BERGISCHE UNIVERSITÄT WUPPERTAL

Bergische Universität Wuppertal
Fakultät 4
Mathematik und Naturwissenschaften
Fachgruppe Physik

Irreducible proton background for the UHE photon search at the Pierre Auger Observatory

Master-Thesis

zur Erlangung des akademischen Grades

Master of Science (M.Sc.) im Studiengang Physik

vorgelegt von **Jannis Pawlowsky**

Erstprüfer:	Prof. Dr. Karl-Heinz Kampert
Zweitprüfer:	Prof. Dr. Christian Zeitnitz
Abgabedatum:	09.02.2021

Abstract

This thesis explores the possibility of proton induced air showers being misinterpreted as photon induced at energies above 10 EeV. This hadronic background hypothesis was tested with CORSIKA simulations using the EPOS-LHC and QGSJetII-04 hadronic interaction models. The probability of photon-like events, including relativistic effects, is calculated and discussed. It will be shown that the probability for misidentifying protons as photons drops sharply at the highest energies due to the Lorentz boosting of π^0*s. The probabilities at lower energies however, indicate the likely presence of photon-like hadronic events in the data from current astrophysical experiments. Nonetheless, this thesis shows that some of these photon-like events can still be discriminated from photon air showers. Finally, consistency between photon flux limits and proton background for a generic, idealized detector as well as for the Pierre Auger Observatory is shown. However, the hadron hypothesis alone cannot serve as explanation for the currently observed photon candidates.*

Contents

1 Structure and goals of this thesis **1**

2 Air shower physics **2**

2.1 Particle interactions 2

2.2 Cosmic rays 3

2.3 Gamma rays 5

2.4 Motivation for identification of the primary 6

2.5 The Pierre Auger Observatory 7

2.6 Discrimination parameters 8

2.7 Theory of photon-like protons 10

2.8 CORSIKA simulations 11

3 Photon-like event probability **14**

3.1 Probability for highly energetic π^0 resulting from proton primaries 14

3.2 Probability for highly energetic π^0 resulting from heavier primaries . . . 19

3.3 The effect of boosting on highly energetic π^0 21

4 Differentiation by shower characteristics **26**

4.1 Discrimination of single and split high energy π^0 26

4.2 Discrimination with X_{max} and N^{μ}_{max} 27

4.3 Discrimination with X_{max} 31

4.4 Discrimination with a generic, idealized ground based detector 32

4.4.1 Detector design 32

4.4.2 Validation of the detector simulation 33

4.4.3 Discrimination parameter distributions 37

4.4.4 Discrimination using a random forest 45

4.5 Conclusion on the irreducible background 54

5 Analysis of photon-like protons with the Pierre Auger software framework **56**

6 Conclusion **61**

A Discrimination power of X_{max} **68**

B Discrimination parameter of the SSD **69**

List of Figures

2.1 Features of the cosmic ray flux . . . 5
2.2 Shower developments of different shower components . . . 6
2.3 Pictures of the Pierre Auger Observatory . . . 8
2.4 Risetime smearing caused by distinct X_{max} . . . 11
2.5 Muon deficit in simulations . . . 11
3.1 Distribution of k_{π^0} of the most energetic π^0 . . . 14
3.2 Energy dependent probabilities for different k_{π^0} of the leading pion . . . 15
3.3 Correlation of $k_{\Sigma\pi^0}$ and k_{π^0} . . . 17
3.4 Distribution of $k_{\Sigma\pi^0}$ . . . 18
3.5 Energy dependent probabilities for different $k_{\Sigma\pi^0}$ . . . 18
3.6 Probability ratios of QGSJetII-04 and EPOS-LHC . . . 19
3.7 Re-weighted probabilities for different $k_{\Sigma\pi^0}$ . . . 20
3.8 Helium-proton comparison . . . 21
3.9 Path lengths of π^0s in the atmosphere . . . 21
3.10 Regarded shower developments . . . 22
3.11 Effects of the Lorentz boosting . . . 23
3.12 Energy and zenith angle dependence of the decay probabilities . . . 24
3.13 $k_{\Sigma\pi^0}$ probabilities for π^0 interactions . . . 24
3.14 Lorentz boost corrected probabilities . . . 25
3.15 Final corrected and re-weighted probabilities . . . 25
4.1 Shower comparison between single and multiple energetic π^0 . . . 26
4.2 X_{max} and N^{μ}_{max} distributions of photons and protons . . . 28
4.3 Fisher Discriminant distribution . . . 29
4.4 Discrimination power of X_{max} and N^{μ}_{max} . . . 30
4.5 Discrimination power of X_{max} . . . 31
4.6 Proton background for a X_{max} separation . . . 32
4.7 Risetime comparison of real data and simulations . . . 34
4.8 Correlation between the risetime and X_{max} . . . 34
4.9 Station signal comparison of real data and simulations . . . 35
4.10 Comparison of photon and proton LDFs . . . 36
4.11 Distribution of the R_{NKG} parameter . . . 38
4.12 Asymmetry effects caused by the zenith angle . . . 39
4.13 Distribution of the γ parameter . . . 39
4.14 Distribution of the risetime parameter . . . 40
4.15 Distribution of the β parameter . . . 41

4.16 Correlation of γ and β . . . 42
4.17 Distribution of the S(1000) parameter . . . 43
4.18 Effects of photonuclear interactions and preshowering . . . 43
4.19 Distribution of the S_b parameter . . . 44
4.20 Correlation between the WCD and SSD parameters . . . 45
4.21 Feature importance of the RF . . . 47
4.22 Precision and recall score of the RF . . . 50
4.23 ROC-Curve of the RF . . . 50
4.24 Photon probabilities for the test data . . . 51
4.25 Irreducible proton induced background . . . 53
4.26 Comparison of photon limits and distinct fluxes . . . 55
5.1 Results from the PCA . . . 57
5.2 Comparison of photon limits and expected fluxes for the Pierre Auger Observatory . . . 59

List of Tables

1 Proton background for a $X_{\max}$ and $N_{\max}$ separation . . . 28
2 Confusion matrices of the RF . . . 48
3 Summary from the PCA . . . 58
4 Proton background for a $X_{\max}$ separation . . . 68

1 Structure and goals of this thesis

The detection of ultra-high energy (UHE) photons is one of the main goals of the current generation of astrophysical experiments. These photons could point to the sources of galactic and extra-galactic cosmic rays. Recent studies helped to set new limits on the flux of UHE photons. A consistency with popular scenarios was shown, but maybe even more importantly, several photon candidates were identified by the Pierre Auger Observatory. However, it is not certain whether these candidates represent photons or a possibly misinterpreted background of hadronic events.

This thesis explores the scenario of photon-like events induced by protons. These photon-like proton air showers produce one or multiple highly energetic π^0s in their first interaction, which via the decay to two photons directly contribute to the electromagnetic part of the shower. If the π^0s receive the main fraction of the primary energy, the shower can resemble a photon induced event.

This thesis is structured as follows:
In Chapter 2, the basic properties of particle interactions and cosmic rays are described. Additionally, a short introduction to the air shower primary distinction is given. Lastly, the scenario of proton induced photon-like events is introduced.

The photon-like event probability is calculated in Chap. 3 for proton primaries. Additionally, the variation of the primary type is examined. Afterwards, the effect of Lorentz boosting on the π^0 secondaries is discussed.

These photon-like events are further explored in Chap. 4. Here, the discrimination between photon induced air showers and proton induced photon-like air showers is presented. It especially focuses on the muonic shower component and maximum shower depth as discrimination parameters. This is also done with detector simulations inspired by the detector at AugerPrime. At the end of this chapter, the consistency between the proton induced background and photon limits is discussed.

Finally, in Chap. 5, the number of expected proton induced photon candidates is calculated for the Pierre Auger Observatory. A conclusion is then presented on whether the proton induced background can serve as explanation for the observed photon candidates.

2 Air shower physics

This chapter explores the basic concepts of particle interactions and cosmic rays. Common detection methods of cosmic rays (CR) by the Pierre Auger Observatory are also shown. Finally, the properties and conditions of photon-like protons are discussed.

2.1 Particle interactions

For the purposes of this thesis, it is important to explain the differences between hadronic and electromagnetic (EM) interactions.

Hadronic interactions

Hadronic interactions are caused, as the name indicates, by hadrons e.g. protons, neutrons, pions, etc. This also includes heavier nuclei, as they consist of multiple nuclei. A highly energetic hadron traveling through the atmosphere will interact at some point with the surrounding matter. It collides with the nuclei of the atmosphere, which results in the production of multiple secondary particles. The multiplicity and type of particles depend on the energy and primary. A typical hadronic interaction could be

$$p + A \rightarrow p + n + \pi^+ + \pi^- + \pi^0 + B,$$

where A and B are nuclei. This holds true as a basic idea, but interactions leading to multiplicities > 100 and more diverse secondaries are more realistic, especially at the highest energies. Another important hadronic process is particle decay. For example, pions mostly decay [1] as follows:

$$\begin{aligned}
\pi^+ &\rightarrow \mu^+ + \nu_\mu, \\
\pi^- &\rightarrow \mu^- + \bar{\nu_\mu}, \\
\pi^0 &\rightarrow \gamma + \gamma.
\end{aligned}$$

The produced muons can then further decay:

$$\begin{aligned}
\mu^+ &\rightarrow e^+ + \nu_e + \bar{\nu_\mu}, \\
\mu^- &\rightarrow e^- + \bar{\nu_e} + \nu_\mu.
\end{aligned}$$

Electromagnetic interactions

EM interactions at the highest energies are dominated by Bremsstrahlung and pair production. A reaction chain can start with a photon, which then produces a electron-positron pair:

$$\gamma + A \rightarrow e^{+} + e^{-} + A.$$

The resulting charged particles interact again with the surrounding matter via Bremsstrahlung:

$$e^{\pm} + A \rightarrow e^{\pm} + \gamma + A.$$

It can be observed that the produced particles differ for hadronic and EM interactions. The only shared processes are the decay of π^0s and EM interactions. The distinct characteristic interactions are fundamental to the discrimination of hadron and photon primaries.

2.2 Cosmic rays

CR are highly energetic subatomic particles coming from space. The CR flux up to an energy of 10 GeV mainly originates from the sun [2]. For higher energies (see figure 2.1a), the origin of CR is less certain. In the range of 10 GeV to 1 PeV, the spectrum is expected to be dominated by galactic sources like supernova remnants [3]. From the PeV up to ZeV energies, a shift to extragalactic sources is predicted by various theories. Current scenarios favour starburst galaxies and active galactic nuclei as possible sources [2]. In particular, the acceleration mechanisms of these sources, which result in CR energies up to 1 ZeV, are a key aspect investigated by modern astroparticle physics. The flux of these UHE CRs is suppressed due to various astrophyical processes.

Figure 2.1a shows the energy spectrum (scaled by the energy cubed), which exhibits a range of features. For energies at the 10 - 100 PeV range, a bump can be observed. This is the so-called (second) knee, which is expected to be caused by the transition from galactic to extragalactic sources. At even higher energies, the EeV range, a second dip can be seen. The origin of this is not fully understood. The hardening is called the ankle of the CR flux with the 'cut-off' following. The most probable explanations for the cut-off at 60 EeV are a maximum acceleration energy of CR, photo disintegration of heavy nuclei, or the possible influence of the GZK-effect. To increase understanding of sources, the CR flux has to be measured and subdivided by primary particle type.

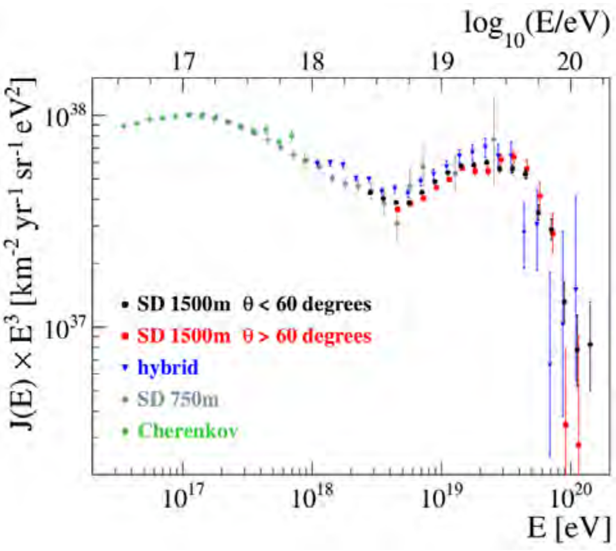
Hence, a detailed composition information is needed (on an event-by-event basis). When CRs hit the atmosphere, we may distinguish two cases [4]:

I) The CR is a hadron or light nucleus. It undergoes the already mentioned hadronic interactions, whereby the resulting secondaries also interact or decay. Thus, a cascade of secondary particles develop. This cascade consists of three components. First there is a hadronic component including mainly protons, neutrons, pions and kaons, but also possible spallation products from the interaction media. Decaying charged pions then feed a second, muonic component. Many muons do not decay due to relativistic time dilation before reaching Earth's surface.

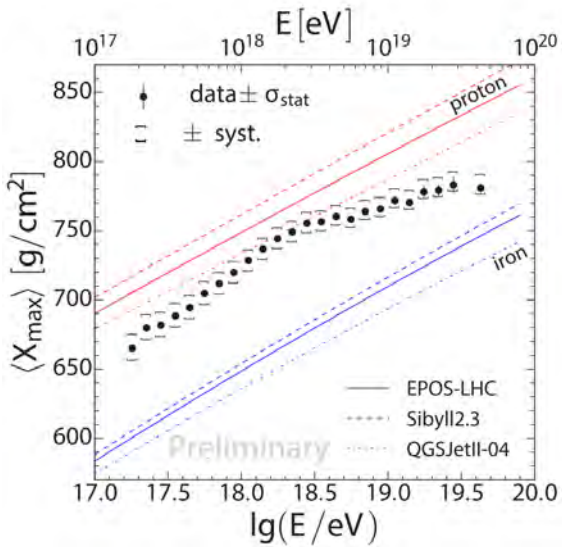
With increasing atmospheric depth, the secondary particles, in average, have a lower energy so that theit decay starts to dominate over reinteraction and other processes begin to dominate shower development. The cascade then stops and in particular the EM component begins to die out. This atmospheric depth marks the depth of shower maximum X_{max} [g/cm^2] with the maximum number of particles, N_{max}.

II) In the second case, the CR consists of a heavier nucleus like iron. The basic principle is the same as for case **I)**. Nevertheless, they differ along one important aspect. The hadronic interaction cross section of nuclei grows with the mass. This increased cross section results in an earlier first interaction in the atmosphere. Additionally, $\pi^{\pm}$s statistically receive less energy from heavy nuclei with respect to protons as the energy of the nuclei is split on A nucleons. Both factors favor the early decay of the $\pi^{\pm}$s and reduce the number of $\pi^{\pm}$ hadronic interactions. This results in an increased number of muons, whereas the amount of EM particles stays approximately the same. Thus, heavy nuclei have a larger ratio between the number of muonic and EM particles. In addition, the earlier first interaction leads to a lower $\langle X_{\mathrm{max}} \rangle$ in terms of $\mathrm{g\,cm^{-2}}$.

Both cases can be compared with real data measured by the Pierre Auger Observatory, as shown in Fig. 2.1b. The measured data is consistent with the simulations [5], but two things bear mentioning. First, the measured $\langle X_{\mathrm{max}} \rangle$ lies between the expectations for a pure proton and pure iron flux. This among other features leads to the conclusion of a mixed composition at these energies. The increasing deviation from the proton scenario indicates a tendency towards heavier nuclei at the highest energies. Even though differences between light and heavy nuclei can be observed, their shower developments are relatively similar, much more significant differences are visible for photon showers. This can also be seen in Fig. 2.2.



(a)



(b)

Figure 2.1: **(a)** The CR flux is shown for energies $> 10\,\mathrm{PeV}$. The flux is weighted by the energy cubed in order to show specific spectrum features, namely the knee and ankle. Plot from [2]. **(b)** The comparison of $\langle X_{\mathrm{max}} \rangle$ between simulations and real data is shown. The data indicates a mixed spectrum of light and heavy nuclei with a tendency towards heavier nuclei for increased energies. Plot from [5].

2.3 Gamma rays

The description of air showers with photon primaries is relatively simple. According to the Heitler model [6], the photon produces an electron and positron and, in rare cases, a muon and anti muon pair. The produced e^{+}-e^{-}-pairs undergo Bremsstrahlung, resulting in new photons, which then pair-produce again. Therefore, a photon induced air shower develops a large electromagnetic and a negligible muonic cascade with a well defined multiplicity of two for each interaction. This behaviour is well described by the Heitler model, where the number of particles doubles after each radiation length until ionization processes dominate the development. Besides the shower composition, another difference from hadron cascades is the depth of shower maximum. Even though the radiation length of photons and electrons is significantly smaller than the interaction length of hadrons, photon showers have a higher X_{max} stemming from the lower multiplicity for EM interactions.

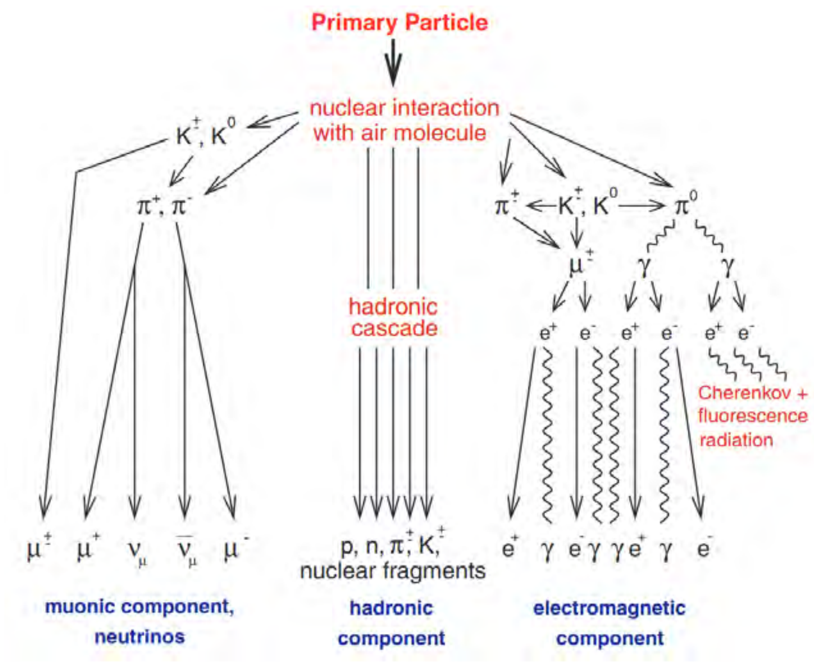


Figure 2.2: The three types of shower development are depicted. Plot from [7].

2.4 Motivation for identification of the primary

The value of primary particle type distinction to astroparticle physics should be discussed. As previously mentioned, the origin of CRs at the highest energies is still unknown. This stems from the fact, that the trajectories of charged particles like protons are deflected by the galactic and extragalactic magnetic fields. Thus, CRs do not point back to their sources. The paths of photons, however, are only marginally deflected due to gravitational effects. Therefore, the detection and identification of photon primaries could point directly to the accelerators of highly energetic particles. The identification of these sources in turn provides information about the source density in the universe and the consistency of distinct astrophysical scenarios. Hence, the detection of photons at the highest energies could accelerate discovery. Therefore, recent studies like [8], in which photon candidates were identified, are of special interest.

In the following, we discuss how experiments like the Pierre Auger Observatory measure the important discrimination parameters $X_{\max}$ and $N_{\max}$ and identify photon candidates. As the specific properties of the observatory are not relevant for this thesis, just a brief overview of the detection method will be presented. It will especially be focused on the detection of the particles reaching Earth's surface. More extensive discussions about the setup and results of the Pierre Auger Observatory can be found e.g. in [2] and [9].

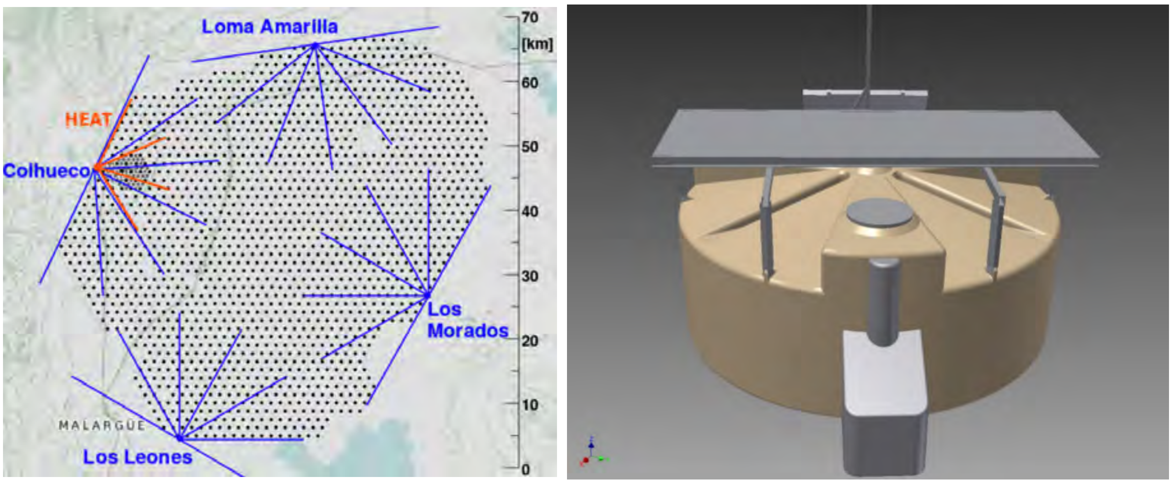
2.5 The Pierre Auger Observatory

The Pierre Auger Observatory is the largest experiment measuring UHE CR. It is based in the Argentinian pampas and covers an area of 3000 km^2, which is about the size of Luxembourg [9]. It is fully operational since 2008 and has a surface detector consisting of 1600 Water Cherenkov Detectors (WCD) with a regular spacing of 1500 m. In addition to the WCD, the Pierre Auger Observatory also consists of five Fluorescence Detector sides (FD), which detect the fluorescence light emitted by charged particles from the air shower moving through the atmosphere. Various other components like radio detection and lasers for atmospheric monitoring contribute to measurement and shower reconstruction accuracy. Complete description goes beyond the scope of this work, but can be found in [10]. Only the parts most important to this thesis will be introduced.

Water Cherenkov Detectors

The WCD grid consists of 1660 (1600 with a regular spacing and 60 in a more dense area) water tanks. Each tank contains 12 m^3 of highly purified water (see Fig. 2.3). Charged particles hitting the WCD produce Cherenkov radiation when travelling through the water. The emitted light is detected by three PMTs inside of the tank and then digitized. This results in a time binned (25 ns binning) signal. In order to increase the data quality, multiple trigger properties need to be fulfilled by the time trace. The most important trigger is a lower threshold for the signal amplitude of 3 VEM (=Vertical Equivalent Muon). Other aspects to increase the data quality include the use of low- and high-gain channels of the PMTs and the requirement of signal coincidences of the PMTs and neighbouring WCDs.

The WCD is able to detect the EM as well as the muonic component. Compared to other detectors, it is quite sensitive to the muonic component. Nonetheless, a significant fraction can traverse the full path through the WCD and exit the detector at the bottom or sides. The EM particles, however, lose nearly all their energy inside of the WCD and are stopped. The sensitivity to the muonic component already results in a possible discrimination of the primaries. Nonetheless, a superior separation power can be achieved, if the EM and muonic components are measured independently and not only summed up in the WCD. Hence, the deployment of Surface Scintillator Detectors (SSD) as part of the upgrade AugerPrime [9] is ongoing.



(a) **(b)**

Figure 2.3: **(a)** The structure of the Pierre Auger Observatory is shown. The black dots represent the WCD grid with a more dense region at the upper left side. The blue and red lines indicate the field of view of the FDs. Picture from [11]. **(b)** A WCD with a SSD mounted on the top. Picture from [9].

Surface Scintillator Detectors

The SSD consists of a scintillator plate with $\approx 3.8\,\mathrm{m}^2$ of active area and it is placed on top of the WCD (see Fig. 2.3b). Incoming particles hit the SSD first. EM particles like electrons can interact in the scintillator and one PMT detects the emitted radiation. Muons, however, are more likely detected in the WCD. This feature leads to separation of the EM component measured by the SSD, and of the muonic component measured by the WCD. Muons sometimes also interact in the SSD and electrons can enter the WCD from the side. Nonetheless, a tendency towards the separation is expected. The ratio of the detector amplitudes is then used to determine the primary particle. An additional surplus of the SSD can be seen for the detection of air showers with close shower cores. Here, the signal amplitudes in the WCD have observable saturation, whereby the SSD with its comparably low particle detection efficiency increases its dynamic range.

2.6 Discrimination parameters

The individual station signals of the WCD and SSD are combined to create powerful discrimination parameters. Here, the Lateral Distribution Function (LDF) parameters and the station risetimes are of special interest.

LDF-Fit parameters

The LDF describes the particle density of an air shower as a function of the distance from the shower core of the air shower. Photons have a steep LDF. The main part of the signal, the EM component, is concentrated close to the shower core. With increasing distance, the signal drops steeply. In contrast, the LDF of a regular proton is less steep. The EM component is weaker, therefore, the signal at the core is lower. But with growing distance, there are still a significant amount of muons contributing to the signal. This contribution is negligible for photons. The LDF of the showers can be parameterized by the following function [12]:

$$S(r) = S(1000) \cdot \left(\frac{r}{1000\,\mathrm{m}}\right)^{\beta} \cdot \left(\frac{r + 700\,\mathrm{m}}{1700\,\mathrm{m}}\right)^{\beta+\gamma}, \tag{1}$$

where r is the distance from the core, and β and γ are steepness parameters. It will be shown that especially γ has a strong discrimination of photons and hadrons. S(1000) describes the expected station signal at a distance of 1000 m and is therefore related to the muonic component. Two other related parameters can be calculated with the LDF.

$\mathbf{S_b}$

$\mathrm{S_b}$ is related to the steepness of the LDF and especially correlated to S(1000). It is calculated as follows:

$$S_b = \sum_{i=1}^{N} \mathrm{S}_i \cdot \left(\frac{r_i}{1000\,\mathrm{m}}\right)^{b} \tag{2}$$

$\mathbf{R_{NKG}}$

Another parameter resulting from the LDF is the R_{NKG}:

$$\mathrm{R_{NKG}} = \frac{1}{N} \sum_{i=1}^{N} \frac{S_i}{S(r)} \tag{3}$$

It describes the averaged relative fit deviation from the station signals. To improve the discrimination power, only stations with distances to the shower core ≥ 1000 m are taken into account. This parameter, as well as the following risetime parameter, were used to identify the photon candidates in [8].

Risetime

The risetime parameter is not related to the LDF. It is calculated as follows [12]:

$$t_{1/2} = \frac{1}{N}\sum_{i=1}^{N} t_{1/2,i}, \tag{4}$$

where $t_{1/2,i}$ is the individual station risetime. It is the time span, defined by the points, when the station signal grows from 10 to 50 % of its integral. The individual stations are required to meet the following requirements in order to be used in the calculation:

- The integrated station signal exceeds a certain threshold (filtering noise).
- The station signal is not saturated.
- The distance to the shower core is $> 1000\,\mathrm{m}$ and $\leq 2000\,\mathrm{m}$.

The differences in the risetime parameter for photons and hadrons results from two facts. First, the heights of X_{max} result in distinct path differences for two points in the shower maximum. This can be seen in Fig. 2.4. Second, the muons arrive the earliest at the detector stations. Muon rich showers can exceed the 10 and 50 % signal thresholds just with the muonic component. However, the muonic component of EM dominated showers will cause a signal exceeding the 10 % threshold, but does not exceed the 50 % threshold until the EM particles arrive at the detector. The discrimination powers of the individual parameters will be shown in chapter 4.4.3. More information on the parameters can be found in [12].

2.7 Theory of photon-like protons

The major differences between air showers with distinct primaries were discussed and how the showers are discriminated in real experiments. Nonetheless, given the statistical nature of the discrimination parameters, there is a certain probability of misidentifying the primary type. In particular, the cases where hadrons are labeled wrongly as photon candidates are of special interest.

