

New approaches for gamma-hadron separation at the IceCube Neutrino Observatory

Federico Bontempo^{1,*} for the IceCube collaboration**

¹Karlsruhe Institute of Technology, Institute for Astroparticle Physics, 76021 Karlsruhe, Germany

Abstract. The IceCube Neutrino Observatory located at the geographic South Pole is composed of two detectors. One is the in-ice optical array, which measures high-energy muons from air showers and charged particles produced by the interaction of high-energy neutrinos in the ice. The other is an array of ice-Cherenkov tanks at the surface, called IceTop, which is used both as veto for the in-ice neutrino measurements and for detecting cosmic-ray air showers. In the next decade, the IceCube-Gen2 extension will increase the surface coverage including surface radio antennas and scintillator panels on the footprint of an extended optical array in the ice. The combination of the current surface and in-ice detectors can be exploited for the study of cosmic rays and the search for PeV gamma rays. The in-ice detector measures the high-energy muonic component of air showers, whereas the signal in IceTop is dominated by the electromagnetic component. The relative size of the muonic and electromagnetic components is different for gamma- and hadron-induced air showers. Thus, the gamma-hadron separation of cosmic rays is attempted using machine learning techniques including deep learning. Here, different approaches are presented. Finally, the prospects for the detection of PeV photons with IceCube-Gen2 will be discussed.

1 Introduction

The IceCube Neutrino Observatory [1] at the Amundsen–Scott South Pole Station in Antarctica is both, a neutrino and cosmic-ray detector, made by two arrays: one on the surface and one in the ice. The surface array, called IceTop [2], covers an area of about 1 km² and is composed of 81 stations with two ice-Cherenkov tanks per station. The in-ice array occupies a 1 km³ volume with 86 strings, each with 60 optical modules attached, which measure the Cherenkov light generated by high-energetic particles, such as muons.

The surface detector is constructed to be used as partial veto for the in-ice neutrino detection and for the measurement of cosmic-ray air showers. The veto capability allows to discriminate between muons that are generated in the atmosphere and the ones generated by neutrino interactions. For cosmic-ray studies, the secondary particles, generated by the interaction of the primary with air molecules in the atmosphere, are detected at the surface. IceTop measures primarily the electromagnetic component of the air shower, and through a log-likelihood minimization it is possible to reconstruct the primary information such as

*email: federico.bontempo@kit.edu

**<https://icecube.wisc.edu/collaboration/authors/>

energy and direction [2]. In addition, muons generated during the shower development with energy above a few hundreds GeV can be detected in the in-ice array.

2 Gamma-hadron separation

Air showers have different developments depending on the primary particle type. Showers with the same energy and direction can show differences in the shower development, for example, in the muon content and lateral spread. In particular, gamma-ray-induced air showers are expected to be almost exclusively electromagnetic thus producing fewer muons and have a narrower lateral spread with less shower-to-shower fluctuations. In contrast, hadronic air showers are richer in muon content and have a wider lateral spread with larger shower-to-shower fluctuations [3]. For the analysis presented here, they will be classified in two categories: gamma-ray or hadronic-induced air showers.

The two categories can be discriminated using the total charge measured in the in-ice detector as an estimator of the high-energy muon component of the air shower. Other parameters taken into account are the reconstructed energy-proxy, S_{125} , as described in the next section, and the reconstructed direction of the primary.

A previous search for PeV gamma rays in IceCube was performed using a random forest classifier [3]. It focused on two main topics: a search for a diffuse flux from the galactic plane and a point source search. The diffuse flux study set an upper limit on the galactic plane gamma-ray emission flux of $2.61 \times 10^{-19} \text{cm}^{-2} \text{s}^{-1} \text{TeV}^{-1}$ at a 90% confidence level. The sky search of gamma-ray point sources was performed after the selection of gamma-ray candidates and was found to be consistent with the background expectation [3].

