



UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Física Gleb Wataghin

ALLAN MACHADO PAYERAS

**POTENTIAL OF THE AUGERPRIME SURFACE DETECTOR FOR
COMPOSITION STUDIES OF HIGH-ENERGY COSMIC RAYS**

**POTENCIAL DO DETECTOR DE SUPERFÍCIE AUGERPRIME PARA ESTUDOS
DE COMPOSIÇÃO DE RAIOS CÓSMICOS DE ALTÍSSIMAS ENERGIAS**

Campinas

2023

ALLAN MACHADO PAYERAS

**POTENTIAL OF THE AUGERPRIME SURFACE DETECTOR FOR
COMPOSITION STUDIES OF HIGH-ENERGY COSMIC RAYS**

**POTENCIAL DO DETECTOR DE SUPERFÍCIE AUGERPRIME PARA ESTUDOS
DE COMPOSIÇÃO DE RAIOS CÓSMICOS DE ALTÍSSIMAS ENERGIAS**

Tese apresentada ao Instituto de Física Gleb Wataghin da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Ciências, na área de Física.

Thesis presented to the Institute of Physics Gleb Wataghin of the University of Campinas in partial fulfillment of the requirements for the degree of Doctor in Sciences, in the area of Physics.

Supervisor: Prof. Dr. Anderson Campos Fauth

ESTE TRABALHO CORRESPONDE À
VERSÃO FINAL DA TESE DEFENDIDA
PELO ALUNO ALLAN MACHADO
PAYERAS E ORIENTADA PELO PROF.
DR. ANDERSON CAMPOS FAUTH

Campinas

2023

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Física Gleb Wataghin
Maria Graciele Trevisan - CRB 8/7450

M18p Machado Payeras, Allan, 1992-
Potential of the AugerPrime surface detector for composition studies of high-energy cosmic rays / Allan Machado Payeras. – Campinas, SP : [s.n.], 2023.

Orientador: Anderson Campos Fauth.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Física Gleb Wataghin.

1. Raios cósmicos de altíssima energia. 2. Raios cósmicos. 3. Chuveiros de raios cósmicos. I. Fauth, Anderson Campos, 1957-. II. Universidade Estadual de Campinas. Instituto de Física Gleb Wataghin. III. Título.

Informações Complementares

Título em outro idioma: Potencial do detector de superfície AugerPrime para estudos de composição de raios cósmicos de altíssimas energias

Palavras-chave em inglês:

Cosmic rays of very high energy

Cosmic rays

Cosmic ray showers

Área de concentração: Física

Titulação: Doutor em Ciências

Banca examinadora:

Anderson Campos Fauth [Orientador]

Carola Dobrigkeit Chinellato

Sandro Guedes de Oliveira

Rita de Cássia dos Anjos

Helio Nogima

Data de defesa: 15-12-2023

Programa de Pós-Graduação: Física

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0001-7635-4348>

- Currículo Lattes do autor: <http://lattes.cnpq.br/6435143158304362>



INSTITUTO DE FÍSICA
GLEB WATAGHIN

MEMBROS DA COMISSÃO EXAMINADORA DA DISSERTAÇÃO DE DOUTORADO DO ALUNO ALLAN MACHADO PAYERAS - RA 164178 APRESENTADA E APROVADA AO INSTITUTO DE FÍSICA GLEB WATAGHIN, DA UNIVERSIDADE ESTADUAL DE CAMPINAS, EM 15/12/2023.

COMISSÃO JULGADORA:

- Prof. Dr. Anderson Campos Fauth – Presidente e orientador (IFGW/UNICAMP)
- Profa. Dra. Carola Dobrigkeit Chinellato (IFGW/ UNICAMP)
- Prof. Dr. Sandro Guedes de Oliveira (IFGW/ UNICAMP)
- Profa. Dra. Rita de Cássia dos Anjos (Universidade Federal do Paraná)
- Prof. Dr. Helio Nogima (Universidade do Estado do Rio de Janeiro)

OBS.: Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

CAMPINAS

2023

This thesis is dedicated to my lovely wife Paula, who supported me unconditionally during my doctoral work.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisor, Professor Anderson Campos Fauth, for providing me with the unique opportunity to pursue a doctoral degree under his guidance. I appreciate the freedom he granted me to choose and conduct my research, always with his unwavering support and keen interest in my academic and personal growth.

I extend great appreciation to Dr. Darko Veberič, whose guidance paved the way to an incredible experience at the Karlsruhe Institute of Technology, where I acquired invaluable knowledge. Special thanks to Dr. David Schmidt and Dr. Markus Roth for their inspiring interactions, which significantly enriched the quality of my work.

I am grateful to my friends, with whom I shared many amazing moments and who supported me through challenging times. During my stay in Germany, I want to thank Emily, Fabio, Tobias, Marina, Joaquin, Steffen, Alex, Max, Martin, Flavia, Fabian, Paul, Fiona, Varada, Carmina, Christoph, Quentin, Francesco, Damian, Katrin, Felix, Michael, Tanja, Selina, Marc, and Niklas. To my friends in Brazil, I express my gratitude to Elker, Samantha, Vinicius, Manuela, Kauê, Julia, Zé, Marina, Theodora, Pedro, and Aline.

Minha família é a parte mais importante da minha vida, e eu devo minhas conquistas a cada um de vocês: Paula, Antonio, Sandra, Rodrigo, Enilda, Walter, Vênus e Júpiter. Amo todos vocês.

I would like to acknowledge the financial support for the development of this research from the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, Brazil) through processes n° 2019/15899-5, 2021/08164-9, and 2019/10151-2. Additionally, this doctoral work was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil) – Finance Code 001.

RESUMO

O Observatório Pierre Auger tem como objetivo principal investigar a natureza dos raios cósmicos com energias superiores a 1 EeV, conhecidos como raios cósmicos de altíssimas energias (UHECRs). O Observatório emprega duas técnicas experimentais para detectar chuveis atmosféricos extensos (EASs) induzidos por raios cósmicos primários. O detector de superfície (SD) inclui uma rede de 1660 detectores Cherenkov (WCDs), abrangendo uma área de aproximadamente 3.000 km². Além disso, 24 telescópios de fluorescência estão distribuídos ao redor do SD. Desde o início de sua operação em 2004, o Observatório avançou significativamente a compreensão dos UHECRs. Contudo, verificou-se a necessidade de um maior volume de dados sensíveis à composição dos raios cósmicos para energias acima de 40 EeV. Esta informação é a chave para entender as fontes astrofísicas e os processos capazes de produzir partículas tão energéticas. Para coletar mais dados sensíveis à composição, o Observatório está passando por um upgrade de seu detector de superfície, chamado AugerPrime. A principal componente é a instalação de placas cintiladoras (SSD) em cima de cada WCD e suas novas eletrônicas associadas, chamada de upgraded unified board (UUB). Nesta pesquisa de doutorado, analisamos os dados das primeiras estações AugerPrime no campo, com o objetivo de validar o upgrade do Observatório. Comparando os dados das estações equipadas com UUB com os da eletrônica anterior, denominada unified board (UB), verificamos que o sinal das duas eletrônicas é compatível dentro da incerteza esperada, com não linearidade e viés abaixo de 3% em ambos os canais existentes. Em relação à reconstrução de eventos, demonstramos que a determinação da energia fornecida pelas estações AugerPrime apresenta viés dentro de 5% em relação às estações UB. Adicionalmente, verificamos que os ângulos de zênite e azimute dos eventos são determinados sem viés entre as duas eletrônicas. Estes resultados validam o upgrade do Observatório e garantem compatibilidade e uma transição suave entre os dados coletados antes e após a implementação do AugerPrime. Para alcançar os objetivos científicos do upgrade em relação à identificação de composição, é essencial separar as componentes muônica e eletromagnética dos EASs. Na segunda parte de nossa pesquisa, desenvolvemos um método que realiza tal separação utilizando os

sinais dos detectores AugerPrime. O método fornece a fração muônica com viés abaixo de 6%. Em seguida, reconstruímos observáveis que quantificam o conteúdo muônico dos EASs, a saber, o tamanho muônico (S_{38}^{μ}) e a fração muônica (f_{38}^{μ}) do chuvaire no ângulo de referência de 38° . Demonstramos que S_{38}^{μ} é capaz de distinguir composições leves de pesadas em eventos individuais, desconsiderando o viés de composição na estimativa de energia dos eventos. Embora f_{38}^{μ} não forneça esse poder de discriminação para eventos individuais, constatamos que este observável não é afetado pelo viés de composição na energia, portanto, seu primeiro e segundo momento podem ser utilizados para identificar tendências na composição dos primários, especialmente na faixa de energia de interesse.

ABSTRACT

The main objective of the Pierre Auger Observatory is the investigation of the nature of cosmic rays with energies exceeding 1 EeV, known as ultra-high energy cosmic rays (UHECRs). The Observatory employs two experimental techniques to detect extensive air showers (EASs) induced by primary cosmic rays. The surface detector (SD) consists of an array of 1660 water-Cherenkov detectors (WCDs), spanning an area of approximately 3 000 km². Additionally, 24 fluorescence telescopes are distributed around the SD. Since the beginning of its operation in 2004, the Observatory has significantly advanced the understanding of UHECRs. However, the necessity for a larger volume of data sensitive to the composition of cosmic rays for energies above 40 EeV became evident. This information is the key to understanding astrophysical sources and processes capable of producing such energetic particles. To collect more data sensitive to composition, the Observatory is undergoing an upgrade of its surface detector, called AugerPrime. The main component is the installation of scintillator plates (SSD) on top of each WCD and their associated new electronics, called the upgraded unified board (UUB). In this doctoral research, we analysed the data from the first AugerPrime stations in the field, with the aim of validating the Observatory upgrade. Comparing the data from stations equipped with UUB with those from the previous electronics, called unified board (UB), we found that the signal from both electronics is compatible within the expected uncertainty, with non-linearity and bias below 3% in both channels. Regarding event reconstruction, we demonstrated that the energy determination provided by the AugerPrime stations exhibits biases within 5% relative to UB stations. Additionally, we verified that the zenith and azimuth angles of the events are determined without bias between the two electronics. These results validate the Observatory upgrade and ensure compatibility and a smooth transition between the data collected before and after the implementation of AugerPrime. To achieve the scientific goals of the upgrade regarding composition identification, it is essential to separate the muonic and electromagnetic components of EASs. In the second part of our research, we developed a method that accomplishes this separation using the signals from AugerPrime detectors. The method provides the muonic fraction

with a bias below 6%. Next, we reconstructed observables that quantify the muonic content of **EASs**, namely, the muonic size (S_{38}^{μ}) and the muonic fraction (f_{38}^{μ}) of the shower at the reference angle of 38° . We demonstrated that S_{38}^{μ} is able to distinguish light from heavy compositions in individual events, disregarding the composition bias in the energy estimate of the events. Although f_{38}^{μ} does not provide such discriminatory power for individual events, we found that this observable is unaffected by the composition bias in energy; therefore, its first and second moments can be used to identify trends in the composition of the primaries, especially in the energy range of interest.

List of Acronyms

ADC Analog-to-digital Converter.

ADST Auger Data Summary Trees.

CCJDR Centro de Computação John David Rogers.

CDAS Central Data Acquisition System.

CIC Constant Intensity Cut.

CMB Cosmic Microwave Background.

CORSIKA COsmic Ray Simulation for KAscade.

DAQ Data Acquisition.

DC Direct Current.

EAS Extensive Air Shower.

FD Fluorescence Detector.

FNAL Fermi National Accelerator Laboratory.

FPGA Field Programmable Gate Array.

GPS Global Positioning System.

GZK Greisen-Zatsepin-Kuzmin Effect.

HG High Gain.

KIT Karlsruhe Institute of Technology.

LDF Lateral Distribution Function.

LED Light-Emitting Diode.

LFF Lateral Fraction Function.

LG Low Gain.

LPMT Large Photomultiplier Tube.

MIP Minimum Ionising Particle.

NKG Nishimura-Kamata-Greisen Function.

PMT Photomultiplier Tube.

PPA Pre-Production Array.

PVC PolyVinyl Chloride.

RD Radio Detector.

SALLA Short Aperiodic Loaded Loop Antenna.

SD Surface Detector.

SiPM Silicon Photomultiplier.

SPMT Small Photomultiplier Tube.

SSD Scintillator Surface Detector.

UB Unified Board.

UHECR Ultra-High-Energy Cosmic Ray.

UMD Underground Muon Detector.

UUB Upgraded Unified Board.

UV Ultraviolet.

VEM Vertical Equivalent Muon.

W3C World Wide Web Consortium.

WCD Water-Cherenkov Detector.

XML Extensible Markup Language.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 17 |
| 2 | Cosmic rays | 23 |
| 2.1 | The discovery of cosmic rays | 23 |
| 2.2 | Extensive air showers | 24 |
| 2.2.1 | Hadronic and muonic components | 24 |
| 2.2.2 | Electromagnetic showers | 27 |
| 2.2.3 | Characteristics of extensive air showers | 30 |
| 2.3 | Experimental detection of cosmic rays | 33 |
| 2.4 | Fundamental properties of cosmic rays | 35 |
| 3 | The Pierre Auger Observatory | 41 |
| 3.1 | General description of the Observatory | 41 |
| 3.2 | The surface detector array | 45 |
| 3.2.1 | Hardware of the water-Cherenkov detectors | 45 |
| 3.2.2 | Calibration of the photomultiplier tubes | 47 |
| 3.2.3 | Surface detector triggers | 51 |
| 3.3 | Reconstruction of events with the SD | 53 |
| 3.4 | The <u>Offline</u> framework | 60 |
| 4 | AugerPrime: upgrade of the Observatory | 64 |
| 4.1 | Motivation for the upgrade | 65 |
| 4.2 | Description of the AugerPrime detectors | 68 |
| 4.2.1 | Scintillator Surface Detector | 70 |
| 4.2.2 | Small Photomultiplier Tube | 71 |

| | | |
|-----------|---|------------|
| 4.2.3 | Radio detector | 72 |
| 4.2.4 | Underground Muon Detector | 72 |
| 4.2.5 | Upgraded Unified Board | 74 |
| 5 | Calibration of UUB stations and signal compatibility | 76 |
| 5.1 | Calibration of UUB stations | 76 |
| 5.1.1 | New algorithm for fitting calibration histograms | 79 |
| 5.1.2 | Modifications to the <u>Offline</u> framework | 81 |
| 5.1.3 | Performance of the new algorithm | 84 |
| 5.2 | Compatibility of UUB and UB signals | 88 |
| 5.2.1 | Integrated signals | 88 |
| 5.2.2 | Signal timing | 96 |
| 6 | Compatibility of event reconstruction | 101 |
| 6.1 | Selection of events | 101 |
| 6.2 | Shower-size and energy estimation | 106 |
| 6.3 | Shower-geometry reconstruction | 111 |
| 7 | Simulations | 119 |
| 7.1 | Simulation of extensive air showers | 119 |
| 7.2 | Simulation of event reconstruction | 122 |
| 7.3 | Execution of simulations and external libraries | 125 |
| 8 | Reconstruction of signal components within AugerPrime stations | 128 |
| 8.1 | Formulation of the matrix formalism | 129 |
| 8.2 | Behaviour of the reconstruction parameters | 132 |
| 8.3 | Prediction of signal components | 142 |
| 8.4 | Reconstruction of trace components | 150 |
| 9 | Application for event reconstruction and composition studies | 154 |
| 9.1 | Reconstruction of muonic observables | 155 |
| 9.2 | Application for composition separation | 163 |
| 10 | Summary and conclusions | 174 |
| 10.1 | On the validation of AugerPrime data | 176 |

| | |
|--|------------|
| 10.2 On the application of AugerPrime detectors for separation of signal components | 179 |
| 10.3 On the reconstruction of muonic observables and their application for mass separation | 181 |
| Bibliography | 185 |
| Appendix A Model for the reconstruction parameters | 193 |

CHAPTER 1

Introduction

Cosmic rays comprise extraterrestrial particles that permeate the Universe, continuously impacting our planet. Extensive research has been conducted on these particles for over a century, resulting in a comprehensive understanding of many of their properties [1, 2]. However, certain aspects concerning ultra-high-energy cosmic rays (UHECR) remain unclear and continue to present challenges for scientific exploration [2]. These open questions primarily pertain to the origin of UHECR, the astrophysical processes responsible for accelerating these particles to energies above 10^{18} eV and their propagation through the interstellar and intergalactic environments.

The study of cosmic rays employs various experimental techniques, and the choice of methods is strongly dictated by the energy range of the particles under investigation. The flux of cosmic rays exhibits a characteristic power-law decrease with energy [3, 4]. At energy levels below 10^{14} eV, the flux is sufficiently high to enable their direct detection. In these scenarios, detectors are taken to high altitudes in the Earth's atmosphere [5] or even deployed in space [6, 7]. Direct detection techniques are often similar to those found in high-energy physics experiments, providing precise measurements of cosmic ray energy and composition within the lower energy range of the cosmic ray spectrum.

For energies exceeding 10^{15} eV, the low flux of cosmic rays renders direct detection unfeasible. However, when these high-energy cosmic rays penetrate the atmosphere, they undergo successive interactions with atmospheric molecules produ-

cing a cascade of particles known as extensive air showers (**EASs**). Distinct experimental methods are employed to detect and study these **EASs** [8–12], with the collected data used to reconstruct properties of the particle cascade and, consequently, make inferences about the primary cosmic ray, including its energy, arrival direction, and composition. It is essential to note that, in this case, since the detection is indirect, the associated uncertainties are significantly greater compared to experiments involving direct detection.

Amongst the experiments to investigate **UHECRs**, the Pierre Auger Observatory [12] plays a central role. Located in Argentina, it is the largest cosmic ray observatory in the world, spanning an extensive area of approximately 3 000 km². The primary objective of the Observatory is to detect and study cosmic rays with energies surpassing 10¹⁸ eV, constituting the most energetic region of the cosmic ray spectrum.

To accomplish this goal, the Pierre Auger Observatory employs two distinct experimental techniques for **EAS** detection. Although these methods are independent, they are integrated to complement each other, enhancing measurement accuracy. The first technique relies on an array of water-Cherenkov detectors (**WCDs**) spaced at 1500 m intervals from their nearest neighbours, collectively referred to as the Surface Detector (**SD**) [13]. When **EAS** particles reach the observation level, the signals recorded by the **WCDs** reflect the particle densities at their specific locations, facilitating the reconstruction of the lateral profile of the shower. Given that the **WCDs** are individually powered by solar panels, the **SD** operates nearly 100% of the time.

The second experimental method at the Pierre Auger Observatory involves a set of fluorescence telescopes encircling the **SD**, known as the Fluorescence Detector (**FD**) [14]. The **FD** records the emission of fluorescence light by atmospheric nitrogen molecules when excited by **EAS** particles. This technique enables the observation of the longitudinal development of the shower in the atmosphere and provides an energy estimation of the primary cosmic ray that is independent of **EAS** simulations. However, due to the requirement of dark nights for operation, the duty cycle of the **FD** is limited to approximately 15%.

Events that are simultaneously detected by both the **SD** and **FD** are termed hybrid events. These occurrences allow for the cross-calibration of energy estimations from the **SD** using direct measurements from the **FD**. Given that the majority of de-

tected events originate exclusively from the surface detector, the hybrid design of the Observatory significantly enhances the precision and reliability of the data obtained.

Since the beginning of the operations of the Pierre Auger Observatory in 2004, several important findings have emerged from its data, significantly advancing our comprehension of **UHECRs**. Notably, the discovery of the suppression of the cosmic ray flux at energies beyond 4×10^{19} eV has been a significant breakthrough [15]. This suppression serves as a strong indicator of limitations within the particle acceleration mechanisms, coupled with propagation effects. Whilst various **UHECR** models have anticipated this observed phenomenon [16, 17], unravelling the true astrophysical scenario necessitates knowledge about the composition of cosmic rays in this energy range.

Currently, the most precise data concerning the composition of **UHECRs**, specifically beyond energies of 10^{18} eV, is derived from measurements of the **FD** at the Pierre Auger Observatory [18]. These measurements rely on the correlation between the atomic mass of the primary cosmic ray and the corresponding point of maximum shower development within the atmosphere, as recorded by the **FD**. It is important to note that composition data is interpreted utilising **EAS** simulations [19] which employ models that extrapolate the cross section of hadronic interactions as measured by accelerator experiments [20, 21].

The latest findings [22, 23] indicate that **UHECRs** are predominantly composed of lighter elements up to 3×10^{18} eV. Beyond this point, there is a discernible trend toward a heavier composition. Nevertheless, due to the limited duty cycle of the **FD**, the statistical volume of the available data in the suppression region is currently insufficient to determine the composition of **UHECRs** in this energy range definitively. Acquiring this missing information is essential for establishing a comprehensive understanding of the nature and underlying astrophysical mechanisms governing these extremely energetic particles.

To enhance the quality and quantity of data for discerning the composition of **UHECRs**, the Pierre Auger Observatory has initiated a major upgrade of its experimental setup, known as AugerPrime [24]. Central to this upgrade is the integration of Scintillator Surface Detectors (**SSDs**) on top of each water-Cherenkov detector. Additionally, the **WCDs** will be equipped with an extra photomultiplier tube (**PMT**)

featuring a smaller transverse area. To support the new configuration, the electronics of the stations will be modernised, replacing the existing systems with a more advanced and efficient board known as the Upgraded Unified Board (UUB).

The rationale behind the design of AugerPrime relies on the distinct responses of SSDs and WCDs to electromagnetic particles and muons within EASs. The combined signals from these detectors can be leveraged to disentangle these shower components [25]. Of particular interest is the muonic component of EASs, as it exhibits a direct correlation with the composition of the primary particle. Thus, the muonic signals recorded by AugerPrime stations will facilitate the reconstruction of critical observables sensitive to the mass of primary cosmic rays. Furthermore, the small PMT will mitigate the effects of signal saturation, extending substantially the dynamic range of the AugerPrime stations. This upgrade will significantly enhance the Observatory's capabilities to delve deeper into the fundamental properties of UHECRs.

In this doctoral work, our primary objective was to contribute to the upgrade of the Pierre Auger Observatory and assess its potential for physical analyses related to the discrimination of distinct cosmic ray species. These efforts add to the ongoing progression of the understanding of UHECRs.

The final version of the upgraded components, including the scintillator surface detectors, the small photomultiplier tube, and the upgraded unified board, was deployed in the field after a series of tests conducted between 2016 and 2021. This deployment, initiated in 2021, is nearing completion, with the official commencement of data collection by the AugerPrime detectors, marking the Phase II of the Observatory, scheduled for 2024.

Prior to using data from the upgraded array for physical analyses, it is imperative to ensure the proper functioning of these detectors. Consequently, the first part of our work involved the analysis of data from AugerPrime stations in the field during the Observatory's transition phase. Our main aim was to contribute to the establishment of a stable array of upgraded detectors.

The introduction of the new UUB electronics brought changes in signal processing, particularly in detector calibration. As a result, we conducted a comprehensive analysis of signals from the new stations, constructing a framework to ensure their accurate processing. This framework was validated through comparisons with calibrated

signals from detectors equipped with the older electronics, referred to as the Unified Board (UB).

Subsequent to the analysis of signals from AugerPrime detectors, we investigated the procedure of event reconstruction utilising their data. The validation of event reconstruction is of utmost importance since it provides essential information about the primary cosmic rays, including their arrival direction and energy, enabling subsequent physical analyses. For this critical study, we compared the results of reconstructions using AugerPrime stations with those obtained from the former detectors with UB electronics. These comparisons serve to ensure the absence of biases between the two electronic systems, ensuring data continuity and a seamless transition between Phase-I and Phase-II operations of the Pierre Auger Observatory.

Following the validation studies of the AugerPrime upgrade, we have explored the potential of its data for differentiating cosmic ray compositions, particularly within the energy range of the flux suppression. These analyses align closely with the core objective of the Observatory's upgrade, aimed at resolving the composition of cosmic rays in the most energetic segment of their spectrum, consequently leading the way for a definitive comprehension of the astrophysical processes underlying the production of these extreme particles.

As previously mentioned, the crucial element of the upgrade design hinges on the disentanglement of muonic and electromagnetic signals within these detectors. Different methods can be employed to achieve this goal. In this work, we have scrutinised one such method known as the matrix formalism. We derived the formulation and thoroughly assessed its application and accuracy using Monte Carlo simulations of both EASs and their subsequent detection by the AugerPrime stations.

Possessing a method for reconstructing muonic signals, we proceeded to apply it to the reconstruction of observables that reflect the muonic content of EASs. The definition and evaluation of these reconstructed observables are essential for composition studies, given their potential correlation. These investigations were also conducted using Monte Carlo simulations, considering that the official collection of Phase-II data commences in 2024.

In the final part of our work, we conducted a detailed analysis of the correlation between the reconstructed muonic observables and the composition of primary

cosmic rays, quantifying their capacity for differentiation. The methods developed in the second part of our work can be applied to Phase-II data in the near future, thereby contributing to a deeper understanding of the nature of ultra-high energy cosmic rays.

In this thesis, Chapters 2 to 4 lay the essential groundwork required to comprehend the analyses conducted within our research work. Chapter 2 treats the physics of cosmic rays, emphasising extensive air showers, whilst also providing a brief overview of our current understanding of cosmic rays and the experimental techniques employed in their study. Chapter 3 is devoted to the experimental setup of the Pierre Auger Observatory, with a particular focus on the Surface Detector and the methodologies applied to extract information about the primary particles. In Chapter 4, we discuss the upgrade of the Pierre Auger Observatory, presenting a detailed description of the newly implemented detectors.

Chapters 5 to 9 constitute the core of our research, where the development and results of our work are presented. Chapter 5 outlines our study on the calibration and evaluation of signals provided by the AugerPrime stations. The critical validation of event reconstructions derived from data collected by these upgraded detectors is the focus of Chapter 6.

Following our validation studies, Chapters 7 to 9 address the analyses we performed concerning the application of AugerPrime data for composition identification. In Chapter 7, we detail the Monte Carlo simulations produced to facilitate our investigations. The formulation and evaluation of the matrix formalism for extracting the muonic and electromagnetic signals within AugerPrime stations is thoroughly discussed in Chapter 8. Finally, Chapter 9 is dedicated to our concluding analyses, which centre on applying the matrix formalism to reconstruct muonic observables and evaluate their potential for distinguishing different cosmic ray species.

In conclusion, Chapter 10 summarises the primary contributions and findings derived from our research within the context of the Pierre Auger Observatory and its crucial role in the advancement of the comprehension of ultra-high-energy cosmic rays.

CHAPTER 2

Cosmic rays

This chapter serves to establish fundamental concepts concerning the physics of cosmic rays, with a particular focus on their relevance to the detection of ultra-high energy cosmic rays (UHECRs). Commencing with a concise historical overview, we delve into the physics of extensive air showers through the application of simplified models, facilitating comprehension of their fundamental characteristics. Subsequently, we explore various detection methodologies. Finally, the chapter concludes with a discussion of the most crucial properties exhibited by UHECRs.

2.1 THE DISCOVERY OF COSMIC RAYS

The history of cosmic rays dates back to 1900 when physicists observed the presence of electrical conductivity in the Earth's atmosphere [26, 27]. This discovery suggested the existence of ionising agents, which, initially, was theorised to be caused by the contamination of the environment with radioactive elements.

To delve deeper into this phenomenon, Victor Hess conducted balloon flights to measure ion densities at varying altitudes. His experiments revealed that the ionisation levels increased with altitude. In 1912, Hess reached the crucial conclusion that the ionisation of air molecules must be attributed to ionising particles originating from outer space [28]. This landmark observation is now recognised as the discovery of cosmic rays.

The breakthrough finding of Victor Hess was subsequently validated by Werner Kolhörster, who developed an enhanced measurement equipment and conducted balloon flights at higher altitudes during 1913 and 1914. In recognition of his groundbreaking work, Victor Hess was awarded the Nobel Prize in 1936 for the discovery of cosmic radiation, which he shared with Carl D. Anderson, recognised for his discovery of the positron.

Since these foundational discoveries, extensive efforts have been dedicated to unravelling the nature and origins of cosmic rays. Numerous theoretical models have been developed to elucidate their astrophysical sources, mechanisms of acceleration, and their propagation through interstellar and intergalactic media [2, 16, 17]. On the experimental front, researchers have constructed increasingly precise and large-scale detectors aimed at testing the various models proposed to explain the physics of cosmic rays [9, 11, 12, 29].

2.2 EXTENSIVE AIR SHOWERS

As previously mentioned, the interaction of cosmic rays with atmospheric molecules gives rise to cascades of particles, commonly referred to as *extensive air showers* [30]. This phenomenon holds significant importance in the experimental detection of high-energy cosmic rays, facilitating comprehensive studies of these particles.

It is worth noting that distinct primary cosmic rays, such as protons or heavier atomic nuclei, generate showers with distinct characteristics. Nevertheless, all of these showers share common components, including electromagnetic, hadronic, and muonic constituents. In the following sections, we will delve into the physics of extensive air showers by elucidating these components through the use of simplified models.

2.2.1 HADRONIC AND MUONIC COMPONENTS

Hadronic cascades are initiated when protons or heavier nuclei originating from outer space collide with atmospheric molecules. Following the first interaction with the atmosphere, various species of hadronic particles, including pions, kaons, η , ρ , and

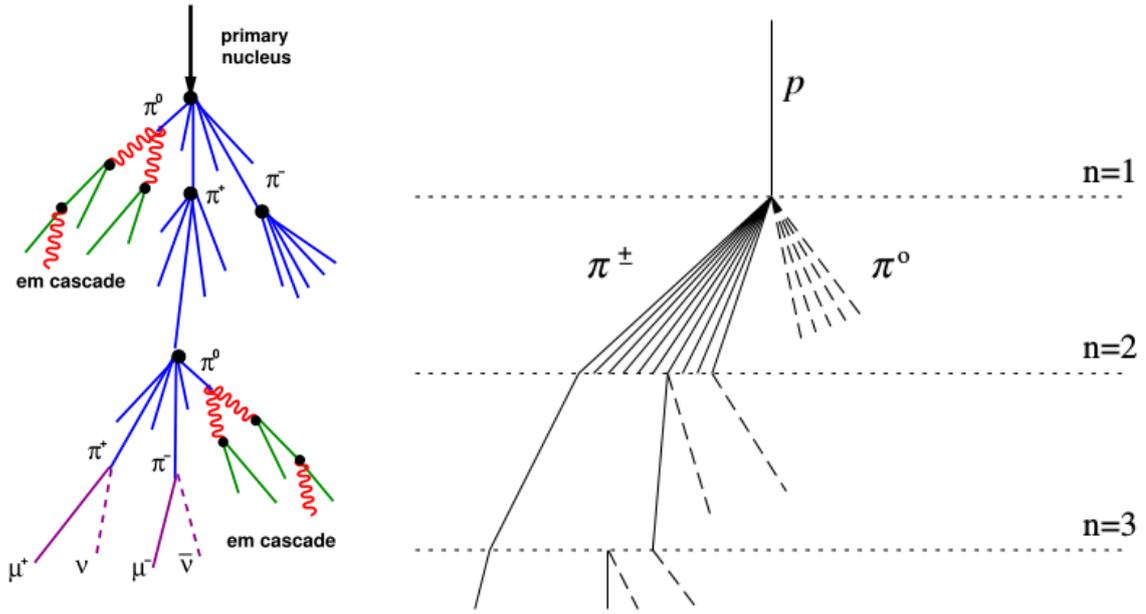


Figure 2.1: *Left*: Illustration of the development of a hadronic cascade [31]. A primary nucleus interacts with atmospheric molecules, predominantly yielding charged and neutral pions. Neutral pions decay, giving rise to electromagnetic cascades. Charged pions further collide with air molecules, leading to the generation of additional hadrons. As the energy of charged pions falls below a critical threshold, they undergo decay, producing muons and neutrinos. *Right*: Schematic representation of Matthews' model for a hadronic shower. In this model, each interaction yields N_{ch} charged pions and $N_{\text{ch}}/2$ neutral pions, with the energy evenly divided among the daughter particles [32].

heavier baryonic resonances, are generated. It is important to note that, among these particles, neutral and charged pions (π^0 and π^\pm , respectively) predominate, produced in roughly equal quantities.

The charged hadrons continue to interact with air molecules, yielding a continuous generation of additional hadronic particles, a process similar to the initial interaction. This ongoing process forms the foundation of what we term the *hadronic component* of the extensive air shower. In contrast, the produced neutral pions rapidly decay into two photons, $\pi^0 \rightarrow \gamma + \gamma$, thereby initiating electromagnetic cascades, which are discussed in the subsequent section. As a result, with each interaction, a portion of the energy within the hadronic component is transferred to the electromagnetic component of the extensive air shower.

The attenuation of the hadronic cascade begins as the characteristic interaction length of charged pions exceeds their decay length into muons and neutrinos, $\pi^\pm \rightarrow \mu^\pm + \nu_\mu/\bar{\nu}_\mu$. This critical energy, denoted as ξ_c^π , at which the probability of pion decay surpasses that of interaction with atmospheric molecules, can be approximated

as roughly 20 GeV. A visual representation of the development of hadronic showers is provided in Figure 2.1 (left).

To describe hadronic showers, we present the simplified Matthews' model, which is influenced by Heitler's work [32]. This model presumes a proton with energy E_0 entering the Earth's atmosphere and colliding with air molecules after traversing an atmospheric length X_0 . This interaction yields N_{ch} and $N_{\text{ch}}/2$ charged and neutral pions, respectively. The initial energy of the proton is evenly shared among these daughter particles. The neutral pions promptly decay into two photons, initiating electromagnetic cascades. Denoting the interaction length between charged pions and air molecules as λ_I , these particles travel a length of $\lambda_I \ln 2$ in the atmosphere, thereby generating additional N_{ch} charged pions and $N_{\text{ch}}/2$ neutral pions. This iterative process, depicted in Fig. 2.1, continues until the energy of the charged pions reaches ξ_c^π , at which point these particles are assumed to decay.

After undergoing n interactions, the quantity of charged pions is given by $N_\pi = N_{\text{ch}}^n$. The energy these particles collectively carry, which represents the energy of the hadronic component, is calculated as:

$$E_{\text{had}} = \left(\frac{2}{3}\right)^n E_0. \quad (2.1)$$

Conversely, the energy E_{em} transferred to the electromagnetic cascades through the decay of the neutral pions can be expressed as:

$$E_{\text{em}} = \left[1 - \left(\frac{2}{3}\right)^n\right] E_0. \quad (2.2)$$

Eq. (2.2) reveals that approximately 90% of the primary energy resides in the electromagnetic component after only six interactions. By dividing Eq. (2.1) by the number of charged pions following n interactions, the individual energy of these pions can be determined as

$$E_\pi = \frac{E_{\text{had}}}{N_\pi} = \frac{E_0}{\left(\frac{3}{2}N_{\text{ch}}\right)^n}. \quad (2.3)$$

This expression can be employed to ascertain the number of interactions n_c required

for the energy of the pions to reach the critical energy ξ_c^π :

$$n_c = \frac{\ln(E_0/\xi_c^\pi)}{\ln(\frac{3}{2}N_{\text{ch}})}. \quad (2.4)$$

Note that the value of n_c exhibits weak dependence on N_{ch} , the number of charged pions generated per interaction. This multiplicity also varies gradually with primary energy, increasing as $E^{1/5}$ for pp and p \bar{p} collisions [3]. Hence, adopting a constant value of $N_{\text{ch}} = 10$ serves as a suitable approximation [32].

Muons within an extensive air shower primarily originate from the decay of charged pions, especially as they attain the critical energy ξ_c^π . These muons constitute what is referred to as the *muonic component* of the shower. In the framework of Matthews' model, the number of muons is assumed to arise solely from the decay of charged pions after they reach the critical energy, thereby expressed by $N_\mu = (N_{\text{ch}})^{n_c}$. By substituting Eq. (2.4) into

$$\ln N_\mu = n_c \ln N_{\text{ch}}$$

one obtains

$$N_\mu = \left(\frac{E_0}{\xi_c^\pi}\right)^\beta, \quad (2.5)$$

where

$$\beta = \frac{\ln N_{\text{ch}}}{\ln(3N_{\text{ch}}/2)}.$$

Adopting a value of $N_{\text{ch}} = 10$ yields $\beta = 0.85$. Hence, the number of muons exhibits a relationship with the primary energy that is less than linear. Results from simulation corroborate values for β ranging from 0.85 to 0.92 [33].

2.2.2 ELECTROMAGNETIC SHOWERS

As previously mentioned, neutral pions undergo decay into a pair of high-energy photons ($\pi^0 \rightarrow \gamma + \gamma$). These energetic photons, at this stage, undergo interactions with atmospheric molecules, generating an electron-positron pair ($\gamma \rightarrow e^- + e^+$). In this situation, it is worth noting that energy loss due to Compton scattering is negligible. Subsequently, the electrons and positrons, through the radiation process known as *Bremsstrahlung*, emit additional photons. This sequence repeats, giving rise to an electromagnetic cascade, visually depicted in Fig. 2.2 (left).

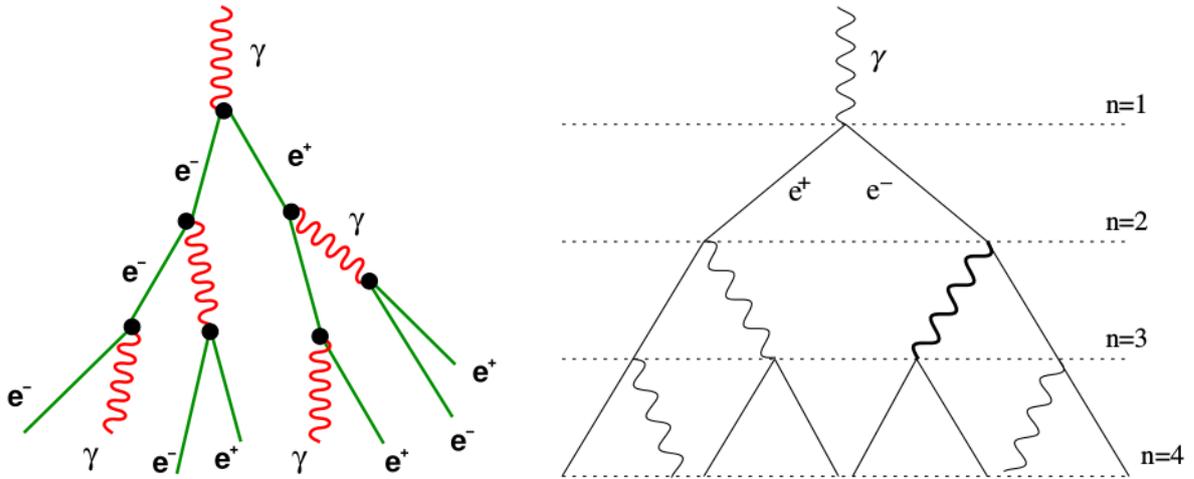


Figure 2.2: *Left*: Visualisation of an electromagnetic cascade [31]. Photons generate electron-positron pairs upon interaction with air molecules. Subsequently, these particles emit photons through the Bremsstrahlung process, giving rise to a cascade of electromagnetic particles. *Right*: Schematic representation of the Heitler model for describing electromagnetic cascades [32]. In this model, each particle interacts after travelling a fixed distance in the atmosphere, leading to the creation of two daughter particles that evenly share the energy of the parent particle.

The term *electromagnetic component* of an extensive air shower encompasses the sum of individual electromagnetic cascades produced by the photons resulting from the decay of neutral pions, which are themselves generated during the development of the hadronic component of the shower.

As the electrons and positrons emit photons, their energy diminishes. The attenuation of an electromagnetic cascade commences when energy losses due to ionisation and excitation of air molecules become more significant than radiative losses. This transition occurs at a critical energy of roughly $\xi_c^e = 85 \text{ MeV}$, in the case of propagation through the atmosphere.

Whilst Monte Carlo simulations provide a precise depiction of electromagnetic cascades, a simplified model can help elucidate their fundamental characteristics. In this model, originally introduced by Heitler [34], a photon with an initial energy of E_0 propagates through the atmosphere. In its first interaction, it generates both a positron and an electron, with each possessing half of the initial energy of the photon. These newly created particles, after travelling a distance $d = \lambda_r \ln 2$, where the radiation length in the medium is denoted by λ_r , give rise to two more particles, each carrying half of the parent particle's energy. Thus, in this simplified model, every particle in-

teracts after travelling a distance d , resulting in two outgoing particles, each sharing an equal portion of the parent particle's energy. A visualisation of an electromagnetic cascade in accordance with Heitler's model is depicted in Fig. 2.2 (right).

Following n successive interactions, occurring at a cumulative distance $x = nd = n\lambda_r \ln 2$, the population of particles within the electromagnetic cascade can be expressed as $N = 2^n = \exp(x/\lambda_r)$, and their individual energies as $E = E_0/2^n = E_0/\exp(x/\lambda_r)$. The proliferation of particles in the cascade persists until the energy of the electrons and positrons diminishes to the critical value ξ_c^e . At this stage, the particle count reaches its maximum, indicated by

$$N_{\max} = \frac{E_0}{\xi_c^e}. \quad (2.6)$$

This expression implies that the number of interactions needed to reach the maximum count N_{\max} , n_c , can be calculated as

$$n_c = \frac{\ln(E_0/\xi_c^e)}{\ln 2}. \quad (2.7)$$

Subsequently, the depth at which the maximum development of the electromagnetic cascade occurs, X_{\max}^{em} , is determined as

$$X_{\max}^{\text{em}} = x(n_c) = \lambda_r \ln \left(\frac{E_0}{\xi_c^e} \right). \quad (2.8)$$

This model-based estimation notably concurs with results from Monte Carlo simulations [32].

The number of particles projected at the stage of maximum shower development, as denoted by Eq. (2.6), is not entirely consistent with empirical findings. This discrepancy can be attributed to the failure of Heitler's model to account for electrons and positrons with reduced energy and the potential creation of multiple photons via Bremsstrahlung. The consistent ratio between the electromagnetic particle count (N_{em}) anticipated by the model and that observed in simulations, irrespective of energy and propagation medium, allows for a reasonable approximation of the order of magnitude for the number of electrons and positrons (N_e). This is achieved by scaling N_{em} with a

constant correction factor, denoted as $g = 10$, such that

$$N_e = \frac{N_{em}}{g}. \quad (2.9)$$

Notwithstanding the identified limitations, Heitler's model aptly describes two important attributes of electromagnetic cascades: the proportionality of the maximum particle count to the primary energy (Eq. (2.6)) and the logarithmic growth of the depth of maximum development of the cascade with respect to the energy (Eq. (2.8)) [32].

2.2.3 CHARACTERISTICS OF EXTENSIVE AIR SHOWERS

Having elucidated the main characteristics of the hadronic, muonic, and electromagnetic cascades through simple models, we will now apply the derived results to formulate expressions that approximate the estimation of energy and the maximum development of extensive air showers initiated by protons. Subsequently, these expressions will be extended to encompass showers generated by heavier nuclei.

To commence, we seek an estimate of the energy of the primary proton, relating it to the total number of muons and electromagnetic particles generated within the extensive air shower. At the point of maximum development of the particle cascade, the model assumes that all particles attain their respective critical energies, denoted as ξ_c^π and ξ_c^e for the hadronic and electromagnetic components, respectively. At this point, the charged pions decay into muons, thereby giving rise to $N_\mu = N_\pi$. The total number of electromagnetic particles following the peak development of the shower can be correlated to the number of electrons through Eq. (2.9), namely, $N_{max}^{em} = gN_e$. Consequently, the primary energy is expressed as follows:

$$\begin{aligned} E_0 &= \xi_c^e g N_e + \xi_c^\pi N_\mu \\ &= g \xi_c^e \left(N_e + \frac{\xi_c^\pi}{g \xi_c^e} N_\mu \right). \end{aligned} \quad (2.10)$$

Utilising the values $\xi_c^e = 85 \text{ MeV}$, $\xi_c^\pi = 20 \text{ GeV}$, and $g = 10$ [32] results in

$$E_0 = 0.85(N_e + 24N_\mu) \text{ GeV}. \quad (2.11)$$

It is important to emphasise that Eq. (2.10) represents an approximation, as during the development of the shower, the energy of a parent particle is typically not evenly distributed among its offspring. Additionally, the model does not account for the energy carried by some remaining hadrons close to the shower core, as well as neutrinos produced in the development of the particle cascade. Nevertheless, this expression aligns remarkably well with energy reconstructions performed by the CASA-MIA experiment [32, 35]. A key takeaway from Eq. (2.10) is that the primary energy exhibits a linear relationship with the number of particles in the shower.

Determining an accurate estimation for the depth of the shower maximum, denoted as X_{\max} , within the simplified models employed is a rather challenging task. As previously mentioned, following a few interactions, a proton-initiated shower is predominantly comprised of the electromagnetic component. A more precise treatment would entail accounting individually for each electromagnetic cascade. However, in the framework of Matthews' model, it is possible to gain an approximation by considering solely the initial electromagnetic cascade generated.

In this approach, the primary proton is postulated to interact at an atmospheric depth of X_0 . Given that one-third of its products constitute neutral pions, each with an individual energy of $E_{\pi^0} = 2E_0/3N_{\text{ch}}$, and since neutral pions decay into two photons, with the assumption that these photons equally share the energy of the parent π^0 , each photon is endowed with energy $E_\gamma = E_0/3N_{\text{ch}}$. According to Eq. (2.8), these photons induce electromagnetic cascades that reach their maximum development at a depth expressed as

$$X_{\max} = X_0 + \lambda_r \ln \left(\frac{E_0}{3N_{\text{ch}} \xi_c^e} \right). \quad (2.12)$$

We emphasise that Eq. (2.12) should not be regarded as a precise prediction but rather as a lower limit to the real X_{\max} . When compared to simulation results, it yields an estimate of approximately 100 g cm^{-2} shallower [32]. This discrepancy arises from the omission of the subsequent generation of electromagnetic cascades and its failure to account for the non-uniform distribution of energy amongst the daughter particles. Nonetheless, Eq. (2.12) underscores the dependence of X_{\max} on primary energy and interaction multiplicity.

Heavier nuclear primaries

Thus far, we have delineated the characteristics of extensive air showers generated by protons. It is possible to extend these descriptions to encompass heavier nuclei by employing the *superposition model*. This approach considers that primary particles with an atomic number of A and an energy of E_0 are composed of A independent nucleons, each possessing an energy of E_0/A . These nucleons initiate hadronic showers, as described by Matthews' model.

Employing this concept, we can readily derive expressions for the number of muons, the energy, and the depth of the shower maximum, akin to Eqs. (2.5), (2.10) and (2.12), respectively.

We commence by examining the number of muons generated by a nucleus with an atomic number A and an energy of E_0 . Utilising the superposition model, this can be expressed as

$$N_{\mu}^A(E_0) = AN_{\mu}^P(E_0/A) = A \left(\frac{E_0}{A\xi_{\pi}^c} \right)^{\beta} = A^{1-\beta} N_{\mu}^P(E_0). \quad (2.13)$$

Here, the superscripts A and p denote a nucleus and a proton, respectively. By employing the previously determined value of $\beta = 0.85$ in Eq. (2.13), we arrive at the relationship

$$N_{\mu}^A(E_0) = A^{0.15} N_{\mu}^P(E_0). \quad (2.14)$$

Consequently, a shower initiated by a nucleus yields more muons than an equivalent shower initiated by a proton with the same energy. For instance, an iron nucleus will generate approximately 1.8 times more muons than an equally energetic proton. This variation arises due to the nonlinearity of the energy dependence of N_{μ} , as evidenced in Eqs. (2.5) and (2.13).

In a similar fashion, we can extend our analysis to calculate the maximum depth of shower development, denoted as X_{\max}^A , for nuclei as follows:

$$X_{\max}^A = X_{\max}^P - \lambda_r \ln A. \quad (2.15)$$

Here, X_{\max}^P represents the depth of the shower maximum for a primary proton with the same energy as the nuclear cosmic ray. It is important to acknowledge that Eq. (2.15)

originates from Eq. (2.12), which is known to lack accuracy in describing X_{\max}^{P} for the aforementioned reasons. Nevertheless, Eq. (2.15) yields the prediction that heavier nuclei will exhibit a shallower X_{\max} than an equally energetic proton. As an example, iron nuclei will manifest a shower maximum on average 150 g cm^{-2} higher than proton primaries with the same energy. Remarkably, this outcome aligns with the findings obtained from simulations.

Furthermore, the relationship between energy and the number of particles generated remains consistent with what was presented in Eq. (2.10), represented by

$$E_0^{\text{A}} = g \xi_c^e \left(N_e + \frac{\xi_c^\pi}{g \xi_c^e} N_\mu \right). \quad (2.16)$$

This consistency arises from the fact that, in both scenarios, the total energy of the electromagnetic and hadronic components is encompassed through the summation of the total number of particles. However, it is essential to note that, as emphasised in Eq. (2.13), heavier primary particles generate cascades with a larger number of particles, contributing to these established relationships.

2.3 EXPERIMENTAL DETECTION OF COSMIC RAYS

The experimental detection of cosmic rays can be categorised into two distinct groups: direct and indirect detection methods, primarily contingent on the energy range under consideration.

For cosmic rays with energies below approximately 10^{14} eV, the flux of these particles is sufficiently high to enable direct detection. Detectors employed in this context may include calorimeters, emulsion stacks, or transition radiation detectors, akin to technologies used in high-energy physics experiments with particle accelerators. These detectors are situated in various locations, including the International Space Station (e.g., ISS-CREAM [36] and AMS [6], as schematically depicted in the left panel of Fig. 2.3), satellites such as PAMELA [7], or high-altitude balloons, as observed in the ATIC [5] and TRACER [37] experiments. Direct detection techniques yield highly precise measurements of both the energy and composition of cosmic-ray particles.

Conversely, for cosmic rays with energies surpassing 10^{15} eV, the flux of

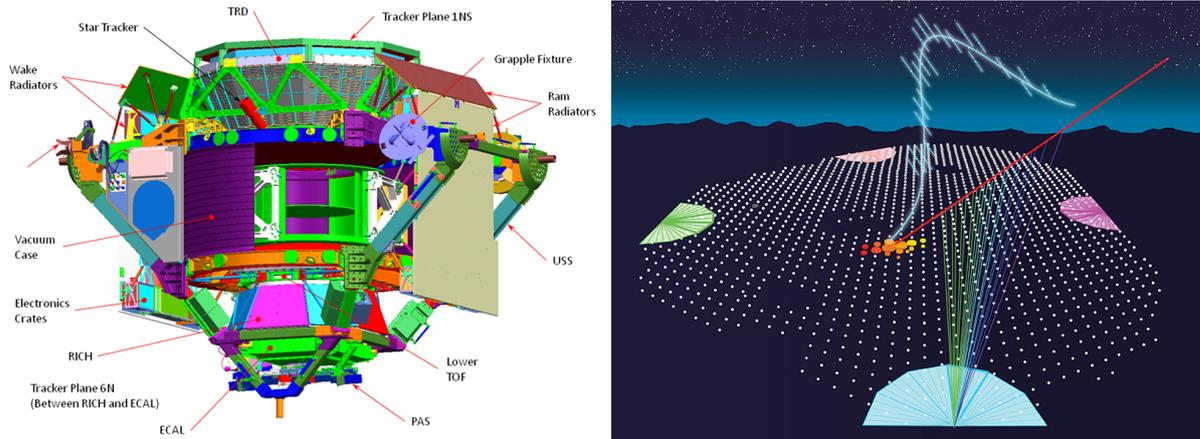


Figure 2.3: The experimental detection of cosmic rays can be categorised as either direct or indirect, contingent upon their flux. *Left:* The AMS detector [38], situated aboard the International Space Station, investigates cosmic-ray particles with energies below 10^{15} eV. It employs a direct detection approach akin to those utilised in high-energy physics experiments with particle accelerators. *Right:* The Pierre Auger Observatory, on the other hand, probes extensive air showers induced by primary cosmic rays with energies exceeding 10^{17} eV. This observatory utilises a hybrid technique that combines an array of water-Cherenkov detectors with surrounding fluorescence telescopes [39].

cosmic rays becomes exceedingly low, rendering direct detection practically unfeasible. To exemplify, at energies exceeding 5×10^{15} eV, the detection rate amounts to less than one particle per square meter per year. In such cases, information about the primary cosmic rays is derived indirectly through the detection of the secondary particles generated within the extensive air showers they generate, as elaborated upon in the preceding section.

Ground-based arrays of detectors spanning a significant area serve as the primary means to detect the particles generated within air showers. These detectors encompass various technologies, including scintillators, as utilised by initiatives like AGASA [29] in Japan and KASCADE [10] in Germany, and water-Cherenkov tanks, notably pioneered by the Haverah-Park [40] experiment in the United Kingdom. An additional noteworthy technique is the detection of atmospheric light emissions. As relativistic charged particles from air showers traverse the atmosphere, they emit highly collimated Cherenkov radiation. Furthermore, they excite atmospheric nitrogen molecules, resulting in the emission of fluorescence light within the ultraviolet spectrum. Some experiments, exemplified by the Fly's Eye [9] and Hi-Res [41] initiatives in the United States, exploit the detection of such light emissions.

In contemporary times, observatories such as the Pierre Auger Observat-

ory [12] in Argentina, depicted in Fig. 2.3 (right), and the Telescope Array [11] in the United States employ a hybrid approach, combining ground-based arrays with fluorescence-light detectors. This hybrid method significantly enhances the quality of data obtained. Nevertheless, it is important to note that the uncertainties in measurements, particularly regarding parameters like energy and composition, remain significantly larger when compared to direct detection methods.

2.4 FUNDAMENTAL PROPERTIES OF COSMIC RAYS

Numerous experiments, as mentioned in the previous section, have been conducted to unravel the fundamental properties of cosmic rays. These investigations primarily aim to elucidate the origin of these particles, the mechanisms responsible for their acceleration, and their propagation dynamics within the interstellar and intergalactic mediums. The examination of the energy spectrum of cosmic rays, investigations of their chemical composition, and other methodologies, including anisotropy analyses, collectively serve as indispensable means for probing these inquiries.

In Fig. 2.4, the differential flux of cosmic rays, as measured by different experiments, is depicted as a function of energy, which is also referred to by the term *cosmic ray spectrum*. The flux has been multiplied by $E^{2.6}$, with E denoting the energy of the cosmic rays, to enhance the visibility of spectral features. Below energies of 10^{10} eV, the flux experiences attenuation due to solar winds, which displace these less energetic particles from the solar system. Hence, this region of the spectrum is significantly influenced by solar activity.

In a first approximation, the cosmic ray spectrum is characterised by three prominent regions, where each can be expressed by a power-law function as defined by

$$\frac{d\phi}{dE} \propto E^{-\gamma}, \quad (2.17)$$

where E represents the energy, and ϕ signifies the flux of the corresponding cosmic rays. The first region extends to the *knee* of the spectrum at roughly 4×10^{15} eV, exhibiting a spectral index of $\gamma \sim 2.7$. Beyond this point, the differential flux diminishes more rapidly with $\gamma \sim 3.1$ continuing until the energy of 5×10^{18} eV, referred to as the *ankle*. Between the knee and the ankle, a *second knee* emerges at 10^{17} eV, with γ

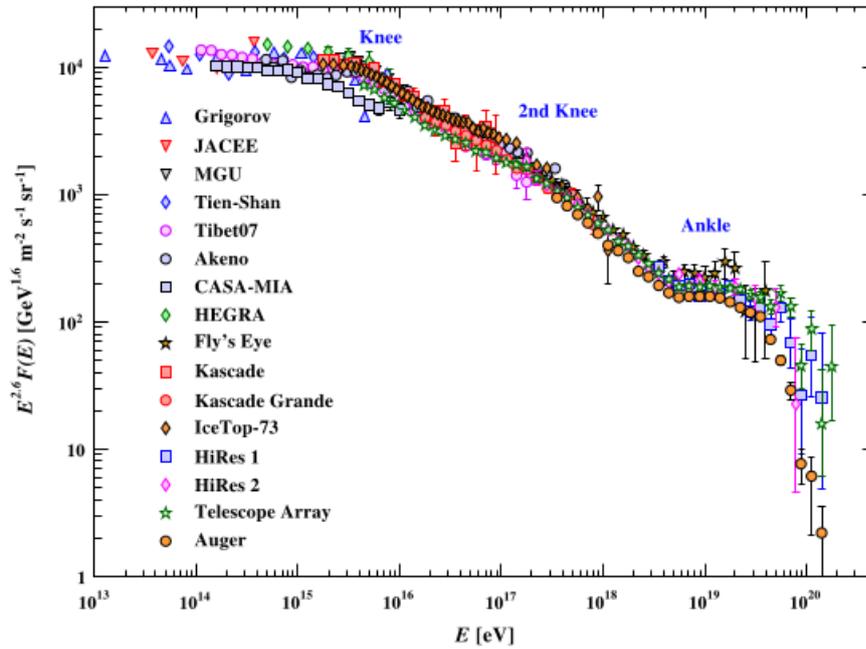


Figure 2.4: Cosmic-ray spectrum consisting of the differential flux as a function of the energy [42]. To highlight the distinguishing features of the spectrum, the differential flux has been multiplied by energy to the power of 2.6. The spectrum exhibits three prominent regions, marked by the points known as the knee and ankle. Each of these regions can be adequately described by a power-law function (refer to Eq. (2.17)), reflecting the underlying acceleration processes of these particles. The flux suppression becomes evident at energies exceeding 4×10^{19} eV.

approximately equal to 3.0 on its left and 3.3 on its right, marking a further steepening. Subsequently, the spectrum experiences a hardening, presenting $\gamma \sim 2.5$ [3]. Finally, a strong suppression of the flux is observed above a few tens of EeV [15].

The power-law shape exhibited by the cosmic-ray spectrum indicates that their acceleration processes are of a non-thermal nature. Enrico Fermi initially postulated that cosmic rays traversing magnetised regions in motion could result in such a spectrum [43]. A more refined understanding emerges when considering shock waves passing through magnetised regions, often generated by phenomena like supernova explosions. Additionally, potential sources of cosmic rays include the vicinity of black holes and neutron stars [44].

The data gathered by the Pierre Auger Observatory has facilitated the measurement of the differential flux of cosmic rays within the highest energy range of the spectrum [4], achieving an unprecedented level of accuracy. Notably, a distinct suppression of the flux becomes evident for energies surpassing 3.9×10^{19} eV. The exact cause of this suppression remains unclear, with distinct astrophysical scenarios proposed to

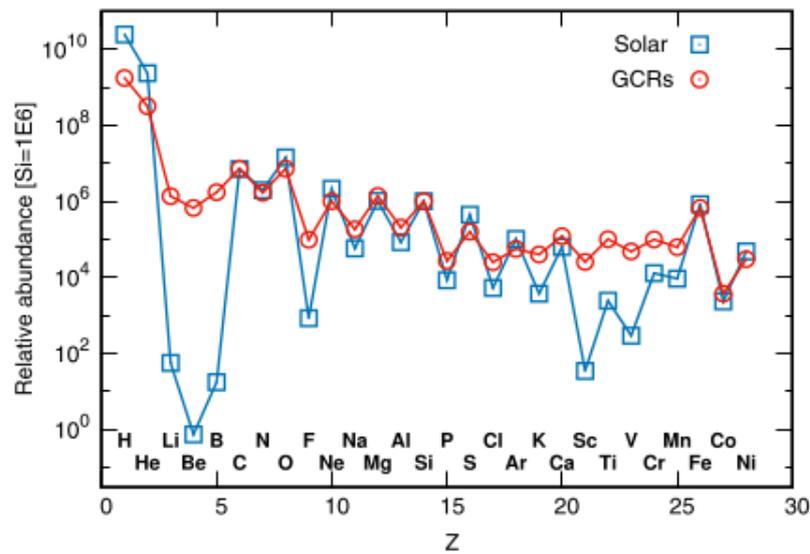


Figure 2.5: Relative abundances of elements in low-energy cosmic rays and in the solar system. The abundances were normalised so that they are 10^6 for Si [31]. The overall abundances are similar for cosmic rays and the solar system. The differences are mainly due to the spallation of heavier nuclei producing lighter ones.

account for this phenomenon.