These events would need to have characteristics resembling those of a photon, which is not possible for regular hadronic shower developments. As described before, in the case of a hadronic primary one obtains mainly hadronic or muonic components after the first interaction. But different papers [15] [16] [17] suggest that this does not hold true in all cases. If one of the produced secondaries is a π^0 and in case the π^0 receives a main fraction of the primary energy, the resulting shower will have a significantly higher

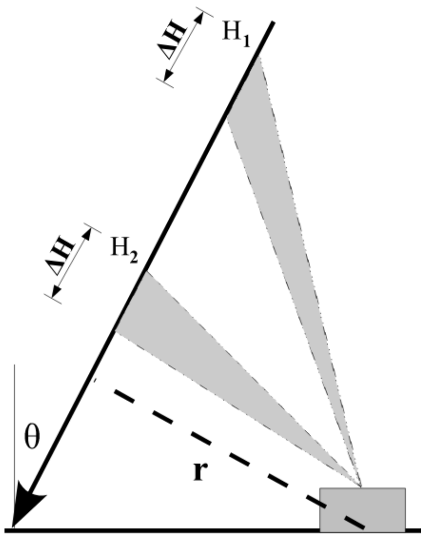


Figure 2.4: Distinct atmospheric heights cause different smearing of the signals . Plot from [13].

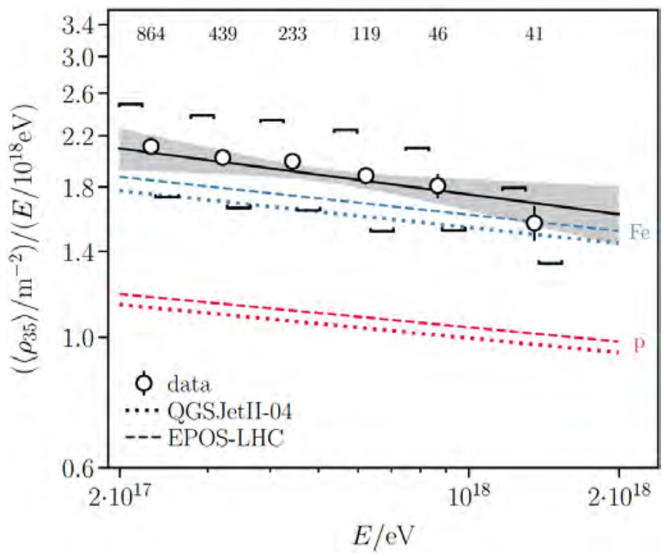


Figure 2.5: The muon density ρ_{35} for different interaction models are shown. A deficit compared to real data can be observed . Plot from [14].

EM component. The π^0 decays into two photons, which in turn produce a regular EM shower. These events will be referred as events with a high k_{π^0} or events with high π^0-inelasticity.

The mentioned papers all agree on the existence of these events, although their probability is quite low. However, it was not discussed, which fraction of primary energy needs to be transferred to the π^0 for the shower to be photon-like. Additionally, it is not clear if these events can really not be discriminated from actual photons and might serve as explanation for the measurement of photon candidates. To test this hadron hypothesis, events with highly energetic π^0 secondaries have been simulated and analysed.

2.8 CORSIKA simulations

The simulations are performed with the program CORSIKA [18]. CORSIKA is the abbreviation for **CO**smic **R**ay **SI**mulations for **KA**scade. It is the most popular program to simulate air showers.

Muon deficit of CORSIKA simulations for hadronic primaries

CORSIKA uses different selectable interaction models. Dependent on these models, the cross sections changes with energies as well as the produced particles. In general, all models result in reasonable shower developments and their simulated detector signals are similar to those of real data. However, all models show a deficit in the number of muons. in [14] and [19], the muon deficit was determined by comparing simulations with measurements. In particular important for this work, in [19] the energy range from 6 to 16 EeV was investigated. The ratio R of the muon density between measured data and simulations is model and composition dependent. It was shown, that for the EPOS-LHC model with proton primaries the correction factor is $R = 1.45 \pm 0.16 \pm 0.08$, whereby the QGSJetII-04 model shows a increased factor of $R = 1.59 \pm 0.17 \pm 0.09$. This will be important for the interpretation of this work. As the number of muons is a important discrimination parameter, the muon deficit can lead to an overestimation of the proton induced flux and therefore more hadronic photon candidates. Even though [19] only determines the muon deficit in the range of 6 to 16 EeV, a similar behaviour and therefore lower muonic component is expected for higher energies.

Simulation geometries

CORSIKA includes different hadronic interaction models [18]. In this thesis, the CORSIKA version 7.7100 and the Monte Carlo hadronic interaction software FLUKA 2011.2 is used. Additionally, the CERN interaction models QGSJetII-04 and EPOS-LHC are both utilized and compared. It is important to mention that the simulations for both models are carried out with the same initial conditions, including shower geometry, random seed and primary energy. The following shower properties will be simulated:

- The primary energy is between $10^{18.8}$ and $10^{20.5}$ eV with a spectral index of -1.
- The zenith angle θ ranges from 0 to 60 ° (convention: 0 ° points vertical with respect to Earth's surface).
- The azimuth angle ϕ covers the whole range of 360 °.
- The geomagnetic field is $B_x = 19.812\,\mu$T and B_z = -14.3187 μT.
- The thinning rate is 10^{-6}, both for hadronic and electromagnetic particles.
- The seeds for the internal randomization are created with the Twister-Algorithm.
- The primary is a proton except for the simulations in chapter 3.2.

Since the development of the EM shower depends on the creation of π^0 and its energy at the first interaction, it is not always necessary to simulate the entire shower development. In the following chapter, it will be explored under which conditions only the first interaction, and under which the full shower is simulated.

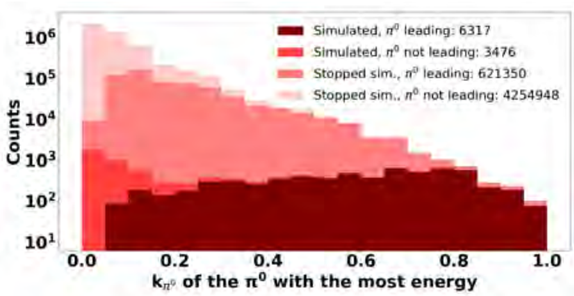
3 Photon-like event probability

3.1 Probability for highly energetic π^0 resulting from proton primaries

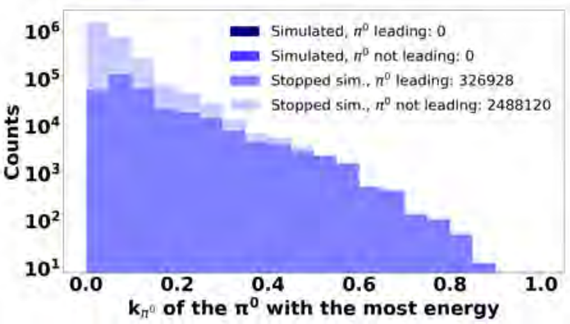
To study and quantify a result with high statistical confidence, a large quantity of simulated events is required. This is especially true for high k_{π^0} events, which are rare and therefore require even larger event statistics. For example, in approximately five million generated events using EPOS-LHC, around 1350 events with k_{π^0} greater than 0.8 were found (see Fig. 3.1). Simulation of all five million events fully to only gain 1350 high k_{π^0} events would be a unreasonable CPU time and would deliver mostly redundant data. Instead, two things can be done to achieve the requested large statistics more efficiently.

The first possibility is to stop and analyze simulations at the first interaction before carrying on and simulating the whole shower. This way, one can decide to only fully simulate the shower depending on the desired k_{π^0}. In Fig. 3.1a it can be seen that the fully simulated showers are chosen such that a roughly equal number of simulated events appear in each bin over the whole k_{π^0}-range. In the last k_{π^0} bins a dip can be seen. This is because events with k_{π^0} greater than 0.9 are exceptionally rare with there are only about 150 events found in five million simulated EPOS-LHC showers. In case QGSJetII-04 is used instead, not a single k_{π^0} above 0.9 was observed.

In the first simulation run, all first interactions from the events shown in Fig. 3.1b were regarded for both models. From this, it was clear that events with large k_{π^0} are more likely to occur in EPOS-LHC. Therefore, in order to further reduce simulation time,



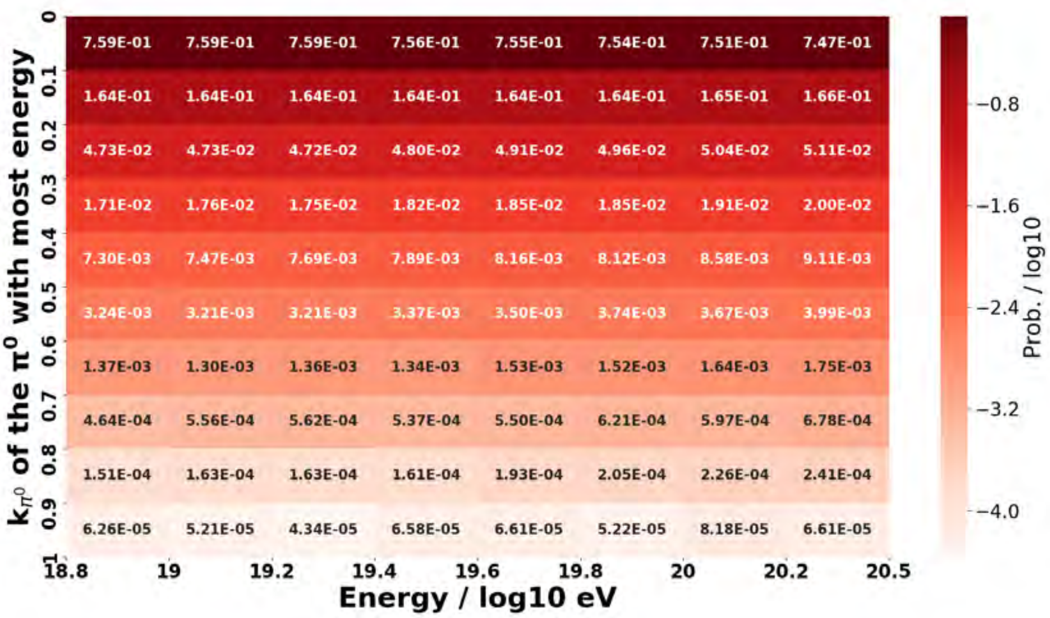
(a) EPOS-LHC



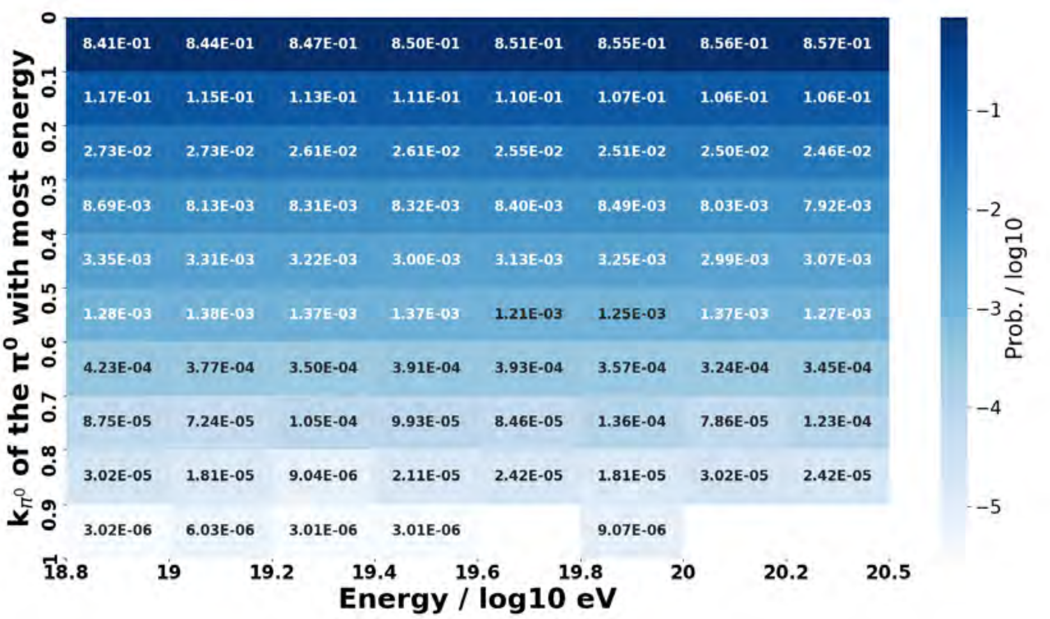
(b) QGS-JetII-04

Figure 3.1: The distribution of the k_{π^0} for the pion with the most energy in each simulated shower is shown. The different opacities indicate whether a shower simulation was stopped after the first interaction and whether the pion was the leading particle.

only EPOS-LHC was used for the large production in order to increase the statistics. The results of the full production can be seen in Fig. 3.1a, which includes all events from both runs.



(a) EPOS-LHC



(b) QGS-JetII-04

Figure 3.2: The probabilities for the different k_{π^0}-bins are shown. The plot is normalized column wise, so that the probabilities of each energy bin sum up to one.

These histograms can also be binned energy wise to give the energy dependent probability for the different k_{π^0}. This is shown in Figs. 3.2a and 3.2b. Two things evident in these plots bear mentioning. First, as was previously mentioned, the probability of having a large k_{π^0} is greater for EPOS-LHC. Second, the probabilities for the $k_{\Sigma\pi^0}$ bins are independent of the energy in QGSJetII-04 and slightly vary with energy in EPOS-LHC.

The other way to increase statistics results from a physical consideration. Events can be observed, which have two or even more significantly inelastic π^0s that sum up to a high $k_{\Sigma\pi^0} = \sum_{i=1}^{N_{\pi^0}} k_{\pi^0,i}$. This feature is shown in Figs. 3.3a to 3.3f. These events are expected to act similarly to a single high k_{π^0} in the shower.

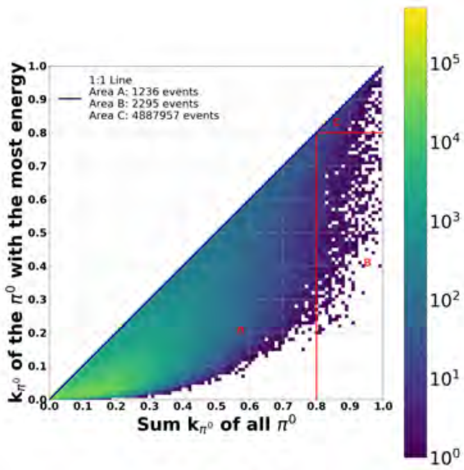
It can be observed, especially in Figs. 3.3a and 3.3b, that the most probable case is that a single π^0 makes up nearly the whole $k_{\Sigma\pi^0}$ of the first interaction. With a lower k_{π^0} for the most energetic π^0, the probability for a high $k_{\Sigma\pi^0}$- event decreases. There are, however, also many high $k_{\Sigma\pi^0}$ events which consist of two or more significantly contributing π^0s, but rarely events with five or more.

In [15] only the protons from area C with $k_{\pi^0} > 0.8$ were considered. However, in the case that events from area B with $k_{\Sigma\pi^0} > 0.8$ are not distinguishable from those of area C, the flux of possible photon-like protons would be significantly greater, almost by a factor of three leading to a significant underestimation. The indistinguishability of the events from area B and C will be discussed in chapter 4.1.

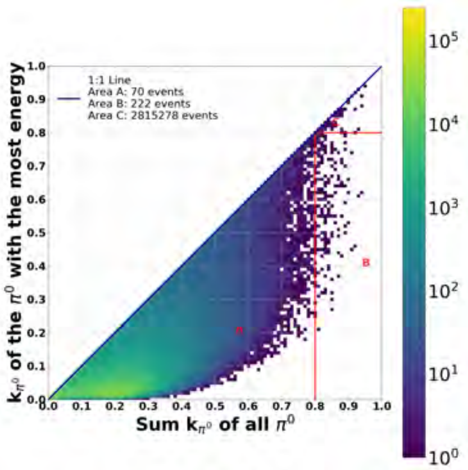
Another aspect to mention regarding
[15] is the selected lower limit of $k_{\pi^0} = 0.8$ as a benchmark for photon-like protons. The limit was chosen without considering whether lower $k_{\Sigma\pi^0}$ could contribute significantly to the proton background. Analyzing these events will lead to a discrimination power depending on the $k_{\Sigma\pi^0}$ and possibly results in a higher flux of photon-like events .

Focusing on $k_{\Sigma\pi^0}$ instead just on the π^0 with the highest energy, the distributions shown in Fig. 3.4 are found. Comparing Figs. 3.1 and 3.4, one can identify that there are many more events with large $k_{\Sigma\pi^0}$ and also that the distribution is shifted to larger $k_{\Sigma\pi^0}$. After this change, the amount of fully simulated events is still more or less equally distributed over the $k_{\Sigma\pi^0}$ range. Also, there are now events observable in QGSJetII-04 in the second last bin. However, the probability for high $k_{\Sigma\pi^0}$ is still greater EPOS-LHC. The probability as a function of $k_{\Sigma\pi^0}$ and primary energy are energy independent in QGSJetII-04 and energy dependent in EPOS-LHC, shown in Figs. 3.5 and 3.6.

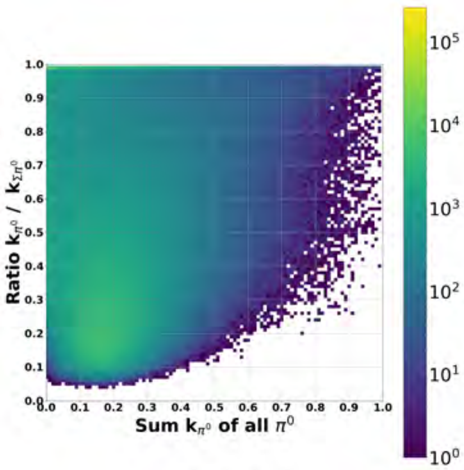
Fig. 3.6 shows that the ratio between the model probabilities increases for lower $k_{\Sigma\pi^0}$ and decreases for higher $k_{\Sigma\pi^0}$. In the last $k_{\Sigma\pi^0}$ bins, a 10 to 20 times higher probability for EPOS-LHC can be seen.



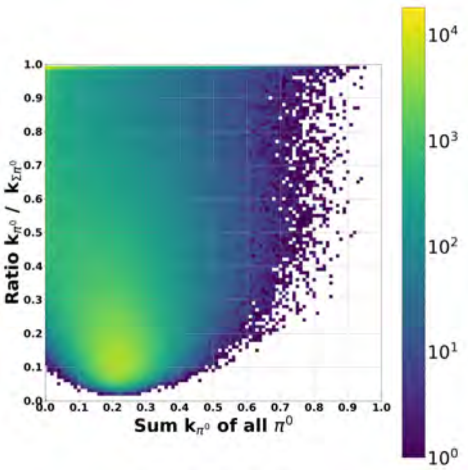
(a) EPOS-LHC



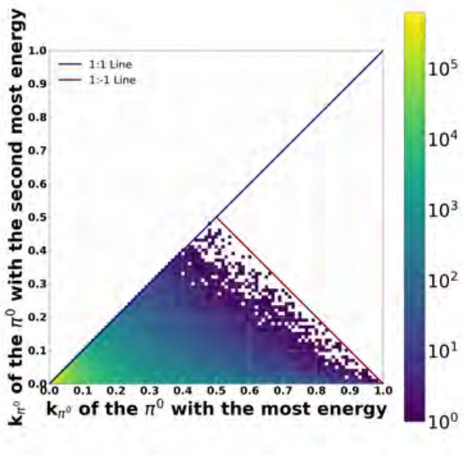
(b) QGS-JetII-04



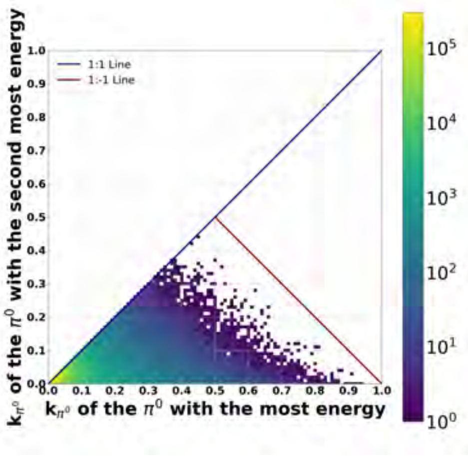
(c) EPOS-LHC



(d) QGS-JetII-04

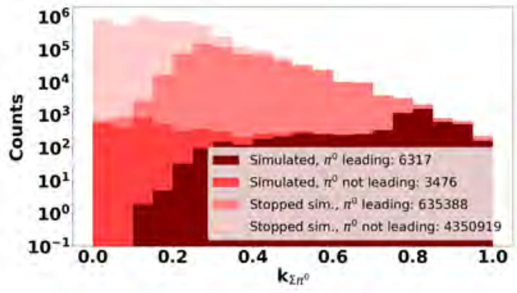


(e) EPOS-LHC

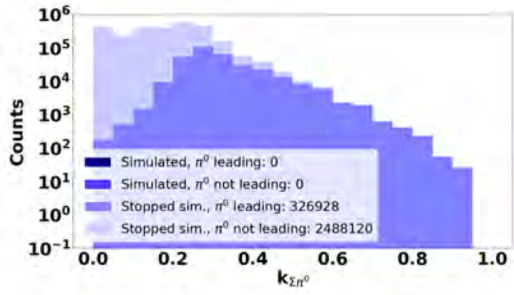


(f) QGS-JetII-04

Figure 3.3: The correlation between the k_{π^0} of the pion with most and second most and the sum of all k_{π^0} is shown. Both models deliver similar results.

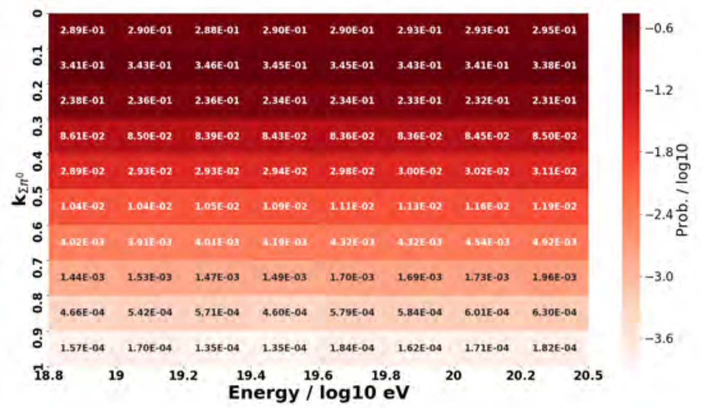


(a) EPOS-LHC

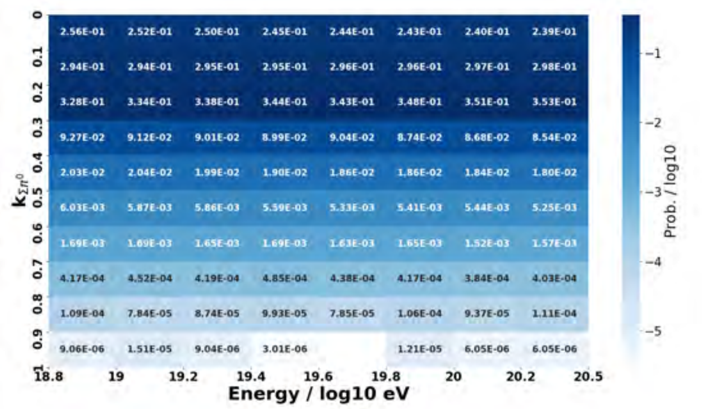


(b) QGSJetII-04

Figure 3.4: The distribution of the $k_{\Sigma\pi^0}$ is shown. The different opacities indicate if the shower simulations were stopped after the first interaction and if the leading particle was a pion.



(a) EPOS-LHC



(b) QGSJetII-04

Figure 3.5: The probabilities for the different $k_{\Sigma\pi^0}$-bins are shown. The plot is normalized column wise, the probabilities of each energy bin sum up to one.

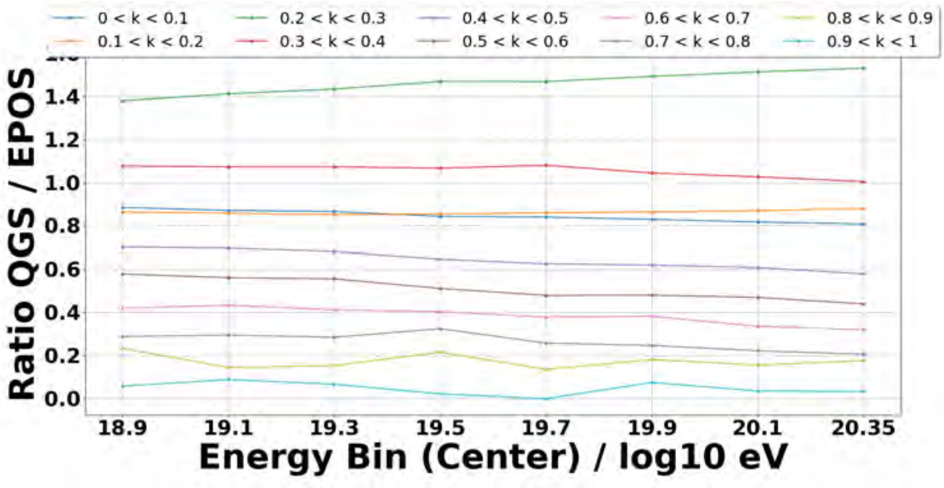


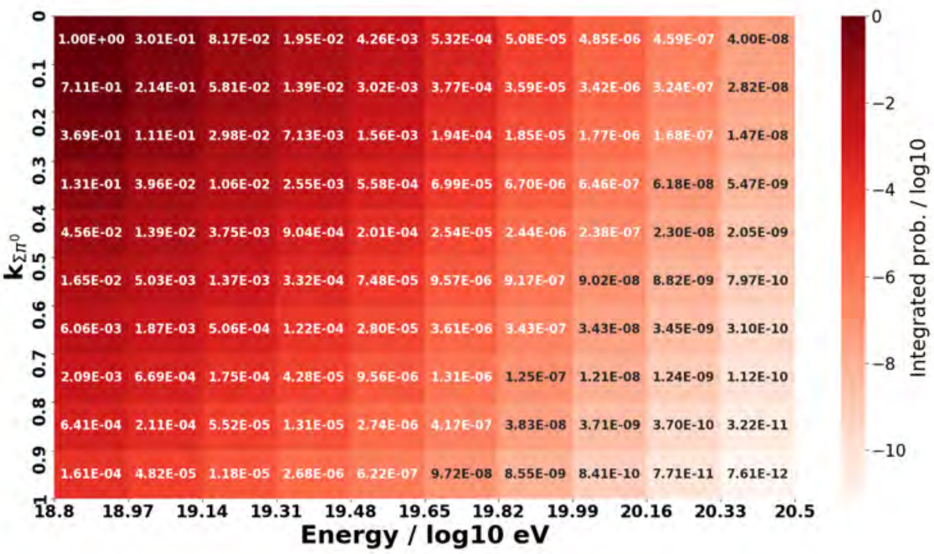
Figure 3.6: The ratio between QGSJetII-04 and EPOS-LHC is shown for each energy bin. For higher energies and $k_{\Sigma\pi^0}$ the ratio decreases while for small $k_{\Sigma\pi^0}$ it increases.

As was previously mentioned, both the proton and photon simulations were performed with an E^{-1} spectrum. As this is flat in $\log(E)$, one can easily re-weight the spectrum of simulated events to the measured spectral indices from [20]. One can then get an impression of the proton flux required for the occurrence of an event with a high $k_{\Sigma\pi^0}$. Figure 3.7 shows this, the re-weighted, integrated probability for a proton in a specific energy and $k_{\Sigma\pi^0}$ bin over an energy range of $10^{18.8}$ to $10^{20.5}$ eV. These probabilities show that millions of highly energetic proton events would be needed for the occurrence of just one high $k_{\Sigma\pi^0}$ proton with an energy greater than $10^{19.65}$ eV. Recalling the limits on the proton flux at these energies, it is clear that the measurement of one such proton is unlikely to occur with the currently existing experiments.

