal as the main observable, i.e. $z_j \equiv S^{\mathrm{mock}}_{1000,j} - S^{\mathrm{MC}}_{1000,j}$, because this variable is a natural indicator of the discrepancy between data and MC. Moreover, the discrepancy should ideally be zero. Another interesting feature arises from equations (1) and (2). For $R_{\mathrm{EM}} = 1$, by simple subtraction we obtain

$$S^{\mathrm{MC}}_{\mu,ij} = \frac{z_{ij}}{R_{\mu,ij} - 1}, \tag{3}$$

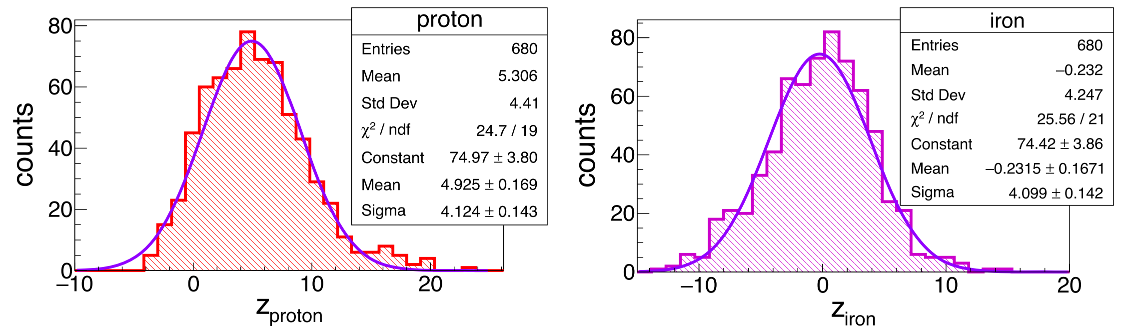

Figure 4. z_i distributions for proton (left) and iron primary (right) based on mock data (EPOS-LHC) and QGSJetII-04 simulations.

for a primary i and an event j. This is the key equation for the method presented in this paper. The formula shows that the muon MC signal is proportional to the difference between data and MC signal, i.e. variable z_{ij}, where a proportionality coefficient depends on the muon scaling factor $R_{\mu,ij}$.

Equation (3) can also be rewritten as

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

Using the mock data set previously built and the QGSJetII-04 MC simulations, the z_{ij} distributions are obtained and shown in figure 4.

The total average muon signal of the mock data set $\langle S_{\mu,\mathrm{tot}}^{\mathrm{mock}}\rangle$ can be expressed as

$$\langle S_{\mu,\mathrm{tot}}^{\mathrm{mock}}\rangle = \frac{1}{N_{\mathrm{tot}}}\sum_i \sum_j^{N_i} S_{\mu,ij}^{\mathrm{mock}}, \tag{5}$$

by summing over all primaries $i \in \{\mathrm{p, He, N, Fe}\}$ and where $N_{\mathrm{p}} = 10$, $N_{\mathrm{He}} = 26$, $N_{\mathrm{N}} = 31$, and $N_{\mathrm{Fe}} = 1$ are the number of proton, helium, nitrogen, and iron events, respectively, that have been used to create the mock data set, and $N_{\mathrm{tot}} = N_{\mathrm{p}} + N_{\mathrm{He}} + N_{\mathrm{N}} + N_{\mathrm{Fe}}$. Equation (5) can be expressed as

$$\langle S_{\mu,\mathrm{tot}}^{\mathrm{mock}}\rangle = \frac{1}{N_{\mathrm{tot}}}\sum_i N_i \langle S_{\mu,i}^{\mathrm{mock}}\rangle, \tag{6}$$

where $\langle S_{\mu,i}^{\mathrm{mock}}\rangle = \frac{1}{N_i}\sum_j^{N_i} S_{\mu,ij}^{\mathrm{mock}}$ is the average over N_i events. Since N_i/N_{tot} is simply the fraction f_i, $\langle S_{\mu,\mathrm{tot}}^{\mathrm{mock}}\rangle$ can be rewritten as

$$\langle S_{\mu,\mathrm{tot}}^{\mathrm{mock}}\rangle = \sum_i f_i\, \langle S_{\mu,i}^{\mathrm{mock}}\rangle. \tag{7}$$