For instance, it is proposed that high-energy protons may interact with photons from the cosmic background radiation, giving rise to pions. This process, known as the Greisen-Zatsepin-Kuzmin (GZK) effect, predicts a spectral cutoff [16, 17]. Alternatively, the measured suppression could stem from intrinsic energetic limitations within astrophysical sources [45, 46]. To address this question conclusively, more precise data pertaining to the composition of ultra-high energy cosmic rays is essential, as will be discussed later.

Fig. 2.5 illustrates a comparative analysis of the relative composition of low-energy cosmic rays overlaid with the elemental composition of the solar system. These compositions were normalised to a value of 10^6 for silicon (Si). The striking resemblance between these compositions implies that the constituents of low-energy cosmic rays are likely formed via similar processes to those that shape planetary systems, specifically, stellar nucleosynthesis.

Despite the overall similarity, some notable disparities are evident. Specifically, the abundances of hydrogen (H) and helium (He) are more pronounced in the solar system compared to cosmic rays, which can likely be attributed to the higher ionisation potential of these elements, rendering their acceleration from source regions more

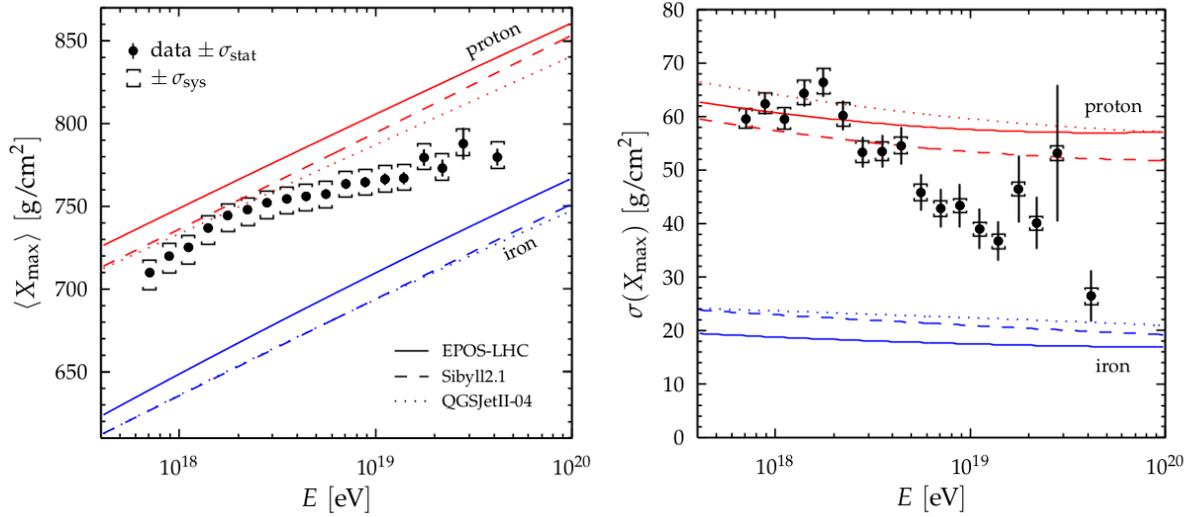


Figure 2.6: Average X_{\max} (on the left) and the corresponding standard deviation (on the right) depicted as functions of energy. The simulation results illustrate the behaviour of proton- and iron-induced showers. A discernible trend in X_{\max} is evident above energies of 3 EeV, indicating a transition for a heavier composition of cosmic rays. Taken from [24].

challenging. Another noticeable distinction lies in the elevated abundances of lithium (Li), beryllium (Be), and boron (B) within cosmic rays, attributable to their production through spallation events when heavier elements, such as carbon (C), nitrogen (N), and oxygen (O), interact with the interstellar medium, resulting in their fragmentation into lighter elements. This process similarly occurs for neon (Ne), yielding fluorine (F), and for iron (Fe) and nickel (Ni), generating elements spanning from scandium (Sc) to manganese (Mn) [31].

For high-energy cosmic rays, as opposed to their low-energy counterparts, determining their composition presents a greater challenge due to the fact that the available data originates from observations of extensive air showers initiated by the primary particles. Consequently, this task demands reliance on simulations and the extrapolation of data related to hadronic interactions, spanning several orders of magnitude, to encompass ultra-high-energy cosmic rays. Through these simulations and extrapolations, it becomes possible to interpret the collected data and derive estimates regarding the composition of the primary cosmic rays.

The Fluorescence Detector at the Pierre Auger Observatory (see Chapter 3) provides the longitudinal profile of the detected EASs, enabling the determination of the depth at which the maximum development of the shower occurs, denoted as

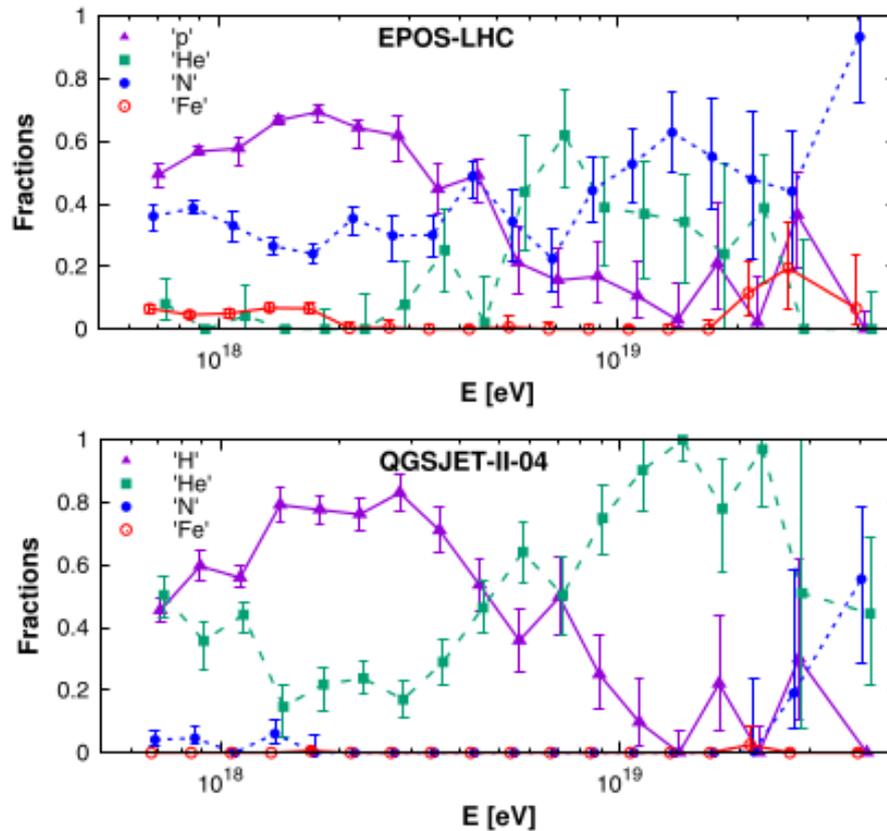


Figure 2.7: Variation in the composition of high-energy cosmic rays relative to their energy. Interpretation of the data was conducted using the hadronic interaction models EPOS-LHC and QGSJET-II-04. Whilst the choice of model influences the results, both models suggest a shift towards a heavier composition for energies exceeding 10 EeV. Taken from Ref. [31].

X_{\max} [18]. This observable, in turn, serves as a tool for estimating the mass¹ of the primary cosmic ray. The panels in Fig. 2.6 depict the first and second moments of X_{\max} as functions of shower energy. The plots include the expected outcome for proton and iron primaries derived from Monte Carlo simulations [24].

Notably, the mean X_{\max} values align with a preference for lighter cosmic ray compositions up to energies of 3×10^{18} eV. However, beyond this energy threshold, there is a noticeable shift towards heavier compositions. This trend is confirmed by the standard deviation in X_{\max} . It is important to note that the Fluorescence Detector operates with a relatively low duty cycle of around 15%, resulting in limited statistics for high-energy events exceeding 3×10^{19} eV, particularly within the region of the flux suppression. The final data point in Fig. 2.6 represents all events with energies

¹The origin of the correlation between X_{\max} and the atomic mass of the primary cosmic ray can be understood from Eq. (2.15).

surpassing 3×10^{19} eV.

Fig. 2.7 illustrates an estimation of the composition of high-energy cosmic rays, employing two distinct models for hadronic interactions, EPOS-LHC [20] and QGSJET-II-04 [21], when analysing the data. It is worth noting that the elements H, He, N, and Fe should be interpreted as groups of elements with atomic masses in proximity to these values, given the current limitations in data resolution for element-by-element differentiation.

When comparing the two plots, the influence of the chosen hadronic interaction model on the results becomes evident. However, despite the model dependency, both EPOS-LHC and QGSJET-II-04 indicate a shift towards heavier cosmic ray compositions for energies exceeding 10^{19} eV. Ongoing efforts aimed at refining the precision of high-energy cosmic ray data are underway, notably led by the Pierre Auger Collaboration, as discussed in Chapter 4.

CHAPTER 3

The Pierre Auger Observatory

The Pierre Auger Observatory is at the heart of all the research detailed in this thesis. Hence, in this chapter, we describe its fundamental components before the upgrade initiative started. To begin, an overview of the Observatory and its instrumentation is presented in Section 3.1. Given that our work predominantly centres on a specific subset of detectors within the Observatory, known as the surface detector, Section 3.2 delves into their characteristics with greater depth. To conclude, we elucidate how data from this detector is utilised for the reconstruction of events triggered by high-energy cosmic rays, accompanied by insights into the computational framework employed to accomplish this and other tasks within the workflow of the Auger Collaboration.

3.1 GENERAL DESCRIPTION OF THE OBSERVATORY

In 1991, Jim Cronin and Alan Watson conceived the Pierre Auger Observatory with the primary objective of investigating cosmic-ray particles exceeding energies of 10^{17} eV, the most energetic particles observed in the universe. Precise data is essential for scrutinising hypotheses concerning the sources, acceleration dynamics and propagation across the interstellar space of these ultra-high energy cosmic rays.

The Pierre Auger Observatory addresses the very low flux of cosmic rays in the energy range of interest by deploying detectors across an extensive area of approximately 3000 km^2 , in the province of Mendoza, Argentina. Such an impressive dimen-

sion entitles the Pierre Auger Observatory as the current largest cosmic-ray experiment in the world. The configuration of the Observatory enables the detection of extensive air showers generated when primary cosmic particles interact with Earth's atmosphere. Beyond the primary aim of facilitating the study of cosmic rays, this approach additionally offers the possibility of investigating fundamental particle interactions, exploiting energy ranges far beyond the capability of human-made accelerators. For instance, a cosmic-ray particle with 10^{19} eV is equivalent to a centre-of-mass energy of 100 TeV.

The Observatory utilises two complementary and independent techniques to measure the properties of the detected extensive air showers [12]. The apparatus involves an array of 1660 water-Cherenkov detectors known as the surface detector array (SD), complemented by 24 fluorescence telescopes distributed across four observation sites, each equipped with six telescopes. An overview of the Observatory is depicted in Fig. 3.1, where each dot represents a water-Cherenkov detector. Additionally, the four fluorescence observation sites are illustrated by lines demarcating the field of view for each telescope. Fig. 3.2 displays images of a surface detector station and a fluorescence detector site.

The water-Cherenkov detectors (WCDs) are positioned in a triangular grid arrangement. In the regular array, the 1600 stations are distributed at intervals of 1500 m from their nearest neighbours, whilst an additional 60 detectors maintain a more compact spacing of 750 m, forming what is referred to as the *infill* array. Each station is essentially comprised of a water-filled tank housing the required electronics for signal collection and processing.

When charged particles from extensive air showers traverse the water volume of a detector at velocities surpassing that of light propagating in water, the emission of Cherenkov radiation occurs, yielding a signal proportional to the count of particles crossing the station. In essence, the water-Cherenkov detectors provide a measurement of the particle density of the shower at ground level.

The fluorescence detector (FD) operates during moonless nights, resulting in a duty cycle of approximately 15%, a notable contrast to the nearly continuous operation of the surface detector. The telescopic mirrors are composed of smaller hexagonal or rectangular mirror segments, directing light towards the camera. This camera consists of 440 photomultiplier tubes (PMTs), specifically the Photonis model XP3062, organised

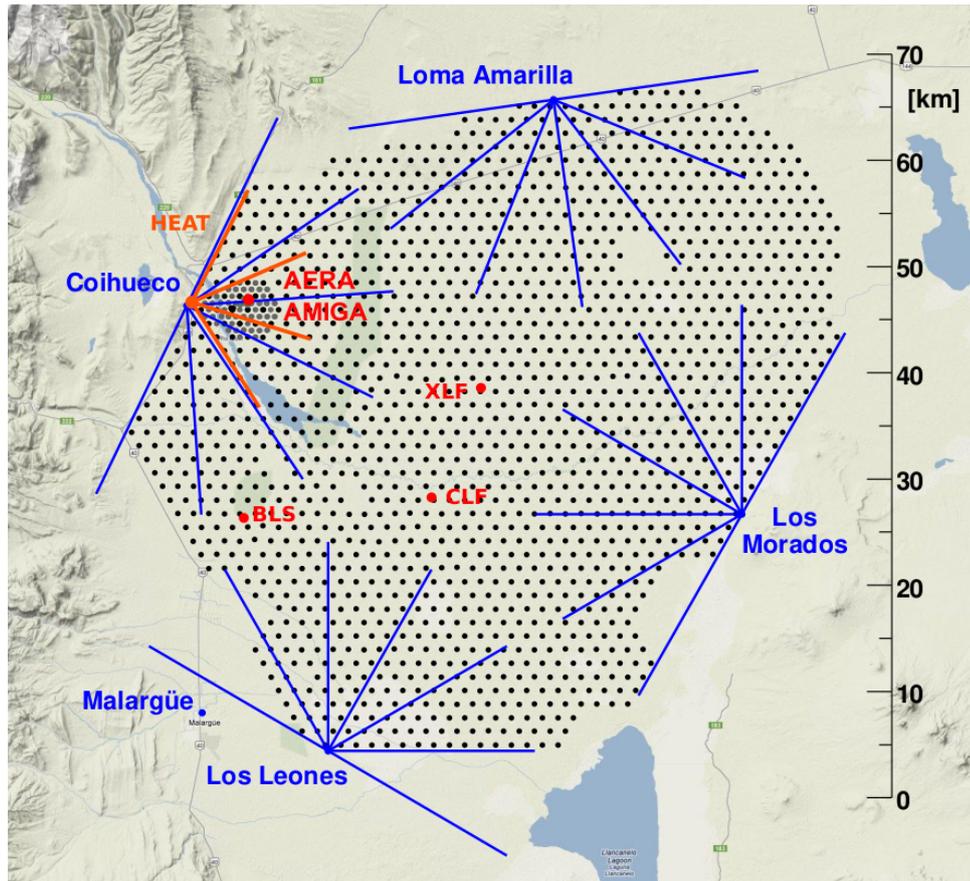


Figure 3.1: Schematic representation of the Pierre Auger Observatory located in Argentina (taken from Ref. [47]). Each red dot corresponds to a water-Cherenkov detector constituting the surface detector array, which itself spans an approximate area of 3000 km^2 . The depiction also illustrates the four fluorescence detector sites, accompanied by the individual field of view for each telescope. The detectors of the infill array are displayed near the Coihueco site.

in a grid of 22 rows and 20 columns, where each photomultiplier serves as an individual pixel. Positioned just beyond the telescope aperture, an ultraviolet filter permits the transmission of over 80% of the light within the wavelength range of 330 to 380 nm. Subsequently, the aggregated electronics digitises and processes the output signal from the camera.

Each fluorescence telescope spans a field of view of 30° in both azimuth and elevation (refer to Fig. 3.1). The interaction between the electromagnetic component of extensive air showers and nitrogen molecules in the atmosphere yields isotropic fluorescence emissions within the ultraviolet spectrum. The fluorescence telescopes record these emissions as the shower progresses through the atmosphere, allowing the longitudinal profile of the shower to be observed. With the intensity of the fluorescence light intrinsically related to the energy deposited by the shower particles, integrating

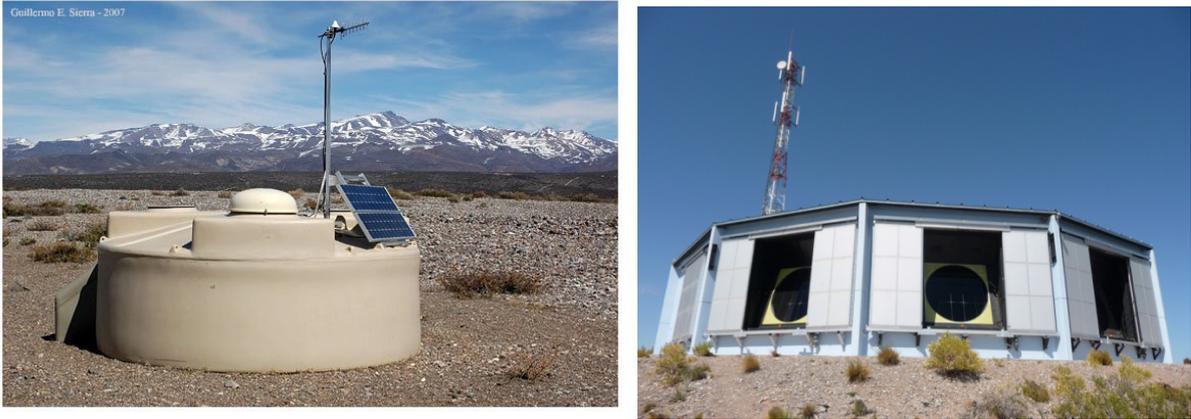


Figure 3.2: Left panel: Water-Cherenkov detector of the surface detector array [48]. Right panel: One of the four fluorescence detector sites housing six telescopes [12], with each telescope encompassing a field of view spanning 30° in both azimuth and elevation.

the energy deposit along the shower axis provides an almost calorimetric estimate of the energy of the primary particle.

By recording the profile of the particle cascade, the determination of the maximum development of the shower becomes straightforward. This observable is closely correlated to the composition of the primary particle, namely, lighter primaries result, on average, in a deeper maximum within the atmosphere. As the primary focus of this thesis does not involve the fluorescence detector, readers are directed to Ref. [14] for a more comprehensive description of this detection system.

The hybrid configuration of the Observatory strategically combines the two distinct techniques, surface and fluorescence detectors, resulting in data of improved precision compared to the capabilities of each method in isolation. A crucial aspect of this hybrid design is its application to the estimation of energy, particularly concerning events exclusively detected with the surface array.

As previously noted, the fluorescence detector offers a near-calorimetric measurement of the energy, however, its operational window constitutes merely around 15% of the time. Conversely, the duty cycle of the surface detector nears 100%. Given that the surface detector measures the particle density at ground level, estimating energy solely via the SD would necessitate reliance on simulations of particle cascades. These simulations, however, introduce substantial systematic uncertainties due to the extrapolation of accelerator-derived data concerning hadronic interactions to the significantly higher energy range of the detected events.

At this point, the advantage provided by the hybrid configuration becomes evident: by using events registered by both detectors, referred to as *hybrid events*, the observables reconstructed with measurements of the surface detector can be calibrated to the direct estimations of the energy obtained from the **FD**. Consequently, even when the fluorescence detector is inoperative, a reliable determination of energy can be conducted using solely the **SD**. Beyond its implications for energy estimation, this hybrid setup facilitates cross-validation of measurements encompassing mass composition and arrival direction. Further details of the procedure for energy calibration are discussed in Section 3.3.

3.2 THE SURFACE DETECTOR ARRAY

Each water-Cherenkov detector within the **SD** array is composed by a cylindrical tank housing 12 000 litres of ultra-pure water encased in a liner. Positioned atop this tank are three photomultiplier tubes symmetrically arranged to observe the water volume. These **PMTs** capture the Cherenkov light emitted during the passage of relativistic charged particles through the water. To ensure sustained operation, each station is self-powered, harnessing energy from two solar panels coupled with auxiliary batteries, resulting in the aforementioned nearly continuous functioning. A visualisation of the constituents of an **SD** station is depicted in Fig. 3.3.

3.2.1 HARDWARE OF THE WATER-CHERENKOV DETECTORS

The water-Cherenkov detectors possess a diameter of 3.6 m and a height limited to 1.6 m. Constructed from high-density polythene via a process of rotational moulding [12, 13], the structure of the station is characterised by a dual-layered wall. The outer beige layer is backed by a black inner layer, occupying two-thirds of the wall thickness, which results in a dark interior.

Contained within a liner manufactured from low-density polythene film, the water volume is shielded by an interior Tyvek[®] layer, which diffusely reflects **UV** Cherenkov light produced within the water. The liner, when filled, accommodates a water height of 1.2 m and also serves the purpose of barring external light from

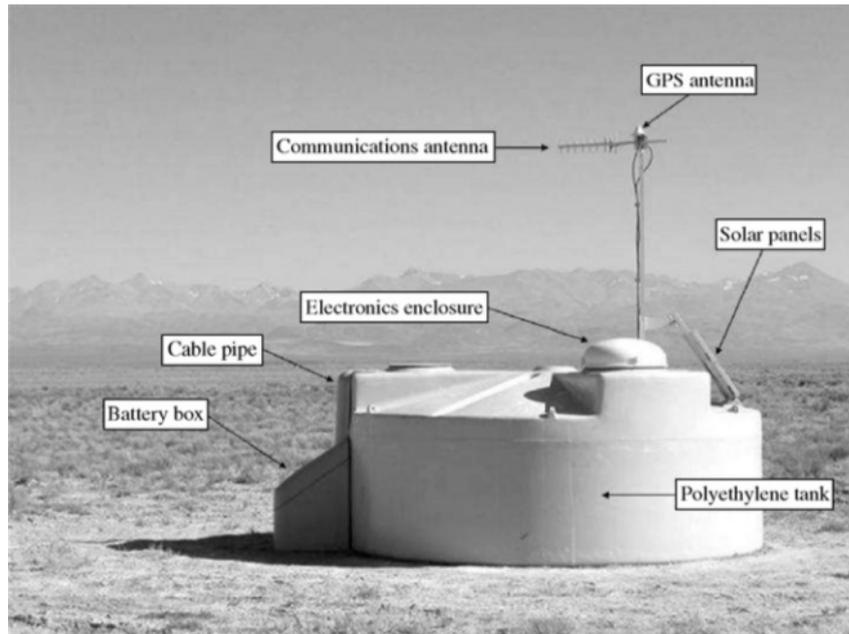


Figure 3.3: Picture of a water-Cherenkov detector and its constituents [13]. The WCD consists of a cylindrical polythene structure housing a liner that contains 12 000 litres of ultra-pure water. Three PMTs collect Cherenkov photons produced in the water volume. Their signals are digitised by a 10-bit ADC at a sampling rate of 40 MHz. A GPS receiver provides event timing with a precision of 10 ns. The station is powered by solar panels attached to two 12 V batteries.

infiltrating the water volume. The liner features three dome windows, providing optical access for the photomultiplier tubes, whilst five additional smaller ports enable water filling and also serve as windows for LED flashers, employed for PMT testing.

The three photomultiplier tubes within the detector (the Photonis XP1805/D1 type with eight dynodes and nine-inch diameter [49]) are symmetrically positioned at a distance of 1.2 m from the central axis of the station. Enclosed to isolate external light and environmental influences, the PMTs reside within a protective case.

Enabling access to the station interior, three hatches are located at the top of the station, a large hatch measuring 560 mm in diameter, alongside two smaller 450 mm diameter hatches. The large hatch is covered by a dome that encloses the electronics of the station (as visible in Fig. 3.3).

The battery box, also made from polythene, is thermally insulated with 50 mm layers of polystyrene foam. The power is sourced from a pair of 55-Wp (Watt-peak) solar panels, which charge two 12 V batteries connected in series. This arrangement furnishes the required 10 W for the station electronics, thus supporting its operation over 97% of the time.

Power cables run from the solar panels through the electronics enclosure and traverse the interior of the station, reaching the battery box. Sensors are present to monitor battery and photomultiplier tube voltages, electric currents, and temperatures every six minutes. If necessary, the electronics board allows the station, and consequently the entire array, to be remotely shut down. Serving the purpose of communication with the Central Data Acquisition System (CDAS), the station is equipped with an antenna. A GPS receiver is also present for synchronisation and event timing with precision of 10 ns (see Fig. 3.3).

The PMTs receive high voltage via a module mounted in their base. This high voltage is proportional to a DC control voltage, which is locally supplied by the slow control system. Each PMT generates two signals: one from the anode and another from the last dynode. The latter is inverted and amplified 32 times the anode charge gain. For this reason, it is referred to as the high gain (HG) channel, whilst the anode signal is denoted as the low gain (LG) channel.

After filtered by a 5-pole Bessel filter, the signals undergo digitisation utilising a dedicated semi-flash analog-to-digital converter (ADC) with 10-bit precision, operating at a frequency of 40 MHz. Combining the 32-time amplification of the dynode signals with the 10-bit ADCs, the system achieves a dynamic range of 15 bits [12].

The digitised outputs are stored in a buffer memory via a programmable logic device, which promptly informs the station microcontroller of any triggers. The microcontroller, interfaced by an IBM PowerPC 403 GCX CPU with a clock rate of 80 MHz, subsequently communicates the local triggers to CDAS. In case of temporal coincidence with nearby stations, CDAS requests local data to construct an event. A *unified board* (UB) incorporates the station microcontroller, slow control functions, event timing and communications system, thus yielding the front-end interface.

3.2.2 CALIBRATION OF THE PHOTOMULTIPLIER TUBES

The photomultiplier tubes within the stations are calibrated to establish a consistent reference point for all WCDs of the surface detector array. This calibration process effectively converts the charge of signals, as measured in hardware units (integrated ADC channels) into a physical unit reflecting the number of particles traversing the detector. The chosen physical unit is termed the *vertical equivalent muon* (VEM), representing the

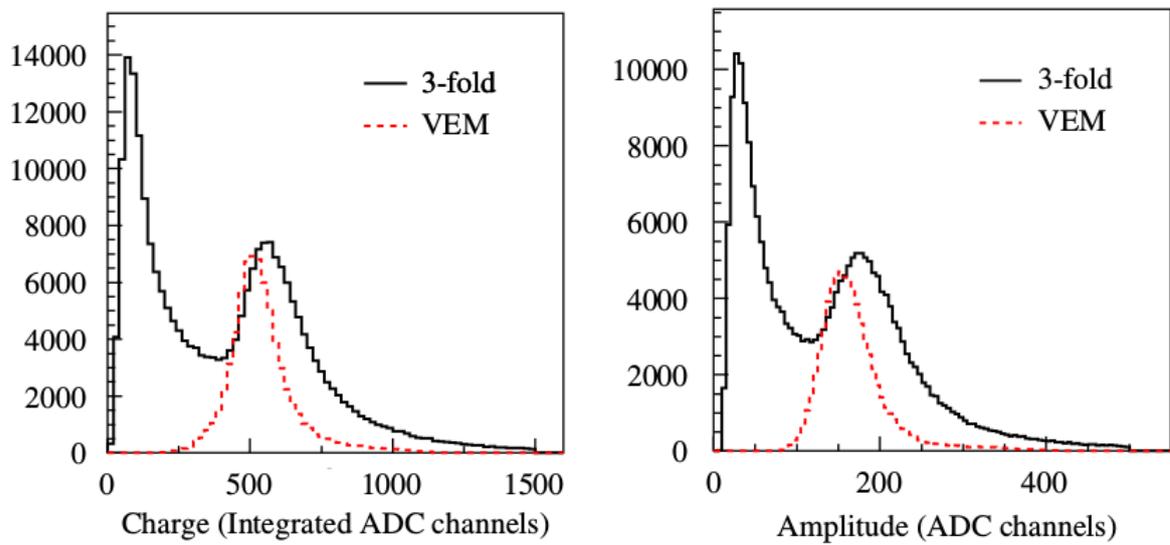


Figure 3.4: Histogram of signal charge (left) and amplitude (right) generated by atmospheric charged particles traversing a reference station (black line). The red dashed lines depict the corresponding results for vertically centred muons, determined using plastic scintillators positioned both above and below the reference station [51]. The peak of the charge distribution for omnidirectional muons is 1.03 (1.09) times that of a VEM for individual PMTs (all PMTs).

average charge generated in the PMTs when a vertical muon traverses the centre of a station.

Additionally, the calibration process determines the peak of the distribution of signal amplitude, measured in ADC channels, arising from atmospheric muons. This calibrated value plays an essential role in establishing the local triggers for each station, thereby ensuring uniform trigger conditions throughout the entire surface detector array.

When operating individually, a station lacks the capability to select vertical muons for calibration. Hence, an indirect yet reliable approach becomes necessary [50–52]. The charge spectrum of background charged particles was acquired utilising a reference station, as depicted in Fig. 3.4. A 3-fold coincidence, requiring simultaneous signals from all three PMTs, served as the triggering criterion, with the charge values representing the sum registered by these three PMTs. In the distribution, two discernible peaks are evident.

The first peak originates from electrons and high-energy gammas, which are capable of generating electron-positron pairs within the water volume. The second peak, on the other hand, corresponds to atmospheric muons. To obtain the charge dis-

tribution of vertically centred muons, plastic scintillators were strategically positioned both above and below the reference station. This distribution is represented by the dashed red line in Fig. 3.4.

This analysis revealed that the peak of the charge distribution for atmospheric muons corresponds to approximately 1.09 times the charge of a vertical centre-going muon, when considering the sum of the three PMTs and 1.03 times a VEM for each individual PMT [50, 51]. The shift in the peak, observed in the distribution of atmospheric muons as compared to that of strictly vertical muons, can be attributed to varying track lengths traversed by background muons, which arrive at the station from various angles, as opposed to the consistent path length of vertical muons [52].

Using the relationship between the peak charge in the atmospheric muon distribution (Q_{peak}^{μ}) and the average charge of a vertically centred muon (VEM or Q_{VEM}), the calibration of the PMTs is achieved through the following steps.

Initially, the high voltage applied to the PMTs is adjusted so that the rate of singles, 150 ADC channels above baseline, reaches 100 Hz. This adjustment aligns the peak of the amplitude distribution generated by atmospheric muons (I_{peak}^{μ}) at approximately 50 ADC channels. Consequently, the gains of the stations may vary, for instance, if the water quality in one tank facilitates better photon propagation, that station will exhibit lower gain.

Following the gain adjustment, fluctuations in I_{peak}^{μ} from the reference point of 50 ADC channels occur. To rectify these drifts and determine I_{peak}^{μ} , an ongoing procedure is implemented. The intuitive approach involving the creation of an amplitude histogram (as the one illustrated in the right panel of Fig. 3.4, black line) to directly determine I_{peak}^{μ} in ADC channels, extends excessively the dead time of the station. Instead, an estimation of I_{peak}^{μ} ($I_{\text{peak}}^{\text{est.}}$) is derived by enforcing that the event rate satisfying a ‘‘calibration trigger’’ reaches 70 Hz. This calibration trigger is defined as a threshold trigger of $2.5I_{\text{peak}}^{\text{est.}}$ for the given PMT and $1.75I_{\text{peak}}^{\text{est.}}$ for all three, with these threshold values derived from the reference station. A convergence algorithm is applied to determine the value of $I_{\text{peak}}^{\text{est.}}$, with the complete algorithm detailed in Ref. [51]. The determination of $I_{\text{peak}}^{\text{est.}}$ is precise to within 6% of I_{peak}^{μ} .

Once $I_{\text{peak}}^{\text{est.}}$ is established, a threshold trigger of $0.1I_{\text{peak}}^{\text{est.}}$ is applied to collect events over a 60-second period, yielding approximately 150 000 events. From these

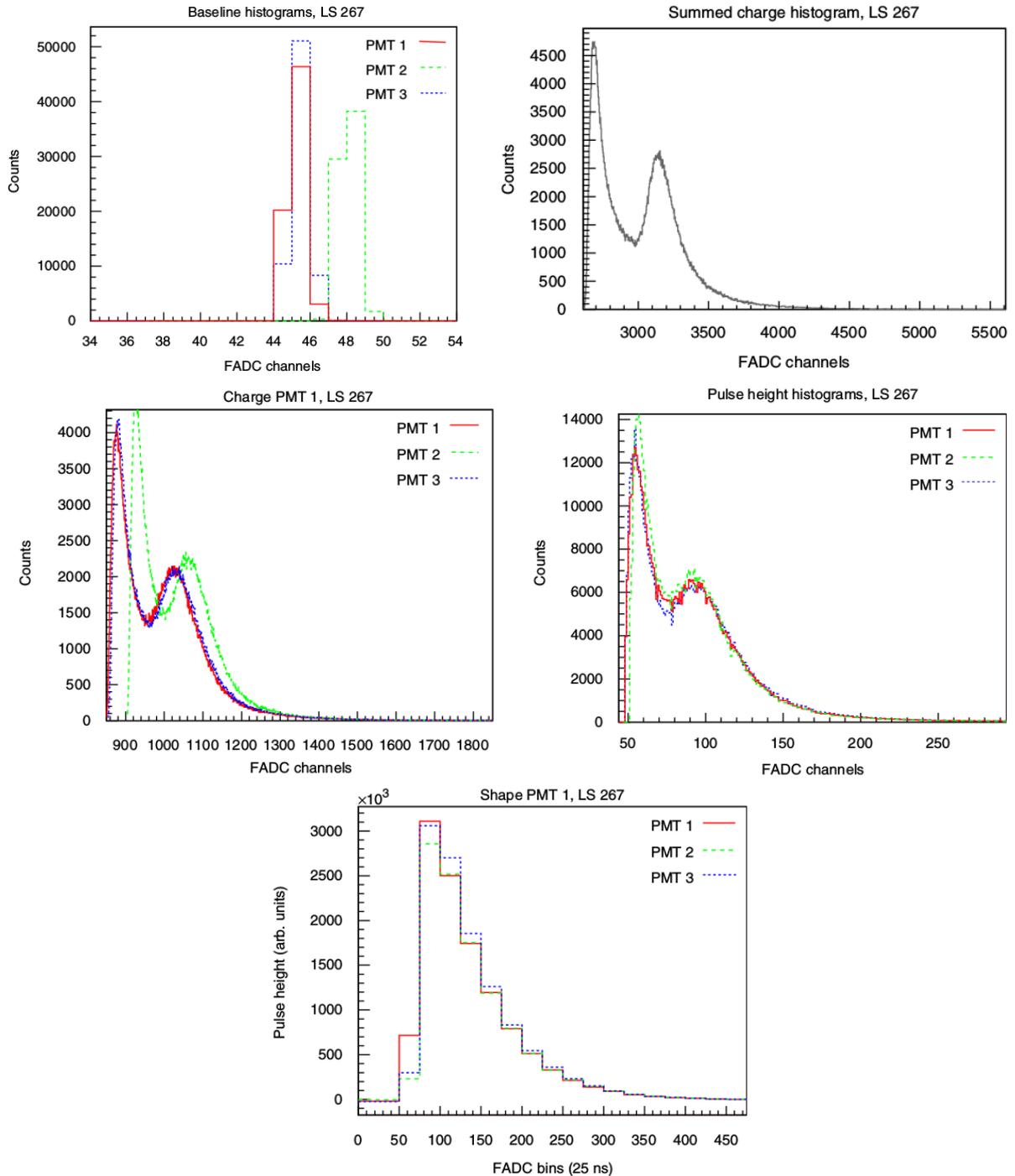


Figure 3.5: Example of calibration histograms sent to CDAS by a WCD. These histograms include baseline values (top left), signal charge (middle left), and amplitude (middle right) for individual PMTs, along with the summed charge (top right) from all PMTs. The bottom panel illustrates the average signal shape for events featuring a charge within the range of $(1.0 \pm 0.1)Q_{\text{VEM}}$. These histograms are constructed based on the signals produced by background atmospheric particles that satisfy the calibration threshold trigger of $0.1I_{\text{peak}}^{\text{est.}}$ over a time window of 60 seconds.

events, several types of histograms relevant to the calibration of the station and its signal post-processing are generated. These include histograms of the individual charge, amplitude and baseline values for each **PMT**, as well as the sum of the signals from all three **PMTs**. Additionally, the average pulse shape of events with a charge within the range of $(1.0 \pm 0.1)Q_{\text{VEM}}$ is computed. An illustrative example of these histograms is provided in Fig. 3.5.

When CDAS requests an event, the corresponding calibration histograms, created within the last minute, are sent as attachments and stored alongside the data of the stations participating in the event. Consequently, all the information necessary for the calibration of the signals during the offline process of event reconstruction (as described in Section 3.3) is available.

3.2.3 SURFACE DETECTOR TRIGGERS

Several stations within the surface array register particles originating from high-energy showers. Trigger conditions are systematically configured to identify such showers, distinguishing them from random coincidences originating from background particles. Here, we provide a concise overview of the different levels of triggers defined for the **SD** and their application in event selection.

Each **WCD** incorporates two levels of local triggers, denoted as T1 and T2. T1 triggers consist of two distinct types. The first type, known as the threshold trigger (T1-TH), demands that all three **PMTs** within the station record a signal with an amplitude surpassing $1.75I_{\text{peak}}^{\mu}$. The second type, the time-over-threshold trigger (T1-ToT), necessitates that, for two out of the three **PMTs**, the signal amplitude exceeds $0.2I_{\text{peak}}^{\mu}$ in at least 13 bins in a sliding window of 3 μs .

The T1-TH trigger is effective at detecting very inclined showers, as their signals tend not to exhibit significant temporal dispersion. Conversely, the T1-ToT trigger tends to favour the selection of vertical showers, particularly those characterised by either low energy and proximity to the core or high energy at a considerable distance from the shower axis. This preference arises because the signals of such showers typically exhibit lower amplitudes and large temporal spreading caused principally by the scattering of electromagnetic particles in the atmosphere.

The second level of local trigger is T2. All T1-ToT triggers are automatically

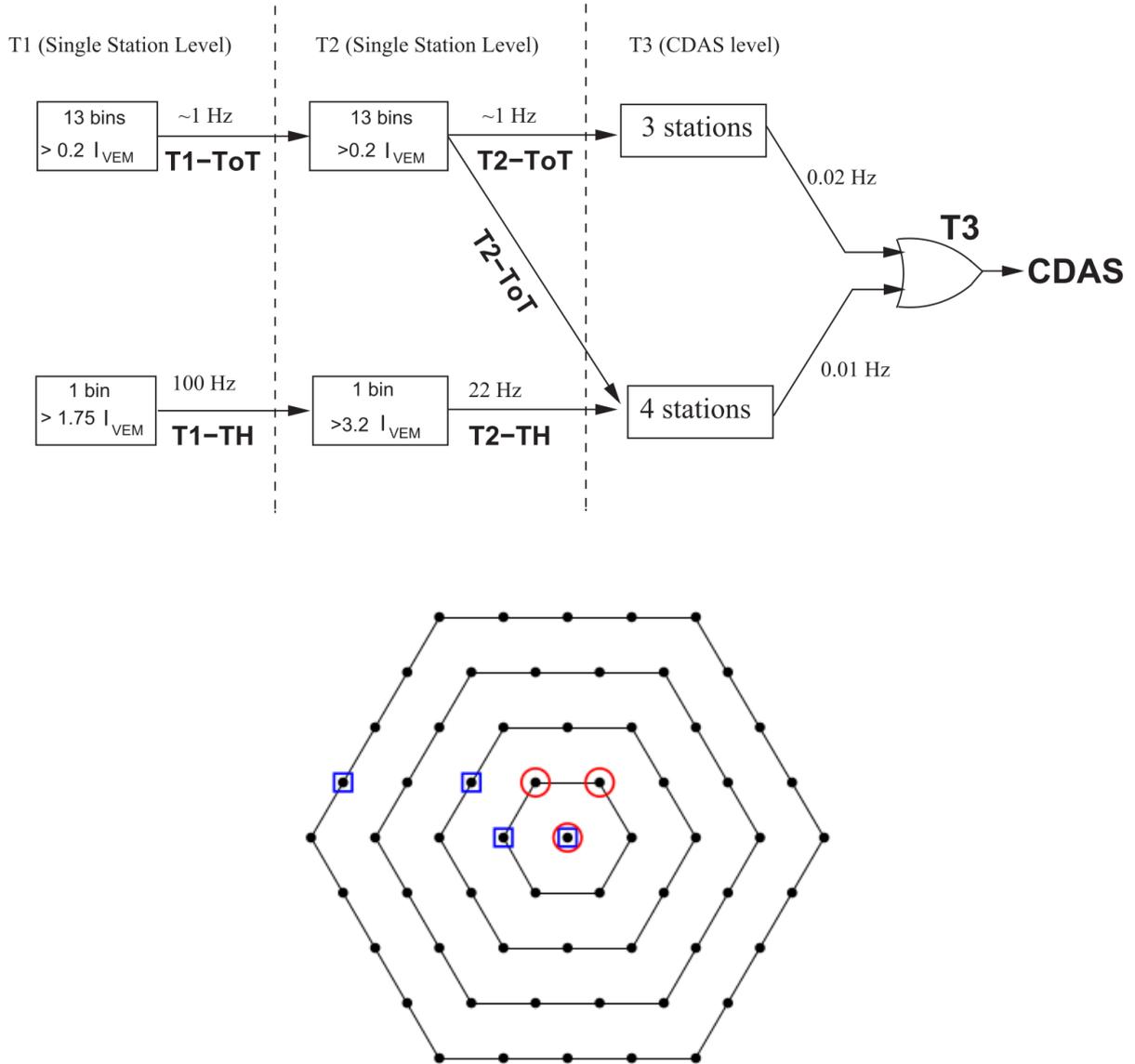


Figure 3.6: Top: Conditions and respective event rates for the different trigger levels (T1, T2, and T3) are outlined. Bottom: A system of concentric hexagons centred on one of the stations triggered at the T2 level. This system is employed during the spatial coincidence analysis. The two spatial modes are depicted by the presence of red circles and blue squares (refer to the text for a detailed explanation). Illustrations taken from Ref. [12].

elevated to T2 status, denoted as T2-ToT. Conversely, for a T1-TH trigger to be promoted to T2, designated as T2-TH, it necessitates that the signal in all three PMTs surpasses $3.2I_{\text{peak}}^{\mu}$. When a station registers a T2 trigger, its timestamp is transmitted to CDAS.

At CDAS, an analysis of the received T2 triggers is executed to identify spatial and temporal coincidences among the surface detectors. This analysis aims at the generation of a level-3 trigger (T3) and the identification of a shower event. Initially, the received T2 triggers are temporally clustered, employing a time interval

of $25 \mu\text{s}$ centred on each T2 trigger. Groups comprised of three or more stations exhibiting clustered T2 triggers are selected for subsequent spatial analysis. For the spatial analysis, a framework of concentric hexagons is utilised, centred on each station within the clustered group, as illustrated in Fig. 3.6 (bottom).

Two modes implement spatial conditions that the clustered group of stations may satisfy to yield a T3 trigger. The first consists of at least three detectors exhibiting a T2-ToT trigger, with one detector positioned within the first hexagon of another and the second detector not farther than the second hexagon. The second mode mandates a coincidence involving four stations with T2 triggers of any type, satisfying the spatial requirement that one station may be as distant as the fourth hexagon if another station falls within the first hexagon and a further one no farther than the second hexagon. Fig. 3.6 (bottom) visually represents an example conforming to the first criterion, marked by red circles, while the blue squares exemplify a scenario aligned with the second mode.

Upon meeting one of these spatial criteria, the T2 triggers must fall within a temporal window of $(6 + 5n) \mu\text{s}$ from the central station, where n denotes the hexagon number, to trigger a T3 event. Once a T3 trigger is validated, CDAS requests all ADC traces within a window of $30 \mu\text{s}$ of the participating stations in the T3 event, which includes stations with only T1 triggers. Considering hadronic primaries, the described configuration of triggers results in full efficiency for the detection of events with energy above $10^{18.5} \text{ eV}$ [53]. In the upper part of Fig. 3.6, a graphical summary outlines the distinct trigger level criteria alongside the corresponding event frequency for each condition met.

3.3 RECONSTRUCTION OF EVENTS WITH THE SD

In this section, we will provide an overview of the fundamental aspects of the standard reconstruction of events detected by the SD. This procedure is applied to the T3 data, as recorded by CDAS, utilising the official software framework of the Pierre Auger Collaboration, known as Offline (further details are given in Section 3.4).

The standard reconstruction is conducted for showers with zenith angle smaller than 60° , which have been detected by the regular array (detector spacing of

1500 m). It is worth noting that the process for showers detected by the 750-m array is rather similar.

The primary objectives of this reconstruction process are summarised below

1. To determine the geometry of the event, thereby enabling the characterisation of the arrival direction of the primary cosmic ray in terms of zenith and azimuth angles.
2. To estimate the energy of the event through its correlation with an observable that reflects the size of the shower.

In the following, the procedures employed to achieve these objectives will be delineated in more detail. Conversely, a comprehensive description is found in Refs. [12, 54].

In the schematic illustration of the detection of an event in Fig. 3.7 (left), a particle front traverses several stations, and the recorded signal times from these stations are subjected to a fitting process using a model that describes the propagation of this particle front. For events involving only a few triggered stations, a plane front model is employed. However, when multiple stations participate in the event, a model considering a spherical front expanding at the speed of light is applied. This model is expressed by the equation:

$$|\vec{x}_i - \vec{x}_{\text{sh}}| = c(t_i - t_0), \quad (3.1)$$

where \vec{x}_{sh} represents the point where the particle cascade initiated in the atmosphere, occurring at time t_0 , whilst \vec{x}_i denotes the position of the i^{th} station struck by the particle front at time t_i . An example of a fit is displayed in Fig. 3.7 (right), where the time difference recorded by the stations to that of a reference plane shower front is plotted against the perpendicular distance from the shower axis. This technique enables the determination of \vec{x}_{sh} , providing an initial approximation of the arrival direction of the primary particle.

In the subsequent stage of the reconstruction, the integrated signals from the stations participating in the event are fit through a maximum likelihood method [54]. The fit function describes the dependency of the particle density within the extensive air shower with respect to its perpendicular distance from the shower core and is referred to as the *lateral distribution function* (LDF). In Fig. 3.8, the calibrated signals for an event are plotted alongside the LDF fit, with the signals expressed in VEM units.

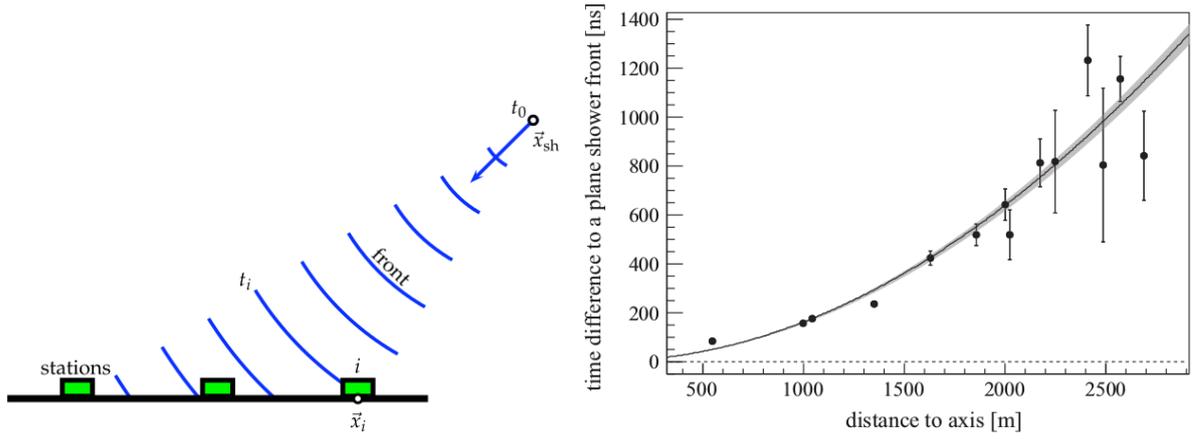


Figure 3.7: Left: Illustration depicting the propagation of a spherical shower front. The shower originates at (\vec{x}_{sh}, t_0) and intersects station i at (\vec{x}_i, t_i) . Right: Fitting process applied to the recorded times of the stations using a model based on spherical front propagation. The plot illustrates the difference in trigger times between stations and a reference plane front, plotted as a function of the perpendicular distance from the shower axis. Taken from Ref. [12].

Regarding the functional form of the **LDF**, a modified Nishimura-Kamata-Greisen (**NKG**) function is employed [55], expressed as:

$$S(r) = S(r_{opt}) \left(\frac{r}{r_{opt}} \right)^\beta \left(\frac{r + r_1}{r_{opt} + r_1} \right)^{\beta+\gamma}, \quad (3.2)$$

where $r_1 = 700$ m, and r_{opt} represents the optimal distance at which variations in the function arising from the choice of the **LDF** shape are minimised. It primarily depends on the geometry of the detector, and for the 1500 m array, $r_{opt} = 1000$ m [56]. $S(r_{opt})$ denotes the signal at this optimum distance. This is the chosen observable to characterise the shower size due to its minimal dependence on the **LDF** used for the fitting procedure. Since $r_{opt} = 1000$ m, $S(r_{opt})$ is equivalent to $S(1000)$, representing the estimated signal at 1000 m from the shower core, as indicated in Fig. 3.8.

In Eq. (3.2), the parameter β varies with the zenith angle, as inclined events reach ground level at a later shower development than vertical ones. At large distances from the shower core, the parameter γ accounts for deviations of the **LDF** from a simple power-law function. Both β and γ are parameterised as a function of the zenith angle and $S(1000)$, employing a lever-arm criterion [54].

From this fitting process, the impact point of the shower on the ground, \vec{x}_{gr} (where the shower core intersects the detector level), is determined. Using the position

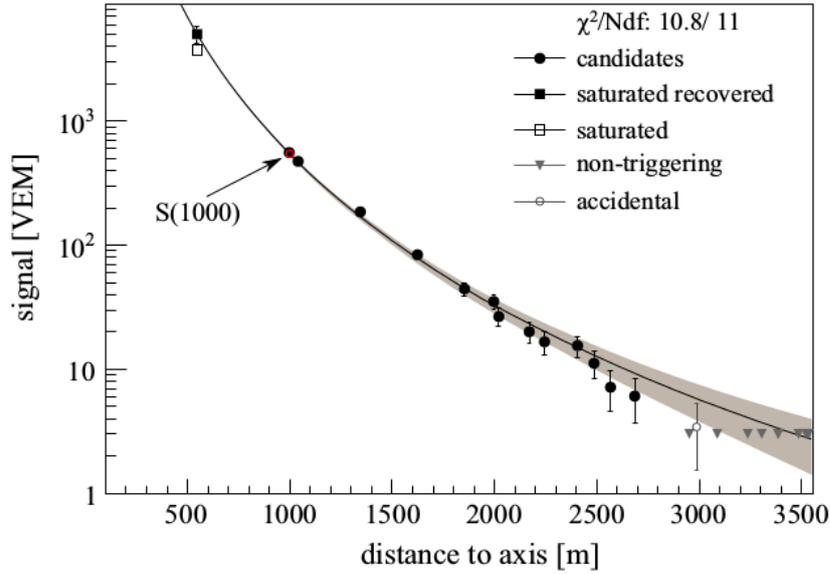


Figure 3.8: For a representative event, the signals recorded by participating stations are displayed as a function of their respective distances from the shower axis. This radial dependence is fitted to a lateral distribution function (LDF), as defined by Eq. (3.2). The LDF provides the estimation of the signal at a fixed distance of 1000 m from the shower core, denoted as $S(1000)$, which serves as the chosen observable for quantifying the shower size. Subsequently, $S(1000)$ is employed in the energy estimation of the primary cosmic ray. Extracted from Ref. [12].

of the shower origin, \vec{x}_{sh} (refer to Fig. 3.7) obtained from the time fit, the arrival direction of the primary cosmic ray can be estimated using:

$$\hat{a} = \frac{\vec{x}_{\text{sh}} - \vec{x}_{\text{gr}}}{|\vec{x}_{\text{sh}} - \vec{x}_{\text{gr}}|}. \quad (3.3)$$

The accuracy in the determination of the arrival direction improves with increasing zenith angles and is dependent on the number of stations triggered in the event: a greater number of stations leads to better resolution. Above full trigger efficiency, the resolution improves from 1.4° for nearly vertical showers to approximately 0.8° for inclinations of 60° [54].

The determination of the energy for an event detected exclusively by the surface detector array relies on a data-driven calibration process derived from hybrid events. Consequently, we will briefly describe the energy reconstruction procedure using data from the fluorescence detector.

As an extensive air shower develops in the atmosphere, it generates fluorescence light emissions from atmospheric nitrogen molecules alongside Cherenkov

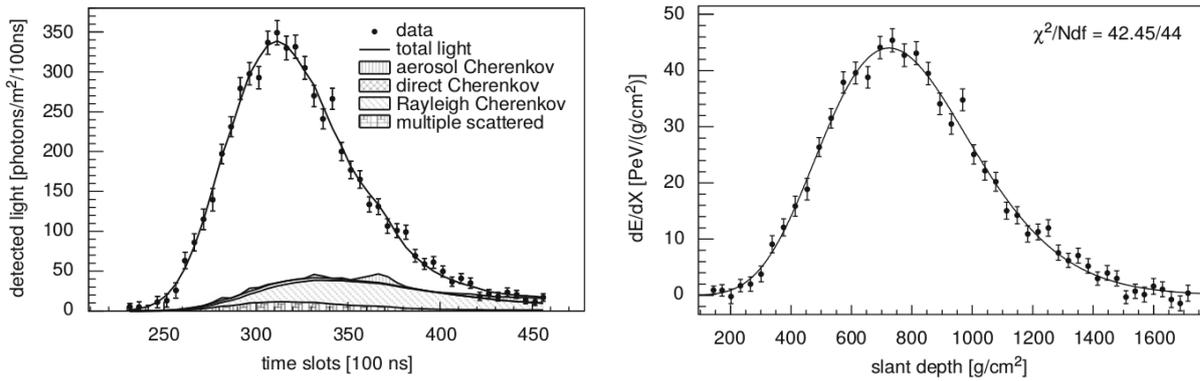


Figure 3.9: Left: Light flux measured by a fluorescence telescope as a function of time for an example event, with estimated contributions from different components of the total detected light. Right: Corresponding energy deposit profile obtained by converting light flux and time to energy and slant depth, respectively. The integration of this profile yields the energy estimation of the event after correction for energy carried away by neutrinos and high-energy muons. Taken from Ref. [14].

radiation. These emissions are detected by the fluorescence telescopes, registering signals in various pixels of the telescope camera as the shower progresses. By analysing the timing of these signals in the camera pixels, a precise determination of the position of the shower axis is achieved. Combining this information with timing data from at least one surface detector station significantly enhances the accuracy of the arrival direction estimation, yielding a typical resolution of 0.6° [14].

Fig. 3.9 (left) illustrates the amount of light collected by a telescope aperture over time for an example event, allowing for the observation of the longitudinal profile of the particle cascade. To convert light flux into energy deposition in the atmosphere, factors like the light attenuation during its path from the shower to the telescope and the contributions of various light components (fluorescence, direct and indirect Cherenkov radiation, and multiple-scattered light) must be considered. These components are visible in Fig. 3.9 (left). Subsequently, Fig. 3.9 (right) displays the energy deposition as a function of atmospheric slant depth. A Gaisser-Hillas function [57] is fitted to the energy deposit profile and integrated to obtain the total energy. A correction derived from Monte Carlo simulations is applied to account for the “invisible energy” carried by neutrinos and high-energy muons. The energy resolution due to statistical uncertainties stands at 10%, whilst the systematic uncertainties cumulatively reach 22% [14].

As previously mentioned, the shower size is quantified by $S(1000)$, repre-

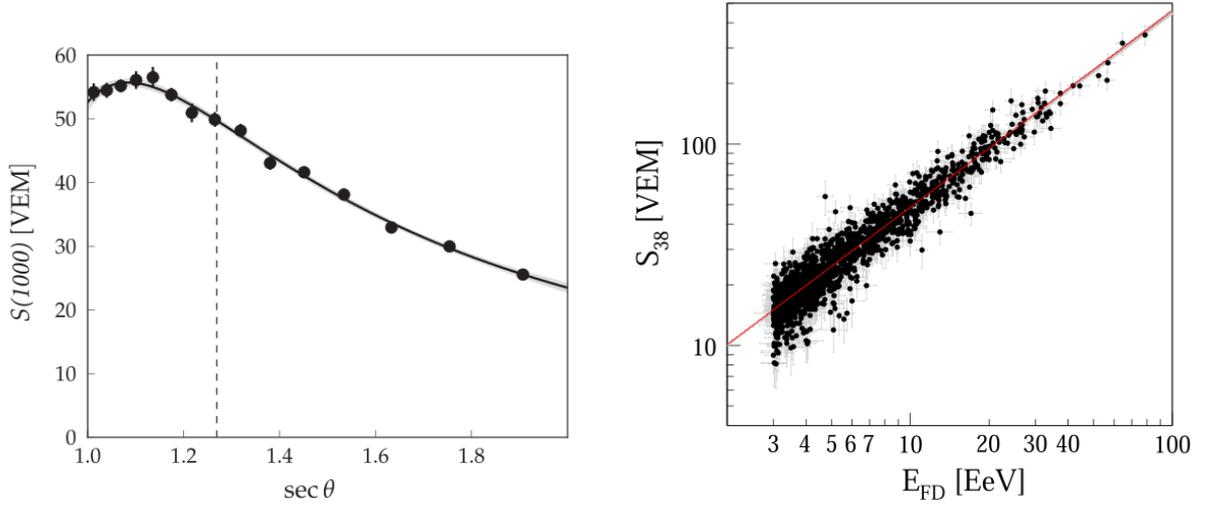


Figure 3.10: Left: Dependence of $S(1000)$ on zenith angle θ . A third-degree polynomial in $x = \cos^2 \theta - \cos^2 \bar{\theta}$ is employed to model the attenuation (see Eq. (3.4)). The chosen reference angle $\bar{\theta} = 38^\circ$ represents the median within the isotropic distribution of events (dashed line). Right: Correlation between the observable S_{38} and the energy measured by the fluorescence detector. A single-power law (Eq. (3.6)) is fitted to the data. This relationship is used to estimate the energy of events exclusively detected by the surface detector. Both plots extracted from Ref. [12].

senting the signal at a distance of 1000 m from the shower core, determined through the **LDF** fit. For a given primary energy, $S(1000)$ diminishes as the zenith angle increases, primarily because inclined events take longer to reach the ground during the shower development compared to vertical events. Assuming an isotropic distribution of incident cosmic rays, the dependency of $S(1000)$ with respect to the zenith angle θ is derived from experimental data employing the *constant intensity cut* (**CIC**) method [58]. The attenuation curve is fitted with a third-degree polynomial in $x = \cos^2 \theta - \cos^2 \bar{\theta}$, represented as:

$$f(\theta) = 1 + ax + bx^2 + cx^3, \quad (3.4)$$

where the coefficients were determined as $a = 0.980 \pm 0.004$, $b = -1.68 \pm 0.01$, and $c = -1.30 \pm 0.45$ [12]. In Fig. 3.10 (left), the relationship between $S(1000)$ and $\sec \theta$ is illustrated, along with the fitted curve. Here, the reference angle $\bar{\theta} = 38^\circ$ represents the median of the isotropic distribution considering the planar geometry of the Observatory. Next, Eq. (3.4) is employed to convert $S(1000)$ into S_{38} , defined as:

$$S_{38} \equiv \frac{S(1000)}{f(\theta)}. \quad (3.5)$$

Therefore, S_{38} can be interpreted as the $S(1000)$ value the shower would possess if it arrived with an inclination of 38° .