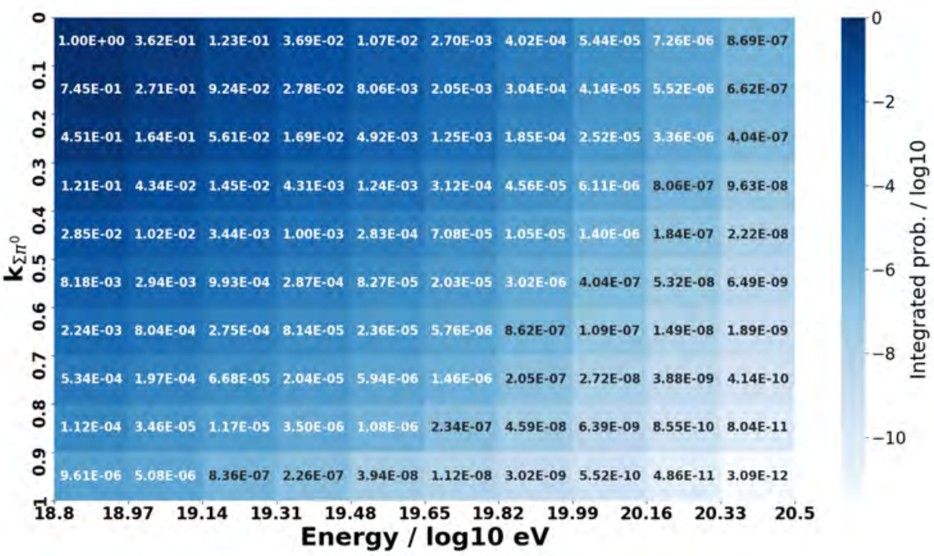
3.2 Probability for highly energetic π^0 resulting from heavier primaries

Until now, only proton primaries were considered when looking for photon-like air showers. However, it is possible that heavier nuclei could lead to these events as well. In the end, it is clear that the probability is quite low, with respect to protons, as the probability of having a large $k_{\Sigma\pi^0}$ decreases with primary mass. This is shown for helium in Fig. 3.8. Clearly, the distributions of the primaries differ significantly, with the curve for helium being much steeper than the proton curve.

For this plots, the first interactions for both primary types were simulated only with EPOS-LHC. Even so, only one helium event with a $k_{\Sigma\pi^0}$ greater than 0.7 is observed for 200k simulated events, whereas a few hundred protons were seen. Obviously, protons



(a) EPOS-LHC



(b) QGSJetII-04

Figure 3.7: The spectrally re-weighted probabilities for different $k_{\Sigma\pi^0}$ and energies. The plot is normalized so that all entries add up to one. The energy binning has to be changed to better match the development of the spectral indices and to equally split the energy range.

are the dominant source of hadronic primary photon-like events compared to helium or even heavier nuclei. Frankly, it can be argued, that a pure proton spectrum at these energies is unlikely. Recent measurements[4][5] and scenarios[20] indicate a tendency

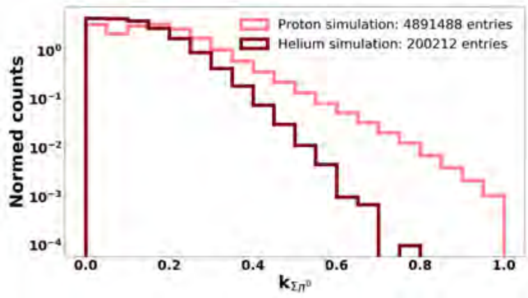


Figure 3.8: The distribution of the $k_{\Sigma\pi^0}$ is shown for EPOS-LHC for proton and helium primaries. The plot is normalized for a better comparison of the curves.

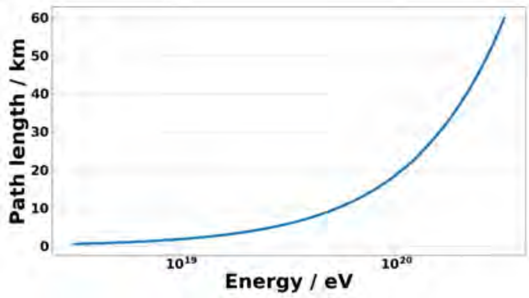


Figure 3.9: The path lengths of π^0s in the atmosphere for different energies are shown.

towards heavier nuclei for higher energies. Nonetheless, a pure proton spectrum will be assumed in this work in order to provide the most conservative scenario in terms of discrimination power.

3.3 The effect of boosting on highly energetic π^0

Until now, high $k_{\Sigma\pi^0}$ and photon-likeness were treated as equivalent. This assumption holds true for low energy events, but at the highest energies one also has to account for relativistic effects. For example, in [15, p.155], there are events visible which have a large $k_{\Sigma\pi^0}$, but differ strongly from other high $k_{\Sigma\pi^0}$ events in that they display the characteristics of a regular proton shower. This feature, which can also be found in the events simulated for this study, was not explained until now.

This behaviour is clearly correlated to the energy of the produced neutral pion. In the case of π^0-energies below 10^{18} eV, it can be assumed that any π^0 decays immediately into two photons due to its short lifetime. However, for higher energies, the Lorentz factor reaches values of $\gamma_{\text{Lorentz}} > 10^{10}$, which significantly increases the proper time of the π^0s. With a mean lifetime of $\mathrm{t}_{\pi^0} = 8.52 \cdot 10^{-17}\,\mathrm{s}$[1] and the speed of light as the velocity, one can calculate the mean path length of the π^0s:

$$l_{\text{mean}} = t_{\pi^0} \cdot c \cdot \frac{E_{\pi^0}}{E^0_{\pi^0}}.$$

This leads to the results shown in Fig. 3.9. While a 10^{19} eV π^0 travels ≈ 2 km, a $10^{20.5}$ eV π^0 can reach distances of ≈ 60 km. Clearly, these are path lengths in which the π^0s re-interact with the atmosphere before decaying. This interaction can then lead back to a regular proton shower or to a photon-like development. This feature is also described in [21]. All regarded possibilities are shown in Fig. 3.10.

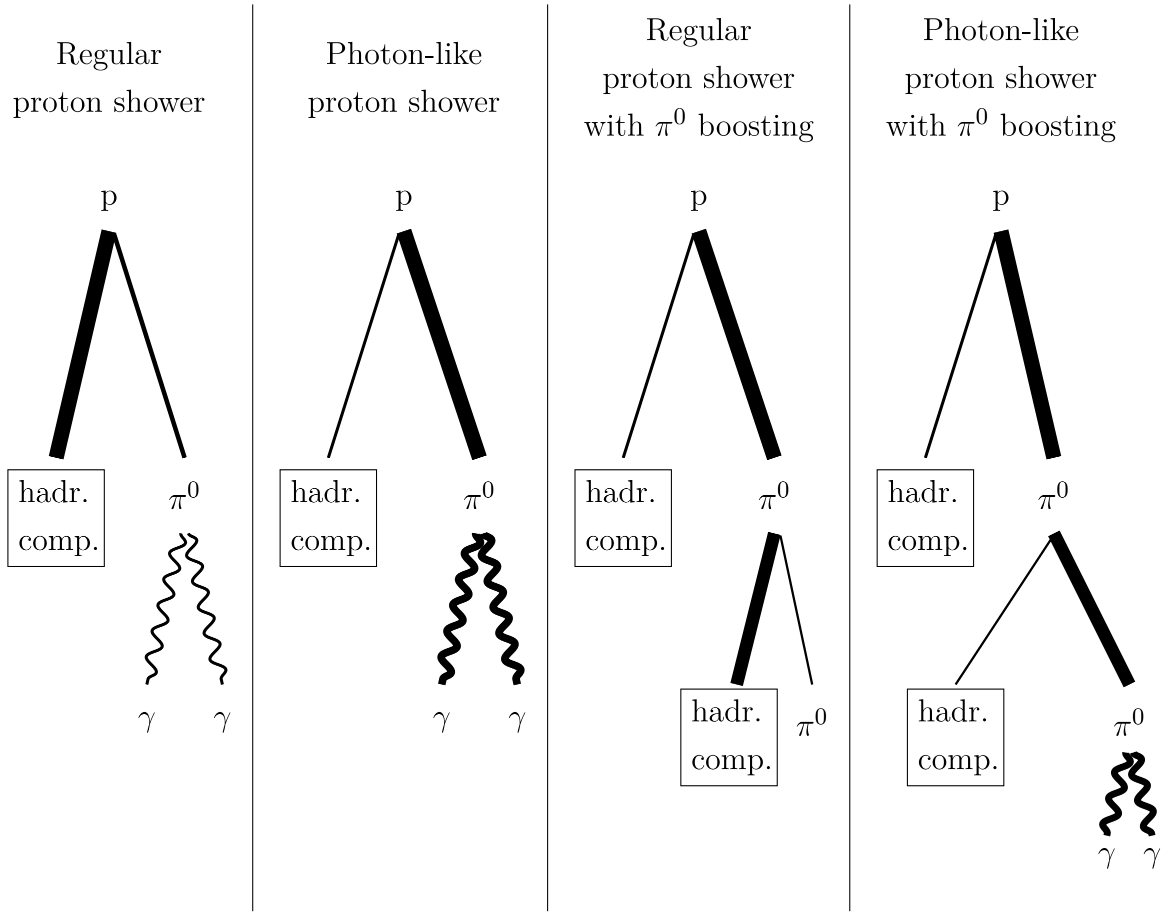


Figure 3.10: The most important shower development scenarios for this paper are shown. The widths of the lines symbolize the received fraction of primary energy. The scenarios are ordered from more probable (left) to less probable (right).

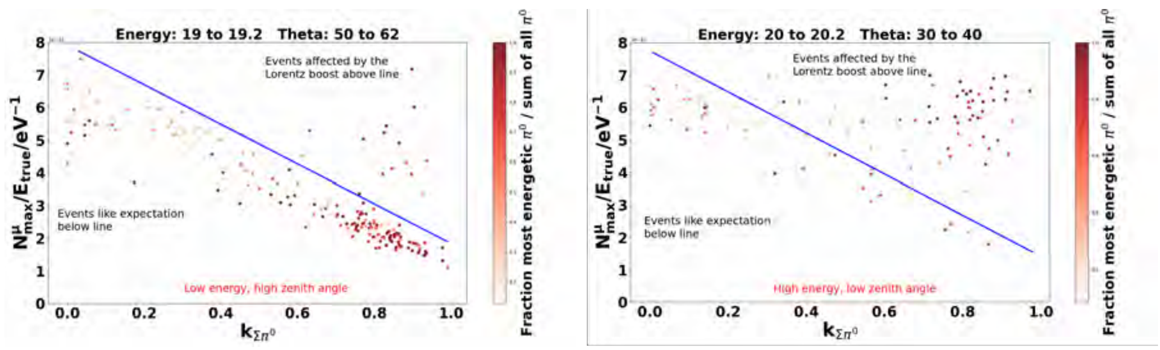
The re-interaction probability is dependent on the following parameters:

1. The maximum possible Lorentz boost depends on **the energy of the primary**
2. **and on** $k_{\Sigma\pi^0}$.
3. **The zenith angle** of the primary, which is closely correlated to the angle of the π^0, as higher zenith lead to larger distances in less dense atmosphere.
4. **The first interaction height** as height is directly correlated with atmospheric density.

A visual example for the importance of the first three properties is shown in Fig. 3.11. A linear decrease in $N^{\mu}_{\rm max}/E_{\rm true}$ is expected for increasing $k_{\Sigma\pi^0}$. The two figures demonstrate the difference for distinct primary energies and zenith angles. It can be seen that for higher $k_{\Sigma\pi^0}$, zenith angles and energies the fraction of events differing from expectation increases as the π^0s are more boosted.

Particularly at high energies, one can see events with high $k_{\Sigma\pi^0}$ and only slightly increased muon number. This can be interpreted as follows. The color of the data points indicates the ratio between $k_{\Sigma\pi^0}$ and k_{π^0}. Events with a small ratio imply that the $k_{\Sigma\pi^0}$ consists of multiple significant contributing π^0s. Because these π^0s share the primary energy their boosts will be lower. So it is possible that some of these π^0s decay while others interact. This results in a small increase in the number of muons, which is still below that of a regular proton shower.

To quantify this effect, the height of first interaction for the simulated proton primaries were obtained and used to produce 1 million CORSIKA showers simulated with π^0 primaries. A zenith angle distribution flat between 0 and 60° and a primary energy range of $10^{18.5}$ to $10^{20.5}$ eV with an E^{-1} spectrum were used. The probability of the π^0 to decay is obtained by outputting the particle stack after the first incident. The results of this procedure are shown in Fig. 3.12. A strong dependence on the energy (and therefore on the $k_{\Sigma\pi^0}$) and a smaller but not negligible zenith angle dependence are clearly visible. The next thing to look at are the probabilities of the π^0s to again produce highly energetic π^0s in these interactions. This is shown in Fig. 3.13.



(a) Low energy, high zenith angle **(b)** High energy, low zenith angle

Figure 3.11: Effects of different energies, $k_{\Sigma\pi^0}$ and zenith angles on the normed muon count. Without boosting effects, a linear decrease like in (a) is expected.

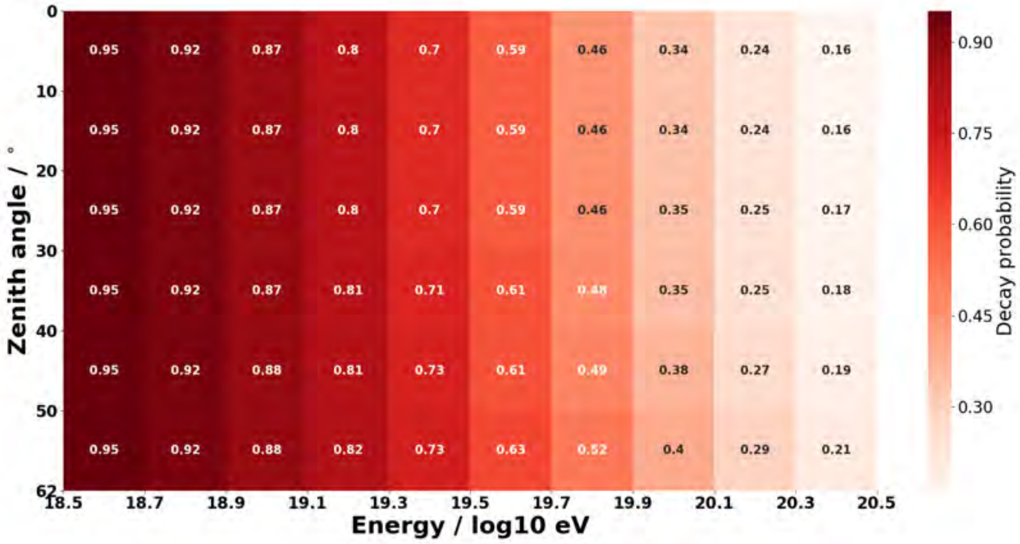


Figure 3.12: Energy and zenith angle dependence of the decay probability.

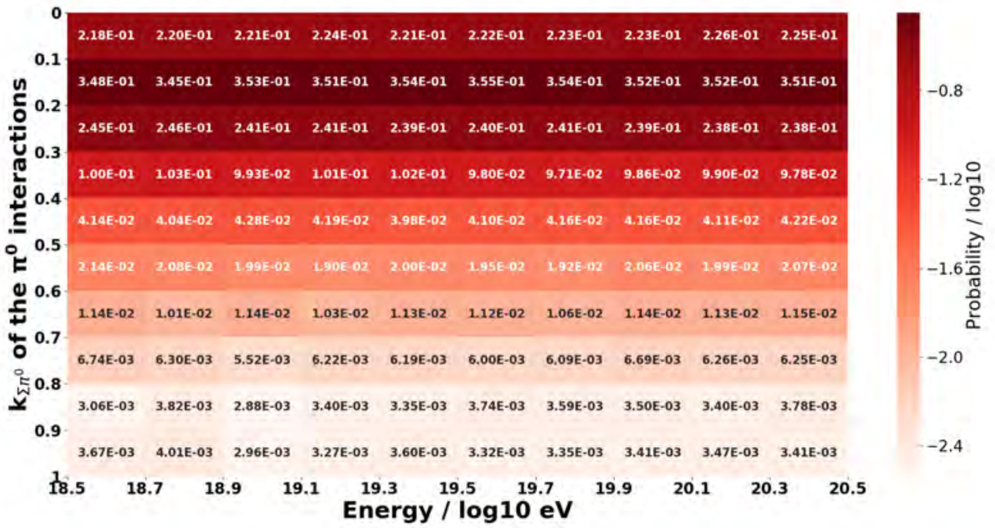


Figure 3.13: Probability of the $k_{\Sigma\pi^0}$ in case of a π^0-interaction.

When these probabilities are applied to the measured $k_{\Sigma\pi^0}$ spectrum from proton primaries, the corrected and final probabilities for photon-like showers are obtained. They are shown in Fig. 3.14, whereas Fig. 3.15 depicts the integrated energy dependent $k_{\Sigma\pi^0}$. For simplicity, the expression $k_{\Sigma\pi^0}$ will continue to be used, but it should be noted, that these events can still look like regular proton showers due to boosting effects.

From the results presented in this chapter, an important question arises with respect to a photon search, and that is whether these photon-like showers can be distinguished from actual photon showers. The next chapter discusses how well these events can be identified using the maximum number of muons and depth of shower maximum.

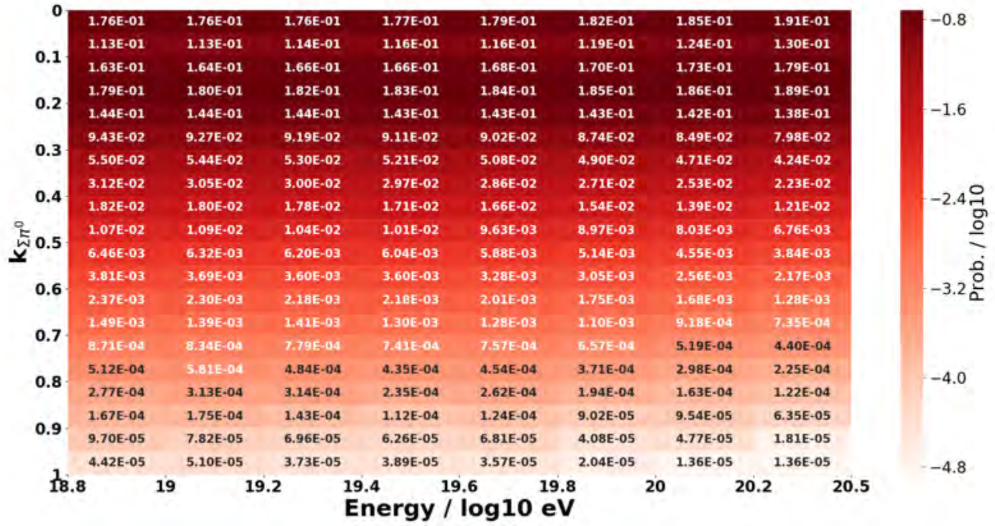


Figure 3.14: Boosting corrected probability for different $k_{\Sigma\pi^0}$ after the second interaction. The plot is normalized column wise.

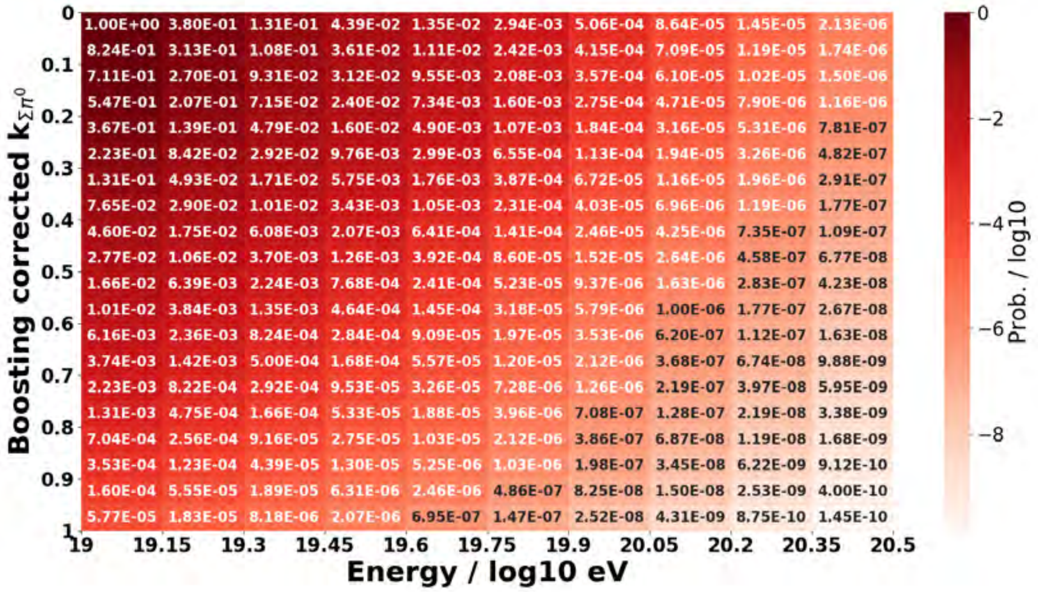


Figure 3.15: Energy and $k_{\Sigma\pi^0}$ integrated probabilities. Each bin gives the probability to have an energy or $k_{\Sigma\pi^0}$ equal or higher than the bin value.

4 Differentiation by shower characteristics

4.1 Discrimination of single and split high energetic π^0s

Before the discrimination of photons and protons is investigated, the shower-to-shower fluctuations of protons are studied. As mentioned before, the $k_{\Sigma\pi^0}$ can consist of one or in other cases multiple π^0s. At best, these two possibilities should only differ in the $X_{\max}$ of the resulting showers, as the number of particles produced in the shower should be equal in both cases as it only depends on primary energy. In case the high $k_{\Sigma\pi^0}$ comes from multiple π^0s, the most likely case is that their energy is split between two π^0 as the probability decreases with increasing π^0-multiplicity. As a limit, the case where both π^0 receive the same energy will be explored. This equal splitting should result in a maximum difference in $X_{\max}$ between events with single and multiple contributing π^0. To do this, high $k_{\Sigma\pi^0}$ events were simulated, half of them with a single significant contributing π^0 and the other half with an equal splitting. All other shower geometries and secondaries are the same. The $X_{\max}$ distribution for both cases is shown in Fig. 4.1.

One can see that their means are similar with $\langle X_{\max}^{\text{single}} \rangle = 949.8 \pm 1.1\,\text{g}\,\text{cm}^{-2}$ and $\langle X_{\max}^{\text{double}} \rangle = 933.9 \pm 0.6\,\text{g}\,\text{cm}^{-2}$. The difference of $15.9 \pm 1.3\,\text{g}\,\text{cm}^{-2}$ is smaller than the radiation length $\lambda_{e^\pm} = \frac{9}{7}\lambda_{\text{brems}} = 25.4\,\text{g}\,\text{cm}^{-2}$[1]. One $\lambda_{e^\pm}$ would be the expected difference if the Lorentz boosting effect is neglected. This relativistic effect causes the five events at the lower edge of the single π^0 distribution. Here, the π^0 interacts and results in a regular proton shower with a significantly lower $\langle X_{\max}^{\text{single}} \rangle$ of around $800\,\text{g}\,\text{cm}^{-2}$. This shifts $\langle X_{\max}^{\text{single}} \rangle$ to lower values. The same occurs for the split π^0 events, but it is less probable as the energy is lower and both π^0s would need to decay. Note that the

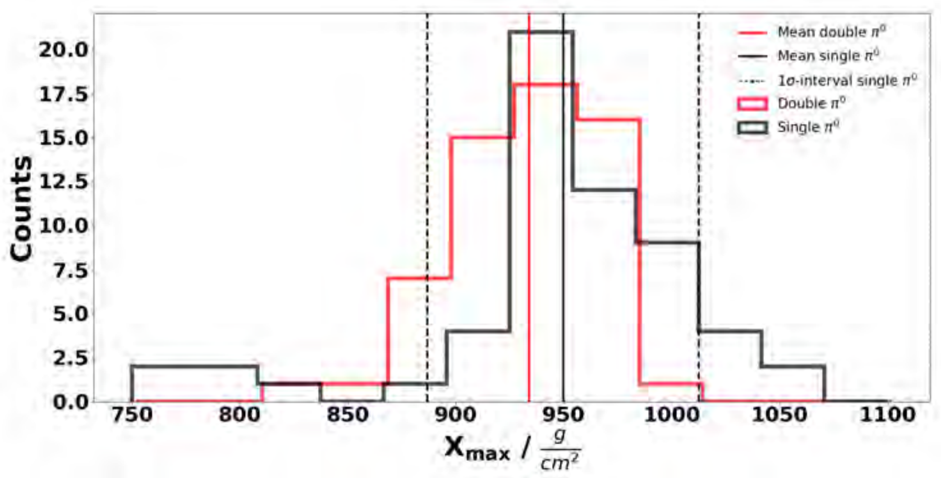


Figure 4.1: Distributions for $X_{\max}$ of air showers with single and split π^0s are shown.

1σ-band of the single π^0-distribution includes most of the split π^0 events. This makes single and split π^0 events hard to distinguish.

4.2 Discrimination with X_{max} and N^{μ}_{max}

The difference between photon and proton showers is much larger than the shower-to-shower fluctuations of protons. The best way to discriminate between photons and protons is to use X_{max} and N^{μ}_{max}. Nearly all other discrimination parameters, regardless of being atmospheric or ground based, result from the differences of these two shower properties and therefore should be less powerful. It is important for both X_{max} and N^{μ}_{max} that the shower maximum of the shower is above ground and only these events are regarded in the following. Hence, zenith angles below 30 ° are not included, as their shower maximum is often below ground at these energies.

The differences in X_{max} and N^{μ}_{max} are shown in Fig. 4.2. Both parameters are energy dependent, with a smaller dependence for X_{max}, which can be resolved by binning in energy. For the binning used, an energy related difference of, at maximum, 2.6 radiation lengths is expected. This smears the distribution slightly, but is smaller than the observable shower-to-shower fluctuations. The energy dependence is larger for N^{μ}_{max}, which can be resolved by normalizing by the energy. Quality cuts were made by using the 5 and 95 % quantiles for X_{max} and the 5 and 90 % quantiles for N^{μ}_{max}. In order to reduce the dimension, a Fisher Discriminant analysis is applied. The distribution of the resulting Fisher Discriminant, as an example, for one bin in E and θ is shown in Fig. 4.3. One can already see good discrimination power. The discrimination power gets better if one re-weights the entries by energy and $k_{\Sigma\pi^0}$. This leads to the $k_{\Sigma\pi^0}$ dependent discrimination, shown in Fig. 4.4b.

Nearly perfect discrimination can be observed. As an example, in Fig. 4.4a, a cut at 1σ results in an impurity of $(4 \pm 3)\cdot 10^{-7}$. This means one would need to measure more than 10 M protons to expect to see a single non separable proton. This large number of required protons is unrealistic given the observed fluxes at these energies. This still holds true when the cut is shifted toward the proton distribution by a few σ. Therefore, one can obtain an efficiency near 100 % with a negligible proton background. The proton background varies depending on energy and zenith angle range. For lower energies and zenith angles the discrimination gets even better as shown in Fig. 4.4b. This results, amongst other parameters, from the previously described Lorentz boosting. The discrimination power for each of the tested energy and zenith angle ranges is shown in Tab. 1. The uncertainties were calculated following the method in [22].

Table 1: Proton background for a $X_{\max}$ and $N_{\max}$ separation.

