PROCEEDINGS OF SCIENCE

ONLINE ICRC 2021
THE ASTROPARTICLE PHYSICS CONFERENCE
Berlin | Germany
37th International Cosmic Ray Conference
12–23 July 2021

# Feasibility Studies on improved Proton Energy Reconstruction with IACTs

**Alicia Fattorini**[a,*] **and Wolfgang Rhode, Dominik Elsaesser, Maximilian Noethe, Dominik Baack**[a]

[a] *Department of Physics, TU Dortmund University*
*Otto-Hahn-Str. 4, 44227 Dortmund, Germany*
*E-mail:* alicia.fattorini@tu-dortmund.de

Air showers induced by cosmic protons and heavier nuclei constitute the dominant background for very high energy gamma-ray observations of Imaging Air Cherenkov Telescopes (IACTs). Even for strong very high energy gamma-ray sources the signal-to-background ratio in the raw data is typically less than 1:5000. Therefore, a very large statistic of events, induced by cosmic protons and heavier nuclei, is easily available as a byproduct of gamma-ray source observations. In this contribution, we present a feasibility study on improved reconstruction of the energy of primary protons. For the latter purpose, we used a random forest method trained and tested by using Monte Carlo simulations of the MAGIC telescopes, for energies above 70 GeV. We employ the aict-tools framework, including machine learning methods for the energy reconstruction. The open-source Python project aict-tools was developed at TU Dortmund and its reconstruction tools are based on scikit-learn predictors. Here, we report on the performance of the proton energy regression with the well-tested and robust random forest approach.

*37th International Cosmic Ray Conference (ICRC 2021)*
*July 12th – 23rd, 2021*
*Online – Berlin, Germany*

*Presenter





PoS(ICRC2021)237

## 1. Cosmic Rays

Unlike photons or neutrinos, charged cosmic rays pose the challenge that their place of origin is difficult to identify. On the way from their source to Earth, the protons and heavier nuclei are deflected by galactic and intergalactic magnetic fields, but also by solar winds and the magnetic field of the Earth itself. This leads to a diffuse cosmic-ray flux being measured on Earth. The question of the sources of cosmic rays is still existent and multi-messenger campaigns are underway to solve the puzzle. Although the diffuse spectrum of cosmic rays is now studied in many experiments, the energy range of IACTs is relatively unexplored. Figure 1 shows the spectra of the most prominent particles, protons, helium nuclei and iron nuclei, based on the measurements of selected experiments. The spectra are given in energy per nucleus.

