Method to calculate the $\beta$ exponent of the Heitler–Matthews model of hadronic air showers

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

*Institute of Nuclear Physics PAN, Radzikowskiego 152, Krakow, Poland
bKarlsruhe Institute of Technology (KIT), Institute for Astroparticle Physics, Karlsruhe, Germany

E-mail: Dariusz.Gora@ifj.edu.pl

The number of muons in an air shower is a strong indicator of the mass number $A$ of the primary cosmic-ray, increasing as a small power of it, $N_\mu \sim A^{(1-\beta)}$, where the exponent $\beta$ is slightly less than 1. This behaviour can be explained in terms of the Heitler–Matthews model of hadronic air showers. In this paper, we present a method for calculating $\beta$ from the Heitler–Matthews model. The method has been successfully verified with a series of simulated events corresponding to events observed by the Pierre Auger Observatory at $10^{19}$ eV. To follow real measurements of the mass composition at this energy, the generated sample consists of certain fractions of events produced with p, He, N and Fe primary particles. Since hadronic interactions at the highest energies can differ from those observed at energies reached by terrestrial accelerators, we generate a mock dataset with $\beta = 0.92$ (the canonical value) and $\beta = 0.96$ (a more exotic scenario). The method can be applied to measured events to determine the muon signal for each primary particle as well as the muon scaling factor and the $\beta$ exponent. Determining the $\beta$ exponent can effectively constrain parameters that govern hadronic interactions and help resolve the so-called muon problem, where hadronic interaction models predict too few muons relative to the observations. In this paper, through a simulation study, we lay foundations for future analyses of measured data from the Pierre Auger Observatory.
1. Introduction

Simulations of extensive air showers using current hadronic interaction models predict too small a number of muons compared to observations by the air-shower experiments, which is known as the muon deficit problem. To study the muon deficit we use the top-down (TD) method [1–4] – this chain of simulations and reconstructions enables us to calculate signals in the fluorescence (FD) and surface detectors (SD) of cosmic ray experiments like the Pierre Auger Observatory or Telescope Array. For each observed hybrid 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. The results of the simulations depend on the properties of the hadronic interaction models that are included in the simulation software. Therefore, 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 to obtain new information that will enable us to refine interaction models and thus reduce the discrepancy between the observations and simulations [1, 4].

In this paper a method is proposed to calculate the $\beta$ exponent of the Heitler–Matthews model [5] by including also the muon deficit problem. The idea of the method is to find a set of muon rescaling parameters $\varepsilon_k$ for different primaries $k$, which is a function of only two parameters: $\varepsilon_p$ and $\Delta \beta$. These two parameters indicate how much we need to scale the proton signal ($\varepsilon_p$ term) and by how much to modify the $\beta$ exponent ($\Delta \beta$) in the Heitler–Matthews formula in order to match the observed numbers of muons in data and in simulations. The method requires the first two moments of the individual $z_k$-distributions (our model) and overall $z$-distribution (the measured observable) to match. In addition we require that the $\varepsilon_k$ parameters should follow the Heitler–Matthews progression. The $z_k$-distribution is essentially the difference between the total signal at 1000 m of a real hybrid event and of the total signal at 1000 m of the Monte Carlo (MC) dataset. In other words, the method tells us by how much each individual $z_k$-distribution must be shifted, rescaled and then, weighted and summed, in order to retrieve the measured $z$-distribution. In TD-analysis we have the input dataset, which are real or mock hybrid events, and the matched dataset, which is produced via Conex/Corsika/Offline Monte-Carlo simulations [6, 7]. The input dataset contains $N$ events and the events will be indexed as $n = 1, \ldots, N$. The multiple profile-matched MC events, simulated with primary $k$, corresponding to an input event $n$ are indexed with $i = 1, \ldots, M$ and are thus denoted with the triplet subscript $nki$.  

