A composition sensitive log-likelihood ratio for cosmic rays and gamma rays

The IceCube Collaboration[†]

[†] *http://icecube.wisc.edu/collaboration/authors/icrc17_icecube*
E-mail: hershal@udel.edu

IceTop, the surface component of the IceCube Neutrino Observatory detects extensive air showers (EAS) initiated by cosmic rays and gamma rays in the energy range of PeV to EeV. IceTop is sensitive to air shower characteristics such as the muon content, local fluctuations and shower-front curvature. These characteristics correlate with the mass of the primary particle for a given energy and arrival direction. Using IceTop observables, we construct three two-dimensional probability distribution functions (PDFs) that reflect these shower-front properties. A log-likelihood ratio is constructed using these PDFs, assuming a pair of hypotheses corresponding to primaries with distinct mass. We show the effectiveness of this parameter for discriminating gamma rays from cosmic rays and heavy from light cosmic ray primaries.

Corresponding author: Hershal Pandya[*][1]

[1]*Dept. of Physics and Astronomy, University of Delaware, Newark, DE 19711, USA*

35th International Cosmic Ray Conference — ICRC2017
10–20 July, 2017
Bexco, Busan, Korea

---

[*]Speaker.

1. Introduction

High energy cosmic rays and gamma rays arriving at Earth, interact in the atmosphere creating cascades of secondary particles. These particle air showers can be detected using ground based air Cherenkov telescopes, fluorescence telescopes, or surface particle detectors. The sources of high energy cosmic rays and the mechanism responsible for accelerating them are still not fully understood. Measurement of the cosmic ray composition across the whole cosmic ray spectrum is important in order to constrain models for the cosmic ray acceleration and propagation. Cosmic ray composition is also important input for deciphering cosmic ray spectrum features such as the knee, the ankle, and the transitions between various cosmic ray populations which are expected to produce these features [1]. Cosmic ray interactions in

Figure 1: Muon spectra for near vertical 1 PeV photon, proton, and iron CORSIKA showers simulated using SIBYLL 2.1.

their source environment or in the interstellar medium lead to emission of gamma rays and neutrinos. Hence, the detection of high energy gamma rays [2, 3] can give valuable insight into the cosmic ray acceleration mechanisms as well as the sources of the IceCube astrophysical neutrinos [4]. In this work, we present a method to discriminate between different primary particle types, based on the shower signals detected by an array of surface particle detectors. Although this method was developed for analysis of IceTop data, it is generally possible to adapt this technique for other surface arrays.

Surface particle detectors are able to reconstruct the energy and direction of the primary particle, but the challenge lies in reconstruction of the mass of the primary particle based on the shower imprint. Extensive air showers generated by photons, protons, and heavier cosmic rays differ mainly in two aspects for a given altitude: the shower age and the hadronic content. The approximate shower maxima, for example, for 2 PeV showers generated by iron, protons, and photons occur at 460, 580 and 620 g/cm^2 respectively [5]. Hence, lighter particles generate younger showers at the detector altitude. In addition, these showers differ in the hadronic content of the cascades, which can be derived from the local fluctuations and muon content of the shower-front. Muon content comparison for photons, protons, and iron is shown in Fig. 1. Measurement of these properties is contingent on the geometry of the array, and response of the detectors to electrons, photons, and muons within the EAS.

The mapping from detector observables to the mass of the primary particle depends on the hadronic interaction models used during simulation of the air showers. For this work, air showers were simulated with CORSIKA [6], using SIBYLL 2.1 [8] for high energy, and FLUKA [7] for low energy hadronic interactions.
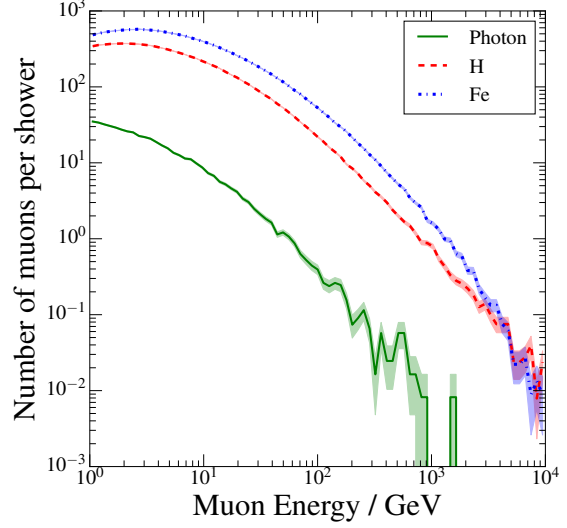
2. IceTop