Next, the observable S_{38} can be effectively correlated with the shower energy, exploiting the hybrid design of the Observatory. This design allows events detected concurrently by both the surface and fluorescence detectors to serve as calibration for the surface detector array. Consequently, a reliance on Monte Carlo simulations is not necessary to determine the energy of events solely detected by the surface stations.

The determination of the energy calibration utilises hybrid events wherein all six nearest neighbours to the station with the highest signal are operational at the time of the event [53], forming a complete hexagon. In Fig. 3.10 (right), a plot illustrates the correlation between S_{38} and the energy obtained from the fluorescence detector, E_{FD} . The relationship between these two quantities can be described by a single-power law function:

$$E_{\text{FD}} = A(S_{38}/\text{VEM})^B \quad (3.6)$$

where the coefficients A and B , obtained from a maximum likelihood fit, were found to be $(1.90 \pm 0.05) \times 10^{17}$ eV and 1.025 ± 0.007 , respectively [59].

By combining Eqs. (3.4) to (3.6), the energy of an event detected by the surface detectors can be estimated as follows:

$$E_{\text{SD}} = A \left(\frac{S(1000)}{1 + ax + bx^2 + cx^3} \right)^B, \quad (3.7)$$

where $x = \cos^2 \theta - \cos^2(38^\circ)$ and $S(1000)$ is in **VEM** units. The energy resolution from the surface detector stands at 16% for low energies and 12% for high energies [12]. This effect is visually evident in Fig. 3.10 (right), where the distribution of S_{38} becomes narrower as energy increases.

In summary, when a high-energy shower is detected, the surface detector records signals and their associated arrival times as the particle front traverses these stations. By fitting the signal times to a model describing the propagation of the shower particle front, and by modelling the radial dependency of the integrated signals with a lateral distribution function, we can accurately determine the arrival direction of the primary particle.

For the energy estimation, we convert the estimated signal at 1000 m from

the shower axis, obtained from the **LDF** fit, into the zenith-independent observable S_{38} . This quantity represents the $S(1000)$ signal the shower would exhibit if it arrived with a zenith angle of 38° . Utilising the established correlation between S_{38} and the energy measured by the fluorescence detector, derived from hybrid events, we can determine the energy of primary cosmic rays exclusively detected by the surface detector. This approach eliminates the need for Monte Carlo simulations, enhancing the reliability of the energy estimation process.

3.4 THE OFFLINE FRAMEWORK

With the primary goal of analysing and interpreting the vast dataset collected by the Pierre Auger Observatory, the Auger Collaboration has developed a versatile framework known as Offline. This framework serves as the backbone for processing and analysing the data, incorporating numerous algorithms contributed by various collaborators. It is designed with a strong emphasis on flexibility, enabling users to effortlessly test and implement new code. This inherent adaptability empowers researchers to create customised applications efficiently, tailored to their specific needs. Moreover, Offline is continuously refined and can be readily expanded to accommodate detector upgrades within the Observatory. It possesses the capability to handle diverse data formats from various sources. Throughout our research, we have extensively utilised and made contributions to Offline. Consequently, in this section, we provide a description of this framework (further information can be found in Refs. [12, 60]).

The fundamental structure of the Offline framework is centred around the following key components: modules, the run controller, the event, and detector description interfaces. An illustration depicting the presence of these elements in an Offline application is provided in Fig. 3.11 (left). In the context of this framework, a *module* represents a discrete block of code implemented as a C++ class, each designed to execute specific tasks. For instance, the fitting of signals from the **SD** stations to an **LDF** is achieved using the `LDFFinder` module. These modules are organised sequentially to accomplish more comprehensive objectives, such as complete event reconstructions, as described in the preceding section.

The sequencing of modules is handled by the *run controller*, which is im-

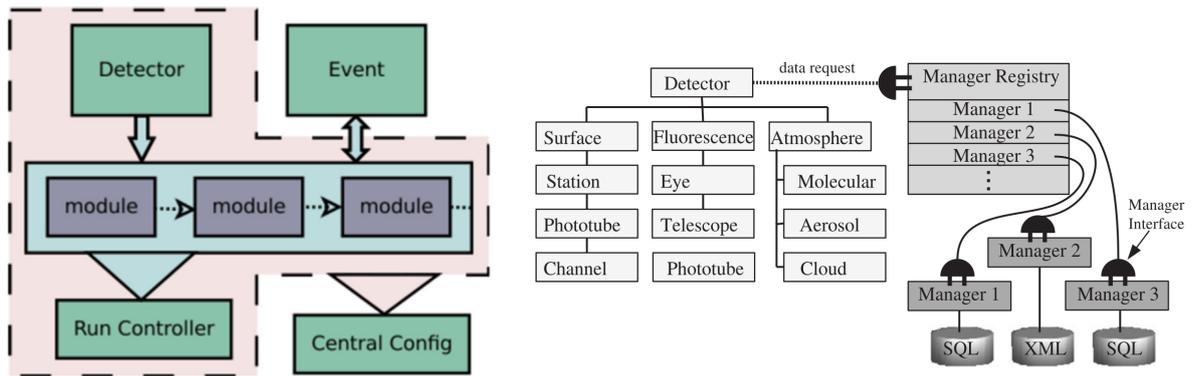


Figure 3.11: Left: A general representation of an Offline application, with modules for specific tasks communicating via the event interface and accessing detector-related data from the detector interface. Module sequencing is handled by the run controller, and module configuration is provided by the central configuration. Right: A schematic depiction of the detector description interface featuring a hierarchical structure that acquires data from various sources via specialised managers. Both illustrations extracted from Ref. [12].

plemented through an **XML** file. The run controller not only dictates the sequence of modules but also enables the possibility of looping through these sequences. Depending on different conditions, modules can issue instructions to the run controller, directing it to conclude its execution, proceed to the next iteration, or halt the application in the event of errors. This feature proves advantageous when iterating through datasets containing numerous triggered events, allowing for the reconstruction of one event in each iteration.

Aligned with the principles of customisation and flexibility, the behaviour of modules within the framework can be finely tuned using *configuration files* written in **XML** format. These files define specific options, constant values, parameter ranges, and other relevant settings, which are then employed by their corresponding modules. The framework includes a *central configuration* that references a set of default configuration files for its existing modules.

Within an Offline application, users define the configuration to be utilised in a *bootstrap file*. This bootstrap file specifies the configuration options for the application. Whilst it is possible to make specific alterations to default configurations directly within the bootstrap file, users also have the option to provide custom configuration files for individual modules, with these custom files being declared in the bootstrap. Furthermore, to maintain data integrity and robustness, the framework employs the

W3C XML Schema standard to validate the **XML** configuration files, facilitating the early detection of configuration errors.

The *event* interface serves as a structured container for various event-related data. Its hierarchical organisation is predefined and aligns with the layout of the detectors of the Observatory. This interface is capable of accommodating both real observational data and data generated through Monte Carlo simulations. The event interface plays an essential role within the framework, enabling communication among the various modules of an application. These modules can read data from and write data to the event interface, facilitating the exchange of information. Whilst the fixed nature of this interface does impose certain limitations, this design choice significantly enhances the interchangeability of modules, which is preferred for data analysis. Furthermore, the data stored within the event interface can be made persistent by saving it to a file. This functionality is achieved through the use of the ROOT toolkit [61], which provides the necessary tools for data serialisation.

Within the framework, the other crucial interface is known as the *detector description*. This interface is dedicated to housing information pertinent to the detectors themselves, particularly data that remains static or changes at an exceptionally slow rate over time. The detector interface also adheres to a structured hierarchy mirroring the layout of the Observatory's detectors. Given that such data is often dispersed across distinct sources, queries to this interface are delegated to specialised entities known as *managers*. Each manager is tasked with retrieving data from a specific source. Static detector information, which seldom undergoes modifications, is typically stored in **XML** files. In contrast, data that experiences gradual changes over time, such as monitoring and calibration data, is typically stored in MySQL databases. An illustration of the elements comprising the detector interface and their interactions is depicted in the right panel of Fig. 3.11.

In addition to its inherent functionalities, the Offline framework offers a set of *Standard Applications* thoroughly crafted by the Auger Collaboration. These applications serve as dedicated tools for the official data reconstruction across the different detectors within the Observatory. This reconstructed data is utilised in high-level physics analyses, culminating in the published research findings by the Collaboration. Furthermore, these standard applications encompass programs designed to simulate the

detector responses to Monte Carlo data generated by software like CORSIKA, AIRES, SENECA, and CONEX. To summarise, the Offline framework provides a robust and versatile toolkit indispensable for the research endeavours of the Auger Collaboration.

CHAPTER 4

AugerPrime: upgrade of the Observatory

Since its inception in 2004, the Pierre Auger Observatory has been continuously collecting invaluable data concerning ultra-high energy cosmic rays (UHECRs). The analyses of this data have greatly advanced our comprehension of the most energetic particles in the Universe. A suppression of the all-particle flux beyond energies of 40 EeV [15] coupled with anisotropy findings [62] and the surprising tendency of heavier composition for increasing energy [18, 22] mark important breakthroughs in the field of UHECRs, obtained by the Pierre Auger Collaboration. Conversely, these remarkable findings result in new mysteries that should be resolved to yield a complete understanding of the astrophysical scenarios pertaining to the nature of UHECRs.

To address these open questions, a major upgrade of the Pierre Auger Observatory, dubbed AugerPrime [24], is currently nearing its completion. The new phase of data taking, known as Phase-II operation, will begin in 2024. Since the upgrade of the Observatory is the central focus of our research project, in this chapter, we will discuss in more detail the motivation for its implementation as well as the means to achieve the targeted scientific objectives. Subsequently, the design of AugerPrime and its new detectors are delineated in more depth.

4.1 MOTIVATION FOR THE UPGRADE

In Section 2.4, we have discussed the differential spectrum of cosmic rays as well as their composition for energies surpassing 1 EeV. One of the most important open questions is determining the origin of the observed flux suppression of cosmic rays above 40 EeV. As also exposed in Section 2.4, candidate models include maximum-rigidity [45] and photodisintegration scenarios [16, 17]. The first attributes the suppression to the maximum output of the sources, accelerating particles proportionally to their charge. On the other hand, the photodisintegration model assumes that sources inject particles with an energy larger than the one of the observed flux suppression, which is attributed to the interaction of these particles with photons from the cosmic microwave background (CMB).

These models predict specific fractions of element groups at different energies, leading to the suppression point. However, neither model is able to provide a satisfactory description of the composition data measured by the Pierre Auger Observatory. As these measurements are derived from data from the fluorescence detector, its low duty cycle of approximately 15% does not yield the necessary composition statistics at the energy of the flux suppression to elucidate this question.

An additional crucial point of investigation is the identification of astrophysical sources of UHECRs. Although important findings were derived from anisotropy analyses [62], pinpointing specific sources poses as a challenging task [63, 64], considering the current experimental configuration. The main difficulties are attributed to insufficient knowledge about intergalactic magnetic fields as well as the composition of the UHECRs.

Since the deflection of these particles is proportional to their charge (tightly related to the composition), light primaries, such as protons, are expected to present minor deflection during their propagation, demonstrating the potential to directly determine their source location. Furthermore, the search for composition-based anisotropies offers an enhanced capability of identifying sources of UHECRs.

The quantity of muons generated within a particle cascade is intricately linked to the underlying hadronic interactions occurring during the development of the shower. Analysing inclined events, it becomes feasible to quantify the number of

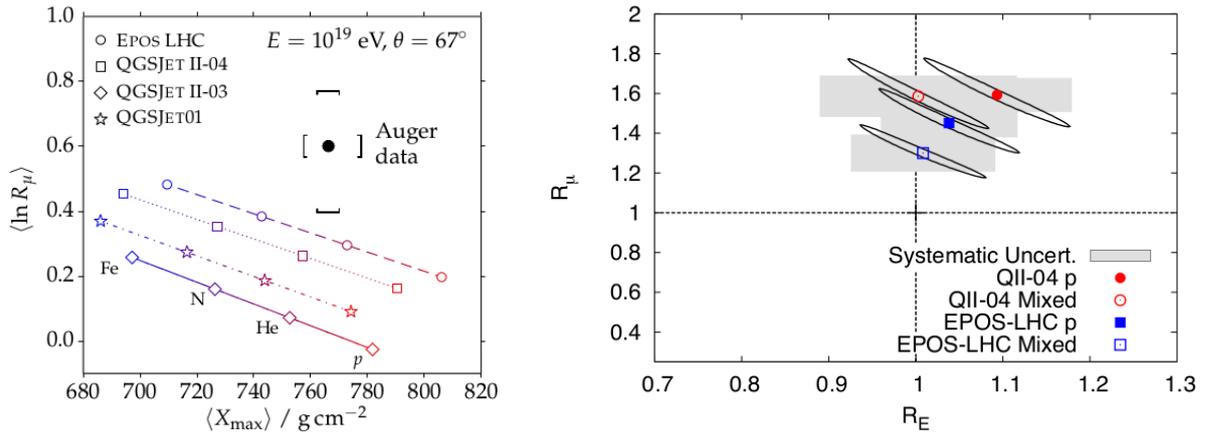


Figure 4.1: *Left*: The quantity of muons, normalised to that of a proton-induced shower with an energy of 10 EeV, presented as a function of the average atmospheric depth of the shower maximum X_{\max} . The results corresponding to different hadronic interaction models are also depicted. *Right*: Scaling factors denoted as R_μ and R_E , necessary for aligning simulations with the muon count and energy observed by the Pierre Auger Observatory. The current simulation models are not in agreement with the available data, which is referred to as the muon puzzle. Taken from Ref. [24].

muons utilising the surface detector of the Pierre Auger Observatory indirectly. This is made possible because, in such cases, the electromagnetic component of the shower is nearly completely absorbed by the atmosphere, resulting in a predominant muonic signal in the water-Cherenkov detectors.

Fig. 4.1 (left) depicts a diagram of the mean number of muons¹ and the corresponding maximum development of the particle cascade in the atmosphere. The results from simulations employing different models of hadronic interactions are displayed. It becomes evident that none of the simulation models concurs with the measured data. This conclusion is endorsed by the plot in Fig. 4.1 (right), where the scaling factors R_μ and R_E , necessary to accurately represent the number of muons and corresponding energy as measured by the Pierre Auger Observatory, are displayed for the different hadronic models. This disparity is known as the *muon puzzle* [65, 66].

At this point, we emphasise that these simulation models extrapolate the current data from particle accelerators to energies far beyond what is attainable in the laboratory. The possibility of new hadronic physics cannot be disregarded. Nevertheless, the exposed results strongly indicate that the current comprehension of hadronic interactions is incomplete. Furthermore, probing extensive air showers initiated by

¹Relative to the mean number of muons generated in showers initiated by protons featuring an energy of 10 EeV.

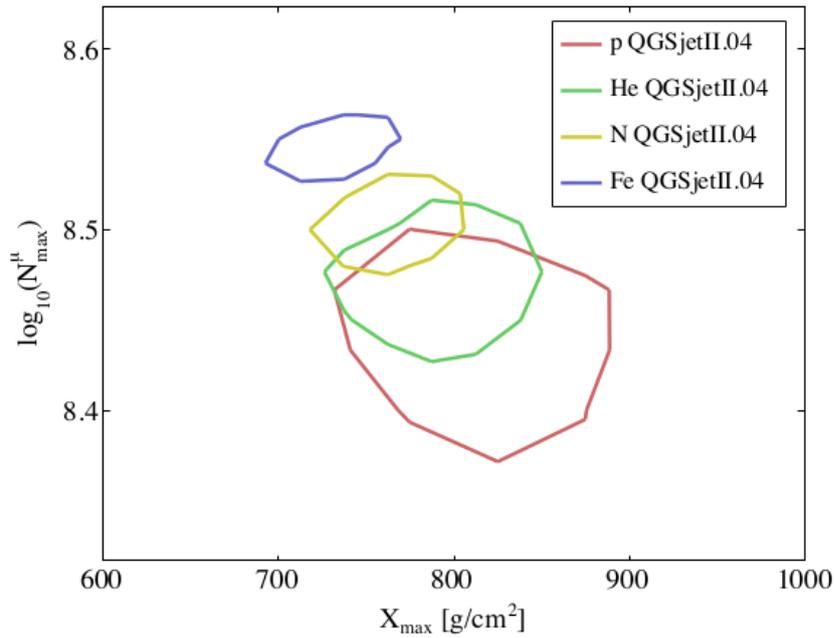


Figure 4.2: The contours shown represent the 1σ region of muon counts as a function of X_{\max} , derived from simulations involving various primary particles with an energy of 50 EeV and an inclination of 38° . These simulations employed the QGSJetII.04 hadronic interaction model. The distinctive separation between light and heavy primaries is evident, underscoring the capability of the muonic component of extensive air showers to discern the composition of the primary cosmic ray. Extracted from Ref. [24].

UHECRs is currently the only instrument to study fundamental particle interactions in such extreme energies.

The above discussion illustrates that data sensitive to the composition of **UHECRs** is the key to advancing our understanding of the nature of these extreme particles. To achieve these scientific goals, the Pierre Auger Collaboration has designed a substantial upgrade of its Observatory. Since the fluorescence detector exhibits an intrinsically low duty cycle, efforts have been directed towards the surface detector to leverage its nearly continuous operation.

Fig. 4.2 illustrates a diagram of the number of muons at maximum development of the shower versus the corresponding atmospheric depth of the maximum, where the lines denote the 1σ contour for distinct primary cosmic rays. The separation between light and heavy primaries is evident, demonstrating the correlation between composition and the muonic component of the induced shower.

For the upgrade of the Pierre Auger Observatory, it was decided to employ a new surface detector capable of separating the muonic and electromagnetic com-

ponents of extensive air showers, thereby attaining sensitivity to the composition of the primary cosmic rays. The advantages of such component separation extend to investigations of hadronic interactions, offering valuable insights into the muon puzzle. Furthermore, the implementation of new triggers in the upgraded detectors holds the potential to place stricter constraints on the existence of neutral particles. Additionally, these enhancements open the door to exploring physics beyond the standard model, including the examination of phenomena such as dark matter decay [67] and potential Lorentz invariance violations [2, 68].

4.2 DESCRIPTION OF THE AUGERPRIME DETECTORS

To achieve the separation of the muonic and electromagnetic contents of extensive air showers, the main component of the upgrade consists in the installation of scintillator plates atop each of the existing water-Cherenkov detectors (WCD). These plates are referred to as Scintillator Surface Detectors (SSD). Leveraging the different responses of the SSD and WCD to muons and electromagnetic particles, the development of methods to reconstruct the corresponding signals of these components in the detectors is possible. The investigation of one of such methods is a crucial part of our research work, which is detailed in Chapter 8.

Moreover, the stations of the surface detector (SD) will be equipped with a radio antenna to directly measure the electromagnetic component of events presenting an inclination larger than 65° . This pioneering technique constitutes the Radio Detector (RD) of the AugerPrime Observatory [69].

The WCDs will receive an additional photomultiplier tube (PMT) featuring a small area. This PMT, dubbed small PMT (SPMT), serves the main purpose of extending the dynamic range of the WCDs, providing unsaturated signals when the electronic readout of their conventional counterparts saturates.

With the addition of these enhanced detectors, it became imperative to upgrade the electronics of the AugerPrime stations. The Upgraded Unified Board (UUB) digitises the signals of both SSD and WCD and manages the local triggers, the calibration of these detectors and their monitoring data. An illustration of an AugerPrime station and its main components is presented in Fig. 4.3.

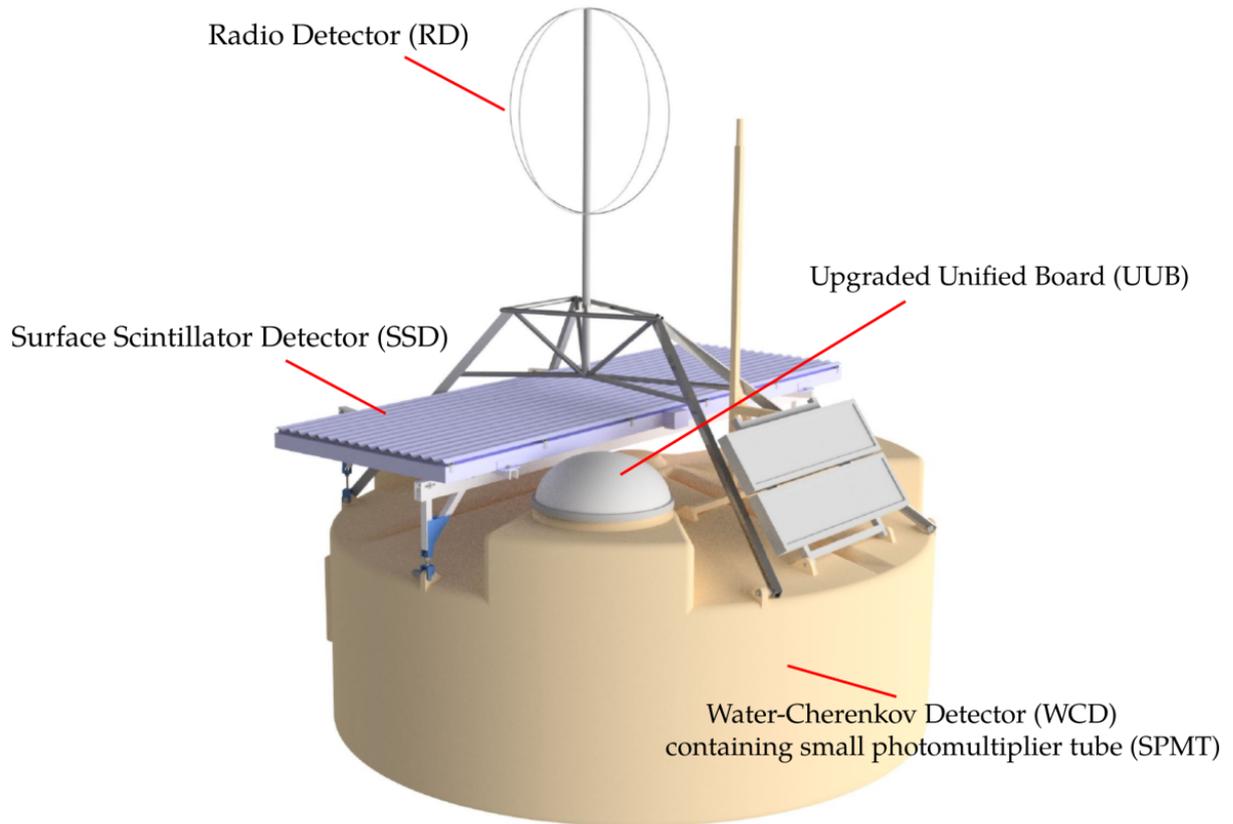


Figure 4.3: Schematic representation of an AugerPrime station and its elements. The main component of the upgrade is the installation of the Scintillator Surface Detector (SSD) on top of the water-Cherenkov detector (WCD). An additional photomultiplier tube with small area (SPMT) extends the dynamic range of the WCD. The Radio Detector (RD) measures the electromagnetic component of showers in inclined events. The Upgraded Unified Board (UUB) digitises the signals of the WCD and SSD, manages calibration and local station triggers, and serves as an interface to the new detectors.

The upgrade of the Observatory additionally introduces the installation of Underground Muon Detectors (UMD) next to the SD stations of the 750-m array. The shielding provided by the soil absorbs most of the electromagnetic component of extensive air showers, enabling the UMD to estimate the muonic content directly. This detector serves as a critical cross-check of the muonic estimations yielded by different methods employing the combination of the SSD and WCD.

Furthermore, the duty cycle of the fluorescence detector will be increased by reducing the high voltage of the camera PMTs. In the following subsections, we will provide further details about the different elements of the AugerPrime Observatory.

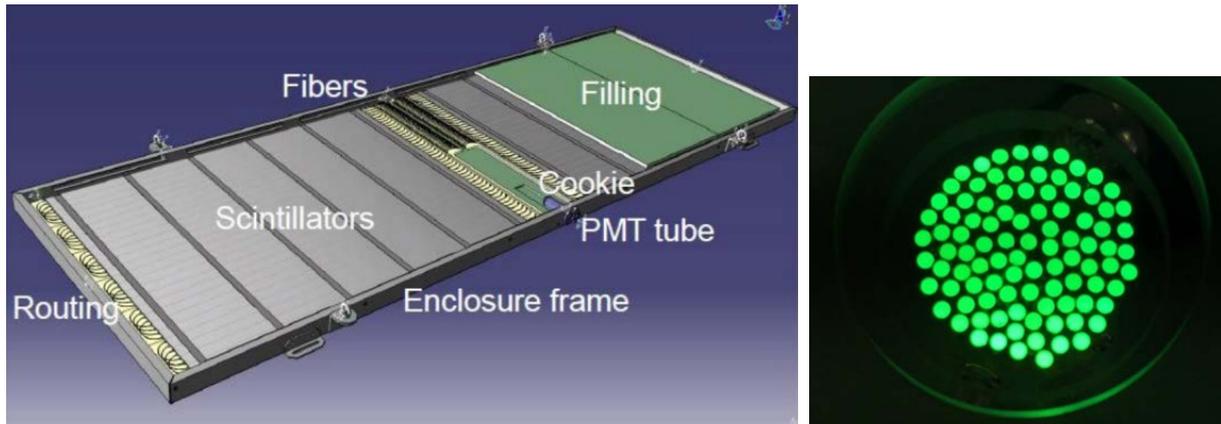


Figure 4.4: *Left*: A Scintillator Surface Detector (SSD) module and its components. The unit is composed of two scintillator panels with area of 1.9 m^2 , each formed by 24 bars containing two channels hosting the wavelength shifting fibres which transport the generated photons to a photomultiplier tube located between the scintillator panels. *Right*: Picture of the “cookie”, consisting of the 96 fibres of the scintillator panels glued together for optical coupling with the SSD PMT.

4.2.1 SCINTILLATOR SURFACE DETECTOR

A unit of the surface scintillator detector, which is mounted on top of each water-Cherenkov detector, consists of two plastic scintillator panels, collectively presenting a surface area of 3.8 m^2 . The light produced in the scintillators is guided through optical fibres to a photomultiplier tube located between the two panels. Fig. 4.4 (left) depicts an SSD unit and its components.

The equipment is housed inside a light-tight enclosure made of aluminium with dimensions $380 \times 128 \times 10 \text{ cm}^3$. The PMT is accessed independently from the scintillator panels for the purpose of simplifying maintenance work in the field. The remaining space inside the module is filled with styrofoam, and an aluminium sunroof is added during deployment to reduce temperature variations.

The scintillator panels are individually composed of 24 scintillator bars, each with dimensions $160 \times 5 \times 1 \text{ cm}^3$. Each bar is made of STYRON 663-W polystyrene, doped with 0.03% POPOP and 1% of PPO. The bars are extruded with a reflective layer of TiO_2 featuring 0.25 mm of thickness. Two parallel channels run along the length of the bar to house the optical fibres. The scintillator panels are manufactured in the Fermi National Accelerator Laboratory (FNAL) in the United States.

The optical fibres are of a wavelength-shifting kind, model Kuraray Y11(300)M S-type. Each fibre is arranged in a “U” shape, running through the channels of adja-

cent bars separated by a distance of 10 cm. The fibres are directed through styrofoam routers to the so-called 'cookie' (see Fig. 4.4, right), which consists of the optical coupling between the 96 fibre ends and the bi-alkali **PMT**, model Hamamatsu R9420 with 1.5 inches of diameter and more than 18% of quantum efficiency at a wavelength of 500 nm.

To ensure the proper working of the **SSD** modules, they were all tested in parallel using the facilities of participating institutions in Germany, France, Italy, Poland and the Netherlands. Once the quality of the modules was established, all 1519 units were shipped to the Observatory site in Argentina. It is important to note that the stations at the border of the array are not equipped with **SSDs**.

4.2.2 SMALL PHOTOMULTIPLIER TUBE

Even with the separation of the signal of the **WCD** photomultipliers into the high and low gain channels, stations that are in proximity to the shower core often present saturation of the read-out electronics due to the large density of particles in these regions. This issue results in a loss of information about the event, especially when fitting a lateral distribution function.

To overcome this limitation, the upgrade of the Observatory includes the installation of an additional photomultiplier tube (**PMT**), model Hamamatsu R8619, in each **WCD**. The new **PMT** has a diameter of one inch, considerably smaller than the regular three **PMTs** of the **WCD**, all with nine inches of diameter. For this reason, the first is referred to as Small **PMT** (**SPMT**), whilst the latter is called the Large **PMTs** (**LPMTs**) of the upgraded **WCD**.

The smaller area of the **SPMT** allows it to record drastically reduced signals when compared to the **LPMTs**. When the latter saturates, typically around 1000 **VEM**, the **SPMT** is capable of providing unsaturated signals up to more than one order of magnitude larger. However, due to the reduced dimensions of the **SPMTs**, it is not possible to calibrate them utilising atmospheric muons (as employed for the **LPMTs**). Therefore, an inter-calibration method is applied using signals of larger showers when both **LPMTs** and the **SPMT** yield clear and unsaturated signals.

The implementation of the **SPMT** considerably extends the dynamic range of the **WCDs**, providing measurements of particle densities much closer to the shower

core. Besides enabling an enhanced reconstruction of high-energy events, the additional data yields information about the most energetic hadronic interactions within the development of the particle cascade.

4.2.3 RADIO DETECTOR

The radio detector (**RD**) consists of two aluminium rings, forming a Short Aperiodic Loaded Loop Antenna (**SALLA**), mounted on a mast made of fibreglass. The antenna is fixed to the **WCD** through an aluminium frame, which does not touch the **SSD**. One of the rings is aligned with the Earth's magnetic field in the **WCD** location, whereas the other ring is arranged perpendicular to the former. The diameter of the rings is 122 mm, aiming at the target frequencies in the interval between 30 and 80 MHz.

The antenna is connected to the readout electronics using a coaxial cable. Signals of the two channels are digitised by an analog-to-digital converter (**ADC**) at a sampling rate of 250 MHz and a dynamic range of 12 bits. When a trigger in the **WCD** occurs, the signal from the **RD** is collected, employing a field-programmable gate array (**FPGA**) to communicate between the antenna **ADC** and the Upgraded Unified Board (refer to Section 4.2.5).

The new detectors, especially the **RD**, considerably increased the power demand of an AugerPrime station. Therefore, the formerly employed solar panels of 55 Wp (Watt-peak) were replaced by the more modern Luxor LX-200m panels, which feature the same size as the former but deliver 200 Wp.

4.2.4 UNDERGROUND MUON DETECTOR

The purpose of the Underground Muon Detector (**UMD**) is to measure the muon density of extensive air showers directly by employing scintillator detectors buried 2.3 m deep under the ground. This setup serves as shielding for the majority of the electromagnetic particles of a shower. Due to cost limitations, the **UMD** is deployed solely in the 750-m array, where a unit is installed for each **WCD** of this region.

A **UMD** unit is composed of three scintillator modules arranged in an "L" configuration, as depicted in Fig. 4.5. A module consists of 64 scintillator bars, each with dimensions $400 \times 4 \times 1 \text{ cm}^3$. The bars embed wavelength-shifting fibres, which guide the



Figure 4.5: Installation of an Underground Muon Detector (**UMD**) unit at a depth of 2.3 m. The three modules are arranged in an "L" shape, collectively covering an area of 30 m². Each module is composed of 64 scintillator bars with embedded wavelength-shifting fibres that guide the produced light to 64 **SiPMs**. The equipment is housed in a **PVC** enclosure, where a **PVC** tube is connected for installation and easy maintenance.

light produced within the scintillator towards an array of 64 Silicon Photomultipliers (**SiPMs**), model Hamamatsu S13361-2050 featuring 1584 micro-cells. The equipment is protected by a **PVC** enclosure, where a **PVC** tube (also visible in Fig. 4.5) is connected for maintenance and installation purposes.

The signals in the **SiPMs** are read out by two different electronic modes. The first is called the binary mode, where the 64 outputs are read separately with a sampling frequency of 320 MHz. Each sample can be either zero or one, depending on whether the signal is below or above the discriminator threshold. In the second mode, termed **ADC** mode, the signals of the 64 **SiPMs** are summed and digitised at a frequency of 160 MHz by high and low gain channels, yielding waveforms with 1024 bins. The reason for the implementation of these distinct modes is to provide an accurate estimation of the number of muons both when few or numerous quantities of these particles cross the **UMD**.

The **UMD** is triggered by the corresponding **WCD**, using a cable for communication between the electronics of the two detectors. Such cable introduces a delay in the communication, which is effectively mitigated using a memory depth of 6.4 μ s in the **UMD**. Additionally, the power is provided by dedicated solar panels independently

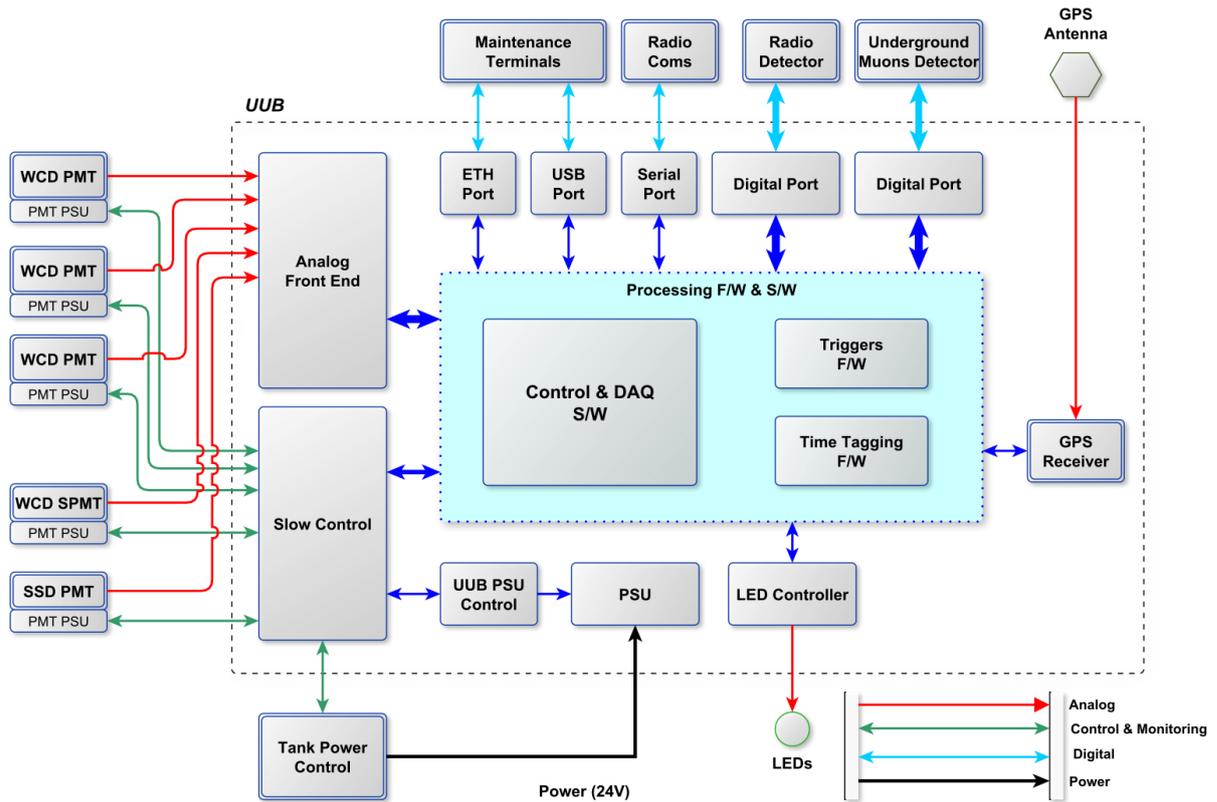


Figure 4.6: Schematic representation of the Upgraded Unified Board (**UUB**) and its high-level components. The **UUB** handles signal processing from the **WCD** and **SSD**, their calibration, local station triggers, monitoring information about the operation of the station, interface for the **RD** and **UMD** as well as communication with CDAS.

of those of the AugerPrime stations.

4.2.5 UPGRADED UNIFIED BOARD

With the introduction of the **SSD**, small **PMT**, **RD** and **UMD**, it became imperative to upgrade the electronics of the AugerPrime stations to comport the new detectors. The new electronics is implemented via the so-called Upgraded Unified Board (**UUB**), which is responsible for processing the signals of the **SSD** and **WCD**, their calibration, managing the different triggers of the local stations, monitoring their operation and providing an interface for the **RD** and **UMD**. A schematic description of the **UUB** and its interaction with the different detectors is provided in Fig. 4.6.

The different AugerPrime stations are kept synchronised by monitoring the variations of the 120 MHz clock in relation to the one pulse per second (PPS) of the **GPS** receiver. As part of the upgrade, the former Motorola Oncore UT+ **GPS** was replaced by the compatible Synergy SSR-6TF receiver, with an accuracy of approximately 2 ns.

A 16-bit CPU micro-controller, model MSP430, controls the high voltage supply necessary for the proper operation of the **PMTs**. The corresponding firmware also manages more than 90 monitoring variables, which provides valuable insights on the stability and overall health of a station. The main processor of the **UUB** is the Cortex A9 Dual 333 MHz ARM coupled with the Artix-7 **FPGA**. This combination manages and controls all the associated systems for signal processing, slow controls, triggering logic and interfaces.

A central function of the **UUB** is handling the analog output from the **PMTs** of the **WCD** and **SSD**. The signals of the **LPMTs** are still divided into low and high gain channels, however, only the signals from the anode of the **PMTs** are used. This signal is electronically splitted into two and the high gain output is amplified by a factor of 32. A similar process is applied to the output of the **SSD PMT**, but, in this case, the amplification ratio is 0.25 and 32 in the low and high gain, respectively, yielding a factor of 128 between these channels. Regarding the **SPMT**, no splitting or amplification are performed. The digitisation of the signals is achieved with dedicated flash analog-to-digital converters (**FADCs**) with sampling frequency of 120 MHz and resolution of 12 bits.

The implementation of local triggers in the **UUB** is of crucial importance for the experiment. To ensure a consistent behaviour of the surface array with respect to the Phase-I Observatory, the signals of the **LPMTs** are filtered and downsampled for triggering purposes, mimicking the pre-upgrade signals in what is called the *compatibility mode*. Utilising these signals, the previously established T1 and T2 triggers are implemented in the firmware of the **UUB**. Recent analyses have demonstrated similar trigger rates between the new and former electronics.

The **UUB** additionally contains ports for communication with the **RD** as well as with the **UMD** (when present). The board is powered by 24 V supplied by the existing batteries. The transmission of data to **CDAS** happens with the previously established bandwidth of 1200 bits per second.

CHAPTER 5

Calibration of UUB stations and signal compatibility

Henceforth, we will describe the analyses we have thoroughly developed and the consequential outcomes of our research work. Within this chapter, our primary focus is on the calibration of the AugerPrime stations equipped with the Upgraded Unified Board (UUB). We begin introducing a new and enhanced algorithm for the calibration and then we will delineate its implementation within the Offline framework and rigorously assess its performance when applied to the upgraded detectors. Furthermore, we expose our investigations regarding the signal compatibility between the AugerPrime stations and their pre-upgrade counterpart. This examination is crucial not only for validating the calibration procedure but, more importantly, the success of the Surface Detector upgrade itself.

5.1 CALIBRATION OF UUB STATIONS

In Section 3.2.2, we described the calibration process for the water-Cherenkov detectors during Phase I of the Observatory's operation. The introduction of the new Upgraded Unified Boards has brought significant changes that directly affect the calibration of the AugerPrime stations. In the case of the WCDs, the procedure of using the signal from vertical background muons as the calibration reference is still applied. However, a

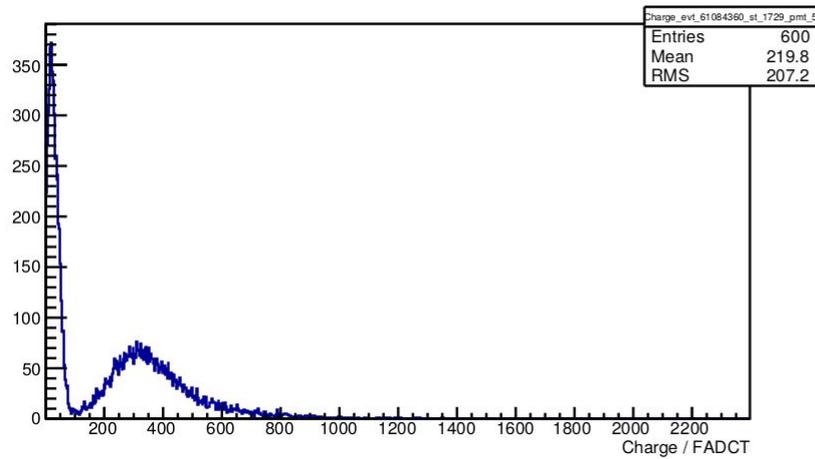


Figure 5.1: Example of a charge calibration histogram for an **SSD**. The prominent peak on the left results from the integration of the baseline and electronics noise when events triggering the **WCD** are not detected by the **SSD** due to its reduced dimensions. The second peak is generated by atmospheric particles with sufficient energy to produce a signal, specifically minimum ionising particles.

key distinction from its pre-upgrade configuration lies in the binning of the calibration histograms for signal amplitude and charge. This change is the result of the enhanced digitisation resolution of the **UUBs** compared to the former Unified Board.

The Scintillator Surface Detector (**SSD**) is calibrated using a procedure analogous to that of the **WCD**. However, since the response of the **SSD** to background electrons and muons is similar, the calibration is performed in terms of the signal produced by a *minimum ionising particle* (**MIP**). When the calibration trigger for a **PMT** of the **WCD** is satisfied, the signal in the **SSD** is also recorded to collect data for its amplitude and charge calibration histograms. Due to the smaller area of the **SSD** compared to the **WCD**, most of the events triggering the latter yield no physical signal in the **SSD**. Consequently, the calibration histograms of the **SSD** exhibit a prominent peak at low values, primarily reflecting the baseline and electronic noise. Nevertheless, a distinct second peak emerges due to particles energetic enough to produce a signal in the **SSD**, specifically, minimum ionising particles. An example of a charge calibration histogram for an **SSD** is presented in Fig. 5.1 to illustrate these characteristics.

Regarding the two electronics, the binning of the calibration histograms can be expressed in a generalised manner. For the amplitude histograms, which consist of

Table 5.1: Step and stride parameters for amplitude and charge calibration histograms sorted by electronics and detector type. The case of the SSD connected to a UB corresponds to the test stations known as the pre-production array.

| Electronics | Detector | Calibration Version | p_A | q_A | p_C | q_C |
|-------------|----------|---------------------|-------|-------|-------|-------|
| UB | WCD | all | 1 | 3 | 1 | 3 |
| | SSD | all | 1 | 3 | 1 | 3 |
| UUB | WCD | all | 4 | 4 | 8 | 4 |
| | SSD | < 262 | 4 | 4 | 8 | 4 |
| | | ≥ 262 | 2 | 4 | 2 | 4 |

150 bins, the bin limits are defined as follows:

$$\begin{cases} p_A i & \text{for } i \leq 100 \\ p_A(100 + q_A i) & \text{for } i \in (100, 150], \end{cases} \quad (5.1)$$

where i represents the bin index, and the parameters p and q are referred to as "step" and "stride", respectively, with the subscript "A" denoting the amplitude histogram. Conversely, the charge histograms, indicated by the subscript "C", possess 600 bins organised according to the following scheme:

$$\begin{cases} p_C i & \text{for } i \leq 400 \\ p_C(400 + q_C i) & \text{for } i \in (400, 600]. \end{cases} \quad (5.2)$$

In Table 5.1, the step and stride parameters for the calibration histograms are outlined for both electronics and detectors. We stress the differing values for the two boards, particularly concerning the WCDs. In Fig. 5.2, this distinction in binning is also evident in the charge histograms, whilst the overall shape remains similar for both electronics.

During Phase I of the Observatory operation, the extraction of the peak value from the charge histograms to convert it into the charge of a vertical equivalent muon was accomplished using an algorithm implemented in the module SdCalibrator of the Offline framework. This algorithm was optimised to operate on the specific shape of the calibration histograms provided by the UBs. Consequently, its application to the SSD and the new WCD histograms with different binning not only becomes inadequate but also leads to failure in event reconstruction encompassing AugerPrime stations.

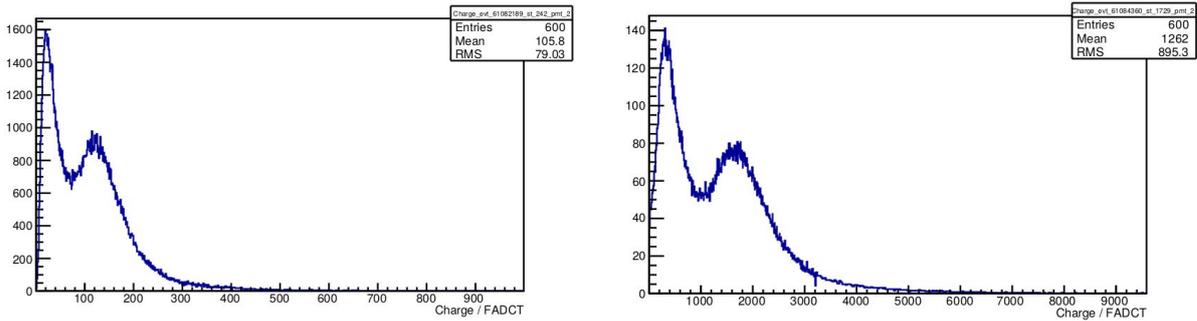


Figure 5.2: Example of charge calibration histograms of WCDs. The left panel displays a histogram of a UB station, whilst on the right a UUB histogram is presented. Although the general shape is similar in both cases, the different binning is evident from the horizontal axes. The "T" in the unit "FADCT" indicates that the flash ADC channels of the signal are integrated in time.

Fortunately, a new calibration algorithm has been developed, which is compatible with data from the UUBs. In the following section, we will provide a detailed description of this algorithm, and in Section 5.1.2, we will outline the modifications implemented to the Offline framework to incorporate this new calibration procedure, as well as to ensure its compatibility with the reconstruction of Phase II data.

5.1.1 NEW ALGORITHM FOR FITTING CALIBRATION HISTOGRAMS

The fitting algorithm for calibration histograms discussed in this section was developed by Dr. Alexander Streich as part of his PhD work. A thorough documentation of this enhanced algorithm and its application to UB data can be found in Ref. [70]. The algorithm was initially designed to be employed for Phase-I data. Over time, the original algorithm experienced an increasing failure rate, primarily attributed to deviations from the expected histogram shapes caused by the ageing of the WCDs. Since effectively addressing this issue involves improving the robustness with a reduced reliance on the precise structure of the calibration histograms, the new algorithm was engineered to additionally cover the upgraded stations.

In the initial stage of the enhanced algorithm, the histograms undergo a pre-selection process. To effectively filter out problematic histograms, a criterion is set based on the total number of entries (N), requiring it to exceed 10 000 counts. However, this condition alone does not provide insight into the internal structure of the histogram. To gain further understanding of the presence of potential minima and maxima within

the histogram, the Shannon's information entropy (S) is employed. Here, the entropy is defined as:

$$S = -\frac{1}{N} \sum_i c_i \ln c_i + \ln N, \quad (5.3)$$

where c_i represents the number of entries in the i^{th} bin. An empirical analysis has determined that dysfunctional histograms can be effectively excluded by applying the two-dimensional condition of $S > 5$ and $N > 50\,000$.

Following the aforementioned histogram selection, the search for the muon peak is initiated. In this phase, the initial step involves normalising the bin entries in relation to their size. This normalisation mitigates the artificial discontinuity introduced by the change in bin size for the final 200 bins, ensuring it does not interfere with the subsequent muon peak identification process. Moreover, a smoothing procedure is applied to the histogram by computing the average of each bin¹ with both its preceding and succeeding neighbours, yielding the respective densities d_i . This smoothing operation serves to minimise the bin-to-bin fluctuations, thereby facilitating the identification of the muon peak. Additionally, during this phase, the corresponding standard deviations (σ_{d_i}) for these densities are computed.

Utilising the normalised and smoothed histogram, the search for the muon peak employs a bin-to-bin comparison, starting from the rear of the histogram and proceeding in reverse order, with the exclusion of the last five bins, as these bins are allocated for metadata transmission. Once the maximum is identified, a specific condition regarding the left and right limits is enforced. Denoting the standard deviation of the maximum bin as $\sigma_{d_{\text{max}}}$, the limit bins should be below a threshold of $f\sigma_{d_{\text{max}}}$, where f -values of 10, 8 and 5 are iteratively tested. If the threshold test fails for all tested f values, the histogram is rejected, and the fitting procedure is terminated. Conversely, if the threshold condition is met, a fitting process is performed within the region of the identified maximum, utilising a second-order polynomial described by:

$$y(x - x_0) = (c_2(x - x_0) + c_1)(x - x_0) + c_0. \quad (5.4)$$

In addition to assessing the reduced χ^2 , the resulting curve must exhibit negative curvature, indicated by a negative value for the coefficient c_2 .

¹The averaging is not performed for the first and last bins.

To ensure the identified maximum corresponds to the characteristic second peak of the calibration histograms, the bin-to-bin search continues from the left limit in an attempt to locate the anticipated minimum to the left of the maximum. Upon finding such a minimum, a similar threshold condition is imposed, specifically, both the left and right limits must exceed $f\sigma_{d_{\min}}$, where $\sigma_{d_{\min}}$ represents the standard deviation of the corresponding minimum bin. If this condition is satisfied, a quadratic fit is applied to this region, and the result is inspected for positive curvature. In instances where a minimum to the left cannot be identified, the fitting procedure is terminated.

Despite the identification of a maximum with an associated minimum to its left, further investigations revealed the persistence of a few problematic histograms. Consequently, a final condition has been introduced to effectively eliminate such outliers. This condition stipulates that the ratio between the charge of the maximum and that of the minimum must exceed 1.3. Only histograms that successfully pass through all the specified steps are retained, and the charge corresponding to the muon maximum is stored for subsequent processing of the PMT signals.

5.1.2 MODIFICATIONS TO THE OFFLINE FRAMEWORK

Whilst the fitting algorithm outlined in Section 5.1.1 was expected to be applicable to the calibration histograms of UUB stations, it had not been integrated into the official Offline framework at that point. Additionally, the existing software was incomplete for correctly handling UUB data, particularly for signal processing and event reconstruction. In this section, we detail the modifications and implementations we performed to adapt the framework, ensuring compatibility and readiness for the processing and reconstruction of UUB data.

Prior to the upgrade, the charge and amplitude histograms had a fixed binning, which was hard-coded into the `Station` interface within the framework. Our initial step involved modifying this interface to enable it to accept parameters denoting the type of electronics and detector as well as their calibration version. This modification allows the interface to dynamically provide the appropriate binning for distinct detector configurations, as outlined in Table 5.1. Note that the table includes entries for an **SSD** connected to a **UB**. These entries correspond to a set of test stations collectively referred to as the **SSD Pre-Production Array (PPA)**. During the period of 2019 to 2021, these

stations were equipped with an SSD connected to one of the PMTs of the WCD. This configuration was designed to evaluate the behaviour and performance of the SSDs in the field environment. Importantly, our implementation of the Station interface also extends support for the calibration histograms of the SSD PPA.

The original fitting algorithm, designed for calibrating histograms during the pre-upgrade phase, is encapsulated within an Offline module known as SdCalibrator. This module, while responsible for estimating calibration factors, also handles all the signal processing tasks. This comprehensive functionality posed challenges for making major modifications and evaluating their impacts effectively.

To address this situation, prior to implementing the improved fitting algorithm, we divided the SdCalibrator module into four separate modules, each dedicated to specific tasks yet collectively achieving the same outcomes. A brief overview of these new modules, along with their respective names, is outlined below:

- SdStationCheckerOG – This module is responsible for identifying stations that exhibit anomalies or issues, such as missing calibration data, trigger errors, absent GPS data, and other related problems.
- SdHistogramFitterOG – Employing the old calibration algorithm, this module conducts the fitting of the calibration histograms, yielding their peak charge. The charge is corrected from omnidirectional to vertical incidence of particles (refer to Section 3.2.2), resulting in the conversion factor for the signals from hardware units (ADC) to the physical unit VEM.
- SdBaselineFinderOG – This module computes the baseline of the signal pulses for each gain channel of the PMTs. It achieves this by identifying flat segments within the pulse and subsequently applying a linear interpolation.
- SdTraceCalibratorOG – For each station in an event, this module analyses the structure of the PMT pulses to identify fragments containing signals related to the detected event. It determines the start and stop times within the pulse, effectively excluding any signal attributed to background particles. Subsequently, it calculates and calibrates the total signal.

In the context of the SdTraceCalibratorOG module, and to ensure consistency in our terminology going forward, we will refer to a pulse from a PMT, which

describes the temporal structure of the signal, as a *trace* ($T(t)$). The signal (S) is determined by integrating the trace, with the baseline ($B(t)$) subtracted, over the specified interval defined by the start (t_{start}) and end (t_{end}) times:

$$S = \sum_{t_{\text{start}}}^{t_{\text{end}}} (T(t) - B(t)). \quad (5.5)$$

The use of the summation symbol emphasises the discrete nature of the traces, which arises from the fixed sampling rate of the electronics. Subsequently, the calibrated signal is obtained by dividing the integrated trace by the calibration factor, which is provided by the `SdHistogramFitter0G` module.

The compatibility of the four new modules with the `SdCalibrator` was subject to thorough investigation through event reconstructions spanning the entire year of 2004 and subsequently for the 7th, 14th, 21st, and 28th days of each month until 2019. This extensive time frame was chosen to encompass all the calibration versions utilised in the `SD` of the Observatory. Our assessment involved comparing the results of reconstructions carried out using both the four new modules and the `SdCalibrator`. We meticulously ensured that identical numerical values for calibration factors, calibrated signals, event multiplicity, reconstructed energy, and arrival direction were consistently achieved with both module configurations.

Following the successful validation of the four modules, the enhanced algorithm for fitting calibration histograms, outlined in Section 5.1.1, was incorporated into a new module named `SdHistogramFitterKG`. It is important to remember that this algorithm is designed to function with both `UB` and `UUB` data. Moreover, in cases where the histogram fitting process fails, we utilise the online estimate of the `VEM` charge, which is included in the raw data. Specifically for `WCD` equipped with `UUB`, we check that the online value falls within the expected range of 1 000 to 2 000 `ADC` channels; otherwise, the `PMT` is rejected. For the `SSD`, since an online estimate of the `MIP` charge is currently unavailable, we have defined a nominal value of 157 `ADC` based on the mean value observed in the data. This nominal value is employed when the fitting procedure fails.

The modules we implemented have been successfully integrated into the main branch of the `Offline` software, rendering it compatible with the reconstruction

of UUB data. Through extensive validation efforts, as detailed in this chapter and described in Ref. [70] for UB stations, we have confirmed the robustness and reliability of these modules. Notably, the four modules, now featuring the newly introduced `SdHistogramFitterKG` replacing the former `SdHistogramFitterOG` module, are integral components of the Standard Applications used by the Auger Collaboration. These modules play an essential role in the reconstruction of SD events in Phase II of the Observatory, which has recently commenced.

5.1.3 PERFORMANCE OF THE NEW ALGORITHM

The assessment of the performance of the new calibration algorithm and its validation was thoroughly conducted for the UB stations in Ref. [70]. However, it was mandatory to extend this investigation to the UUB stations, thereby demonstrating its compatibility, efficiency, and reliability in the context of their calibration. This essential evaluation serves to establish the viability of the algorithm for processing data in the Phase II of the Observatory operation.

To initiate this evaluation, we have undertaken a comprehensive examination of the success rate achieved by the new algorithm when applied to the calibration of the UUB stations. A successful outcome is realised when the algorithm correctly identifies a peak due to the atmospheric muons within the calibration histogram, and subsequently, it yields a conversion factor between the hardware and physical units.

To conduct this analysis, we have utilised data collected between 01 January 2022 and 06 February 2023. This dataset comprises a pool of 16 463 403 calibration events corresponding to WCD PMTs, alongside 5 895 146 events with respect to the SSD PMTs within UUB stations. To facilitate a comparative analysis, we concurrently examined the 39 370 933 calibration events recorded for UB stations within the same temporal window.

We quantify the efficiency of the calibration algorithm simply as the proportion of successfully accomplished histogram fits within the subset of PMTs that have been flagged as working under the expected conditions, also called *unmasked PMTs*. In Table 5.2, we present the efficiency of the algorithm across distinct electronics and PMT types. Furthermore, the rate of masked PMTs is displayed for each scenario.

For WCD PMTs, our investigation reveals a rate of 8.7% of dysfunctional

Table 5.2: Efficiency of the algorithm for fitting calibration histograms alongside the rate of masked PMTs for the distinct electronics and PMT types. The algorithm successfully fits more than 99.5% of the calibration histograms for both electronic boards and detectors.

| Electronics | PMT type | Rate of masked tubes (%) | Fitter efficiency (%) |
|-------------|----------|--------------------------|-----------------------|
| UB | WCD | 8.7 | 99.7 |
| UUB | WCD | 8.3 | 99.7 |
| UUB | SSD | 0 | 99.8 |

PMTs for UB stations, whilst UUB exhibit a closely corresponding rate of 8.3%. We emphasise that despite the replacement of the electronic boards, the PMTs of the stations remain the same, which accounts for the similar rate of masked PMTs observed. In the case of both electronics, the efficiency of the new algorithm is 99.7%, confirming its robustness and applicability to calibrate histograms of WCD PMTs provided by stations equipped with the UUB.

In the case of SSD PMTs within UUB stations, the operational flag indicating whether a tube is masked or not is not yet implemented in the process of data acquisition, hence the zero rate of masked SSD PMTs observed. To ensure that we precisely compute the efficiency considering only the PMTs that are operating normally, we have merged the SSD calibration data to the corresponding monitoring data closest in time. The efficiency is calculated for the calibration events whose monitoring data is within 400 s and that the high voltage in the SSD PMT is within the range of normal operation.² Utilising the data selected by the aforementioned considerations, we calculated an efficiency of 99.8% for the application of the new calibration algorithm to SSD histograms. It is noteworthy that in over 99.5% of calibration events the new calibration algorithm consistently provides calibration factors regardless of the type of electronics and detector, underscoring its efficacy.

Whilst the efficiency of the calibration algorithm has been demonstrated, a complementary analysis is essential to scrutinise the meaningfulness of the calibration factors derived from the histogram-fitting procedure. We illustrate in the left column of Fig. 5.3 the distributions of the mean VEM (MIP) charge of the stations, encompassing distinct electronics configurations and PMT types. Specifically for the WCDs, the

²The range of normal high voltage in the SSD PMT is considered between 800 and 900 V.

