Dr. rer. nat. Jonas Glombitza

Erlangen Centre for Astroparticle Physics
Erwin-Rommel-Str. 1
91058 Erlangen

Thesis review Margita Majerčáková

The master's thesis "*Application of deep machine learning for the studies of the muon air-shower component using the Pierre Auger Surface Detector (SD) data*" from Margita Majerčáková is generally well written. In the thesis, a deep-learning-based approach to estimating the muon signals of air-showers using the Surface Detector of the Pierre Auger Observatory is investigated. This is scientifically very interesting since the muonic component gives information about the mass of the particle initiating the shower, and the detector cannot directly disentangle different shower components. Whereby the ansatz of studying deep-learning-based algorithms in this context belongs to the field of scientific novelties. In detail, in this thesis, two different approaches were developed that rely on the work of Ref. [1]:

- The reconstruction of the total muonic part of the measured SD signals using a simple neural network (NN) architecture.
- The reconstruction of the time-dependent muonic signal using recurrent neural networks (RNNs).

The writing style is linguistically good, the formulations are precise. The literature is well organized, and the references are given according to scientific standards. The thesis structure is meaningful and guides the reader very well through. Starting with introducing ultra-high-energy cosmic rays (UHECRs) and the Pierre Auger Observatory, a short discussion on the UHECR mass composition follows. The chapters are detailed and give an excellent intro to the 'UHECR universe'. Subsequently, machine learning and the used data are discussed shortly. The fundamental concept of supervised learning is well presented, but it could be a bit more detailed as a few essential references are missing.
The next two parts (chapters six and seven) are dedicated to the two different machine learning techniques (NN and RNN). In both chapters, the network design is described together with its application. The carefully-performed studies cover a variety of tests, e.g., the dependence on various air-shower kinematics. Plenty of plots are presented that quantify the performance in much detail; their discussion, however, could be more elaborate. Finally, the RNN is applied to other hadronic models and measured Auger data, and a conclusion is drawn.

The approach presented within the thesis extends the method developed in Ref. [1] that relies on the measured SD signals and reconstructed air-shower observables to estimate the muonic signals. Unfortunately, the work of Ref. [1] is not introduced in much detail; instead, it is referenced in several places, which is less descriptive. Nevertheless, the method is presented in a very understandable way. Primarily, in this thesis, those techniques are extended to the low-gain channel of the SD for stations with saturated high gain channels. Additionally, the used simulations are updated together with a meaningful update of the used cuts.

The achieved performance is very good, competitive with the work of Ref. [1,2], and features a relative error of a few percent. However, the obtained results are among each other not well comparable due to different evaluations and dissimilar plotting strategies. In the future, in-depth comparisons to the studies performed so far [1,2], including a detailed investigation of the method's resolution, would be interesting.
When discussing the NN-based reconstruction of the integrated muon signal, a clear dependence on the zenith angle and the distance to the core is visible. This looks like an artifact of the training, as both shower observables are part of the network input. Here, improvements are probably possible. However, these artifacts are not visible in the more sophisticated RNN training. Here, the overall performance is improved, the biases are symmetric around zero, and the results are consistent. Therefore, the inconsistencies found in the NN approach are of minor relevance.

To conclude, the well-written and well-structured thesis gives an excellent description of the reconstruction of muons signals using deep learning algorithms. The discussed neural network architectures were successfully designed, trained, and tested during the performed studies. Diverse crosschecks were performed, and the according plots are presented. However, the checks should be discussed more extensively, and also, a comparison to the work of Ref. [1,2] is apparently missing. Besides the study using simulations, even two data applications (the muon rise time and the dependency of the muon signal to the distance to the shower axis) of the RNN method are presented and briefly discussed. In fact, they show the expected behavior that the current generation of hadronic models generates too less muons.
As the RNN results are consistent and many careful evaluations of the algorithms were performed and are shown in the thesis, including the application to another hadronic model and measured Auger data, combined with very detailed introductory chapters and a clear thesis structure, the work can still be rated as excellent (A) despite a slight lack of comparisons and partly missing discussions of the obtained results.

Erlangen, 24.05.2022 _____

Dr. rer. nat. Jonas Glombitza

References:

[1] J. M. Carceller, *"A study of the signals measured with the water-Cherenkov detectors of the Pierre Auger Observatory to infer the mass composition of ultra-high energy cosmic rays"*, Ph.D. dissertation, University of Granada, 2020.
[2] A. Aab (Pierre Auger Collaboration) et al., *"Extraction of the muon signals recorded with the surface detector of the Pierre Auger Observatory using recurrent neural networks"*, JINST 16 P07016 (2021), doi: 10.1088/1748-0221/16/07/p07016.