In this work, a new approach with the same goals is now attempted encouraged by the recent results of the LHAASO collaboration: the first observation of a PeV photon and multiple ultra-high-energy gamma-ray sources [4]. The galactic plane, observed in the Northern Hemisphere, turned out to be rich in multiple TeV to PeV gamma-ray sources. Thus, sources might be expected also in the Southern Hemisphere. For this reason, an estimation of the flux of a potential LHAASO-like source in IceCube's field of view (FOV) is calculated. The flux of three LHAASO sources is extrapolated in the energy range $5.7 < \log_{10}(E/\text{GeV}) < 6.3$ and compared to the expected background, as shown in Figure 1, with an estimated 4° angular resolution for gamma-ray air showers. Four million cosmic-ray events are expected in this energy range per year at IceTop. The number of gamma rays expected in IceTop for a LHAASO-like source depends strongly on the flux assumption. A back of the envelope estimation gives approximately between 10 and 50 events per year. Thus, the expected flux of a LHAASO-like sources in IceCube is roughly 5 orders of magnitude lower than the cosmic-ray background flux.

The discovery of LHAASO-like sources with 10 years of data is expected with a significance of $N_s / \sqrt{N_b} > 5\sigma$, where N_s is the number of photons from the source and N_b is the estimated number of background events. Therefore, the separation between gamma rays and hadrons must be at least between 3 and 4 orders of magnitude. The previous analysis had a background suppression of approximately 2 orders of magnitude and therefore no sufficient sensitivity was reached. Two ways are currently explored for the background suppression. The first one is an ongoing study to improve the angular resolution, which will provide a better pointing to the potential sources, a smaller opening angle, and consequently, a reduced background. The second way is to discriminate between gamma-ray and hadron-induced air showers via the utilization of machine learning techniques. In this preliminary work, both a random forest regressor and a fully connected neural network are investigated [5]. The random forest regressor is a machine learning technique which consists of a combination of many simple decision trees with a binary splitting being combined together at the end of

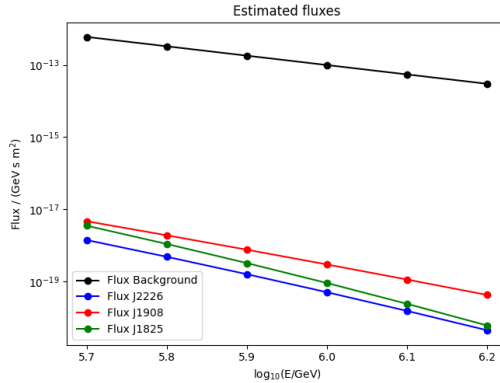


Figure 1: The extended flux of the 3 LHAASO sources [4] up to 2 PeV (colored lines), compared with the background (black line) is shown, an energy range $5.7 < \log_{10}(E/\text{GeV}) < 6.3$ and a 4 degree angular resolution for PeV photons at IceCube as circular region in the sky.

training, whereas the deep learning fully connected neural network uses weights and non-linear functions for the same purpose. For the discrimination of signal (gamma rays) and background (hadrons), both models have their output constrained between 0 and 1. Thus, it can be interpreted as a probability for the a posterior gamma-hadron-separation.

Both models are provided with the same input features. Namely, the total in-ice measured charge is used as the indicator for the total high-energy muonic component of the air shower. A containment parameter is utilized for an estimation on how contained the shower axis is in the in-ice detector. The shower size, S_{125} which is the signal expectation at a perpendicular reference distance of 125 meters from the shower axis [6], is taken as the energy estimator parameter. The sine of the declination angle is the indicator of the inclination of the air shower. Finally, a log-likelihood parameter is considered, which estimates how likely it is for an air shower to be induced by a gamma-ray or a hadronic primary [3].