2. Two-parameter nonlinear scaling model

Both observations of air showers and simulations are in agreement on that the number of muons $N_\mu$ grows almost linearly with the shower energy $E$, and it also increases with a small power of the

1Hybrid event is seen simultaneously by the SD and FD detectors.
2All $S^x$ symbols will be referring to the signal at 1000 m from the shower core so that the 1000 subscript can be dropped entirely. The signals at 1000 m for the input dataset will have no decorations, i.e. just $S$, and the signals from the matched dataset will be denoted with $\bar{S}$. 

Figure 1: Average logarithm of the muon signal for EPOS-LHC [8] and QGSJetII-04 [9]. Solid lines are fits of the function $\bar{S}_\mu = \text{const} A_1^{1-\beta}$ to the TD simulation. From the fit, we obtain $\beta = 0.925 \pm 0.003$ for EPOS-LHC (red line), and $\beta = 0.918 \pm 0.003$ for QGSJetII-04 (blue line).

primary mass $A_k$. These relations can be reproduced in the framework of the Heitler–Matthews model of hadronic air showers [5]. This model predicts

$$N^k_{\mu} = A_k \left( \frac{E}{A_k} \right)^{\beta},$$

where $\beta \approx 0.9$. More precisely, MC simulations yield $\beta_{\text{mc}} = 0.927 \pm 0.002$ for Epos-LHC and $\beta_{\text{mc}} = 0.925 \pm 0.002$ for QGSJetII-04 [10]. For any fixed energy Eq. (1) describes how the muon number depends on the primary mass: $N^k_{\mu} = N^0_{\mu} A_k^{1-\beta}$. Simulations have shown that muon number depends on various properties of hadronic interactions (e.g. multiplicity, charge ratio, baryon–anti-baryon pair production) [11]. Therefore, estimating the $\beta$ exponent from data would be helpful in constraining the parameters of hadronic interactions and improving the accuracy of models. On the other hand, results obtained from the Pierre Auger Observatory and other 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. To account for this effect, we can formulate a scaling ansatz in Eq. (1) by:

$$N^k_{\mu} = \bar{N}^k_{\mu} A_k^{1-\beta} e^{\varepsilon_p} A_k^{-\Delta \beta},$$

where the scaling factor can be defined as: $r_{\mu,k} := 1 + \varepsilon_k := e^{\varepsilon_p} A_k^{-\Delta \beta} = \exp(\varepsilon_p - \Delta \beta \ln A_k)$. Thus, having MC values of the $\beta_{\text{mc}}$ for the given hadron interaction model and the value of the parameter $\Delta \beta$, we can calculate the $\beta$ exponent from $\beta = \beta_{\text{mc}} + \Delta \beta$.

In the context of this work, this corresponds to saying that the number of muons $N^k_{\mu}$ in the input dataset is proportional to the muon number $\bar{N}^k_{\mu}$ in the matched dataset, with the usual Matthews-Heitler progression with mass $A_k$, but with a slight scaling $e^{\varepsilon_p}$ and modification of $\beta$ by $\Delta \beta$. In this work, the input dataset is constructed from Epos-LHC simulations (mock dataset) and is built by taking MC simulations from the TD simulation chain obtained with Epos-LHC around $10^{19}$ eV. The matched dataset is obtained from QGSJetII-04 simulations. Details regarding these two datasets can be also found in Ref. [4].
Since these simulations were performed for p, He, N, and Fe primaries for both Epos-LHC and QGSJetII-04, we can plot the evolution of the average muon signal as a function of the primary mass for both hadronic models, as shown in Fig. 1. Since QGSJetII-04 predicts, on average, fewer muons than Epos-LHC, one can imagine that the muon problem can be recreated by comparing the two hadronic models. Therefore, we can try to figure out what is the best way to rescale QGSJetII-04 in order to match the muon signal of the mock dataset built with Epos-LHC.