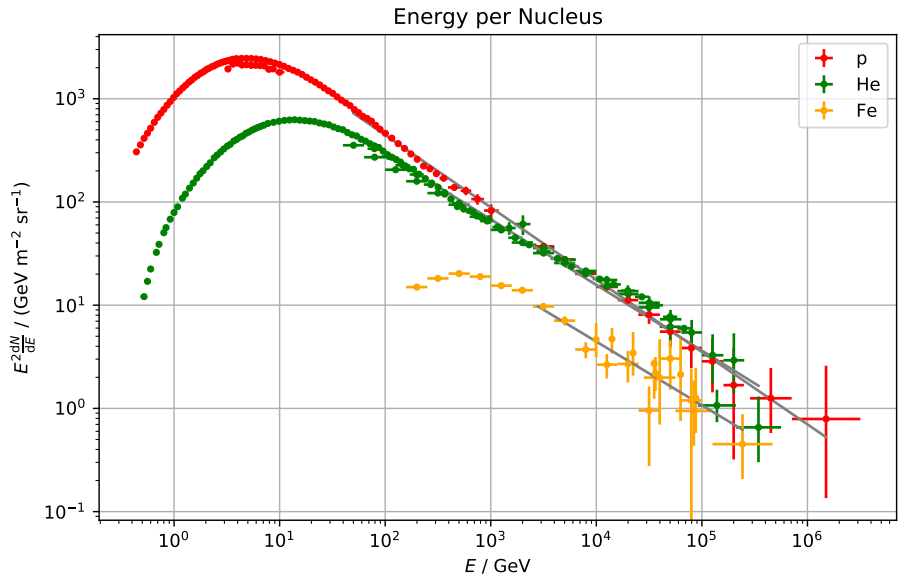


**Figure 1:** Spectra of cosmic protons, helium and iron nuclei. The data is extracted from the database [1]. For the proton spectrum, measurements of CAPRICE98, RUNJOB, PAMELA and CREAM are used. The data points of the helium spectrum are provided by RUNJOB, ATIC-2, PAMELA, CREAM and DAMPE, the iron spectrum is based on measurements of RUNJOB, CRN, SANRIKU and ATIC-2. In grey a fit of a power law function to the data points is shown.

The measurements are mainly from balloon experiments, but IACTs have also started to contribute to measuring the spectra [2], [3] [4]. As an example, simulations of protons and heavier nuclei from the MAGIC telescopes are used in this work to build models for the classification of particles detected by IACTs and reconstruct the energies of the particles classified as protons.

PoS(ICRC2021)237

## 2. Imaging Air Cherenkov Telescopes

While high-flying balloons can directly measure cosmic rays in the high-energy range, ground-based experiments use the Earth's atmosphere as a detector medium and observe the secondary radiation produced by the interaction between incoming cosmic rays and Earth's atmosphere nuclei. The advantage of ground-based experiments is that the measurement apparatus can be built much heavier and larger. This makes it possible to increase the sensitivity of the telescopes and to detect even low particle fluxes. Work is currently underway to build the next generation IACTs, the Cherenkov Telescope Array (CTA), two arrays of IACTs located in the northern and southern hemispheres. Although CTA is primarily being built to detect gamma rays, it will also be more sensitive to cosmic nuclei than any IACT before.

### 2.1 Air Showers

The detection technique of IACTs is based on the fact that highly relativistic charged particles emit Cherenkov light when they pass from a less dense medium into a denser medium. When cosmic rays enter the Earth's atmosphere, the highly relativistic particles interact with the nuclei of the atmosphere. This interaction leads to the production of further particles, so that a cascade of interactions, a so-called air shower, is produced. The charged particles in the showers emit the characteristic Cherenkov light. Depending on the primary particle, characteristic showers are produced, which are difficult to distinguish manually on the basis of the Cherenkov light image in the telescope camera. While a gamma ray entering the atmosphere triggers an electromagnetic shower, a greater variety of interactions can occur in showers generated by hadronic primary particles. Hadronic showers can be divided into three components: an electromagnetic, a muonic and a hadronic component. The electromagnetic component leads to similar shower images in the camera as those triggered by gamma rays, making the distinction between the different types of particle a challenge. In this contribution we will show that with the help of machine learners, it is possible to estimate the type of the primary particles with a high statistical accuracy. The energy of the primary particle can also be determined with a certain probability using the same method.

### 2.2 The MAGIC Telescopes

The Major Atmospheric Gamma-Ray Imaging Cherenkov (MAGIC) Telescopes [5] are a system of two Imaging Air Cherenkov Telescopes located at an altitude of 2200 m at the Roque de los Muchachos Observatory at La Palma (Canary Islands). With its light structure and its strong motor, MAGIC originally was built for the detection and investigation of gamma-ray bursts. Since the commissioning of the first of the two telescopes in 2004, it has made a significant contribution to progress in modern high-energy astrophysics, by detecting the first GRB with an IACT, but also in the search for neutrino sources and dark matter, or in variability studies of gamma ray sources. The telescopes have a mirror surface of 17 m diameter each and a field of view of 3.5 deg in the stereoscopic operation mode. The telescopes are sensitive to gamma rays from 30 GeV and up to 100 TeV and the detection range of nucleons is in the same order of magnitude. The ratio of photons to hadronic particles can be 1:5000 for for weak gamma-ray sources in the MAGIC's energy regime. Hence the measurements of IACTs offer besides the analysis of gamma-ray sources a great opportunity to investigate the cosmic-ray spectrum in the TeV energy range.