Using the values of $\langle S_{\mu,i}^{\mathrm{mock}}\rangle$ given in section 2, we obtain $\langle S_{\mu,\mathrm{tot}}^{\mathrm{mock}}\rangle = (19.78 \pm 0.22)$ VEM. We can then rescale the MC data set to retrieve $\langle S_{\mu,\mathrm{tot}}^{\mathrm{mock}}\rangle$,

$$\langle S_{\mu,\mathrm{tot}}^{\mathrm{mock}}\rangle \equiv \langle R_{\mu,i}\rangle\, \langle S_{\mu,i}^{\mathrm{MC}}\rangle. \tag{8}$$

where $\langle S_{\mu,i}^{\mathrm{MC}}\rangle = \frac{1}{N_i}\sum_j^{N_i} S_{\mu,ij}^{\mathrm{MC}}$ and $\langle R_{\mu,i}\rangle = \frac{1}{N_i}\sum_j^{N_i} R_{\mu,ij}$ are the averages over N_i events. In other words, the average total muon signal in mock data corresponds to the average muon signal obtained from MC simulations for a given primary, multiplied by the average muon scaling factor obtained for that primary. The values obtained for the right-hand side of equation (8) are also reported in table 1, along with a summary of the results presented until now. The accuracy with which this equivalence is obtained can be calculated through the ratio

$$k \equiv \frac{\langle R_{\mu,i}\rangle\, \langle S_{\mu,i}^{\mathrm{MC}}\rangle - \sum_i f_i\, \langle S_{\mu,i}^{\mathrm{mock}}\rangle}{\sum_i f_i\, \langle S_{\mu,i}^{\mathrm{mock}}\rangle}. \tag{9}$$

Values of k are reported in the last column of table 1. This method allows us to recover the average muon signal of the mock dataset within ~9%. The results shown in table 1 are also a validation of the MC simulation for each primary, as we can recover the total muon signal for each primary.

4. Calculation of the β exponent

The number of muons in an air shower is a powerful tracer of the mass of the primary particle. Simulations and measurements have confirmed that the number N_μ of muons produced rises almost linearly with the primary

Table 1. Mean value of the muon rescaling parameters $R_{\mu,i}$ for different primaries i. Also, the corresponding mean values of the total muon signal $S_{\mu}^{\rm MC}$ from QGSJetII-4 model (MC), reconstructed muon signal at 1000m expected in the mock data set $\langle R_{\mu,i}\rangle\ \langle S_{\mu,i}^{\rm MC}\rangle$ and the ratio k are listed. The errors shown in the fourth column are the maximum error calculated from $\langle R_{\mu,i}\rangle\ \delta S_{\mu}^{\rm MC} + \langle S_{\mu}^{\rm MC}\rangle\ \delta R_{\mu,i}$, where $\delta R_{\mu,i}$ and $\delta S_{\mu}^{\rm MC}$ are the errors listed in the second and third column, respectively.

i	$\langle R_{\mu,i}\rangle$	$\langle S_{\mu,i}^{\rm MC}\rangle$	$\langle R_{\mu,i}\rangle\ \langle S_{\mu,i}^{\rm MC}\rangle$	k
p	1.35 ± 0.02	15.57 ± 0.17	21.02 ± 0.54	6%
He	1.24 ± 0.01	17.25 ± 0.19	21.39 ± 0.41	8%
N	1.11 ± 0.01	19.37 ± 0.20	21.50 ± 0.41	9%
Fe	1.00 ± 0.01	21.62 ± 0.23	21.62 ± 0.44	9%

energy E, and increases with a small power of the cosmic-ray mass A. This behavior can be understood in terms of the Heitler-Matthews model of hadronic air showers [1], which predicts $N_{\mu}^{A} = A(E/A\epsilon_{\rm c}^{\pi})^{\beta} = N_{\mu}^{\rm p}\ A^{1-\beta}$, with $\beta \simeq 0.92$ [4]. Detailed simulations of β show multiple dependencies on hadronic-interaction properties, like multiplicity, charge ratio and baryon anti-baryon pair production [28]. Thus, measurements of the β exponent can effectively constrain the parameters governing hadronic interactions and improve the accuracy of hadronic models. Assuming that the average reconstructed muon signal $\langle S_{\mu,i}^{\rm rec}\rangle$ (see below) is proportional to N_{μ} and calculating the average logarithm of the muon number $N_{\mu,i}$ for primary i and iron (A= 56), we get the expression of β given by

$$\beta_i = 1 - \frac{\ln\langle S_{\mu,\rm Fe}^{\rm rec}\rangle - \ln\langle S_{\mu,i}^{\rm rec}\rangle}{\ln A_{\rm Fe} - \ln A_i}, \tag{10}$$

where A_i is the mass number of all considered primaries i (except iron).