Figure 2 shows the suppression of the background and the loss of signal for the two different approaches for the gamma-hadron separation. Figure 2a shows the performance of the random forest. The passing fraction of events is plotted as a function of S_{125} , the energy proxy, where $\log_{10}(S_{125}) = 0$ corresponds to approximately 1 PeV. The background is given by a small subset of data, which might contain a negligible fraction of gamma rays. In these plots the selection cut is set at 0.999, this means that the all the outputs with values less than the cut value are labeled as background and not passing the selection. Figure 2b uses the same data and input as the random forest, the only difference is that in this case a deep learning fully connected neural network is used for the discrimination. In both plots, the fraction of gamma rays passing the selection cut is approximately constant over the entire energy range. We find that, the deep learning model has a higher passing fraction compared to random forest. However, in order to compare the two methods up to the highest energies, a larger background testing sample is required to enhance the statistics in the relevant region.

3 Conclusion and Outlook

The combination of IceCube’s in-ice and surface detectors enables gamma-hadron separation, which can be performed using machine learning methods such as random forest or deep learning. This can be achieved by combining the reconstructed properties of the air shower

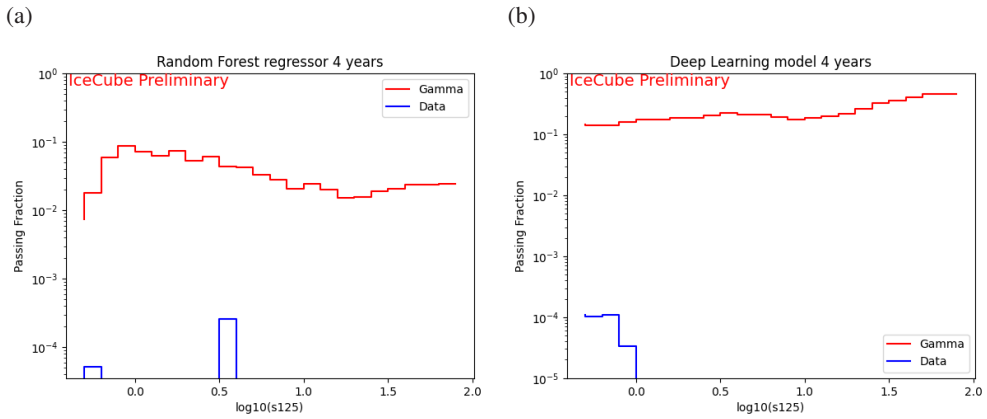


Figure 2: Passing fractions of gamma rays (red line) and data (blue line) as function of S_{125} . The two different models investigated for the gamma-hadron separation are shown. Both use the same input features and have the selection cut set at 0.999. The period of time taken into account is from 2012 to 2015. The missing line in Data means a zero value, obtained on the test sample which counts ~ 10.000 events per year.

such as energy and direction of the primary, the total in-ice charge for the estimation of high-energy muons, and by the log-likelihood parameter. The discovery of gamma-ray sources is expected with 10 years of data available, with a significance of 5σ , if the separation between gamma rays and hadrons is at least between 3 and 4 orders of magnitude.

In the near future, the planned extension of IceCube, called IceCube-Gen2 [7], will have 8 times more statistics for the current field of view (FoV) since it will cover a surface of $\sim 6 \text{ km}^2$ with 150 stations, each composed by 8 scintillator panels and 3 antennas. Together with the increased FoV, the overall aperture will increase by a factor of about 30. This will further increase the discovery potential for gamma-ray sources and the capability for the classification of the primary particle will improve via the combination of ice Cherenkov tanks, scintillator panels, and antennas, providing a better estimation of the muon content of the air showers.

References

- [1] M.G. Aartsen et al. (IceCube), *JINST* **12**, P03012 (2017), 1612.05093
- [2] R. Abbasi et al. (IceCube), *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **700**, 188 (2013)
- [3] M.G. Aartsen et al. (IceCube), *Astrophys. J.* **891**, 9 (2019), 1908.09918
- [4] Z. Cao et al. (LHAASO), *Nature* **594**, 33 (2021)
- [5] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, M. Chica-Rivas, *Ore Geology Reviews* **71**, 804 (2015)
- [6] M.G. Aartsen et al. (IceCube), *Phys. Rev. D* **100**, 082002 (2019), 1906.04317
- [7] M. Ackermann et al. (IceCube), *Journal of High Energy Astrophysics* **36**, 55 (2022)

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 802729).