IceTop is the surface component of the Ice-Cube neutrino observatory which is located at the geographic South Pole. As shown in Fig. 2, it is a square kilometer array of 162 ice tanks, each equipped with two digital optical modules (DOMs). The DOMs detect Cherenkov radiation emitted by relativistic charged particles passing though the tank. A pair of tanks forms a station, and spacing between two neighboring stations is 125 m on average, with an additional denser in-fill region in the center of the array. IceTop is located at an altitude of 2835 m above sea level which corresponds to an atmospheric depth of around 692 g/cm^2. IceTop is most sensitive to air showers generated by cosmic rays in the energy range of roughly 1 PeV to 1 EeV. IceTop triggers at a

Figure 2: IceTop array geometry [9].

rate of about 30 Hz, with a rate of successfully reconstructed events above 1 PeV of about 3 Hz.

IceTop DOMs are calibrated using vertical muons from low energy air showers. Muons deposit roughly the same amount of energy, characterized by their constant ionization loss in ice and their path length inside the tank. Thus, the signals are measured in the units of vertical equivalent muons (VEM). Shower direction, core location, and shower energy are reconstructed by simultaneously fitting the measured charges with an LDF, and signal times with a function which describes the shower-front curvature. The lateral distribution function is defined as [9]

$$S(R) = S_{125} \left(\frac{R}{125 \; m} \right)^{-\beta - \; 0.303 \; \log_{10} \left(\frac{R}{125 \; m} \right)}, \tag{2.1}$$

where S(R) is the signal measured at a lateral distance R from shower axis, β is the slope of the logarithmic LDF at 125 m. The signal measured at 125 m, S_{125}, is the shower size or energy proxy. Snow accumulates on top of the IceTop tanks over time, which reduces the measured signal in the tank. This attenuation is accounted for while reconstructing the shower size, by taking the snow height on each tank into account.

The simulation datasets used in this work are generated using snow heights measured in October of 2012, and the observed data used for testing is a randomly selected 10% subsample of the IceTop data recorded between May 2012 and May 2013. We restrict our analysis to the $\log_{10}(S_{125})$ range of 0 to 2, which corresponds to energy range of roughly 1 PeV to 78 PeV for vertical showers. Quality cuts are applied to all events to ensure that the events used in the final sample are successfully reconstructed.

3. Construction of Probability Distribution Functions

For every successfully reconstructed IceTop event, the following observables are available: a.) measured tank charge, signal time with respect to core arrival time, and lateral distance of the tanks
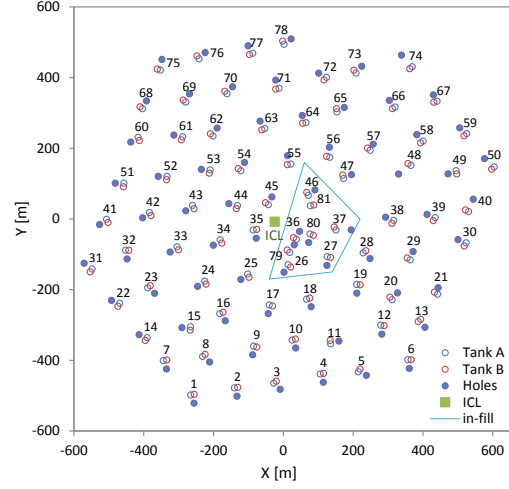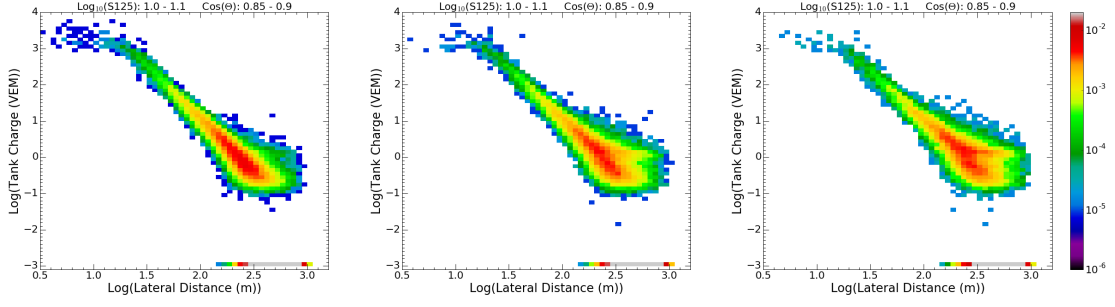
3