## 3. Analysis

The analysis is based on Monte Carlo simulations of protons, helium nuclei and iron nuclei. The showers are simulated with CORSIKA [6] and processed with MAGIC internal software [7] until the shower images have been cleared of electrical noise, gone through image cleaning and parameterized. To improve the quality of the data set, some cuts are made on the simulations. Figure 2 shows the distributions of events before and after the quality cuts.

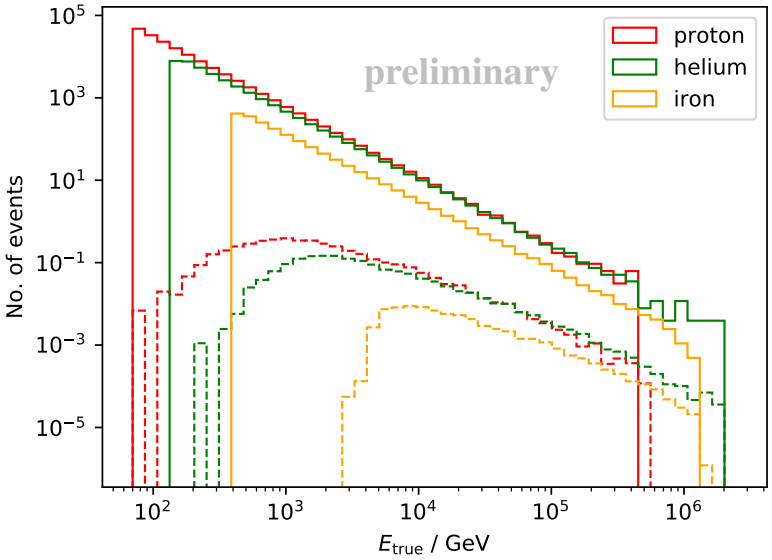


**Figure 2:** Simulated events for the Analysis, exemplified by one of the three subsets and weighted to the empirical spectra of the particles. All simulated particles are shown by the solid lines, the transparent lines represent the particles surviving the quality cuts.

During the parameterization of the shower images, Hillas [8] and stereo parameters are determined. For example, the length and width as well as the total charge of the shower image or the height of the shower maximum based on geometric calculations are used as important attributes in the analysis. For the analysis, the data sets are divided into three equally sized subsets. The first subset is needed to train the random forests for the following particle classification and energy regression. The second subset is used to validate the random forests and calculate the migration of the events. The third subset becomes the test data set, on which the entire analysis chain is tested and its performance evaluated.

### 3.1 The open-source Python Project aict-tools

The open-source software aict-tools [9] was developed at the TU Dortmund and is a tool to train and apply random forests especially for the reconstruction of physical parameters of IACT data. The project is based on the Python module scikit-learn [10] and is publicly available on github. As input, aict-tools uses event files in hdf5 format and a configuration file in which the attributes to be used for training the models must be listed. In addition to training and applying random forests, the tool has many uses, such as applying quality cuts and cross-validating models.

### 3.2 Particle Identification

In this analysis, two random forests are needed for particle classification. A first random forest is trained to separate iron nuclei from the lighter nuclei, here protons and helium nuclei. The random forest is trained on all three particle types and estimates a parameter called Ironess. An event with a high Ironess has more likely an iron nucleus as a primary particle than an event with a lower Ironess. In the second step, a second random forest separates the remaining particles after the first step into protons and rest particles. This random forest is trained on protons and helium nuclei and applied to the remaining data. The parameter estimated by the second random forest is called Heliumness. An event with a high value of Heliumness is more likely to not be a proton compared to an event with a lower Heliumness. In fact, all particles surviving the first cut but not the second are classified as helium nuclei.