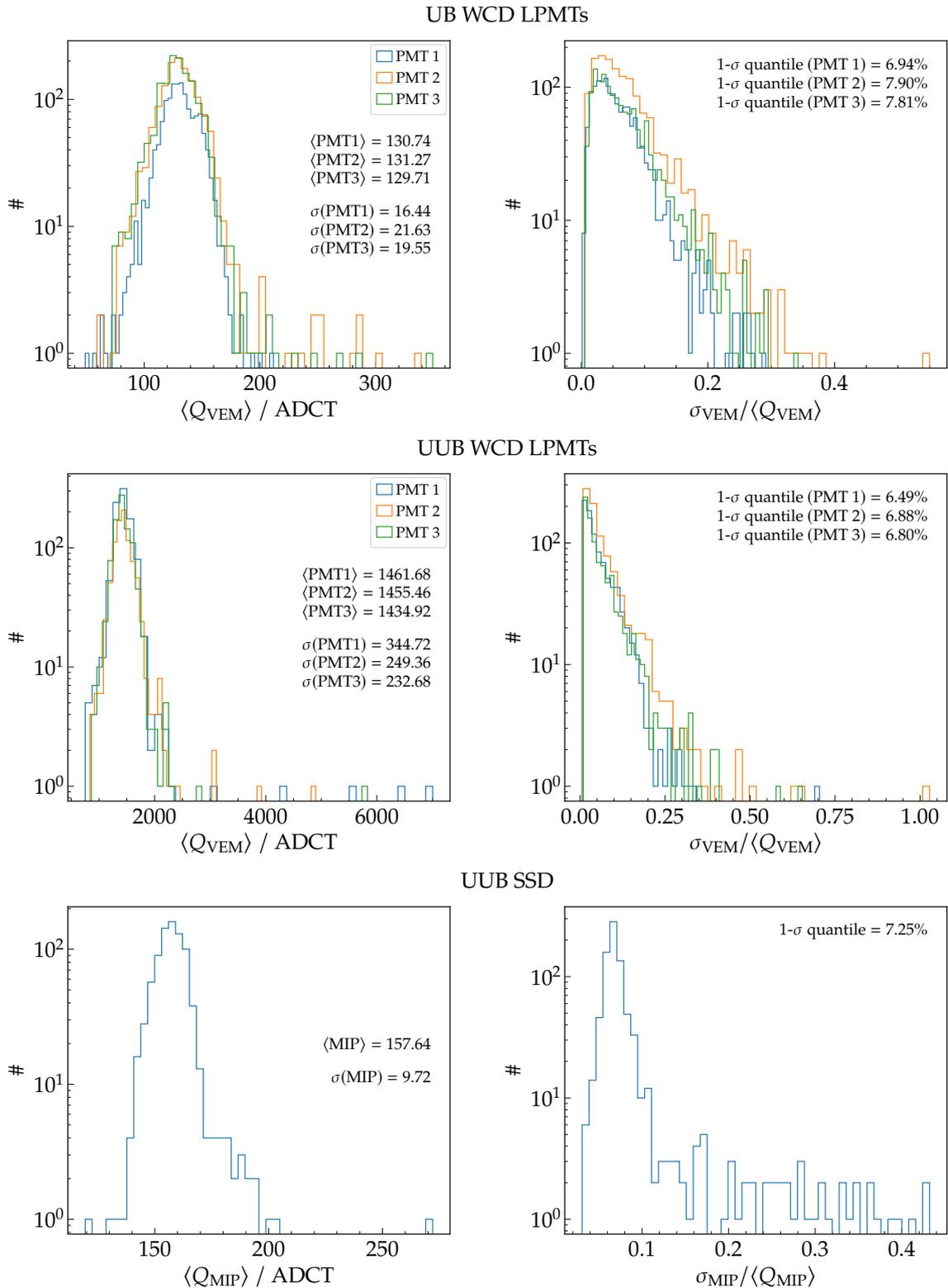


Figure 5.3: Distributions for the mean charge calibration factor (left column) and relative deviation (right column) for individual PMTs. The three different WCD PMTs are separately shown for UB (top row) and UUB (middle row). The results for the SSD PMTs (bottom row) are also presented. The distributions are similar for the different WCD PMTs. A relative variation of 7.6%, 6.7% and 7.2% in the 1- σ quantile is observed for WCD UB, UUB and SSD, respectively.

graphs display the individual distributions corresponding to each of their three PMTs.³ Evidently, the three PMTs exhibit a normal distribution which are similar between each other in both the cases of stations equipped with UB and UUB. This finding further reinforces the robustness of the new algorithm.

The inherent station-to-station fluctuations, contributing to the overall distribution, predominantly stem from the disparities in the high voltage applied to the PMTs. This discrepancy in high-voltage settings emerges during the calibration process undertaken at the local stations. Notably, these deviations amount to approximately 15% and 19% for UB and UUB electronics, respectively. It is important to stress that these variations do not reflect calibration inconsistencies. Rather, they merely indicate the distinct station-specific conditions necessitating correspondingly distinct values of VEM charge for optimal PMT calibration.

To rigorously assess the consistency of the VEM charge values provided by the histogram-fitting algorithm, we considered, individually for each PMT, the relative deviation of their VEM charge with respect to their mean value. The resulting distributions are presented in the right column of Fig. 5.3. For the WCDs, 68.3% of the PMTs exhibit a relative deviation of VEM charge of approximately 7.6% and 6.7% for UB and UUB electronics, respectively. It is again worth mentioning that the behaviour of the distributions for the three PMTs is similar. This outcome not only attests to the consistency of the new algorithm for the WCD PMTs of UUB stations, but also verifies an enhanced performance relative to their UB counterparts.

The lowermost row of Fig. 5.3 showcases the corresponding results concerning the SSD PMTs. The qualitative discussion previously established for WCD PMTs extends to the calibration process applied to the SSD PMTs. However, the station-to-station variation of the MIP charge is approximately 5.7%. This discrepancy, relative to the WCDs, could be attributed to the specific selection on high voltage applied to the analysis of the SSD PMTs. Moreover, a relative variation of 7.2% in the MIP charge is observed for the $1\text{-}\sigma$ quantile. To summarise, it is unequivocally verified that the histogram-fitting algorithm maintains its consistency when applied to the SSD PMTs as well, thereby reinforcing its comprehensive applicability across the WCD and SSD of the upgraded surface detector of the Auger Observatory.

³For sake of clarity, we refer to the large PMTs (LPMT) of the WCDs. The calibration of the small PMT follows a different procedure, as described in Ref. [71].

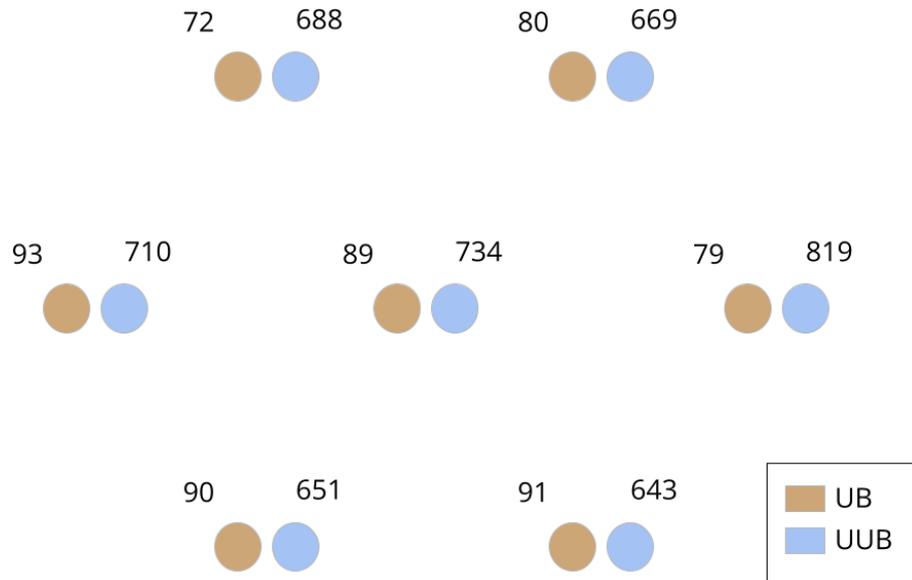


Figure 5.4: Schematic view of the hexagon with doublet stations used for compatibility studies. Each doublet has one station with UUB and the other with UB. Their separation is approximately 11 m and the grid spacing is 1500 m. The numbers indicate the identification number of the stations.

5.2 COMPATIBILITY OF UUB AND UB SIGNALS

In the following, we studied the compatibility of the signals of WCDs between UUB and UB. Since the event reconstruction with the surface detector derives from the processing of the station signals, it is crucial to show that they are compatible between UUB and UB. This study not only contributes to the validation of the electronics upgrade but it is also a step towards ensuring compatibility between Phase-I and Phase-II data.

5.2.1 INTEGRATED SIGNALS

Our compatibility studies use data from the doublet hexagon. This is a hexagon in the infill region composed of doublet stations, but the distance between doublets is that of the regular 1500-m array. In January 2022, the setup of this hexagon was modified so that each doublet consists of one station with UUB and the other with UB, as displayed in Fig. 5.4. The stations of a doublet are separated by a distance of approximately 11 m. Taking advantage of the doublet hexagon being situated in the infill region, we have used an Offline application for event reconstruction of infill stations. This provides us with a larger number of events to analyse than if we had used the standard

reconstruction for the 1500-m array. We used data from 01 January 2022 until 02 April 2023 for these analyses.

The stations of a doublet sample the same part of the shower, giving a nearly-direct comparison of their integrated signals. When the doublet is close to the shower axis, their 11-m separation causes a significant difference in signal, as the LDF is very steep in this range of distances [72]. This introduces a bias for our subsequent analyses. To correct for the radial dependency of the signals, we calculate the corrected signal as the one that the station would have if it was in the mean distance between the two of the doublets with respect to the shower core, i.e.,

$$S_{\text{corr}}(r) = S(r) - [S_{\text{LDF}}(r) - S_{\text{LDF}}(\bar{r})], \quad (5.6)$$

where $S_{\text{corr}}(r)$ is the corrected signal of the station at a distance r from the shower core, $S(r)$ is the actual signal measured by the station, S_{LDF} is the expected signal from the LDF and $\bar{r} = (r_1 + r_2)/2$ is the mean distance of the two stations of the doublet to the shower core.

In Fig. 5.5, we present the mean relative difference between UUB and UB signals for one of the doublets as a function of the difference between the distance of the stations to the shower core. We only considered the events whose mean UB/UUB signal is larger than 100 VEM so that these stations are close to the shower core where the LDF is steeper. For uncorrected signals, we clearly observe an increasing bias as the distance difference between the stations and the shower core becomes larger. As expected, when the UUB station is closer to the shower core, its measured signal is larger than the UB station, and vice versa. However, once the correction is applied, the bias due to the separation distance vanishes, and we are able to compare fairly their signals.

Given the fluctuations due to the statistical nature of the physical processes involved in the detection of particles, we expect the signals registered by each of the stations of a doublet to be, in general, slightly different for the same event. However, if the VEM calibration of UUB and UB is compatible, their calibrated signals must be within their expected uncertainty. In Fig. 5.6, we show the scatter plot of both high- and low-gain signals in the UUB and UB stations of one of the doublets. The aforementioned correction of the signals was applied. In the background of the scatter plot we present

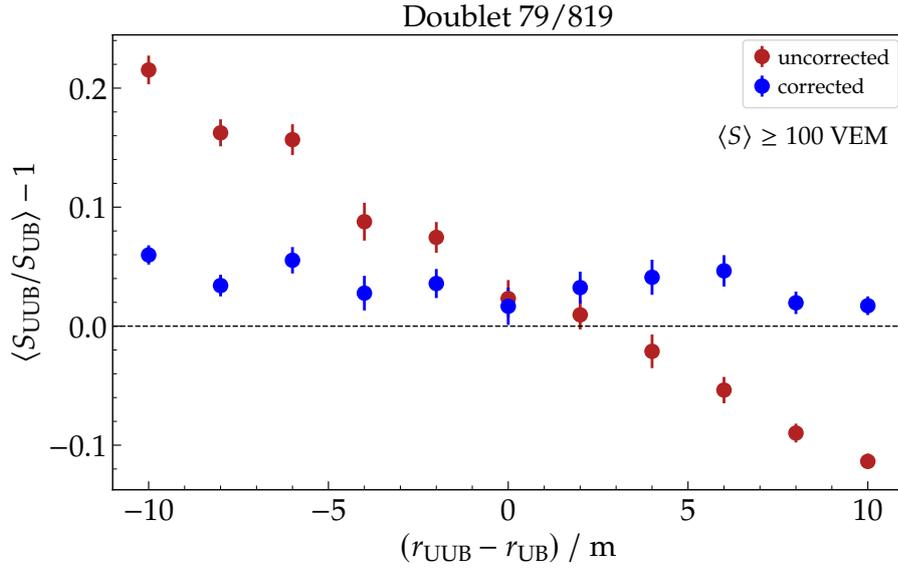


Figure 5.5: For the doublet stations 79 and 819, the mean relative difference of UUB and UB signals is presented as a function of the distance difference of the stations to the shower core. Only events with mean UB/UUB signals larger than 100 VEM are considered. The signal correction given in Eq. (5.6) successfully accounts for the signal bias introduced by the separation distance between the stations of a doublet.

the 1σ and 3σ bands of the expected signal uncertainty around the equality of the UB and UUB signals, which is represented by the black dashed line. We observe that the data is mainly contained in the expected region of signal variation which indicates that the UUB and UB signals are compatible within their expected fluctuations.

To ensure the compatibility of the UUB and UB signals, we need to show whether any biases exist between them. Our approach is to fit both high- and low-gain signals of UB and UUB for each doublet and study how close they are to the equality line, as the one shown in Fig. 5.6. Unfortunately, a simple χ^2 fit would not be appropriate in this case, because effects such as differences in trigger, transition from high gain to low gain and saturation for UUB and UB should be accounted for to avoid systematic biases. To obtain an unbiased fit, we have developed and applied a maximum likelihood method.

We choose a power law, $S_{\text{UUB}} = A(S_{\text{UB}})^B$, as the fit function to describe the relation between UB and UUB signals. When the parameters are determined from the fit, B will indicate the linearity between UUB and UB, and A possible biases. The likelihood function \mathcal{L} is defined as the product of the probability density function (p.d.f.) $f(S_{\text{UUB}i}, S_{\text{UB}i})$ of measuring a pair of signals in a UUB/UB doublet for a given

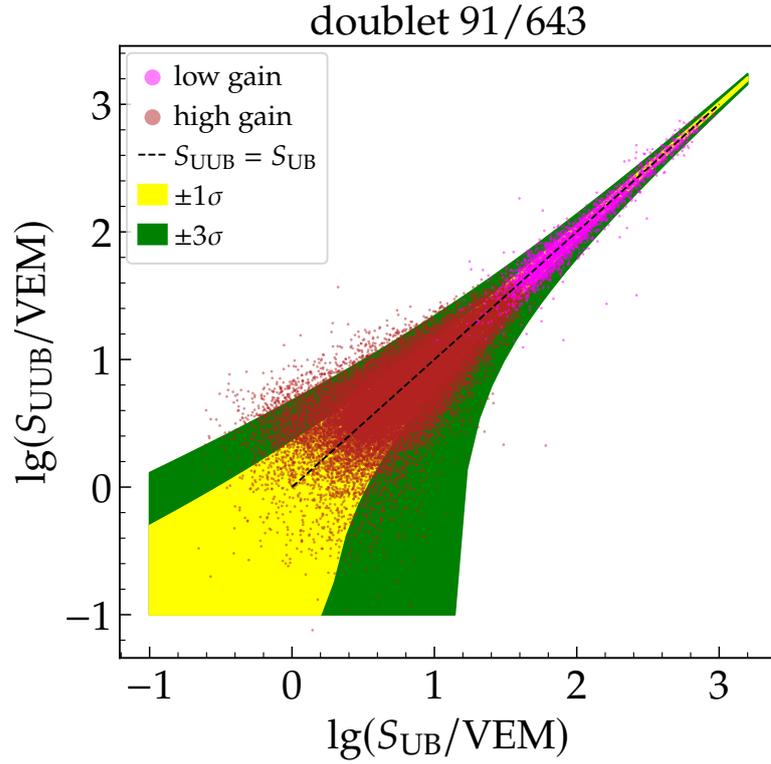


Figure 5.6: Scatter plot of the UUB and UB signals registered by the doublet 91/643. The bands represent the expected 1σ and 3σ variation around the dashed black line which corresponds to the equality of UB and UUB signals. The variation of the signals of the two electronics are compatible within their expected uncertainty.

event i ,

$$\mathcal{L} = \prod_i^N f(S_{\text{UUB}i}, S_{\text{UB}i}). \quad (5.7)$$

Then, we only have to minimise⁴ $-\ln \mathcal{L}$ to obtain the parameters A and B of the fit function.

Instead of parameterising the trigger efficiency of the stations and incorporating it to the probability of our likelihood function, we have simplified this step by performing a data selection. We will avoid the biases due to trigger effects by using only those events for which both stations of the doublet present signals above full T2-trigger efficiency. For UB stations, it was shown that the T2 trigger efficiency is 100% above 16 VEM [73]. The trigger-efficiency curve for UUB stations presents a similar behaviour and we also take 16 VEM as a cut for the UUB signals.

For each of the gains, we can write the conditional p.d.f. $g(S_{\text{UUB}}|S_{\text{UB}})$ of measuring a signal in the UUB station, given that a certain signal is measured by the

⁴Minimising $-\ln \mathcal{L}$ is equivalent to maximising \mathcal{L} .

UB station as

$$g(S_{\text{UUB}}|S_{\text{UB}}) = \frac{f(S_{\text{UUB}}, S_{\text{UB}})}{f_{\text{UB}}(S_{\text{UB}})}, \quad (5.8)$$

where $f_{\text{UB}}(S_{\text{UB}})$ is the marginal p.d.f. of UB signals, which is simply the projection of the $S_{\text{UUB}}:S_{\text{UB}}$ space onto the S_{UB} axis. The conditional p.d.f. $g(S_{\text{UUB}}|S_{\text{UB}})$ describes the fluctuation of the UUB signals around the fit function for a given value of S_{UB} . On the other hand, we can also write the conditional p.d.f. $h(S_{\text{UB}}|S_{\text{UUB}})$ of measuring a UB signal given that the UUB station has measured a certain signal

$$h(S_{\text{UB}}|S_{\text{UUB}}) = \frac{f(S_{\text{UUB}}, S_{\text{UB}})}{f_{\text{UUB}}(S_{\text{UUB}})}, \quad (5.9)$$

where $f_{\text{UUB}}(S_{\text{UUB}})$ is the marginal p.d.f. of UUB signals and $h(S_{\text{UB}}|S_{\text{UUB}})$ describes the variation of UB signals around the fit function. From Eqs. (5.8) and (5.9), we find that $f(S_{\text{UUB}}, S_{\text{UB}})$ is given by $g(S_{\text{UUB}}|S_{\text{UB}}) f_{\text{UB}}(S_{\text{UB}})$ and $h(S_{\text{UB}}|S_{\text{UUB}}) f_{\text{UUB}}(S_{\text{UUB}})$. For our final expression of $f(S_{\text{UUB}}, S_{\text{UB}})$, we average on the two conditional p.d.f. to symmetrically account for the variations around the fit function and include both UB and UUB signal distributions, which leads us to

$$f(S_{\text{UUB}}, S_{\text{UB}}) = \frac{g(S_{\text{UUB}}|S_{\text{UB}}) f_{\text{UB}}(S_{\text{UB}}) + h(S_{\text{UB}}|S_{\text{UUB}}) f_{\text{UUB}}(S_{\text{UUB}})}{2}. \quad (5.10)$$

For both $g(S_{\text{UUB}}|S_{\text{UB}})$ and $h(S_{\text{UB}}|S_{\text{UUB}})$, we assume that below 20 VEM, the fluctuation of the respective signals around the fit function follows a Poisson distribution and above that value, a normal distribution is used. For modelling the UB and UUB signal distributions $f_{\text{UB}}(S_{\text{UB}})$ and $f_{\text{UUB}}(S_{\text{UUB}})$, we used data from the doublets themselves. In Fig. 5.7, we present the distributions of signals for high and low gain, as well as for saturated signals in the UB and UUB stations of one of the doublets. The similarity between the UB and UUB distributions is also an indication of their compatibility. We observe the transition from high-gain to low-gain signals around 45 VEM, and then they mostly saturate above 800 VEM. However, these transitions are clearly not sharp, and, to account for them in our likelihood function, we have fitted them for each gain and individually for the two electronics. The high-gain distribution was fit to a generalised exponential function in the $\lg = \log_{10}$ of the signals, whilst the distributions for the low gain and saturated signals were fitted with an exponentially-modified normal distribution.

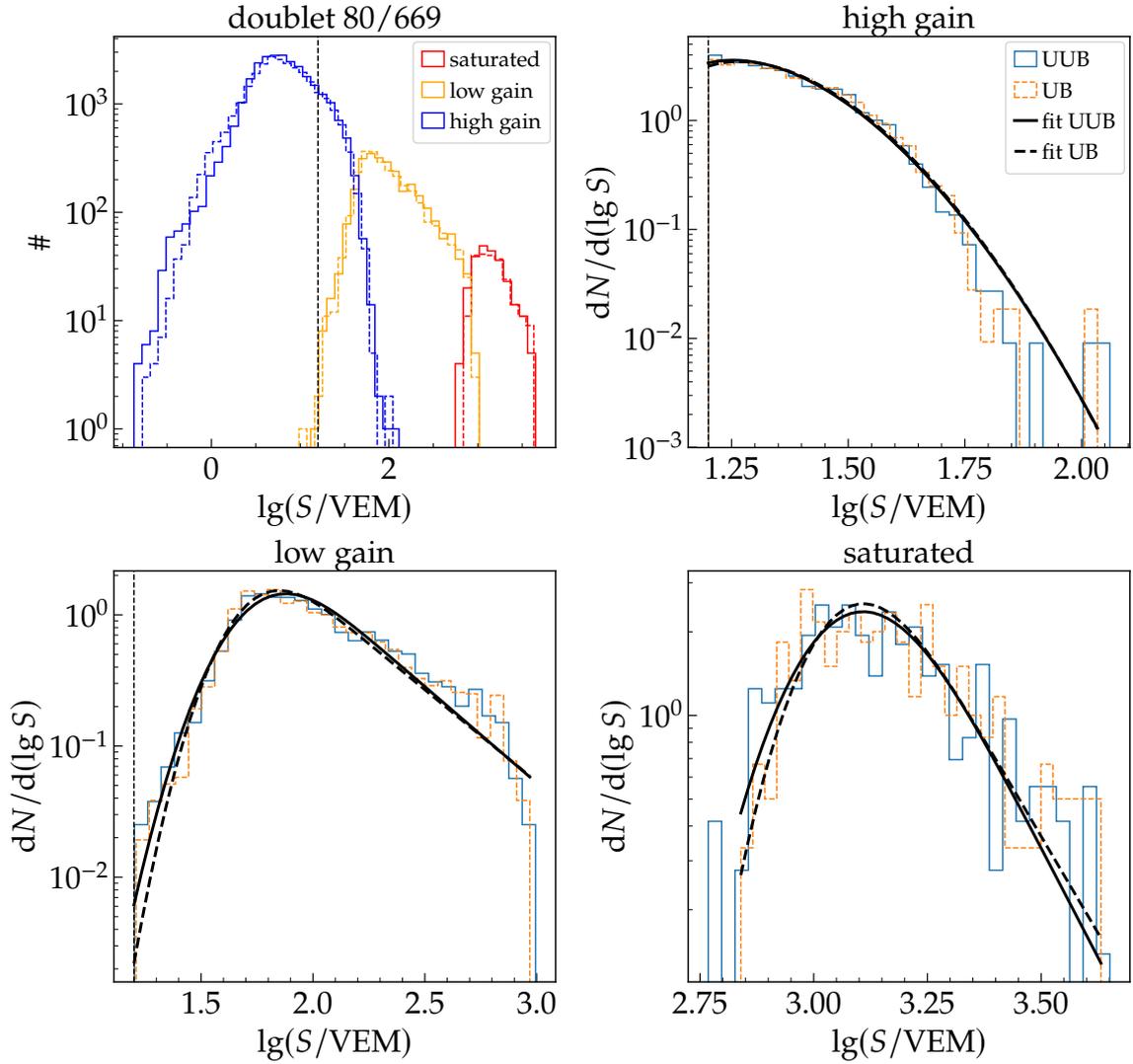


Figure 5.7: Distribution of UB and UUB signals in doublet 80/669 for high-gain, low-gain and saturated signals. Similar distributions are observed for the different electronics. The high-gain distribution was fitted with a generalised exponential function, whereas the low-gain and saturated signals with an exponentially-modified normal distribution, in all cases, separately for UB and UUB.

After formulating the likelihood function, we conducted a maximum-likelihood fit for each of the doublets in our analysis. An illustrative example of this fit for one of the doublets is depicted in Fig. 5.8 (left). We define y as the functional relationship between the UUB and UB signals, i.e., $y \equiv S_{\text{UUB}}(S_{\text{UB}})$. Accordingly, the equality line is $y_{\text{eq}} \equiv S_{\text{UUB}} = S_{\text{UB}}$, whilst the fit line is represented by $y_{\text{fit}} \equiv S_{\text{UUB}}(S_{\text{UB}}) = A(S_{\text{UB}})^B$. The fitted curves for both low- and high-gain signals are displayed along with the equality line. To better distinguish the disparities between the fit results and the anticipated equality, their relative differences are presented as a function of the signal.

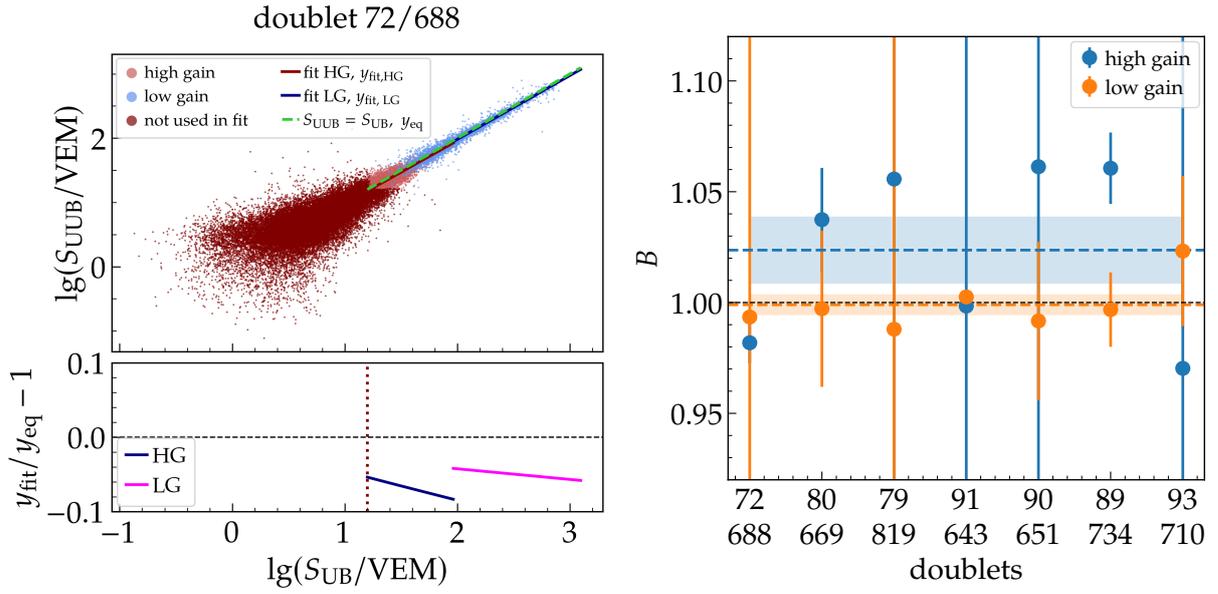


Figure 5.8: *Left*: Result for the maximum-likelihood fit for both gain channels in the doublet 72/688. The relative difference between the fits and the equality function are presented. *Right*: Parameter B obtained from the maximum-likelihood fit for each of the doublet stations. The non-linearity between the signals of the two electronics were found to be $(2.4 \pm 1.5)\%$ and $(0.1 \pm 0.4)\%$ for the high and low gain, respectively.

Using our chosen fit function, the relative difference is given by

$$\frac{y_{fit}}{y_{eq}} - 1 = \frac{A(S_{UB})^B}{S_{UB}} - 1 = A(S_{UB})^{B-1} - 1. \quad (5.11)$$

Note that the slope of the relative-difference curve is primarily driven by parameter B . For $B > 1$, the curve increases with signal, for $B < 1$ it decreases, and $B = 1$ translates into a constant relative difference and a perfectly linear relation between UUB and UB signals. We define, thus, the non-linearity of the two electronics as $|B - 1|$. The different values of parameter B obtained from the fit⁵ are shown in Fig. 5.8 (right) for both gain channels of all doublets. The dashed line and the band matching the marker colours represent the mean and its standard error, respectively, for each gain considering all the doublets. Due to the statistical uncertainty of the fit and also station-to-station variations, we use the mean value of B to calculate the non-linearity between UUB and UB signals. We found for the high gain a value of $(2.4 \pm 1.5)\%$ and $(0.1 \pm 0.4)\%$ for the

⁵In Fig. 5.8 (right), the error bars for some of the doublets appear incompatible with the station-to-station variation. Note that the error bars were obtained from the inverse Hessian matrix, which is computed by the minimiser used for the maximum-likelihood fit, in this case, `scipy.optimize.minimize` [74], from the Scipy package. This is an approximation for the uncertainty by evaluating the second derivative of the likelihood function at the minimum found. Therefore, the incompatible error bars likely come from the shape of this function at the minimum.

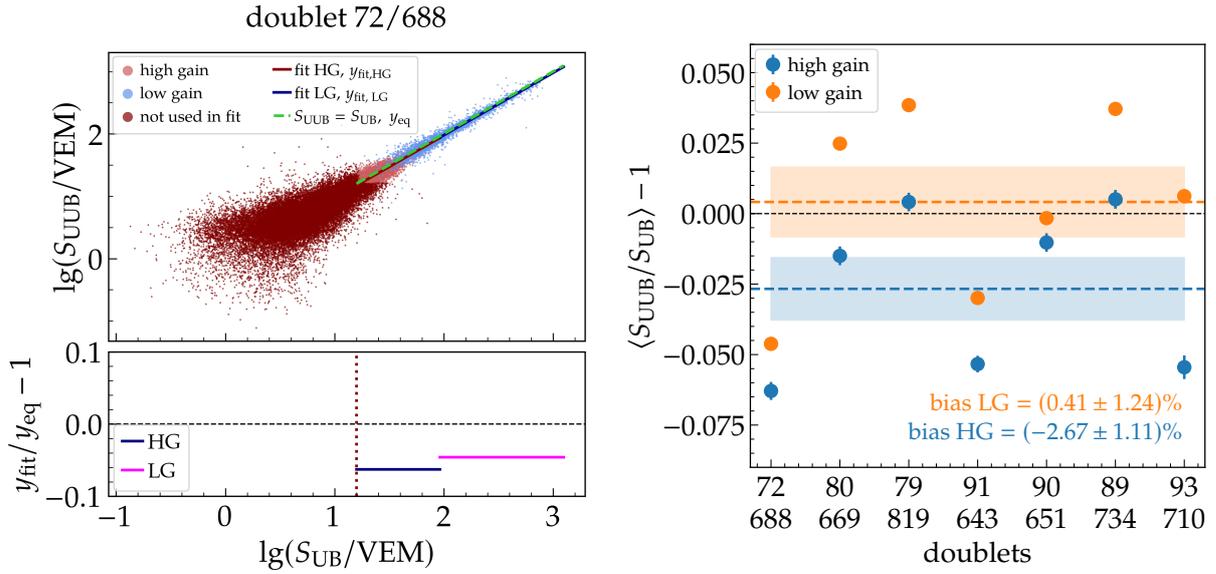


Figure 5.9: *Left*: Result of the maximum-likelihood fit when fixing parameter B to 1 for high- and low-gain signals of the doublet 72/688. In the bottom plot, the relative difference between the fit and the identity function is displayed for both gain channels. *Right*: Bias between UUB and UB signals for all doublets. The mean bias (coloured dashed lines) for the high- and low-gain channels are $(-2.7 \pm 1.1)\%$ and $(0.4 \pm 1.2)\%$, respectively.

low gain. From this result we can conclude that the signals of the two electronics have an overall linear relation.

To determine the bias between the UUB and UB signals, we performed the maximum-likelihood fit once more, however now with the parameter B fixed to 1. Since the fitted line is providing the best representation of the average signals in the UB and UUB stations, its relative difference to the equality function can be interpreted as the bias of the signals of the two electronics (see bottom plot of Fig. 5.9 (left)). Therefore, when we take Eq. (5.11) with $B = 1$, the signal bias is simply given by $A - 1$. In Fig. 5.9 (right), the signal bias of both gain channels is displayed for each doublet. The mean bias and its standard error are represented by the coloured dashed line and corresponding band, respectively. For the high gain, from where the signals used for the VEM charge calibration are taken, a small bias of $(-2.7 \pm 1.1)\%$ is present between UUB and UB. Since the calibrated signals are the integrated ADC traces divided by the VEM charge, this result suggests that the VEM charge is overestimated for the UUB. Indeed, at the time of writing of this thesis, we verified that the histogram-based calibration of UUB stations are not corrected for the baseline of the atmospheric muon traces in Offline, which could explain the observed small bias. On the other hand, the low-gain

channel has a bias of $(0.4 \pm 1.2)\%$ between the signals of the two electronics. Although the overestimation of the VEM charge for the UUB should also impact the low-gain signals, we also have the systematic biases between high- and low-gain signals for the individual electronics, which are discussed in Refs. [75, 76]. The difference in bias in the high and low gain of Fig. 5.9 (right) is in agreement with the reported systematic differences between signals of these channels in each of the electronics.

5.2.2 SIGNAL TIMING

The signal timing plays a central role in the geometrical reconstruction of showers. Therefore, it is important to show that it is consistent among the UUB stations. Given the different electronics, we expect some timing differences between the UUB and UB. Once the entire surface-detector array is upgraded, such timing differences will not matter for the event reconstruction anymore, but they are important for the interim period of a heterogeneous array when events will have to be reconstructed with signals provided by both electronics. Thus, it is relevant to study also the UB/UUB signal-time differences.

For the following analyses, we also used data from a triplet consisting of the stations 1739, 56 and 59, which are all equipped with UUBs. These stations form an isosceles triangle with a distance of approximately 10 m between stations 56 and 59 and 15 m between 1739 and the other two. We reconstructed their data from 01 January 2022 to 03 February 2023. Since they are in the standard 1500-m array, the number of events that include these stations is significantly smaller than that of the hexagon with the doublets, which we also have used for the timing analysis.

Before we can compare the signal times of any pair of stations, we must correct them for the time that the shower front takes to travel between the two stations. We approximate the shower front as a plane perpendicular to the shower axis so that the time difference Δt_{ij} between the signals of a station pair ij is given by

$$\Delta t_{ij} = t_i - t_j = t_i^{\text{start}} - t_j^{\text{start}} - \Delta t_{ij}^{\text{pf}}, \quad (5.12)$$

where t_i^{start} and t_j^{start} are the start times of the signal, as defined in Ref. [54], and $\Delta t_{ij}^{\text{pf}}$ is the time the plane front takes to travel from station i to station j . If the shower core,

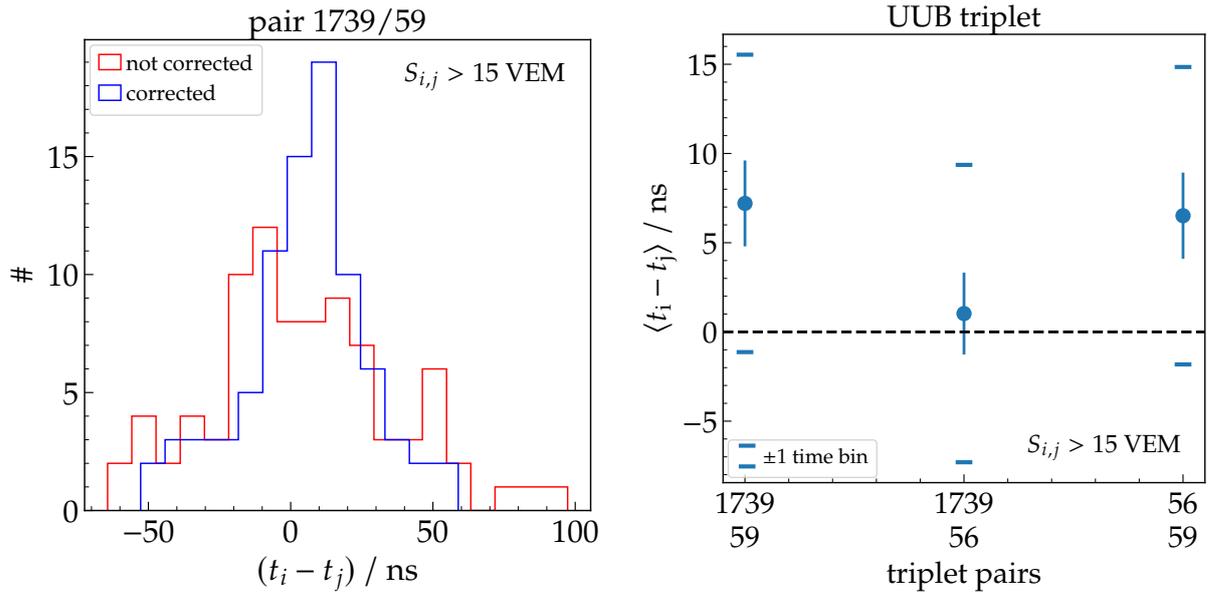


Figure 5.10: *Left*: Distribution of signal time differences $\Delta t_{ij} = t_i - t_j$ for pair $ij = (1739, 59)$ of the UUB triplet stations. The correction for the propagation of the shower front yields a narrower distribution. *Right*: Mean time offsets $\langle \Delta t_{ij} \rangle = \langle t_i - t_j \rangle$ for all possible pairs ij of the UUB triplet. All of them are within one time bin, indicating that no significant offset is present for these stations.

the direction of which is given by the unit vector \hat{a} , hits the detector ground level at time t_c and position \vec{x}_c , then the time that the plane front passes through the station at position \vec{x}_i is

$$t_i^{\text{pf}} = t_c - \frac{1}{c} \hat{a} \cdot (\vec{x}_i - \vec{x}_c), \quad (5.13)$$

where c is the speed of light. Using Eq. (5.13), we can show that $\Delta t_{ij}^{\text{pf}}$ becomes

$$\Delta t_{ij}^{\text{pf}} = \frac{1}{c} \hat{a} \cdot (\vec{x}_j - \vec{x}_i). \quad (5.14)$$

Hence, knowledge about the time and position that the core hits the ground is not necessary for the correction of the signal time difference.

In Fig. 5.10 (left) and Fig. 5.11, we observe the distribution of time differences with and without the shower-propagation correction for station pairs with different electronics setup. The effect of the correction is to narrow the distribution, given that it subtracts from the difference in start times. However, the mean offset should be similar to that of the uncorrected time difference, since the distribution of arrival directions of the showers is symmetric with respect to the local zenith.

In Fig. 5.10 (right), we show the mean time offset calculated with the possible

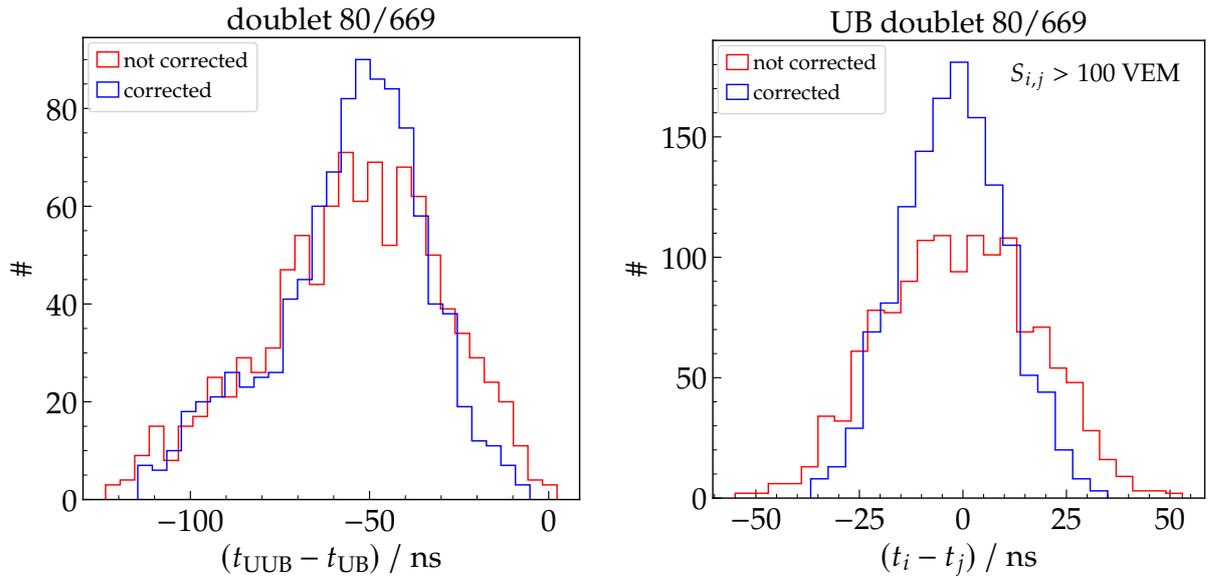


Figure 5.11: Distribution of time differences between the doublet stations 669 and 80 when they were a UUB/UB doublet (left) and when both had UB installed (right). The plot on the right is done with the data of the year 2013. Only signals larger than 100 VEM are used. The left tail in the distribution for the UB/UUB doublet is due to the different sampling rates of the two electronics. For the UB doublet, the distribution is symmetric, and its width reflects the 25-ns resolution of the UB electronics.

pairs of the UUB triplet stations. Since the start time is more precisely determined for larger signals, we perform a cut at 15 VEM, despite the limited statistics. Given the sampling rate of 120 MHz of the UUB electronics, the signal timing of the UUB triplet is consistent, as their differences are all within one time bin with size of $8.\bar{3}$ ns. This is an encouraging result which indicates that the UUB signal timing is correct for use in event reconstructions and it also counts as the validation of the Observatory upgrade. However, it would be beneficial to have this kind of analysis done in the future with a larger number of UUB doublet stations as well as with increased statistics.

As mentioned previously, we are also interested in studying the timing differences between UB and UUB, which is relevant for the reconstruction of events where we have both UUB and UB stations participating. In Fig. 5.11 (left), the distribution of time differences between one of the UB/UUB doublets is displayed. Since the statistics is much more abundant for these doublets, we have increased the signal cut to 100 VEM, which allows for a more precise determination of the signal start time. Although the distribution is mostly Gaussian, a tail is present on the left side. We can understand it as due to the different sampling rates of the two electronics. For the UUB, the size of the trace bin is $8.\bar{3}$ ns, whereas for UB it is 25 ns. Therefore, $t_{\text{UUB}} - t_{\text{UB}}$ is not symmetric

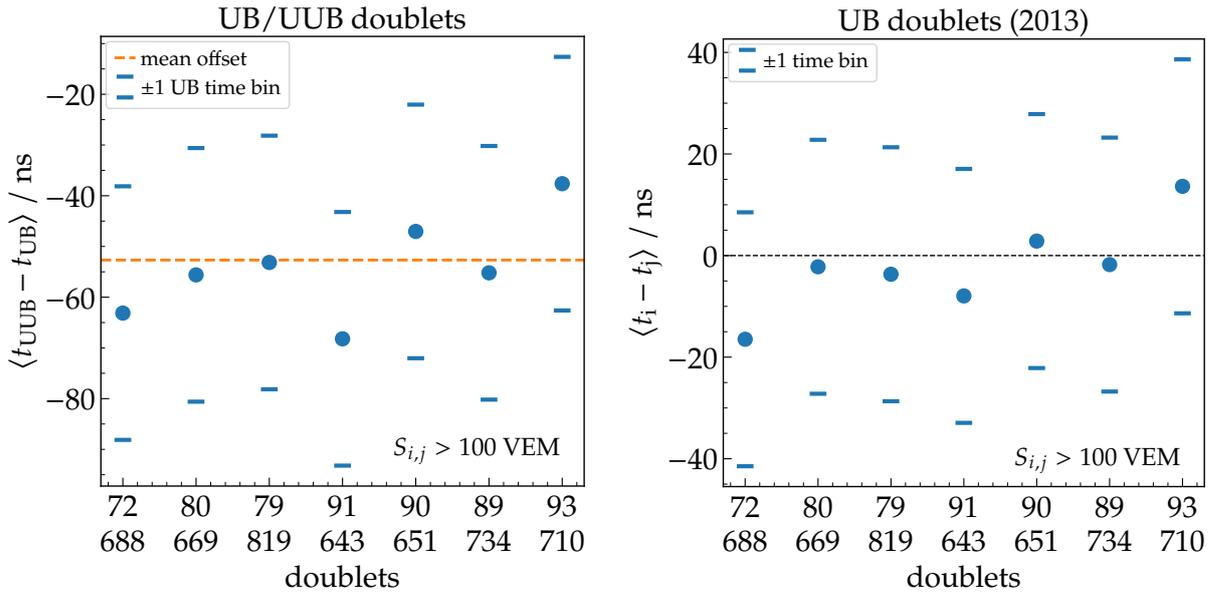


Figure 5.12: Mean time difference for all the UB/UUB doublets (left) and using data from the year 2013 when the stations had UBs (right). In both cases, the differences are compatible within the UB timing resolution. The mean time difference between UUB and UB stations is (-53.3 ± 0.2) ns. The doublet-to-doublet variation is similar in both plots suggesting that it is not mainly caused by the different electronics in the case of the left plot.

under an inversion of start times between the two electronics. When we are comparing the signals of doublet stations with the same electronics (see Fig. 5.10 (left) and Fig. 5.11 (right)), the distribution of the time differences is symmetric. In Fig. 5.11 (right), we have the distribution for the same doublet as in Fig. 5.11 (left). However, it was plotted for data of the year 2013 when both stations had UBs installed. Besides the approximately Gaussian shape, note that the width of the distribution is compatible with the 25-ns resolution of the UB electronics.

In Fig. 5.12 (left), we display the mean time offset for each of the UB/UUB doublets. Since its resolution is determined by the larger time bin of the UB station, we added horizontal bars to represent a ± 1 UB time bin. We verify that the time differences between the UB/UUB doublets are compatible within the resolution of the UB sampling rate. This consistency indicates a synchronised timing of the signals for the stations employing the same electronics. The computed mean time difference is $\langle t_{\text{UUB}} - t_{\text{UB}} \rangle = (-53.3 \pm 0.2)$ ns, which, as previously stated, is anticipated given the different electronics. In the reconstruction of events involving both UB and UUB stations, it is important to note that UUB signals will be earlier in time by about 2

UB time bins. To ensure coherence among all signals and reference them to the same absolute time frame, this temporal difference should be appropriately corrected in the signals of one of the electronics. For comparison, in Fig. 5.12 (right), we present the same analysis as done for the UB/UUB doublets however, we used data from the year 2013, when all the stations composing these doublets had UBs installed. Firstly, we see that their time difference is compatible with zero within the UB resolution, but more importantly, the doublet-to-doublet variation of the time difference is very similar to the one we see when the setup of these stations was changed into the UB/UUB doublets. This means that the variations observed in Fig. 5.12 are not mainly attributed to the different electronics but are likely due to other properties of the WCDs, such as water quality or differences of their PMTs.

CHAPTER 6

Compatibility of event reconstruction

Whilst our analyses, as described in Chapter 5, have confirmed signal compatibility in the water-Cherenkov detectors, our primary concern revolves around the event-level observables derived from shower reconstruction, specifically energy and arrival direction. Thus, in this chapter, we discuss the careful investigation we have undertaken to scrutinise the compatibility of event reconstruction between the SD configuration before and after the upgrade of its stations. Our primary objective is to ensure the absence of any systematic discrepancies that could compromise the data integrity.

In Section 6.1, we provide an overview of the instrumental configuration employed for these investigations, highlighting its performance in reconstructing events. We additionally explain our strategy for selecting events that are utilised in the subsequent analyses. The reconstruction of $S(1000)$ and its corresponding energy is examined in Section 6.2, whilst Section 6.3 is dedicated to the geometric reconstruction of the detected showers. This comprehensive study is crucial to guarantee the seamless transition and continuity between Phase-I and Phase-II data within the Pierre Auger Observatory, thereby preserving the quality and reliability of its dataset.

6.1 SELECTION OF EVENTS

In this study, our analyses were based on data collected by the same hexagon of doublet stations as used for the analyses of signal compatibility, which was described in Sec-

tion 5.2.1 (refer to Fig. 5.4 for a schematic representation of the hexagon configuration). We remind that each doublet consists of one station equipped with the **UB** and the other with the **UUB**, where the **UUB** serves as the master station for the trigger. Notably, this hexagon is situated in the infill region, resulting in the presence of stations from the 750-m array in between. In addition, detectors from a denser array, with a station spacing of 433 m, are in the northwest region of the doublets hexagon. The dataset utilised in our analysis spans from 01 January 2022, when the **UB/UUB** configuration of the doublets was implemented, up to 09 June 2023.

In this study, we performed event reconstructions employing the Offline framework, exclusively utilising stations with the same type of electronics from the doublet hexagon. This approach ensured independent reconstructions of the same events, facilitating a direct comparison between **UB** and **UUB**. In Fig. 6.1, we present the event count reconstructed by each electronics type as a function of time. Notably, when considering all events above 0.1 EeV, a larger number of events were reconstructed by the **UB** hexagon throughout the entire time range. Initially, this observation may appear counterintuitive, given that the **UUB** stations serve as the masters. However, it is important to note that stations from the **SD-750** and **SD-433** arrays are also present within the doublet hexagon. These additional stations can form a T3 trigger, allowing them to incorporate stations from the doublet hexagon provided they have a T1 within a specific time window of the T3 (as described in Section 3.2.3). Consequently, trigger differences between the **UB** and **UUB** may lead to reconstructed events with the **UB** stations without the **UUB** masters triggering. Despite the higher number of events reconstructed by the **UB** hexagon above 0.1 EeV, we observe a similar number of reconstructions for both electronics when considering events above the full trigger efficiency (i.e., events with energy above 3 EeV, as depicted at the bottom of Fig. 6.1).

Furthermore, we observed a more significant disparity in the number of events before 24 August 2022 (indicated by the dark-red dotted line in Fig. 6.1). Upon investigation, we traced this discrepancy to issues with **UUB** station 710 (refer to Fig. 5.4), which experienced difficulties in transmitting its data to the central data acquisition system. Consequently, this problem led to a substantial reduction in event reconstructions within the western side of the **UUB** hexagon, as evident in the top-right histogram of reconstructed positions of the shower cores in Fig. 6.2. After identifying the issue, it

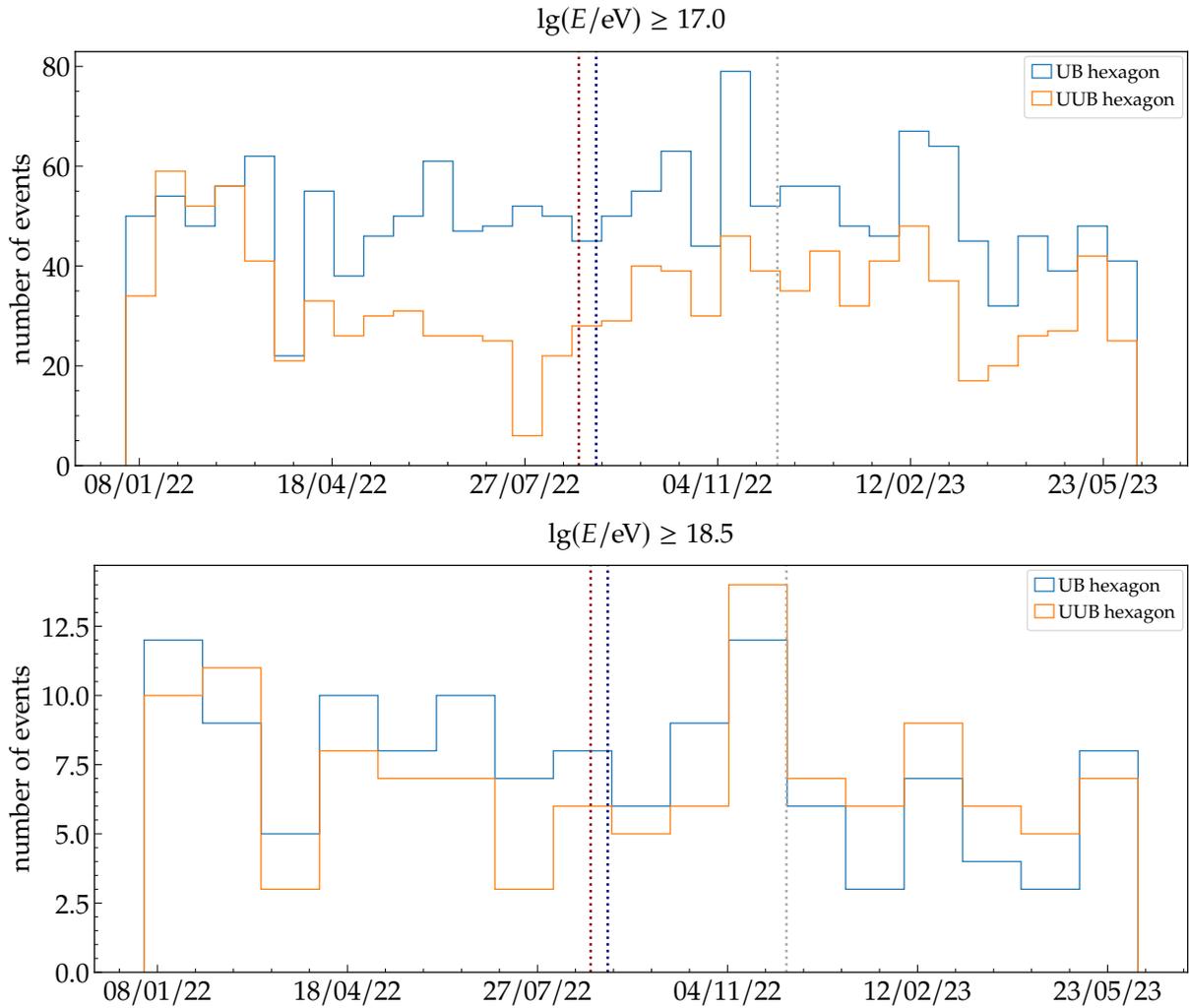


Figure 6.1: Number of events reconstructed separately by the **UB** (blue) and **UUB** (orange) stations of the doublet hexagon as a function of time. At the top, we consider all events, whereas the plot at the bottom includes only events with energy larger than $10^{18.5}$ eV, where the **UB** array reaches its full trigger efficiency. The dark-red dotted line marks the replacement of the communication hardware of **UUB** station 710. The dark-blue and grey dotted lines represent the deployment of DAQ versions V128R0B0P12 and V128R0B0P15, respectively (read text for more details).

was addressed by replacing the communication hardware in station 710. As a result, we observed a uniform distribution of shower core positions in both the **UB** and **UUB** hexagons, as shown in the bottom of Fig. 6.2.

Additionally, it is pertinent to mention two versions of the data acquisition (**DAQ**) software used in the local stations, which also had an impact on the **UUB** triggers. Firstly, an issue with the **GPS** caused the time stamp of the trigger to be offset by 1 second in certain **UUB** stations, resulting in loss of events across the entire **SD** array. To rectify this problem, a solution was implemented in **DAQ** version V128R0B0P12, deployed on

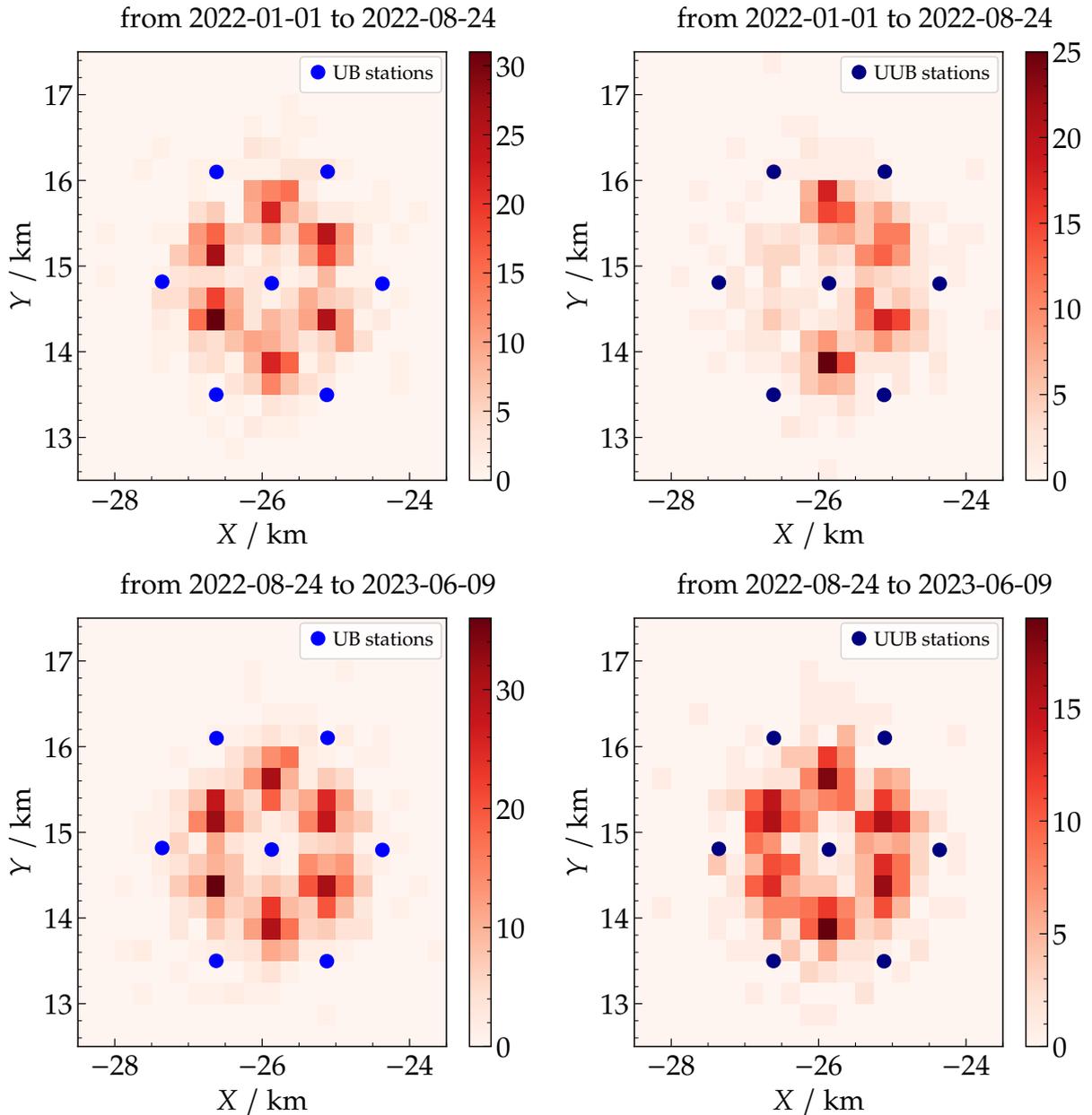


Figure 6.2: Two-dimensional histograms of the position of shower cores reconstructed by the **UB** (left) and **UUB** (right) stations of the doublet hexagon. The plots at the top show the histograms before the replacement of the communication hardware of **UUB** station 710. Note the lack of reconstructed events on the western side of the **UUB** hexagon and how it becomes homogeneous after the fix, in the corresponding histogram at the bottom.

02 September 2022 (indicated by the dark-blue dotted line in Fig. 6.1). Secondly, the DAQ version V128R0B0P15, which was deployed on 05 December 2022 (indicated by the dotted grey line), utilised downsampled traces for calculating the **VEM** peak [77]. Consequently, this alteration changed the rate of threshold T1 and T2, as well as the ToT triggers of **UUB** stations.