However, the β exponent can also be calculated using the reconstructed muon signal for each primary i, e.g. $S_{\mu,i}^{\rm rec} \equiv r_{\mu,i}\ \langle S_{\mu,i}^{\rm MC}\rangle$. Ideally, we should have $S_{\mu,i}^{\rm rec} = S_{\mu,i}^{\rm mock}$. *Here, the definition of the rescaling factor* $r_{\mu,i}$ *is slightly different from the one discussed in the previous section as it rather corresponds to the weight needed to be applied to the MC muon signal of each primary in order to recover the muon signal from the mock data set.* Here $S_{\mu,i}^{\rm rec}$ is by definition the contribution of the signal for each primary i to the total muon signal. In this case, the exponent β_i can be given by

$$\beta_i = 1 - \frac{\ln(r_{\mu,\rm Fe}\ \langle S_{\mu,\rm Fe}^{\rm MC}\rangle) - \ln(r_{\mu,i}\ \langle S_{\mu,i}^{\rm MC}\rangle)}{\ln A_{\rm Fe} - \ln A_i}. \tag{11}$$

In the following, we show how to compute the β_i exponent for a set of hybrid events that consists of certain fractions of events with different primaries. The total signal for the mock and the MC datasets can be expressed as

$$\langle S_{1000}^{\rm mock}\rangle = \sum_i f_i\ \langle S_{1000,i}^{\rm mock}\rangle = \sum_i f_i(\langle S_{{\rm EM},i}^{\rm mock}\rangle + \langle S_{\mu,i}^{\rm mock}\rangle), \tag{12}$$

$$\langle S_{1000}^{\rm MC}\rangle = \sum_i f_i\ \langle S_{1000,i}^{\rm MC}\rangle = \sum_i f_i(\langle S_{{\rm EM},i}^{\rm MC}\rangle + \langle S_{\mu,i}^{\rm MC}\rangle). \tag{13}$$

Again, assuming that in TD-simulations, the electromagnetic component is correctly reproduced, i.e. the scaling factor for electromagnetic part is $R_{{\rm EM},i} = 1$, we can define the overall $z^{\rm mix}$ variable as

$$\langle z^{\rm mix}\rangle = \langle S_{1000}^{\rm mock}\rangle - \langle S_{1000}^{\rm MC}\rangle = \sum_i f_i(\langle S_{\mu,i}^{\rm mock}\rangle - \langle S_{\mu,i}^{\rm MC}\rangle), \tag{14}$$

$$\langle z^{\rm mix}\rangle = \sum_i f_i\ \langle S_{\mu,i}^{\rm MC}\rangle\ (r_{\mu,i} - 1). \tag{15}$$

Therefore, for a single event, we can calculate the $z^{\rm mix}$ variable defined as

$$z_j^{\rm mix} \equiv S_{1000,j}^{\rm mock} - \sum_i f_i\ S_{1000,ij}^{\rm MC}. \tag{16}$$

We consider the same primary fractions for the MC dataset as for the ones used to generate the mock data set, i.e. $f_{\rm p} = 0.15, f_{\rm He} = 0.38, f_{\rm N} = 0.46$, and $f_{\rm Fe} = 0.01$. The distribution of $z^{\rm mix}$ variable is shown in figure 5 (left). The $z^{\rm mix}$ histogram can be fitted with a Gaussian function described by

[4] The $N_{\mu}^{\rm p}$ is the number of muons for proton shower and $\epsilon_{\rm c}^{\pi}$ is the critical energy at which pions decay into muons.