Figure 3: PDF based on lateral distribution of charges for photons, protons, and iron (left to right). For events with $1.0 \leq \log_{10}(S_{125}) < 1.1$ and $0.85 \leq \cos\theta < 0.9$.

hit by the shower-front, b.) lateral distance for tanks that did not record any signal (unhit tanks). The residual time ΔT for each tank is obtained by subtracting the expected signal time, according to a median shower curvature, from the observed signal time. Thus for each event, we obtain the following observables:

$$\{(Q_i, \Delta T_i, R_i)\}, \forall \text{ Hit tank,}$$
$$\{R_i\}, \forall \text{ Unhit tank,}$$
$$N_{\text{Hit}} + N_{\text{UnHit}} = 162, \qquad (3.1)$$

where Q_i is the charge measured, ΔT_i is the residual time, and R_i is the lateral distance of ith tank. Using these observables we construct three two-dimensional probability distribution functions (PDFs). Since different events have varying number of tanks hit by the shower-front, we include the unhit tanks in the PDF by assigning them with a false charge of 10^{-3} VEM and a false residual time of 10^{-2} ns. This ensures a correct normalization for the PDF. The shower properties vary with energy of the primary particle as well as the zenith angle. Hence, the PDFs are generated separately for various $\log_{10}(S_{125})$ and $\cos\theta$ bins. The values of $\log_{10}(S_{125})$ range from 0.0 to 2.0 in steps of 0.1, and $\cos\theta$ from 0.8 to 1.0 in steps of 0.05.

3.1 Lateral Distribution of Charges

As discussed in Sec. 2, muons going through a tank deposit around 1 VEM energy depending on their path length inside the tank. The signal in tanks near the shower core is dominated by the electromagnetic component of the air shower. Beyond a large lateral distance, the electromagnetic LDF falls off, exposing the roughly 1 VEM signal from GeV muons. This muon feature is visible in the PDF based on the lateral distribution of charges as shown in Fig. 3. The muon signal far from the shower core gets more prominent for heavier masses, whereas it is almost absent for the photon induced air showers. The PDF shown in Fig. 3 also contains information on the shower age and hadronic content. The slope of the LDF correlates with the shower age, and the width of the charge distribution for a given radius, arising from local fluctuations, is correlated with the hadronic content of the shower.
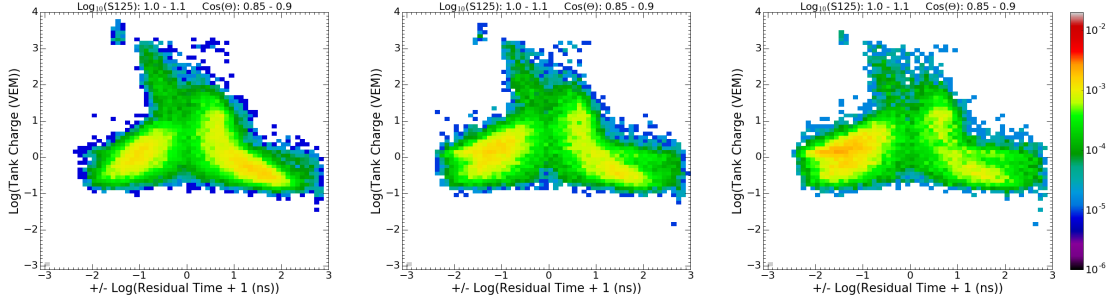
4