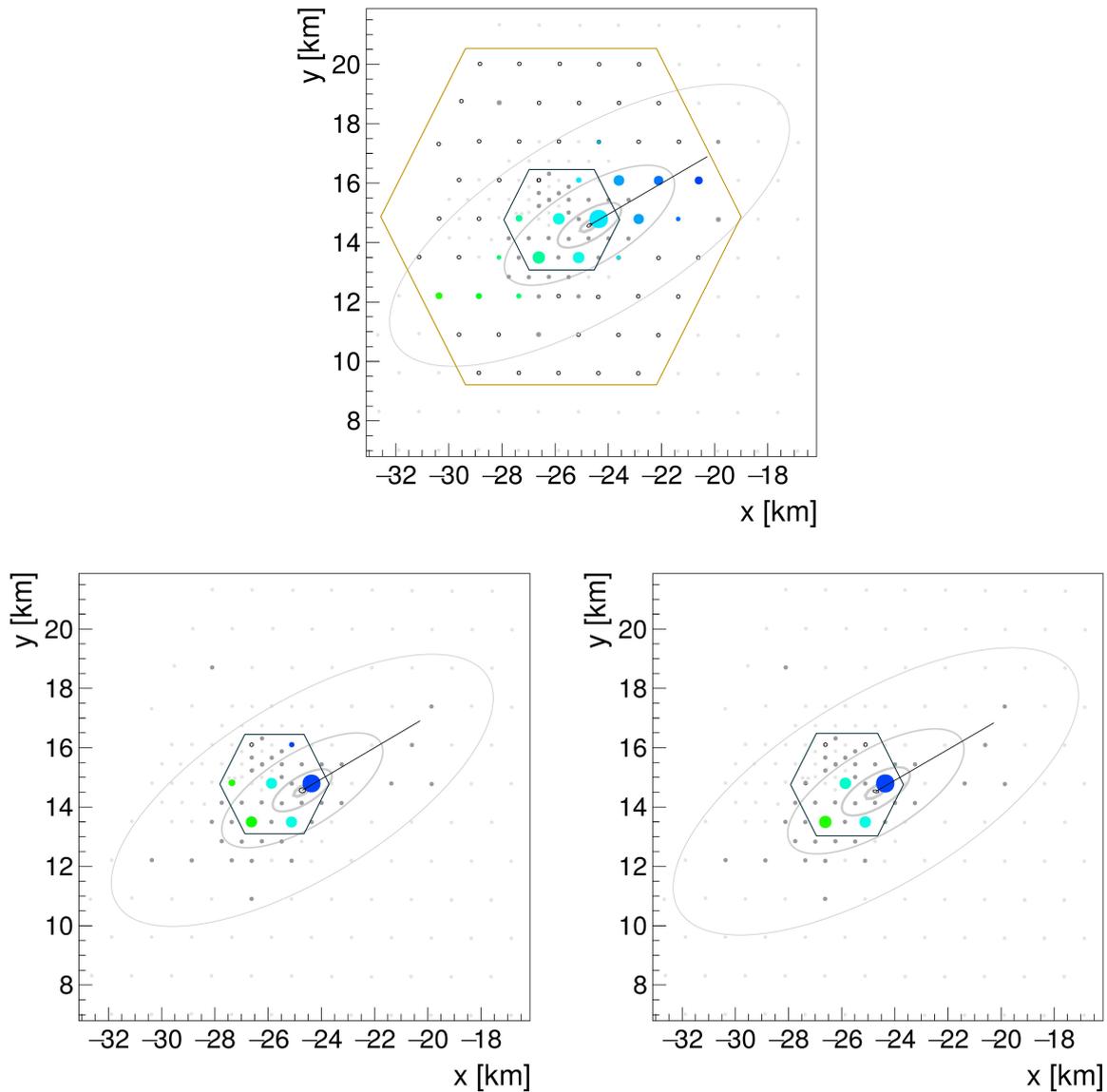


Figure 6.3: Example of an event selected for the compatibility studies. At the top, the event is first reconstructed in the region of the doublet hexagon (inner black hexagon), including the stations of the three crowns (red hexagon) around it. Since the event has its core position inside the doublet hexagon, it is independently reconstructed by the **UB** (bottom left) and **UUB** (bottom right) stations of the doublets. Note that, in this example, the independent reconstructions have different multiplicities.

To ensure unbiased comparisons between reconstructions with **UB** and **UUB** stations, a straightforward independent reconstruction of events using the doublet hexagon is insufficient. Simply excluding the stations around the hexagon may lead to biased reconstructions, especially when the true shower core lies outside the doublet hexagon. A careful event selection process is employed to address this issue.

Initially, all events within the region of the doublet hexagon, including three

crowns of stations from the 1500-m array surrounding it, are reconstructed. From this set of events, those with the shower core located inside the doublet hexagon are identified. By focusing on these events, we prioritise the doublet stations with the largest signals in the event, which play the primary role in the reconstruction, effectively mitigating potential biases. Subsequently, independent reconstructions are performed using each electronics of the doublet hexagon for the selected events. An example of such a selection is illustrated in Fig. 6.3. The top panel shows an event reconstructed with the inclusion of the stations from the three crowns surrounding the doublet hexagon, confirming the presence of the shower core inside it. The bottom panels display the independent reconstructions of the selected event by each electronics of the doublets.

Using this selection procedure, a total of 515 events are obtained. However, it is noteworthy that the independent reconstructions may exhibit different multiplicities for individual events, stemming from trigger differences between **UB** and **UUB**, as well as statistical fluctuations due to the number of particles and the inherent detection processes. To address these scenarios, a further selection is made, considering only events where the stations from the same doublets participate in the reconstruction. This process reduces the statistics to 313 events. Nevertheless, extensive verification demonstrates that most of the results of subsequent analyses are consistent when considering all events or only those with the same multiplicity. Consequently, we present the results using all events to ensure higher statistical significance. In cases where the use of events with the same multiplicity significantly impacts the results, we explicitly specify their usage in the analysis.

6.2 SHOWER-SIZE AND ENERGY ESTIMATION

We initiate the comparison of reconstructed observables between the **UB** and **UUB** stations by analysing the two closely related quantities $S(1000)$ and energy (their correlation was previously exposed in Section 3.3). Consequently, we anticipate the compatibility results for $S(1000)$ and energy to be comparable, assuming equivalent zenith-angle reconstructions, which will be discussed in Section 6.3.

In Fig. 6.4, we display the distributions of $S(1000)$ and energy as reconstruc-

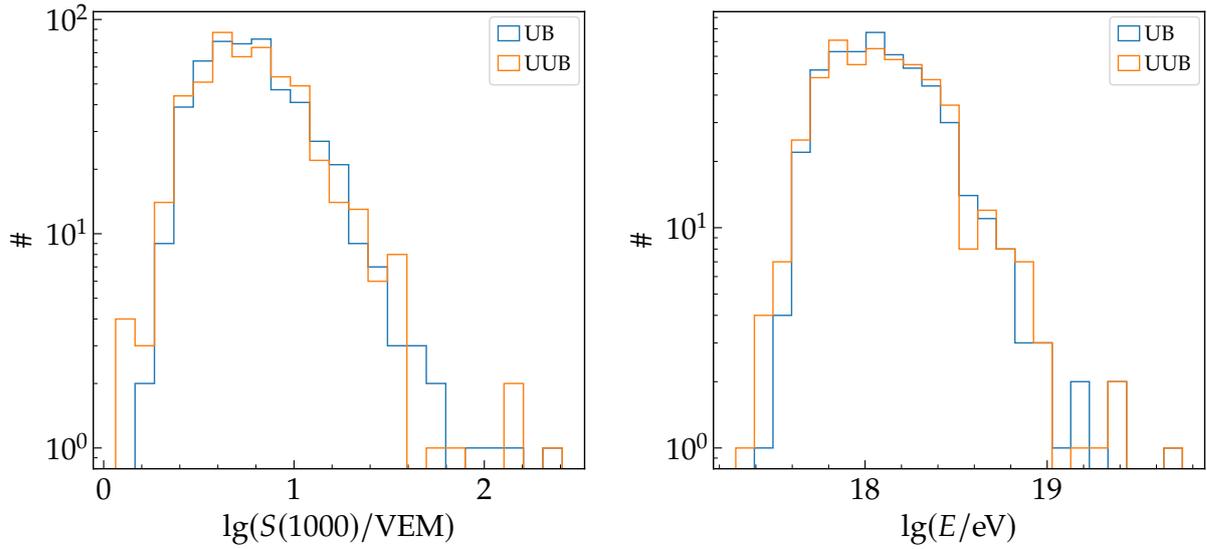


Figure 6.4: Distribution of $S(1000)$ (left) and energy (right) reconstructed using the **UB** (blue) and **UUB** (orange) hexagons. The resemblance observed between the histograms obtained from the two electronics serves as a preliminary indication of their compatibility. Notably, the majority of the analysed events exhibit low energy, with $S(1000)$ values falling below 10 VEM .

ted by each electronics. The striking similarity between the distributions of the two electronics serves as a promising initial indicator of their compatibility. It is worth noting that a significant portion of the reconstructed events exhibit $S(1000)$ values smaller than 10 VEM , corresponding to a majority of low-energy events, particularly below $10^{18.2} \text{ eV}$. For energies exceeding $10^{18.5} \text{ eV}$, which is above full trigger efficiency of the **UB** array, we have a total of 47 events in all selected reconstructions, with only 15 of these events having the same multiplicity. While having a larger statistics of events above full trigger efficiency would significantly enhance the compatibility studies, the current dataset has enabled us to draw meaningful and insightful conclusions, as we shall demonstrate.

A more direct comparison of these quantities is presented in Fig. 6.5, where we analyse their relative differences. Notably, both $S(1000)$ and energy display narrower distributions for larger cuts on these quantities, as expected due to reduced variation at higher values. The mean relative differences, which characterise the bias between the reconstructions of the two electronics, are found to be less than 5% for both $S(1000)$ and energy. In the bottom part of Fig. 6.5, we plot the relative differences as a function of the corresponding mean $S(1000)$ and energy values. Here, the aforementioned decrease in fluctuation for larger values of these quantities becomes evident. For energies above

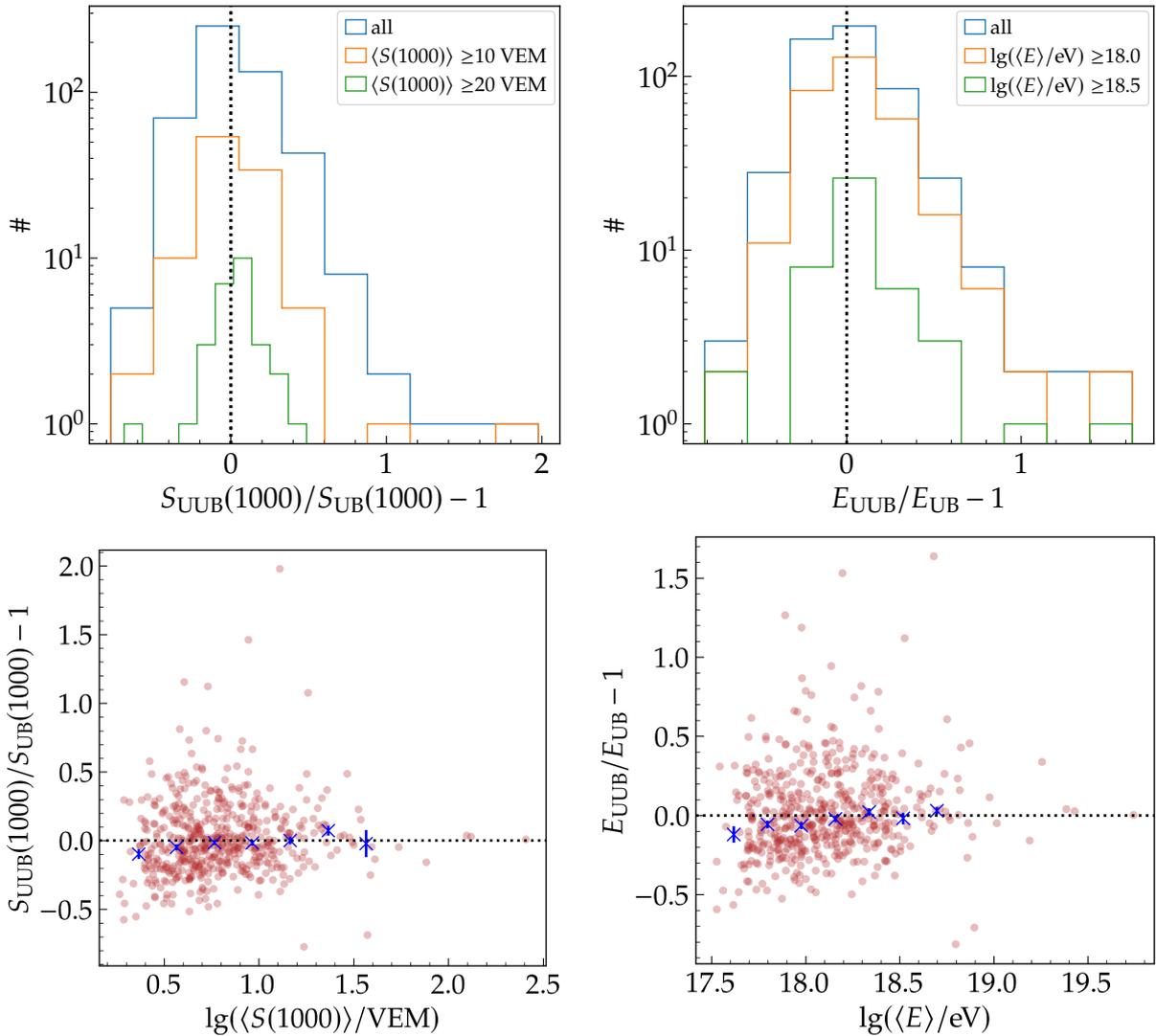


Figure 6.5: The top panels show histograms of the relative difference of $S(1000)$ (left) and energy (right) reconstructed with the two electronics. Different cuts on these quantities are applied, and in all cases, the bias is smaller than 5%. In the bottom panels, the relative difference of $S(1000)$ and energy is plotted as a function of the mean reconstructed $S(1000)$ and energy, respectively. Whilst some bias is observed for small values of these quantities, the majority of the data points are nearly unbiased within the available data range.

approximately $10^{18.15}$ eV, we observe an unbiased reconstruction of energy using both electronics. However, below this energy threshold, a bias towards smaller energy values with **UUB** is present. A similar situation is observed for $S(1000)$. These biases at low values of $S(1000)$ and energy are likely attributed to the trigger differences between the two electronics, as mentioned previously. Additionally, the bias of 2.7% we found between high-gain signals of **UB** and **UUB** stations (see Fig. 5.9), in our performed study on station-level compatibility could also contribute to the observed bias for low values

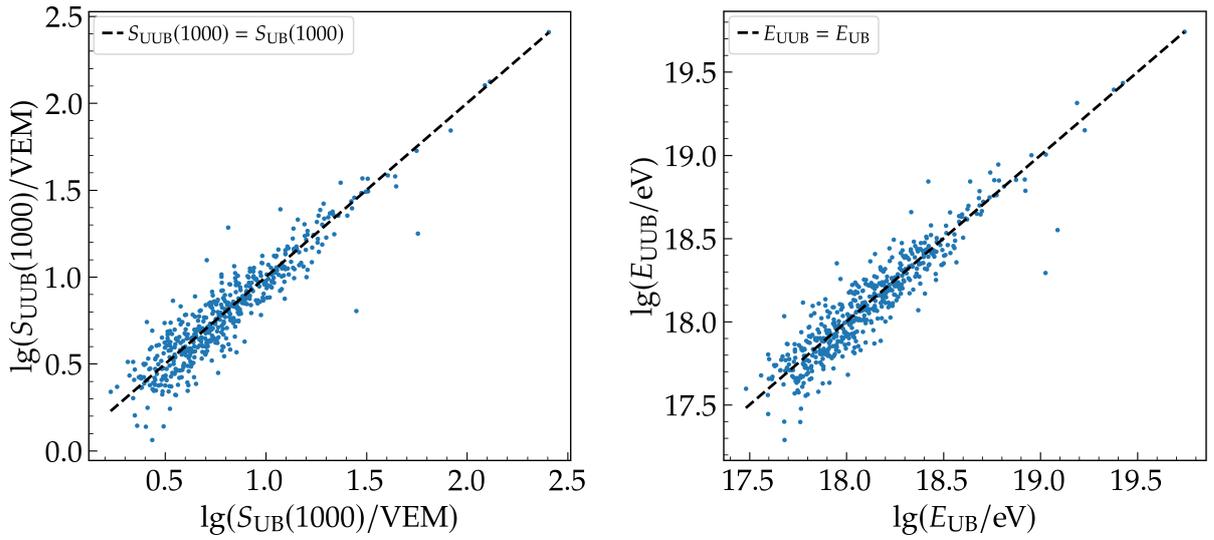


Figure 6.6: Scatter plots showing the **UUB:UB** space for the reconstructed $S(1000)$ (left) and energy (right). The data fluctuates around the equality lines, indicating compatibility of the two electronics. The two outlier points in both plots are likely due to the use of online **VEM** calibration in these events (see text and corresponding footnote).

of $S(1000)$ and energy.

As an additional means of validation, the corresponding scatter plots are depicted in Fig. 6.6. Generally, the reconstructed quantities obtained using the different electronics exhibit compatibility, as evidenced by their fluctuations around the presented equality lines. Notably, it should be mentioned that both plots contain two outlier points, whose occurrence can likely be attributed to the utilisation of online¹ **VEM** estimations during signal calibration. Nevertheless, further investigation is beneficial to fully comprehend these deviations. Moreover, a constructive approach to enhance the analysis involves employing a maximum-likelihood fit to account for the trigger differences between the electronics, hence providing a more precise estimation of the biases. Consequently, a simple fit was not performed in this study for this particular reason.

As pointed out earlier, the width of the distributions in the top panel of Fig. 6.5, along with the fluctuations observed in the bottom panel and around the equality lines in Fig. 6.6, can be attributed to the uncertainty associated with the meas-

¹Despite the demonstration of the **VEM** calibration being carried out with the calibration histograms in over 99.5% of the cases, as presented in Section 5.1.3, there exists a possibility that a few calibrations utilised the online **VEM**, which is acknowledged to be inaccurate by a factor of 2 during the period of the analysed data.

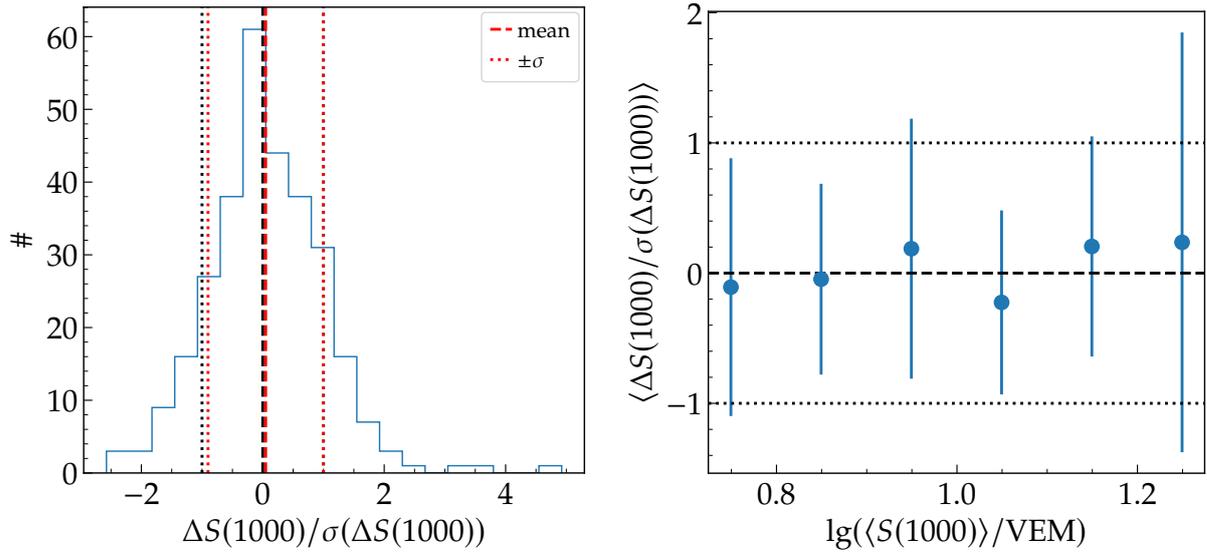


Figure 6.7: Left: distribution of $\Delta S(1000)/\sigma(\Delta S(1000))$. Its mean (dashed red line) and standard deviation (dotted red line) are respectively compatible with zero (dashed black line) and one (dotted black line). Right: Mean $\Delta S(1000)/\sigma(\Delta S(1000))$ as a function of the mean reconstructed $S(1000)$. The error bars represent the standard deviation, which is close to one in the presented range.

urement of $S(1000)$ and energy. Consequently, it is crucial to assess whether the differences in the reconstruction of these quantities with the two electronics align with the expected uncertainties. Given that the uncertainty in $S(1000)$ propagates to the energy reconstruction, we will focus this analysis on $S(1000)$.

The uncertainty on $S(1000)$, denoted as $\sigma(S(1000))$, is comprised of both statistical and systematic contributions [54]. The statistical uncertainty, $\sigma_{\text{stat}}(S(1000))$, arises from the maximum-likelihood fit of the station signals to the lateral distribution function. This component is mainly affected by the uncertainty in the signals of individual stations and the event multiplicity. On the other hand, the systematic uncertainty, $\sigma_{\text{syst}}(S(1000))$, stems from the lack of precise knowledge of the true LDF and is characterised by the uncertainty in its parameter β . Combining both contributions, the total uncertainty of $S(1000)$ is given by the equation

$$\sigma^2(S(1000)) = \sigma_{\text{stat}}^2(S(1000)) + \sigma_{\text{syst}}^2(S(1000)). \quad (6.1)$$

For this analysis, we use the absolute difference between $S(1000)$ reconstructed with the two electronics, $\Delta S(1000) = S_{\text{UB}}(1000) - S_{\text{UUB}}(1000)$. Since the reconstruction with UB and UUB are independent, the uncertainty for each, obtained from

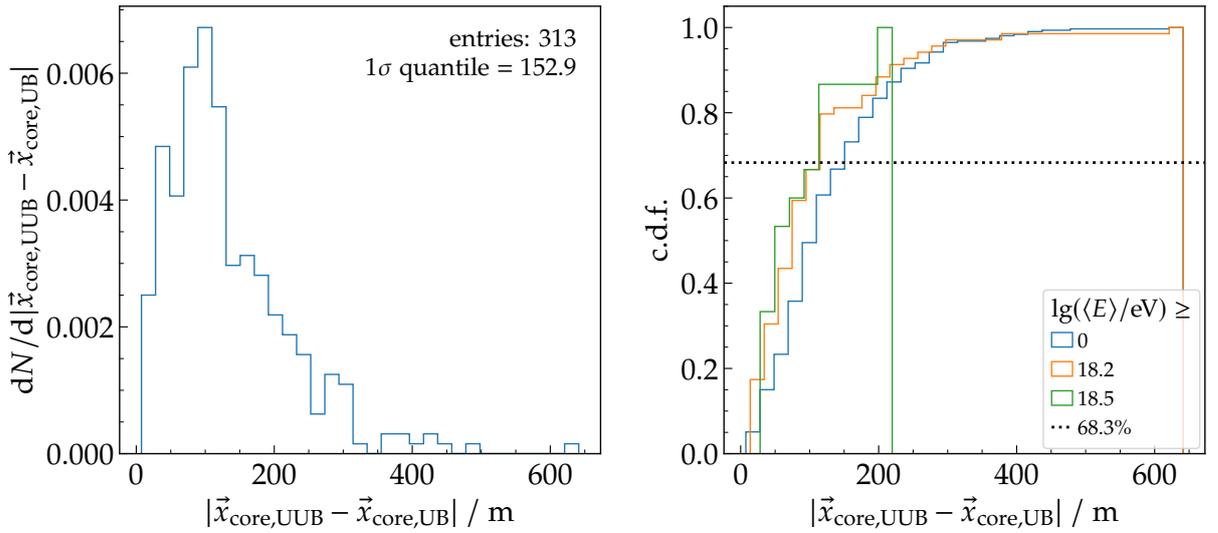


Figure 6.8: Normalised distribution (left) of the difference between reconstructed position of the shower core with **UUB** and **UB**, as well as its corresponding cumulative distribution function (right). The dotted line represents 68.3% which defines its resolution between the two electronics. The resolution improves from 153 m, with no applied cuts, to 116 m and 106 m when employing cuts of $10^{18.2}$ eV and $10^{18.5}$ eV, respectively.

Eq. (6.1), can be added in quadrature

$$\sigma^2(\Delta S(1000)) = \sigma_{\text{UUB}}^2(S(1000)) + \sigma_{\text{UB}}^2(S(1000)). \quad (6.2)$$

Fig. 6.7 (left) displays the distribution of $\Delta S(1000)/\sigma(\Delta S(1000))$, revealing an initial observation of the mean value in proximity to zero, which concurs with the expected outcome, considering the previous demonstration of unbiased reconstruction with the distinct electronics. A crucial aspect to highlight is the proximity of the standard deviation of the distribution to unity, indicating that the variations in $\Delta S(1000)$ are consistent with their corresponding uncertainties. In other words, the discrepancies observed between $S(1000)$ with **UB** and **UUB** fall within the range of their anticipated uncertainties. This consistency extends across the entire range of available reconstructed $S(1000)$, as further depicted in Fig. 6.7 (right).

6.3 SHOWER-GEOMETRY RECONSTRUCTION

The geometry reconstruction of showers holds important significance in various studies, including investigations of anisotropy and the identification of potential sources of

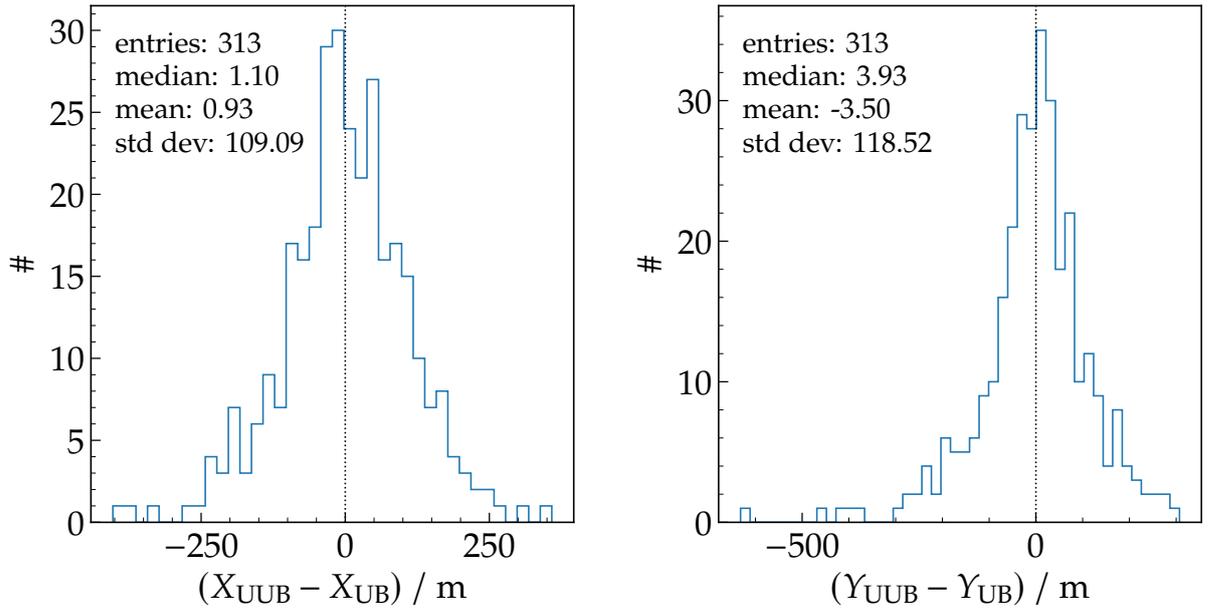


Figure 6.9: Comparison of the reconstructed coordinates, X (left) and Y (right), for the shower core position between the **UUB** and **UB** hexagons. The coordinates are defined within the shower plane of the **UB** reconstruction, with the y-axis oriented towards the upstream direction of the shower. Both mean values closely approximate zero, signifying an unbiased reconstruction between the two electronics.

high-energy cosmic rays. Consequently, it becomes imperative to undergo a thorough validation to ensure that the new electronics yield equivalent quantities, such as the shower-core position, its axis, and the arrival direction of the primary particle when compared to the results obtained from the old electronics. This validation process further establishes the reliability and consistency of the data and supports accurate future analyses with AugerPrime.

We initiated our investigation of geometric compatibility by examining the position of the impact point of the shower axis at the detector level, which we will hereafter refer to as the shower core position. It is important to note that, in the detector plane, even slight angular disparities in the reconstructed shower axis can lead to significant variations in the position of the shower core, especially for more inclined events. Consequently, such discrepancies may appear more pronounced than they actually are. To ensure a fair and unbiased comparison, we calculate the difference in the shower core position within the shower plane of the reconstruction using the **UB** stations. Furthermore, this analysis is conducted on events with the same multiplicity, as the number of stations significantly influences the determination of the shower core.

The normalised distribution depicting the absolute difference between the

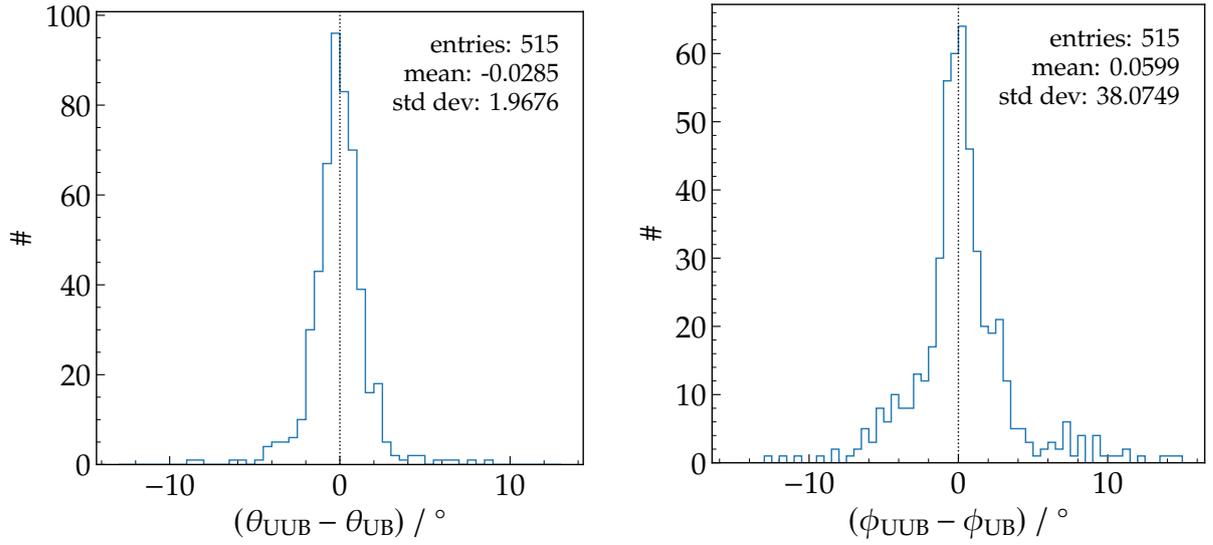


Figure 6.10: Distribution of the difference in the reconstruction of zenith (left) and azimuth (right) angles between **UUB** and **UB** stations. Both distributions exhibit mean values in close proximity to zero, indicating a compatible reconstruction of these angles using the distinct electronics. It is noteworthy that the distribution of azimuth angles displays a larger standard deviation, which can be attributed to nearly-vertical events where the azimuth angle becomes degenerate.

position of the reconstructed shower cores obtained using the two distinct electronics, along with its corresponding cumulative distribution function, is presented in Fig. 6.8. To quantify the resolution of the shower core between **UB** and **UUB**, we employ the 68.3% quantile (1σ) of their difference distribution, represented as the dotted black line in the cumulative plot. Notably, it is observed that the resolution displays an improvement with increasing energy cuts. Without any applied cuts, the resolution stands at 153 m, whilst with energy cuts set at $10^{18.2}$ eV and $10^{18.5}$ eV, the resolution enhances to 116 m and 106 m, respectively.

In Fig. 6.9, the disparity in the x and y coordinates of the shower cores, reconstructed utilising the different electronics, is presented. Once again, it is essential to note that these coordinates exist within the plane of the shower reconstruction accomplished using the **UB** hexagon, where the positive y-axis aligns with the upstream direction of the shower. Notably, the proximity of the distribution to a mean of zero signifies that the **UUB** achieves an unbiased reconstruction of the shower core in comparison to its **UB** counterpart.

In this next investigation, we have examined the compatibility concerning the arrival direction, characterised by the zenith and azimuth angles of the reconstructed

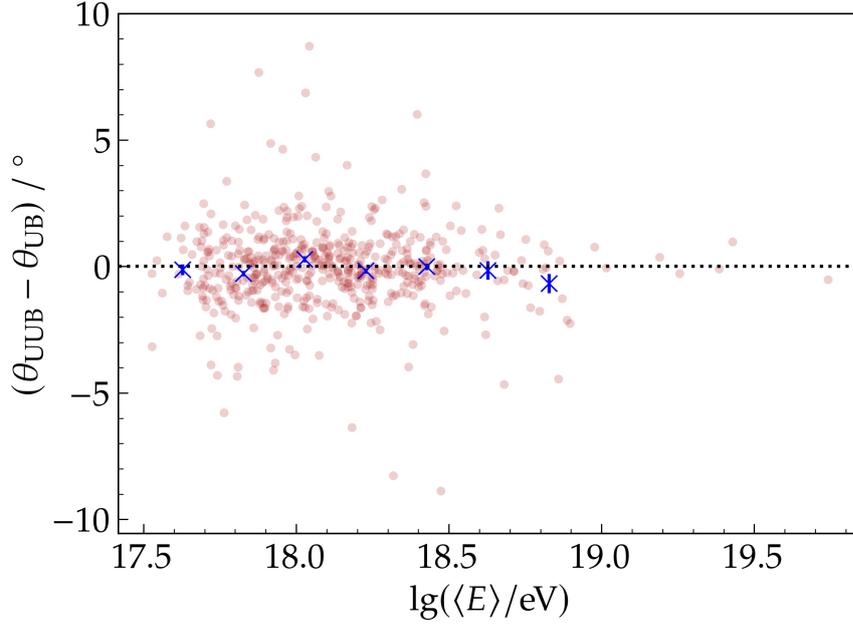


Figure 6.11: Difference in zenith angle reconstructed using **UUB** and **UB** hexagons plotted against the mean reconstructed energy. The blue markers denote mean values for bins with a width of 0.2 in the base-10 logarithm of the energy. Notably, the **UUB** demonstrates unbiased results across the range of available energies.

shower axis. The corresponding differences in the reconstructions provided by the distinct electronics are displayed in the left and right panels of Fig. 6.10. Both distributions have their mean value in proximity to zero, also indicating an unbiased reconstruction of the arrival direction. However, it is worth noting that the zenith-angle distribution displays a standard deviation of approximately 2° , whilst the azimuth-angle distribution exhibits a larger standard deviation of 38° . This significant deviation in azimuth angles can be attributed to nearly vertical showers, where the azimuth angle becomes degenerate.

As an additional verification, Fig. 6.11 portrays the difference in the zenith angle as a function of the mean reconstructed energy. For larger energies, the fluctuation in the difference diminishes, in line with the expectation that improved reconstruction occurs at higher energies due to the increased multiplicity of these events. Furthermore, we emphasise that the reconstruction of the zenith angle using the **UUB** stations remains unbiased across the entire range of reconstructed energies available.

In this next study, we address the disparities in the reconstruction of the shower axis achieved by each of the electronics. To accomplish this objective, we introduce the angle η between the shower axes of the two reconstructions using the

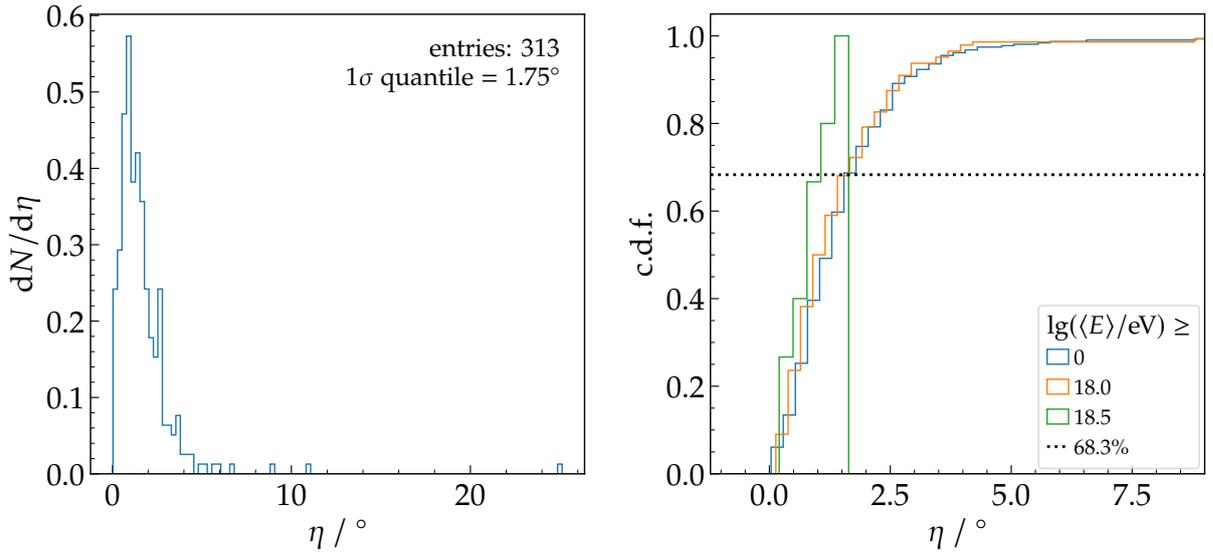


Figure 6.12: Normalised (left) and corresponding cumulative (right) distributions of the angle η , as defined in Eq. (6.3). Different cuts on energy are applied in the cumulative distribution. The angular resolution, defined as the 1σ quantile of the η distribution, is initially 1.8° without any applied cuts. Subsequently, it improves to 1.7° and 1.1° for cuts on 1 EeV and 3 EeV, respectively. Only events reconstructed with the same multiplicity are considered in these distributions.

expression

$$\sin \eta = |\hat{a}_{\text{UUB}} \times \hat{a}_{\text{UB}}|, \quad (6.3)$$

where $\hat{a}_{\text{UUB(UB)}}$ represents the unit vector pointing in the direction of the reconstructed shower axis with the **UUB** (**UB**) hexagon.

Fig. 6.12 depicts the normalised distribution of η and its corresponding cumulative distribution function. It is important to emphasise that we consider only events with the same multiplicity in these plots. Similar to the treatment of the shower core position, we define the angular resolution between the two electronics as the 1σ quantile of the η distribution. This value corresponds to the point at which the cumulative distribution reaches 68.3%, as indicated by the dotted black line. Notably, the resolution improves from 1.8° (when no cuts are applied) to 1.7° and 1.1° for cuts above 1 EeV and 3 EeV, respectively.

Moving on to Fig. 6.13, we analyse the dependency of the angular resolution on the zenith angle. In this analysis, we utilise all events selected to ensure sufficient statistical significance for binning the data in $\sin^2 \theta$. Evidently, we observe an enhancement in resolution for more inclined showers, which is likely associated with the larger multiplicity of such events. Furthermore, there are significant improvements for larger

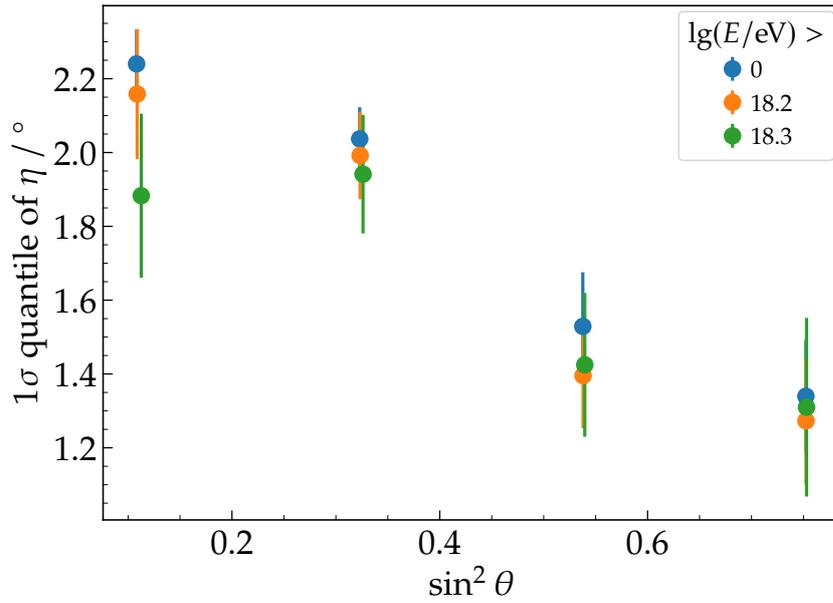


Figure 6.13: Angular resolution as a function of the zenith angle of the shower. All 515 selected events are included to ensure sufficient statistics for binning in $\sin^2 \theta$. The angular resolution exhibits enhancements for larger zenith angles, as well as for higher energy cuts, with a more pronounced effect observed for less inclined showers.

energies, especially in showers with smaller zenith angles.

In the final analysis of this chapter, we aimed to estimate the angular resolution of the reconstructions with **UUB** stations. Considering that the reconstructions with **UB** and **UUB** are independent, the measured angular resolution (AR) with the **UB/UUB** doublets can be related to the resolutions of each electronics through the expression

$$\text{AR}_{\text{UB:UUB}}^2 = \text{AR}_{\text{UB}}^2 + \text{AR}_{\text{UUB}}^2. \quad (6.4)$$

Therefore, we can estimate the angular resolution of the **UUB** if we also have information about the resolution of the **UB**. Fortunately, the angular resolution of **UB** reconstructions was previously estimated in a study of the Collaboration [54], where three data-driven methods were employed. Among these methods, the one that yielded the best angular resolution utilised doublet stations, both equipped with **UB**. Consequently, we chose the result from this approach for our analysis.

In the aforementioned Ref. [54], the angular resolution is also provided as a function of $\sin^2 \theta$, similar to Fig. 6.13. Therefore, to estimate the angular resolution of **UUB** reconstructions, we applied a linear fit to both sets of data for **UB:UUB** and **UB:UB** doublets, which were subsequently used in Eq. (6.4).

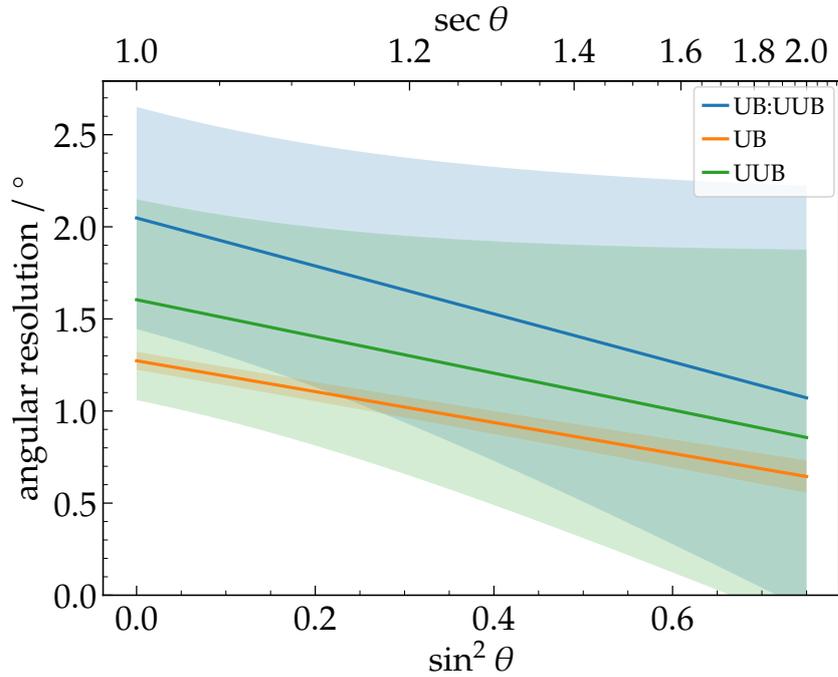


Figure 6.14: Angular resolution as a function of zenith angle. The blue and orange lines represent the linear fits to the data obtained from **UB:UUB** and **UB** doublets, respectively. The red line denotes the estimated angular resolution for **UUB**, derived from Eq. (6.4). The results show that the angular resolutions of **UB** and **UUB** are statistically compatible with the current dataset.

For the resolution study with **UB** stations, 913 events were utilised, and a cut at 3 EeV, above the full trigger efficiency of the array, was applied. As previously described in Section 6.1, our data selection for **UB/UUB** doublets comprises 515 events, of which only 47 are above 3 EeV. Consequently, we could only apply a cut at 2 EeV, resulting in 126 events with sufficient statistical significance for binning in $\sin^2 \theta$.

The results for the angular resolution of **UUB**, along with those of **UB:UUB** and **UB:UB** doublets, are presented in Fig. 6.14. Whilst the **UB** resolution may seem better than that of **UUB**, the former falls within the statistical uncertainty of the latter. Therefore, we can only conclude that the angular resolution of both electronics is statistically compatible with each other, considering the current dataset. Furthermore, our analysis supports the indication of a potential tendency towards better angular resolution with the **UUB** stations. This tendency can be mainly justified by the fact that a less energetic cut is applied for the **UB/UUB** doublets. Therefore, a significant improvement of the resolution is expected by utilising events with the larger energy cut. Whilst the current dataset offers valuable insights, a more definitive conclusion

regarding this trend would necessitate a larger dataset from UB/UUB doublets. The availability of such a dataset would not only reduce the statistical uncertainty but also enable a more robust and assertive result.

CHAPTER 7

Simulations

The investigations presented in Chapters 8 and 9 necessitate in-depth simulations of extensive air showers, followed by their detection and reconstruction employing the enhanced AugerPrime stations. These simulations are crucial for exploring the physics potential offered by the upgraded detectors. In this concise chapter, we provide an overview of the computational tools essential to conducting these simulations for our research studies.

Section 7.1 delves into the software used for simulating extensive air showers, highlighting specific parameters tailored to meet our research requirements. In Section 7.2, we outline the process of simulating the response of the AugerPrime surface detector to the generated particle cascades, accomplished through the use of the Offline framework. Lastly, in Section 7.3, we discuss the execution of our simulations using the available computational infrastructure, as well as our utilisation of external libraries to enhance our studies.

7.1 SIMULATION OF EXTENSIVE AIR SHOWERS

For simulating extensive air showers (EAS), we have chosen the software **CORSIKA** (COsmic Ray Simulation for KAscade) [19], originally developed in the context of the KASCADE [78] experiment in Karlsruhe, Germany. Over time, it underwent improvements, transforming it into a versatile tool for EAS simulations, widely adopted by

various research groups for interpreting their experimental data.

CORSIKA employs Monte Carlo calculations, tracking every particle generated during the development of the cascade, accounting for their physical interactions. These interactions encompass ionisation energy loss, the decay of unstable particles, hadronic interactions among cascade particles and air molecules, multiple scattering, and electromagnetic interactions like Bremsstrahlung emission and pair production. The simulation also incorporates factors such as the Earth's magnetic field, atmospheric models, and the potential to record the production of Cherenkov radiation by **EAS** particles.

The software comes with a compilation script that enables users to choose the hadronic models for both low and high-energy interactions. Once an executable is generated, numerous options can be specified through a *steering file*. This file configures the conditions and settings for the simulated extensive air showers. These parameters encompass details such as the type and energy of the primary particle, its zenith and azimuth angles, the number of showers for simulation, the altitude of the observation level, and so forth. When a simulation run is completed, the program outputs a binary file containing, for each shower, the information regarding position, momentum and arrival time of each particle that reaches the observation plane.

To investigate the reconstruction of signal components within the AugerPrime stations, as detailed in Chapter 8, it is crucial to understand how these components correlate with specific shower properties such as energy and zenith angle. To facilitate these analyses, we opted to create a library of simulated extensive air showers with fixed values of energy and zenith angle.

The energy range was chosen to align with the highly energetic events probed by the Pierre Auger Observatory, particularly in the region of cosmic ray flux suppression. Thus, we simulated primaries with fixed energies using their base-10 logarithm ($\lg = \log_{10}$) as $\lg(E/\text{eV}) = [18.0, 18.5, 19.0, 19.5, 20.0]$. Regarding the zenith angles, we selected the values $[0, 8^\circ, 18^\circ, 28^\circ, 38^\circ, 48^\circ, 58^\circ]$. Moreover, to scrutinise the potential for composition separation, our simulation library contains primaries of proton and iron nuclei. For each combination of primary particle species, energy and zenith angle, we have simulated 80 showers.

We utilised version 7.6900 of the **CORSIKA** software for our simulations.

Table 7.1: Parameters and their corresponding values in our simulations with the **CORSIKA** software. These options are defined in the steering file passed as input for the simulation program. A description of each of the parameters is presented.

| Option name | Value | Description |
|-------------|----------------------------|--|
| OBSLEV | 1400×10^2 | Altitude (in cm) of the observation plane with respect to the sea level. We use that of the Pierre Auger Observatory. |
| MAGNET | 19.71, -14.18 | Earth's magnetic field (in μT), where the first value is the horizontal component pointing to the North and the second is the vertical component downwards. The presented values correspond to Malargüe, Argentina. |
| ATMOD | 18 to 29 | Model of the atmospheric profile. Randomly chosen amongst the indicated values, which correspond to each month of the year from January to December, respectively, in Malargüe, Argentina. |
| ECUTS | 0.1, 0.1, 0.00025, 0.00025 | Lower limit for the energy (in GeV) that defines when a particle is no longer tracked. The values presented are respectively for hadrons, muons, electrons and photons. |
| ELMFLG | False True | Flags to enable or disable the treatment of the electromagnetic cascade analytically with the NKG approach (first flag) or utilising a Monte Carlo treatment employing the package EGS4 [84] (second flag). We opted for the latter exclusively. |

For high-energy hadronic interactions, we employed the EPOS-LHC model (version 3400) [20, 79, 80], whilst low-energy interactions were modelled using FLUKA [81–83]. The transition between the two models occurs at 80 GeV. The azimuth angles were randomly selected from the complete interval $[-180^\circ, 180^\circ]$ at the start of each simulation run. In Table 7.1, we present various options used in our steering files, along with descriptions for each.

It is important to emphasise that these simulations are highly time-consuming and demand substantial storage space. As an example, a test simulation of one hundred EASs generated by vertical protons with an energy of 10^{16} eV took approxi-

ately 102 hours to complete, producing an output file of 25.8 GB. Given the significantly larger energy interval required for our studies (two to four orders of magnitude larger), we anticipated a substantial increase in both simulation duration and total storage size.

To address these challenges, our simulations employ a thinning algorithm designed to reduce both simulation time and data volume to manageable levels. This algorithm involves tracking only one of the product particles emerging from an interaction and assigning it a corresponding weight when the energy of the particle falls below a specified threshold. Within the **CORSIKA** framework, this is achieved through the **THIN** option. The first argument of this option sets the threshold energy fraction, and the second argument determines the maximum weight factor assigned to particles.

In our simulations, we set the threshold energy fraction to 10^{-6} and the maximum weight to be the threshold value multiplied by the numeric value of the primary particle energy, as recommended in the **CORSIKA** User's Guide [85]. This approach substantially reduces computational resources whilst maintaining acceptable accuracy in the description of the particle cascades.

7.2 SIMULATION OF EVENT RECONSTRUCTION

After generating the simulated extensive air showers, our research necessitates simulating how the AugerPrime surface detector responds to these events. To simulate the response of the upgraded detectors to the resulting particles from the generated **EASs**, we employed the **Offline** framework. The framework provides a specific module that implements the simulation of both the water-Cherenkov detector and scintillator surface detector using the Geant4 toolkit [86, 87].

This module defines the geometry and materials that compose both the **WCD** and **SSD**, as illustrated in the top panels of Fig. 7.1. The Geant4 package simulates the passage of particles through the AugerPrime detectors, accounting for the various physical processes involved. However, it is important to note that this simulation tool is used differently for the **WCD** and **SSD** to extract their corresponding signals.

In the case of the **WCD**, the implemented simulation tracks the Cherenkov photons produced within the water volume (see Fig. 7.1, bottom). Subsequently, the arrival time distribution of these photons at the location of the windows of the three

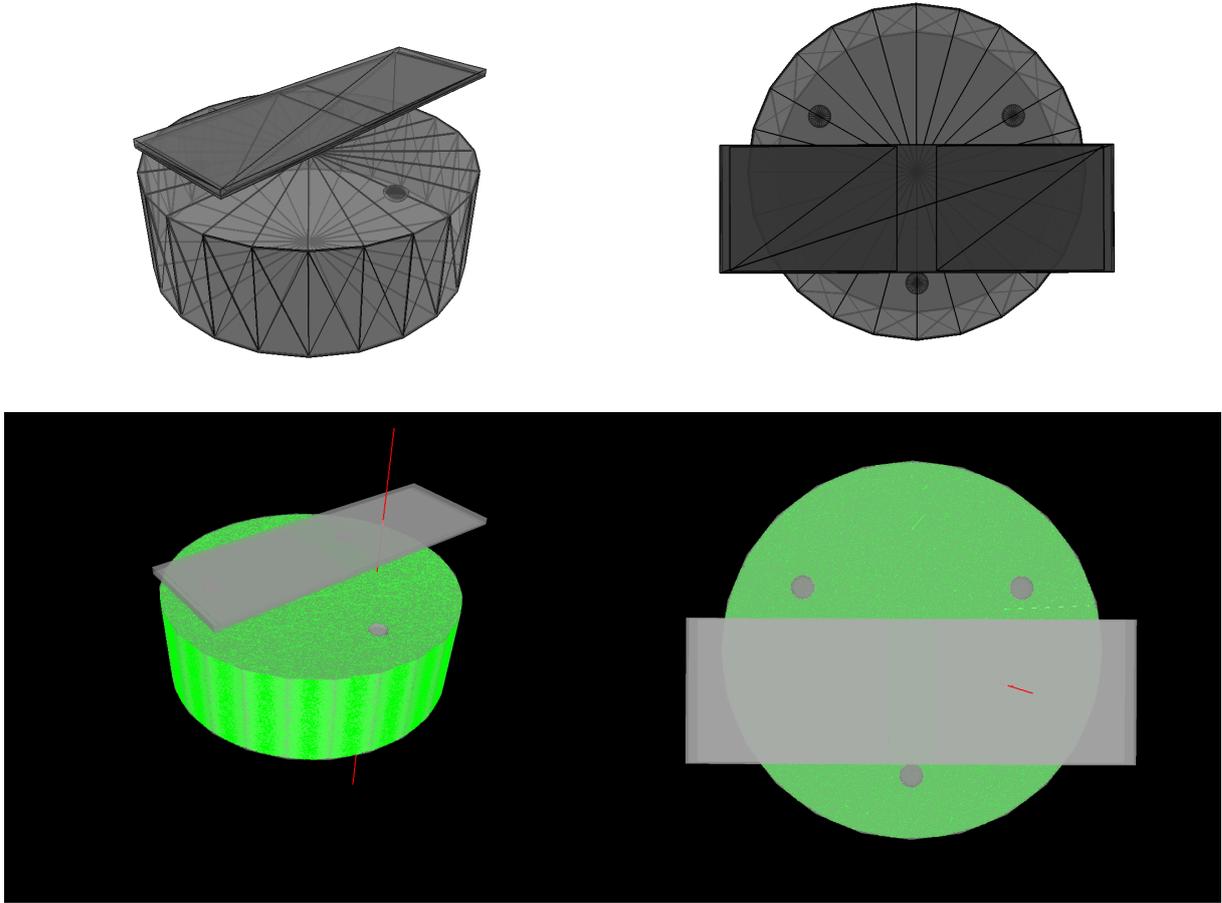


Figure 7.1: Top: Visualisation of the geometry of the **WCD** and **SSD** defined within the simulation module of the **Offline** framework, incorporating the Geant4 toolkit. Bottom: Response of the station when a vertical muon, represented by the red line, traverses the detectors. The green lines within the water volume illustrate the Cherenkov photons produced, which reach the **PMT** windows. For the **SSD**, the deposited energy is extracted from the simulation.

PMTs is converted into a corresponding distribution of photoelectrons generated in the photocathode of the **PMT**s.

For the **SSD** [88], the total energy deposited (E_{dep}) by each impinging particle is extracted from the Geant4 simulation. Then, the number of photoelectrons in the **SSD PMT** (N_{pe}) is estimated for individual particles using the expression:

$$N_{\text{pe}}(\vec{x}) = N_{\text{ref}} \frac{E_{\text{dep}}}{E_{\text{ref}}} f_{\text{att}}(\vec{x}), \quad (7.1)$$

where E_{ref} represents the reference energy produced by vertical minimum ionising particles, as estimated from simulation analysis. The function $f_{\text{att}}(\vec{x})$ describes the at-

tenuation in the number of photons during their propagation in the optical fibres, which depends on the position \vec{x} where the particle interacts with the SSD. The parameters of this function are derived from measurements obtained from the actual detector. N_{ref} represents the reference number of photoelectrons produced by vertical minimum ionising particles after accounting for the attenuation correction.

After obtaining the temporal distribution of photoelectrons, the corresponding responses of the PMTs, as well as the subsequent digitisation with the upgraded board, are simulated. This treatment considers the appropriate gain and employs a convolution procedure with the transfer function of the electronics. The result is a series of ADC pulses sampled at a rate of 120 MHz.

Within the Offline framework, a standard application facilitates the simulation of EAS detection and reconstruction using the AugerPrime surface detector. We will now provide an overview of this application, which we have extensively utilised in our research.

The Offline framework supports the reading of CORSIKA output files, enabling the direct utilisation of the EAS simulations detailed in Section 7.1. The simulated showers are positioned randomly within the surface detector array and subsequently propagated through the simulation of the AugerPrime station.

The resulting ADC signals are utilised to replicate the trigger procedure implemented by the Central Data Acquisition System (CDAS). If the simulated shower triggers the surface detector, signal processing and event reconstruction, as elucidated in Section 3.3, are executed. It is worth noting that the identical modules employed for actual data are applied to the simulated events, ensuring methodological consistency.

Upon reconstruction completion, event-related information, encompassing simulated station signals (distinguished by their respective shower components), as well as reconstructed and true Monte Carlo values (e.g., energy and shower geometry), is recorded in an output file. These output files adhere to the ROOT format, featuring a tree-like data structure mirroring the hierarchy of the event interface within the framework. These files are designated Auger Data Summary Trees (ADST), and the term also denotes the C++ interface used for data retrieval from such files.

7.3 EXECUTION OF SIMULATIONS AND EXTERNAL LIBRARIES

The **EAS** simulations and subsequent detection and reconstruction with the AugerPrime surface detector were conducted at the Centro de Computação John David Rogers (**CCJDR**), the computing centre of the Physics Institute at the University of Campinas.

At the time of execution of our simulations, the **CCJDR** computing cluster featured the following configuration:

- Headnode with 8 cores and 96 GB of RAM.
- 8 nodes, each equipped with 8 cores and 36 GB of RAM.
- 1 node with 8 cores and 144 GB of RAM.
- 1 node with 8 cores and 72 GB of RAM.
- Storage comprising 8 cores with 96 GB of RAM, encompassing volumes of 30 TB and 36 TB.

All processors within the cluster were Intel Xeon 5520 with a clock speed of 2.4 GHz.

Following the software setup on the cluster, we developed shell scripts to manage the execution of simulations. These scripts allocated each available core in the cluster to execute a **CORSIKA** simulation, with the configuration of the primary particle specified as command-line arguments. Immediately following the simulation of an **EAS**, the resulting output file was processed by the Offline application to simulate the response of the AugerPrime detectors, ultimately producing an **ADST** file containing the relevant information. Due to storage limitations, the script deletes the **CORSIKA** file, retaining only the **ADST** file, which is transferred to a designated directory for future analysis. The design of these scripts maximised the utilisation of the multiple cores of the cluster, enabling the parallel execution of our simulations.