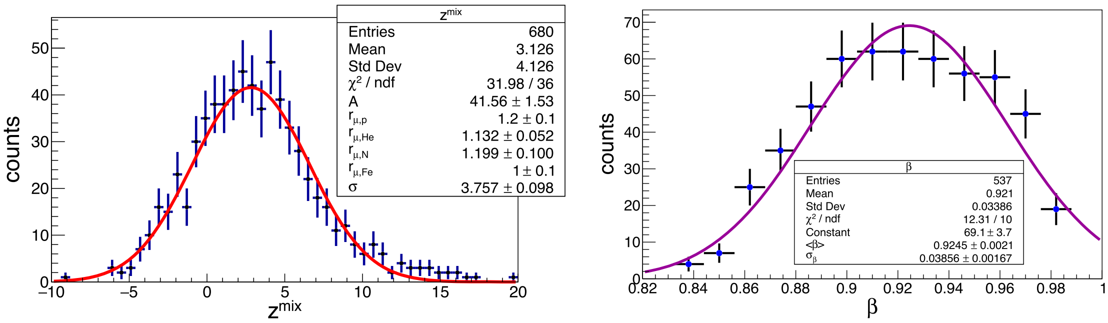

Figure 5. Left: z^{mix} distribution as described by equation (16) with $f_\text{p} = 0.15, f_\text{He} = 0.38, f_\text{N} = 0.46$, and $f_\text{Fe} = 0.01$. The distribution is fitted with the function described by equation (17), with an example of a possible set of fitting parameters $\{A, \sigma, \mathbf{r}_\mu\}$. **Right:** Distribution of the average β parameter as described by equation (10). The histogram is fitted with a Gaussian function with a mean $\langle\beta\rangle$ and a standard deviation σ_β represented by the solid purple line.

Table 2. Mean values of the muon rescaling factors obtained with the fitting procedure, and of the MC muon signal, the reconstructed and the mock dataset muon signals, for all primaries considered and with $f_\text{p} = 0.15, f_\text{He} = 0.38, f_\text{N} = 0.46$, and $f_\text{Fe} = 0.01$. The overestimation $\delta = (\langle S^{\text{rec}}_{\mu,i}\rangle - \langle S^{\text{mock}}_{\mu,i}\rangle)/\langle S^{\text{mock}}_{\mu,i}\rangle$ of the reconstructed muon signal compared to the one from the mock data set is also provided.

i	$\langle r_{\mu,i}\rangle$	$\langle S^{\text{MC}}_{\mu,i}\rangle$/VEM	$\langle S^{\text{rec}}_{\mu,i}\rangle$/VEM	$\langle S^{\text{mock}}_{\mu,i}\rangle$/VEM	δ
p	1.142 ± 0.004	15.57 ± 0.17	17.78 ± 0.25	17.30 ± 0.25	2.7%
He	1.167 ± 0.001	17.25 ± 0.19	20.13 ± 0.24	19.03 ± 0.30	5.8%
N	1.153 ± 0.001	19.37 ± 0.20	22.33 ± 0.25	21.12 ± 0.38	5.7%
Fe	1.148 ± 0.004	21.62 ± 0.23	24.82 ± 0.35	23.42 ± 0.25	6.0%

$$P(A,\ \sigma,\ \mathbf{r}_\mu) = A\exp\left[-\frac{(z^{\text{mix}}(\mathbf{r}_\mu) - \langle z^{\text{mix}}(\mathbf{r}_\mu)\rangle)^2}{2\sigma^2}\right], \tag{17}$$

where fitting parameters are the amplitude A, the standard deviation σ and four rescaling parameters $\mathbf{r}_\mu = \{r_{\mu,\text{p}}, r_{\mu,\text{He}}, r_{\mu,\text{N}}, r_{\mu,\text{Fe}}\}$. Note that following equation (15), the mean of the total muon signal will be proportional to the $\langle z^{\text{mix}}\rangle$, and *the factor* $r_{\mu,i}\ \langle S^{\text{MC}}_{\mu,i}\rangle$ *is by definition the contribution of the primary i to the total muon signal*. The CERN ROOT [29] routine Minuit [30] used to fit the histogram requires all these parameters to have initial values when using a user-defined function with multiple parameters. Multiple fits are therefore performed with $r_{\mu,i}$ between 1 and 2 with steps of 0.025. In this example, correct fits are selected based on physical conditions such as

$$r_{\mu,\text{p}}\ \langle S^{\text{MC}}_{\mu,\text{p}}\rangle < r_{\mu,\text{He}}\ \langle S^{\text{MC}}_{\mu,\text{He}}\rangle < r_{\mu,\text{N}}\ \langle S^{\text{MC}}_{\mu,\text{N}}\rangle < r_{\mu,\text{Fe}}\ \langle S^{\text{MC}}_{\mu,\text{Fe}}\rangle, \tag{18}$$