Figure 4: PDF based on tank charge versus residual time for photons, protons, and iron (left to right). For events with $1.0 \leq \log_{10}(S_{125}) < 1.1$ and $0.85 \leq \cos\theta < 0.9$.
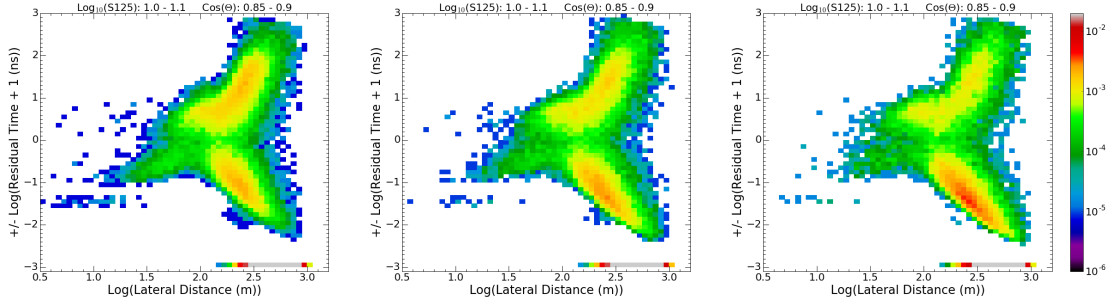
Figure 5: PDF based on shower-front shape after shower curvature subtraction for photons, protons, and iron (left to right). For events with $1.0 \leq \log_{10}(S_{125}) < 1.1$ and $0.85 \leq \cos\theta < 0.9$.

3.2 Time Distribution

In addition to the PDF described in Sec. 3.1, we construct two more PDFs that utilize the temporal distribution of the shower particles. The time residuals ΔT_i can have negative (early) as well as positive (late) values. Hence, the transformation of residual time to the logarithmic scale is done as $\text{Sign}(\Delta T)\log_{10}(|\Delta T|+1)$. From Fig. 4 one can observe that muons arrive relatively early at the detector surface from their signature 1 VEM charge accumulation at negative times. PDFs shown in Fig. 5 represent the shower-front of three primary masses with an average proton shower curvature subtracted out of all of them. It can be noted that heavier masses have a larger fraction of particles arriving earlier in time.

3.3 Log-Likelihood Ratio

For a given pair of hypotheses, H_1 and H_2, a log-likelihood ratio is calculated based on one of the three PDFs, described in the previous section. The pair of hypothesis could be H_γ and H_{CR} for the purpose of separating gamma ray air showers from hadronic showers or it could be H_H and H_{Fe} for the aim of cosmic ray composition. The log-likelihood ratio using lateral distribution of charges, for example, is calculated as

$$\lambda_{QR} = \log_{10}\left(\frac{L(\{(Q_i,\ R_i)\}|H_2)}{L(\{(Q_i,\ R_i)\}|H_1)}\right), \tag{3.2}$$
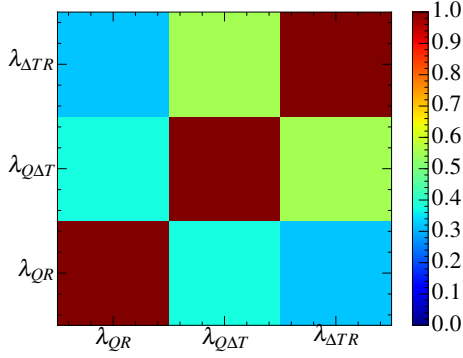
Figure 6: Pearson correlation coefficient matrix for log-likelihood ratios obtained using three PDFs as described in Sec. 3.
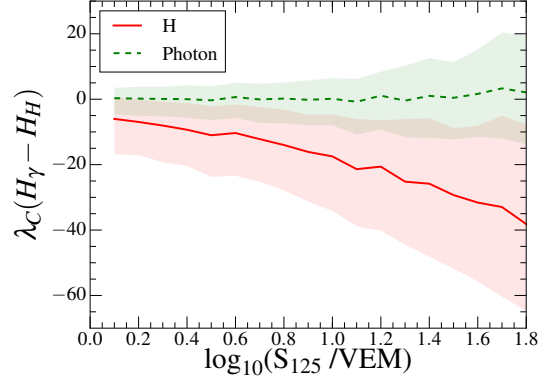

Figure 7: $\lambda_C(H_\gamma$ - $H_H)$ versus $\log_{10}(S_{125})$ for simulated gamma ray and proton showers with $0.8 \leq \cos\theta < 1.0$. Shaded regions contain 95% of the events.

where the likelihood L is defined as

$$L(\{(Q_i, R_i)\}|H) = \prod_{i=1}^{162} P(Q_i, R_i|H),$$ (3.3)

with $P(Q_i, R_i|H)$ being the probability of having a tank with measured charge Q_i(VEM), at a lateral distance R_i(m) from the shower axis, for the hypothesis H. Similarly, one can calculate $\lambda_{Q\Delta T}$ and $\lambda_{\Delta TR}$ from the PDFs based on the time distribution of charges and the shower-front shape.