Utilising the described machinery, we conducted the simulations detailed in Section 7.1, which are also summarised in Table 7.2. During the later stages of this doctoral research, we gained access to simulation libraries created at the Karlsruhe Institute of Technology (**KIT**), Germany. Dr. Steffen Hahn generated these libraries during his PhD studies [89], employing a methodology closely aligned with ours,

Table 7.2: Summary of the characteristics of the simulation library produced at the University of Campinas. A total of 80 unique showers were simulated for each combination of parameters. Although a substantial part of our studies were conducted using these simulations, the final results reported in this thesis were produced with the libraries detailed in Tables 7.3 and 7.4, benefiting from their superior statistics.

| Characteristic | Values |
|-------------------------|---|
| Primary type | Proton, Iron |
| Energy / eV | $10^{18.0}, 10^{18.5}, 10^{19.0}, 10^{19.5}, 10^{20.0}$ |
| Zenith angle / ° | 0, 8, 18, 28, 38, 48, 58 |
| Azimuth angle / ° | Uniformly distributed between -180 to 180 |
| Total number of showers | 5600 |

employing **CORSIKA** and the $\overline{\text{Offline}}$ software. Leveraging the superior computational resources of **KIT**, these libraries exhibit significantly larger statistics.

Therefore, once we obtained access to these libraries, we favoured their utilisation. The final results reported in Chapters 8 and 9 were obtained using these **KIT** libraries. The first **KIT** library is similar to our approach, featuring fixed energies and zenith angles. However, its distribution in inclination is uniform in $\sin^2 \theta$, resulting in the zenith angles outlined in Table 7.3. For each combination of primary particle, energy, zenith angle, and atmospheric model (comprising one model for each month of the year), ten distinct showers were simulated. This approach resulted in nearly twice as many showers compared to our library.

The second library was generated by applying the $\overline{\text{Offline}}$ application to a continuous **CORSIKA** simulation library produced by the Pierre Auger Collaboration, stored at the IN2P3 Computing Centre in Lyon, France. These simulated showers possess uniform distributions in $\sin^2 \theta$ between 0° and 65° , as well as in the logarithm of energy, covering the range from 18 to 20.2. The number of successfully reconstructed simulated showers using the $\overline{\text{Offline}}$ application is presented in Table 7.4, organised by logarithmic energy bins. It is important to note that the original **CORSIKA** library spans energies below the full trigger efficiency of the Observatory array, explaining the reduced number of events below $10^{18.5}$ eV.

Table 7.3: Fixed library of **CORSIKA** showers and their subsequent reconstruction using the Offline simulation application generated at the Karlsruhe Institute of Technology [89, 90]. The event distribution is uniform in $\sin^2 \theta$, where θ represents the zenith angle of the showers. A total of 120 events were simulated for each combination of the parameters listed, with a uniform distribution across the 12 atmospheric models, each corresponding to a month of the year in Malargüe, Argentina.

| Characteristic | Values |
|-------------------------|---|
| Primary type | proton, iron |
| Energy / eV | $10^{18.0}, 10^{18.5}, 10^{19.0}, 10^{19.5}, 10^{20.0}$ |
| Zenith angle / ° | 0, 12, 22, 32, 38, 48, 56, 65 |
| Azimuth angle / ° | Uniformly distributed between -180 to 180 |
| Total number of showers | 9600 |

Table 7.4: Number of **CORSIKA** showers reconstructed with the Offline simulation application for the different bins in the logarithm of the energy of the primary particles. The utilised **CORSIKA** library is continuous and uniform in logarithmic energy and $\sin^2 \theta$. The reduced number of reconstructed events in the lower bins reflects the fact that the Observatory array of detectors is not fully efficient for these energies.

| | 18-18.5 | 18.5-19 | 19-19.5 | 19.5-20 | 20-20.2 |
|--------|---------|---------|---------|---------|---------|
| Proton | 25 997 | 57 735 | 50 126 | 50 020 | 19 434 |
| Iron | 34 623 | 48 645 | 50 009 | 48 739 | 19 477 |

CHAPTER 8

Reconstruction of signal components within AugerPrime stations

The primary objective behind the installation of scintillators atop the water-Cherenkov detectors is the combination of their signals to enable the extraction of the corresponding muonic component within the AugerPrime stations, a critical step in the pursuit of reconstructing shower observables closely correlated to the composition of primary cosmic rays. Following our validation studies detailed in Chapters 5 and 6, where we successfully demonstrated the compatibility of signals and event reconstruction between Phase-I and Phase-II data, our focus has turned towards utilising the enhanced signals of the AugerPrime surface detector to assess its potential in achieving the objectives of the upgrade.

In this chapter, we discuss our investigation and the development of a method for disentangling the muonic and electromagnetic signals within the upgraded stations. This method is referred to as the *matrix formalism*. We commence by presenting its general formulation in Section 8.1. Subsequently, we scrutinise the behaviour of its reconstruction parameters in Section 8.2. We meticulously assess the accuracy of the method for predicting muonic signals in Section 8.3. Lastly, in Section 8.4, we employ this formalism to reconstruct the temporal structure of the signal components. We emphasise that all these studies have been conducted using the simulations elaborated in Chapter 7.

8.1 FORMULATION OF THE MATRIX FORMALISM

The method of matrix formalism was initially conceived for a layered surface detector [25], where the signals recorded in each of its segments are utilised to distinguish their respective muonic (denoted with " μ ") and electromagnetic (denoted with "em") components. The application of this method to the configuration of the AugerPrime surface detector has been studied by the Auger Collaboration in [90–93]. In our investigation, we have revised the method with the aim of providing a definite formulation as described in the following.

In our approach, we still consider that the total signals within the detectors can be decomposed simply into their muonic and electromagnetic components. The latter encompasses the combined contribution of electrons, positrons and photons. When using a more complicated separation of particle components, for instance, the four components of the Universality formulation, the components should be bundled together according to the most similar behaviour [94].

In the context of an atmospheric shower front passing through a station and locally exhibiting a certain flux \mathcal{F}_μ of muons and a certain flux \mathcal{F}_{em} of electromagnetic particles, we can assert that the resulting signals in the water-Cherenkov detector (**WCD**, abbreviated as "w") and in the scintillator detector (**SSD**, abbreviated as "s") of the station are both fundamentally linearly proportional to these two fluxes. This relation can be expressed with a matrix equation

$$\begin{bmatrix} S_w \\ S_s \end{bmatrix} = \mathcal{M} \begin{bmatrix} \mathcal{F}_\mu \\ \mathcal{F}_{\text{em}} \end{bmatrix} \quad \text{where} \quad \mathcal{M} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad (8.1)$$

and S_w and S_s are the signals in the **WCD** and **SSD**, respectively. The coefficients a , b , c , and d of matrix \mathcal{M} describe the aforementioned proportionality and account for the unit conversions, as the two signals are measured in different units, **VEM** and **MIP**, respectively.

From Eq. (8.1), it can be deduced that the signal registered in the **WCD** can be decomposed into two distinct components, which we denote as $S_w^\mu = a \mathcal{F}_\mu$ and

$S_w^{\text{em}} = b \mathcal{F}_{\text{em}}$. Extending this decomposition to the **SSD**, we obtain:

$$\begin{aligned} S_w &= a \mathcal{F}_\mu + b \mathcal{F}_{\text{em}} = S_w^\mu + S_w^{\text{em}}, \\ S_s &= c \mathcal{F}_\mu + d \mathcal{F}_{\text{em}} = S_s^\mu + S_s^{\text{em}}. \end{aligned} \quad (8.2)$$

The Eq. (8.1) can be inverted to solve for the fluxes:

$$\begin{bmatrix} \mathcal{F}_\mu \\ \mathcal{F}_{\text{em}} \end{bmatrix} = \mathcal{M}^{-1} \begin{bmatrix} S_w \\ S_s \end{bmatrix} = \frac{1}{D} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} S_w \\ S_s \end{bmatrix} \quad (8.3)$$

where $D = \det \mathcal{M} = |\mathcal{M}| = ad - bc$, iff $D \neq 0$. The condition $D \neq 0$ ensures that the two detector responses to the two fluxes have to be *different*. Employing the definitions presented in Eq. (8.2), we can rewrite this solution in a form that solely involves quantities that are, at least in Monte-Carlo data, directly accessible:

$$\begin{bmatrix} S_w^\mu \\ S_w^{\text{em}} \\ S_s^\mu \\ S_s^{\text{em}} \end{bmatrix} = \frac{1}{D} \begin{bmatrix} ad & -ab \\ -bc & ab \\ cd & -bc \\ -cd & ad \end{bmatrix} \begin{bmatrix} S_w \\ S_s \end{bmatrix} = \begin{bmatrix} \alpha & -\beta \\ -\gamma & \beta \\ \delta & -\gamma \\ -\delta & \alpha \end{bmatrix} \begin{bmatrix} S_w \\ S_s \end{bmatrix}, \quad (8.4)$$

where we introduced $\alpha = ad/D$, $\beta = ab/D$, $\gamma = bc/D$, and $\delta = cd/D$. As the electromagnetic and muonic components of signals have to add up to the total signal, i.e. $S_w = S_w^\mu + S_w^{\text{em}}$ and $S_s = S_s^\mu + S_s^{\text{em}}$, we obtain the additional relations $\alpha - \gamma = 1$ and $\alpha\gamma = \beta\delta$.

Since the electromagnetic signals S_w^{em} and S_s^{em} are not independent quantities (as they can be derived from the muonic and total signals), we can omit them at this stage. This provides us with the final form of the matrix formalism, which establishes the connection between the muonic signals in the two detectors and their total signals:

$$\begin{bmatrix} S_w^\mu \\ S_s^\mu \end{bmatrix} = \mathcal{A} \begin{bmatrix} S_w \\ S_s \end{bmatrix} \quad \text{where} \quad \mathcal{A} = \begin{bmatrix} \alpha & -\beta \\ -\frac{\alpha(1-\alpha)}{\beta} & 1-\alpha \end{bmatrix}, \quad (8.5)$$

iff $\beta \neq 0$. Note that $\det \mathcal{A} = |\mathcal{A}| \equiv 0$, which is a consequence of the fact that the Eq. (8.5) cannot be inverted so that the total signals in **WCD** and **SSD** cannot be expressed only in terms of their muonic signals. The two rows of matrix \mathcal{A} are linearly dependent, which

is evident from the fact that the lower row is obtained by multiplying the upper row with a factor $-(1 - \alpha)/\beta$. Furthermore, it implies a proportional correlation between the muonic signals in the two detectors.

Although Eq. (8.5) allows us to determine the muonic component in the detectors from their corresponding total signals, it is crucial to firstly establish the coefficients α and β . Since we have access to the signals on both sides of Eq. (8.5) in Monte-Carlo data, we are able to derive the following linear system of equations

$$\begin{bmatrix} S_w^\mu \\ S_s^\mu \end{bmatrix} = \underbrace{\begin{bmatrix} S_w & -S_s \\ S_w^\mu & -S_s^\mu \end{bmatrix}}_{\mathcal{K}} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}. \quad (8.6)$$

The solution of this system yields the following expressions for α and β :

$$\alpha = \frac{S_w^\mu(S_s - S_s^\mu)}{S_w^\mu S_s - S_w S_s^\mu} \quad \text{and} \quad \beta = \frac{S_w^\mu(S_w - S_w^\mu)}{S_w^\mu S_s - S_w S_s^\mu}, \quad (8.7)$$

iff $\det \mathcal{K} \neq 0$.

Introducing the muonic fractions in the detectors as well as a signal ratio through the expressions

$$f_w^\mu = S_w^\mu/S_w, \quad f_s^\mu = S_s^\mu/S_s, \quad \text{and} \quad s = S_w/S_s, \quad (8.8)$$

allow us to further simplify Eq. (8.7) into

$$\alpha = \frac{1 - f_s^\mu}{1 - f_s^\mu/f_w^\mu} \quad \text{and} \quad \beta = s \frac{1 - f_w^\mu}{1 - f_s^\mu/f_w^\mu}, \quad (8.9)$$

which mathematically enforce the condition $S_s \neq 0$, reflecting the obvious requirement of having signal in the SSD. It is noteworthy that the necessity of $\det \mathcal{K} \neq 0$, directly implies that the application of the formalism requires $f_s^\mu \neq f_w^\mu$. This reinforces the physical requirement of having distinct responses of the detectors to the muonic and electromagnetic components of the particle cascade. Consequently, the elements of the matrix \mathcal{A} solely depend on the relative amounts f_w^μ and f_s^μ of the muonic signal in each detector and the ratio s of the two total signals.

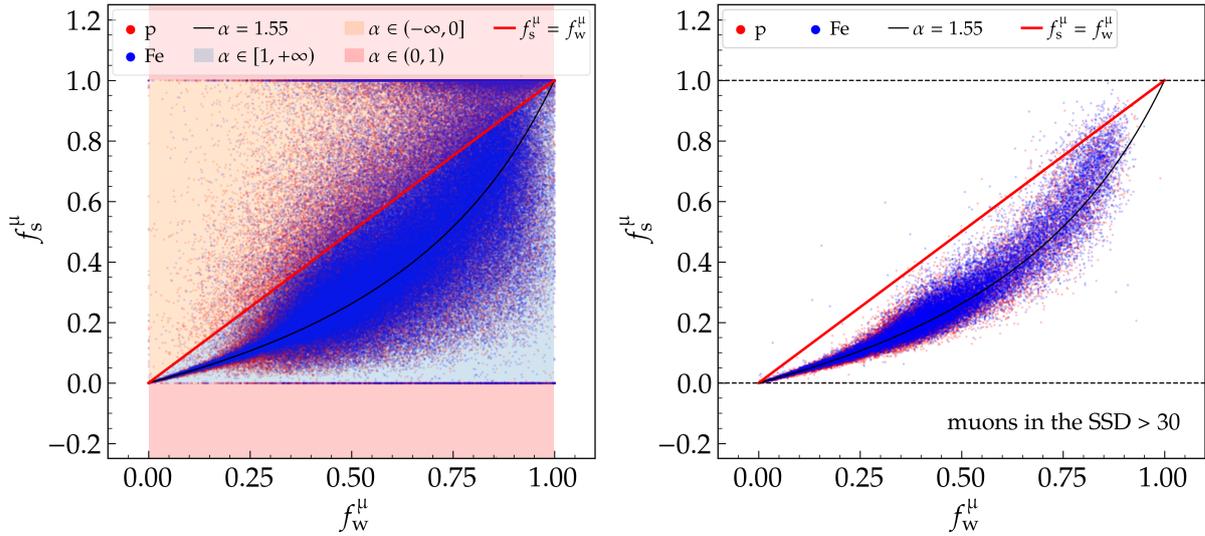


Figure 8.1: Distribution of the Monte-Carlo data in the $f_w^\mu: f_s^\mu$ space for the complete data set (left) and when more than 30 muons impinge on the **SSD** (right). The different colour regions in the left panel correspond to the curves given by Eq. (8.10) with different values of the parameter α . The cut applied on the right removes stations where the particle fluxes in the detectors are significantly different. This causes a drastic clean up in the space, and we are left with a narrower distribution.

8.2 BEHAVIOUR OF THE RECONSTRUCTION PARAMETERS

The Eq. (8.5) implies that, once we have the appropriate parameters α and β , the muonic¹ component in either detector can be obtained from the total signal in the **WCD** and **SSD**. Therefore, it is crucial to study the behaviour of the parameters α and β and determine the situations in which the formalism can be applied.

In Fig. 8.1 (left), we display the muonic fractions in the detectors as obtained from Monte-Carlo data. The data is mainly distributed in an arc, which is more apparent for smaller muonic fractions. However, a large amount of data is scattered in the $f_w^\mu: f_s^\mu$ space. There are also events in which no muons cross the **SSD** ($f_s^\mu = 0$) and others in which the **SSD** signal is purely muonic ($f_s^\mu = 1$). From Eq. (8.9), we can express f_s^μ as a function of f_w^μ ,

$$f_s^\mu = \frac{1 - \alpha}{1 - \alpha/f_w^\mu}. \quad (8.10)$$

Hence, different values of α produce different curves in the $f_w^\mu: f_s^\mu$ space. One such curve is represented in Fig. 8.1 by the black line (using $\alpha = 1.55$). The red line shows the

¹and also the electromagnetic component, using the sum rule of the signals.

situation $f_s^\mu \equiv f_w^\mu$, which implies an undefined value of α (and β , see Eq. (8.9)). When $\alpha \geq 1$, the curves will be contained in the purple region (lower diagonal). On the other hand, for $\alpha \leq 0$, the curves are in the orange region (upper diagonal). Interestingly, $0 < \alpha < 1$ corresponds to the "forbidden" region in red, where the muonic fraction in the SSD would assume the nonsensical values of $f_s^\mu < 0$ or $f_s^\mu > 1$.

When deriving the matrix formalism, we assumed that the muonic (and electromagnetic) flux is the same in the SSD and WCD. Such a situation is not met when few particles impinge upon the detectors, given their different surface areas. In Fig. 8.1 (right), we present the $f_w^\mu : f_s^\mu$ space when more than 30 muons cross the SSD. This cut removes the situations where, due to a sampling effect, the fluxes in the detectors are significantly different. It is striking how the data is distributed in a much narrower and better-defined region. We shall return to the discussion of the $f_w^\mu : f_s^\mu$ space after the following investigations.

To study in a more quantitative manner the variations of the parameters α and β , in Fig. 8.2, the plots in the middle and the right present in their bins the standard deviation for these parameters, which are calculated using Eq. (8.9) with Monte-Carlo data. The plots in the left column depict the corresponding number of entries in each bin.

In Fig. 8.2 (top), we study the variation of the parameters as a function of the number of muons that cross the detectors. Given that the SSD has a smaller area than the WCD, we naturally expect to have a larger number of muons in the latter. When a small number of muons impinge upon the detectors, a large variation of the parameters is observed. We can think of two effects that explain this behaviour. The parameters α and β could present a dependency on shower properties such as the zenith angle, the distance of the detectors to the shower core as well as energy. The other possible reason, as mentioned previously, is that due to the low number of particles in the detectors, their fluxes are not the same, and since it is the base assumption of the formalism, the use of Eq. (8.9) provides values that are not consistent. However, even with more than 30 muons traversing the SSD, the variation of the parameters is still appreciable, which suggests that dependencies on shower properties play a more important role.

The detectors cannot directly measure the number of muons passing through them. Therefore, we must rely on other measurable variables to determine when the

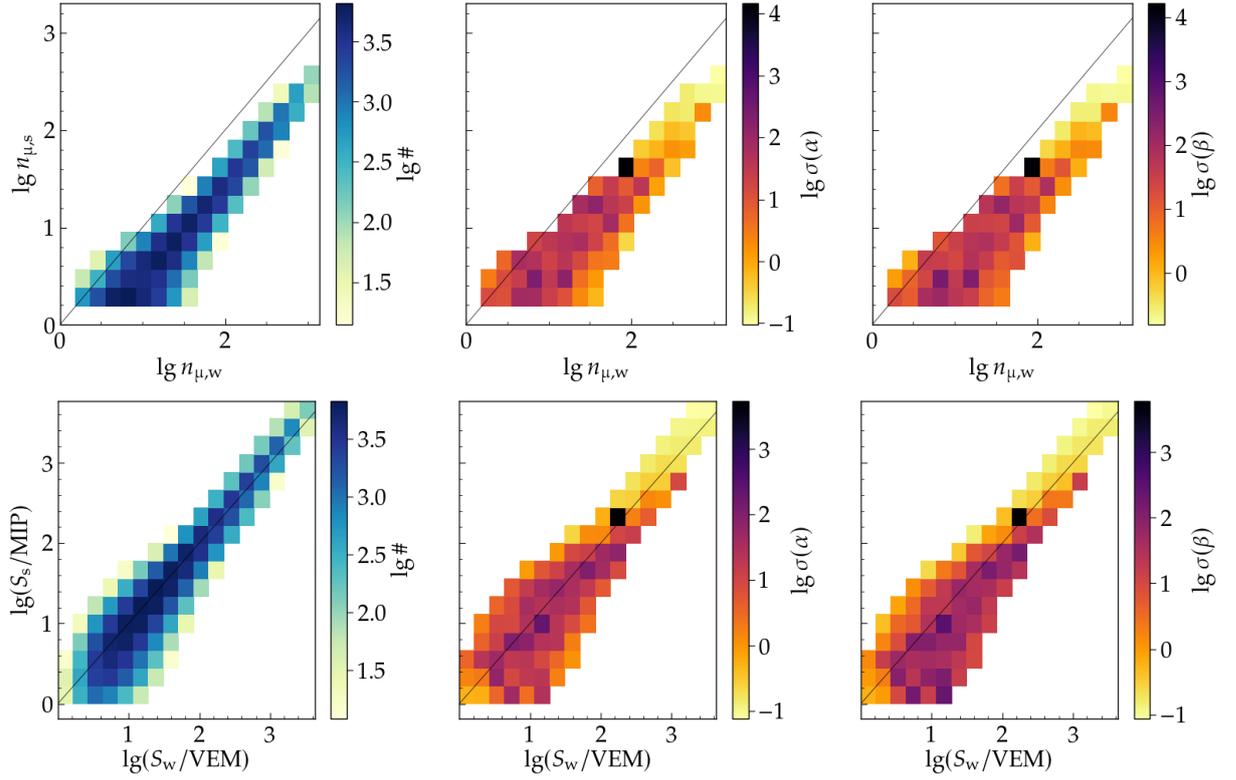


Figure 8.2: Standard deviation of parameters α (mid column) and β (right column) as a function of the number of muons (top row) and signals (bottom row) in the detectors. On the left side, the corresponding number of entries for each bin is displayed. We only consider bins with more than ten entries.

matrix formalism is applicable. A logical starting point is to consider the total signals recorded by the detectors, as they are proportional to the number of particles. As shown in the lower part of Fig. 8.2, we can observe that larger signals correspond to smaller variations in the parameters α and β . It is worth noting that saturated signals were excluded from this analysis.

In Fig. 8.3, we provide similar plots to those found in Fig. 8.2, but here we focus on shower properties. It becomes evident that when the signal in the SSD exceeds 200 MIP, the parameters α and β exhibit considerably reduced variation, regardless of the shower properties. As depicted in the lower row of Fig. 8.3, such substantial signals correspond to stations located near the shower core, typically within 1000 m. This observation underscores the necessity of parameterising α and β to account for variations related to shower properties. Attempting to approximate them with constants would necessitate imposing significant signal cuts on the stations and would greatly reduce the number of events for which the approximation could be effectively applied.

Concerning Eq. (8.9), which demonstrates the dependence of parameters α

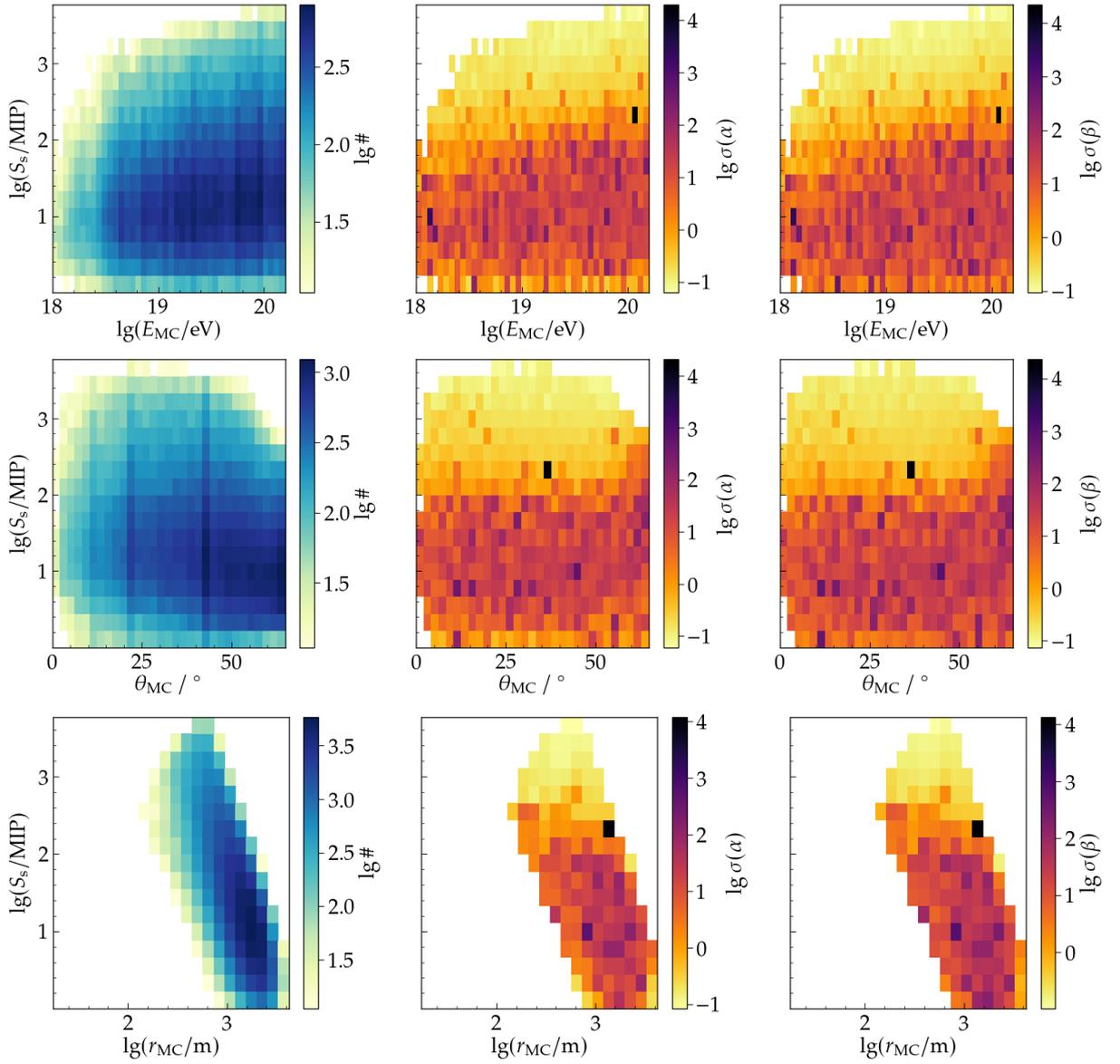


Figure 8.3: Standard deviation of parameters α (mid column) and β (right column) as a function of the signal in the **SSD** and energy of the primary (top row), zenith angle (mid row), and distance to the shower core (bottom row). The corresponding number of entries for each bin is displayed in the plots in the left column. Only bins with more than ten entries are presented.

and β on the muonic fractions and the total signal ratio of the detectors, together with the insights gained from Figs. 8.2 and 8.3, we investigated how these quantities relate to shower properties.

Fig. 8.4 depicts the behaviours of muonic signal fractions in the detectors, their ratio f_s^μ/f_w^μ , and the ratio of total signals in the **WCD** and **SSD** as functions of the distance to the shower core. For this analysis, we utilised primaries with an energy of $10^{19.5}$ eV. Different colours correspond to fixed zenith angles, and open and solid

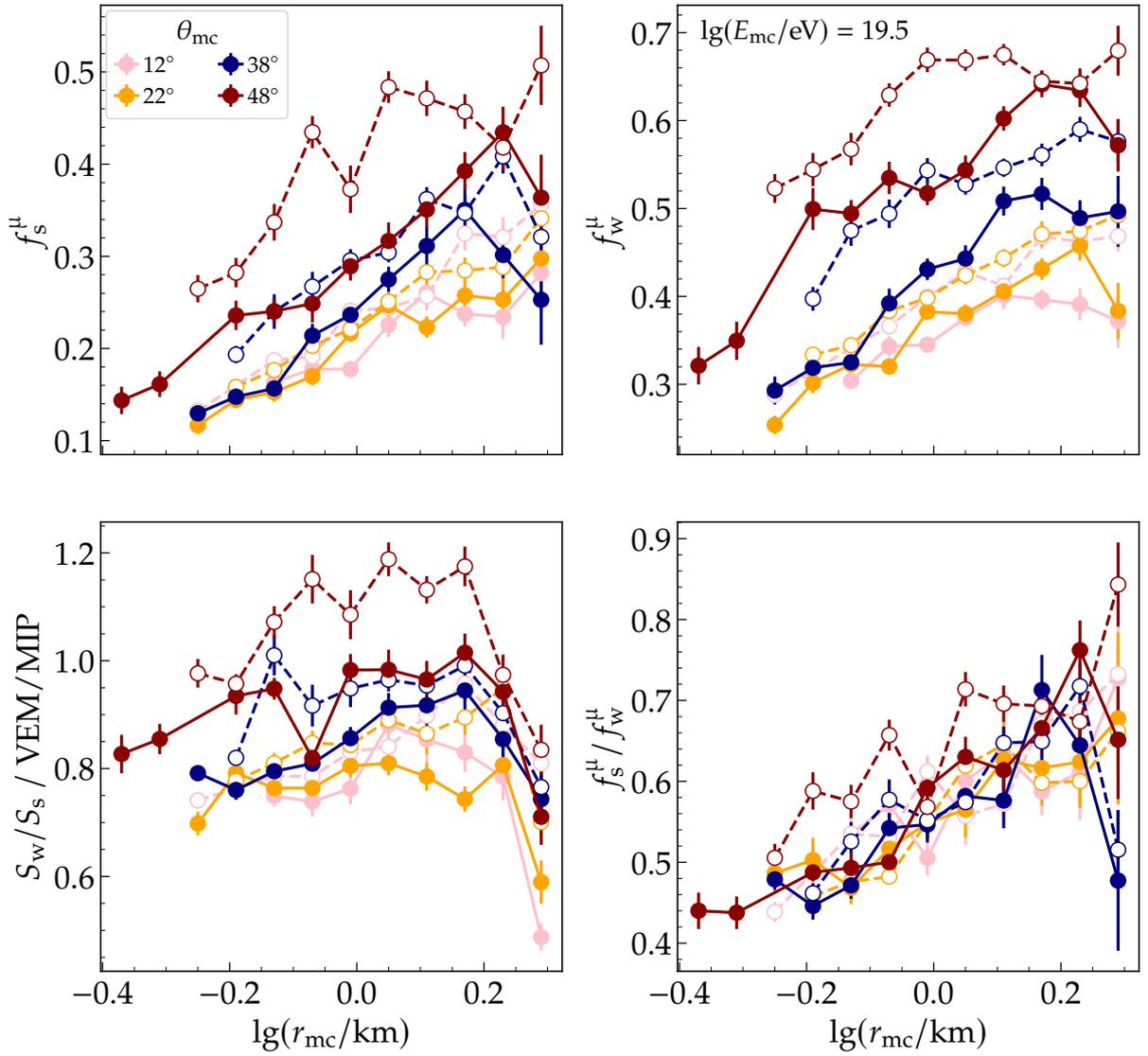


Figure 8.4: Radial behaviour of the muonic signal fraction in the **SSD** (top left) and in the **WCD** (top right), as well as their ratio (bottom right) and the total signal ratio in the detectors (bottom left). The different colours represent different zenith angles, whereas open markers are for iron primaries and closed ones for protons. Only primaries with energy of $10^{19.5}$ eV were utilised.

markers represent iron and proton primaries, respectively.

As expected, the muonic fraction in the detectors increases with distance from the core due to the greater atmospheric absorption of the electromagnetic component further away from the shower core. Moreover, we observe larger muonic fractions for showers with higher zenith angles, a consequence of the increased absorption of the electromagnetic component during their development to the observation level. Comparing different primary particles with the same zenith angle, iron primaries yield larger muonic fractions than protons. This disparity arises from the greater mass of

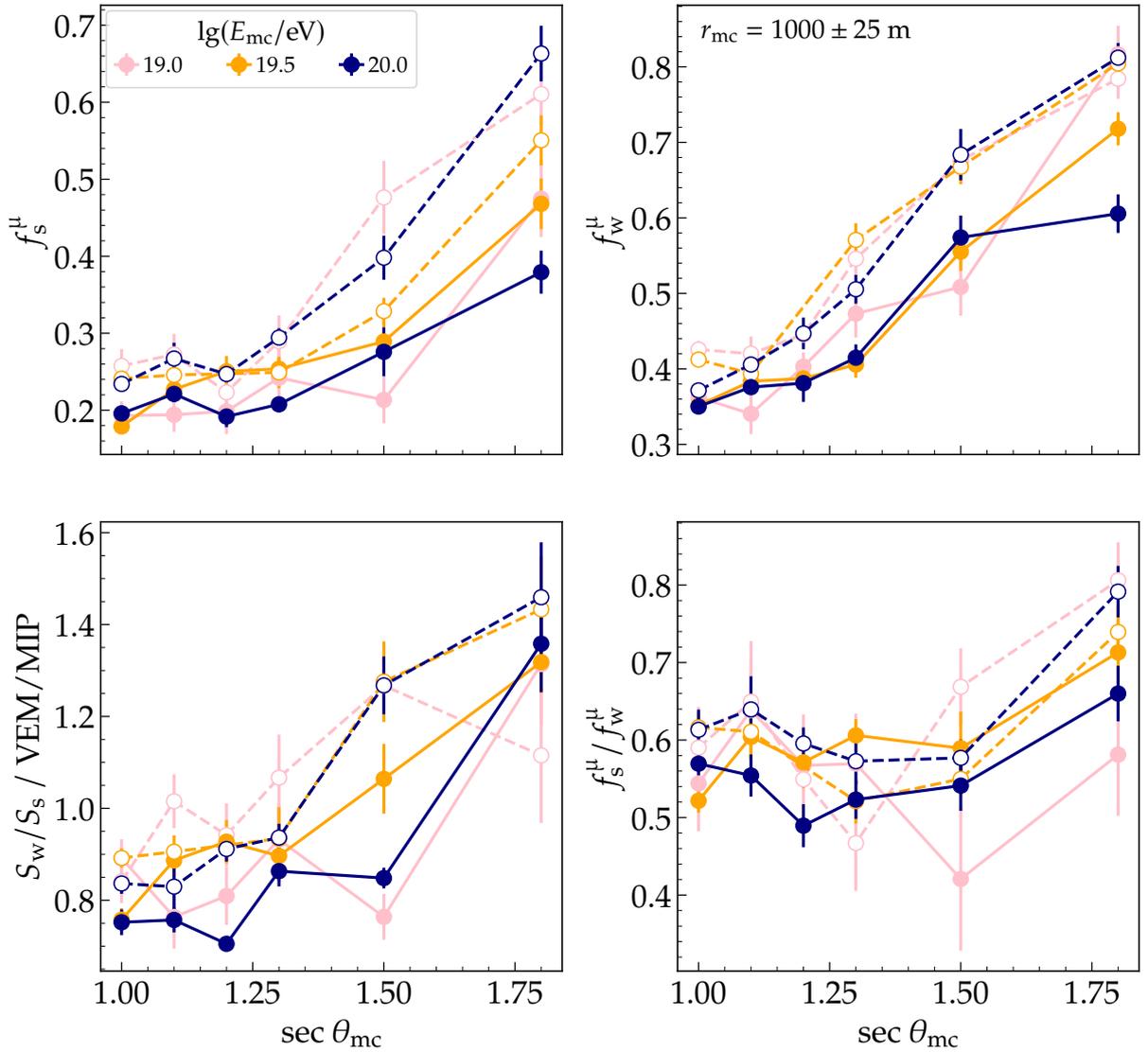


Figure 8.5: Behaviour with the zenith angle of the muonic signal fraction in the **SSD** (top left) and in the **WCD** (top right), as well as their ratio (bottom right) and the total signal ratio in the detectors (bottom left). The different colours represent different energies of the primaries, whereas open markers are for iron and closed ones for proton. Only signals of stations at (1000 ± 25) m from the shower core were utilised.

iron nuclei, which leads to a larger production of muons in the atmospheric cascade of particles.

The ratio of muonic fractions f_s^μ / f_w^μ is consistently smaller than one, indicating that the **WCD** is more sensitive to the muonic component compared to the **SSD**. Moreover, this ratio tends to increase as the distance from the shower core grows. This effect reveals that f_s^μ increases more rapidly with distance than f_w^μ . In contrast, the ratio of total signals, S_w/S_s , exhibits a relatively weak dependency on distance until it experiences a noticeable drop at around 1400 m.

In Fig. 8.5, we examine more thoroughly the dependency of f_s^μ , f_w^μ , f_s^μ/f_w^μ and S_w/S_s on zenith angle. We only consider the signal of stations at (1000 ± 25) m from the shower core to eliminate the dependency on distance. Besides the increase of f_s^μ and f_w^μ with zenith angle, which was discussed previously, we note it is rather weak until about 40° , then it becomes much stronger. This behaviour demonstrates that the suppression of the electromagnetic component of the shower is greatly increased beyond the aforementioned inclination.

Furthermore, we observe that the ratio f_s^μ/f_w^μ is quite constant with zenith angle, which is a consequence of the similar dependency of f_s^μ and f_w^μ on this quantity (as apparent in their individual plots). The total signal ratio S_w/S_s has a similar behaviour to the individual muonic fractions f_s^μ and f_w^μ . Regarding the same primary type, the behaviour of the muonic fractions and ratios is very similar for the distinct primary energies.

The plots in Fig. 8.6 confirm that the dependency of these quantities on energy is weak. This outcome is consistent with what we expect since, in first approximation, primaries with larger energies produce a larger number of particles, hence the muonic fractions and ratios are not significantly affected by the rough scaling on the number of muons and electromagnetic particles being sampled in the detectors.

The results in Figs. 8.4 to 8.6 demonstrate that the quantities which the parameters α and β depend on in Eq. (8.9) vary more importantly with distance to the shower axis and zenith angle. The dependency on primary energy plays a secondary role. Hence, we should also expect dependencies of α and β on these shower properties.

In Fig. 8.7, these parameters are displayed as a function of the distance to the shower core. We distinctly observe a stronger dependency for distances larger than 850 m. At closer proximity to the shower core, particularly for zenith angles larger than 38° , the parameters are rather constant, which is in agreement with the results presented in Fig. 8.3.

Notably, the error bars in Fig. 8.7 (top) become markedly large for growing distances. Analysing the corresponding scatter plots (bottom row of Fig. 8.7), we trace the source of these pronounced variations to stations presenting a considerable deviation from the mean. In these stations, which are typically located at larger distances from the shower core, the muonic component dominates, meaning that these events

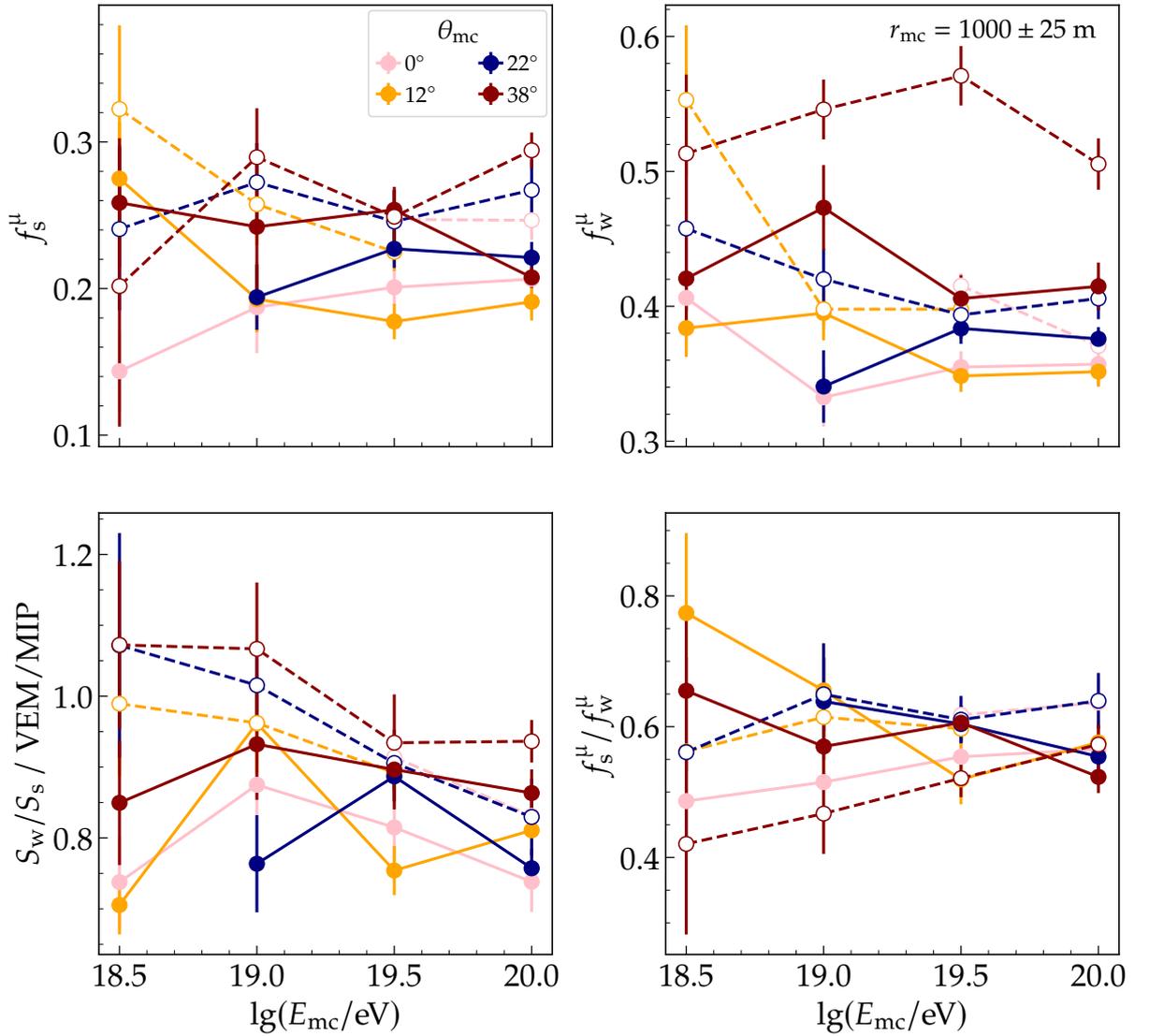


Figure 8.6: Behaviour with primary energy of the muonic signal fraction in the **SSD** (top left) and in the **WCD** (top right), as well as their ratio (bottom right) and the total signal ratio in the detectors (bottom left). The different colours represent different zenith angles, whereas open markers are for iron primaries and closed ones for protons. Only signals of stations at (1000 ± 25) m from the shower core were used.

lie in the upper-right region of the $f_w^\mu:f_s^\mu$ space of Fig. 8.1, where the data crosses the equality line $f_s^\mu = f_w^\mu$, a situation that causes an undefined behaviour of parameters α and β (refer to Eq. (8.9)), hence the outliers in Fig. 8.7 (bottom).

It is relevant to comment that purely muonic signals in both detectors also cause the parameters to be undefined. This interestingly demonstrates the necessity of the presence of the electromagnetic component for the formalism to be effectively employed.

The zenith and energy dependencies of parameters α and β are illustrated

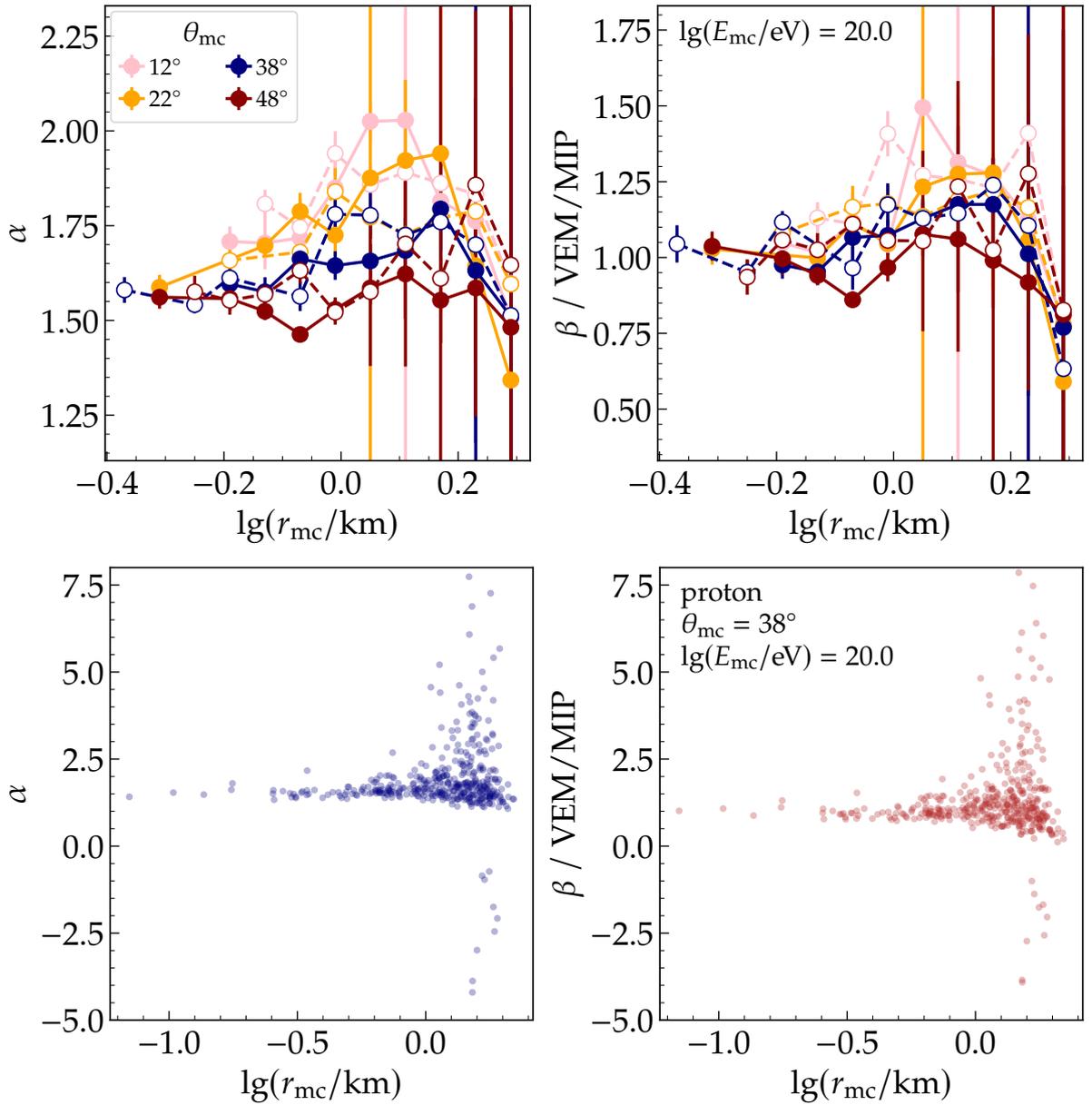


Figure 8.7: Parameters α (left) and β (right) as a function of the distance to the shower core. Both scatter plots (bottom row) and profiles (top row) are presented. The "outlier" points in the scatter plots, far from the shower core, likely correspond to stations where the muonic fraction in the detectors is similar, which implies an undefined behaviour of the parameters (from Eq. (8.9)).

in Fig. 8.8. We notice that α tends to have larger values for showers with smaller zenith angles. Conversely, the parameter β exhibits a slightly weaker dependence on the zenith angle. In terms of energy dependency, both parameters demonstrate a weak relationship, consistent with the behaviour of muonic fractions and total signal ratio, as shown in Fig. 8.6.

For completion of the discussion of Fig. 8.1, we conclude that, besides the

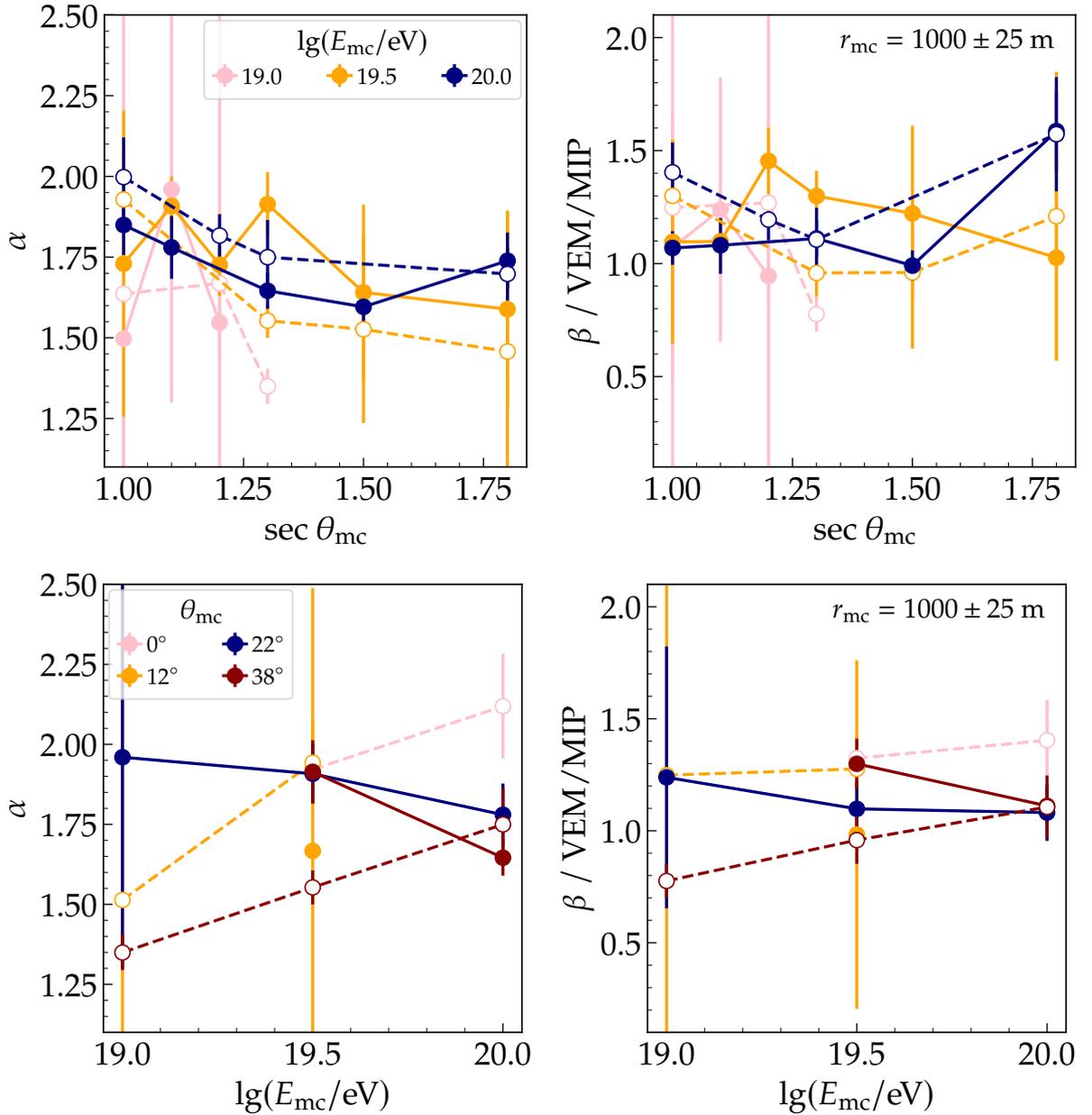


Figure 8.8: Parameters α (left) and β (right) as a function of the zenith angle (top row) and primary energy (bottom row). The different colours represent different energies (top) and zenith angles (bottom), whereas solid markers are for proton and open for iron primaries. In all plots, only stations at (1000 ± 25) m from the shower core are considered.

clean up of the space with the cut imposed on the number of muons traversing the SSD, which removes situations with different fluxes of particles in the detectors, the crescent-like shape of the region where the data is confined is due to the different values that the parameter α can assume, given its dependencies on shower properties that we have shown.

An interesting approach for understanding the main quantities that drive

the behaviour of the parameters α and β is to model the signal produced by the muonic and electromagnetic components and compare those with the coefficients of the matrix formalism. We have decided to leave this discussion in Chapter A, as the conclusions do not differ from the results described in this section.

8.3 PREDICTION OF SIGNAL COMPONENTS

From the discussion in the preceding section, we verified that the reconstruction parameters α and β need, in general, to be parameterised to account for their dependencies on shower properties, allowing the formalism to be applied in a broader spectrum of situations. However, we additionally showed that within stations where the signal is large enough, typically above 100 MIP in the scintillator, the reconstruction parameters present a small variation. Although such a large cut on signal drastically reduces the number of events that can be utilised, we have explored the possibility of employing constants as the reconstruction parameters α and β , given its simplicity. Subsequently, we will compare the results with those obtained when the reconstruction parameters are parameterised.

In Fig. 8.9, we present the overall distribution of parameters α and β calculated from Eq. (8.9) using Monte Carlo data, where 50% of the events are generated by protons and the other 50% by iron nuclei. When a cut of 100 MIP is applied to the signal of the scintillators, the resulting distribution becomes notably narrower, as anticipated by the information in Figs. 8.2 and 8.3. An estimate of the parameters is obtained by evaluating the median of the resulting distributions, which yields the values 1.67 and 1.09 VEM/MIP for α and β , respectively. Employing the median instead of the mean mitigates the effect of the outliers evident in Fig. 8.7.

To parameterise the reconstruction coefficients, we rewrite the matrix \mathcal{A} of Eq. (8.5) in a generic manner, so that

$$\begin{bmatrix} S_w^\mu \\ S_s^\mu \end{bmatrix} = \mathcal{A} \begin{bmatrix} S_w \\ S_s \end{bmatrix} \quad \text{where} \quad \mathcal{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}. \quad (8.11)$$

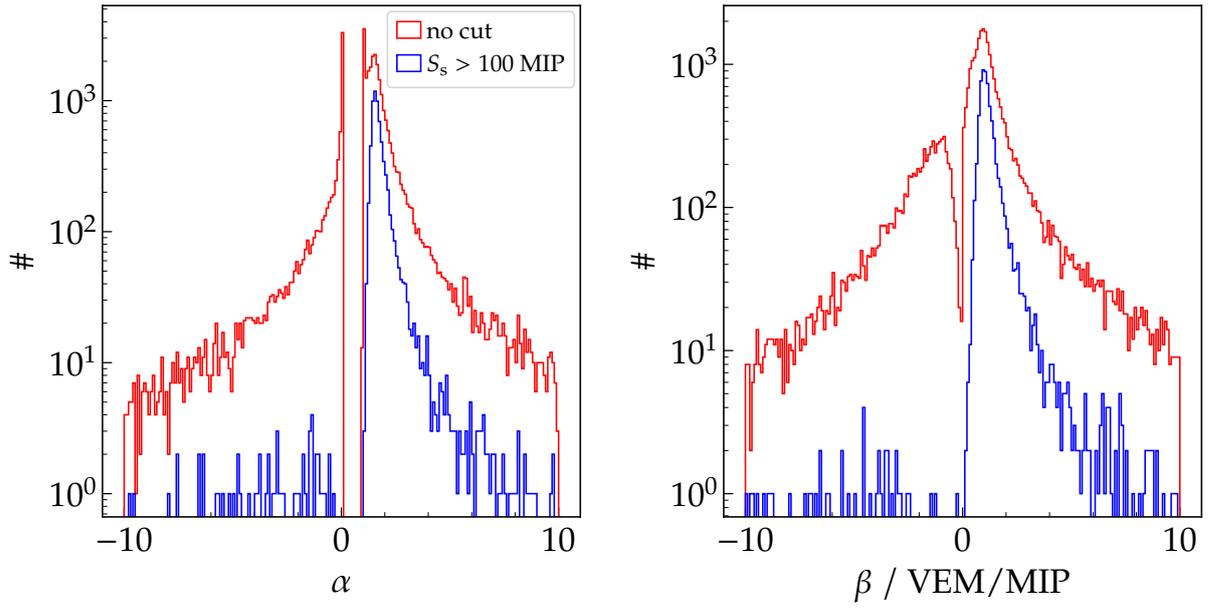


Figure 8.9: Distribution of parameters α and β calculated using Eq. (8.9) with Monte Carlo data. Primaries of proton and iron were used in a 1:1 ratio. The parameters α and β are estimated as the median of the corresponding distributions when the signal in the SSD is larger than 100 MIP, which provides the values 1.67 and 1.09 VEM/MIP, respectively.

Subsequently, each of the coefficients a_{ij} are assumed to be described by the function

$$\begin{aligned}
 a_{ij} = & p_0^{ij} \\
 & + p_1^{ij} \varepsilon + p_2^{ij} \varepsilon^2 \\
 & + p_3^{ij} \rho + p_4^{ij} \rho^2 \\
 & + p_5^{ij} \vartheta + p_6^{ij} \vartheta^2 \\
 & + p_7^{ij} \varepsilon \rho + p_8^{ij} \varepsilon \vartheta + p_9^{ij} \rho \vartheta,
 \end{aligned} \tag{8.12}$$

where the parameters p_k^{ij} with $k = 0, 1, 2, \dots, 9$ are constants to be determined for each of the coefficients of the matrix \mathcal{A} . Note that, in total, we have 40 parameters to reconstruct the muonic signal of the detectors². Considering a station at a distance r from the core of a shower with zenith angle θ and energy E , we defined

$$\varepsilon \equiv \lg(E/\text{eV}) - 19.0, \quad \vartheta \equiv \sec \theta - \sec 38^\circ, \quad \text{and} \quad \rho = \lg(r/\text{km}), \tag{8.13}$$

so that the parameterisation is somewhat centred with respect to our data. To account

²The electromagnetic signals can then be easily obtained by subtracting the muonic component from the total signal.

for the dependency of the reconstruction coefficients on shower properties, we included a second-degree polynomial on ε , ϑ and ρ , as well as their cross terms in the first degree.

In the final step, the parameters p_k^{ij} are determined by minimising the χ^2 -like function

$$\chi_{\text{det}}^2 = \sum_i \frac{(\hat{S}_{\text{det},i}^{\mu} - S_{\text{det},i}^{\mu})^2}{S_{\text{det}}}, \quad (8.14)$$

where "det" represents each of the detectors, **WCD** and **SSD**. $\hat{S}_{\text{det}}^{\mu}$ is the estimated muonic signal in the corresponding detector, which is obtained from Eq. (8.11) with the parameterisation in Eq. (8.12). S_{det}^{μ} is the true muonic signal in the referred detector from the Monte Carlo data. And S_{det} is the total signal of the detector. From the simulations, we use the reconstructed values for the event energy and zenith angle as well as for the distance to the shower core in the parameterisation of Eq. (8.12). The minimisation for the **WCD** provides the parameters of the matrix coefficients a_{11} and a_{12} , whereas a_{21} and a_{22} have their corresponding parameters from the **SSD** minimisation.

To assess the performance of the reconstruction of the muonic signals, we study the quantity $\Delta f_{\text{det}}^{\mu}$, which is defined as the difference between the reconstructed muonic fraction $\hat{f}_{\text{det}}^{\mu}$ and the true Monte Carlo fraction f_{det}^{μ} for each detector. The reconstructed fraction is calculated as $\hat{S}_{\text{det}}^{\mu}/S_{\text{det}}$, where the muonic signal is estimated using Eq. (8.11) with the appropriate reconstruction parameters.

We compared the application of constant and parameterised coefficients for the reconstruction of the signal components of the detectors. Since the constant parameters α and β were estimated using signals larger than 100 **MIP** in the **SSD**, the same cut was utilised when determining the parameterised coefficients for this comparison. We also only reconstruct signal components when the **SSD** total signal exceeds 100 **MIP**.