which simply underlines the fact that the reconstructed muon signal should be larger as the primary gets heavier, and the linearity condition such that

$$\left|\frac{\ln(r_{\mu,\text{p}}\ \langle S^{\text{MC}}_{\mu,\text{p}}\rangle) - \ln(r_{\mu,\text{He}}\ \langle S^{\text{MC}}_{\mu,\text{He}}\rangle)}{\ln A_\text{p} - \ln A_\text{He})} - \frac{\ln(r_{\mu,\text{He}}\ \langle S^{\text{MC}}_{\mu,\text{He}}\rangle) - \ln(r_{\mu,\text{N}}\ \langle S^{\text{MC}}_{\mu,\text{N}}\rangle)}{\ln A_\text{He} - \ln A_\text{N}}\right| < \epsilon \quad \text{and} \tag{19}$$

$$\left|\frac{\ln(r_{\mu,\text{He}}\ \langle S^{\text{MC}}_{\mu,\text{He}}\rangle) - \ln(r_{\mu,\text{N}}\ \langle S^{\text{MC}}_{\mu,\text{N}}\rangle)}{\ln A_\text{He} - \ln A_\text{N}} - \frac{\ln(r_{\mu,\text{N}}\ \langle S^{\text{MC}}_{\mu,\text{N}}\rangle) - \ln(r_{\mu,\text{Fe}}\ \langle S^{\text{MC}}_{\mu,\text{Fe}}\rangle)}{\ln A_\text{N} - \ln A_\text{Fe}}\right| < \epsilon, \tag{20}$$

where $\epsilon = 0.10$ is a tolerance of the non-linearity.

The conditions described by equations (18)–(20) are a consequence of the Heitler-Mathews model, which predicts the linear dependence of the muon signal as a function of the logarithm of the primary mass. The mean values of $\langle r_{\mu,i}\rangle$ distributions, which are reported in table 2 fall within the uncertainties of the mean true rescaling value $r^{\text{MC}}_{\text{true}}$ calculated in section 2, i.e. 1.10 ± 0.04. The uncertainties of the mean of proton and iron primaries are suspected to stem from the fact that our mock data set contains small numbers of proton and iron events, therefore increasing the uncertainties on the fitting procedure. All the values are reported in table 2. With the method proposed in this note, the reconstructed muon signal is overestimated by less than 6% compared to the muon signal from the mock data set, for all primaries. Using values of $\langle S^{\text{rec}}_{\mu,i}\rangle$ given in table 2, we obtain the total reconstructed muon signal $\langle S^{\text{rec}}_\mu\rangle = (20.84 \pm 0.24)$ VEM, which differs by approximately +5% from MC true one, $\langle S^{\text{mock}}_\mu\rangle = (19.78 \pm 0.22)$ VEM.

For each fit fulfilling the conditions described above, we can calculate the averaged $\beta = \frac{1}{3}\sum_i \beta_i$. The β distribution is shown in figure 5 (right). The mean of that distribution, 0.924 ± 0.002, is very close to the EPOS-LHC true value of 0.927 ± 0.003 (within ~1%) [31], therefore supporting the effectiveness of the method to estimate the β parameter governing the number of muons in hadronic showers.

5. Summary and conclusion

The muon problem currently is one of the hot topics in cosmic ray physics, and for a few years some attempts have been made to solve it, but up to now it has not been explained fully. This is because of the inaccessibility of certain phase space regions, which are important for the typical energies of EAS, to accelerator experiments. Exploiting ultra-high energy cosmic rays data, we reach center-of-mas energies up to 400 TeV i.e. more than 30 times of those attainable at the Large Hadron Collider (LHC) [32]. Thus, an extrapolation of hadronic interaction properties to higher energies is necessary, contributing to systematic uncertainties of the final results. On the other hand, even in the simple Mathews-Heitler model, increasing the hadronic energy fraction of interactions by about 5% per generation, can lead to about 30% change in the number of muons after 6 cascade generations. The formation of a Strange Fireball [33], String Percolation [34], Chiral Symmetry Restoration [35], increasing the inelastic cross section [36], or for instance resorting to Lorentz Invariance Violation [37] could also explain the muon excess seen in EAS.