Infrequently it may be possible that $P(Q_i, R_i|H)$, $P(\Delta T_i, R_i|H)$, or $P(Q_i, \Delta T_i|H)$ for the ith tank, may not be defined in a PDF for a particular event. This may occur due to less number of events used for generating the PDF or due to shower-to-shower fluctuations. For such tanks, the P value is obtained by extrapolation. After the extrapolation, the PDF is re-normalized and the log-likelihood ratio for the event is calculated.

4. Results and Discussion

The log-likelihood ratios $\lambda_{Q\Delta T}$, $\lambda_{\Delta TR}$, and λ_{QR}, are expected to be correlated to each other. The Pearson correlation coefficient matrix for the three ratios is shown in Fig. 6. To fully utilize classification power from all three log-likelihood ratios, a dimension reduction technique such as linear discriminant analysis, or a classification algorithm like the decision tree needs to be implemented. However, as a first estimate one can define a combined log-likelihood ratio as

$$\lambda_C = \lambda_{QR} + \lambda_{\Delta TR} + \lambda_{Q\Delta T},$$ (4.1)

which is equivalent to taking a simultaneous log-likelihood ratio, for a given event, using all three PDFs.

One of the applications of this method is to search for PeV gamma rays in the IceTop data. We calculate the λ_C for simulated proton and gamma ray showers, using the hypothesis pair H_γ -
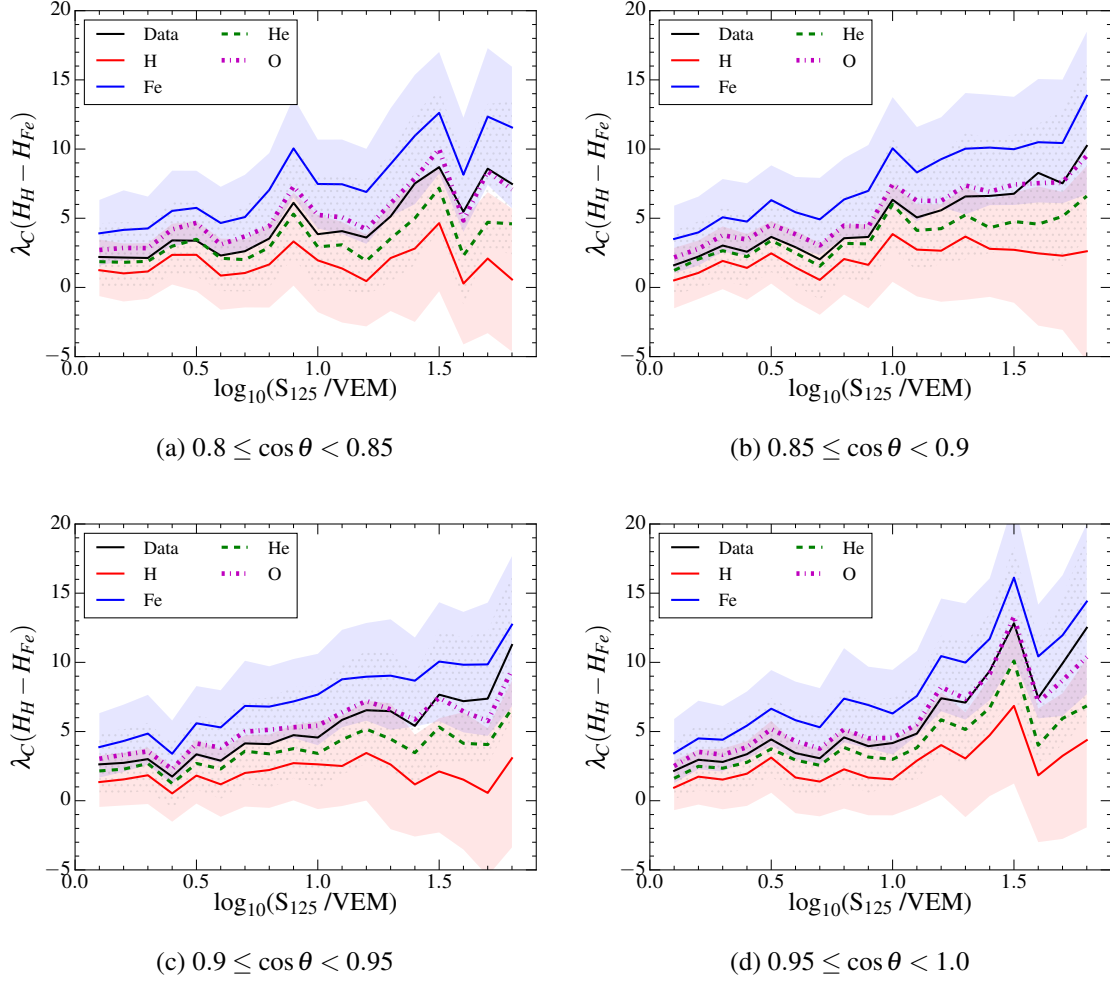
Figure 8: $\lambda_C(H_H - H_{Fe})$ versus $\log_{10}(S_{125})$ for simulated proton, helium, oxygen, and iron, as well as observed IceTop data. Blue and red shaded regions confine 68.27% λ_C distribution for Fe and H, while the dotted region confines 68.27% of data λ_C distribution. The four sub plots correspond to four different zenith angle bins.