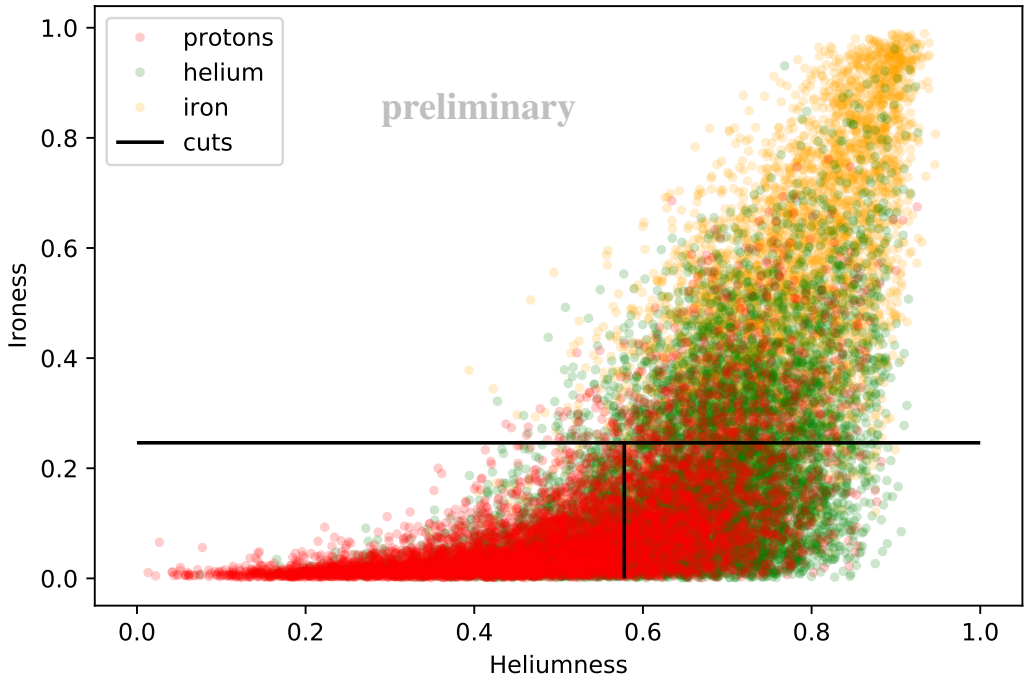


**Figure 3:** Estimated Ironess and Heliumness of the validation subset. The protons are shown in red, the helium and iron nuclei are shown in grey. The black solid lines represent the cuts. All particles above the horizontal line are classified as iron nuclei, the events in the quarter down right are classified as helium nuclei. The quarter down left shows all particles classified as protons.

Figure 3 shows the events of a subset, depending on the parameters Ironess and Heliumness. The black lines represent the cuts, first in Ironess and then in Heliumness. The cuts were chosen for the highest purity and efficiency. For the following analysis, these cuts are applied. All particles above the horizontal line are classified as iron nuclei, the events in the quarter down right are classified as helium nuclei. The quarter down left shows all particles classified as protons. Depending on the goal of the analysis and demands on the data samples, the cuts can be chosen accordingly.

To better present the results of the classification and to quantify the misclassification, a migration matrix is created and shown in figure 4 below.