In Figs. 8.10 and 8.11, we present the resulting $\Delta f_{\text{w}}^{\mu}$ and its mean $\langle \Delta f_{\text{w}}^{\mu} \rangle$, which represents the prediction bias for both reconstruction approaches. Even with a considerable cut on the detector signals, we observe that employing constant values for α and β results in biases on the reconstruction that are more zenith-angle dependent but also evident for distinct primary energies. Conversely, the reconstruction bias is significantly reduced when a parameterisation of the coefficients is applied, being mostly confined within 5% for different zenith angles, energies and ranges of distance

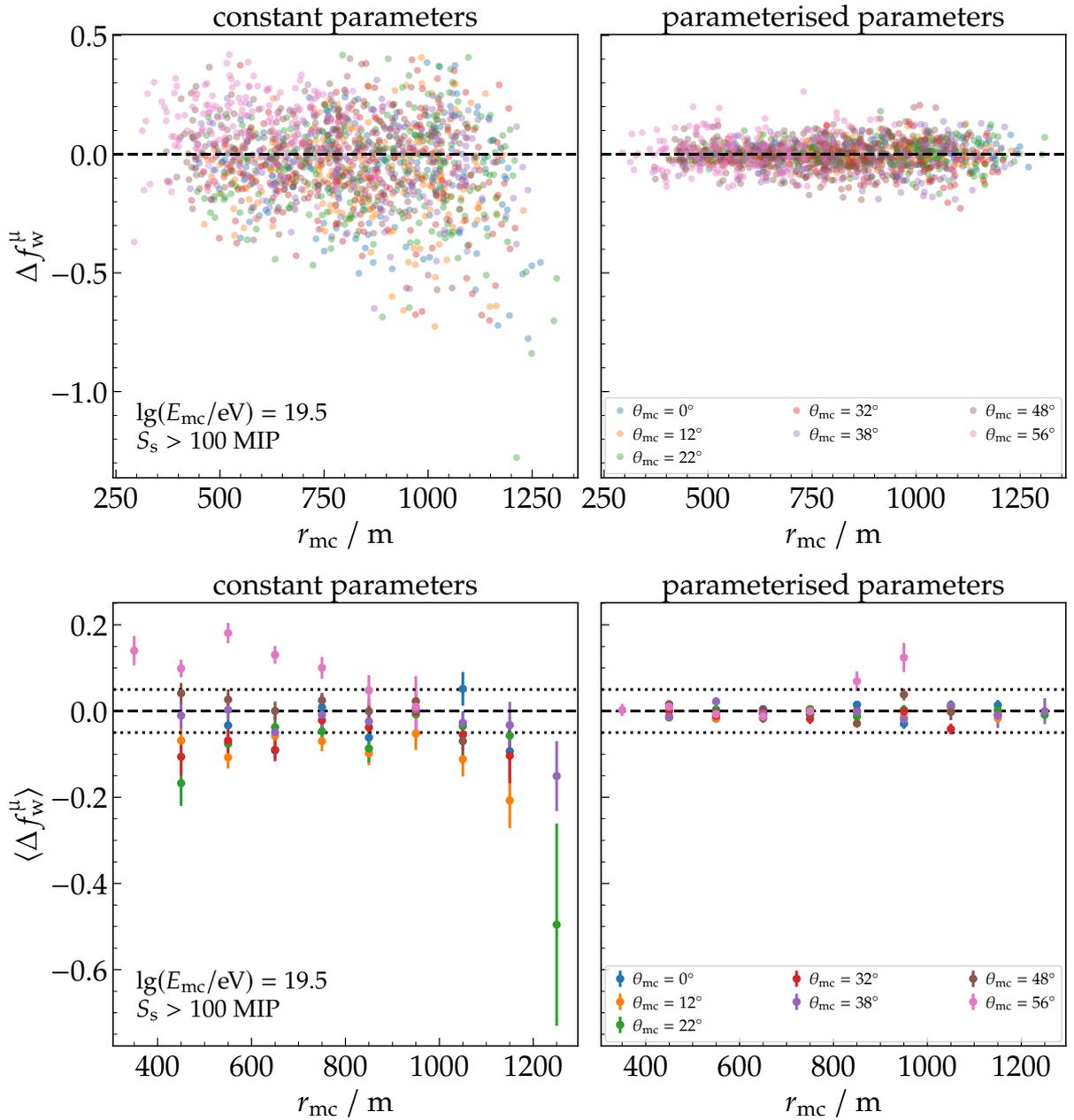


Figure 8.10: Difference between reconstructed and true muonic fraction in the **WCD** (top) and its mean (bottom) as a function of the distance to the shower core. The results using constant (left) and parameterised (right) coefficients are displayed for different zenith angles in showers with an energy of $10^{19.5}$ eV. Only stations with a signal in the **SSD** exceeding 100 **MIP** were considered.

to the shower core. It is noteworthy that the dispersion of the reconstructed fractions is markedly smaller when utilising the parameterisation. This outcome reinforces the dependency of the reconstruction with shower properties and that a parameterisation is necessary for a more accurate estimation of the signal components.

We conducted a more detailed examination of the signal component re-

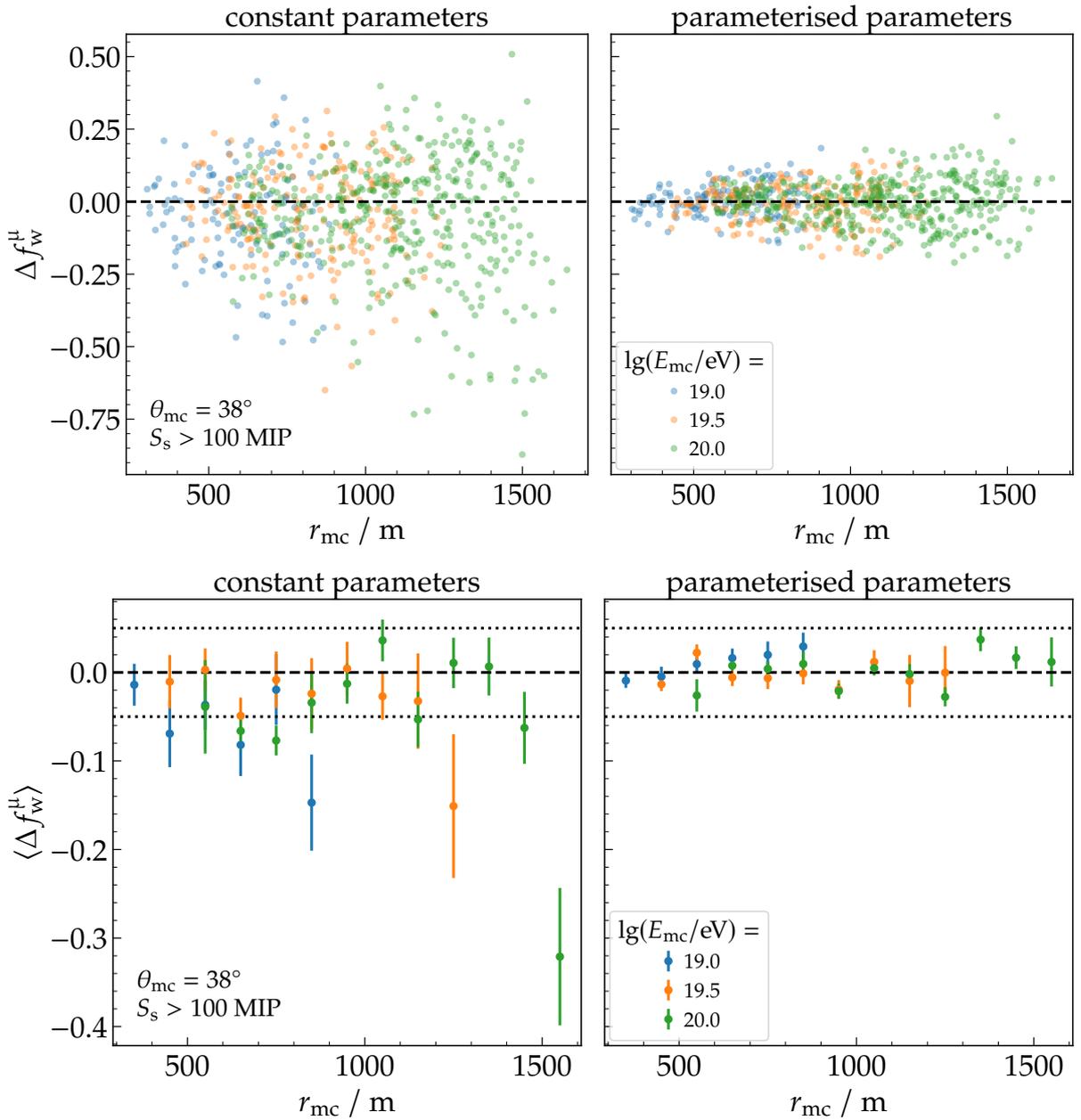


Figure 8.11: Difference between reconstructed and true muonic fraction in the **WCD** (top) and its mean (bottom) as a function of the distance to the shower core. The results using constant (left) and parameterised (right) coefficients are displayed for different energies of the primary particle in showers with a zenith angle of 38° . Only stations with a signal in the **SSD** larger than 100 **MIP** were utilised.

construction using the parameterised coefficients. As the parameterisation effectively accounted for the dependencies on shower properties, we relaxed the cut on the **SSD** signal to 30 **MIP**, thereby extending the applicability of the reconstruction to a larger pool of events. However, it remains necessary to impose a signal cut to maintain roughly equivalent particle fluxes in both the **WCD** and **SSD**, as demonstrated in Fig. 8.1.

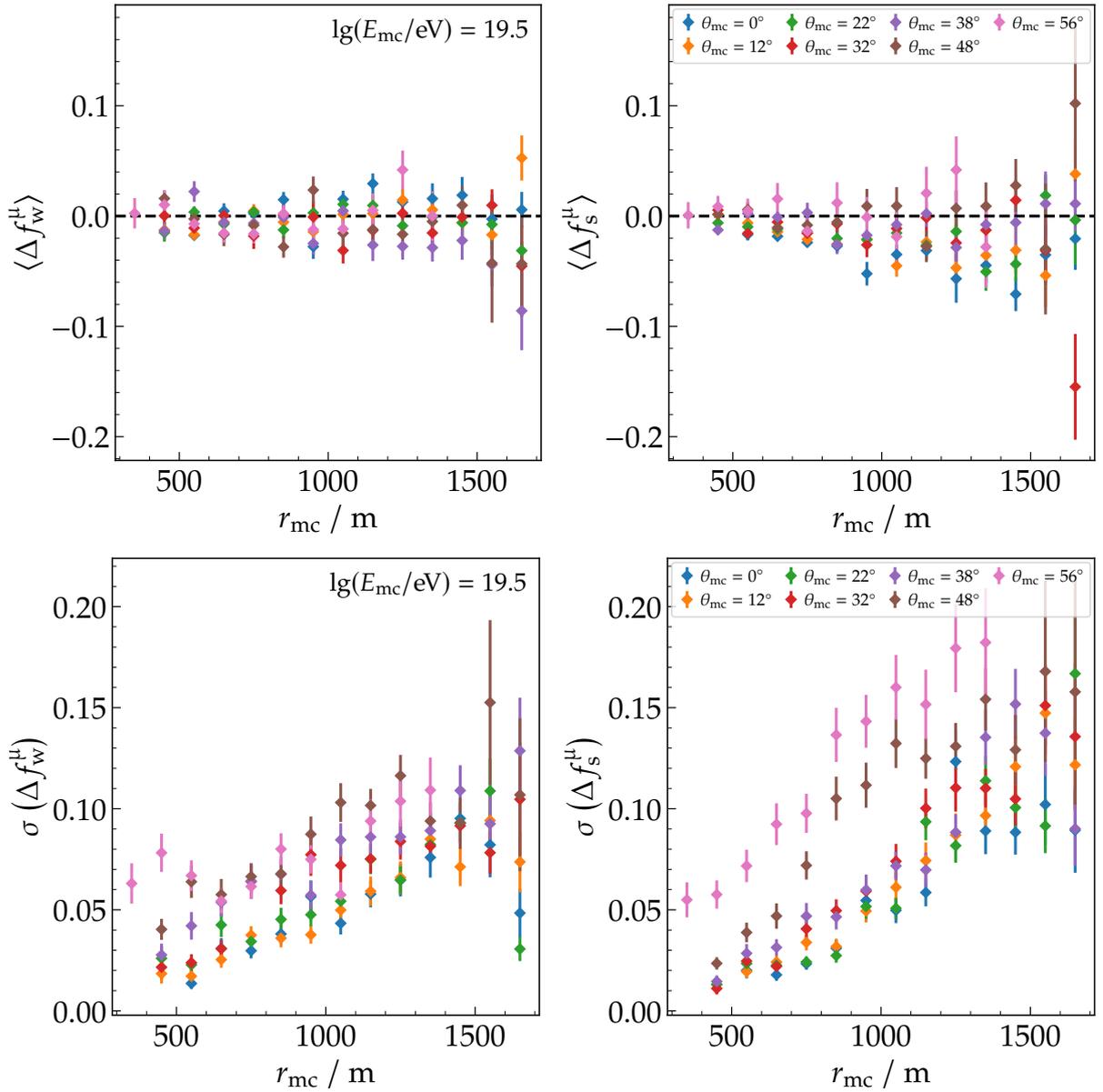


Figure 8.12: Bias (top) and resolution (bottom) of the reconstruction of the muonic fraction in the **WCD** (left) and **SSD** (right) as a function of distance to the shower core. Only showers with an energy of $10^{19.5}$ eV are considered. The results for different zenith angles are given by the different marker colours.

Considering a sample of simulated events where 50% were initiated by proton and 50% by iron nuclei, and selecting only stations presenting unsaturated signals, we estimated the parameters of the reconstruction coefficients (in Eq. (8.12)) by minimising the χ^2 -like function of Eq. (8.14). In the context of the subsequent analysis, we quantify the resolution of the reconstructed muonic fractions as the standard deviation of $\Delta f_{\text{det}}^{\mu}$, denoted as $\sigma(\Delta f_{\text{det}}^{\mu})$.

We commenced the assessment of the reconstruction accuracy investigating

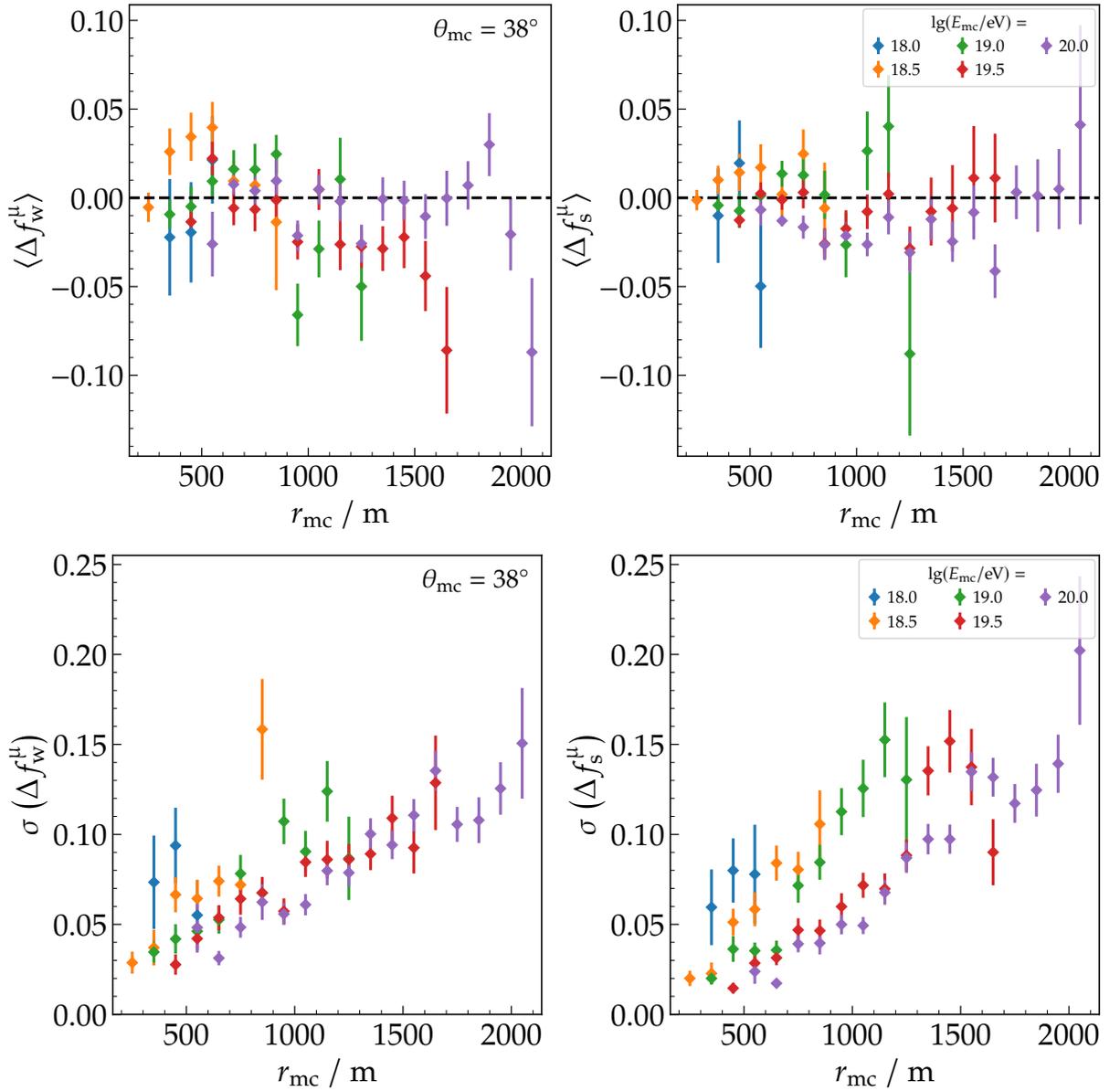


Figure 8.13: Bias (top) and resolution (bottom) of the reconstruction of the muonic fraction in the **WCD** (left) and **SSD** (right) as a function of distance to the shower core. Only showers with a zenith angle of 38° are considered. The results for different primary energies are given by the different marker colours.

the bias and resolution of the predicted muonic fractions in the **WCD** and **SSD**. These quantities are displayed in Fig. 8.12 for distinct zenith angles as a function of the distance to the core of the particle cascade. The bias increases weakly for larger distances, although it is mostly below 6% in both detectors for all different zenith angles.

On the other hand, the resolution evidently increases with distance, and it is also larger for more inclined showers. The resolution is below 5% around 500 m and grows to within 17% at 1500 m, considering the different shower inclinations. Since the

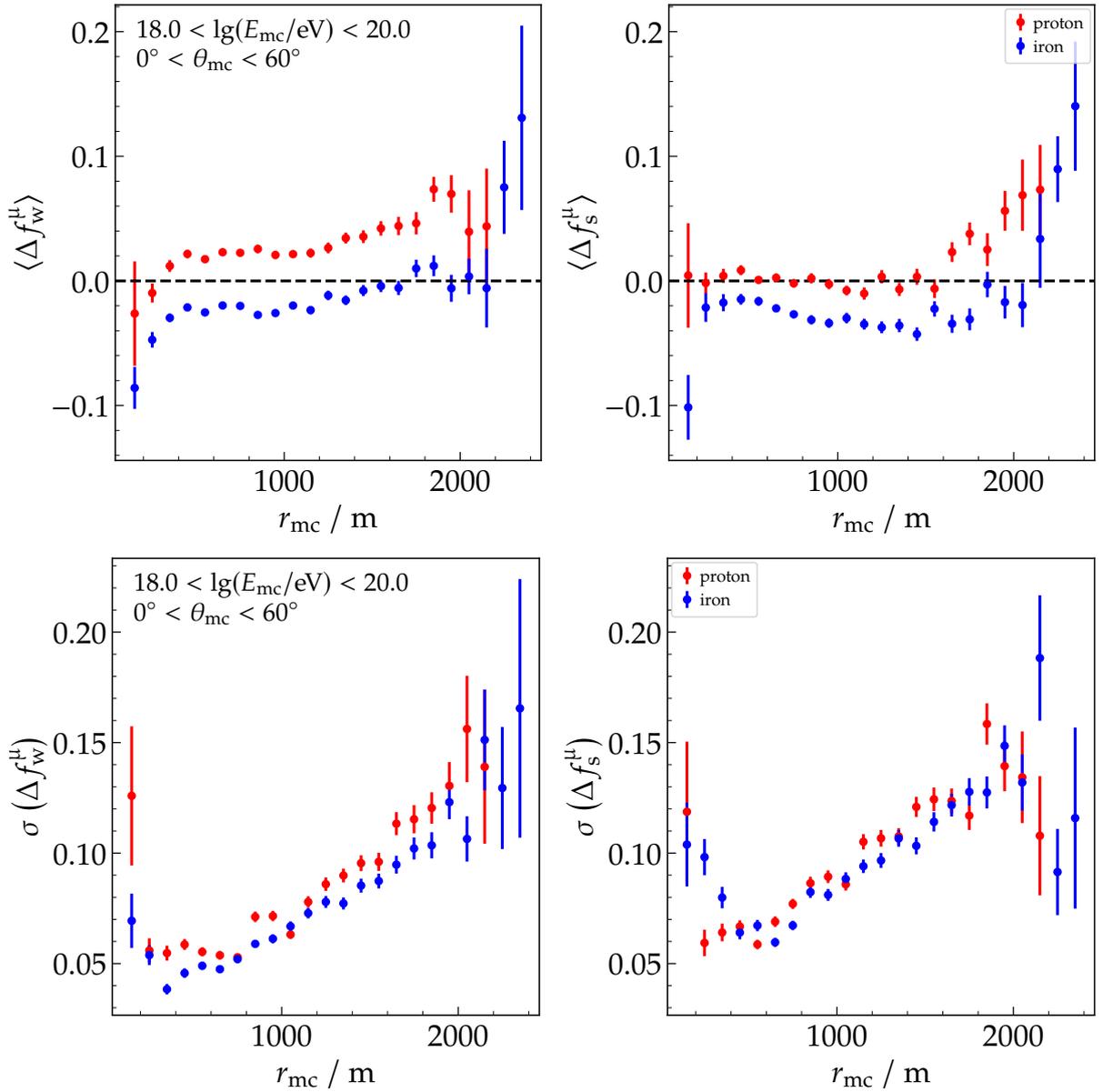


Figure 8.14: Bias (top) and resolution (bottom) of the reconstruction of the muonic fraction in the **WCD** (left) and **SSD** (right) as a function of distance to the shower core. Showers with energy between 1 and 100 EeV and inclination up to 60° were considered. The results for proton (red) and iron (blue) are presented separately. The difference in bias between the two types of primaries is 5% (3%) in the **WCD** (**SSD**).

signals decrease farther from the shower core, their variance grows, given that their Poissonian fluctuations are more pronounced. The same behaviour is exhibited by showers with increasing zenith angles, comparing signals at the same distance to the shower core.

In Fig. 8.13, we depict the bias and resolution for various primary energies whilst keeping the zenith angle constant. The bias of the reconstructed muonic fractions

remains largely within 5% for both the **WCD** and **SSD**. The resolution exhibits an increase with distance, ranging from less than 5% in close proximity to the shower core to approximately 15% at a distance of 2000 m. It is worth noting that showers with higher energy tend to have improved resolution. As discussed earlier, this effect can be attributed to the production of smaller signals by less energetic showers.

We lastly investigated the impact of different compositions of the primary particle on the prediction of the muonic signal. The result is presented in Fig. 8.14, where the bias and resolution are displayed separately for proton and iron nuclei, considering showers with energies between 1 and 100 EeV and inclination up to 60°. We clearly observe that the reconstruction of muonic signals depends on primary composition. Concerning the bias, an effect of approximately 5% (3%) is noted in the **WCD** (**SSD**) between the two primary types. Whilst the resolution is similar for both primaries, iron exhibits a slightly better resolution.

To summarise, the parameterisation of the reconstruction coefficients allowed us to reduce the cut on the signal to 30 **MIP**. This reduction enables the application of the method to a larger number of events. However, even with such a reduction in the signal cut, the reconstruction of muonic fractions presents small biases for showers with different zenith angles, energies, and different distances to the core of the particle cascade. Regarding the unveiled composition bias, we will scrutinise its impact on the reconstruction of events and the subsequent potential for composition separation in Chapter 9.

8.4 RECONSTRUCTION OF TRACE COMPONENTS

We have additionally studied the application of the matrix formalism to reconstruct the traces of the muonic and electromagnetic components. The distribution of arrival time of the different particles contains important information that can be used to infer the composition of the primary, as performed, for instance, in the analyses of muon production depth [95].

The reconstruction of the muonic component of traces consists in the application of Eq. (8.11) to each bin of the total traces of the **WCD** and **SSD**. The electromagnetic traces can be obtained afterwards subtracting the muonic from the total traces for each

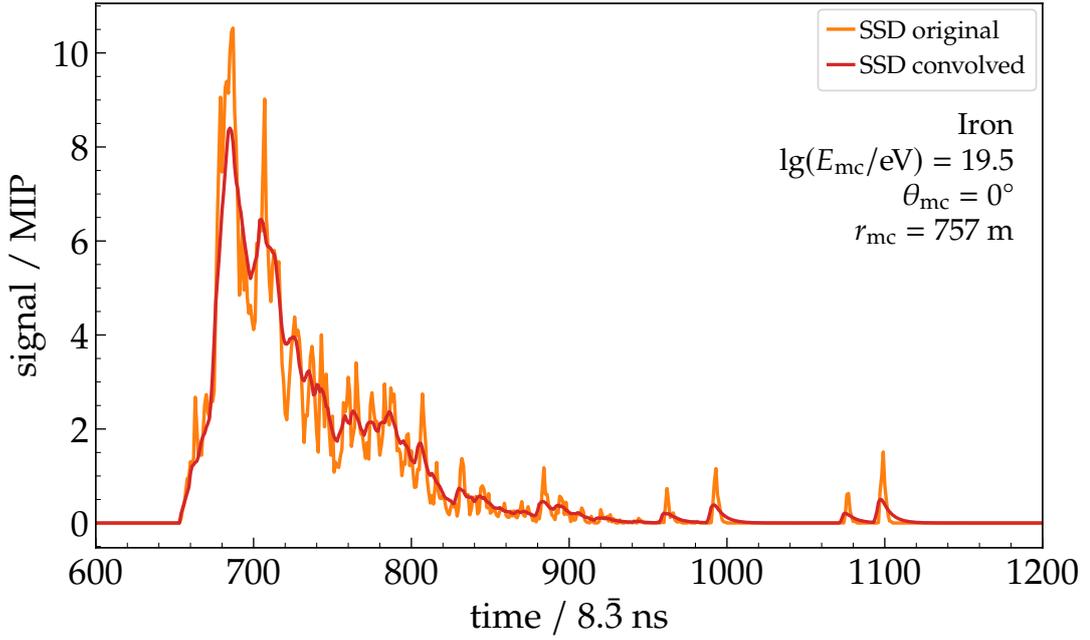


Figure 8.15: Example of an **SSD** trace and the result of its convolution given by Eq. (8.15). The convolved trace is smoother than the original due to the increased response time to match that of the **WCD**.

detector. However, prior to the application of the formalism, it is necessary to preprocess the traces of the detectors.

Firstly, the calibrated traces are typically given in units of **VEM (MIP)** peak, whereas Eq. (8.11) applies to signals in **VEM (MIP)**. Therefore, we renormalise the traces so that their integrals directly give the total signal of the detector in **VEM (MIP)** units.

Secondly, it is mandatory to account for the different response times of the **WCD** and **SSD**. The **SSD** exhibits a faster response compared to the **WCD**, necessitating an adjustment in the **SSD** traces to align with the time bins of the **WCD**. Both detectors respond proportionally to a decaying exponential function, characterised by time constants of approximately 65 ns for the **WCD** and 15 ns for the **SSD**. Therefore, to match the time bins of the **WCD**, the **SSD** trace is expanded through a convolution with an exponential function:

$$T'_s(t) = T_s(t) * \exp\left[-\frac{t + t_0}{\tau_d}\right], \quad (8.15)$$

where $T'_s(t)$ represents the convolved trace, $T_s(t)$ is the original **SSD** trace, and τ_d is the time difference between the **WCD** and **SSD** time constants, calculated as $\tau_w - \tau_s = 65 \text{ ns} - 15 \text{ ns} = 50 \text{ ns}$. To ensure proper alignment, an offset term t_0 is introduced, with $t_0 = 25 \text{ ns}$ found to yield optimal reconstructions through experimentation with

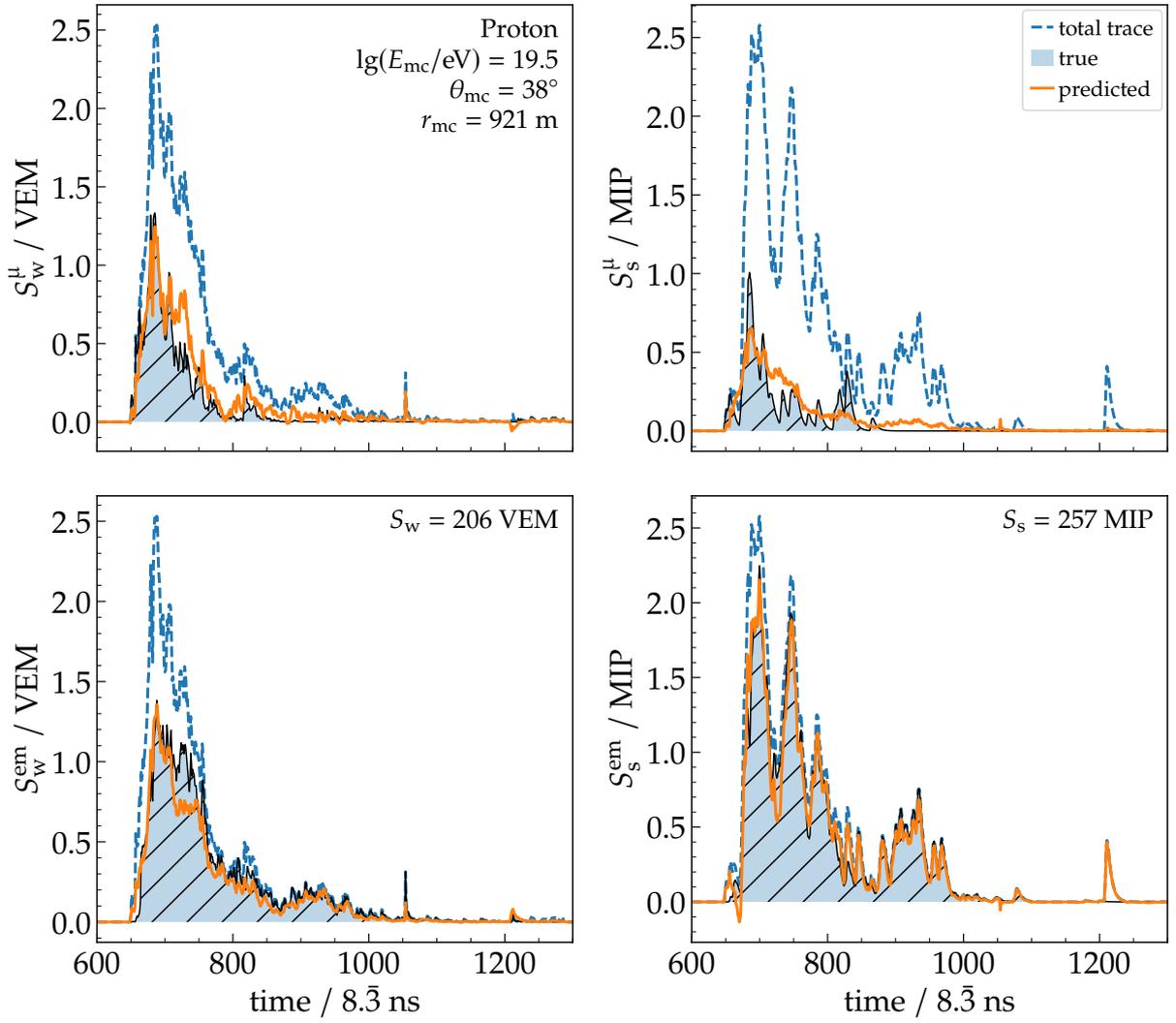


Figure 8.16: Example of reconstruction of the muonic (top) and electromagnetic (bottom) components of the traces in the **WCD** (left) and **SSD** (right). The predicted and true trace components are displayed along with the total one. A visual inspection shows a good prediction using the matrix formalism in this event.

the data. The final trace is renormalised to match the integral of the original **SSD** trace. Fig. 8.15 provides a visual comparison between an **SSD** trace and the result of its convolution as described by Eq. (8.15). It becomes evident that the convolved trace is considerably smoother than the original, a consequence of the adjusted response time to match that of the **WCD**.

After the proper treatment of the total traces of the detectors, we apply Eq. (8.11) to their bins, where the matrix coefficients are given by the parameterisation of Eq. (8.12), as described in Section 8.3. In Fig. 8.16, we present an example of the reconstruction of the muonic and electromagnetic traces by employing this process. Visually, the reconstruction procedure yields a satisfactory prediction of both compo-

nents. However, a more quantitative method is necessary to assess the performance of the trace reconstructions. We additionally verified that small and sparse traces are inadequately reconstructed, likely as a result of their large variation caused by sampling effects. The determination of the situations for which the reconstruction of the trace components is effective should be the subject of further investigation.

CHAPTER 9

Application for event reconstruction and composition studies

In Chapter 8, we demonstrated the applicability of the matrix formalism, which yields accurate estimations of both muonic and electromagnetic signals within the AugerPrime surface stations. In this final chapter of our thesis, we delve into how the separation of these signal components can be leveraged to reconstruct event-level observables characterising the muonic content of cosmic ray showers and their subsequent application for composition identification, especially at the highest energies. This study contributes to the demonstration of the physics potential possible with the upgraded surface detector of the Observatory, aligning with its primary objective.

In Section 9.1, we detail the reconstruction of the muonic shower size and the muonic fraction, additionally introducing zenith-independent observables derived from them. Subsequently, in Section 9.2, we explore the potential utility of these muonic observables for composition separation. It is important to highlight that the analyses presented in this chapter are also conducted utilising the simulation data outlined in Chapter 7.

9.1 RECONSTRUCTION OF MUONIC OBSERVABLES

In the conventional procedure for event reconstruction employing the surface detector, the size of the shower is quantified by the computed signal at the optimal distance of 1000 m from the core of the shower, denoted as $S(1000)$. This quantity, as discussed in Section 3.3, is obtained through a fitting process applied to the signals of the WCDs involved in the event. The functional form employed for this fitting procedure describes the lateral distribution of the particle cascade, commonly referred to as the lateral distribution function (LDF). To characterise the size of the muonic component within the shower, we employ an analogous methodology, reconstructing the muonic signal produced by the particle cascade at a designated reference distance from the shower core.

For each simulated event, we employ the matrix formalism, utilising the parameterisation of the reconstruction coefficients described in Section 8.3 to the total signals of the WCDs and SSDs, thereby yielding their respective muonic component. To maintain the assumption of comparable particle fluxes in both detectors, an essential condition for the applicability of the method, we only consider stations exhibiting SSD signals exceeding 10 MIP.

Having extracted the muonic signals from the participating stations, we proceed to subject their lateral distribution to a fitting procedure. The empirical choice for the functional form of the muonic LDF is a modified version of the Nishimura-Kamata-Greisen function, expressed as follows:

$$S_w^\mu(r) = S_w^\mu(r_{\text{opt}}) \left(\frac{r}{r_{\text{opt}}} \right)^\beta \left(\frac{r + r_s}{r_{\text{opt}} + r_s} \right)^\gamma, \quad (9.1)$$

where $r_s = 700$ m, as in the standard reconstruction. The parameter r_{opt} maintains the established value of 1000 m, analogous to its utilisation in the context of $S(1000)$. This distance is predominantly determined by the spacing between the stations within the array [56]. However, we do explore the impact on the composition separation when varying the value of the optimal distance.

To fit the free parameters $S_w^\mu(r_{\text{opt}})$, β , and γ , specific constraints must be met. A minimal station multiplicity within the event is essential, as is a sparse pattern of

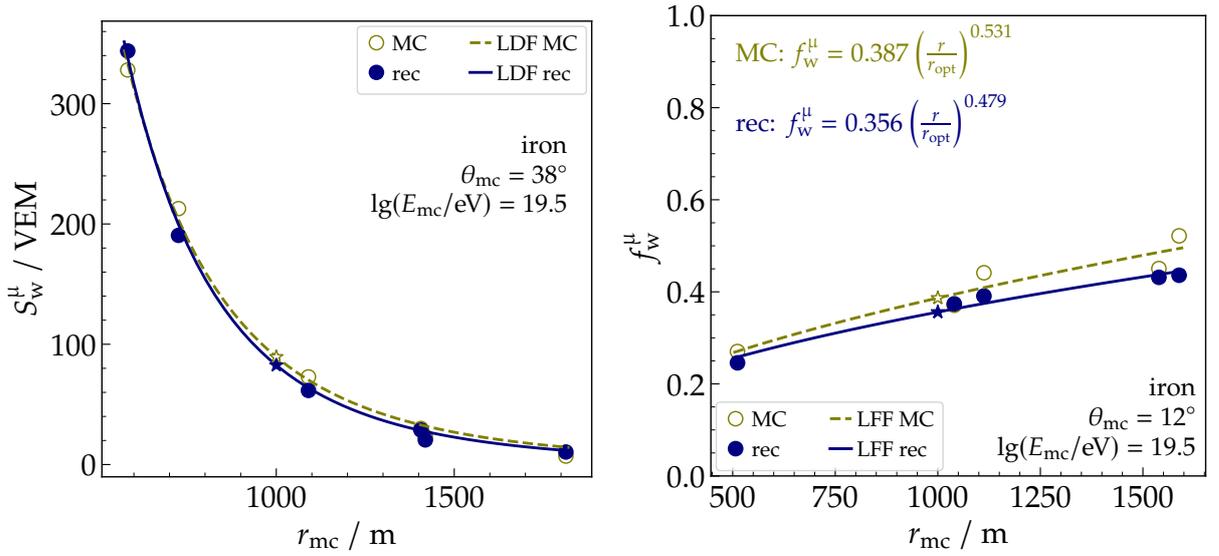


Figure 9.1: Lateral profiles of muonic signals (left panel) and fractions (right panel) in example events. The outcomes of employing the matrix formalism for signal reconstruction (navy blue) as compared to the corresponding true values obtained from Monte Carlo data (olive green) are depicted. Star markers denote the derived estimations of $S_w^\mu(r_{\text{opt}})$ and $f_w^\mu(r_{\text{opt}})$, evaluated at $r_{\text{opt}} = 1000 \text{ m}$, arising from the data fitting procedure to Eqs. (9.1) and (9.2), respectively.

their radial distribution. Consequently, we exclusively consider events featuring three or more stations with reconstructed muonic signals and ensure that the difference between the distance of the nearest and farthest stations from the core of the shower exceeds 500 m.

In addition to quantifying the muonic component of the showers using the absolute observable $S_w^\mu(r_{\text{opt}})$, we have explored the reconstruction of the muonic fraction of the event. Within a WCD, the fraction is defined as the ratio between its muonic and total signals. Notably, the muonic fraction of the shower holds greater significance as it resides within the confined interval of zero to one, directly reflecting the proportion of the muonic component relative to the entirety of the particle cascade.

To characterise the muonic fraction of the shower, we have followed a methodology similar to that employed for the absolute muonic size. By fitting the fractions measured by the WCDs, we establish a functional representation of its radial behaviour, referred to as the lateral fraction function (LFF). In this context, we adopt a power law to model its form:

$$f_w^\mu(r) = f_w^\mu(r_{\text{opt}}) \left(\frac{r}{r_{\text{opt}}}\right)^\delta, \quad (9.2)$$

where both the parameter δ and the muonic fraction at the optimal distance $f_w^\mu(r_{\text{opt}})$ are determined through the fitting process. Consistent with previous discussions, the optimal distance is maintained at 1000 m from the core of the shower. The constraints referring to station multiplicity and their spatial distribution, as imposed in the fitting procedure for the absolute muonic signal, are upheld here as well.

It is important to emphasise that the fitting procedures should be further investigated and refined. As an example, our current approach employs a standard χ^2 minimisation technique. However, the quality of the fits can be considerably enhanced by adopting a maximum likelihood approach that takes into consideration factors such as station trigger and saturation probabilities. Furthermore, a parameterisation of the coefficients β , γ , and δ through methods like Monte Carlo simulations or even analysis of real data may yield improvements for the determination of $S_w^\mu(r_{\text{opt}})$ and $f_w^\mu(r_{\text{opt}})$, particularly for events with a low multiplicity of stations.

In Fig. 9.1, we illustrate examples of events where the radial distributions of absolute muonic signals (left panel) and fractions (right panel) are subjected to the fitting procedures based on the functions outlined in Eqs. (9.1) and (9.2), respectively. In the depiction, both the outcomes of the reconstruction with the matrix formalism (navy blue) and derived from the true Monte Carlo muonic component (olive green) are presented. The plotted star markers denote the estimations of $S_w^\mu(r_{\text{opt}})$ and $f_w^\mu(r_{\text{opt}})$, obtained through the fitting processes employing $r_{\text{opt}} = 1000$ m.

Whilst the reconstructed outcomes depicted in Fig. 9.1 exhibit a striking visual resemblance to their Monte Carlo counterparts, a rigorous assessment of the accuracy in determining $S_w^\mu(1000)$ and $f_w^\mu(1000)$ through the matrix formalism is crucial. To this end, we introduce the metrics $\Delta S_w^\mu(1000)$ and $\Delta f_w^\mu(1000)$ for the muonic shower size and fraction, respectively.

For the muonic shower size, $\Delta S_w^\mu(1000)$ is defined as the relative difference between the reconstructed value $\hat{S}_w^\mu(1000)$ and the true Monte Carlo value $S_w^\mu(1000)$, expressed as:

$$\Delta S_w^\mu(1000) = \frac{\hat{S}_w^\mu(1000) - S_w^\mu(1000)}{S_w^\mu(1000)}. \quad (9.3)$$

Conversely, for the muonic fraction, the quantity $\Delta f_w^\mu(1000)$ is defined as the difference between the reconstructed value $\hat{f}_w^\mu(1000)$ and the corresponding true value $f_w^\mu(1000)$

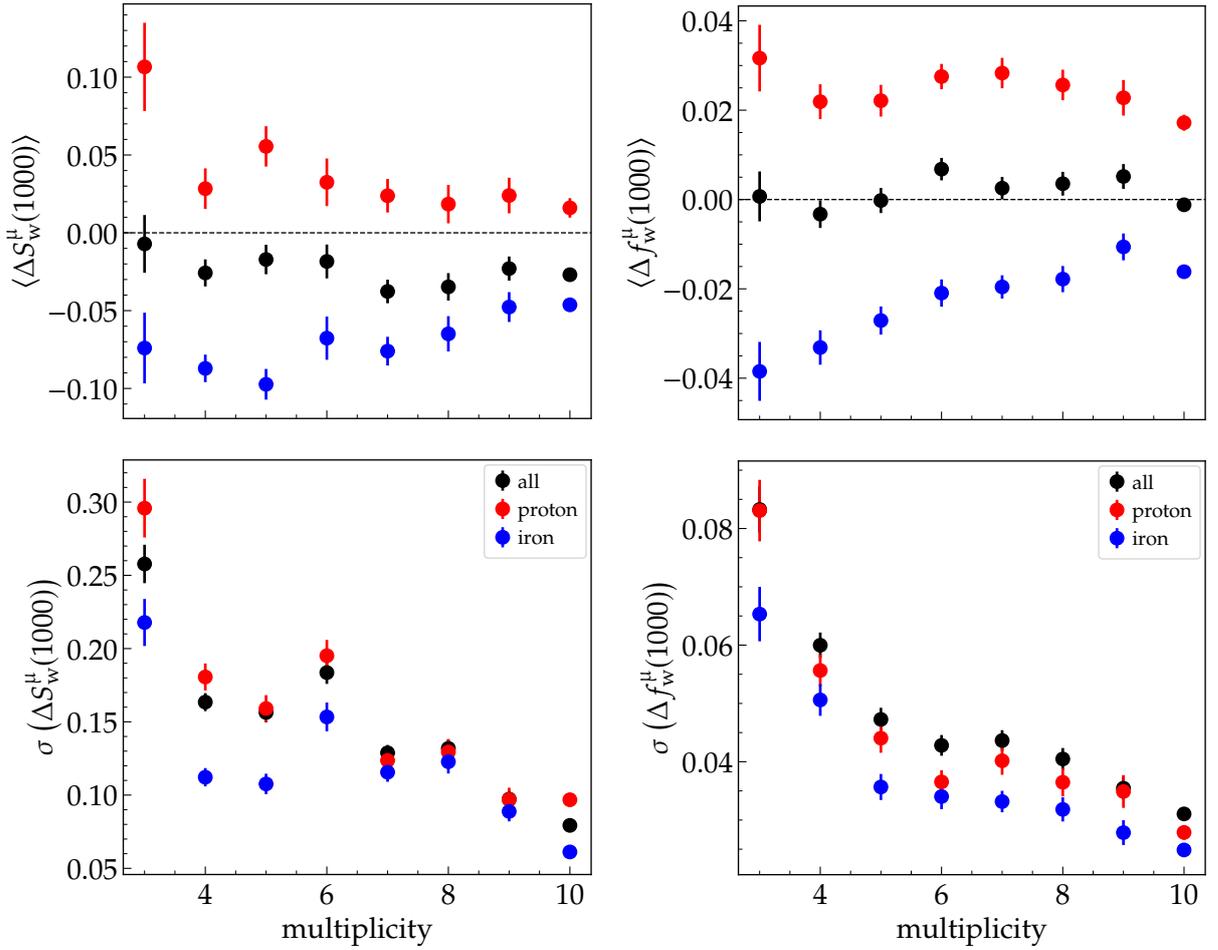


Figure 9.2: Bias (top panels) and resolution (bottom panels) with respect to event multiplicity, depicted for the muonic shower size (left) and fraction (right). Each colour represents a distinct type of primary particle. The final point on the multiplicity axis corresponds to events involving ten or more stations. Markedly, a disparity of approximately 10% (5%) is noticeable between proton and iron reconstructions for $S_w^\mu(1000)$ ($f_w^\mu(1000)$).

derived from the Monte Carlo data:

$$\Delta f_w^\mu(1000) = \hat{f}_w^\mu(1000) - f_w^\mu(1000). \quad (9.4)$$

Utilising the aforementioned definitions, we present in Fig. 9.2 both the bias (top panels) and the resolution (bottom panels) intrinsic to the reconstruction of the muonic shower size (left) and fraction (right) as a function of the event multiplicity for distinct species of primary particles. Upon initial inspection, the resolution of $S_w^\mu(1000)$ ($f_w^\mu(1000)$) demonstrates an evident enhancement, transitioning from approximately 25% (8%) for events featuring a multiplicity of three, to 8% (3%) for events encompassing ten or more stations. This behaviour aligns with expectations, as a higher

event multiplicity facilitates a more robust fitting performance.

Considering all the primary particles, the bias associated with the muonic shower size remains within 5%, whereas it approaches zero for the muonic shower fraction. Nevertheless, these biases exhibit a dependency on composition, as indicated by the discrepancy of approximately 10% between proton and iron for $S_w^\mu(1000)$ and 5% for $f_w^\mu(1000)$. This composition-related bias is a manifestation of the same feature observed in the reconstruction of individual muonic signals through the application of the matrix formalism, as demonstrated in Section 8.3 (refer to Fig. 8.14). We will explore the implications of such biases in the context of identifying distinct primary particles in Section 9.2.

In addition to its sensitivity to the composition of the primary particle, the observable $S_w^\mu(1000)$ also exhibits a dependency on their energy and zenith angle. As the standard event reconstruction yields information regarding the latter two variables, the study of composition separation directly through $S_w^\mu(1000)$ necessitates the arrangement of events based on their energy and zenith angle. This study will be presented in Section 9.2 (refer to Fig. 9.7). However, to reduce the complexity of this analysis, we introduce a parameterisation to characterise the zenith angle dependence of $S_w^\mu(1000)$. This procedure yields a new observable that is independent of the inclination of the primary particle.

Letting $S_w^\mu(1000)$ be denoted as S_{1000}^μ , we assume that its dependence on the zenith angle (θ) can be factored into a distinct function, denoted as $g(\theta)$. Consequently, we express S_{1000}^μ as $\mathcal{S}_\mu(E, A)g(\theta)$, wherein \mathcal{S}_μ encapsulates the zenith-independent muonic size of the particle cascade. It is worth noting that \mathcal{S}_μ continues to depend on the energy (E) and composition (A) of the primary.

By considering the ratio of S_{1000}^μ at a generic zenith angle to that at a reference angle θ_{ref} , whilst maintaining the same energy and composition, we derive

$$\frac{S_{1000}^\mu(\theta, E, A)}{S_{1000}^\mu(\theta = \theta_{\text{ref}}, E, A)} = \frac{\mathcal{S}_\mu(E, A)g(\theta)}{\mathcal{S}_\mu(E, A)g(\theta_{\text{ref}})} = f(\theta) \quad (9.5)$$

$$\implies S_{1000}^\mu(\theta = \theta_{\text{ref}}, E, A) = \frac{S_{1000}^\mu(\theta, E, A)}{f(\theta)}. \quad (9.6)$$

Consequently, the quantity $S_{1000}^\mu(\theta = \theta_{\text{ref}})$ is free from the influence of the shower

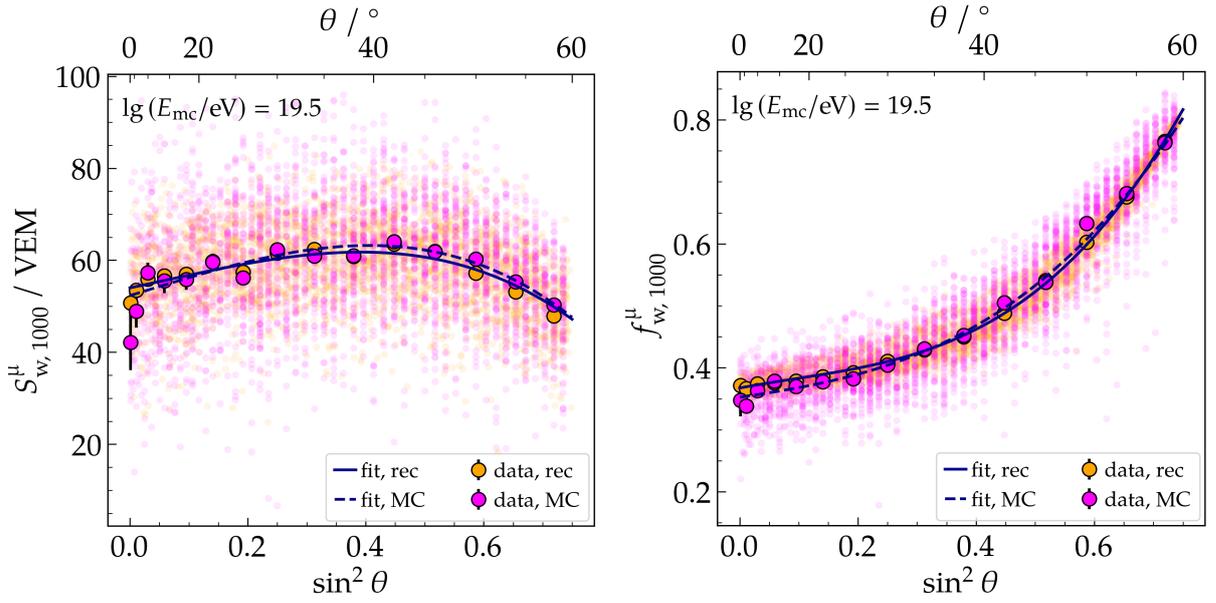


Figure 9.3: Dependency of muonic shower size (left) and fraction (right) on the zenith angle of the shower. The primaries possess a Monte Carlo energy of $10^{19.5}$ eV. The results of the fitting procedure with observables derived from the reconstruction with the matrix formalism (orange) and with the true Monte Carlo values (magenta) are represented by the solid and dashed lines, respectively. The fit process utilises the function detailed in Eq. (9.7).

inclination and can be interpreted as the value of S_{1000}^{μ} that the particle cascade would exhibit had it possessed a zenith angle of θ_{ref} . A similar rationale applies to the muonic fraction of the shower. This allows us to derive the zenith-independent observable $f_{1000}^{\mu}(\theta_{\text{ref}})$.

We choose the reference angle to be 38° , motivated by its close representation of the median of the inclination of events detected with the Observatory. For the functional form of the zenith dependence of both $S_w^{\mu}(1000)$ and $f_w^{\mu}(1000)$, we adopt a third-degree polynomial in terms of the variable $x \equiv \sin^2 \theta - \sin^2 \theta_{\text{ref}}$, thus expressed by

$$f(\theta) = 1 + ax + bx^2 + cx^3, \quad (9.7)$$

where the parameters a , b and c remain subject to determination.

In Fig. 9.3, we depict the dependency of $S_w^{\mu}(1000)$ and $f_w^{\mu}(1000)$ on shower inclination, utilising events with Monte Carlo energy of $10^{19.5}$ eV. Regarding the absolute muonic size, it increases with the zenith angle until approximately 42° , at which point a diminishing pattern is evidenced. The initial ascent can be attributed to the more pronounced development of the particle cascade for larger zenith angles. Conversely,

Table 9.1: Parameters of the function characterising the zenith angle dependency of S_{1000}^{μ} and f_{1000}^{μ} relative to their values at an angle of 38° (as defined by Eq. (9.7)). The presented parameters were derived from fitting the corresponding reconstructed observables employing the matrix formalism.

| | a | b | c |
|------------------------------|-------------------|------------------|------------------|
| $S_{1000}^{\mu}(38^{\circ})$ | 0.050 ± 0.027 | -1.49 ± 0.05 | -1.27 ± 0.30 |
| $f_{1000}^{\mu}(38^{\circ})$ | 0.973 ± 0.009 | 2.25 ± 0.02 | 2.62 ± 0.10 |

the subsequent attenuation after 42° likely stems from transitioning beyond the zenith angle corresponding to the peak of the muonic cascade, when a reduction in its overall content occurs.

On the other hand, the muonic fraction demonstrates an increase as the inclination of the shower grows. This behaviour aligns with expectations, as inclined events experience greater absorption of the electromagnetic component within the atmosphere. It is important to note that although the absolute muonic size diminishes after 42° , the ascent in the muonic fraction becomes more rapid. This observation indicates a stronger suppression of the electromagnetic component for these particular inclinations.

Furthermore, we also present in Fig. 9.3 the result of the fitting process applied individually to the reconstructed and Monte Carlo observables using the function detailed in Eq. (9.7), which reveals a remarkable similarity between the two. The extracted parameters a , b , and c , derived solely from the reconstructed S_{1000}^{μ} and f_{1000}^{μ} utilising the matrix formalism are presented in Table 9.1. We employ these values to compute the corresponding zenith-independent quantities $S_{1000}^{\mu}(38^{\circ})$ and $f_{1000}^{\mu}(38^{\circ})$, which we will refer to as S_{38}^{μ} and f_{38}^{μ} .

To investigate the impact of applying the function $f(\theta)$ to the corresponding S_{1000}^{μ} and f_{1000}^{μ} , we present the calculated S_{38}^{μ} and f_{38}^{μ} as functions of the zenith angle in Fig. 9.4. Upon analysing all primary particles, as utilised in the determination of the parameters of function $f(\theta)$, we verified that the zenith dependency is effectively neutralised for both observables across the different energies examined (compare to Fig. 9.3). However, considering individual primary species, particularly in the case of iron, a residual dependency on the shower inclination remains perceptible.

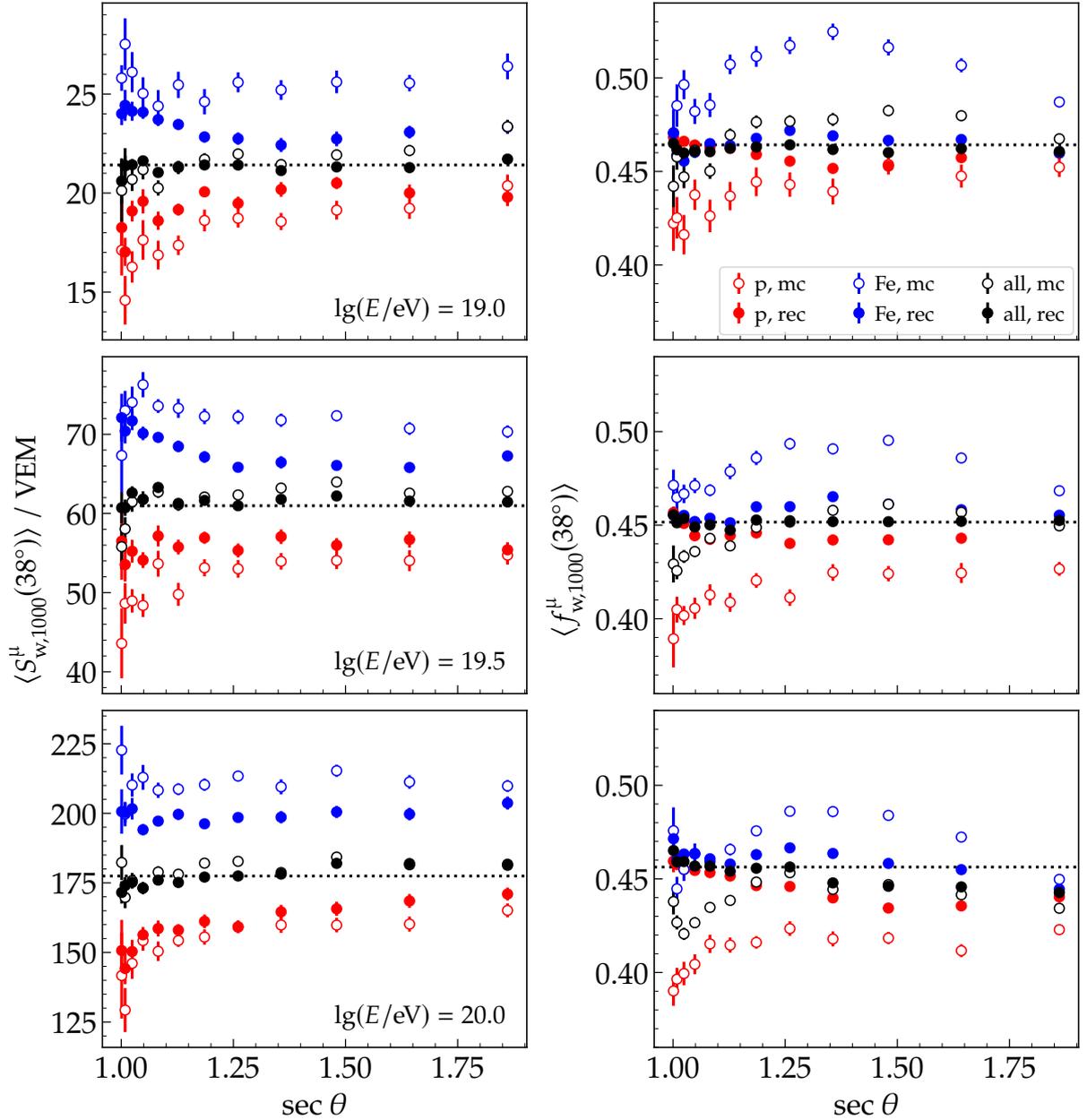


Figure 9.4: Computed observables S_{38}^{μ} (left) and f_{38}^{μ} (right) depicted as functions of the zenith angle of the shower. The panels correspond to primaries with energies of $10^{19.0}$ eV (top), $10^{19.5}$ eV (middle), and $10^{20.0}$ eV (bottom). Different colours indicate distinct primary species. Whilst the inclination-related dependency of S_{1000}^{μ} and f_{1000}^{μ} is eliminated when considering all primary particles, a residual dependency is still noticeable when examining individual primaries.