Cut from median	Zenith angle range / °
Energy / log10(eV)	Value

-1σ	30-40	40-50	50-60
18.8-19.6	0	$(8 \pm 8)\cdot 10^{-7}$	$(4 \pm 4)\cdot 10^{-6}$
19.6-20.5	0	0	0

$\pm 1\sigma$	30-40	40-50	50-60
18.8-19.6	0	$(2 \pm 2)\cdot 10^{-6}$	$(1 \pm 1)\cdot 10^{-6}$
19.6-20.5	0	0	$(4 \pm 3)\cdot 10^{-7}$

$\pm 2\sigma$	30-40	40-50	50-60
18.8-19.6	$(4 \pm 4)\cdot 10^{-6}$	$(7 \pm 6)\cdot 10^{-6}$	$(3 \pm 2)\cdot 10^{-2}$
19.6-20.5	$(6 \pm 6)\cdot 10^{-7}$	$(2 \pm 2)\cdot 10^{-3}$	$(1.3 \pm 0.9)\cdot 10^{-7}$

$\pm 3\sigma$	30-40	40-50	50-60
18.8-19.6	$(8 \pm 8)\cdot 10^{-6}$	$(3 \pm 2)\cdot 10^{-5}$	$(4 \pm 2)\cdot 10^{-5}$
19.6-20.5	$(9 \pm 9)\cdot 10^{-6}$	$(2 \pm 2)\cdot 10^{-3}$	$(2 \pm 2)\cdot 10^{-5}$

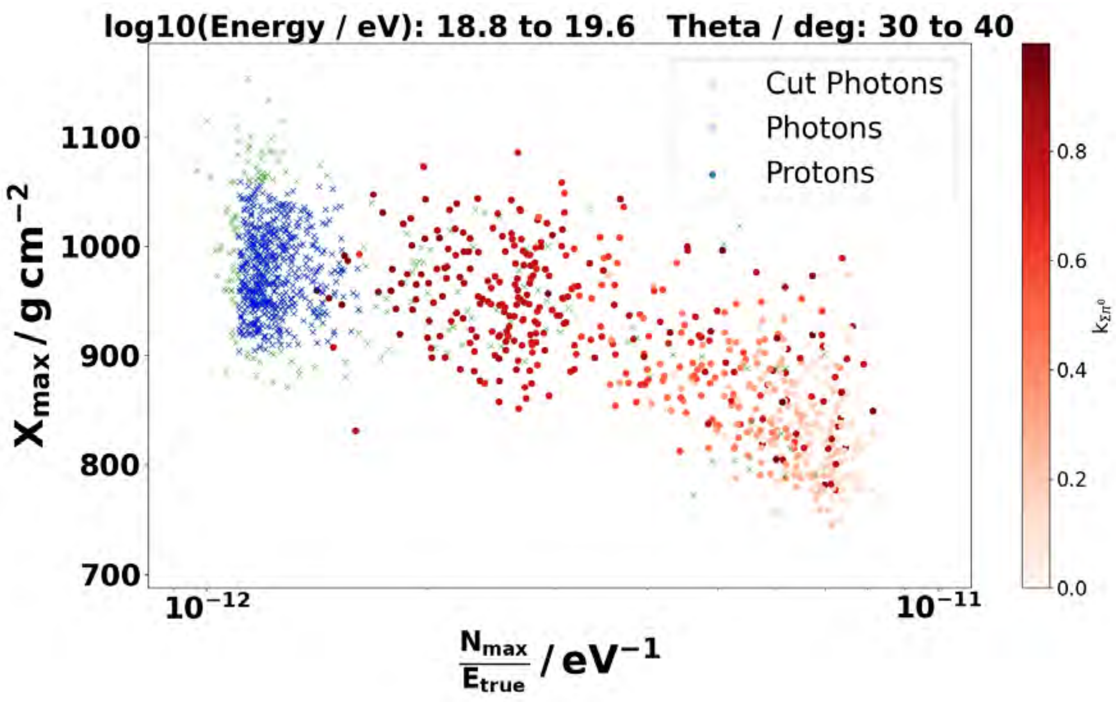


Figure 4.2: The $k_{\Sigma\pi^0}$ dependence of $X_{\max}$ and $N^{\mu}_{\max}$ for protons is shown. The photons are divided by the quality cuts into accepted (blue crosses) and rejected (green crosses) events. Uncertainties are not shown to keep the plot legible.

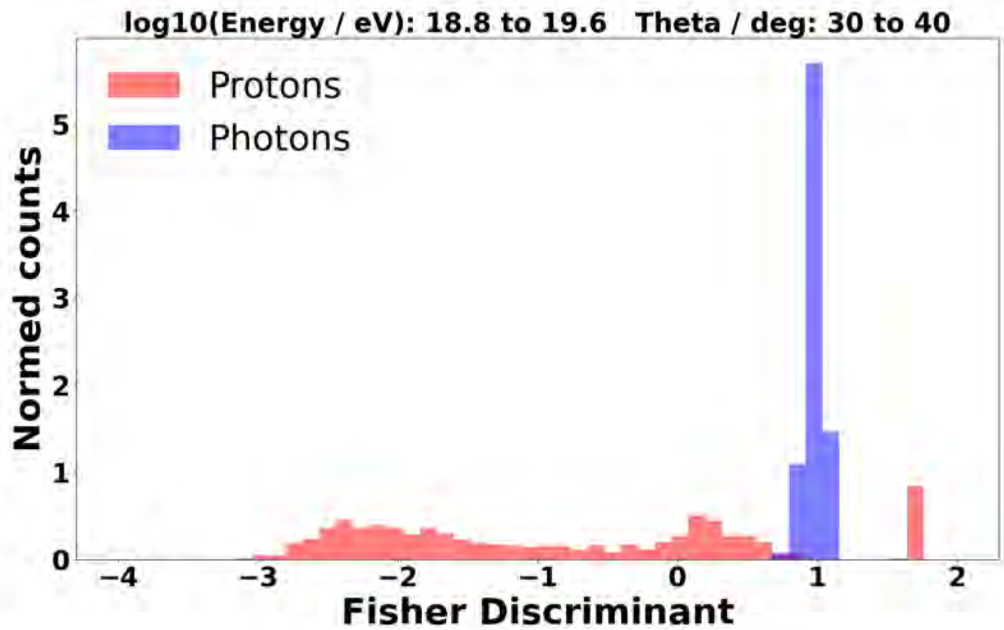
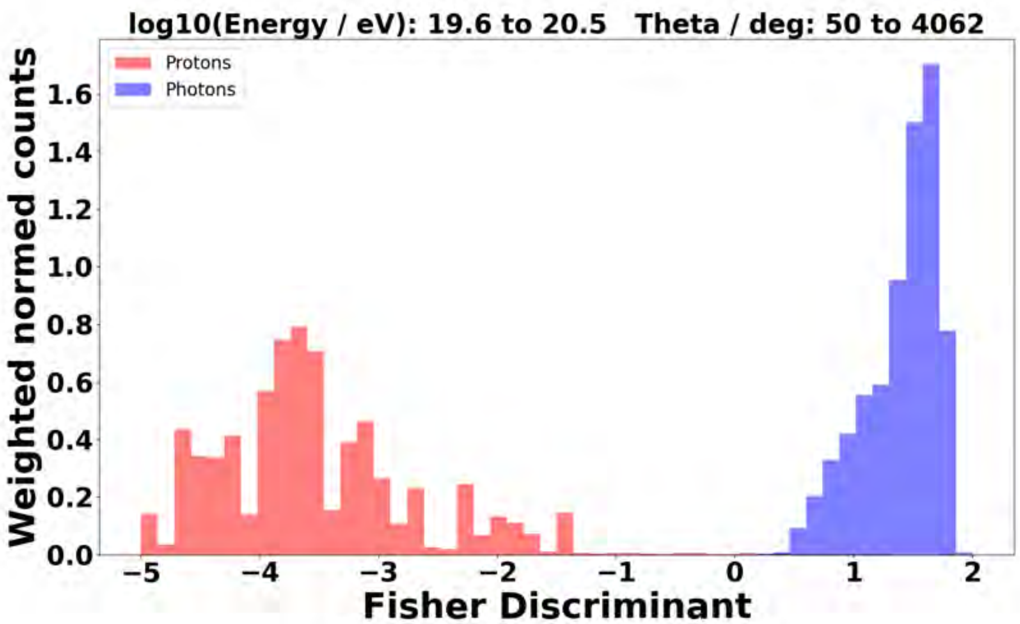
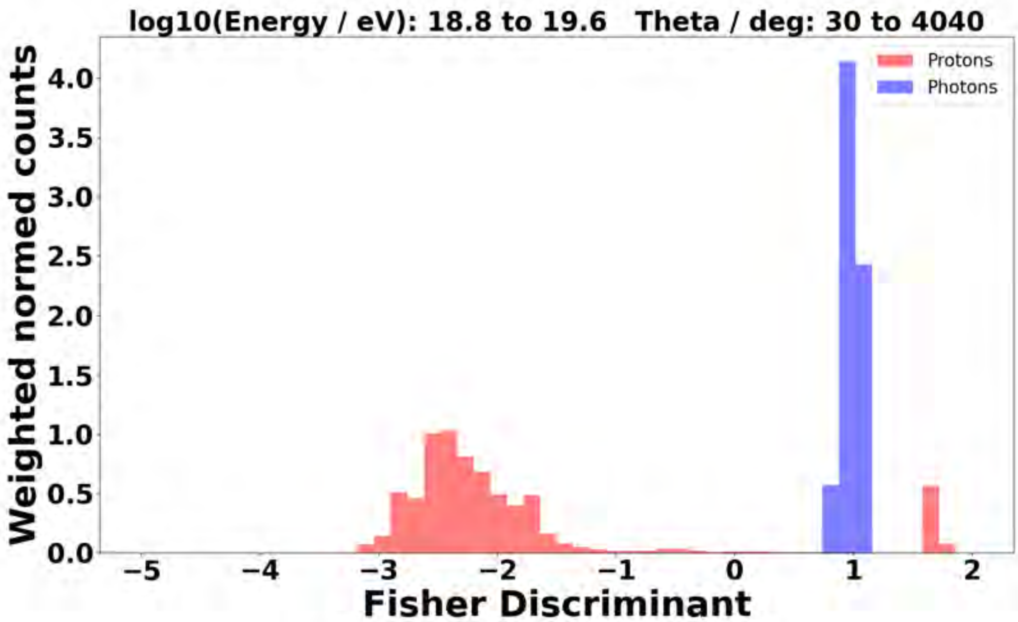


Figure 4.3: The distribution of the Fisher Discriminant is shown for the energy range of $10^{18.8}$ to $10^{19.6}$ eV and zenith angle range of 30 to 40 °. A good separation between photons (blue) and protons (red) is visible.



(a)



(b)

Figure 4.4: The separation power of the Fisher Discriminant analysis is shown depending on the $k_{\Sigma\pi^0}$. **(a)** shows the discrimination for high zenith angles and energies. The separation is weak compared to **(b)**, which shows the same discrimination for low energies and zenith angles.

4.3 Discrimination with $X_{\mathbf{max}}$

Realistically, measuring the maximum number of muons is quite difficult and experiments can only give an estimate by analyzing the shower. The reason for this is simple. Muons are more difficult to measure with respect to other charged particles due to their lower cross section. This means they traverse large distances without interacting. Thus, measuring the muonic component and especially N^{μ}_{max} is difficult. X_{max}, however, is a parameter very descriptive for the shower and more straightforward to measure. Experiments like the Pierre Auger Observatory use the depth of shower maximum to help separate photons from hadrons.

The separation power of X_{max} is lower than N^{μ}_{max}. This is because the ratio between the shower-to-shower fluctuations for a single primary type and the difference in its X_{max} compared to another primary type is low. An example is shown in Fig. 4.5 at low energies and zenith angles. Low $k_{\Sigma\pi^0}$ showers have a mean value which is 3 to 4 σ from of the photon median. High $k_{\Sigma\pi^0}$ showers are more dense 1 to 2 σ apart from the photon median and have approximately the same number of events on the other side of the photon median. Hence, a cut at 1 σ results in a weak discrimination and a proton background of $(4.0 \pm 2.5)\cdot 10^{-2}$. This is approximately five orders of magnitude larger than that of the analysis, which included N^{μ}_{max} as discrimination parameter. The discrimination power further worsens at higher energies and zenith angles. An exemplary case can be seen in Fig. 4.6 and in Tab. 4 in appendix A.

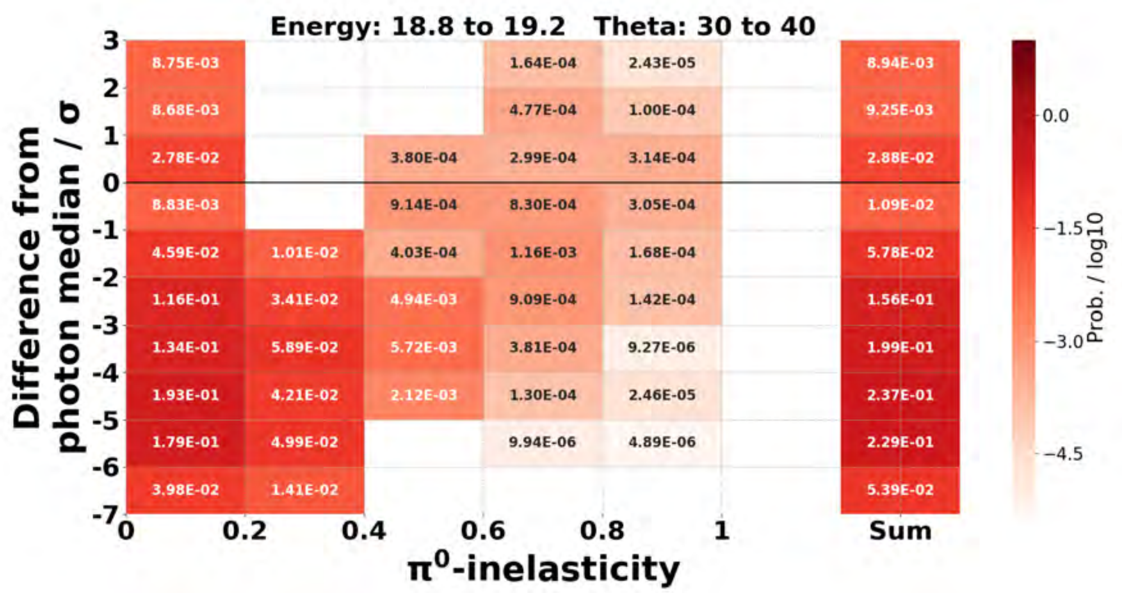


Figure 4.5: The discrimination power of the X_{max} parameter is shown for the energy range of $10^{18.8}$ to $10^{19.6}$ eV and zenith angle range of 30 to 40 °.

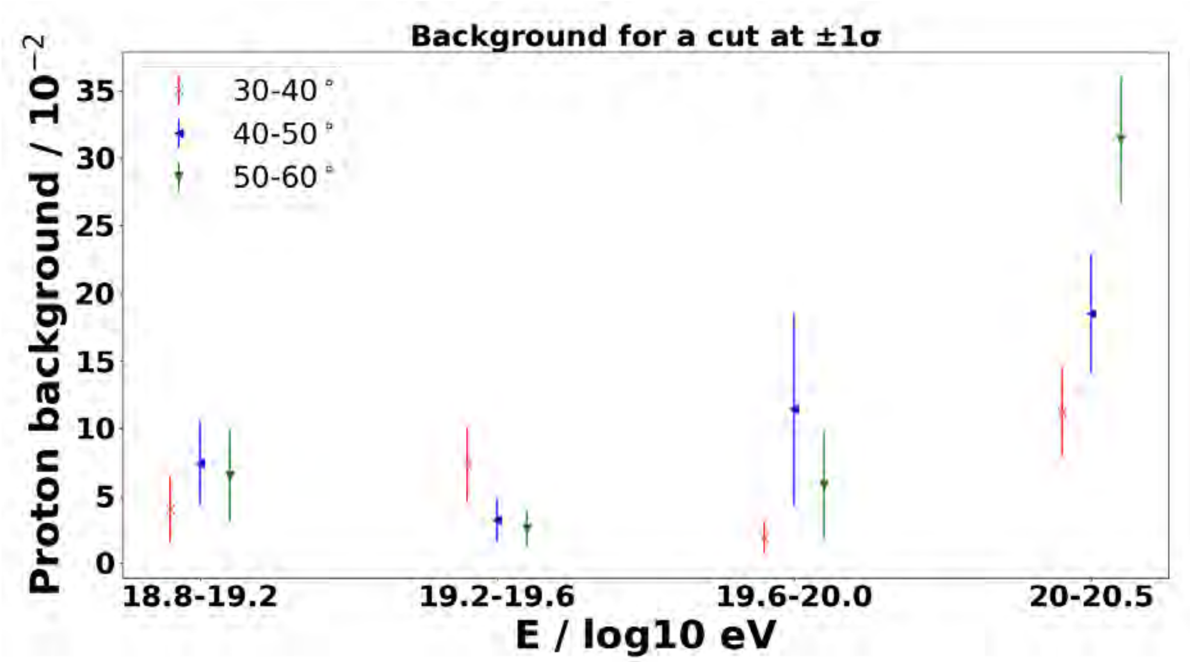


Figure 4.6: Proton background for a X_{max} separation shown for different zenith angle ranges (30-40 °: red, 40-50 °: blue, 50-60 °: green). The data points are shifted energy wise for a more clear plot.

4.4 Discrimination with a generic, idealized ground based detector

As found in chapter 4.2, N^{μ}_{max} is the key for a good photon-proton-discrimination. However, the parameter is difficult to directly measure, hence, shower characteristics instead arising from N^{μ}_{max} must be used. These parameters are in particular the signal risetime and the R_{NKG}, described in chapter 2.4. In order to obtain these parameters from raw simulation data, an idealized detector simulation was used.

4.4.1 Detector design

The idealized detector is designed as follows:

- The detector consists of 108 stations on a star shape grid.
- Stations are separated by an angular distance of $2\pi/12$ rad and radial distances of 200, 300, 400, 500, 750, 1000, 1250, 1600 and 2000 m from the center.
- Each station consists of a WCD with 3.54 m diameter and 1.2 m height with a $1.6 \times 0.6 \times 0.01\,\mathrm{m}^3$ SSD placed on top.

- The sampling rate of both detector components is 120 MHz (8.3 ns).
- Saturation effects are not regarded, triggering thresholds are variable and the detector has a 100 % efficiency.

The detector design was inspired by real setups and the technical properties are based on modern experiments. The absence of inefficiencies and saturation effects should lead to the strongest, achievable discrimination. The unthinning method used is described in [23]. The detector simulation was built for this study and therefore requires validation. To validate the detector simulations and show that the shower simulations return reasonable detector responses, a comparison to real data is needed.

4.4.2 Validation of the detector simulation

Two station properties that determine the discrimination power will be focused on, namely the station risetime and the integrated station signal. The station risetime for real data is discussed in [24]. In that paper, the station risetimes for real data are discussed and a distance dependent fit is done with the fit function

$$t_{1/2} = 40\,\mathrm{ns} + \sqrt{\mathrm{A}^2 + \mathrm{B} \cdot r^2} - \mathrm{A}. \tag{5}$$

These fit results can be used and compared to this work's simulations at the same energy and zenith angle range. This comparison is shown in Fig. 4.7. The fit of simulation data is similar to the observed values from [24]. The B-parameters differ by $0.01\,\mathrm{ns}^2\,\mathrm{m}^{-2}$ with fit values of 0.31 and $0.32\,\mathrm{ns}^2\,\mathrm{m}^{-2}$. The A-parameter is slightly larger with 376.5 ns for the simulated data compared to 344 ns for the real data. It can be seen that the main fraction of risetimes increases with the distance to the shower core. For distances $\geq 1000\,\mathrm{m}$ an increased smearing of the risetime distribution is visible and lower risetimes are again observable. These effects are also present in real data.

That the station risetime is correlated to physical values and is not a detector artifact should also be validated. In Fig. 4.8, a clear correlation between risetime and depth of maximum shower can be observed. Here, the risetime increases nearly linearly with the shower depth. The photons and low $k_{\Sigma\pi^0}$ protons are separated by a distinct offset, while the high $k_{\Sigma\pi^0}$ protons are found within both distributions. This offset between distributions results from the significantly differing muonic fractions.

In addition to the risetime, another important aspect to consider is the integrated station signal. The signal height should also increase with energy. Additionally, the LDF should change based on the muon fraction. To test these, first, a comparison with

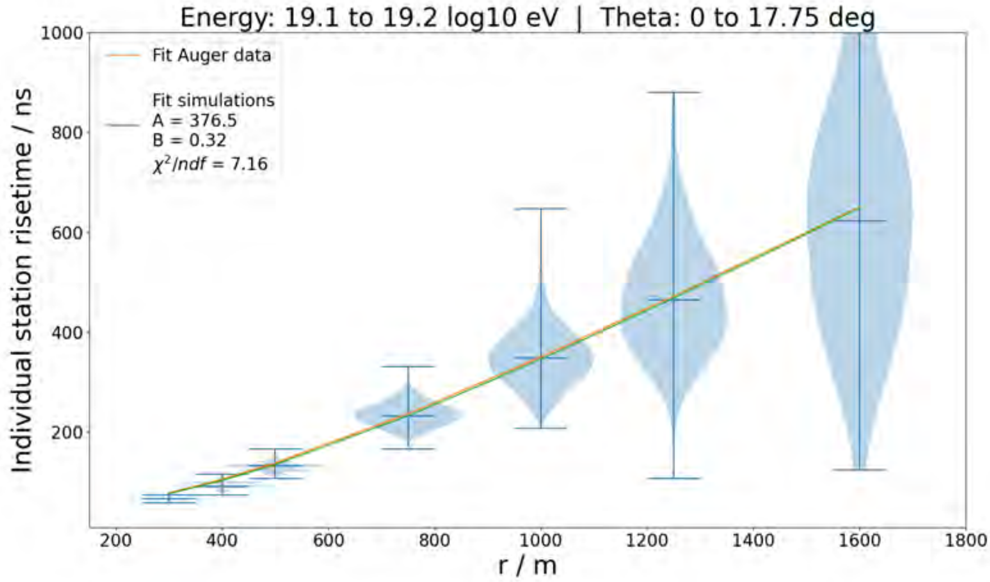


Figure 4.7: The fit (green line) of simulation data (violin plots) compared to the result found in [24] (orange line).

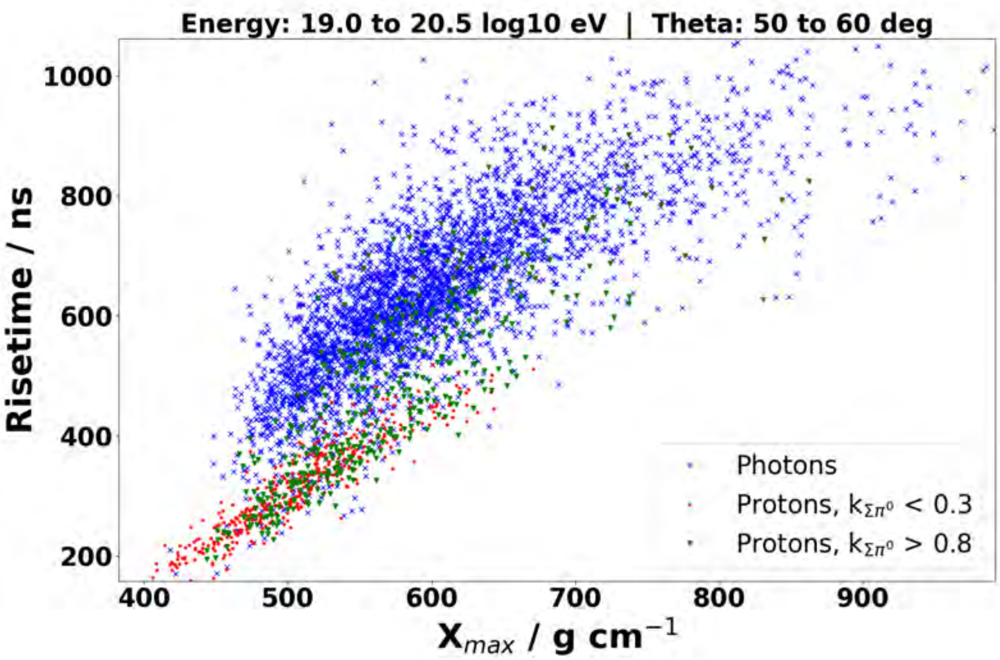


Figure 4.8: The correlation between the X_{max} and risetime is shown. The same slope is seen but an offset can be observed when comparing the photon and proton distributions. Errorbars are not shown to keep the plot legible.

real data is performed. In [25], the energy calibration for real data is shown. The energy is determined with the FD (see Chap. 2.5) at the Pierre Auger Observatory. The FD energy estimation is highly precise and is therefore comparable to the Monte Carlo energy of the simulations. In the following, a comparison between the fit of detector signals

$$E = C \cdot (S(450)/\mathrm{VEM})^D \tag{6}$$

and the results of [25] is performed. Results are shown in Fig. 4.9. Fitting the simulations results leads to an exponent D of one, which matches the fit in [25]. However, the amplitude C of the fit is lower by a factor of approximately five in the simulated data. Nonetheless, the development is similar and the detector seems to result in reasonable signal heights and displays a similar energy dependence. Additionally, the signal amplitude clearly depends on the muon fraction. It can also be observed that the low $k_{\Sigma\pi^0}$ protons have a different amplitude than photons.

The energy dependence is not the only factor which is crucial for a realistic detector response, but the signal heights must also vary with increasing station distance to the

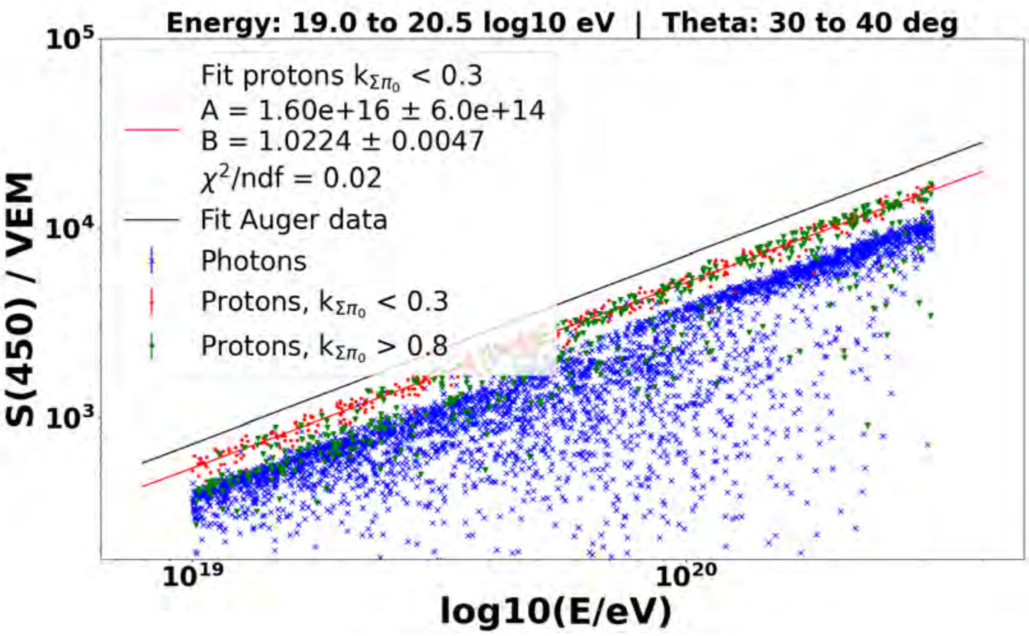


Figure 4.9: The energy dependence of the stations signal (in blue photons, in red and green protons) is shown for the energy range of $10^{18.8}$ to $10^{19.6}$ eV and zenith angle range of 30 to 40 °. The fit of Auger data (black line) is compared to the fit of low $k_{\Sigma\pi^0}$ proton events (red line). Errorbars are not shown to keep the plot legible.

shower core. Showers with a rather large muonic component should have higher signals in more distant stations. Low muonic component shower LDFs should be steeper, but can be higher close to the shower core ($< 100\,\mathrm{m}$). An example of these features is shown in Fig. 4.10.

Summarizing from these results it is clear that the designed detector as well as the simulations return reasonable results and are similar to real data. Additionally, the detector responses correlate well with physical values. Hence, there are no indications that these important behaviors are artifacts of the detector. It is also worthwhile to note that the proton simulations were compared to real data, which of course can be a mixture of protons and heavier nuclei. Additionally, there is no real data available to compare to the photon simulations at the highest energies.