H_H, to show the performance of this method (Fig. 7). As shown in Fig. 3, Fig. 4, and Fig. 5, the differences between photon and proton PDFs are rather pronounced, as compared to differences between proton and iron. Since the expected flux of PeV gamma rays is of the order of 10^{-4} of the cosmic ray flux or less, the final sample of gamma rays needs to have a high degree of purity. Hence, we show a 95% error region around the median in Fig. 7. The quality of separation using λ_C improves with $\log_{10}(S_{125})$. This is partly due to an improvement of the angular resolution, and partly due to increasing differences in the shower properties.

The goal in the case of cosmic ray composition is to separate four different mass groups with comparable fluxes but fairly close shower properties. For the purpose of demonstration, we calculate the λ_C using the hypothesis pair H_H - H_{Fe}. The resulting distributions are shown in Fig. 8. The distributions for helium and oxygen are placed between proton and iron, and helium is placed below oxygen, as expected. The data is close to Helium at $\log_{10}(S_{125}) = 0.1$, it crosses Oxygen

at intermediate energies, and is placed above Oxygen for $\log_{10}(S_{125}) = 1.8$. This indicates that the fraction of heavier mass primaries is increasing in the data with increasing primary energy. Inclined showers pass through more atmosphere resulting in decay of particles before they reach the detector. This affects heavier primaries more than lighter primaries. A trend of data getting lighter towards larger zenith angles can be seen in Fig. 8. Although none of these are statistically significant results, they are indicators of this method's capability to classify cosmic ray primaries.

In conclusion, we present this new log-likelihood ratio as a mass sensitive parameter for extensive air showers detected by surface particle detectors. Analyses described in Refs. [2, 3] utilize this technique for discriminating gamma ray showers from cosmic ray showers. An implementation of this technique is underway for an event-by-event determination of cosmic ray mass composition using IceTop data. In principle, it is possible to adapt this technique for other surface detectors, like the HAWC observatory [10] and the surface detector component of the Pierre-Auger observatory [11]. The choice of variables and PDFs would depend on the detector response to various components of the EAS, detector geometry, and resolution of the reconstructed shower parameters.

References

[1] T. K. Gaisser, T. Stanev, S. Tilav, *Front. Phys.* **8** (2013) 748.

[2] **IceCube** Collaboration, H. Pandya et al., `PoS(ICRC2017)705` (these proceedings).

[3] **IceCube** Collaboration, Z. Griffith et al., `PoS(ICRC2017)715` (these proceedings).

[4] **IceCube** Collaboration, M. G. Aartsen et al., *Phys. Rev. Lett.* **113** (2014) 101101.

[5] M. Risse and P. Homola, *Mod. Phys. Lett. A* **22** (2007) 749.

[6] D. Heck, G. Schatz, T. Thouw, J. Knapp, and J. N. Capdevielle, *Report No. FZKA 6019* (1998) .

[7] A. Ferrari, P.R. Sala, A. Fasso, and J. Ranft, *CERN-2005-10 (2005), INFN/TC_05/11, SLAC-R-773*.

[8] E. J. Ahn, R. Engel, T. K. Gaisser, P. Lipari, and T. Stanev , *Phys. Rev. D* **80** (2009) 094003.

[9] **IceCube** Collaboration, R. Abbasi et al., *Nucl. Instrum. Meth. A* **700** (2013) 188.

[10] A.U. Abeysekara et al., *arXiv* (2017) 1701.01778.

[11] The Pierre Auger collaboration, *Nucl. Instrum. Meth. A* **798** (2015) 172.