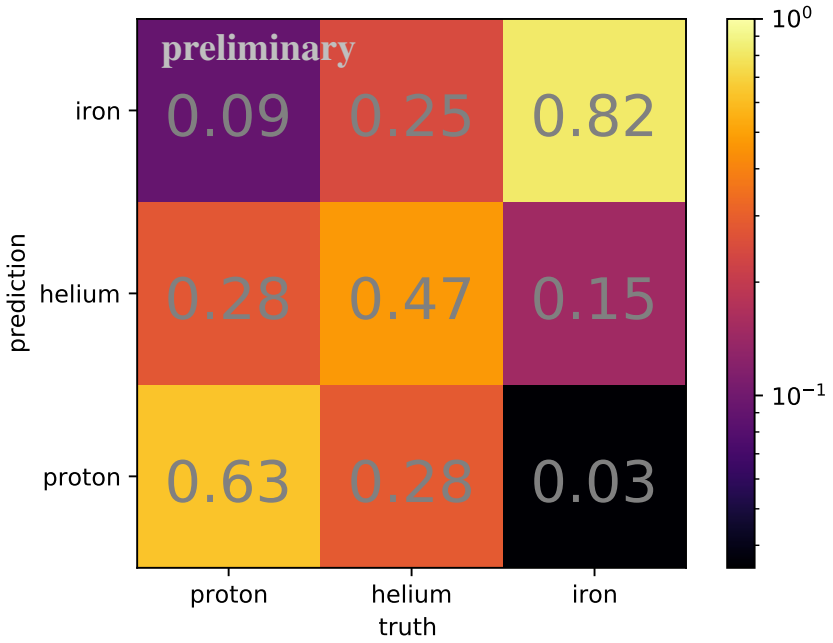


**Figure 4:** Migration matrix of particle classification in two steps. The classified particle type is plotted against the true particle type. The numbers in the squares indicate the migration probability of the particle type during the classification.

The migration matrix indicates the migration probability between the particle types due to the classification uncertainty. The probabilities that a particle is correctly classified can be taken from the diagonal entries of the matrix (proton: 63 %, helium: 47 % and iron: 82 %), while a proton is misclassified with a probability of 37 % (28 % as a helium and 9 % as an iron nucleus).

### 3.3 Energy Regression

After generating a sufficiently pure data set, the energy of the protons is estimated. For this purpose, a random forest is trained on the simulated proton subset. The attributes for the energy estimator are the same as the attributes for the particle classifier. In the case of the energy regressor, the migration probability from the true energy value to the estimated energy value can also be represented using a migration matrix. The migration matrix of the proton energy estimator is shown in figure 5. In general, a random forest can easily be trained for the other particle types, but in this work the focus is on the reconstruction of proton energies.

The migration matrix shows the characteristic cut-off at low estimated energies, since the random forest always returns the mean of true energies in a leaf of the estimator. That leads to the fact that the smallest estimated energy is always larger than the smallest true energy, so that with small energies an overestimation occurs.