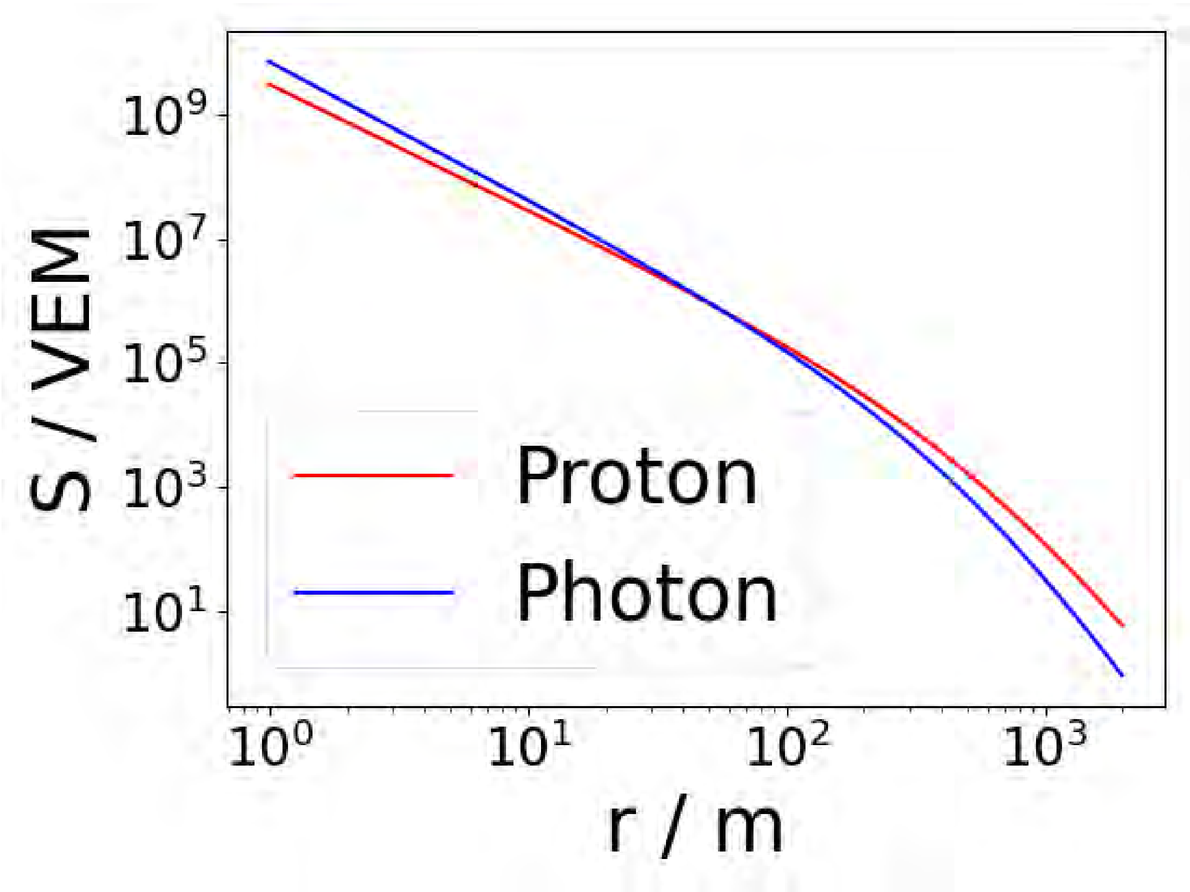


Figure 4.10: LDFs for two example showers with $E \approx 10\,\mathrm{EeV}$ and $\theta \approx 35°$ (red: proton, blue: photon primary). The station signals right at the shower core are interpolated and might not be realistic.

4.4.3 Discrimination parameter distributions

Following the validation of detector, the discrimination parameters can be studied. The parameters are ordered from strongest to weakest in discrimination power. The example plots shown are for the WCD, the SSD plots can be found in Appendix B. The SSD parameters have a smaller discrimination power. This is due to the relatively low muon sensitivity of the SSD, which is an especially important component for discriminating between photons and high $k_{\Sigma\pi^0}$ events.

First considering R_{NKG}, which distribution is depicted in Fig. 4.11. Good discrimination between protons and photons is observed for low zenith angles. For angles up to $50\,°$ perfect discrimination between low $k_{\Sigma\pi^0}$ protons and photons is achieved by cutting at the mode of the photon distribution. High $k_{\Sigma\pi^0}$ proton events have a similar distribution as regular protons only differing by a small shift to lower values. Occasionally, one can obtain few high $k_{\Sigma\pi^0}$ proton events with lower values than the photon mode. The discrimination worsens for zenith angles above $50\,°$. For these more horizontal showers the ratio $\frac{S_f}{S_v}$ increases, where S_f is the signal of stations before the shower core on ground (azimuth wise) and S_v the signal of stations behind the shower core. While for vertical showers this ratio is approximately one, an asymmetry is observed for horizontal showers (see Fig. 4.12). This directly causes a deviation from the expected LDF. While the absolute deviations are rather small for large distances from the shower core, the relative deviations reach values up to 100. As described, R_{NKG} is the averaged relative deviation from the fit, which serves as explanation for increasing R_{NKG} values for both protons and photons at high zenith angles. These larger values at high zenith angles lead to smearing of the distributions and weakens the discrimination power.

The γ parameter distributions shown in Fig. 4.13 are similar to the R_{NKG} distributions. For angles below $50\,°$, there is a nearly perfect separation between low $k_{\Sigma\pi^0}$ protons and photons. However, high $k_{\Sigma\pi^0}$ protons are found at the center of the photon distribution. Similar to R_{NKG}, high zenith angles lead to a significant weakening of discrimination power. It is remarkable that the low $k_{\Sigma\pi^0}$ proton mode remains around zero, whereas the distribution is smeared towards lower values. It is the opposite for the photon distribution, which shifts higher towards the proton distribution. Also worthwhile mentioning is the behaviour of the high $k_{\Sigma\pi^0}$ protons, which shift to even higher values compared to the low $k_{\Sigma\pi^0}$ protons. This feature can be explained by the distinct EM components. The EM component is concentrated at the shower core. At high zenith angles the shower core spreads over larger radial distances in the detector plane. This flattens the LDF and shifts the γ distribution to higher values. This also holds true for photons and high $k_{\Sigma\pi^0}$ protons. Hence, the muonic component vanishes for mid-ranges,

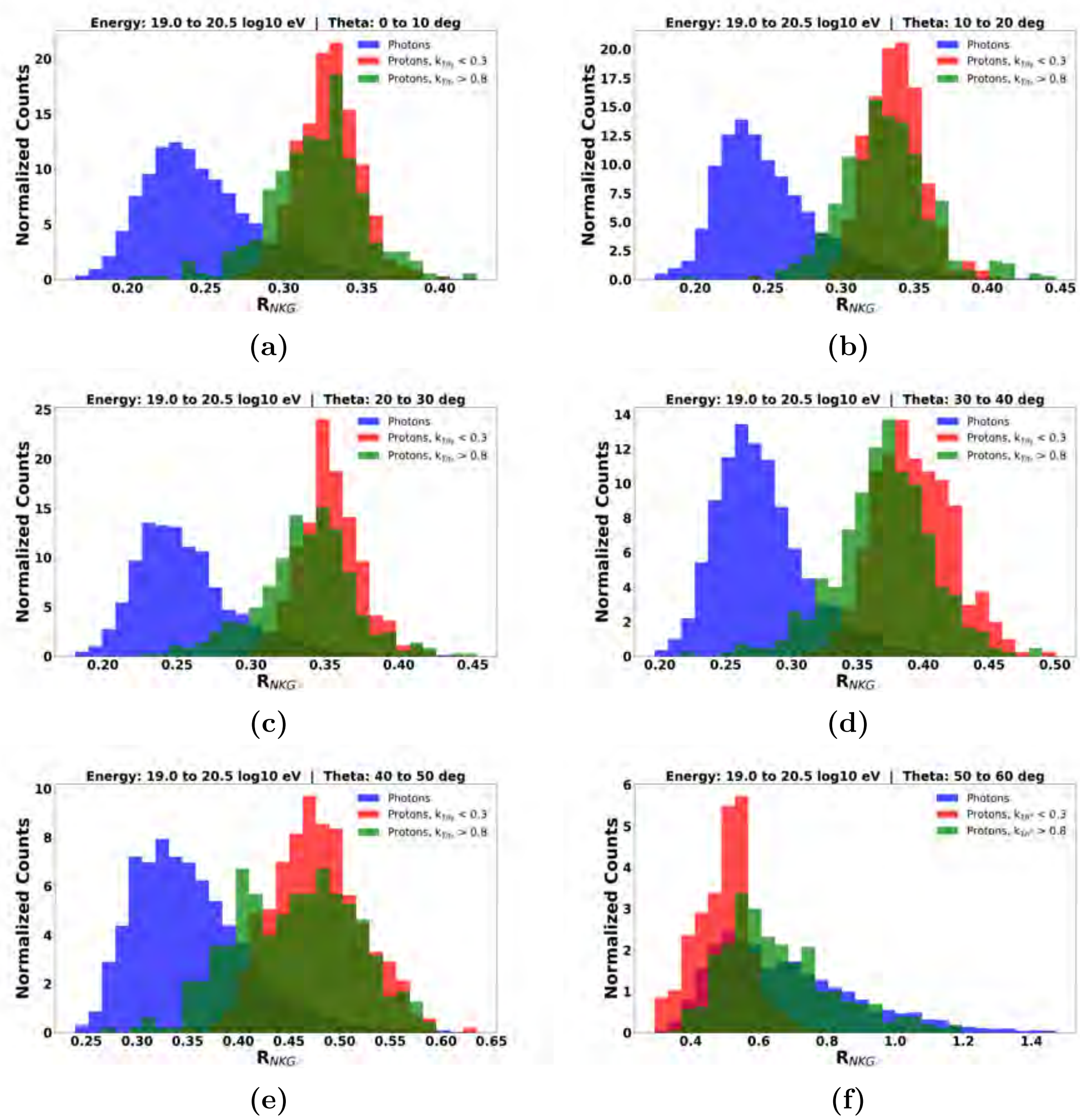
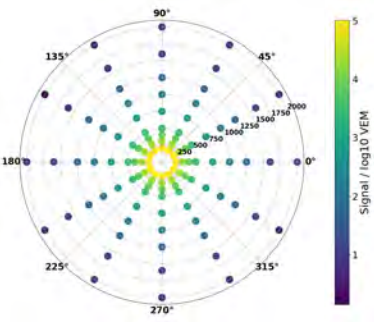


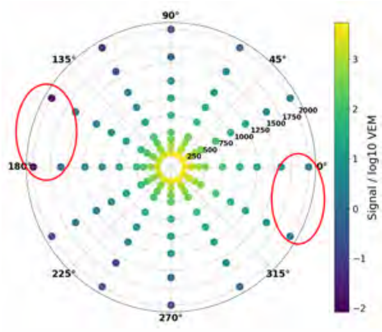
Figure 4.11: The distribution of the R_{NKG} parameter is shown for photons (blue), low (red) and high (green) $k_{\Sigma\pi^0}$ proton events.

while the EM component increases. However, the low $k_{\Sigma\pi^0}$ protons are not dominated by the EM component. Therefore, depending on the ratio of muonic and EM component, one can obtain lower or higher LDF γ value.

The discrimination powers of R_{NKG} and γ both have their weakness at high zenith angles. In contrast, the risetime parameter has its best discrimination power at the highest zenith angles, shown in Fig. 4.14. This arises from the path difference in the atmosphere for particles produced at different parts of the shower. This path difference increases for more horizontal showers, whereas there is no difference for vertical ones. A missing path difference automatically leads to a significant decrease of the difference

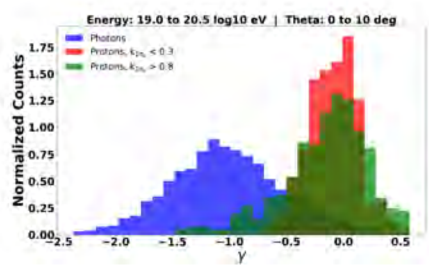


(a)

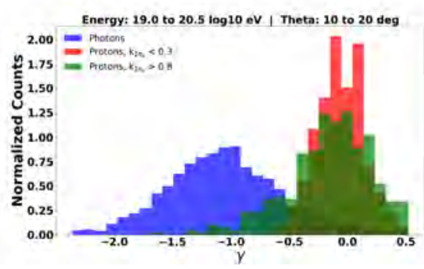


(b)

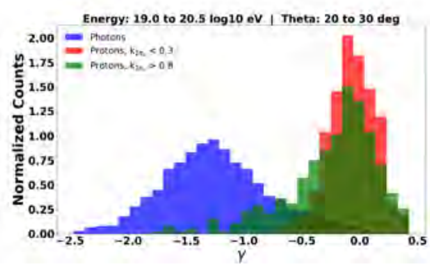
Figure 4.12: **(a)** Integrated signal (colored points) in the detector plane caused by a vertical air shower. No asymmetry effects are observable. **(b)** For inclined shower asymmetry effects are observable marked by the red ellipses.



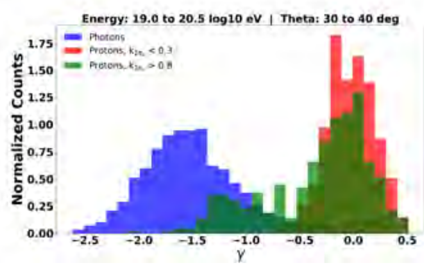
(a)



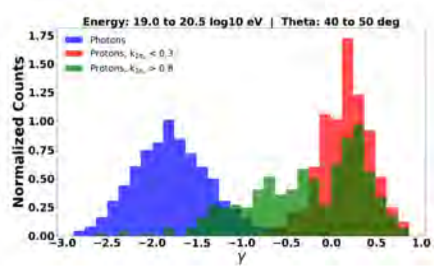
(b)



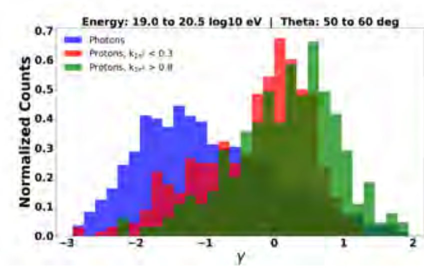
(c)



(d)



(e)



(f)

Figure 4.13: The distribution of the γ parameter is shown for photons (blue), low (red) and high (green) $k_{\Sigma\pi^0}$ proton events. With increasing zenith angles a worse separation can be observed.

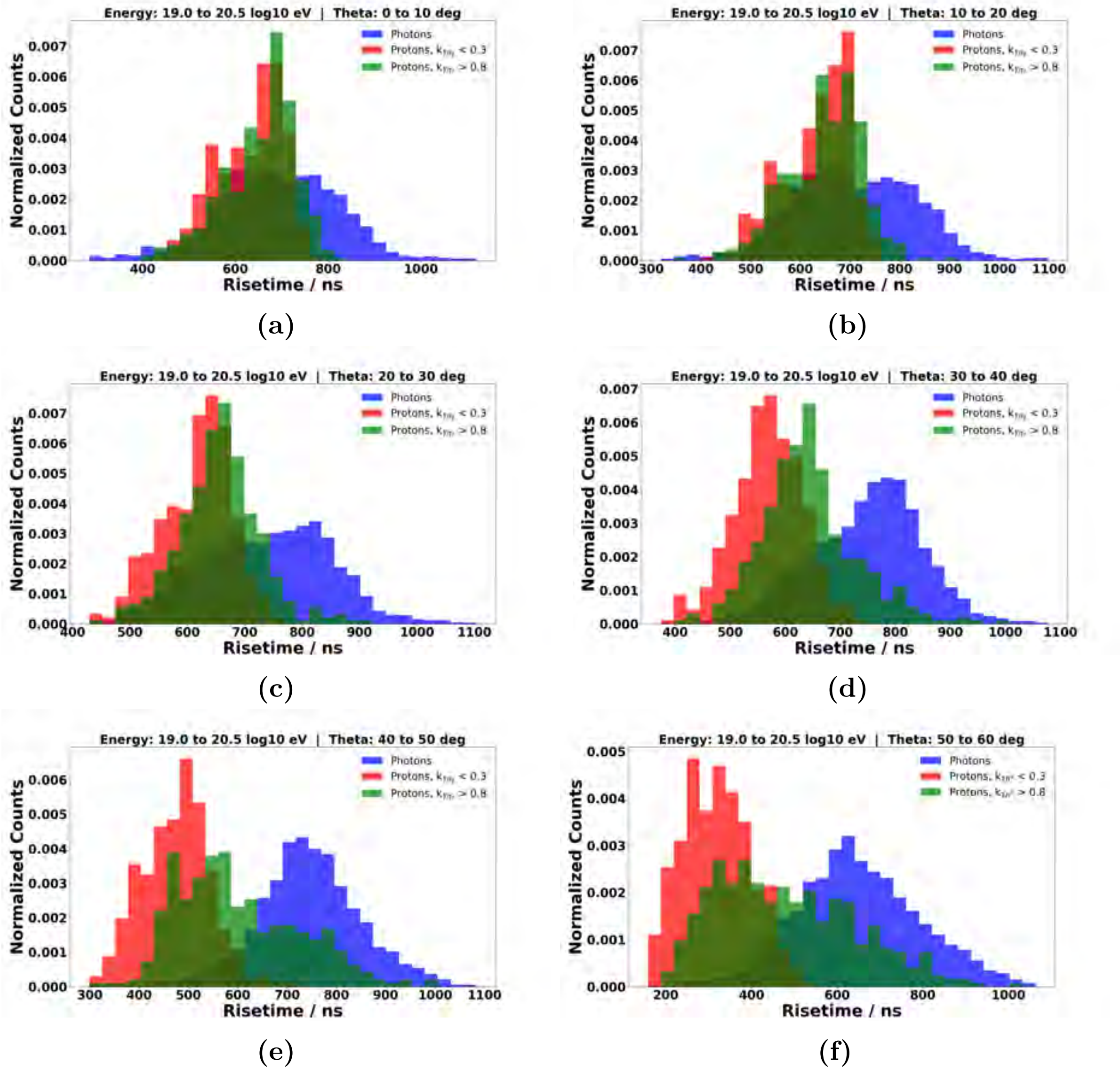


Figure 4.14: The distribution of the risetime parameter is shown for photons (blue), low (red) and high (green) $k_{\Sigma\pi^0}$ proton events. With increasing zenith angles a better separation can be observed.

between photons and protons. Only at zenith angles larger than 30°, one can obtain a discrimination between low $k_{\Sigma\pi^0}$ protons and photons that is as significant as that for R_{NKG} or γ. Note the large overlay of high $k_{\Sigma\pi^0}$ protons and photons, which is more significant than before.

A possible explanation for the large overlap come from the parameter characteristics. The LDF parameters arise from the EM and muonic components. Even for high $k_{\Sigma\pi^0}$ events (for example a $k_{\Sigma\pi^0}$ of 0.9), a non-negligible fraction of the primary energy contributes to the muonic component. This component results in a LDF which is similar

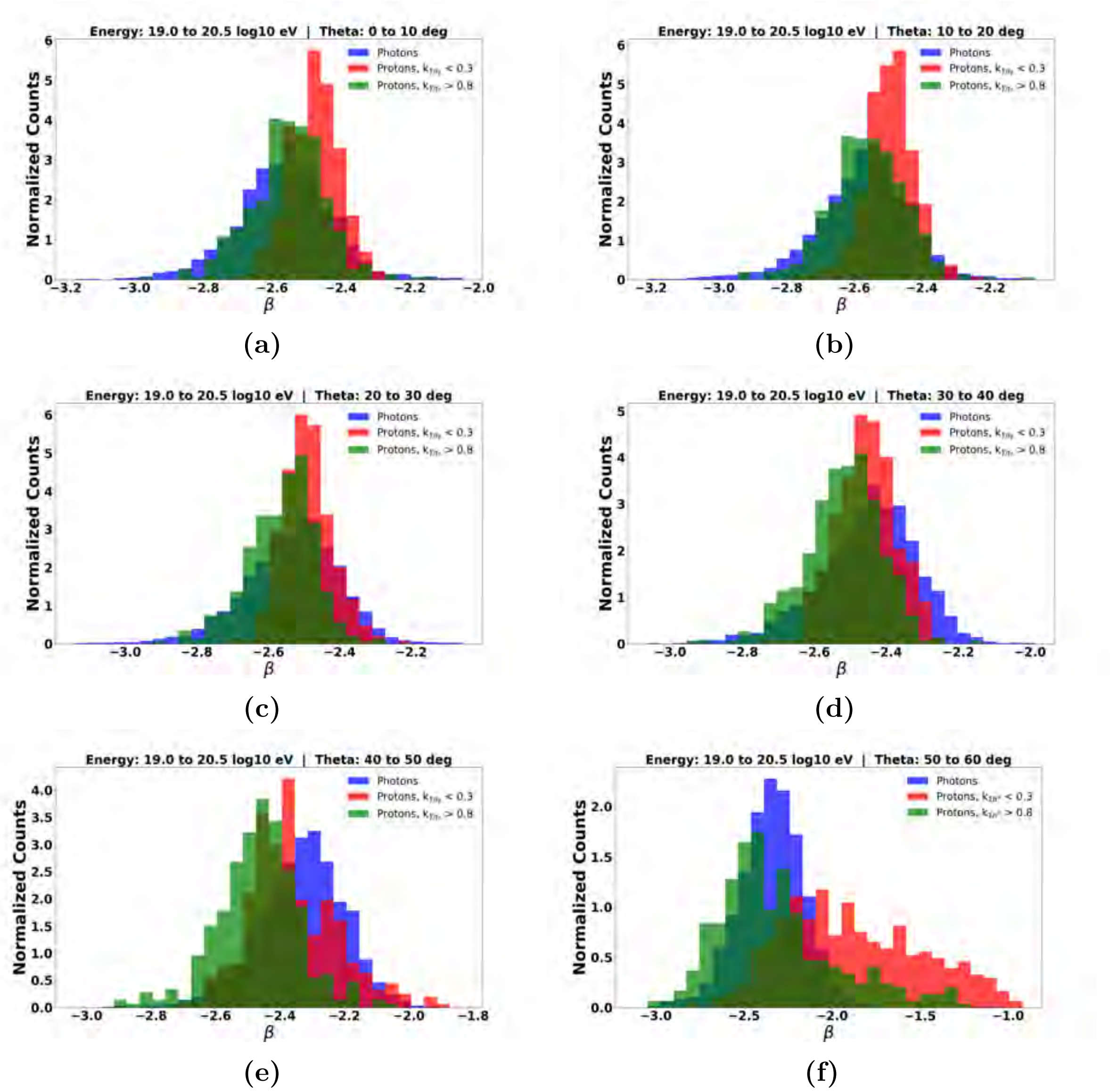
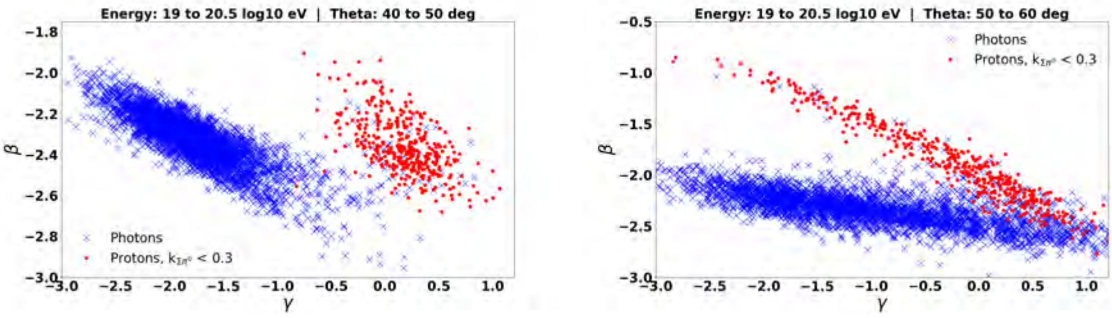


Figure 4.15: The distribution of the β parameter is shown for photons (blue), low (red) and high (green) $k_{\Sigma\pi^0}$ proton events. With increasing zenith angles a better separation can be observed.

to a regular proton. However, this low muonic fraction is not relevant to the risetime parameter. The muonic component is too low to raise the signal over the 20 % threshold of integrated signal. Therefore, here only $X_{\max}$ yields significant discrimination power.

The parameter β has similar and well defined distributions for the different primary types (see Fig. 4.15). However, the distributions nearly completely overlap. Only at the highest zenith angles can one obtain at least a weak separation due to the smearing of low $k_{\Sigma\pi^0}$ protons towards higher values.



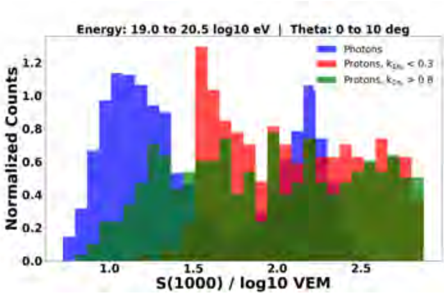
(a) **(b)**

Figure 4.16: Distribution of photons (blue) and protons (red) in the γ-β plane. A clear correlation can be seen.

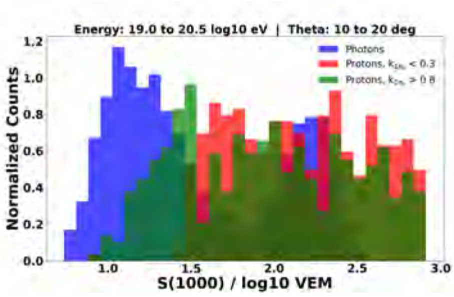
The two fitting parameters β and γ are correlated. Figure 4.16 shows their correlation. It can be observed that distinct selections on just one parameter leads to worse discrimination powers than when a linear or higher dimensional cut in the 2D-space is used. As an example, for angles larger than 50 °, one cannot obtain photon efficiencies $> 50\,\%$ while requiring a proton impurity of 0 %. In 2D-space this can easily be achieved using a linear cut or elliptical selection around $\gamma = -1.5$ and $\beta = -2.2$.

Three parameters arising from the LDF also have rather low separation power. One of the parameters is S(1000), which is shown in Fig. 4.17. At low zenith angles, one obtains low values for photons and larger values for low $k_{\Sigma\pi^0}$ protons. The high $k_{\Sigma\pi^0}$ protons spread over the whole range and are therefore hardly distinguishable. A cut at around 30 VEM would lead to good discrimination to some extent. However, the efficiency would be rather low.

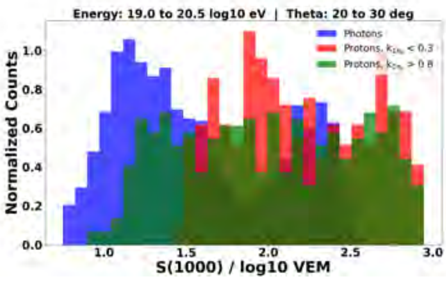
This results from the shape of the photon distribution at zenith angles $< 40\,°$, which consists of two peaks, one at 15 VEM and a smaller maximum at 200 VEM. Two causes of this feature are shown in Fig. 4.18a. At highest energies of 10^{20} eV, two effects have to be taken into account. First, the preshower effect [8] which reduces shower-to-shower fluctuations. Preshowering leads to splitting of primary photons producing many photons and few e^-/e^+-pairs at heights of around 1000 km. This larger number of particles entering the atmosphere leads to lower shower-to-shower fluctuations which yields sharper peaks in the photon distribution. The second reason is the increasing cross section for photonuclear reactions which leads to a station signal difference of air showers with E lower and higher than 10^{20} eV. At $E > 100$ EeV, this cross section even exceeds pair-production cross section. Hence, the hadronic component of the photon shower increases leading to a second peak in the photon distribution. This is observable in both S(1000) and S_b. It also results in higher values for χ^2/ndf, shown in Fig. 4.18b.



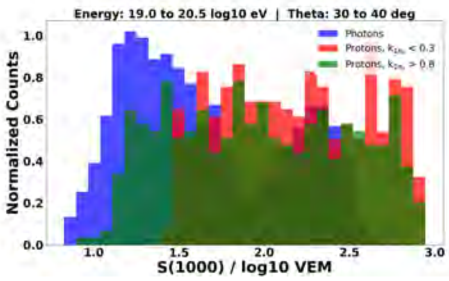
(a)



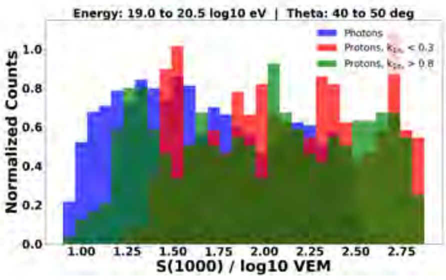
(b)



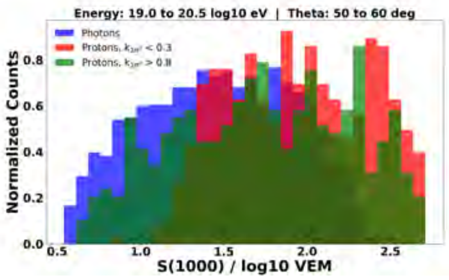
(c)



(d)

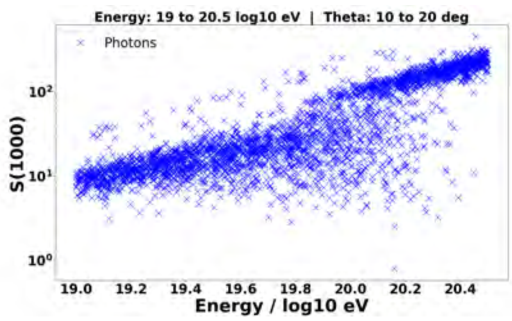


(e)

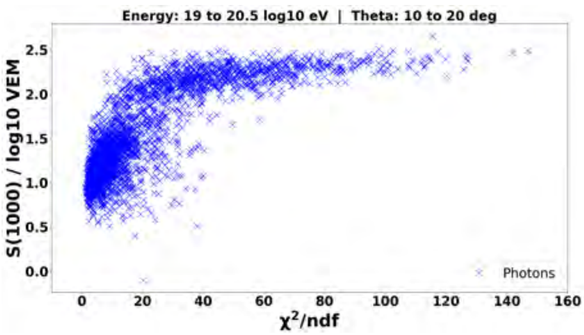


(f)

Figure 4.17: The distribution of the S(1000) parameter is shown for photons (blue), low (red) and high (green) $k_{\Sigma\pi^0}$ proton events.