9.2 APPLICATION FOR COMPOSITION SEPARATION

In this section, we present our investigation concerning the identification of distinct primary cosmic rays, employing the matrix formalism in conjunction with the observables outlined in Section 9.1. Prior to delving into the details of these analyses, we introduce a metric designed to assess the discriminative capability between proton and iron primaries, as offered by a particular observable X . This metric is referred to as the merit factor (M.F.), and is expressed as

$$\text{M.F.}(X) = \frac{|\langle X_p \rangle - \langle X_{\text{Fe}} \rangle|}{\sqrt{\sigma^2(X_p) + \sigma^2(X_{\text{Fe}})}}, \quad (9.8)$$

where $\langle X \rangle$ represents the mean value of the observable X , and $\sigma(X)$ denotes its standard deviation. The subscripts "p" and "Fe" correspond to proton and iron primaries, respectively. It is important to note that the merit factor provides a quantitative measure of the discrepancy that a given observable exhibits between proton and iron primaries, in terms of the associated uncertainty in this disparity. Consequently, this metric considers the impact of the inherent fluctuations of the observable for each primary, alongside the difference in their mean values.

Heavier primary cosmic rays induce the production of a larger amount of muons during the development of their particle cascade. This property enables the utilisation of the muonic signal and its derived observables with the potential for discriminating of distinct primary species. We initiate the study of composition separation by directly investigating the muonic signals recorded by the upgraded surface detector and generated by different primary particles.

The upper panels of Fig. 9.5 unveil the lateral distribution of the true Monte Carlo muonic signal and fraction as detected by the WCDs for primaries of proton and iron with an energy of $10^{19.5}$ eV and zenith angle of 22° . As expected, iron primaries yield larger muonic signals and fractions than protons. The scatter plots allow us to verify that the absolute muonic signal presents a more distinct separation between the two primary types compared to the muonic fraction. In the former case, the impact of significant Poisson fluctuations for lower signals beyond a distance of 1400 m prevents a clear distinction between the two primary types under these conditions.

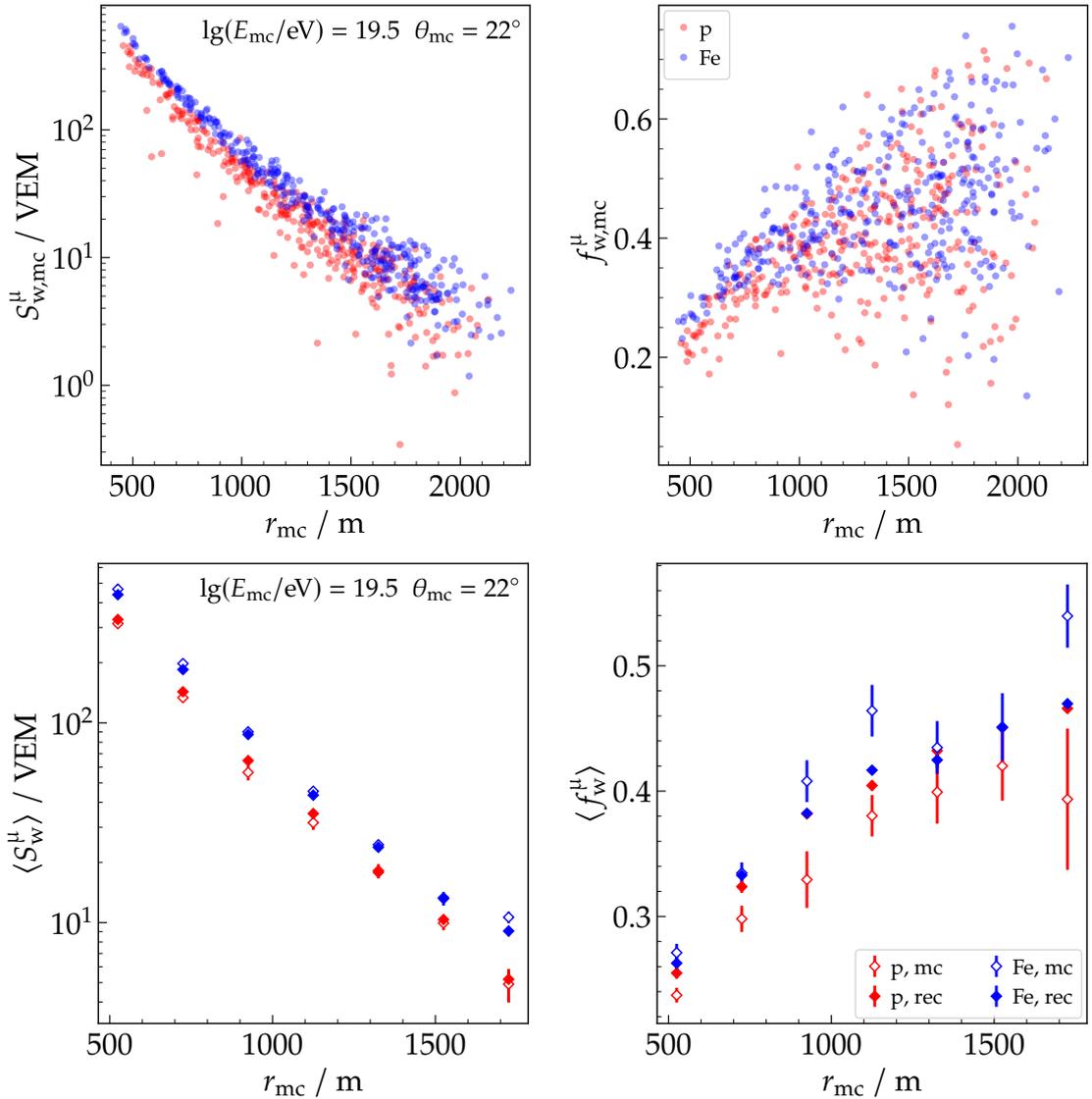


Figure 9.5: Top: Lateral distribution of the true Monte Carlo muonic signals (left) and muonic fractions (right) in the **WCD** for primaries of proton (red) and iron (blue). Bottom: Corresponding mean values (open markers) alongside the result of the application of the matrix formalism for estimation of the muonic signals (closed markers). The primaries have a zenith angle of 22° and an energy of $10^{19.5}$ eV.

In the lower panels of Fig. 9.5, we depict the corresponding mean values juxtaposed with the outcomes of employing the reconstruction using the matrix formalism to estimate the muonic signals in the **WCD**. This analysis reveals that the discrimination between distinct primaries diminishes when utilising the reconstructed signals. This attenuation is anticipated and attributed to the previously noted composition bias, reported in Section 8.3 (and also evident in Fig. 9.2). Whilst this bias exerts a relatively modest influence on the absolute muonic signals, it significantly compromises composition separation when relying on the muonic fractions.

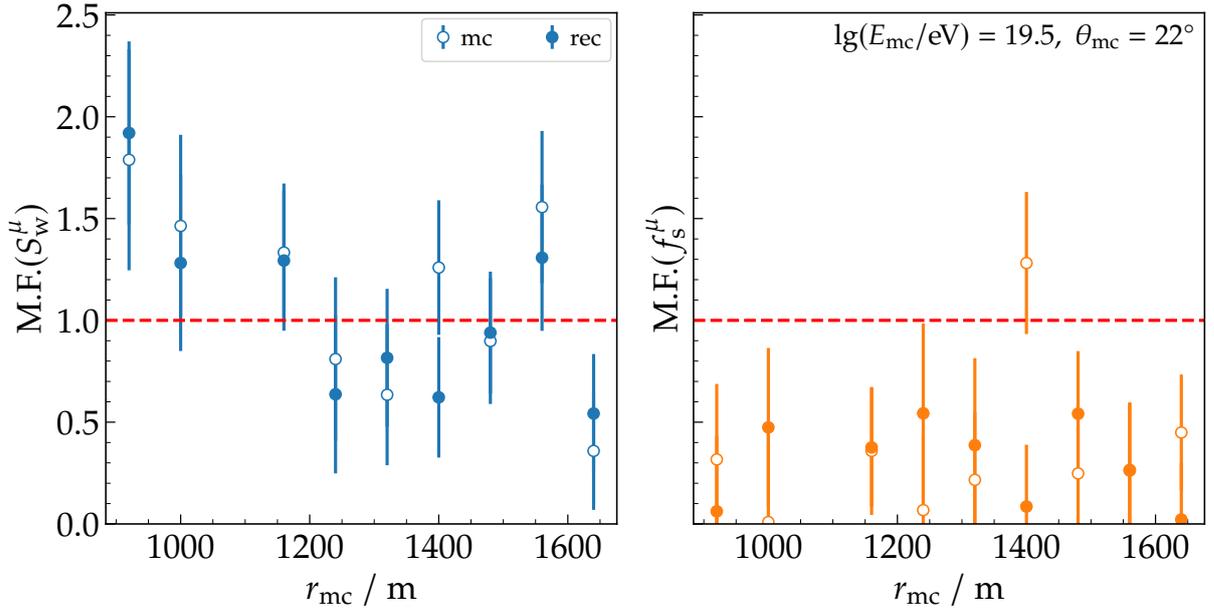


Figure 9.6: Lateral dependence of the merit factor for muonic signals (left) and fractions (right) recorded by the water-Cherenkov detectors. This analysis utilises primary particles possessing an inclination of 22° and an energy of $10^{19.5}$ eV. The results for both true Monte Carlo values (open markers) and those reconstructed using the matrix formalism (closed markers) are displayed. The absolute muonic signals exhibit a more pronounced ability to differentiate between distinct primaries when compared to the muonic fractions.

The aforementioned discussion is quantified by the merit factors displayed in Fig. 9.6, which are also presented as functions of the distance to the shower core. An immediate observation is the limited capacity of composition identification exhibited when relying solely on the muonic fraction of the stations. The associated merit factor remains consistently below 0.5 across nearly the entire range of distances to the shower core. Conversely, the merit factor associated with the absolute muonic signal in the **WCD** demonstrates an enhancement for stations situated closer to the shower core, reaching a value of approximately 1.8. However, even beyond a distance of 1200 m, the merit factor maintains a value above 0.5.

We proceeded with an examination of the applicability of the muonic shower size and fraction, reconstructed as described in Section 9.1, to distinguish between different primary particles. As mentioned previously, the muonic shower size depends not only on the composition of the primaries but also on their energy and zenith angle. Consequently, the left panels of Figs. 9.7 and 9.8 depict the mean values of $S_w^\mu(1000)$ and $f_w^\mu(1000)$, respectively, for proton and iron primaries as a function of the shower

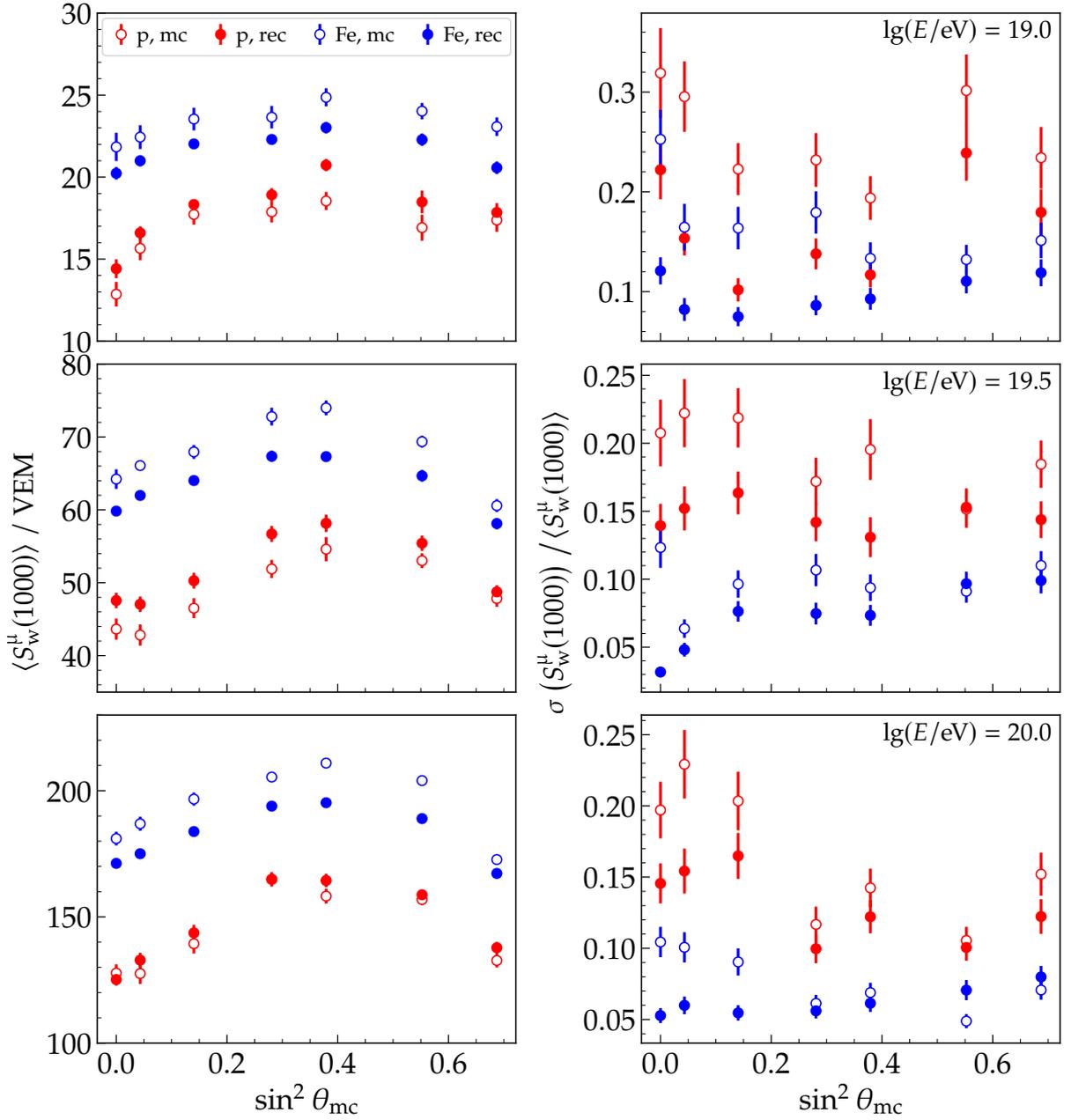


Figure 9.7: Dependence on the zenith angle of the mean value (left) and relative variation with respect to the mean (right) of $S_w^H(1000)$. The analysis encompasses primaries with energies of $10^{19.0}$ eV (top row), $10^{19.5}$ eV (middle row) and $10^{20.0}$ eV (bottom row). The outcomes for proton (red) and iron (blue) primaries are presented individually, and superimposed are the results obtained through the application of the matrix formalism (closed markers) and their corresponding Monte Carlo counterparts (open markers).

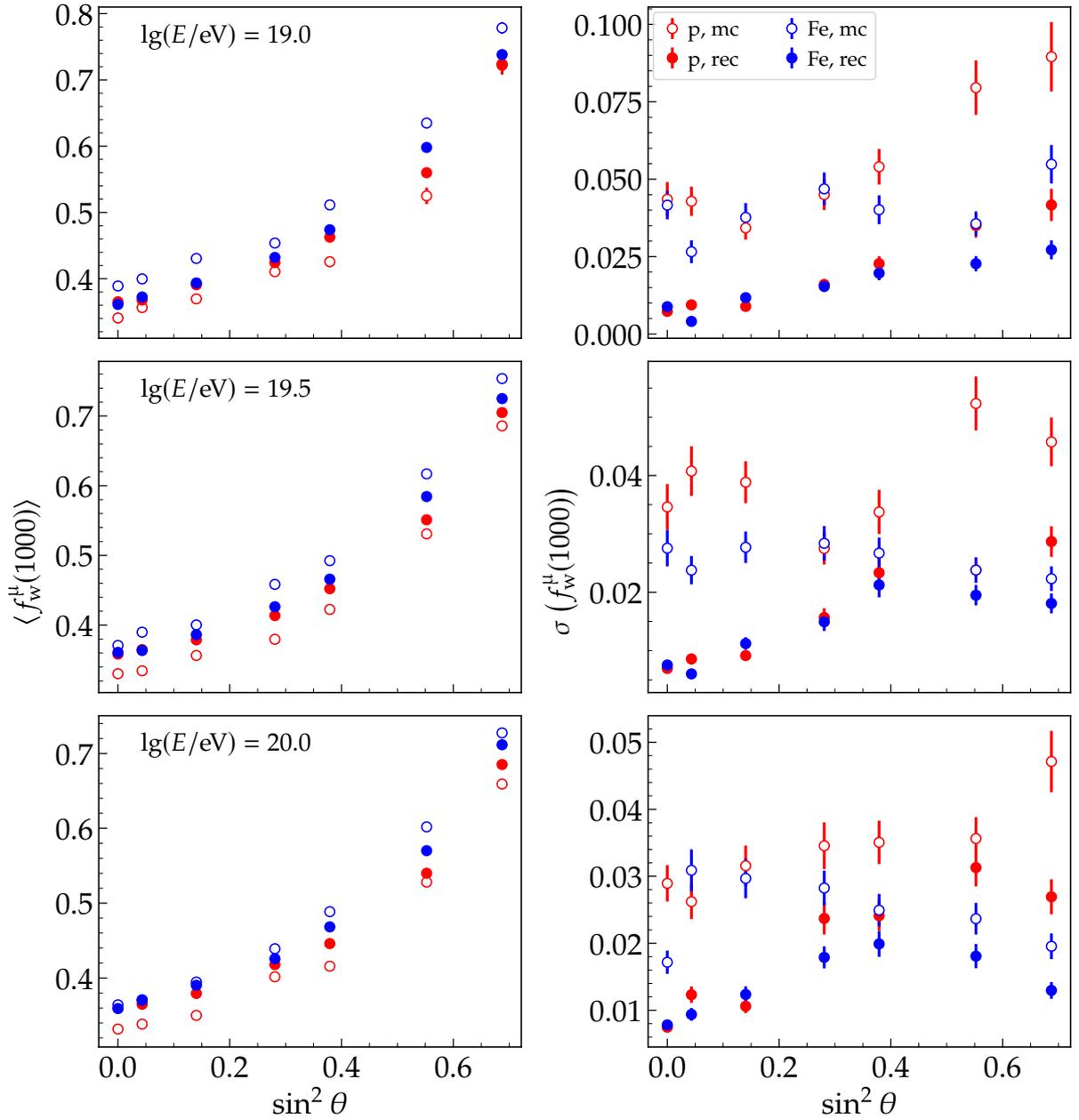


Figure 9.8: Dependence on the zenith angle of the mean value (left) and corresponding standard deviation (right) of $f_w^H(1000)$. The results for proton (red) and iron (blue) primaries are presented independently. The closed markers signify the outcomes obtained through the application of the matrix formalism, whilst their corresponding Monte Carlo counterparts are denoted by the open markers. The analysis covers primaries with energies of $10^{19.0}$ eV (top row), $10^{19.5}$ eV (middle row) and $10^{20.0}$ eV (bottom row).

inclination and separately for different energies.

The behaviour of both types of primary particles concerning their zenith angle is similar and aligns with the prior discussion in Section 9.1, which occurred during the derivation of the corresponding observables S_{38}^{μ} and f_{38}^{μ} . Unsurprisingly, the composition bias highlighted in Fig. 9.2 renders a reduction in the discriminatory capability of both $S_w^{\mu}(1000)$ and $f_w^{\mu}(1000)$, with the latter being affected more pronouncedly.

On the right side of Fig. 9.7, we display the relative variation of $S_w^{\mu}(1000)$ with respect to its mean value. This relative fluctuation is larger for proton primaries in comparison to iron, as the latter primary yields larger signals at a fixed distance from the shower core compared to protons. Additionally, we found evidence that the reconstructed observable exhibits less variability than its Monte Carlo counterpart. This effect is attributed to the parameterisation of the reconstruction coefficients employed by the matrix formalism. These coefficients effectively average the true muonic signal over different configurations of zenith angle, energy, and distance to the shower core.

As the energy of the primary particle grows, the observed variation tends to decrease, an effect stemming again from the greater signals generated by more energetic showers. For primary particles possessing an energy of $10^{19.5}$ eV, the relative fluctuation of the reconstructed $S_w^{\mu}(1000)$ amounts to approximately 14% and 7% for proton and iron, respectively. Regarding the fluctuations of $f_w^{\mu}(1000)$ on the right side of Fig. 9.8, the qualitative discussion outlined above remains applicable. Notably, the observed variations in this case are predominantly contained within 5%.

The corresponding merit factors are illustrated in Fig. 9.9, where we explore the outcomes employing optimal distances of 800 m, 1000 m, and 1200 m. The reduced capability of composition separation exhibited by the muonic shower fraction becomes apparent through its smaller merit factors in comparison to the muonic shower size. It is noteworthy to stress the pronounced impact of the composition bias, stemming from the application of the matrix formalism, on the result of $f_w^{\mu}(r_{\text{opt}})$, with its merit factors consistently assuming values below one.

Conversely, concerning $S_w^{\mu}(r_{\text{opt}})$, we initially observe a closer similarity between the reconstruction performed using the matrix formalism and the results derived from the Monte Carlo values. In this context, the differences between distinct optimal distances remain modest, with no definitive preference for a particular distance.

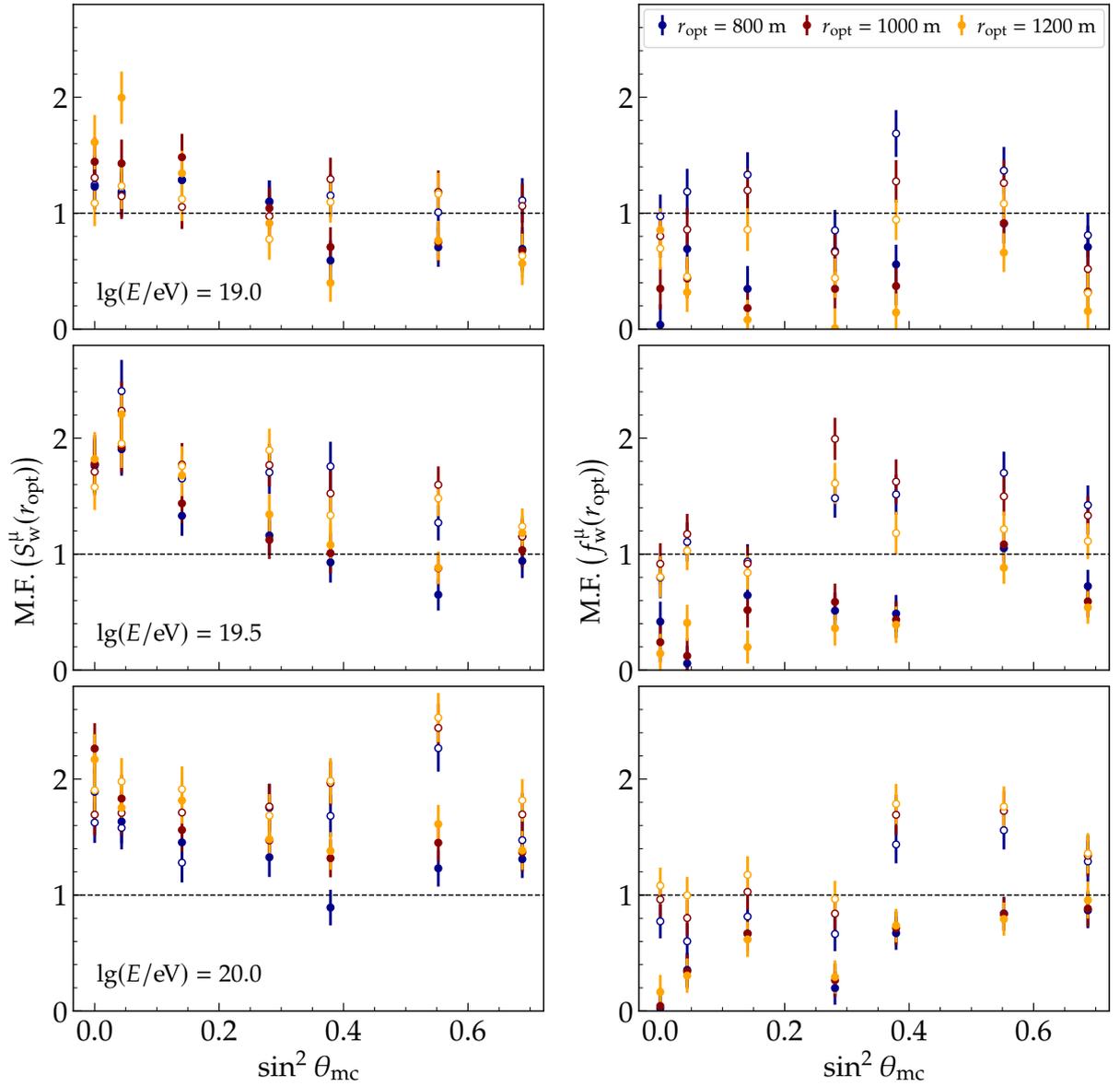


Figure 9.9: Merit factor for the muonic shower size (left) and fraction (right) as a function of the zenith angle. Distinct colours correspond to different reference distances utilised for computing the corresponding observables. Open markers denote the results obtained from Monte Carlo values, whilst closed markers indicate the outcomes from the application of the matrix formalism. The panels individually refer to primary particles with energies of 10^{19} eV (top), $10^{19.5}$ eV (middle) and 10^{20} eV (bottom panels).

Additionally, a tendency of enhanced merit factors becomes evident for increasingly energetic showers and shallower inclinations. This enhancement can be attributed to the larger, less variable signals produced at the designated optimal distance under these conditions. For primary particles featuring energy of 10^{20} eV, the corresponding merit factors vary from approximately two for shallow inclinations to around 1.2 beyond 30° .

Up to this point, our analyses explored the potential for composition diffe-

rentiation using the true Monte Carlo values for the energy of the primary particles. In practice, we solely possess access to the reconstructed energy of the event, which is estimated through its relationship with the shower size estimator, the observable $S(1000)$, in the case of the surface detector. This observable encapsulates both the electromagnetic and muonic components of the shower, which are reflected in the measured signals of the detectors. Although the electromagnetic component is independent of the composition of the primary, the muonic component exhibits a dependence, as demonstrated in Fig. 9.7. Consequently, the reconstructed energy presents a composition bias that is overestimated for heavier primaries [96]. This bias directly impacts our studies concerning composition separation, requiring the sorting of muonic observables by energy before comparisons can be undertaken.

For this reason, the results concerning the reconstructed zenith-independent observables S_{38}^{μ} and f_{38}^{μ} are depicted in Fig. 9.10 binned in energy separately for reconstructed and Monte Carlo values. Although a fair distinction between proton and iron primaries is observed for S_{38}^{μ} when utilising the Monte Carlo energy, this difference is almost completely nullified by the composition bias of the reconstructed energy. The same effect is observed for its relative variation, which is systematically reduced by approximately 6% in the case of proton primaries.

Conversely, the reconstructed f_{38}^{μ} is far less influenced by the composition bias introduced by the energy estimation, yielding strikingly similar outcomes for both reconstructed and Monte Carlo values. This absence of a composition bias in the case of f_{38}^{μ} can be attributed to the computation of the ratio between the muonic and total signals. This ratio effectively cancels out the mass dependence of both these quantities, resulting in an observable that is nearly unaffected by the composition bias of the energy estimation. Analysing the mean value of f_{38}^{μ} for both proton and iron, we note an increasing disparity between the two primary types as energy grows. However, below an energy of $10^{18.8}$ eV, the separation between the primaries is unclear due to low signals at the reference distance of 1000 m. Notably, we verified that utilising a distance of 800 m leads to a more pronounced separation in the lower energy range. As for the standard deviation of f_{38}^{μ} , we also observe a modest influence of the composition bias in the energy estimation, and a discernible discrepancy between proton and iron is present.

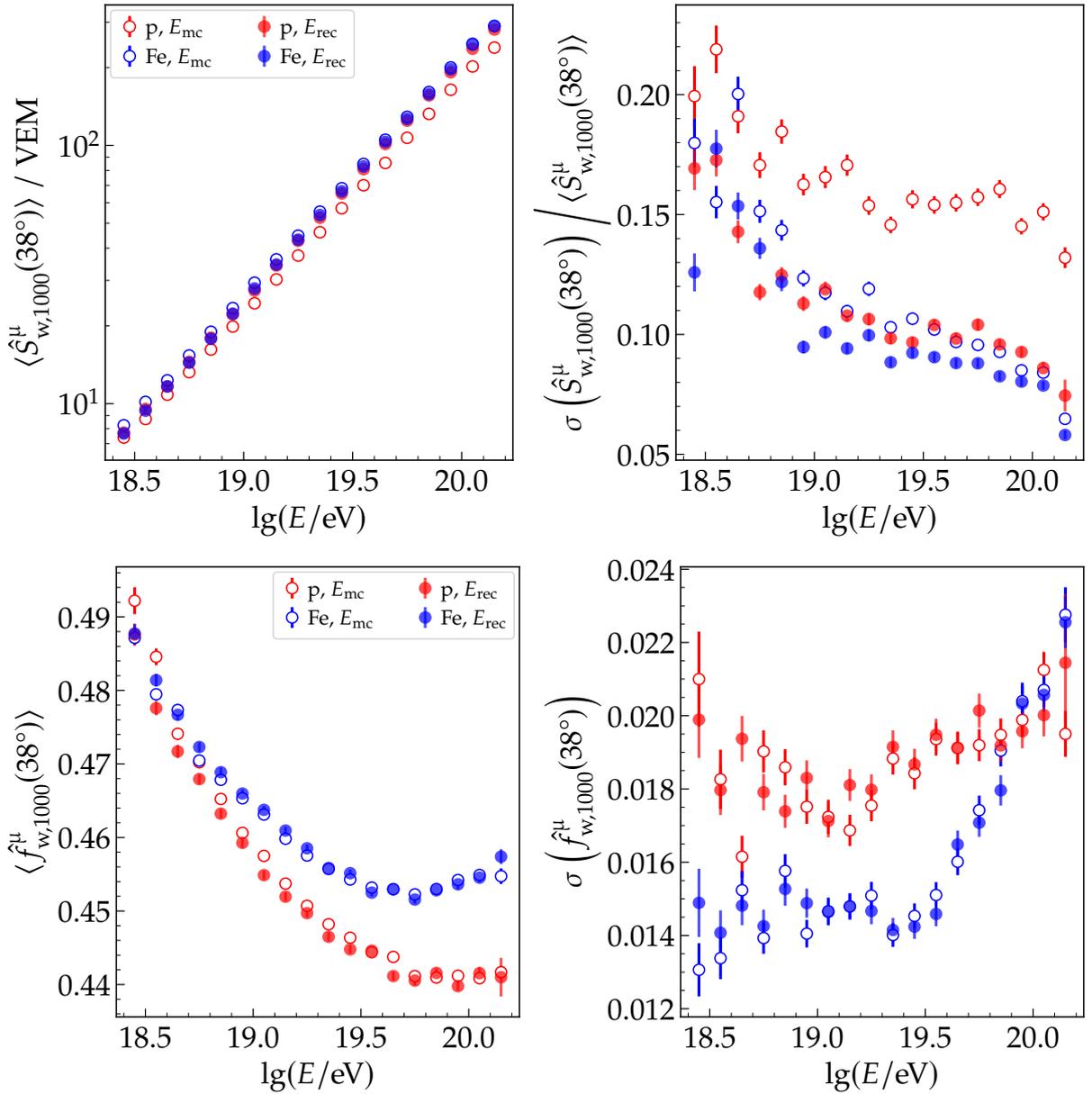


Figure 9.10: Top panels: Mean value (left) and relative variation (right) of S_{38}^{μ} as a function of energy. Bottom panels: Mean value (left) and standard deviation (right) of f_{38}^{μ} as a function of energy. The results are shown for proton (red) and iron (blue) primaries, considering both the reconstructed (closed markers) and Monte Carlo (open markers) values of their energies. This illustrates the contrasting impacts of the composition bias in energy estimation on S_{38}^{μ} and f_{38}^{μ} . Whilst the former is heavily impacted, the latter remains mostly unaffected.

In Fig. 9.11, we present the computed merit factors for S_{38}^{μ} and f_{38}^{μ} , considering reference distances of 800 m, 1000 m, and 1200 m. For S_{38}^{μ} , the drastic impact of the energy-related composition bias significantly diminishes the merit factor from around one to less than 0.25 at energies of approximately $10^{19.5}$ eV. Regarding the Monte Carlo energy, we consistently observe increasing merit factors when estimating S_{38}^{μ} using refe-

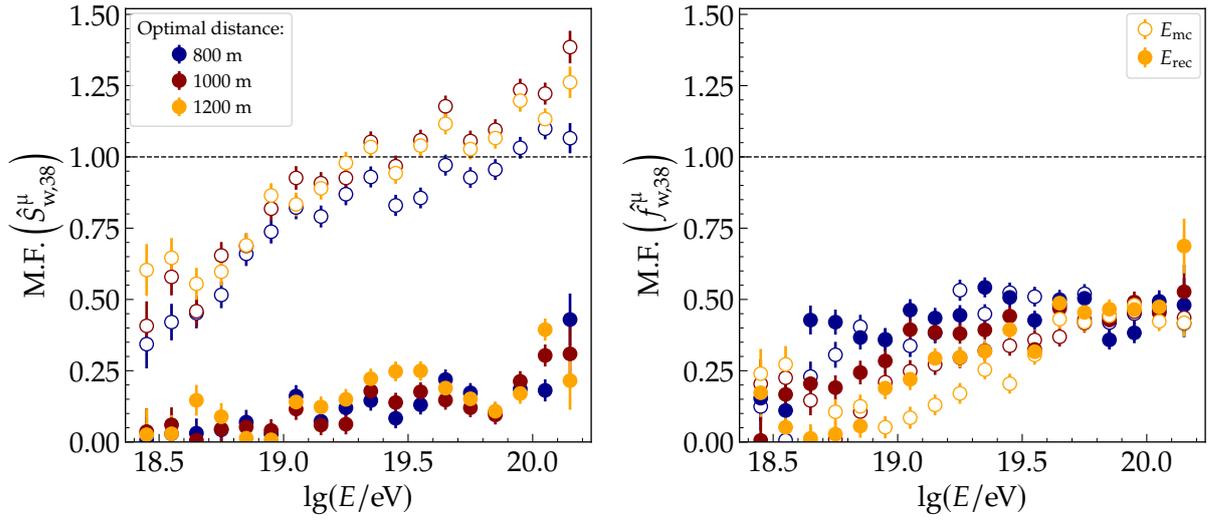


Figure 9.11: Merit factors for S_{38}^{μ} (left) and f_{38}^{μ} (right) as functions of energy. The merit factors are computed for estimations of the observables using reference distances of 800 m (blue), 1000 m (red), and 1200 m (orange). Additionally, the results are presented for both the true Monte Carlo (open markers) and reconstructed (closed markers) values of the energy.

reference distances of 800 m, 1200 m and 1000 m, respectively. Improvement is also verified for increasing energy of the primary, which is associated with the reduced fluctuations in the observable, as demonstrated in the upper right panel of Fig. 9.10. The previously noted lesser impact of the energy-related composition bias on the observable f_{38}^{μ} is reflected by the similar merit factors obtained using reconstructed and Monte Carlo energies. A tendency of better merit factors employing smaller reference distances is evident. However, for energies beyond $10^{19.5}$ eV, the merit factors merely approach a value of 0.5.

Whilst the current results may not support event-by-event identification of primary particles, they do suggest the possibility of using the first and second moments of f_{38}^{μ} for determining composition trends within the highest energy range of the spectrum. As we extend the dynamic range through the addition of the small photomultiplier tubes, the investigation of smaller reference distances, such as 600 m, becomes feasible and offers the potential for improved primary particle separation. We also emphasise the potential of employing the matrix formalism to address or even eliminate the composition bias present in the energy estimations. Given that this bias primarily arises from the muonic component of $S(1000)$, employing the formalism to derive its electromagnetic component could yield an improved calibration procedure

for energy reconstruction when using the upgraded surface detector. This approach has the potential to mitigate the bias and thus improve the accuracy of primary particle identification, particularly concerning the utilisation of S_{1000}^{μ} and S_{38}^{μ} .

CHAPTER 10

Summary and conclusions

The Pierre Auger Observatory initiated its operations in 2004, with the primary goal of investigating cosmic rays with energies exceeding 1 EeV, also referred to as ultra-high energy cosmic rays (**UHECRs**). After more than a decade of data collection, the Auger Collaboration has achieved many breakthroughs in the field of **UHECRs**. Particularly, a suppression of the flux of these particles beyond 40 EeV was measured with unprecedented accuracy. Additionally, an evident tendency of progressively heavier composition above 3 EeV was observed.

These findings are tightly related to the astrophysical sources and propagation dynamics pertaining to these cosmic rays. However, the available data is not sufficient to discriminate between different models aiming at a final description of the nature of **UHECRs**. To arrive at a definitive model, it became clear that more data regarding the composition of **UHECRs**, especially in the suppression region, is imperative.

To gather a larger volume of data sensitivity to the composition of the primary cosmic rays, the Pierre Auger Collaboration designed and initiated a significant upgrade of its Observatory. The main element of the upgrade, dubbed AugerPrime, is the installation of scintillator plates (**SSD**) on top of each water-Cherenkov detector (**WCD**) of the surface array of the Observatory. An additional photomultiplier with a small cross-section will be added in the **WCDs** to extend their dynamic range. New and enhanced electronics will replace the former boards, giving support to the new

detectors with improved precision.

The different responses of the **SSD** and **WCD** to the electromagnetic and muonic components of extensive air showers (**EASs**) initiated by the primary cosmic rays will be exploited to disentangle these components. This separation allows for the reconstruction of observables directly related to the composition of **UHECRs**.

The deployment of the upgraded detectors in the field commenced in 2021. At the end of 2023, the upgrade nears its completion. The new phase of data-taking is termed Phase II and is scheduled to begin in 2024.

We dedicate this final chapter to highlighting the key findings of our research and their significant implications and contributions for the Pierre Auger Collaboration concerning the AugerPrime upgrade and its crucial role in the field of **UHECRs**. To enhance clarity and organisation, we divide this chapter into three sections.

Section 10.1 summarises our investigations and analyses of the first data obtained from the AugerPrime stations during their deployment phase. The primary focus of these studies concerns the critical validation of the performance of the upgraded detectors and their data, especially with respect to calibration and signal processing. We ultimately investigated event reconstructions derived from the AugerPrime stations to ensure the stability of the array and compatibility with the data collected in Phase I of the Observatory operation.

In Section 10.2, we emphasise the essential outcomes obtained from our development of the matrix formalism, a method that utilises the signals of the **SSD** and **WCD** to extract their corresponding electromagnetic and muonic components. This separation is the key aspect behind the upgrade design of the Observatory. This investigation was conducted with Monte Carlo simulations of **EASs** and their subsequent detection with the AugerPrime configuration. We utilised these simulations to understand the relevant parameters when predicting the signal components as well as assess the performance of the method.

Finally, in Section 10.3, we discuss the main results regarding the application of the matrix formalism at an event level, enabling the reconstruction of observables that quantify the muonic content of extensive air showers. Subsequently, we summarise our findings concerning their potential to discriminate among different primary species, which is the crucial measurement envisaged with the upgrade of the Auger Obser-

vatory, facilitating future insights into the origin of UHECRs and the astrophysical scenarios capable of producing these extreme particles.

10.1 ON THE VALIDATION OF AUGERPRIME DATA

To promote the study of data collected by the upgraded stations, aimed at validating and assessing its quality for event reconstruction, and to ensure a seamless transition between Phases I and II of the operation of the Observatory, we performed substantial modifications to the official software for data analyses of the Auger Collaboration: the Offline framework. Our initial objective was to implement the new calibration procedure for stations equipped with the new electronics, called Upgraded Unified Board (UUB).

Firstly, we replaced the previous module within the framework for calibration and signal processing of stations with four new modules designed to perform the same tasks but in a more organised and flexible manner, which allows for future updates and enhancements. Thereupon, we introduced a new module encapsulating a statistics-based algorithm for fitting calibration histograms from both water-Cherenkov detectors and scintillator surface detectors found in stations with either UUB or the former Unified Board (UB) electronics.

We emphasise that our significant modifications, alongside minor adjustments, have rendered the Offline framework fully compatible with processing and event reconstruction of Phase-II data. Thanks to the flexibility of the new calibration algorithm, the existing standard application for event reconstruction with the Surface Detector (SD) can handle data collected by both UB and UUB-equipped stations.

The impact of our contributions to the Offline framework extends beyond the scope of our research analysis. All future physical results obtained by the Auger Collaboration, reliant on event reconstruction of SD data, will derive from the new modules we have implemented in Offline.

With the Offline framework capable of conducting calibration for AugerPrime stations, we achieved a remarkable success rate in fitting calibration histograms using the new algorithm. Specifically, for WCD and SSD stations equipped with UUB electronics, we attained a rate of successful fits of 99.7% for the former and 99.8% for the

latter. These results demonstrate the high efficiency and reliability of the algorithm. Additionally, our validation revealed that the variation in **VEM (MIP)** charge for 68.3% of the stations falls within 6.7% (7.2%) for **WCD (SSD)**, confirming the robustness and consistency of both the upgraded stations and their calibration procedure.

After establishing and assessing the calibration machinery, we turned our attention to examining the compatibility of calibrated signals from **WCDs** between **UUB** and **UB** electronics. This investigation served two main purposes: to further validate the **VEM**-charge calibration of the new electronics and to ensure compatibility between Phase-I and Phase-II data at a station level.

Utilising data from doublets consisting of one station equipped with **UB** and another with **UUB** electronics within a hexagonal arrangement, we compared the signals from both electronics, correcting for the effects of the separation distance between the two stations. Initially, we confirmed that the signals from both electronics were compatible within the bounds of their uncertainties. However, we conducted a more thorough analysis by performing a maximum-likelihood fit of the signals from each doublet.

Our findings indicated that the non-linearity between the signals of **UB** and **UUB** electronics is below 3%. Additionally, we observed a minor bias of 2.7% in the high-gain channel between **UB** and **UUB** signals, leading us to discover that the **VEM** charge estimation for **UUB** electronics lacks correction for the baseline of the calibration traces. Further investigation and correction of this issue are warranted. In the low-gain channel, the bias was found to be 0.4%. The discrepancy between the two gains is consistent with the small residual biases between gain channels in the individual electronics.

To comprehensively assess signal compatibility, we analysed the timing of signals within **WCDs**, given its critical role in the geometrical reconstruction of events. Using data from a triplet of stations equipped with **UUB** electronics, we verified that the differences in start times, corrected for the time of propagation of the shower front between the stations, were compatible with the resolution of the 120 MHz sampling rate. Consequently, no significant timing differences were observed among stations with **UUB** electronics.

During the transition period to full AugerPrime deployment in the field,

some recorded events encompass both **UB** and **UUB** stations in the reconstruction process. To investigate potential timing differences between the two electronics, we utilised data from the **UB/UUB** doublets. Our analysis revealed that the start time offset between the two electronics remained consistent across all doublets, considering the resolution of the lower sampling frequency of the **UB** (40 MHz). The computed mean time offset between the two types of electronics was approximately 53 ns, a factor that should be considered in event reconstructions from the transition phase involving stations with different electronics.

Having established the compatibility of calibrated signals and their timing, we have addressed the utilisation of these signals in event reconstructions, which ultimately produce observables directly relevant to physics studies. Given the substantial changes in the experimental configuration of the Observatory brought about by the upgrade, we conducted a meticulous investigation of the compatibility of event-level observables derived from data collected by **UB** and **UUB** stations. These analyses are critical to confirm the absence of systematic differences in event reconstructions, fundamentally ensuring consistency in the physical analyses conducted during the two operational phases of the Observatory.

Using the hexagon of **UB/UUB** doublet stations, we independently reconstructed the same events with both electronics, facilitating a direct comparison of event-level observables. We demonstrated that the distributions of $S(1000)$ and energy are remarkably similar for both electronics, indicating preliminarily their compatibility. However, most of the analysed dataset consists of low-energy events below $10^{18.2}$ eV and events above full trigger efficiency were limited at the time of the study.

Despite these limitations, we verified that the **UUB** provides unbiased estimates of $S(1000)$ and energy, as the relative differences between the two electronics remain generally below 5%. Nonetheless, some small biases were observed at low energies, below $10^{18.15}$ eV, which can be attributed to differences in triggers between the two electronics. The measured discrepancies in $S(1000)$ were found to be consistent with the expected uncertainties in their determination.

Additionally, we conducted an extensive validation of the geometric reconstruction of showers. The analysis of the shower core position revealed an improved resolution between **UUB** and **UB** for larger energies, reaching 106 m above full trigger

efficiency of the array. Moreover, the **UUB** provides an unbiased determination of the coordinates of the shower core position, confirming its compatibility with **UB**.

Regarding the reconstruction of arrival direction, the **UUB** demonstrated unbiased results for zenith and azimuth angles compared to **UB** reconstructions. However, a larger standard deviation was observed for azimuth angles, mainly attributed to nearly vertical showers, where the azimuth angle becomes degenerate. The investigation of discrepancies in the reconstruction of the shower axis, quantified by the angle η , showed an improvement in angular resolution with increasing energy, reaching 1.1° above 3 EeV. We also observed an enhancement in resolution for more inclined showers, which can be attributed to the increased multiplicity of these events.

Lastly, by utilising the resolution results of **UB** and **UB/UUB** doublets, we estimated the angular resolution achieved by reconstructions with **UUB** stations. Whilst the utilised dataset limits a definitive conclusion, we found that the resolutions of the two electronics are statistically compatible, with a potential tendency for better resolution with **UUB**. Once a larger dataset is available, more energetic cuts can be applied, facilitating the precise quantification of the resolution of the new electronics and the comparison to its former counterpart.

In summary, the analyses presented in this part of our research not only ensured the correct calibration of AugerPrime stations but also validated the use of their calibrated signals for the reconstruction of event-level observables, including shower size, energy, and shower geometry. This validation effort crucially contributes to the compatibility of Phase-I and Phase-II data of the Observatory, ensuring unbiased future physical results for a deeper understanding of ultra-high-energy cosmic rays.

10.2 ON THE APPLICATION OF AUGERPRIME DETECTORS FOR SEPARATION OF SIGNAL COMPONENTS

In the second part of our research, we exploited Monte Carlo simulations of extensive air showers and their subsequent detection and reconstruction using the AugerPrime surface detector to scrutinise a method known as the matrix formalism. This method aims to disentangle the electromagnetic and muonic components of signals in the

upgraded stations. The separation of these components is essential for achieving the primary objective of the upgrade of the Observatory, which is enhancing the sensitivity to the composition of primary cosmic rays within the most energetic range of the spectrum.

To commence this study, we formulated expressions for the muonic and electromagnetic signals in both the **WCD** and **SSD**, written in terms of their respective total signals and two reconstruction parameters. These parameters, in turn, could be expressed as a function of the muonic fractions within the detectors and the ratio of their total signals. This framework allowed us to explore the behaviour of these parameters through Monte Carlo data, revealing a stronger dependence on the distance of the station to the shower core and the zenith angle of the primary, with a comparatively weaker dependence on primary energy. These dependencies naturally extend to the reconstruction parameters, underscoring the need for a parameterisation with respect to shower properties.

Given that the derivation of the formalism assumes approximately constant particle fluxes within the detectors, we demonstrated the necessity of implementing a signal cut to mitigate sampling variations arising from the distinct surface areas of the **WCD** and **SSD**. Applying a cut of 30 **MIP** in the signal of the **SSD** significantly enhances the applicability of the formalism whilst maintaining its accuracy.

Although a parameterisation of the reconstruction coefficients proves essential, our investigations revealed that events with **SSD** signals exceeding approximately 100 **MIP** exhibit small variations in the reconstruction parameters. This suggested the feasibility of employing constant values in such cases. However, even with this substantial signal cut, we found that the reconstruction accuracy significantly improves with the implementation of a parameterisation.

In the process of parameterising the reconstruction coefficients with respect to shower properties and applying a signal cut of 30 **MIP**, we extended the applicability of the formalism to a broader spectrum of events. Across the entire range of distances from the shower core, we observed that the prediction of the muonic component maintained a bias of less than 6% for showers with zenith angles up to 60° and energies exceeding 1 EeV. The resolution exhibited variation, with higher accuracy closer to the shower core, particularly for less inclined and more energetic showers. On average,

it remained below 5% at approximately 500 m from the shower axis, increasing to a maximum of 17% at 1500 m. Additionally, we noted a composition bias, with around a 5% difference between proton and iron primaries in the **WCD** and 3% in the **SSD**.

To further explore the application of the formalism, we considered obtaining the muonic and electromagnetic components of the traces of the detectors. Due to the distinct response times of the **WCD** and **SSD**, a convolution was necessary on the traces of the latter to align their time bins with those of the **WCD**. With the prepared traces, we executed the reconstruction for each bin, utilising the parameterised coefficients of the formalism. The visual assessment indicated satisfactory trace component reconstruction for stations with substantial signals. Conversely, for smaller signals, the trace reconstruction displayed reduced accuracy. Additional investigations are required to quantify the precision when reconstructing trace components and delineate scenarios where the formalism can be applied effectively to the AugerPrime traces.

Our study and advancements in the method of matrix formalism represent a significant contribution, offering an important tool for extracting signal components from AugerPrime stations. Its successful application to real data in Phase II of the Observatory enables the inference of event-level observables intrinsically linked to the cosmic ray composition. Moreover, it presents the possibility of enhancing the accuracy of fundamental reconstructed quantities, such as the primary energy.

10.3 ON THE RECONSTRUCTION OF MUONIC OBSERVABLES AND THEIR APPLICATION FOR MASS SEPARATION

In the concluding phase of our research, we utilised Monte Carlo simulations to explore the reconstruction of observables correlated to the muonic content of extensive air showers. These reconstructions are derived from the muonic signals obtained through the application of the matrix formalism. We then scrutinised the sensitivity of these observables to the composition of the primary cosmic rays, which is the fundamental information aimed by the upgrade of the Auger Observatory.

Initially, we followed a similar approach to the shower size estimator $S(1000)$ by defining the muonic shower size as the muonic signal that a station at a reference

distance from the shower core would detect. This estimation was accomplished by fitting the lateral distribution of muonic signals to a modified Nishimura-Kamata-Greisen function, allowing us to determine the muonic signal at the specified reference distance.

Apart from characterising the muonic shower size, which represents an absolute quantity, we also examined the muonic fraction of the shower. This fraction is similarly defined as the ratio of the muonic signal to the total signal detected by a station at a reference distance. To quantify this observable, we employed the concept of a lateral fraction function, which describes the radial behaviour of the muonic fraction for stations participating in an event. This function was modelled using a power law, enabling us to determine the muonic fraction at the reference distance.

To validate the event-level accuracy of the application of the matrix formalism, we compared the reconstructed muonic shower size and fraction with those obtained from the true Monte Carlo data. Our findings indicate that, at a reference distance of 1000 m, the bias for the muonic shower size remains within 5%, whilst the bias for the muonic fraction is nearly negligible. Nonetheless, a composition bias of around 10% (5%) exists for $S_w^\mu(1000)$ ($f_w^\mu(1000)$). To neutralise the dependence on the zenith angle, we introduced a parameterisation with respect to the reference inclination of 38° , yielding the zenith-independent observables S_{38}^μ and f_{38}^μ .

Proceeding to our investigation of mass separation, we initially observed that the muonic signals at the station level hold the potential for distinguishing between different primary species. The merit factors for these signals display values above one at distances shorter than 1200 m for specific combinations of primary energy and inclination angles. However, the composition bias introduced by the reconstruction of muonic signals degrades the composition resolution, particularly evident in the station-level muonic fractions, which exhibit merit factors below 0.5.

This qualitative description also applies to the utilisation of the muonic shower size and fraction as event-level observables, although these show slightly improved merit factors. It is worth stressing that our previous analyses did not account for the composition bias present in the energy estimation, which was considered in the analysis of the zenith-independent observables S_{38}^μ and f_{38}^μ .

In the case of S_{38}^μ , the aforementioned bias in energy estimation significantly

impacts its discriminatory power, reducing the merit factor from one to below 0.25. On the other hand, we have verified that the observable f_{38}^{μ} is minimally affected by the composition bias in energy determination. Consequently, its first and second moments offer the potential to identify composition trends in the highest energy range, even though its merit factor of approximately 0.45 above 10^{19} eV discourages event-by-event identification of primary particles.

It is important to highlight that this study can be enhanced and further refined. For instance, the fitting process applied to the lateral distribution of muonic signals and fractions could be improved by employing a maximum likelihood approach that takes into account trigger probabilities and saturation, similar to what is currently done for the determination of $S(1000)$. Furthermore, the free parameters of these functions could be parameterised for scenarios involving events with a limited number of participating stations or when these stations are arranged in a compact radial configuration.

Given the observed tendency that utilising smaller reference distances improves the potential of mass discrimination with f_{38}^{μ} , the extension of the dynamic range through the installation of the small PMTs will allow us to explore the capabilities of this observable closer to the shower core.

Additionally, we noted that reducing the composition bias in the energy estimation would significantly enhance the potential for composition discrimination with S_{1000}^{μ} and S_{38}^{μ} . Since this bias is primarily correlated to the muonic component of the shower size, the utilisation of its electromagnetic counterpart, mostly free from composition dependency, for calibrating the energy estimation with the surface detectors could offer a promising approach to mitigating the current composition bias. Again, the application of the matrix formalism represents a robust tool to pursue this goal.

In summary, we have demonstrated that the observables S_{1000}^{μ} , f_{1000}^{μ} , as well as their respective zenith-independent counterparts, S_{38}^{μ} and f_{38}^{μ} , characterise the muonic content of extensive air showers detected by the AugerPrime surface array. Importantly, we have confirmed that these observables are capable of distinguishing between light and heavy cosmic ray species. Whilst further refinements are necessary to fully realise the potential of these observables, our research efforts have contributed significantly to the primary objective of upgrading the Pierre Auger Observatory. This

objective is aimed at enhancing the sensitivity of the experimental setup to the composition of primary particles, ultimately advancing our understanding of ultra-high energy cosmic rays and the astrophysical processes responsible for generating these elusive yet intriguing particles.

Bibliography

The references containing the identifier GAP-year-number refer to internal technical notes of the Pierre Auger Collaboration. These notes can be obtained by contacting the Publication Committee of the collaboration (auger-pc@lists.kit.edu).

1. Gaisser, T. K., Engel, R. & Resconi, E. *Cosmic Rays and Particle Physics* 2nd ed. (Cambridge University Press, 2016).
2. Alves Batista, R. *et al.* Open Questions in Cosmic-Ray Research at Ultrahigh Energies. *Frontiers in Astronomy and Space Sciences* **6** (2019).
3. Workman, R. L. *et al.* Review of Particle Physics. *PTEP* **2022**, 083C01 (2022).
4. Fenu, F. *The cosmic ray energy spectrum measured using the Pierre Auger Observatory in Proceedings of 35th International Cosmic Ray Conference, Busan* (2017).
5. Ahn, H. S. *et al.* ATIC experiment: Preliminary results from the flight in 2002 in *28th International Cosmic Ray Conference* (July 2003), 1853–1856.
6. Aguilar, M. *et al.* The Alpha Magnetic Spectrometer (AMS) on the international space station: Part II — Results from the first seven years. *Phys. Rept.* **894**, 1–116 (2021).
7. Picozza, P. *et al.* PAMELA A payload for antimatter matter exploration and light-nuclei astrophysics. *Astroparticle Physics* **27**, 296–315. arXiv: [astro-ph/0608697](https://arxiv.org/abs/astro-ph/0608697) [[astro-ph](https://arxiv.org/abs/astro-ph)] (Apr. 2007).
8. Linsley, J., Scarsi, L. & Rossi, B. *Suppl. J. Phys. Soc. Japan* **17**, 91 (1962).
9. Baltrusaitis, R. *et al.* The Utah Fly’s Eye detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **240**, 410–428 (1985).

10. Haungs, A. *et al.* The Air-Shower Experiment KASCADE-Grande. *Nuclear Physics B - Proceedings Supplements* **196**. Proceedings of the XV International Symposium on Very High Energy Cosmic Ray Interactions (ISVHECRI 2008), 80–85 (2009).
11. Kawai, H. *et al.* Telescope Array Experiment. *Nuclear Physics B - Proceedings Supplements* **175-176**. Proceedings of the XIV International Symposium on Very High Energy Cosmic Ray Interactions, 221–226 (2008).
12. Aab, A. *et al.* The Pierre Auger Cosmic Ray Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **798**, 172–213 (2015).
13. Allekotte, I. *et al.* The surface detector system of the Pierre Auger Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **586**, 409–420 (2008).
14. Abraham, J. *et al.* The fluorescence detector of the Pierre Auger Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **620**, 227–251 (2010).
15. Abraham, J. *et al.* Observation of the Suppression of the Flux of Cosmic Rays above 4×10^{19} eV. *Phys. Rev. Lett.* **101**, 061101 (2008).
16. Greisen, K. End to the cosmic ray spectrum? *Phys. Rev. Lett.* **16**, 748–750 (1966).
17. Zatsepin, G. T. & Kuzmin, V. A. Upper limit of the spectrum of cosmic rays. *JETP Lett.* **4**, 78–80 (1966).
18. Aab, A. *et al.* Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above $10^{17.8}$ eV. *Phys. Rev. D* **90**, 122005 (2014).
19. Heck, D., Knapp, J., Capdevielle, J. N., Schatz, G. & Thouw, T. *CORSIKA: A Monte Carlo code to simulate extensive air showers* FZKA-6019 (1998).
20. Pierog, T., Karpenko, I., Katzy, J. M., Yatsenko, E. & Werner, K. EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider. *Phys. Rev. C* **92**, 034906 (2015).
21. Ostapchenko, S. Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model. *Phys. Rev. D* **83**, 014018 (2011).

22. Aab, A. *et al.* Depth of maximum of air-shower profiles at the Pierre Auger Observatory. II. Composition implications. *Phys. Rev. D* **90**, 122006 (2014).
23. Aab, A. *et al.* Combined fit of spectrum and composition data as measured by the Pierre Auger Observatory. *Journal of Cosmology and Astroparticle Physics* **2017**, 038 (2017).
24. Aab, A. and others. The Pierre Auger Observatory Upgrade - Preliminary Design Report. *arXiv:1604.03637* (2016).
25. Letessier-Selvon, A., Billoir, P., Blanco, M., Mariş, I. C. & Settimo, M. Layered water Cherenkov detector for the study of ultra high energy cosmic rays. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **767**, 41–49 (2014).
26. Swann, W. F. G. The History of Cosmic Rays. *American Journal of Physics* **29**, 811–816 (1961).
27. Ginzburg, V. L. Cosmic ray astrophysics (history and general review). *Physics-Uspeski* **39**, 155–168 (1996).
28. Hess, V. F. Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten. *Phys. Z.* **13**, 1084–1091 (1912).
29. Chiba, N. *et al.* Akeno Giant Air Shower Array (AGASA) covering 100 km² area. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **311**, 338–349 (1992).
30. Auger, P., Ehrenfest, P., Maze, R., Daudin, J. & Fréon, R. A. Extensive Cosmic-Ray Showers. *Rev. Mod. Phys.* **11**, 288–291 (1939).
31. Mollerach, S. & Roulet, E. Progress in high-energy cosmic ray physics. *Progress in Particle and Nuclear Physics* **98**, 85–118 (2018).
32. Matthews, J. A Heitler model of extensive air showers. *Astroparticle Physics* **22**, 387–397 (2005).
33. Alvarez-Muñiz, J., Engel, R., Gaisser, T. K., Ortiz, J. A. & Stanev, T. Hybrid simulations of extensive air showers. *Phys. Rev. D* **66**, 033011 (2002).
34. Heitler, W. *The Quantum Theory of Radiation* third ed. p. 386 (Section 38) (Oxford University Press, 1954).

35. Glasmacher, M. *et al.* The cosmic ray composition between 10^{14} and 10^{16} eV. *Astroparticle Physics* **12**, 1–17 (1999).
36. Seo, E. S. *et al.* Cosmic Ray Energetics And Mass for the International Space Station (ISS-CREAM). *PoS ICRC2019*, 137 (2020).
37. Ave, M. *et al.* The TRACER instrument: A balloon-borne cosmic-ray detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **654**, 140–156 (2011).
38. <http://www.ams02.org>. Accessed 27-08-2018.
39. Gaisser, T. