

PIOTR KALACZYŃSKI

## RECONSTRUCTION OF MUON BUNDLES IN KM3NeT DETECTORS USING MACHINE LEARNING METHODS

**Abstract** *The KM3NeT Collaboration is installing the ARCA and ORCA neutrino detectors at the bottom of the Mediterranean Sea. The focus of ARCA is neutrino astronomy, while ORCA is optimised for neutrino oscillation studies. Both detectors are already operational in their intermediate states and collect valuable data, including the measurements of the muons produced by cosmic ray interactions in the atmosphere. This work explores the potential of machine learning models for the reconstruction of muon bundles, which are multi-muon events. For this, data collected with intermediate detector configurations of ARCA and ORCA was used in addition to simulated data from the envisaged final configurations of those detectors. Prediction of the total number of muons in a bundle as well as their total energy and even the energy of the primary cosmic ray is presented.*

**Keywords** machine learning, cosmic ray, KM3NeT, MUPAGE, CORSIKA, muon, multiplicity, neutrino, primary, energy, LightGBM, GZK, PMT, Cherenkov

**Citation** Computer Science 26(SI) 2025: 91–107

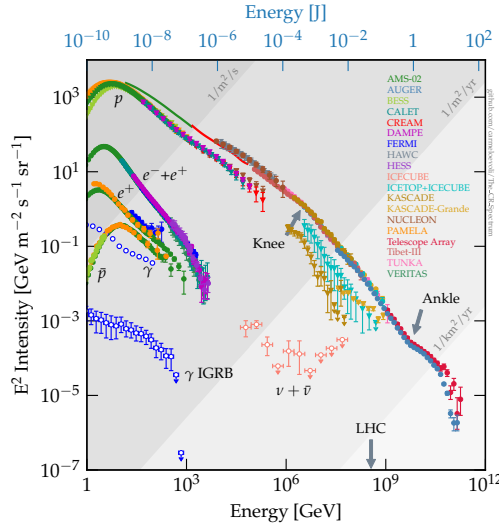
**Copyright** © 2025 Author(s). This is an open access publication, which can be used, distributed and reproduced in any medium according to the Creative Commons CC-BY 4.0 License.

## 1. Introduction

Atmospheric muons produced in cosmic ray (see Sec. 2) interactions to many may seem a *terra cognita*. Upon closer examination it becomes clear there is still much to be explored. The most prominent example is the (in-)famous Muon Puzzle [8,21]: experiments measuring the muon flux consistently report a deficit in muon events observed at high energies and high multiplicities (number of muons from a single cosmic ray interaction). Uncertainties alone cannot explain the discrepancy between the measurements and current state-of-the-art theoretical predictions. The KM3NeT observatory contributes to a better understanding of the problem by measuring the muon flux at the bottom of the Mediterranean Sea. A simulation workflow has been established to study the properties of muon bundles – groups of muons originating from the same primary cosmic ray. With a dedicated muon simulation, it was possible to train a machine learning model to reconstruct several important observables, such as the total number of muons in a bundle, their total energy and the energy of the primary cosmic ray. Measuring those quantities can help to pinpoint the shortcomings of the current theoretical models used to compute the muon flux.

## 2. Cosmic rays

The existence of cosmic radiation was discovered by Victor Hess in 1912 following a series of balloon-borne radiation measurements [17]. Since then, a wide range of experiments have embarked on a quest for better understanding of the nature and origin of cosmic rays (CRs) and pushed to improve the precision of CR measurements [1, 2, 4, 9, 22].



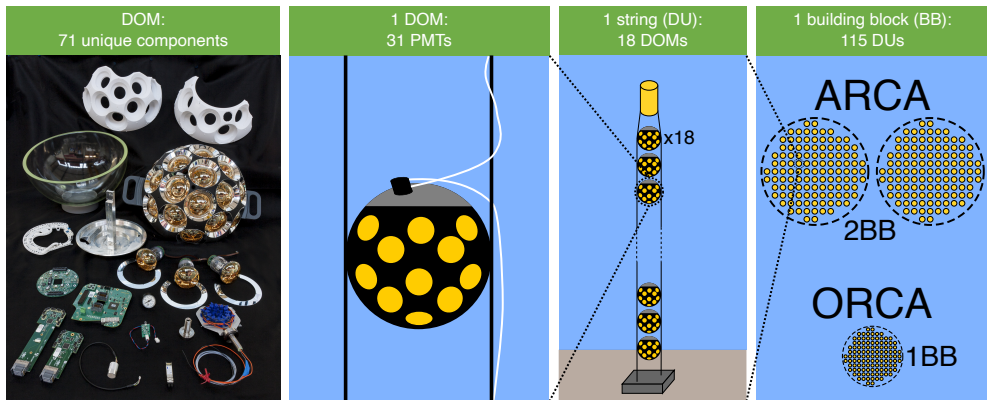
**Figure 1.** Energy spectra of various CR primary particles measured by several experiments.

The “LHC” arrow indicates the kinetic energy achievable with the Large Hadron Collider (as of 2023). The figure was adapted from [12]

A CR is a high-energy particle or atomic nucleus arriving at Earth from outer space. As shown in Figure 1, the vast majority of primary CRs are protons, followed by electrons, positrons and photons. A tiny contribution from the cosmic neutrinos has been observed as well. Direct measurements of primary CRs are only possible with detectors placed at extremely high altitudes. What can be observed at the Earth's surface and below are the products of CR interactions with the atmospheric molecules. Such collisions result in a cascade of particles called an extensive air shower (EAS). A typical EAS starts at an altitude of approximately 30 km; some of the secondary particles will reach sea level and even some distance underground or underwater. The most penetrant secondaries are muons and neutrinos, which are studied using the KM3NeT telescopes.

### 3. The KM3NeT experiment

KM3NeT [3] is a network of two neutrino telescopes at the bottom of the Mediterranean Sea, currently under construction: ARCA near Portopalo di Capo Passero in Italy and ORCA offshore Toulon in France. They share the same hardware, however, their physics focus and thus geometry differ. ARCA is located at a depth of 3500 m and is the more sparse of the two detectors, allowing for the study of cosmic neutrinos in the TeV-PeV energy range. ORCA is placed shallower, at 2500 m underwater. Its more dense instrumentation allows for precise studies of atmospheric neutrino oscillations in the GeV-TeV energy range. The design of the KM3NeT detectors is summarised in Figure 2.



**Figure 2.** Pictorial summary of the design of KM3NeT neutrino telescopes.

From left to right: the components of an optical module, the module integrated into a detection unit, an artist's impression of a detection unit and a building block

The optical modules are composed of many elements; the most important are the photomultiplier tubes (PMTs), which measure the Cherenkov light. Muons and other

electrically charged products of neutrino interactions travelling faster than light in water cause emission of Cherenkov photons by the water molecules upon deexcitation. Each optical module contains 31 PMTs facing different directions, which allows for directional reconstruction [5]. A vertical line composed of 18 buoyant DOMs, anchored to the seabed and kept straight by a buoy is called a detection unit. A building block of a detector consists of 115 detection units. ORCA will comprise a single building block, while for ARCA two building blocks are foreseen. Separation into two separate blocks instead of a single larger one has been driven by cost-efficiency considerations rather than a specific physics goal.

The main interest of KM3NeT is to observe neutrinos. However, most of the collected data consists of muons produced in EASs. The muons do not only constitute the main background for neutrino measurements due to a similar signal produced in the detector but also present a valuable opportunity for indirect study of the primary CRs from which they originate.

## 4. Muon simulation

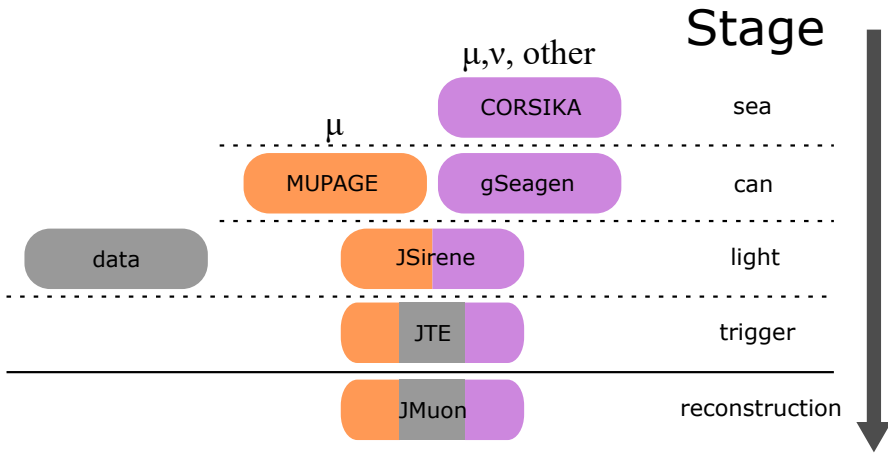
In KM3NeT, two standard muon simulation tools are developed: MUPAGE [10] and CORSIKA [16]. Both were used for this work and are briefly described in what follows.

MUPAGE (atmospheric MUons from PARametric formulas: a fast GENerator for neutrino telescopes) is a code inherited by KM3NeT from the ANTARES experiment. It extensively makes use of parametrised functions, which can be tuned to experimental data and simulations. The original tuning was performed based on the HEMAS [24] Monte Carlo (MC) simulation code and data from the MACRO experiment at Gran Sasso in Italy [14]. The output of MUPAGE consists of muon bundle events sampled on a cylindrical volume surrounding the detector, called the can.

CORSIKA (COsmic Ray SIMulations for KASCADE) originated as an MC simulation code for the KASCADE experiment [9]. With time, it evolved into a state-of-the-art solution for simulating cosmic ray air showers. The code generates EASs starting from the very beginning, i.e. from the first interaction of the primary particle in the atmosphere. It includes a plethora of customization possibilities such as the geometry of the problem, the assumed atmospheric density profile, the choice of low- and high-energy hadronic interaction models and more. Although currently only used for muon simulations by KM3NeT, CORSIKA is not limited to this. All particles produced in the shower are available to the user at a specified observation level (sea level in the case of KM3NeT simulations).

In Figure 3, the processing chain of the muon simulation is shown. As can be seen, the starting points differ, however, both simulation codes and the experimental data go through certain common stages. The sea level is unique to CORSIKA simulation and to transport the muons to the can, the gSeaGen [6] software was adapted by the author of this work. It is the subject of an upcoming paper “gSeaGen code by KM3NeT: an efficient tool to propagate muons simulated with CORSIKA”. JSirene [18] is an application developed to simulate photon emission and detection

by the PMTs, including the signal from environmental background. The JTE code selects potentially interesting events, i.e. performs triggering. Finally, the observables are reconstructed for each event using a specialized algorithm. Here, the standard KM3NeT tool for muon track reconstruction is JMuon. It is a multi-stage fit based on a maximum likelihood approach, assuming a single muon hypothesis. This paper presents a novel reconstruction method utilizing machine learning techniques to predict the muon bundle energy as well as the energy of the primary cosmic ray and the muon multiplicity.



**Figure 3.** The processing chain for muon simulations in KM3NeT. The colour coding distinguishes three possibilities: CORSIKA simulation marked in purple, MUPAGE simulation in orange and experimental data indicated by the grey colour. The order of execution is from top to bottom, along the vertical arrow. At the trigger stage, both simulations and experimental data can be compared against each other

## 5. Reconstruction of muon bundle properties

In this section, the reconstruction of three muon bundle observables is presented: bundle energy, primary CR energy and muon multiplicity. The data used to produce the results is summarized in Table 1. The reconstruction was performed separately for two detector configurations for both ARCA and ORCA. In each case, it was a full building block (ARCA115, ORCA115) and an intermediate state with 6 detection units installed (ARCA6, ORCA6).

The data preprocessing included extracting 46 features (described in detail in [19]), event weights and target (observable) values. Each feature had its mean removed and was scaled to unit variance (standard scaling). This was done for each detector configuration before the division into training, validation and test datasets. Event weight scaling was also considered but proved not to improve the training.

Before reaching the final result of reconstruction, several loops of training and validation were performed to select the best machine learning (ML) regression model for the task, pick the features that improve the performance and tune the hyperparameters (internal parameters of the model). The regressor selection was performed only for bundle energy; the result was used for primary CR energy and muon multiplicity reconstruction. The outcome for each of the observables is presented in the following subsections.

**Table 1**

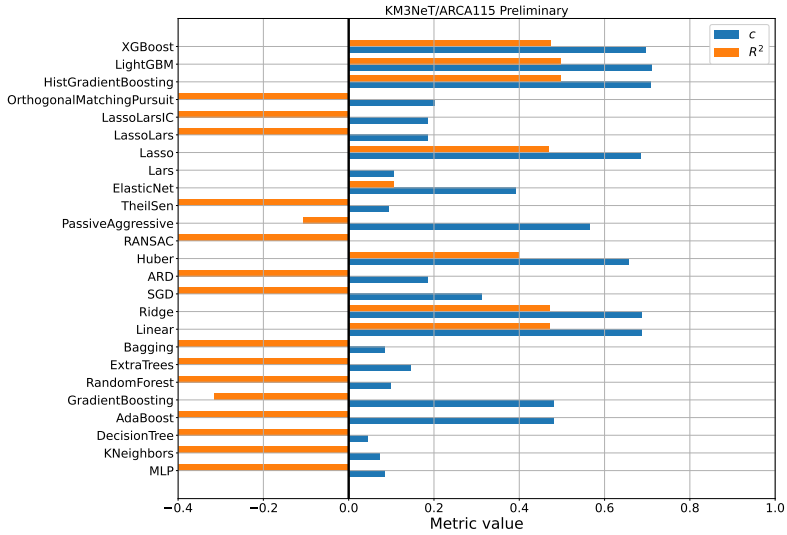
Summary of the datasets used to train and evaluate the machine learning models. The simulated data was split between training, validation and test datasets in proportion 64 : 16 : 20 and was extracted from the official KM3NeT CORSIKA simulation produced by the author of this work. The experimental data comes from the measurements taken by ARCA6 and ORCA6 detectors in 2020 and 2021

	ARCA115	ARCA6	ORCA115	ORCA6
Training data	14 374 415	3 894 293	15 439 946	6 154 895
Validation data	3 598 133	974 371	3 857 715	1 539 988
Test data	4 489 177	1 215 552	4 825 073	1 923 475
Experimental data	×	3 013 982	×	4 574 887

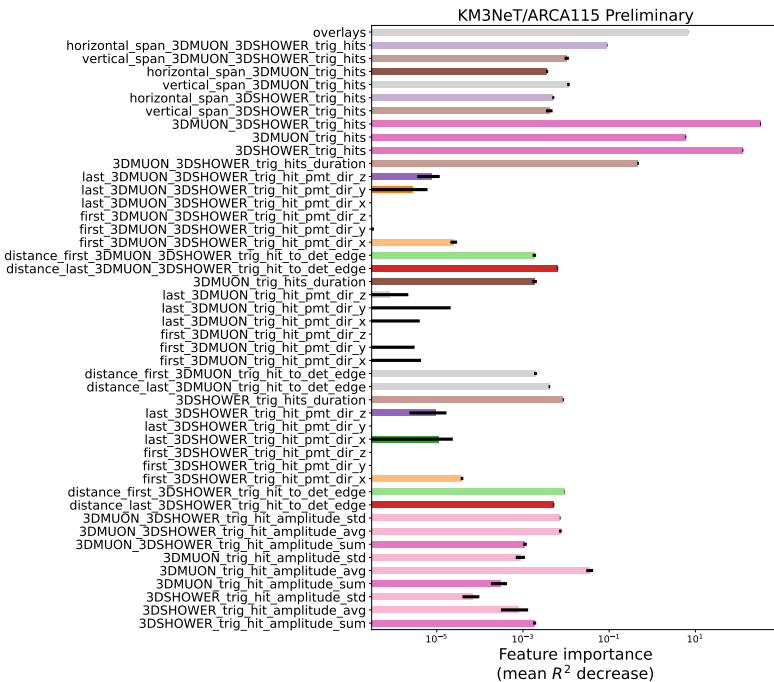
### 5.1. Muon bundle energy

The first step in determining the muon bundle energy was the selection of the best ML model. A range of regression models available in the Scikit-learn library [23], as well as LightGBM [20] and XGBoost [11] were considered. The criteria for determining the best model were the weighted coefficient of determination ( $R^2$ -score) and weighted Pearson correlation coefficient  $c$ . As demonstrated in Figure 4, the best performer was LightGBM, a histogram-based gradient-boosting model. Not only was it superior in terms of achieved scores, but thanks to working on the histograms rather than on the events themselves, it turned out to be one of the fastest models as well (see [19]). The negative  $R^2$  despite quite high  $c$  value in case of some models is a result of the difference in focus of the two metrics. The  $R^2$ -score explicitly tests the difference between the truth and prediction, while the correlation coefficient only checks whether their relation is close to linear. Examples of 2D histograms with such negative  $R^2$  can be found in the appendix of [19].

The next step was to evaluate the usefulness of individual features. Not all were equally important and several were strongly inter-correlated, which could indicate redundancy. The feature importance was defined as the mean decrease of the  $R^2$ -score upon removal of a given feature. The results for ARCA115 are shown in Figure 5.



**Figure 4.** Different machine learning models in terms of achieved weighted  $R^2$ -score and weighted Pearson correlation coefficient  $c$ . The models were trained on a fraction (50k) of the ARCA115 training dataset because of the infeasibly long runtime and RAM memory required by some of the models

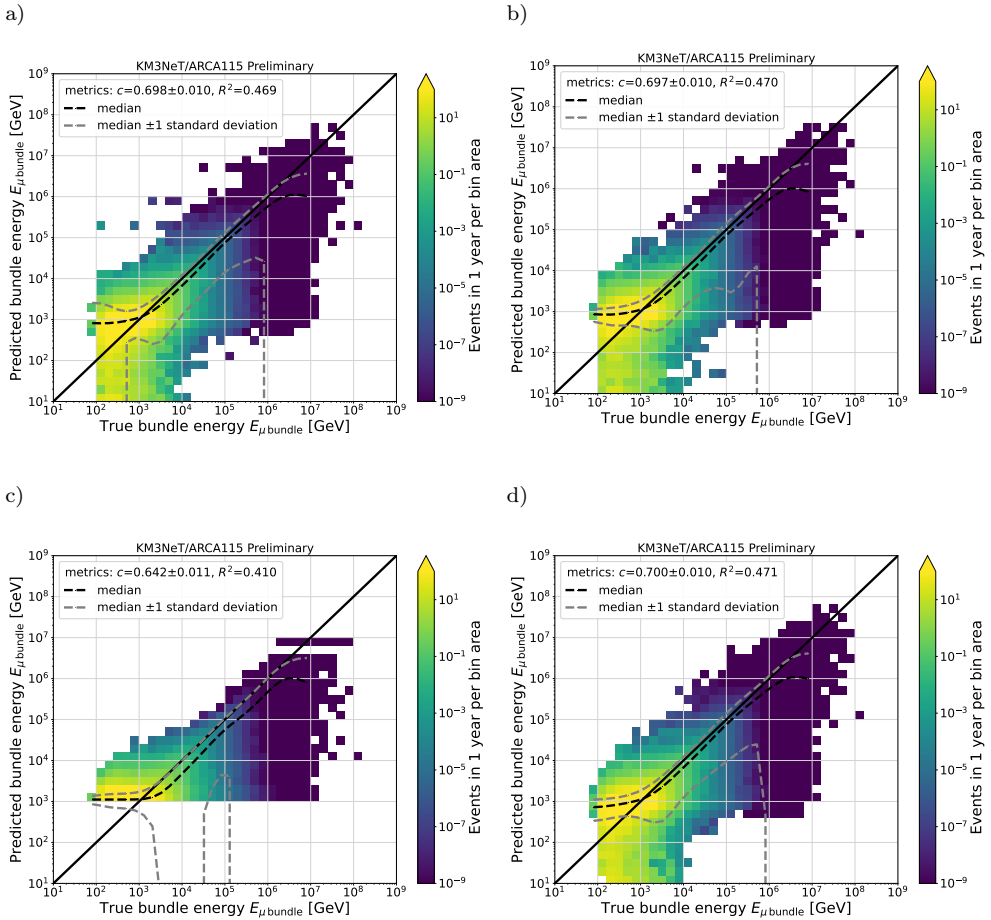


**Figure 5.** Feature importance computed over 10 random trials of removing a given feature. The colour coding groups features into clusters based on their inter-correlation

Four different approaches to feature selection were considered:

1. Using all available features.
2. Taking only features with positive importance and selecting the most important feature from each cluster.
3. Using only the single most important feature.
4. Requiring positive importance for all features.

The comparison between those four categories for ARCA115 is presented in Figure 6.

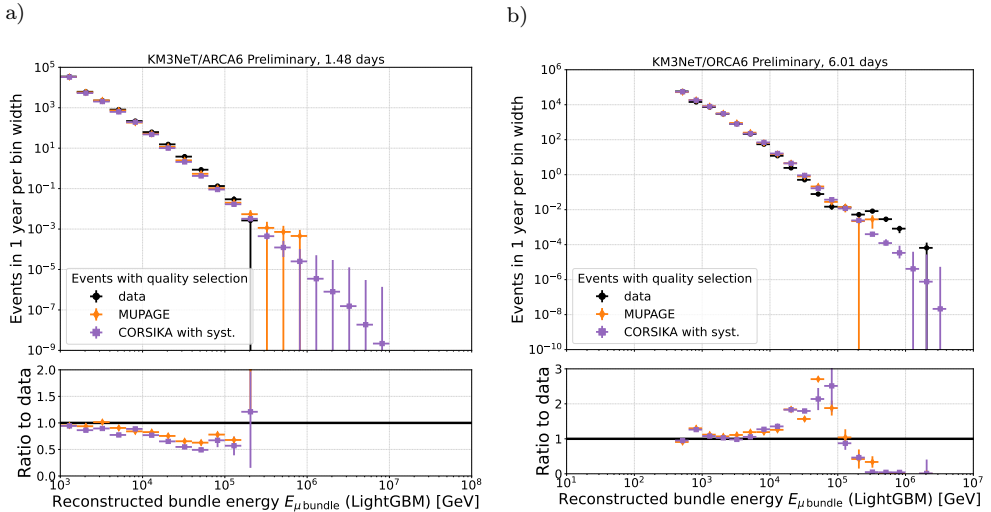


**Figure 6.** Comparison of predicted and true bundle energy for different feature selections for the ARCA115 configuration: all features (a); selected features (importance & clustering) (b); only 3DMUON\_3DSHOWER\_trig\_hits (c); selected features (positive importance) (d). The training target for the reconstruction was the bundle energy, not its logarithm. The logarithmic scale is only used for clearer visualisation

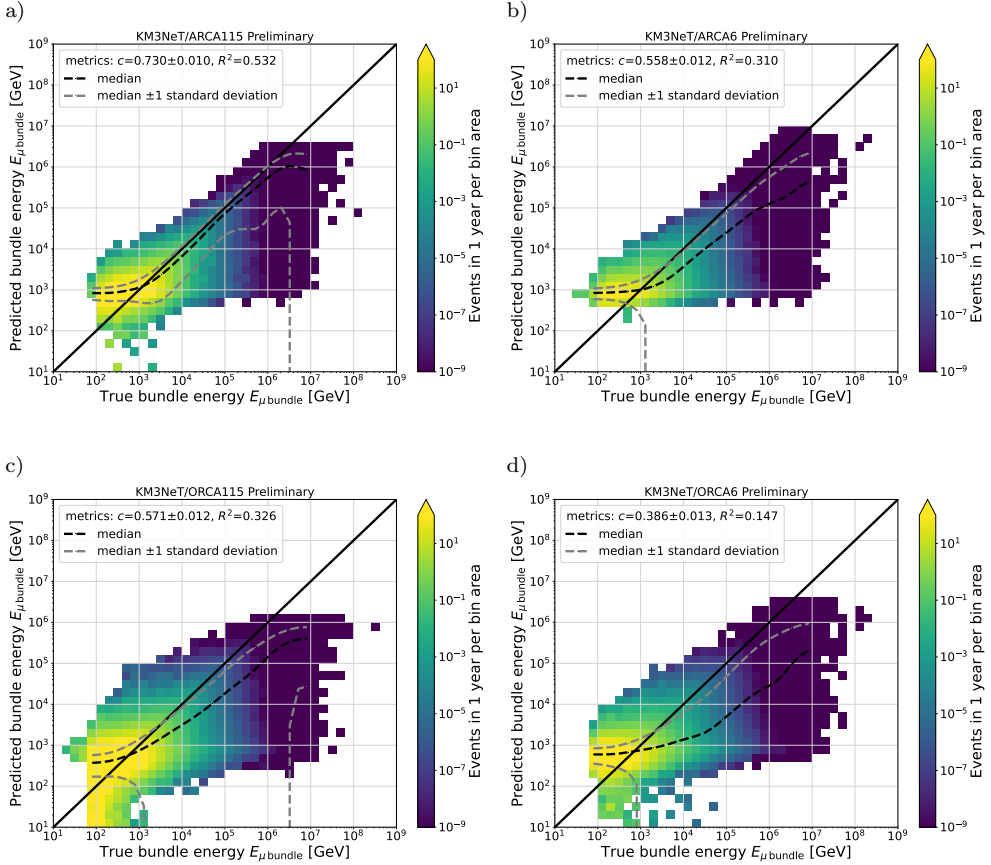
The feature selection method requiring the positive importance of each feature provided the best results. Interestingly, even with a single feature (Fig. 6c), it was possible to achieve a reasonable performance. There was no hope that this task could be reliably done just by counting the number of hits (the feature used is the number of triggered hits), however, it constituted an important sanity check, as hit counting is the basis of most of the energy reconstruction algorithms.

After performing the hyperparameter tuning using the Optuna library [7], the final results presented in Figure 8 were produced for the ARCA115, ARCA6, ORCA115 and ORCA6 configurations. Comparison of Figure 8a against Figure 6d shows that tuning is an important step, which can significantly improve the outcome. As expected, the bundle energy reconstruction is most accurate for ARCA115 – this configuration instruments the largest volume and hence can observe the largest energy deposits. There is a threshold energy of about 1 TeV below which the model cannot make reasonable predictions. Towards the highest energies, there is a growing underestimation of the true bundle energy. This might result from the smaller number of such events in the MC simulation and the smaller chance for such events to be fully contained within the can.

The comparison of the experimental data against the MC simulations for the reconstructed bundle energy is shown in Figure 7. The simulations reproduce the data within the few-TeV energy scale, however, at higher energies discrepancies are visible, similarly as in Figures 10 and 12. The consistency between MUPAGE and CORSIKA behaviour suggests a common issue for both generators, whether it is the theoretical inputs, the way the simulated events are processed or something else.



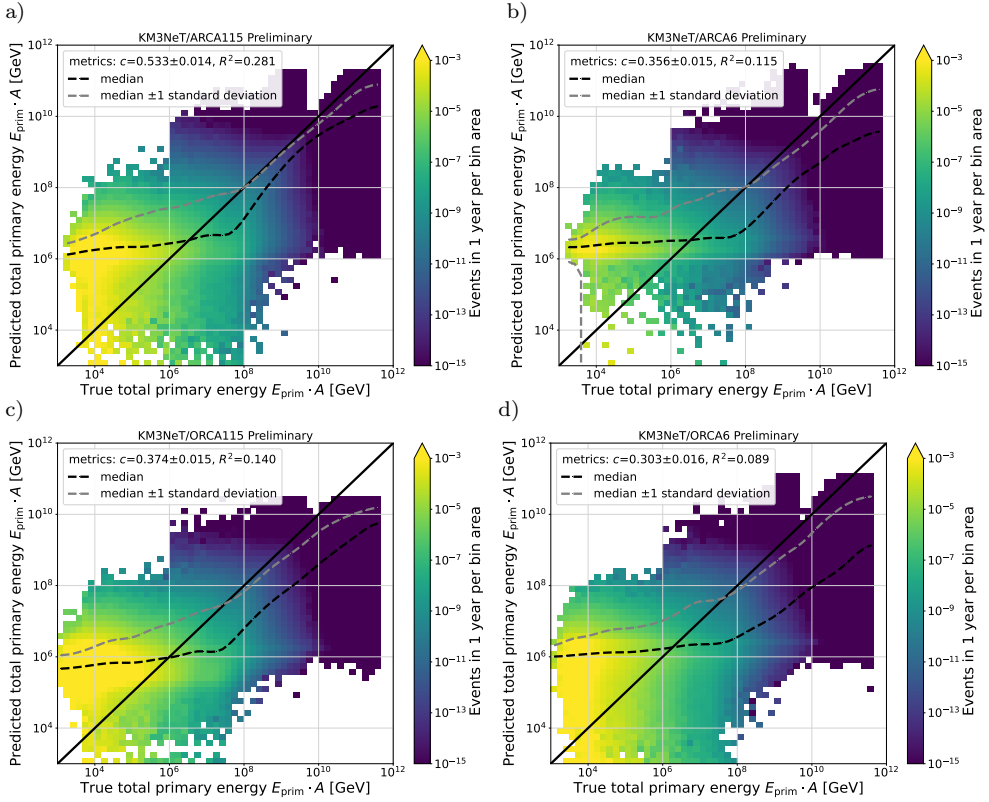
**Figure 7.** ML-based bundle energy reconstruction results for ARCA6 (a) and ORCA6 (b) data. The quality selection is described in detail in [19]



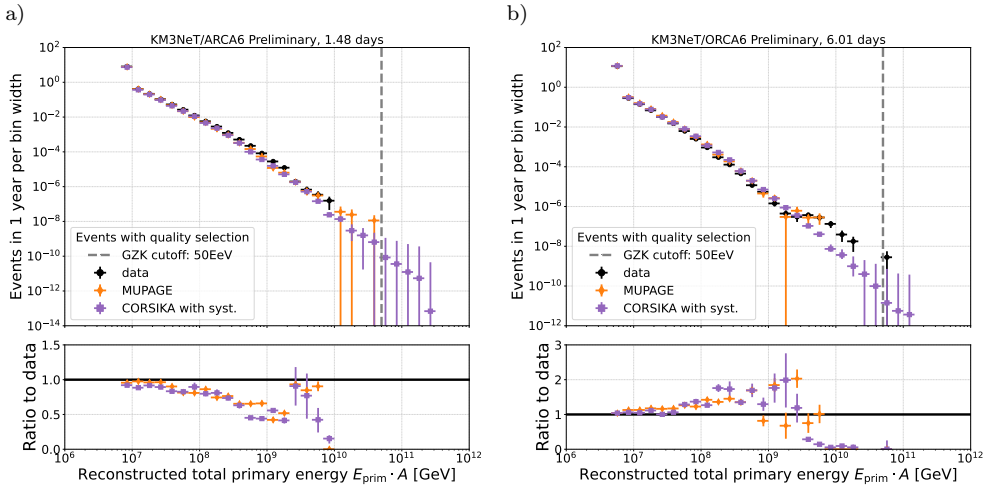
**Figure 8.** Comparison of reconstructed and true muon bundle energy  $E_{\mu}$  for ARCA115 (a), ARCA6 (b), ORCA115 (c) and ORCA6 (d)

## 5.2. Primary cosmic ray energy

The outcome of the previous subsection encouraged an attempt to reconstruct the primary CR energy. The same procedure was followed, with the same feature selection but a different hyperparameter tuning. Despite the bad performance, the results reported in Figure 9 look promising above 100 PeV, given how much of the original information is lost when the EAS has reached the detector. It suggests that KM3NeT may contribute to the primary CR measurements at the highest energies, notably to the study of the Greisen-Zatsepin-Kuzmin (GZK) cutoff [15, 25]. The data from ARCA6 and ORCA6 used in this work is insufficient to make any decisive statements about the (non-)observation of the GZK cutoff (see Fig. 10), both due to limited exposure time and detector size.



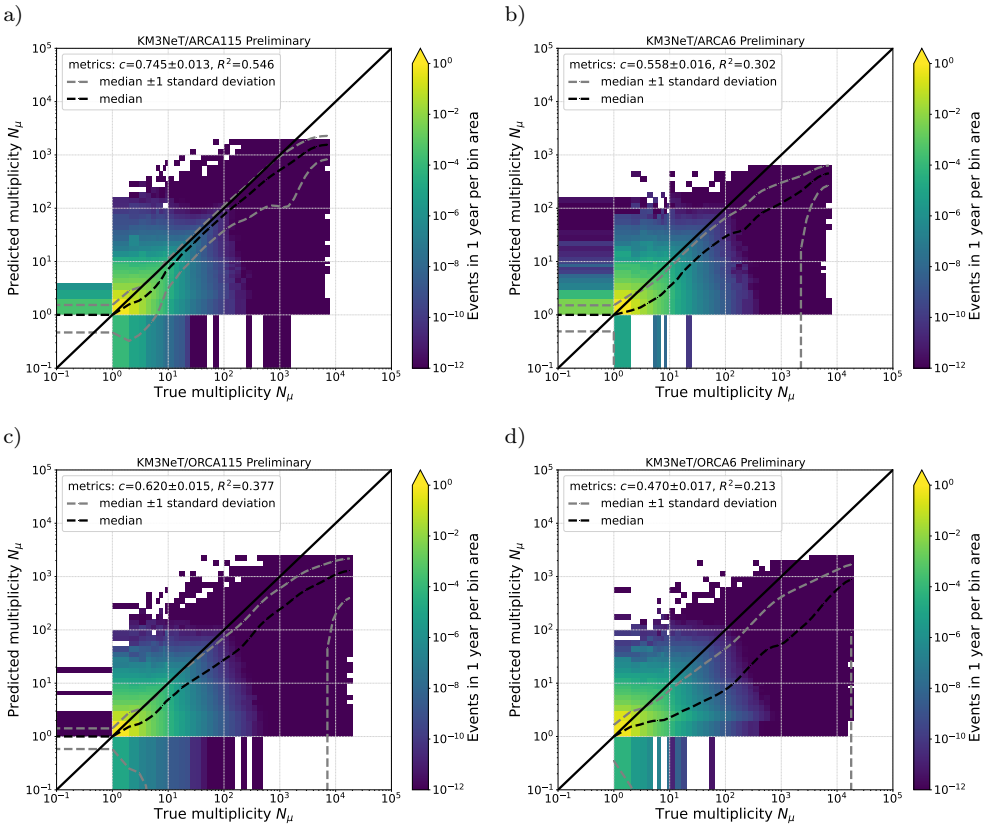
**Figure 9.** Comparison of reconstructed and true total primary energy for ARCA115 (a), ARCA5 (b), ORCA115 (c) and ORCA6 (d)



**Figure 10.** ML-based primary CR energy reconstruction results for ARCA6 (a) and ORCA6 (b) data. The quality selection is described in detail in [19]

### 5.3. Muon multiplicity

Counting the muons in the bundle is a challenging task. Some muons may not be energetic enough or only skim the edge of the sensitive volume, resulting in insufficient illumination of the PMTs. To account for this, the multiplicity definition has been adapted by counting only the muons with energy above 120 GeV for ARCA and 1 GeV for ORCA. An additional spatial criterion was found beneficial for ARCA6, requiring that the closest distance from the muon trajectory to the can centre is at most 269.4 m and the muon pathlength inside the can is at least 240 m. Details are provided in [19]. Such a muon selection results in the possibility of having bundles with multiplicity equal to 0, which is indeed the case, except for ORCA6 (see Fig. 11).



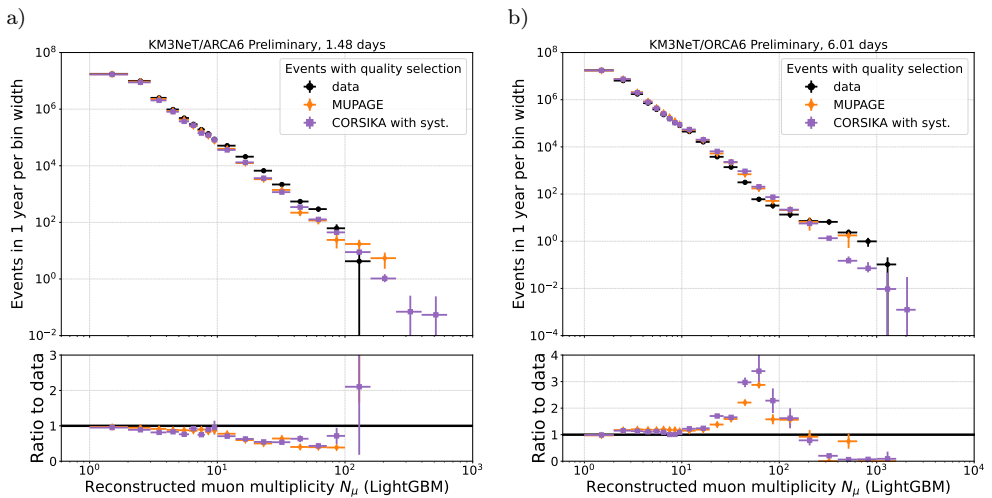
**Figure 11.** Comparison of muon multiplicity reconstructed with the LigthGBM model against the true value for ARCA115 (a), ARCA5 (b), ORCA115 (c) and ORCA6 (d).

Both metrics were computed on the bin values, not individual events

Reconstructing such events proves troublesome for the same reasons that led to discarding the muons in the first place. Nevertheless, the overall performance is im-

proved thanks to more accurate prediction for high-multiplicity events. The multiplicity reconstruction is more accurate for ARCA, thanks to the typical muon signature in a detector (elongated track). For each detector configuration, the best performance was observed for low (except for 0) and intermediate multiplicity, with a growing underprediction for events with a larger number of muons (with similar reasons as for the muon bundle energy).

The application of the trained models to the experimental data is presented in Figure 12. There is good agreement between the data and MC simulations for bundles with multiplicity between 1 and 10. For MUPAGE and CORSIKA at higher multiplicities, the comparison between data and MC simulations indicates a need for further work on CR shower simulations to understand the underlying problem.



**Figure 12.** ML-based energy reconstruction results for ARCA6 (a) and ORCA6 (b) data. The quality selection is described in detail in [19]

## 6. Conclusion

The partial KM3NeT detectors already collect data containing large numbers of muons created in air showers. They could be studied thanks to simulations generated with CORSIKA. Reconstruction of the energy of muon bundles and primary CR primaries using machine learning models is promising.

The potential for further improvement lies in achieving a better understanding of the discrepancies between the data and MC. The model training could potentially benefit from extracting features from even lower-level data to ensure minimal information loss. Considering other machine learning models, like graph neural networks [13], might yield valuable results. For full detector configurations, an attempt to reconstruct individual muon tracks may be considered.

## Acknowledgements

*The authors acknowledge the financial support of the funding agencies:*

*Funds for Scientific Research (FRS-FNRS), Francqui foundation, BAEF foundation.*

*Czech Science Foundation (GAĎR 24-12702S).*

*Agence Nationale de la Recherche (contract ANR-15-CE31-0020), Centre National de la Recherche Scientifique (CNRS), Commission Européenne (FEDER fund and Marie Curie Program), LabEx UnivEarthS (ANR-10-LABX-0023 and ANR-18-IDEX-0001), Paris Île-de-France Region, Normandy Region (Alpha, Blue-waves and Neptune), France.*

*Shota Rustaveli National Science Foundation of Georgia (SRNSFG, FR-22-13708), Georgia.*

*The General Secretariat of Research and Innovation (GSRI), Greece.*

*Istituto Nazionale di Fisica Nucleare (INFN) and Ministero dell'Università e della Ricerca (MUR), through PRIN 2022 program (Grant PANTHEON 2022E2J4RK, Next Generation EU) and PON R&I program (Avviso n. 424 del 28 febbraio 2018, Progetto PACK-PIR01 00021), Italy; IDMAR project Po-Fesr Sicilian Region az. 1.5.1; A. De Benedittis, W. Idrissi Ibsalikh, M. Bendahman, A. Nayerhoda, G. Papalashvili, I. C. Rea, A. Simonelli have been supported by the Italian Ministero dell'Università e della Ricerca (MUR), Progetto CIR01 00021 (Avviso n. 2595 del 24 dicembre 2019); KM3NeT4RR MUR Project National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 3.1, Funded by the European Union – NextGenerationEU, CUP I57G2100040001, Concession Decree MUR No. n. Prot. 123 del 21/06/2022.*

*Ministry of Higher Education, Scientific Research and Innovation, Morocco, and the Arab Fund for Economic and Social Development, Kuwait.*

*Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO), the Netherlands.*

*The National Science Centre, Poland, grant number 2021/41/N/ST2/01177. We gratefully acknowledge the funding support by program “Excellence initiative-research university” for the AGH University in Krakow as well as the ARTIQ project: UMO-2021/01/2/ST6/00004 and ARTIQ/0004/2021. The grant “AstroCeNT: Particle Astrophysics Science and Technology Centre”, carried out within the International Research Agendas programme of the Foundation for Polish Science financed by the European Union under the European Regional Development Fund.*

*Ministry of Research, Innovation and Digitalisation, Romania.*

Slovak Research and Development Agency under Contract No. APVV-22-0413; Ministry of Education, Research, Development and Youth of the Slovak Republic.

MCIN for PID2021-124591NB-C41, -C42, -C43 and PDC2023-145913-I00 funded by MCIN/AEI/10.13039/501100011033 and by “ERDF A way of making Europe”, for ASFAE/2022/014 and ASFAE/2022/023 with funding from the EU NextGenerationEU (PRTR-C17.I01) and Generalitat Valenciana, for Grant AST22\_6.2 with funding from Consejería de Universidad, Investigación e Innovación and Gobierno de España and European Union – NextGenerationEU, for CSIC-INFRA23013 and for CNS2023-144099, Generalitat Valenciana for CIDE-GENT/2018/034, /2019/043, /2020/049, /2021/23, for CIDEIG/2023/20 and for GRISOLIAP/2021/192 and EU for MSC/101025085, Spain.

Khalifa University internal grants (ESIG-2023-008 and RIG-2023-070), United Arab Emirates.

The European Union’s Horizon 2020 Research and Innovation Programme (ChETEC-INFRA – Project no. 101008324).

## References

- [1] Abbasi R.U., Abe M., Abu-Zayyad T., Allen M., Azuma R., Barcikowski E., Belz J.W., et al. (Telescope Array Collaboration): Study of muons from ultrahigh energy cosmic ray air showers measured with the Telescope Array experiment, *Physical Review D*, vol. 98, 022002, 2018. doi: 10.1103/PhysRevD.98.022002.
- [2] Abdul Halim A., Abreu P., Aglietta M., Allekotte I., Almeida Cheminant K., Almela A., Aloisio R., et al.: Constraining models for the origin of ultra-high-energy cosmic rays with a novel combined analysis of arrival directions, spectrum, and composition data measured at the Pierre Auger Observatory, *Journal of Cosmology and Astroparticle Physics*, vol. 2024(01), 022, 2024. doi: 10.1088/1475-7516/2024/01/022.
- [3] Adrián-Martínez A., Ageron M., Aharonian F., Aiello S., Albert A., Ameli F., Anassontzis E.: Letter of intent for KM3NeT 2.0, *Journal of Physics G: Nuclear and Particle Physics*, vol. 43(8), 084001, 2016. doi: 10.1088/0954-3899/43/8/084001.
- [4] Aguilar M., Ali Cavazonza L., Ambrosi G., Arruda L., Attig N., Barao F., Barri L.: The Alpha Magnetic Spectrometer (AMS) on the international space station: Part II – Results from the first seven years, *Physics Reports*, vol. 894, pp. 1–116, 2021. doi: 10.1016/j.physrep.2020.09.003.
- [5] Aiello S., Albert A., Alshamsi M., Alves Garre S., Aly Z., Ambrosone A., Ameli F.: The KM3NeT multi-PMT optical module, *Journal of Instrumentation*, vol. 17(07), P07038, 2022. doi: 10.1088/1748-0221/17/07/P07038.

- [6] Aiello S., Albert A., Alves Garre S., Aly Z., Ameli F., Andre M., Androulakis G.: gSeaGen: The KM3NeT GENIE-based code for neutrino telescopes, *Computer Physics Communications*, vol. 256, 107477, 2020. doi: <https://doi.org/10.1016/j.cpc.2020.107477>.
- [7] Akiba T., Sano S., Yanase T., Ohta T., Koyama M.: Optuna: A next-generation hyperparameter optimization framework. In: *KDD'19: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ACM, New York, 2019. doi: 10.1145/3292500.3330701.
- [8] Albrecht J., Cazon L., Dembinski H., Fedynitch A., Kampert K.-L., Pierog T., Rhoe W.: The Muon Puzzle in cosmic-ray induced air showers and its connection to the Large Hadron Collider, *Astrophysics and Space Science*, vol. 367(3), 27, 2022. doi: 10.1007/s10509-022-04054-5.
- [9] Apel W.D., Arteaga-Velázquez J.C., Bekk K., Bertaina M., Blümer J., Bozdog H., Brancus I.M.: KASCADE-Grande measurements of energy spectra for elemental groups of cosmic rays, *Astroparticle Physics*, vol. 47, pp. 54–66, 2013. doi: 10.1016/j.astropartphys.2013.06.004.
- [10] Carminati G., Bazzotti M., Biagi S., Cecchini S., Chiarusi T., Margiotta A., Sioli M., *et al.*: MUPAGE: a fast atmospheric MUon GEnerator for neutrino telescopes based on PArametric formulas, *ArXiv preprint, arXiv:09075563v*, 2009. doi: 10.48550/arXiv.0907.5563.
- [11] Chen T., Guestrin C.: XGBoost: A Scalable Tree Boosting System. In: *KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, ACM, New York, NY, USA, 2016. doi: 10.1145/2939672.2939785.
- [12] Evoli C.: The Cosmic-Ray Energy Spectrum, 2020. doi: 10.5281/zenodo.4396125.
- [13] Filippini F., Androutsou E., Domi A., Spisso B., Drakopoulou E.: Data reconstruction and classification with Graph neural networks in KM3NeT/ARCA, *PoS*, vol. 444, 1194, 2023. doi: 10.22323/1.444.1194.
- [14] Giacomelli G., Margiotta A.: The MACRO Experiment at Gran Sasso. In: *Charles Peck Fest, ArXiv preprint, arXiv:07071691v1*, 2007. doi: 10.48550/arXiv.0707.1691.
- [15] Greisen K.: End to the Cosmic-Ray Spectrum?, *Physical Review Letters*, vol. 16(17), pp. 748–750, 1966. doi: 10.1103/physrevlett.16.748.
- [16] Heck D., Knapp J., Capdevielle J.N., Schatz G., Thouw T.: CORSIKA: A Monte Carlo code to simulate extensive air showers, Forschungszentrum Karlsruhe GmbH, Karlsruhe, 1998.
- [17] Hess V.F.: Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten, *Physikalische Zeitschrift*, vol. 13, pp. 1084–1091, 1912.
- [18] de Jong M.: Multi-dimensional interpolations in C++, *ArXiv preprint, arXiv:190702597v1*, 2019. doi: 10.48550/arXiv.1907.02597.
- [19] Kalaczyński P.: *The Measurement and Modelling of Cosmic Ray Muons at KM3NeT Detectors*, Ph.D. thesis, NCBJ, *ArXiv preprint, arXiv:240202620v1*, 2024. doi: 10.48550/arXiv.2402.02620. 2402.02620.

- [20] Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., *et al.*: LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (eds.), *Advances in Neural Information Processing Systems*, vol. 30, Curran Associates, Inc., 2017. <https://proceedings.neurips.cc/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf>.
- [21] KM3NeT Collaboration, Aiello S., Albert A., Alshamsi M., Alves Garre S., Ambrosone A., Ameli F., Andre M., *et al.*: Atmospheric muons measured with the KM3NeT detectors in comparison with updated numeric predictions, *The European Physical Journal C*, vol. 84(7), 696, 2024. doi: 10.1140/epjc/s10052-024-13018-8.
- [22] Mikhailov V.V., Adriani O., Bazilevskaya G.A., Barbarino G.C., Bellotti R., Bogomolov E.A., Boezio M.: Galactic Cosmic Ray Electrons and Positrons over a Decade of Observations in the PAMELA Experiment, *Bulletin of the Russian Academy of Sciences: Physics*, vol. 83(8), pp. 974–976, 2019. doi: 10.3103/S1062873819080288.
- [23] Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., *et al.*: Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [24] Scapparone E.: HEMAS: A Monte Carlo code for hadronic, electromagnetic and TeV muon components in air shower, *ArXiv preprint, arXiv:physics/9902043v1*, pp. 238–246, 1998. doi: 10.48550/arXiv.physics/9902043.
- [25] Zatsepin G.T., Kuz'min V.A.: Upper Limit of the Spectrum of Cosmic Rays, *Soviet Journal of Experimental and Theoretical Physics Letters*, vol. 4, 1966. <https://ui.adsabs.harvard.edu/abs/1966JETPL...4...78Z>.

## Affiliations

### Piotr Kalaczyński

AstroCeNT, Nicolaus Copernicus Astronomical Center, Polish Academy of Sciences,  
Rektorska 4, 00-614 Warsaw, Poland, pkalaczynski@camk.edu.pl  
AGH University of Krakow, Center of Excellence in Artificial Intelligence, al. Mickiewicza 30,  
30-059 Krakow, Poland, pkalaczynski@agh.edu.pl

**Received:** 6.03.2025

**Revised:** 12.03.2025

**Accepted:** 12.03.2025