(a)



(b)

Figure 4.18: **(a)** The reason of the double peak structure of S(1000) is shown for the zenith angle range 10 to 20 °. **(b)** The worsening of the χ^2/ndf for larger S(1000) values is shown for the zenith angle range 10 to 20 °.

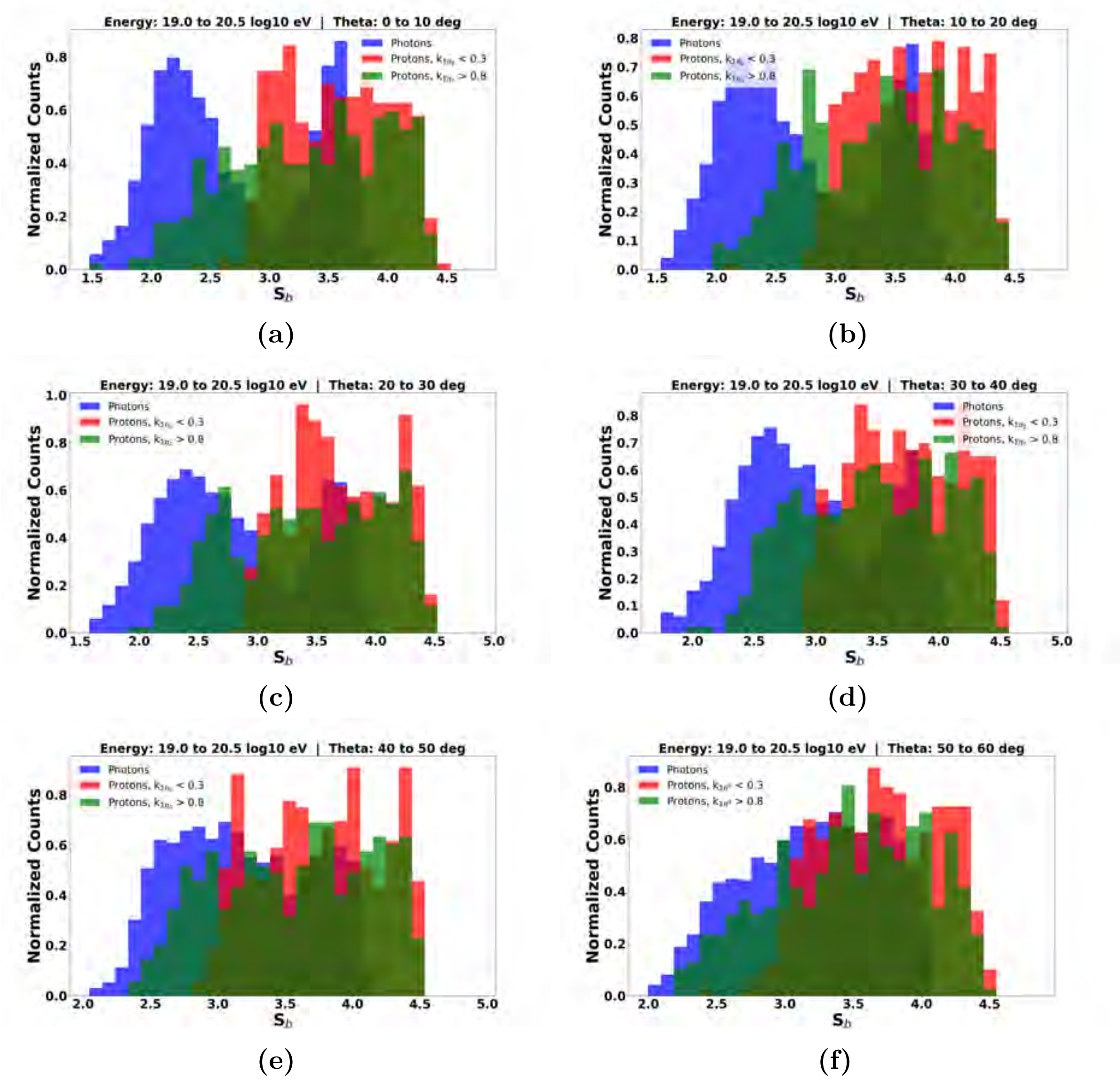


Figure 4.19: The distribution of the S_b parameter is shown for photons (blue), low (red) and high (green) $k_{\Sigma\pi^0}$ proton events.

For the highest zenith angles the two photon peaks merge and the photon and proton distributions are not distinguishable. The distributions of the S_b parameter for the different primaries are analogous to S(1000) and have no clear advantage in terms of discrimination power. The distributions for S_b are shown in Fig. 4.19.

Combined, the analyzed parameters yield a good impression of the WCD parameter discrimination powers. However, as mentioned, the SSD has weaker discrimination powers and is more useful in combination with the WCD. As an example, the γ-parameter, with its weak discrimination power at highest zenith angles, benefits from the connection of WCD and SSD as shown in Fig. 4.20. The marked area in the plot includes

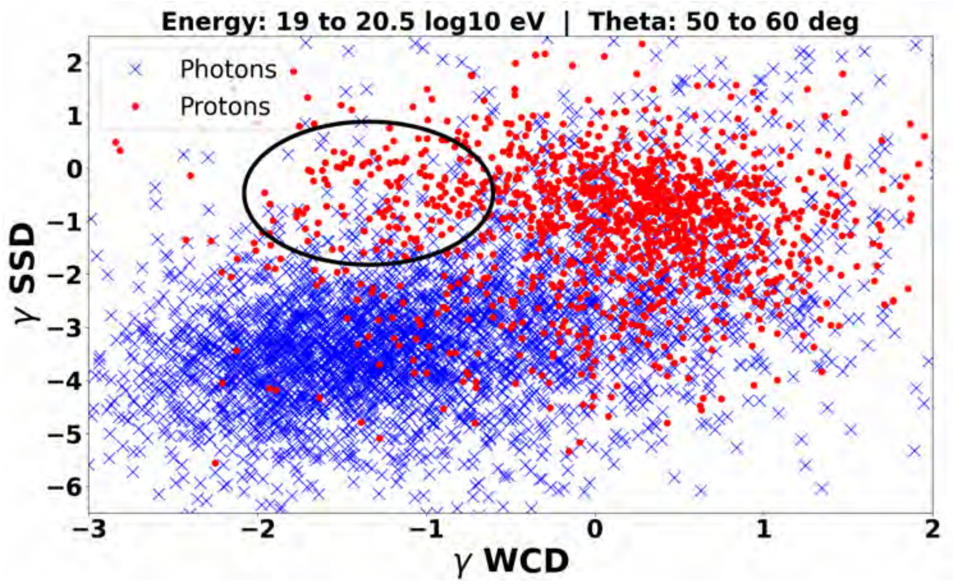


Figure 4.20: The benefit of using a combination of WCD and SSD for the discrimination of protons (red) and photons (blue) is shown. The black ellipse marks the region with events, that can additionally be labeled correctly in the WCD-SSD plane.

protons and photons. On the one hand, protons in the black ellipse can be identified, which are located at the photon mode in the WCD 1D-distribution. These protons can now be classified as such, which results in larger purity for the photon selection. On the other hand, photons might be classified as protons as they are centered in the SSD 1D-distribution, which would reduce the photon efficiency.

Clearly, in order to reduce proton background, the events from the marked area should be excluded. However, it is up to discussion whether the events can be identified correctly by including all parameters. The amount of dimensions and correlations which arise when trying to use all of the presented parameters simultaneously diminishes the possibility of creating cuts manually. Because of this, a machine learning network is used.

4.4.4 Discrimination using a random forest

The random forest machine learning algorithm (RF) [26] was selected. RFs consist of N decision trees. The trees differ by included feature selection and bootstrapping of training data. Each tree is built independently based on the entropy criterion and votes on the dataset, whereby the event is classified by the tree majority. RFs have advantages and disadvantages as follows:

Advantages:

- **Accuracy**: RFs outperform gradient decision trees in terms of accuracy.
- **Robustness**: The RF is robust against overfitting. More precisely, the single trees over- or underfit, but these cancel each other out. Therefore, overfitting is usually not a problem.
- **Stability** Compared to gradient decision trees, which are strongly variable depending on the training dataset, RFs with a large number of trees are quite stable.
- **Speed**: The training speed of a RF is low for a complex machine learning method. The trees are calculated in parallel and which only increases CPU time linearly with the number of trees.

Disadvantages:

- **Visualization**: Single trees and their decisions can be visualized, but it is not feasible for hundreds of trees.
- **Interpretation**: RFs are black boxes. Hence, a clean dataset is needed and training on artifacts should be prevented. The feature importance can be obtained, but it is difficult to understand for which reason certain features are important.

The first three advantages are especially important for this work. The accuracy determines the irreducible background and the stability is contributing to the background uncertainty. The robustness of the method is crucial as the high $k_{\Sigma\pi^0}$ protons in the center of the photon distribution are a risk of overfitting.

Before inspecting the results of the model, the model itself is validated. For each zenith angle range, a model consisting of a RF with 5000 trees is trained. The hyperparameters 'feature selection', 'maximal tree depth' and 'minimal entries required for a split' were found via a grid search. The trees are limited to seven to nine features which depend on the zenith angle, whereas the other hyperparameters do not need a limitation. For each zenith angle range 25 RFs were trained with different random seeds, which determine the split of train and test data sets, make the random feature selection and bootstrap the data. Due to the robustness of RFs, the fluctuations between the trained RFs are rather low. Nonetheless, the average result of all RFs together will be used.

Each RF has its own feature importance. The feature importance translates to the decrease of impurity achieved by that each split. A large value (the feature importance ranges from 0 to 1) indicates a large importance to the classification. The features are expected to have similar rankings as described in 1D-space, and the SSD features should have lower importance in general. However, the correlations mentioned earlier can change this ranking. Figure 4.21 depicts the feature importance. The plot is in agreement with previous observations and validates the model. The LDF parameters, especially those of the WCD, are crucial for the discrimination for angles below 30 °. At these zenith angles, all SSD parameters have feature importance around 0.05 or lower. At angles between 40-50 °, the WCD risetime doubles its importance and continues with a steep increase at the last zenith angle bin. Note the missing rise of the SSD risetime importance. While the WCD risetime dominates the discrimination at the highest zenith angles, the SSD risetime importance remains below 0.05. From this, it seems that the SSD risetime does not contain information, which is not already present in the WCD risetime.

While the risetime importance sharply increases, the LDF parameters γ and R_{NKG} of the WCD lose discrimination power with zenith angle. They have their highest importance from 0 to 30 °, adding up to ≈ 0.7. Subsequent, R_{NKG} loses its importance and drops from approximately 0.5 to 0.1 and thus equaling risetime's importance. The γ parameter, on the other hand, significantly increases its importance for both the WCD and SSD. Here, the WCD risetime is more important than all other features summed.

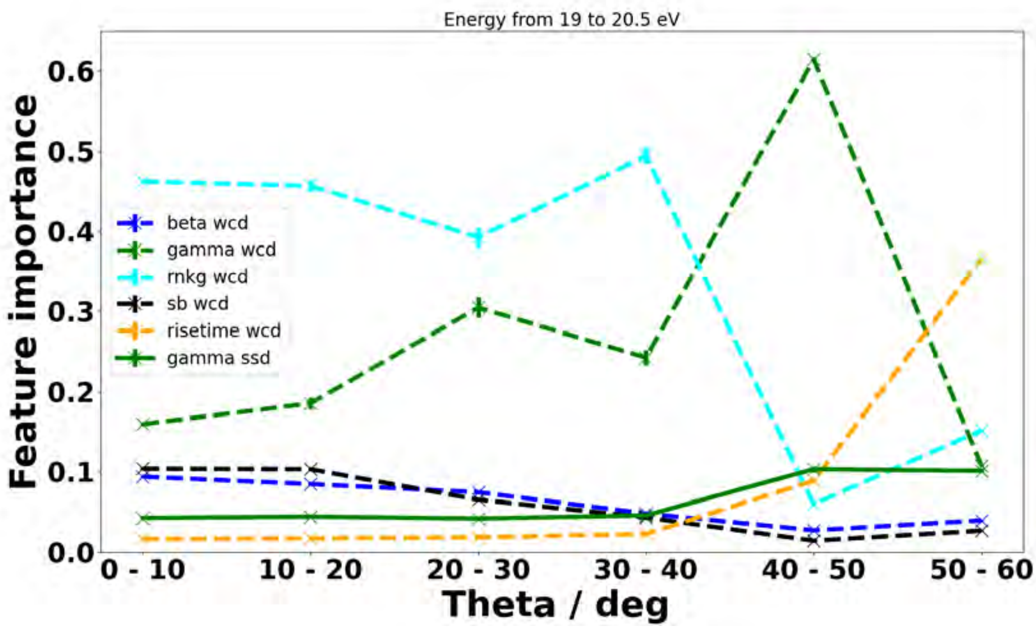


Figure 4.21: The development of the feature importance are shown for the different zenith angles. Features with importance continuously below 0.05 are not shown.

This goes along with the almost perfect discrimination observed at the 1D distributions. The SSD feature of γ is therefore the SSD parameter to become at least the second most important feature. In the last zenith angle bin, a different picture is observed. The WCD R_{NKG} importance increases again, which is the result of the decreases of both γ-parameter importance. The γ WCD feature drops its importance by 0.5 becoming equivalent to the SSD feature.

This model was then applied to the test data, which enables an interpretation of the results. First, the confusion matrices, shown in Tab. 2, are explored for the different zenith angle ranges. The confusion matrices look similar for all zenith angles below 50 °. Especially interesting are the protons classified as photons. All five lower zenith angle ranges have nearly identical values of around 28 falsely predicted protons. A similar behaviour can be observed for the incorrectly classified photons. Here however, a small increase can be seen in the range from 40 to 50 °. The zenith angles $> 50\,°$ show a more significant increase in falsely classified events. More than double the number of incorrectly classified protons can be found as compared to the other zenith angles. This leads an accuracy of $\approx 92\,\%$ at the highest zenith angles, whereas the lower zenith angles have an accuracy of $\approx 95\,\%$.

When interpreting the confusion matrices, one has to be careful as the absolute numbers of events for the different zenith angles vary. Additionally, unbalanced data

Table 2: The unweighted confusion matrices of the RF are shown for different zenith angle ranges.

0 to 10 °		Predicted	
		Proton	Photon
True	Proton	376 ± 4	28 ± 1
	Photon	29 ± 1	829 ± 6

10 to 20 °		Predicted	
		Proton	Photon
True	Proton	386 ± 4	24 ± 1
	Photon	28 ± 1	864 ± 6

20 to 30 °		Predicted	
		Proton	Photon
True	Proton	383 ± 4	28 ± 1
	Photon	28 ± 1	964 ± 6

30 to 40 °		Predicted	
		Proton	Photon
True	Proton	385 ± 4	27 ± 1
	Photon	28 ± 1	1005 ± 6

40 to 50 °		Predicted	
		Proton	Photon
True	Proton	378 ± 4	27 ± 1
	Photon	36 ± 1	998 ± 6

50 to 60 °		Predicted	
		Proton	Photon
True	Proton	346 ± 4	61 ± 1
	Photon	38 ± 1	1020 ± 6

sets are used (ratio of ≈ 3:1 of photons to protons). Hence, it will be focused on metrics, which remain significant for unbalanced data sets. The precision and recall metrics do not lose significance in case of unbalanced data sets. The precision score summarizes the fraction of correctly assigned events, that are classified as photons:

$$\text{Precision} = \frac{\text{correctly classified photons}}{\text{correctly classified photons} + \text{incorrectly classified protons}}.$$

The recall is also referred to as the sensitivity, aka the fraction of photon events, which are classified correctly:

$$\text{Recall} = \frac{\text{correctly classified photons}}{\text{correctly classified photons} + \text{incorrectly classified photons}}.$$

Clearly, the precision is more important for this work. A perfect score of 1 means a non-existent proton background. However, it is also evident that too low sensitivities would reject this method as practicable as too many photon events would be discarded.

Figure 4.22 shows both metrics for the different zenith angle ranges. For zenith angles $< 40\,^\circ$, stable scores around 0.97 are observed. Only a small increase of the scores is visible with increasing zenith angle. While the precision remains at a value of 0.97 for the zenith range from 40-50 °, the recall decreases to 0.96 and stays at that level afterwards. As mentioned, the sensitivity is an important aspect. However, a score larger than 0.9 is more than sufficient. Common photon searches are done with sensitivities of 0.5 or even lower[8]. Hence, the decrease is not important.

The observable drop for the highest zenith range of the precision is more remarkable. Within the last bin, the score drops from 0.97 to 0.94. Even though 0.94 is still a good score, it results in double the number of falsely classified protons. The reason is evident, the loss of discrimination power of the γ and R_{NKG} parameters take its toll, and the larger discrimination power of the risetime parameter cannot compensate it.

Both metrics are helpful for confirming the model. However, to determine the irreducible background, one has modify two aspects of the metrics. The first aspect concerns the classification of the events. Each classified event also receives a probability from the RF. This probability is defined by the fraction of trees that classified the event correctly. The classification of RFs works with a probability benchmark of 0.5, however, it can be changed manually. Increasing this benchmark leads to reduced efficiency while increasing purity, and so also the fraction of correctly classified events. In a search for rare events, the significance is mainly improved by selecting high purity and low efficiency. In best case the decrease of selected photons is smaller than for protons. This aspect is

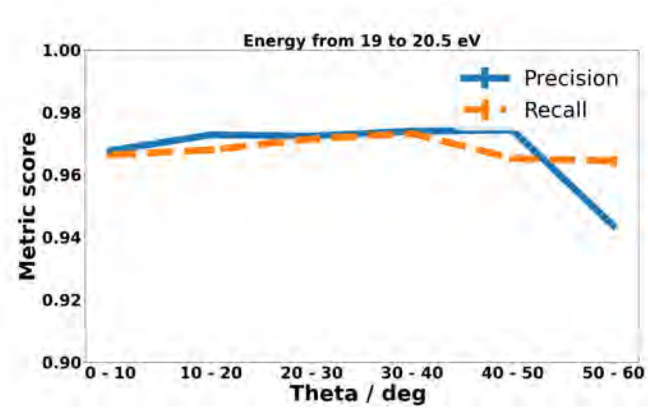


Figure 4.22: The development of the precision and recall score is shown for different zenith angles. Note the scaling of the y-axis.

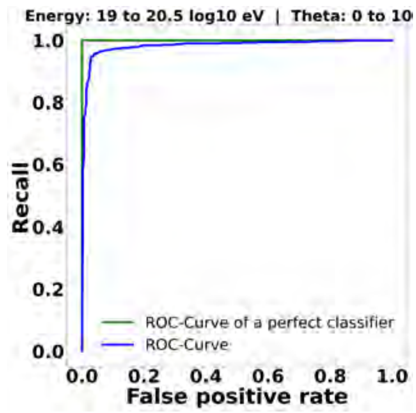


Figure 4.23: An exempla ROC-Curve for one random split of the data for the zenith angle range 0-10 ° (blue) compared to the ROC-Curve of an ideal classifier (green).

best depicted by the receiver operating characteristic curve (ROC-Curve), shown in Fig. 4.23. For a not perfect classifier, an increasing sensitivity results in a increasing false positive rate.

In the following, the photon sensitivity will be displayed when calculating the proton background . One should keep in mind that this is done by increasing the probability threshold. The second aspect concerns the properties of incorrectly classified protons. One has to re-weight them for energy and $k_{\Sigma\pi^0}$ and normalize their weights to calculate the irreducible background. Figure 4.24 shows both aspects. Each plot shows one of the 25 random configurations used to train the model for the different zenith angle ranges.

The observation fits the expectations. Over the whole zenith angle range, only a few events with $k_{\Sigma\pi^0} < 0.7$ have photon probabilities > 0.4. A sharp increase of events with higher photon probabilities is observed for $k_{\Sigma\pi^0} > 0.75$. This is especially true for events with rather low energy that have a high photon probability (different weights at the same $k_{\Sigma\pi^0}$ result from different energies, e.g. a larger energy means a lower weight). This is expected to be an effect of the Lorentz boosting as described before. While the low energy events develop photon-like showers, high energy air showers tend to have re-interacting π^0s and thus are hadronic in character. For higher zenith angles, more events are obtained with a high photon probability. Here, one should keep in mind that

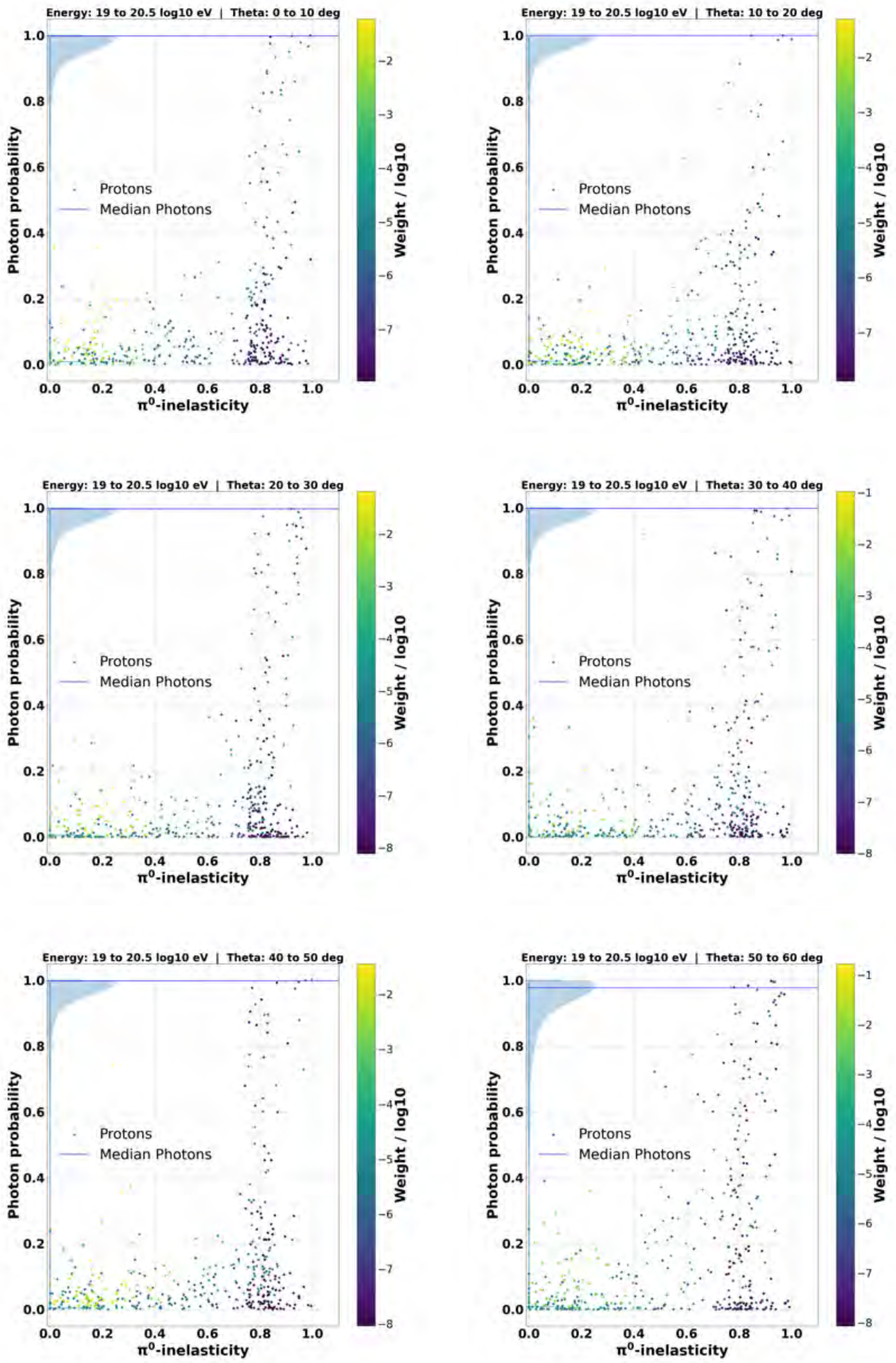


Figure 4.24: The photon probabilities of the proton events dependent on the $k_{\Sigma\pi^0}$ are shown for different zenith angles. A high $k_{\Sigma\pi^0}$ enables a high photon probability. The simulated photons are visualized by the violin plot.

the effects of Lorentz boosting combine with the lower separation power at the highest zenith angles.

The photon distribution is fairly easy to describe. As seen before, more than 94 % of the events have probabilities larger than 0.5. For zenith angles $< 50\,^\circ$ the median of the photon distribution is mostly equal to 1, only few random model configurations have medians < 1. In both cases the majority of photon events have probabilities $>$ 0.8. For the highest zenith angles, however, the median is never equal to 1 and more like 0.97. Additionally, the distribution spreads over a broader range with many photon probabilities between 0.6 and 0.8. It is interesting to see, but not relevant for the irreducible background, are the photons with photon probability ≈ 0. This can be observed for all zenith angle bins. These low probabilities are most likely caused by the photons which have undergone photonuclear interactions.

Finally, the simulated proton events are re-weighted for energy and $k_{\Sigma\pi^0}$ and used to calculate the photon sensitivity dependent irreducible proton background. This is shown in Fig. 4.25. The value is interpreted as probability measuring a photon-like event in case of a proton detection with $E > 10^{19}\,\text{eV}$.

The observed background again looks similar for zenith angles $< 50\,^\circ$ and stand out for the highest zenith angles. For the lower zenith angles a minimal sensitivity (selection on events with photon probabilities = 1) between 0.3 and 0.5 is observed. At a sensitivity = 0.5, an irreducible proton background of $\approx 10^{-5}$ is achieved. For the zenith ranges of 0-10 ° and 20-30 ° a rather slow and steady increase up to a sensitivity of ≈ 0.95 is observed, with a steep increase afterwards. At this sensitivity, the cut on the photon probability is below 0.4 and includes the bulk of protons with high weights. The development is similar for the zenith ranges of 10-20 ° and 30-50 °. However, the curves do not look as smooth as with respect to the previously mentioned zenith angle bins and jumps can be seen. These jumps originate from single protons with high weights, low $k_{\Sigma\pi^0}$ but high photon probability. One of these events was already depicted in Fig. 4.24 [1].

The cause for these events is not completely clear. It is expected that it results from events with a highly elastic proton in the first interaction that produces high energy π^0 in the second interaction. This would not be detected by the analysis, but could also lead to a photon-like event. It is expected that the weight of this type of events is overestimated and would decrease with more statistic. Simulation bugs and false reconstructions can also serve as another explanation.