From Fig. 1 we can also see, that the average muon signal increases as a function of the primary mass. As expected, both hadronic models considered display a similar ratio with the average about \( r^\mu_{\text{true}} = \frac{\bar{S}_\mu^{\text{em}}}{\bar{S}_\mu^{\text{qgsjet}}} = 1.10 \pm 0.04 \); upon closer examination we also see that larger rescaling is needed for protons \((1.12 \pm 0.03)\) than for iron \((1.08 \pm 0.03)\). In Fig. 1 we show fits to the MC muon signal from Epos-LHC and QGSJetII-04, motivated by the Heitler–Matthews model. The calculated value of \( \beta^{\text{mc}} \) from the fit is about 0.92, so it is pretty close to the values from Ref. [10]. This cross-check of \( \beta \)-calculation is a validation of our TD simulations.

### 3. Fitting the \( z \)-histogram

The mean signal \( \langle S \rangle \) of the input dataset is the sum of the mean electromagnetic (em) and muonic components

\[
\langle S \rangle = \sum_k f_k \langle S \rangle_k = \sum_k f_k \left( \langle S^{\text{em}} \rangle_k + \langle S^{\mu} \rangle_k \right) = \langle S^{\text{em}} \rangle + \langle S^{\mu} \rangle,
\]

where \( \langle \cdot \rangle_k \) denotes a mean within a given primary class \( k \). Note that for the input dataset the averages for given \( k \) are not really observable, but it is clear that a sum over the composition fractions \( f_k \) gives then the average in the whole input dataset, a quantity which is fully available. Equivalently, for the mean signal \( \langle \bar{S} \rangle \) in the matched dataset, where the quantities are known for various primary groups \( k \), we can explicitly write

\[
\langle \bar{S} \rangle = \sum_k f_k \langle \bar{S} \rangle_k = \sum_k f_k \left( \langle \bar{S}^{\text{em}} \rangle_k + \langle \bar{S}^{\mu} \rangle_k \right) = \langle \bar{S}^{\text{em}} \rangle + \langle \bar{S}^{\mu} \rangle,
\]

where \( \langle \bar{S} \rangle_k := \left( \sum_n \frac{M_{nk}}{M} \bar{S}_{nk} \right) / \sum_n M_{nk} \) is the signal \( \bar{S}_{nk} \) of the matched dataset averaged over all \( n \) and \( i \) for a given \( k \).
Since we assume a perfect matching of the longitudinal profile and thus the EM component of the signal, all the $\tilde{S}_{nki}^{\text{em}}$ are very close or identical to the $S_n^{\text{em}}$ signals in the corresponding input events. The mean difference $\Delta S$ of the signals in the two datasets thus only depends on the muonic part

$$\Delta S := \langle S \rangle - \langle \tilde{S} \rangle = \sum_k f_k \left( \langle S \rangle_k - \langle \tilde{S} \rangle_k \right) = \sum_k f_k \left( \langle S^\mu \rangle_k - \langle \tilde{S}^\mu \rangle_k \right) = \langle S^\mu \rangle - \langle \tilde{S}^\mu \rangle = \Delta S^\mu. \tag{5}$$

The mean muonic signals $\langle S^\mu \rangle_k$ of the primary $k$ in the input data can be obtained by rescaling the muonic signals $\langle \tilde{S}^\mu \rangle_k$ in the matched dataset with corresponding scaling factors $1 + \varepsilon_k$,

$$\langle S^\mu \rangle_k = (1 + \varepsilon_k) \langle \tilde{S}^\mu \rangle_k. \tag{6}$$

With this scaling we can simplify the difference $\Delta S$ from Eq. (5) into

$$\Delta S^\mu = \sum_k f_k \varepsilon_k \langle \tilde{S}^\mu \rangle_k. \tag{7}$$