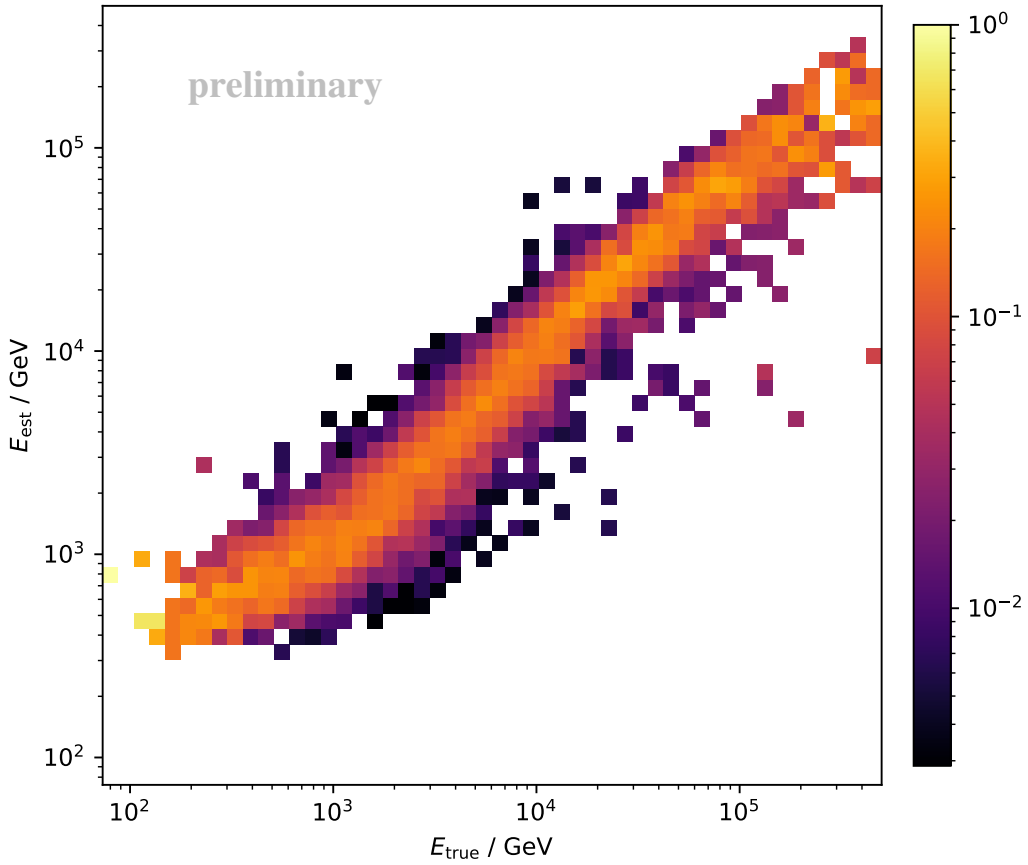


**Figure 5:** Migration matrix of the proton energy regressor. The estimated proton energy is plotted against the true energy of the proton.

At higher energies, the random forest underestimates the energies slightly and systematically. All in all, the migration matrix shows that the resolution of the random forest looks very promising and that the energies are not strongly scattered.

### 3.4 Unfolding

Both particle identification and energy reconstruction with the random forest approach turn out to be very promising. Using the migration matrices, the systematic and the statistical error in the reconstruction can be described quantitatively. A correction of these errors can be done by an unfolding. An unfolding is based on the following formula 1 on the next page.

One approach could be the estimation of background $b(E_{\text{est}})$ (particles wrongly classified as protons) and the efficiency $\epsilon(E_{\text{true}})$ of the protons (ratio of simulated protons and those that are triggered and well classified). With the protons' energy migration matrix $\mathbf{A}(E_{\text{est}}, E_{\text{true}})$ and the estimated energy distribution $g(E_{\text{est}})$ the true energy distribution $f(E_{\text{true}})$ can be unfolded.

$$g(y) = \int_{a_0}^{a_1} \mathbf{A}(y, x) \cdot f(x)dx + b(y) \tag{1}$$

Another approach would be a multi-dimensional migration matrix, which takes into account both the particle migration and the migrations of the energies. In this case, the background would be negligible and $g$ would depend on both, the energy $E_{\text{est}}$ and the particle type $p_{\text{est}}$, and $f$ would depend on the corresponding true values.

## References


[1] Maurin, D., Melot, F. and Taillet, R., *A database of charged cosmic rays*, *A&A* **569** (2014) A32.

[2] HEGRA collaboration, *The Cosmic ray proton spectrum determined with the imaging atmospheric Cherenkov technique*, *Phys. Rev. D* **59** (1999) 092003 [astro-ph/9901160].

[3] HEGRA collaboration, *Energy spectrum and chemical composition of cosmic rays between 0.3-PeV and 10-PeV determined from the Cherenkov light and charged particle distributions in air showers*, *Astron. Astrophys.* **359** (2000) 682 [astro-ph/9908202].

[4] MAGIC collaboration, *Protons spectrum from magic telescopes data*, in *Proceedings of the 37th International Cosmic Ray Conference*, vol. 395, 2021.

[5] MAGIC collaboration, *The major upgrade of the MAGIC telescopes, Part II: A performance study using observations of the Crab Nebula*, *Astropart. Phys.* **72** (2016) 76.

[6] D. Heck, G. Schatz, T. Thouw, J. Knapp and J. Capdevielle, *Corsika: A monte carlo code to simulate extensive air showers," report fzka 6019, forschungszentrum karlsruhe*, Tech. Rep. (1998).

[7] A. Moralejo, M. Gaug, E. Carmona, P. Colin, C. Delgado, S. Lombardi et al., *MARS, the MAGIC Analysis and Reconstruction Software*, *arXiv e-prints* (2009) .

[8] A.M. Hillas, *Cerenkov light images of eas produced by primary gamma*, in *Proceedings of the 19th International Cosmic Ray Conference*, vol. 3, pp. 445–448, 1985.

[9] M. Nöthe, K.A. Brügge and J.B. Buß, "aict-tools." 10.5281/zenodo.3338081.

[10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel et al., *Scikit-learn: Machine learning in Python*, *Journal of Machine Learning Research* **12** (2011) 2825.