[1] zenith range 30-40 °, $k_{\Sigma\pi^0} \approx 0.425$, photon probability ≈ 0.9

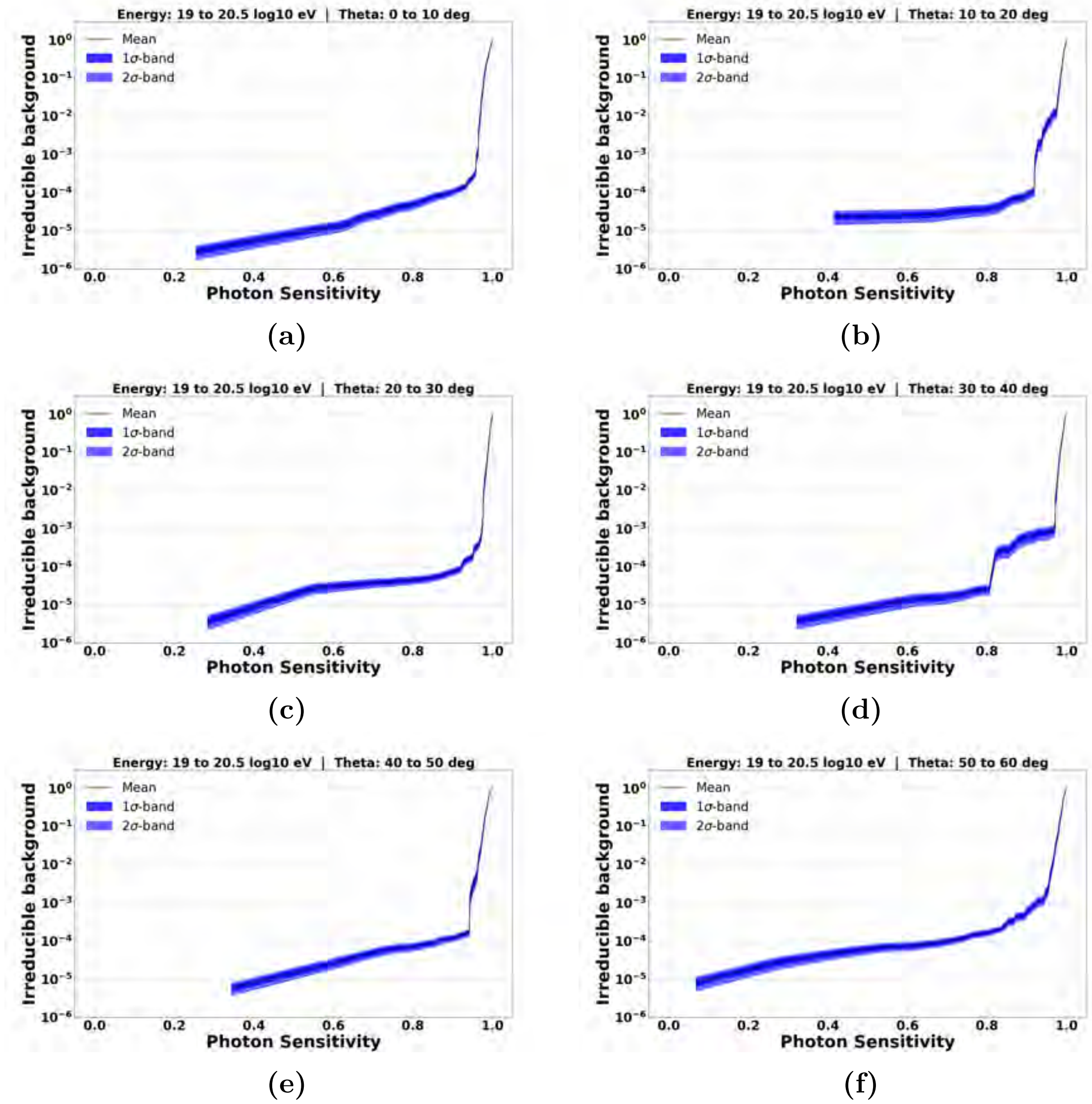


Figure 4.25: The re-weighted irreducible proton background vs the photon sensitivity is shown for different zenith angles ((**a**)-(**f**)). Higher photon sensitivities lead to an increased proton background.

Besides missing smoothness, the developments of all zenith angles below 50° are analogous and their values are comparable. This is in contrast to the curve for the highest zenith angles. The minimum sensitivity is ≈ 0.05. The irreducible background is about 10^{-5}, which is comparable to the value of the lower zenith angles at a sensitivity of 0.5. The curve increases smoothly to a value of nearly 10^{-4} at a sensitivity of 0.5 and then takes a similar development compared to other zenith angles with a steep increase around 0.95.

4.5 Conclusion on the irreducible background

The results are summarized to determine the irreducible proton induced background, shown in Fig. 4.26. The blue curve depicts the spectrum from [20] measured at the Pierre Auger Observatory. The other curves are calculated based on this spectrum. Here, it is important to be reminded that a pure proton spectrum has been assumed, even though recent measurements indicate a tendency to heavier nuclei for high energies. This would lead to a substantially lower flux of these photon-like events, meaning these results are conservative.

The purple and red areas show the 1σ-band of the integrated flux of the high $k_{\Sigma\pi^0}$ protons, whereby the red band takes the boosting effects into account. Therefore, the curves are similar for low energies, but start differing more and more for higher energies. The green and yellow curves indicate the irreducible background, respectively, for a sensitivity of 50 and 80 %. While the flux is 1 to 2 orders of magnitude lower for less energetic events, the curves approach the red band at the highest energies. The arrows mark the most recent limits of integrated photon flux[8][13][27]. It can be observed that even the lowest limits from [8] are orders of magnitudes lower than the fluxes of high $k_{\Sigma\pi^0}$ events. Hence, the results are consistent with current flux limits.

This should be even more clear by remembering that just the fluxes of events with $k_{\Sigma\pi^0} > 0.8$ are compared. These events could still be labeled as protons from the detecting experiment which would lead to the lower fluxes shown by the green and yellow curves in Fig. 4.26. For the ideal detector, the fraction of neglected photon candidates would be reduced by more than a factor of 10. The tested detector indicates that only protons with $k_{\Sigma\pi^0} > 0.8$ can lead to photon-probabilities near 1. In the following chapter, we will investigate whether this hypothesis holds true for measurements at the Pierre Auger Observatory.

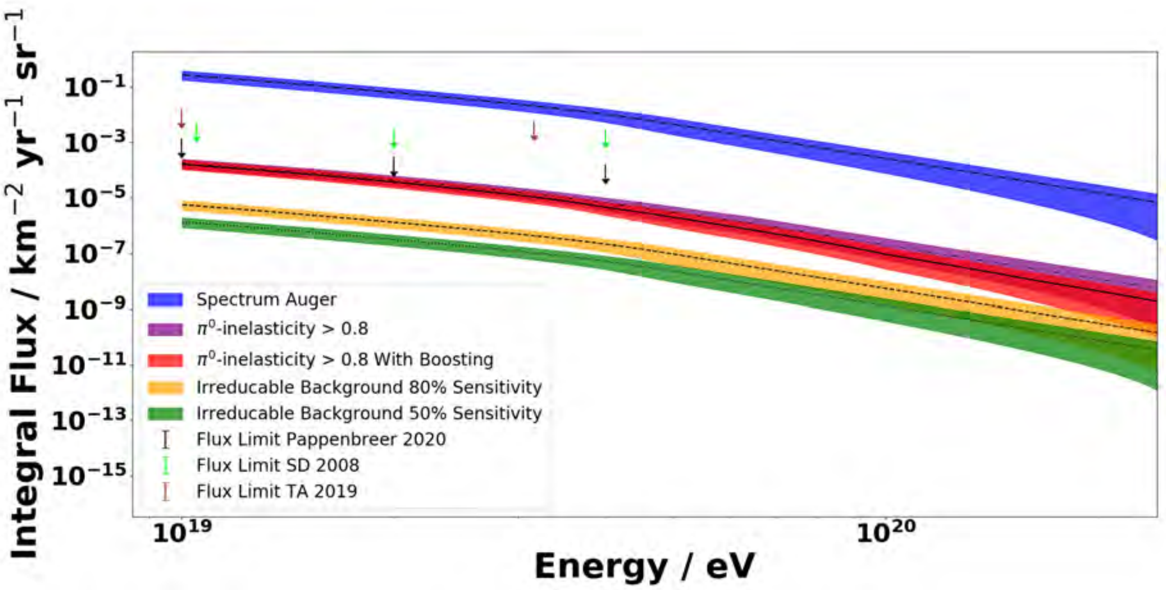


Figure 4.26: The re-weighted integral fluxes of high $k_{\Sigma\pi^0}$ events (purple and boosting corrected red) and irreducible proton backgrounds (green: 50 % photon sensitivity, yellow: 80 % photon sensitivity) based on the spectrum measured by the Pierre Auger Observatory together with the flux limits from [8][13][27] are shown. The limit of [13] at 10 EeV is shifted slightly to higher energies for a clearer view. The plot shows a consistency of the photon limits and irreducible proton background.

5 Analysis of photon-like protons with the Pierre Auger software framework

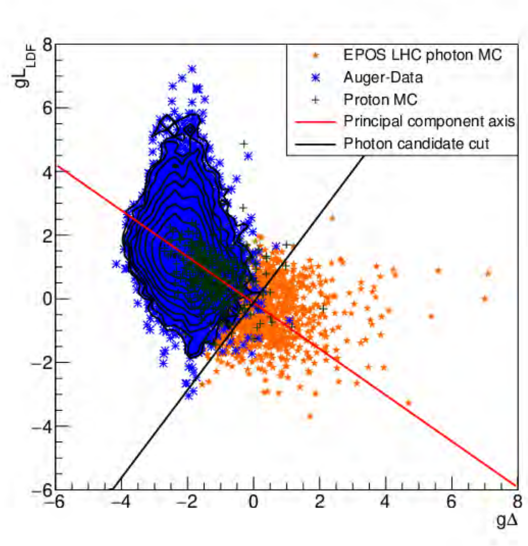
For this analysis, Offline, the software framework[28] of the Pierre Auger Collaboration is utilized. The software simulates the station signals of the Pierre Auger Observatory for CORSIKA showers. For this thesis every shower is thrown five times at different positions with distinct random seeds for the unthinning algorithm. The results are analysed with the method from [8]. Here, modified versions of the R_{NKG} and risetime parameters are used for the proton-photon-discrimination:

$$\begin{aligned}
\Delta_{\mathrm{Leeds}} &= \frac{1}{N}\sum_{i=1}^{N}\frac{t^{i}_{1/2}-t^{\mathrm{bench}}_{1/2}}{\sigma^{i}_{t_{1/2}}}, \\
\mathrm{g}\Delta_{\mathrm{Leeds}} &= \frac{\Delta_{\mathrm{Leeds}}-\langle\Delta^{\gamma}_{\mathrm{Leeds}}\rangle}{\sigma^{\gamma}_{\mathrm{Leeds}}}, \\
\mathrm{gL}_{\mathrm{R_{NKG}}} &= \frac{\mathrm{L}_{\mathrm{R_{NKG}}}-\langle\mathrm{L}^{\gamma}_{\mathrm{R_{NKG}}}\rangle}{\sigma^{\gamma}_{\mathrm{R_{NKG}}}},
\end{aligned}$$

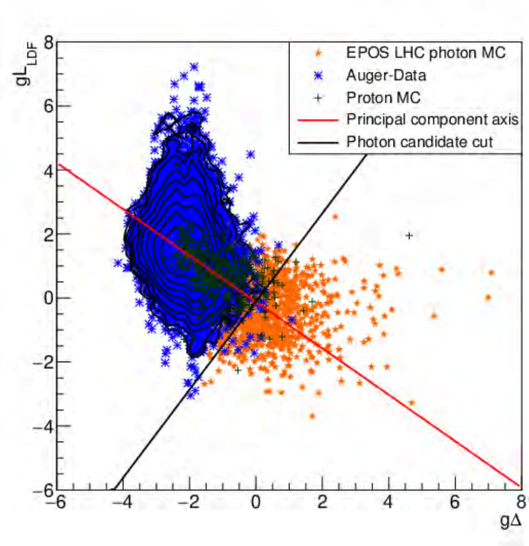
where the σ^i means the standard deviation of the single detector stations. The modified parameters $\mathrm{gL}_{\mathrm{LDF}}$ and $\mathrm{g}\Delta_{\mathrm{Leeds}}$ are then used in a principle component analysis (PCA). A cut at the median PCA value of the photon distribution yields a good discrimination.

For the analysis, 1896 high $k_{\Sigma\pi^0}$ simulated proton air showers with $k_{\Sigma\pi^0} > 0.6$ were selected, resulting in ca. 10k event reconstructions. The zenith angle and energy is limited by the analysis method to $\theta_{\mathrm{MC}} \in [30\,^\circ, 60\,^\circ]$ and $E_{\mathrm{MC}} > 10\,\mathrm{EeV}$. It is important to mention, that the reconstructed zenith angle and energy are used instead of the Monte Carlo values. Hence, events with θ_{true} between 28 and 30 °, and 60 and 62 ° can survive the cuts due to the up to 2 ° zenith angle reconstruction uncertainty at the Pierre Auger Observatory. Hence, additional events with $\theta_{\mathrm{true}} \in [60\,^\circ, 62^\circ]$ were simulated. The PCA results are shown in Fig. 5.1, where the data of the Pierre Auger Observatory, the proton simulations and in [8] simulated photons showers are compared.

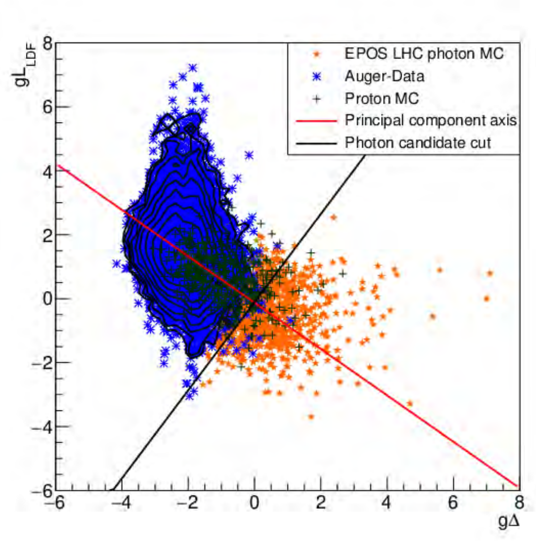
The simulated showers are within expectations. A smearing from the center of the data distribution to the PCA cut can be observed with few events larger than the cut value. These events are classified as photon candidates.It is evident that almost all simulated proton showers are more photon-like than the mean of the data. Indeed, more events with low PCA values would have been expected, which are Lorentz boosted (see again Chap. 3.3) and look like a regular proton shower. The absence of these events is likely caused by the muon deficit of the simulations discussed in chapter 2.8. This



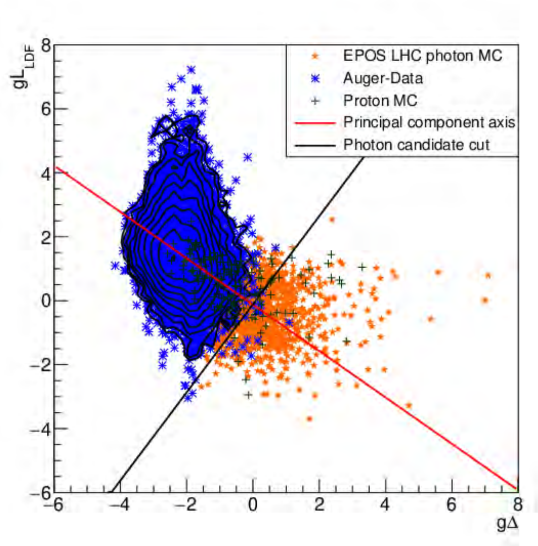
(a) $0.6 < k_{\Sigma\pi^0} < 0.7$



(b) $0.7 < k_{\Sigma\pi^0} < 0.8$



(c) $0.8 < k_{\Sigma\pi^0} < 0.9$



(d) $0.9 < k_{\Sigma\pi^0} < 1$

Figure 5.1: The analysis from [8] applied to the photon simulations (orange) and data sample (blue) from [8] is shown. The PCA-axis (red) and the selection threshold on the PCA-variable (black) are depicted. Events right of the threshold are considered as photon candidates. The analysis applied on the on the simulated proton showers results in the green crosses. For every $k_{\Sigma\pi^0}$ range (**(a)**-**(d)**) photon candidates can be observed. Plots provided by Philipp Papenbreer.

results in lower PCA values and shifts the distribution towards the photon distribution. Thus, the number of photon candidates from protons could be overestimated.

Also, even though photon candidates are observed in Fig. 5.1, they seem to differ from the photon candidates of the data distribution. The simulation candidates are located at larger $g\Delta_{\mathrm{Leeds}}$ values, whereas their gL_{LDF} values are not significant different from the data distribution. They are therefore classified as candidates only due to their larger $g\Delta_{\mathrm{Leeds}}$ value. The candidates from the data, however, are differing from the main dataset mainly by their lower gL_{LDF}, whereas their $g\Delta_{\mathrm{Leeds}}$ does not show any anomalies.

The cause of this feature is not certain. Two explanations are reasonable. First, it could be an artifact of the muon deficit in the simulations. This is consistent with larger risetimes, however, the muon deficit is not expected to shift gL_{LDF} to larger values and therefore more proton like values. Second, it can be argued that the data and simulation candidates are not from the same type. This would indicate, that high $k_{\Sigma\pi^0}$ protons are not the origin of the photon candidates from [8].

To verify the second interpretation, a thorough review of the simulated proton induced photon candidates and their occurrence needs to be done. The fraction of events labeled as photon candidates increases with the $k_{\Sigma\pi^0}$ as shown in Tab. 3. The events are re-weighted by their reconstructed energy and $k_{\Sigma\pi^0}$. This leads to the flux of irreducible high $k_{\Sigma\pi^0}$ events and expected photon candidates. First, the flux is explored, which is shown in Fig. 5.2. The plots shows, that the irreducible proton background is consistent with all photon limits. The flux of high $k_{\Sigma\pi^0}$ events is approximately a factor 10 lower than the latest photon limits. The integrated flux $\mathbf{\Phi}$ leads directly to the number of expected photon candidates. The number is calculated by taking the photon efficiency ϵ and exposure $A = 39002.8\,\mathrm{km}^2\,\mathrm{sr}\,\mathrm{yr}$ at the Pierre Auger Observatory into account:

$$N_{\mathrm{cand}} = \mathbf{\Phi} \cdot \epsilon \cdot A$$

Table 3: The number of events after different analysis steps are shown.

$k_{\Sigma\pi^0}$	# survived θ_{MC} and E_{MC} cuts	# survived quality cuts	# photon candidates	Fraction photon candidates / %
0.6 - 0.7	1724	389	14	3.6
0.7 - 0.8	1646	377	26	6.9
0.8 - 0.9	2053	466	58	12.4
0.9 - 1	625	142	37	26.1

This results in $N_{\text{cand}} = 1.2 \pm 0.5$ for $E > 10\,\text{EeV}$, $N_{\text{cand}} = 0.29 \pm 0.11$ for $E > 20\,\text{EeV}$, $N_{\text{cand}} = 0.047 \pm 0.023$ for $E > 40\,\text{EeV}$. Hence, $\approx 11\,\%$ of the 11 observed photon candidates[8] at energies $> 10\,\text{EeV}$ can be explained by the proton induced background, whereas no candidates are expected for higher energies. This is consistent with the observations from[8] for $E > 40\,\text{EeV}$.

It should be highlighted that this is a conservative estimation. The number of hadron induced photon candidates would be further reduced by three factors. First, the mentioned muon deficit for simulations would most probable lead to lower discrimination. Investigating and resolving this effect can only accomplished by an intensive study and is beyond this work. However, an increased muon component in hadronic air showers is expected to result in less photon candidates. Second, the actual flux of CRs does not consist of a pure proton composition. As seen in Chap. 3.2, a composition of heavier nuclei leads to a negligible probability of high $k_{\Sigma\pi^0}$ events. Generous estimations like a 20 % proton fraction at the highest energies would reduce the number of proton induced photon candidates by a factor of 5. Third, the fluxes and amount of candidates are calculated with the probabilities for EPOS-LHC which produces the largest fraction of high $k_{\Sigma\pi^0}$ events. The usage of other models like QGSJetII-04 would reduce the number of photon candidates again by factors up to 20.

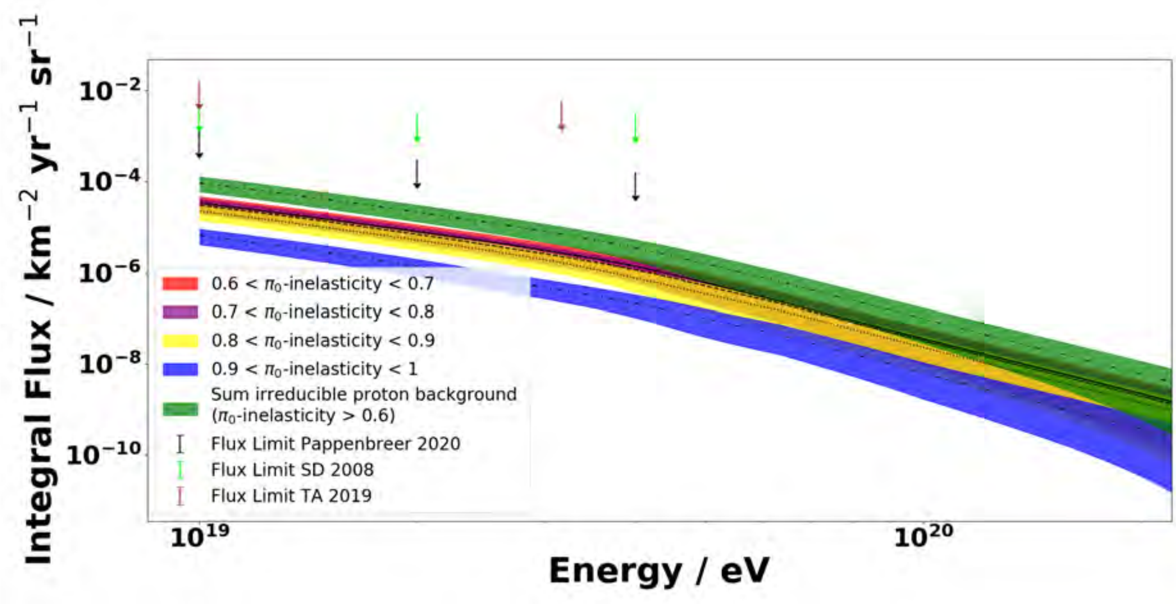


Figure 5.2: The integrated photon flux limits [8][13][27] and expected flux of irreducible protons detected by the Pierre Auger Observatory is shown. The plots shows a consistency of the measurements.

To be clear, the conclusion at this point is **not** that the observed events in [8] have photons as primaries. The interpretation is that the events cannot be explained by high $k_{\Sigma\pi^0}$ events alone. Other possibilities have to be regarded, which can lead to the occurrence of photon candidates. For example in Fig. 5.2, it can be observed that the flux of photon candidates is the highest for the $k_{\Sigma\pi^0}$ between 0.6 and 0.7 due to their more frequent occurrences. It should be discussed whether this trend is continuing to lower $k_{\Sigma\pi^0}$. These photon candidates could occur from reasons other than the low fraction of muonic particles caused by a high $k_{\Sigma\pi^0}$.

6 Conclusion

In this thesis the irreducible proton background at energies $> 10\,\mathrm{EeV}$ caused by high energy π^0 secondaries was studied. A method was developed to efficiently simulate high $k_{\Sigma\pi^0}$ air showers with CORSIKA. Here, two hadronic interaction models were taken into account, EPOS-LHC and QGSJetII-04. From the analysis of the first interaction it was clear that the probability of high $k_{\Sigma\pi^0}$ air showers is significant higher for EPOS-LHC, which was the model further used in the analysis.

It was found that the probability for photon-like events depends not only on the produced secondaries, but strongly varys with energy. For the highest observed energies, Lorentz boosting effects must be considered. Lorentz boosting significantly increases the proper time of π^0s, which then increases probability of their hadronic interaction instead of decay. Thus, the number of photon-like events originating from proton primaries sharply decreases at the highest energies. It was also shown that heavier nuclei as primaries like helium have **no** significant contribution to the number of photon-like air showers.

To analyze the simulations, a generic and idealized ground detector was used. To discriminate between distinct primaries, six detector specific discrimination parameters were introduced. R_{NKG} and γ were found to have strong discrimination powers for zenith angles $< 50\,°$, whereas the station risetime dominates the discrimination for zenith angles from 50-60 °. For a good proton-photon-separation, all parameters should be used via a multivariate analysis. A random forest machine learning algorithm was used to combine all parameters and take their correlations into account. It was found that for this idealized detector, only proton events with $k_{\Sigma\pi^0} > 0.8$ are indistinguishable from photon air showers. These events were used to calculate the irreducible proton background and integrated flux for photon sensitivities of 50 and 80 %. The results were found to be consistent with recent photon flux limits.

Finally, the analysis from [8] was adapted and applied on the simulated proton events. It was shown that the resulting integrated proton background flux is again consistent with the integrated photon flux limits. Additionally, the number of expected photon candidates was calculated for the Pierre Auger Observatory. Here, only 11 % of the photon candidates from [8] at energies $> 10\,\mathrm{EeV}$ can be explained by the proton background, while no proton induced photon candidates are expected for the energies $> 20\,\mathrm{EeV}$.

It was also clarified that for various reasons the background estimations are quite conservative. It is expected that the proton induced background is likely even lower. Hence it is concluded, the irreducible proton background **cannot** serve as explanation for the identified photon candidates.

References

[1] M. Tanabashi and et al. Review of Particle Physics. *Phys. Rev. D*, 98:030001, Aug 2018.

[2] The Pierre Auger Collaboration. The Pierre Auger Observatory: Contributions to the 36th International Cosmic Ray Conference (ICRC 2019), 2019.

[3] D. Caprioli, P. Blasi, and E. Amato. Non-linear diffusive acceleration of heavy nuclei in supernova remnant shocks. *Astroparticle Physics*, 34(6):447 – 456, 2011.

[4] The Pierre Auger Collaboration. Depth of Maximum of Air-Shower Profiles at the Pierre Auger Observatory: Measurements at Energies above $10^{17.8}$ eV. *Physical Review D*, 90(12), Dec 2014.

[5] The Pierre Auger Collaboration. Mass composition of cosmic rays with energies from $10^{17.2}$ eV to 10^{20} eV using surface and fluorescence detectors of the Pierre Auger Observatory. *EPJ Web of Conferences*, 191:08008, 01 2018.

[6] J. Matthews. A Heitler model of extensive air showers. *Astroparticle Physics*, 22(5):387 – 397, 2005.

[7] Andreas Haungs, Heinigerd Rebel, and Markus Roth. Energy spectrum and mass composition of high-energy cosmic rays. *Reports on Progress in Physics*, 66:1145, 06 2003.

[8] Philipp Papenbreer. *Search for Ultra-High-Energy Photons with the Pierre Auger Observatory*. PhD thesis, Bergische Universitaet Wuppertal, 2020.

[9] The Pierre Auger Collaboration. The Pierre Auger Observatory Upgrade - Preliminary Design Report, 2016.

[10] The Pierre Auger Collaboration. The Pierre Auger Cosmic Ray Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 798:172–213, Oct 2015.

[11] The Pierre Auger Collaboration. Spectral Calibration of the Fluorescence Telescopes of the Pierre Auger Observatory. *Astroparticle Physics*, 95, 09 2017.

[12] Nicole Krohm. *Search for Ultra-High Energy Photons with the Surface Detector of the Pierre Auger Observatory*. PhD thesis, Bergische Universitaet Wuppertal, 2017.

[13] J. Abraham. Upper limit on the cosmic-ray photon flux above 10^{19} eV using the surface detector of the Pierre Auger Observatory. *Astroparticle Physics*, 29(4):243–256, May 2008.

[14] The Pierre Auger Collaboration. Direct measurement of the muonic content of extensive air showers between 2×10^{17} and 2×10^{18} eV at the Pierre Auger Observatory. *The European Physical Journal*, (80), Aug 2020.

[15] Lu Lu. *A search for photons of energy above 6×10^{18} eV using data from the Water-Cherenkov detectors of the Pierre Auger Observatory.* PhD thesis, Leeds U., 2014.

[16] Lorenzo Cazon, Ruben Conceição, Miguel Alexandre Martins, and Felix Riehn. Constraining the energy spectrum of neutral pions in ultra-high-energy proton-air interactions. *Physical Review D*, 103(2), Jan 2021.

[17] Corinne Berat, Francois Montanet, and Mariaangela Settimo. Study of proton showers with leading π^0. *Auger internal note, GAP2016-052*, 2016.

[18] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw. *CORSIKA: a Monte Carlo code to simulate extensive air showers.* 1998.

[19] The Pierre Auger Collaboration. Testing Hadronic Interactions at Ultrahigh Energies with Air Showers Measured by the Pierre Auger Observatory. *Physical Review Letters*, 117(19), Oct 2016.

[20] The Pierre Auger Collaboration. Features of the Energy Spectrum of Cosmic Rays above 2.5 $\times 10^{18}$ eV Using the Pierre Auger Observatory. *Physical Review Letters*, 125(12), Sep 2020.

[21] David d'Enterria and et al. Impact of QCD jets and heavy-quark production in cosmic-ray proton atmospheric showers up to 10^{20} eV. *The Astrophysical Journal*, 874, 2019.

[22] Benno List. Statistical error of efficiency determination from weighted events. `www.desy.de/~blist/notes/effic.ps.gz`, 1999. updated 2003.

[23] Pierre Billoir. A sampling procedure to regenerate particles in a ground detector from a "thinned" air shower simulation output. *Astroparticle Physics*, 30(5):270 – 285, 2008.

[24] The Pierre Auger Collaboration. Estimating the Depth of Shower Maximum using the Surface Detectors of the Pierre Auger Observatory. *PoS*, ICRC2019:440, 2019.