The method described in this paper allows us to recover the average muon signal in a hybrid data set, but also offers the possibility to calculate the muon signals for each primary in the considered sample of hybrid events. We show how to compute the β_i exponent for a set of hybrid events that consists of a certain fractions of events with different primaries. By using EPOS-LHC simulation as mock dataset and QGSJetII-04 simulations as MC dataset, we can recover the average muon signal in the mock dataset within ~9%, and within less than ~6% for individual primaries. The average β value calculated from the reconstructed muon signal agrees well with the results shown in [31] for EPOS-LHC and QGSJetII-04. The method can be applied to real events to determine the muon signal for each primary, as well as the scaling factor and the β exponent. Thus, measurements of the β exponent can effectively constrain the parameters governing hadronic interactions, improve the accuracy of hadronic models, and also show that ultra-high energy cosmic rays present a great opportunity to explore particle physics beyond the reach of accelerators.

Acknowledgments

The authors are very grateful to the Pierre Auger Collaboration for providing the tools necessary for simulation for this contribution. The authors would like to thank the colleagues from the Pierre Auger Collaboration for all the fruitful discussions. We thank Armando di Matteo for reading the manuscript and Ruben Conceição for valuable discussions. We want to acknowledge the support in Poland from the National Science Centre grant No. 2016/23/B/ST9/01635, grant No. 2020/39/B/ST9/01398 and from the Ministry of Education and Science grant No. DIR/WK/2018/11 and grant No. 2022/WK/12.

Data availability statement

No new data were created or analysed in this study.

ORCID iDs

Kevin Almeida Cheminant https://orcid.org/0000-0001-6352-5339
Dariusz Góra https://orcid.org/0000-0002-4853-5974
Nataliia Borodai https://orcid.org/0000-0003-1864-937X
Ralph Engel https://orcid.org/0000-0003-2924-8889
Tanguy Pierog https://orcid.org/0000-0002-7472-8710
Jan Pękala https://orcid.org/0000-0002-1062-5595
Markus Roth https://orcid.org/0000-0003-1281-4477
Michael Unger https://orcid.org/0000-0002-7651-0272
Darko Veberič https://orcid.org/0000-0003-2683-1526
Henryk Wilczyński https://orcid.org/0000-0003-2652-9685