On the other hand, as it is clear from Eq. (5), $\Delta S \equiv \Delta S^\mu$. The third term of Eq. (5) can be rewritten as

$$\sum_k f_k \left( \langle S \rangle_k - \langle \tilde{S} \rangle_k \right) = \langle S \rangle - \sum_k f_k \langle \tilde{S} \rangle_k, \tag{8}$$

so that for each event $n$ and match $i$ we can define an observable

$$z_{ni} = S_n - \sum_k f_k \tilde{S}_{nki}. \tag{9}$$

Equivalently, based on Eq. (7) we can define a scaling-dependent quantity

$$\tilde{z}_{ni} = \sum_k f_k \varepsilon_k \tilde{S}_{nki}^\mu = \sum_k f_k \varepsilon_k g_k(\theta_n) \tilde{S}_{nki}, \tag{10}$$

where $\tilde{S}_{nki}^\mu$ is obtained either directly from the MC events or, like here, by using a factor $g$ from universality, $\tilde{S}_{nki}^\mu = g_k(\theta_n) \tilde{S}_{nki}$. The average muon signal as a fraction of the total signal at the ground, $g_k(\theta_n)$ has been calculated in our previous analyses, see for example [4].

For each event $n$ and $i$ we can also define a variable $z_{nki} = S_{nki} - \tilde{S}_{nki}$, which is a simple difference between the total signal for data and MC, for given primary $k$. In Fig. 2 we show corresponding distributions of this variable for the considered primaries obtained from TD simulations with Eros-LHC and QGSJetII-04 (for simplicity we use notation $z_k$ for each individual histogram). As we can see from Fig. 2 for the considered number of events, the corresponding $z_k$-distribution can be quite well described by a Gaussian function, the fit to histograms gives $\chi^2/\text{ndf} \approx 1.5$. From the fit for individual distributions we can get the mean value of signal difference $\langle z_k \rangle$ and the corresponding standard deviation $\sigma(z_k)$. These variables can be used to define the probability density function (PDFs) for each primary $k$, which is given by

$$P_k(z_k, \sigma(z_k)) = \frac{1}{\sqrt{2\pi} \sigma(z_k)} \exp \left[ -\frac{(z_{nki} - \langle z_k \rangle)^2}{2\sigma^2(z_k)} \right]. \tag{11}$$

\*\*It is worth mentioning that this fraction depends on the shower zenith angle and the type of the primary cosmic ray, and only slightly on different hadronic interaction models [12].
Figure 3: (Left): The $z_{ni}$-distribution as described by Eq. (9) with $f_p = 0.15$, $f_{He} = 0.38$, $f_N = 0.46$, and $f_{Fe} = 0.01$ [13] for mock dataset. Since we have 68 mock events (Epos-LHC) and 10 QGSJetII-04 events associated to each of the mock events we have 680 events contained in this histogram. The distribution is fitted (red line) with a Gaussian function in order to get its mean $\langle z_{ni} \rangle = 2.825 \pm 0.16$ and the standard deviation $\sigma(z_{ni}) = 3.80 \pm 0.14$. (Right): Sketch showing the idea of the method i.e. each $z_k$-distribution must be shifted, rescaled, and then weighted and summed, in order to retrieve the $z_{ni}$-histogram.

where again index $k$ spans over different primaries.

Note that according to Eq. (10), the mean position of $z_k$-distribution should be connected with the average ground muon signal expected for a given primary. However, such conversion is possible, if we already know proportionality constants i.e. scaling factors $\epsilon_k$. In other words, if we plot rescaled distribution shown in Fig. 2 in $\langle S^\mu \rangle$ phase-space, the means of such distributions should give us average muon signals on the ground for considered masses. Moreover, we should expect from physics of extensive air showers that position of the mean for lighter element should be smaller that for heavier element i.e. $\langle S^\mu \rangle_p < \langle S^\mu \rangle_{He} < \langle S^\mu \rangle_N < \langle S^\mu \rangle_{Fe}$. 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 by using two-parameter scaling model $\epsilon_k$.

In order to find $\epsilon_k$ and thus to convert the mean of $z_k$-distribution to $S^\mu$ phase-space, we can use the Minuit minimization, where the fitted function is a combination of four Gaussian PDFs, which have the form

$$F(\vec{c}, A_{mpl}) = A_{mpl} \sum_k f_k \frac{1}{\sqrt{2\pi}\sigma(z_k)} \exp \left[ -\frac{(z_{nik} - \epsilon_k \langle S^\mu \rangle_k)^2}{2\sigma^2(z_k)} \right],$$

(12)

where $\epsilon_k = e^{c_r - \Delta \beta \ln A_k} - 1$ and const $= A_{mpl}$. The $f_k$ is fraction of $N = 68$ pure mass samples and const gives possibility to rescale the normalized individual $z_k$-distribution to overall $z_{ni}$-histogram.