[25] The Pierre Auger Collaboration. Measurement of the Cosmic Ray Flux near the Second Knee with the Pierre Auger Observatory. *PoS*, ICRC2019:225, 2019.

[26] Leo Breiman. Random Forests. *Machine Learning*, 45:5 – 32, 2001.

[27] Telescope Array Collaboration. Upper limit on the flux of photons with energies above 10^{19} eV using the Telescope Array surface detector. *Physical Review D*, 88(11), Dec 2013.

[28] J. Allen and et al. The Pierre Auger Observatory offline software. *J. Phys.: Conf. Ser.*, 119, 2008.

Danksagungen

(German)
An dieser Stelle ist es Zeit Danke zu sagen, allen Leuten die mich während meines Studiums begleitet haben. Solltet ihr euch hier nicht namentlich genannt finden, seid euch trotzdem meiner Dankbarkeit gewiss.

An erster Stelle möchte ich Herrn Professor Kampert danken. Dies bezieht sich mit Sicherheit auf die Betreuung und Korrektur dieser Thesis, aber auch, dass ich seit mehreren Jahren in verschiedenen Funktionen seiner Arbeitsgruppe angehören darf. Ich erachte es nicht für selbstverständlich, dass ein Professor sich in seinem Maße für die Studenten einsetzt und Perspektiven bietet.

Ebenso möchte ich Herrn Professor Zeitnitz danken, dass er einmal mehr das Zweitgutachten dieser Thesis übernommen hat.

Drei Personen möchte ich ganz besonders in Verbindung mit meiner Thesis danken.
Als erstes Dr. Julian Rautenberg als meinen Betreuer und Korrekturleser, welcher stets sehr hilfreichen Input geben konnte. Bei jeder Frage konnte ich auf ihn zugehen und auch lange Diskussionen führen. Ich freue mich auf die weitere Zusammenarbeit.
An zweiter Stelle ist Dr. Philipp Pappenbreer zu nennen. Er schrieb die Dissertation, auf der diese Folgearbeit beruht. Nicht nur während der finalen Phase seiner Arbeit, auch nach Abgabe konnte ich mit jeder Frage bezüglich seiner Thesis auf ihn zukommen. Auch danke ich ihm für die Bereitstellung seines Analysecodes und der Erstellung der finalen Plots.
Ebenso möchte ich Monsieur Professor Montanet danken. Ich durfte während meines Auslandssemesters bei ihm mein Projektpraktikum machen. Darüber hinaus blieb die Zusammenarbeit im Zuge dieser Thesis bestehen, er erstellte auch den Detektor aus der Analyse. Natürlich gilt auch mein Dank den anderen Grenobler, namentlich Corinne Berat und Carla Bleve, für ihren Input.

Dr. Eric Mayotte und Alex Kääpä möchte ich hervorheben, da beide eingewilligt haben meine Thesis sprachlich zu korrigieren, was bei meinen sprachlichen (Un)Fähigkeiten wahrscheinlich sehr zeitaufwändig war.

Meinen (ehemaligen) Bürokollegen Simon, Alina, Urs und insbesondere Sonja ist zu danken, dass sie meine noch so doofen Fragen beantworteten, aber auch stets für Abwechslung gesorgt haben. Auch wenn die gemeinsame Zeit im vergangenen Jahr nur sehr beschränkt war, hat es trotzdem Spaß gemacht.
Auch die weiteren Gruppenmitglieder waren immer hilfsbereit und unterhaltsam, daher danke an Karl-Heinz, Fabian, Iona, Karol, Enrico, Jörg, Pia, Marvin, Tobias, Frederik, Wilson, Uwe, Vivek, Christian, Dennis, Sarah, Ievgenii, Anna, Tetiana, Rukje, Ruth, Sven, Shivani, Michael, Srijan, Timo, Pavish. Natürlich auch danke an Frau Starke für die Unterstützung bei bürokratischen Dingen und das Hinüberwegsehen, wenn ich meine Kaffeerechnung mal wieder zu spät gezahlt habe.

Auch wenn sie nichts mit der Physik zu tun haben, wäre ohne zwei Personen diese Thesis nicht in dieser Form möglich gewesen. Danke an meine Eltern Sigrid und Rolf auf die ich in jeglicher Hinsicht seit dem Beginn meines Studiums zählen konnte und mich in der ein oder anderen misslichen Lage unterstützt haben.
Meinen Schwestern Julia und Marisa ist für ihren ununterbrochenen Support zu danken, auch wenn ich das Studium manchmal mehr als die beiden im Kopf hatte.

Als letztes möchte ich noch meinen Freunden für ausreichend Ablenkung vom Stress danken. Auch wenn ich einige Namen nennen müsste, möchte ich mich insbesondere bei Christian und Marcel bedanken, für zahlreiche Kaffees, Gespräche und Sonstiges.

Statutory declaration

I certify that I have written the thesis independently myself, using only the sources and aids documented therein, and that I have indicated all quotations as such.

February 09, 2021 **Jannis Pawlowsky**

A Discrimination power of $X_{\mathbf{max}}$

Table 4: Proton background for a X_{max} separation.

Cut from median	Zenith angle range / °
Energy / log10(eV)	Value / 10^{-2}

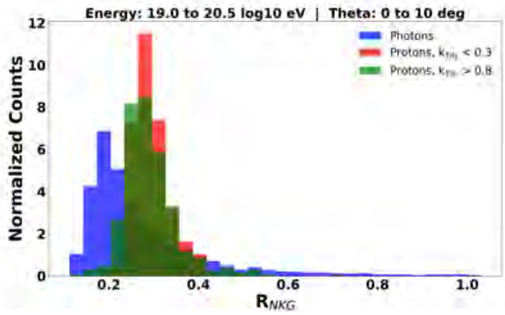
$+1\sigma$	30-40	40-50	50-60
18.8-19.2	2.9 ± 1.6	1.3 ± 0.9	1.2 ± 1.1
19.2-19.6	0.7 ± 0.6	0.09 ± 0.04	0.02 ± 0.01
19.6-20.0	0.4 ± 0.2	0.07 ± 0.04	0.025 ± 0.23
20.0-20.5	2.1 ± 1.0	5.8 ± 1.8	3.1 ± 1.2

$\pm1\sigma$	30-40	40-50	50-60
18.8-19.2	4.0 ± 2.5	7.4 ± 3.1	6.5 ± 3.4
19.2-19.6	7.3 ± 2.8	3.2 ± 1.6	2.6 ± 1.3
19.6-20.0	1.9 ± 1.2	11.4 ± 7.2	5.8 ± 4.0
20.0-20.5	11.2 ± 3.2	18.5 ± 4.4	31.4 ± 4.7

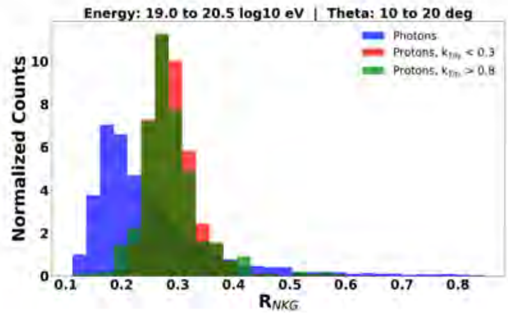
$\pm2\sigma$	30-40	40-50	50-60
18.8-19.2	10.7 ± 5.3	21.7 ± 7.1	18.6 ± 7.6
19.2-19.6	11.7 ± 4.4	10.4 ± 3.9	11.0 ± 3.6
19.6-20.0	49 ± 16	41 ± 17	45 ± 13
20.0-20.5	28.9 ± 7.4	43.3 ± 9.1	80.4 ± 9.4

$\pm3\sigma$	30-40	40-50	50-60
18.8-19.2	26.3 ± 8.6	42.6 ± 10.7	48.8 ± 12.3
19.2-19.6	22.6 ± 7.0	28.5 ± 7.4	35.1 ± 7.0
19.6-20.0	99 ± 30	87 ± 29	100 ± 23
20.0-20.5	46 ± 10	68 ± 13	100 ± 13

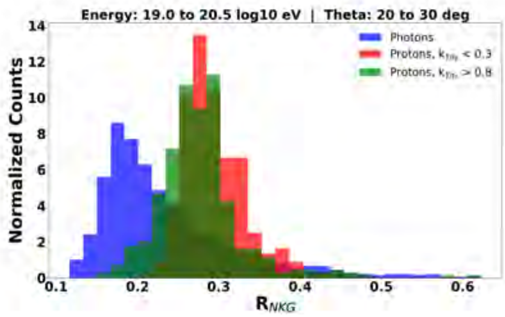
B Discrimination parameter of the SSD



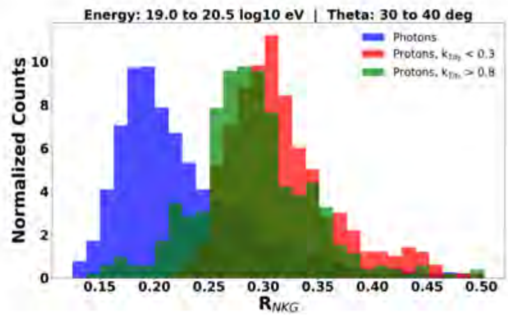
(a)



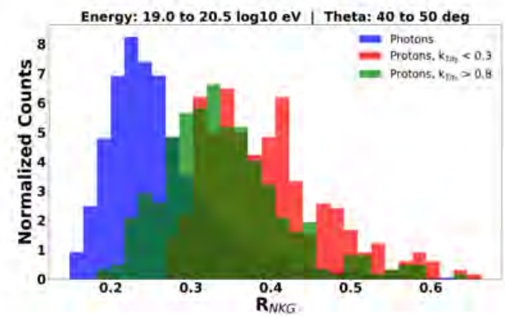
(b)



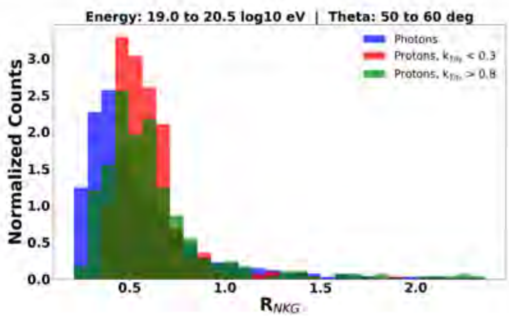
(c)



(d)

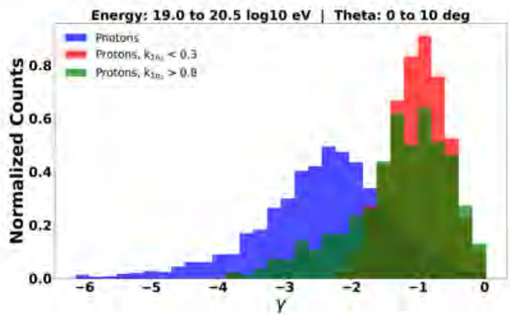


(e)

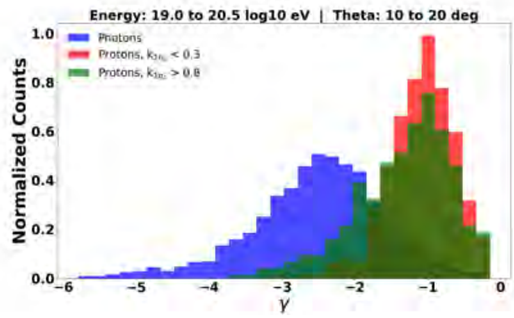


(f)

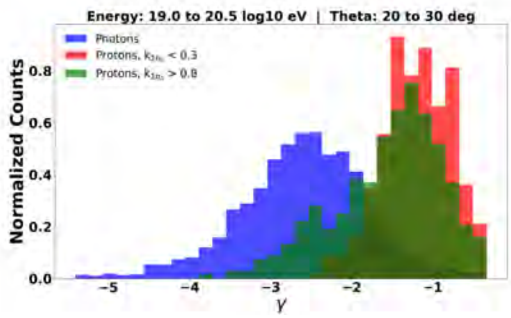
The distribution of the R_{NKG} parameter is shown for photons (blue), low (red) and high (green) $k_{\Sigma\pi^0}$ proton events. With increasing zenith angles a worse separation can be observed.



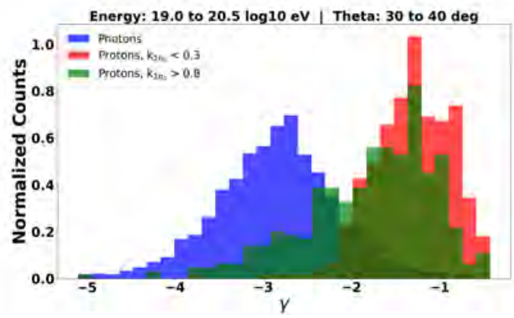
(a)



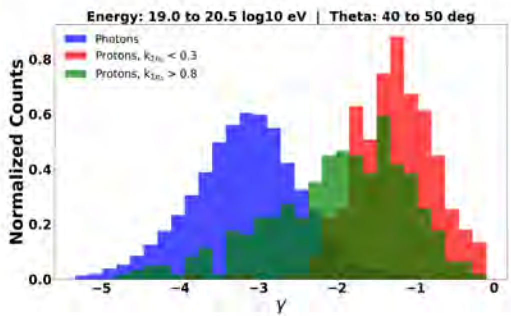
(b)



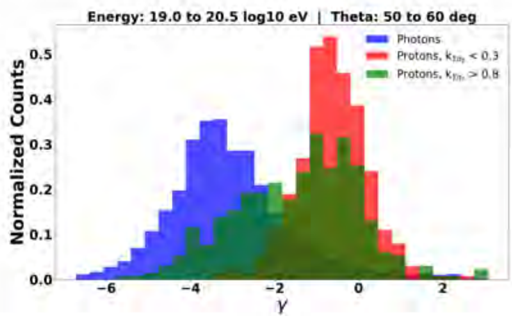
(c)



(d)

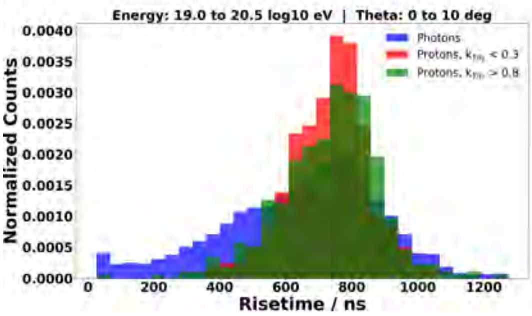


(e)

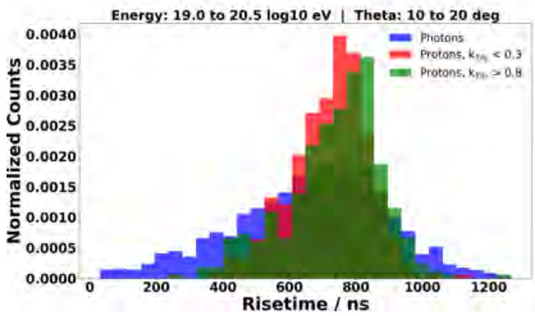


(f)

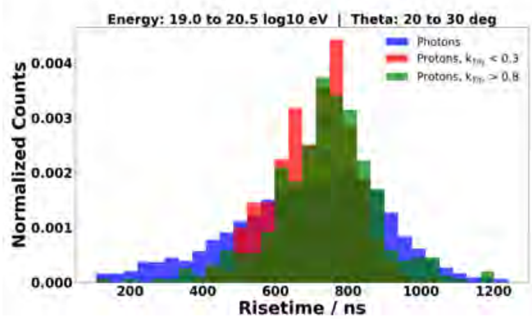
The distribution of the γ parameter is shown for photons (blue), low (red) and high (green) $k_{\Sigma\pi^0}$ proton events. With increasing zenith angles a worse separation can be observed.



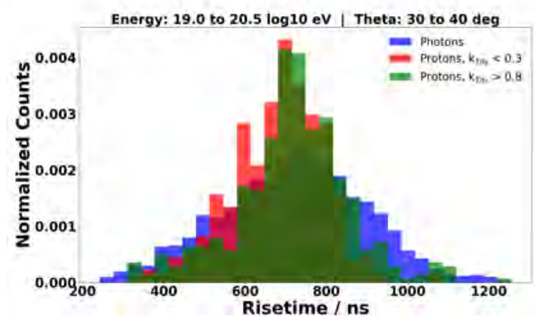
(a)



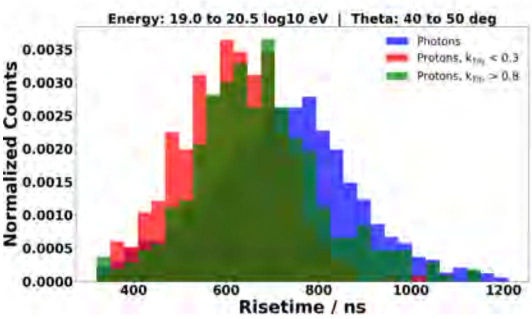
(b)



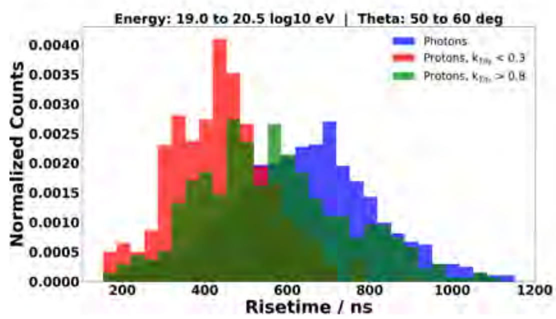
(c)



(d)

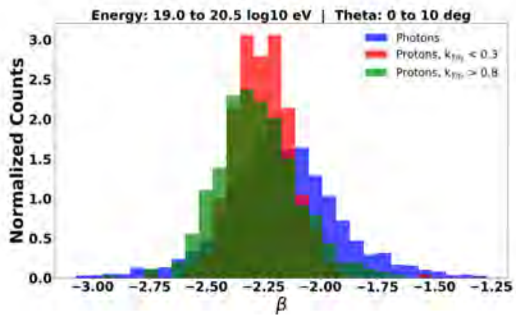


(e)

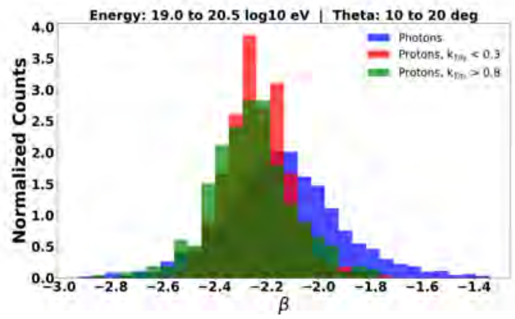


(f)

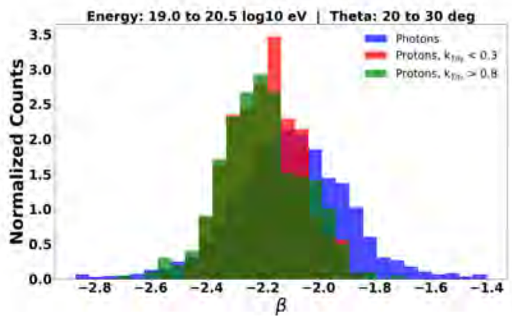
The distribution of the risetime parameter is shown for photons (blue), low (red) and high (green) $k_{\Sigma\pi^0}$ proton events. With increasing zenith angles a better separation can be observed.



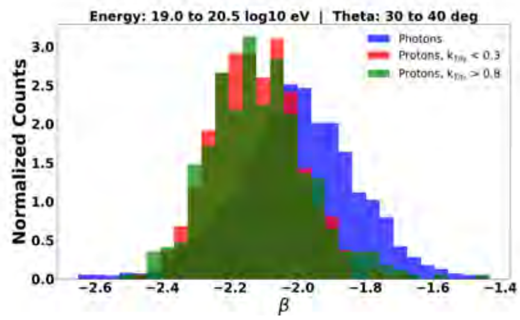
(a)



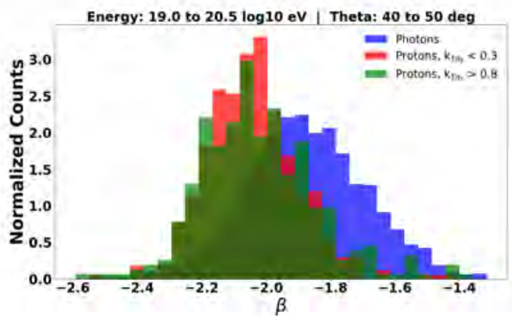
(b)



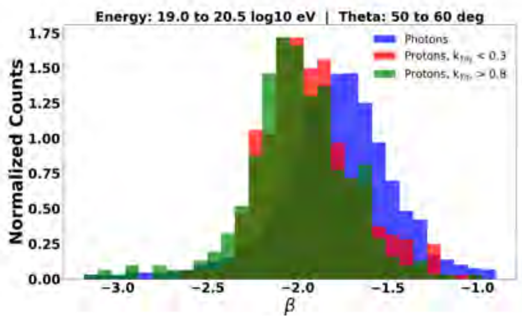
(c)



(d)

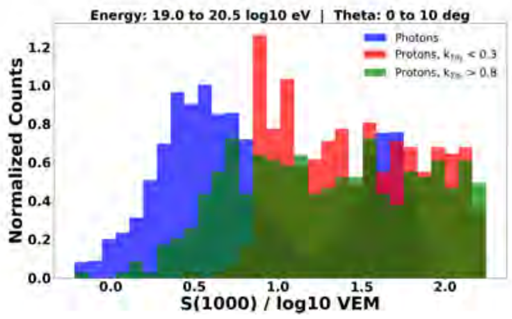


(e)

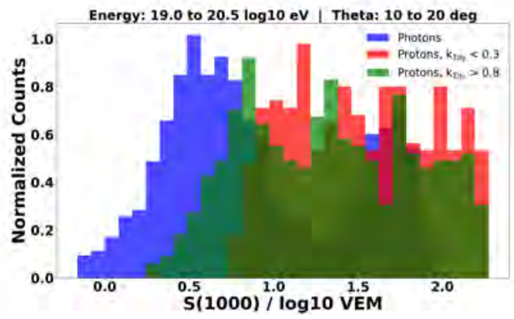


(f)

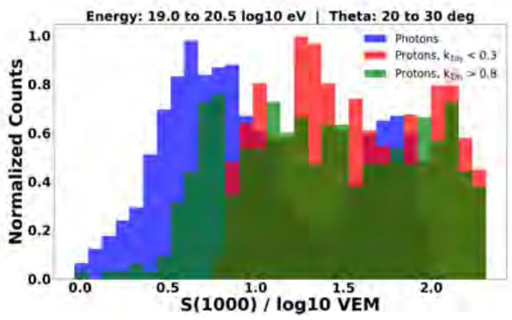
The distribution of the β parameter is shown for photons (blue), low (red) and high (green) $k_{\Sigma\pi^0}$ proton events. With increasing zenith angles a better separation can be observed.



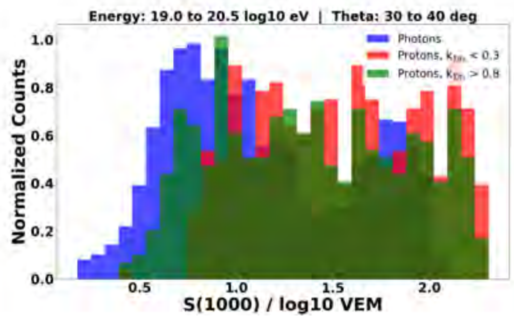
(a)



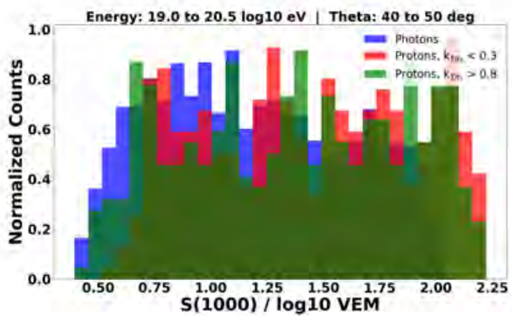
(b)



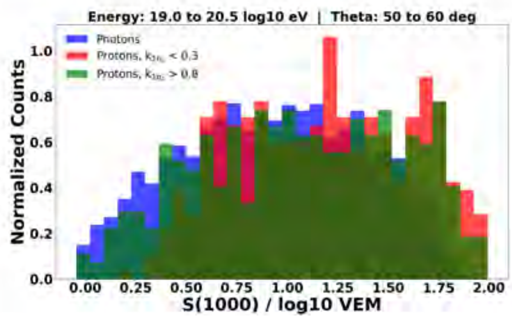
(c)



(d)

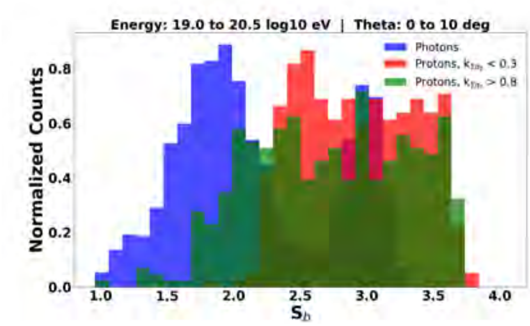


(e)

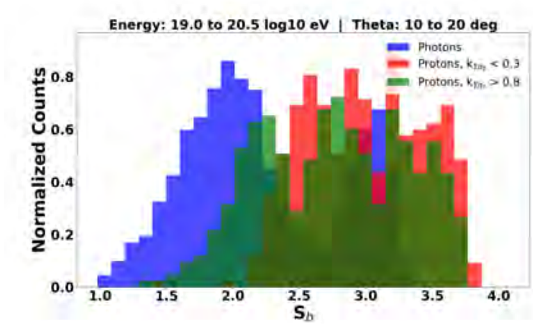


(f)

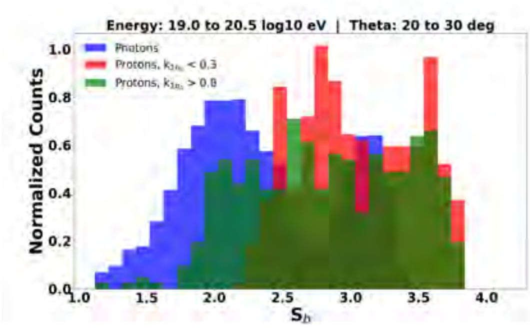
The distribution of the S(1000) parameter is shown for photons (blue), low (red) and high (green) $k_{\Sigma\pi^0}$ proton events. Remarkable is the double peak structure of the photon distribution.



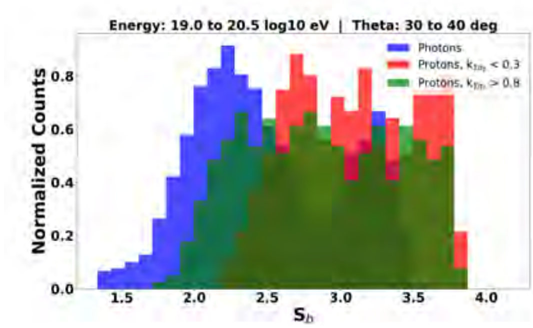
(a)



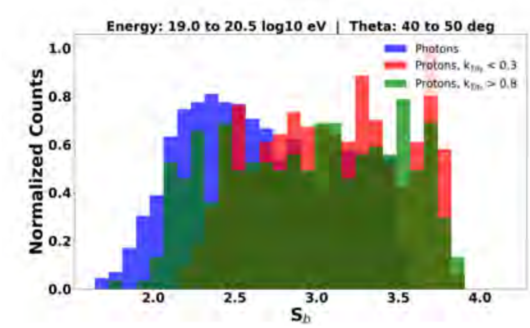
(b)



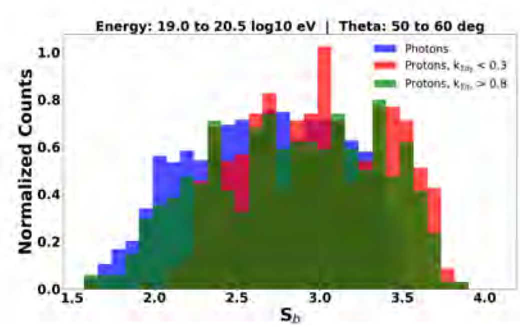
(c)



(d)



(e)



(f)

The distribution of the S_b parameter is shown for photons (blue), low (red) and high (green) $k_{\Sigma\pi^0}$ proton events.