References

[1] Matthews J 2005 A Heitler model of EAS *Astropart. Phys.* **22** 387
[2] H E S S -high energy stereoscopic system (https://mpi-hd.mpg.de/hfm/HESS/pages/about/)
[3] Aleksic J *et al* 2016 The major upgrade of the MAGIC telescopes, Part II: a performance study using observations of the Crab Nebula *Astropart. Phys.* **72** 76
[4] Weekes T C *et al* 2002 VERITAS: the very energetic radiation imaging telescope array system *Astropart. Phys.* **17** 221
[5] Abeysekara A U *et al* 2017 Observation of the Crab Nebula with the HAWC gamma-ray observatory *ApJ* **843** 39
[6] Tokuno H *et al* 2012 New air fluorescence detectors employed in the Telescope Array experiment *Nucl. Instr. Meth.* A **676** 54
[7] (Pierre Auger Collab.), Aab A *et al* 2015 The Pierre Auger cosmic ray observatory *Nucl. Instrum. Meth.* A **798** 172
[8] Soldin D on behalf of the EAS-MSU, IceCube, KASCADE-Grande, NEVOD-DECOR, Pierre Auger, SUGAR, Telescope Array, and Yakutsk EASArray collaborations 2021 Update on the combined analysis of muon measurements from nine air shower experiments *PoS* **ICRC2021** 349
[9] Pierog T *et al* 2015 EPOS LHC: test of collective hadronization with data measured at the CERN Large Hadron Collider *Phys. Rev.* C **92** 034906
[10] Ostapchenko S *et al* 2011 Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model *Phys. Rev.* D **83** 014018
[11] (Pierre Auger Collab.), Aab A *et al* 2016 Testing hadronic interactions at ultrahigh energies with air showers measured by the Pierre Auger observatory *Phys. Rev. Lett.* **117** 192001
[12] (MIA, HiRes Collab.), Abu-Zayyad T *et al* 2000 Evidence for changing of cosmic ray composition between 10^{17} and 10^{18} eV from multicomponent measurements *Phys. Rev. Lett.* **84** 4276
[13] Bogdanovet A G *et al* 2018 Investigation of very high energy cosmic rays by means of inclined muon bundles *Astropart. Phys.* **98** 13
[14] Bellido J A *et al* 2018 Muon content of extensive air showers: comparison of the energy spectra obtained by the Sydney University Giant air-shower recorder and by the Pierre Auger observatory *Phys. Rev.* D **98** 023014
[15] (Telescope Array Collab.), Abbasi R U *et al* 2018 Study of muons from ultrahigh energy cosmic ray air showers measured with the telescope array experiment *Phys. Rev.* D **98** 022002
[16] (KASCADE-Grande Collab.), Apel W D *et al* 2017 Probing the evolution of the EAS muon content in the atmosphere with KASCADE-Grande *Astropart. Phys.* **95** 25
[17] Fominet Y A *et al* 2017 No muon excess in extensive air showers at 100–500 PeV primary energy: EAS-MSU results *Astropart. Phys.* **92** 1
[18] Dembinski H *et al* 2018 Report on tests and measurements of hadronic interaction properties with air showers *EPJ Web Conf.* **210** 02004
[19] (IceCube Collab.), Gonzalez J *et al* 2018 Muon Measurements with IceTop Nagoya, Japan (2018) *EPJ Web Conf.* **208** 03003
[20] (Pierre Auger Collab.), Aab A *et al* 2020 Direct measurement of the muonic content of extensive air showers between 2×10^{17} and 2×10^{18} eV at the Pierre Auger Observatory *Eur. Phys. J.* C **80** 751
[21] (Pierre Auger Collab.), Aab A *et al* 2014 Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above $10^{17.8}$ eV *Phys. Rev.* D **90** 122005
[22] Porowski CZ 2018 The top-down method of extensive air shower reconstruction: its feasibility and prospects for testing nuclear interaction models thesis *Repozytorium IFJ PAN* **289** (http://rifj.ifj.edu.pl/handle/item/289)
[23] Góra D *et al* 2021 Muon number rescaling in simulations of air showers *Proceedings of 37th International Cosmic Ray Conference (ICRC 2021) (July 12th –23rd, 2021)* 207 arXiv:2108.07527
[24] Heck D *et al* 1998 CORSIKA: a Monte Carlo code to simulate extensive air showers *Report FZKA Forschungszentrum Karlsruhe* **6019**
[25] Argiró S *et al* 2007 The offline software framework of the Pierre Auger observatory *Nucl. Instrum. Meth.* A **580** 1485
[26] Pierre Auger Collab., Kégl B 2013 Measurement of muon signal using the temporal and spectral structure of the signals in surface detectors of the Pierre Auger observatory *XXXIII International Cosmic Ray Conference* 33 415
[27] Pierre Auger Collab, Souza V 2016 Measurements of the depth of maximum of air-shower profiles at the Pierre Auger observatory and their composition implications *XXV European Cosmic Ray Symposium* arXiv:1701.06812
[28] Ulrich R *et al* 2011 Hadronic multiparticle production at ultra-high energies and extensive air showers *Phys. Rev.* D **83** 054026
[29] Brun R and Rademakers F 1997 ROOT—an object oriented data analysis framework *Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. Meth. A* 389 81
[30] Jameas F and Roos M 1975 Minuit—a system for function minimization and analysis of the parameter errors and correlations *Comp. Phys. Commun.* **10** 343
[31] Cazon L *et al* 2018 Probing the energy spectrum of hadrons in proton air interactions at ultrahigh energies through the fluctuations of the muon content of extensive air showers *Phys. Lett.* B **784** 68
[32] The Large Hadron Collider (LHC) web page: https://home.cern/science/accelerators/large-hadron-collider
[33] Anchordoqui L A *et al* 2017 Strange fireball as an explanation of the muon excess in Auger data *Phys. Rev.* D **95** 063005
[34] Alvarez-Muñiz J *et al* 2012 Muon production and string percolation effects in cosmic rays at the highest energies arXiv:1209.6474
[35] Farrar G R and Allen J D 2013 A new physical phenomenon in ultra-high energy collisions *EPJ Web Conf.* **53** 07007
[36] Farrar G R 2019 Particle physics at ultrahigh energies arXiv:1902.11271
[37] Aloisio R *et al* 2014 Spectra of astrophysical particles at ultra high energy and the auger data *Frascati Phys. Ser.* **58** 274