In this way from the Gaussian fit given by Eq. (12) to overall $z_{ni}$-histogram, the correction factors $\epsilon_k$ and $\Delta \beta$ for hadronic models can be calculated. In other words, Eq. (12) tells us by how much each $z_k$-distribution must be shifted, rescaled, and then weighted and summed, in order to retrieve the $z_{ni}$-distribution and also its first and second moments, see also Fig. 3.
The results are shown in Fig. 4 and Table 1. We see that the fit can reproduce the ratio of the muon signals of simulations using Epos-LHC (mock data) and QGSJetII-04 within ~5%: as we already previously mentioned, the muon signal ratio for MC-true is \( r_{\mu,\text{true}} = 1.10 \pm 0.04 \) and the average reconstructed ratio (from Table 1) is 1.15 ± 0.06. The difference is caused by the fact that the signal for the mock dataset is not exactly equal to the one for Epos-LHC (Table 1). We also recover the \( \beta \) parameter (average ≈0.92) for the studied set, because parameter \( \Delta \beta \) is zero within its error i.e. \( \Delta \beta = 0.003 \pm 0.035 \). Finally we can check our solution by comparing the mean \( n_\mu \) given by Eq. (10) and that from a Gaussian fit to the \( z \)-histogram shown in Fig. 3. We get 2.74 ± 0.49 VEM vs. \( \langle z_{\mu} \rangle = 2.83 \pm 0.16 \) VEM, which agree very well within the errors limits. We have the standard deviation match by definition, because \( \sigma^2(z_{\mu}) = \int \sum_k f_k z_{\mu,k}^2 P_k \sigma(z_k) dz_{\mu,k} = \sum_k f_k \sigma^2(z_k) \).

Since the true value (\( S^P \)) of the hybrid dataset may differ from that of the hadron interaction models used in this analysis, it would be interesting to perform the same analysis for a sample dataset built from the Epos-LHC sample, but with hadron interaction evolution. For a sample dataset built from the Eros-LHC sample, we constructed a mockdataset with the evolution of the mean muon signal as a function of primary mass, leading to a significantly different exponent value.
\( \beta \approx 0.96 \). This allows us to check whether the fitting procedure is able to recover this value as well. The average muon signal of the new sample set as a function of primary mass is shown in Table 1. Two features of this mock dataset can be noticed: for nitrogen, a slight rescaling from QGSJetII-04 to the mock dataset is needed \( (r_{\mu,N} = 1.01) \) and for iron the average rescaling of the muon signal is lower than 1 for the mock dataset \( (r_{\mu,Fe} = 0.96) \). The results of the fit are shown in Fig. 4 (right) and in Table 1. We can see that down scaling scaling of the primary iron is slightly underestimated, while the signal is well recovered for all other elements. The muon signal from the mock dataset is recovered within 2.4%. Moreover, fitting of the reconstructed muon signal gives a value of \( \Delta\beta = 0.04 \) which agrees quite well with the expectation \( \beta = 0.955 \pm 0.005 \), although the error of \( \Delta\beta \) is quite large (0.02).

5. Summary and Conclusion

The method presented in this work recovers the mean muon signal and provides the ability to calculate muon signals for each element in the considered sample of real-like events. In this work, we have been performed calculations of muon scaling factors and \( \beta \) exponents, by fitting a four-element Gaussian distribution to the overall \( z \)-histogram, with two-parameter scaling model which should follow Heitler–Matthews progression. This work shows that the \( z \)-method can be applied to hybrid events to determine the muon signal, the scaling factor (total and for each element), and the \( \beta \) exponent.

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.

References


\footnote{5We 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, grant No. 2022/45/B/ST9/02163 and from the Ministry of Education and Science grant No. DIR/WK/2018/11 and grant No. 2022/WK/12.}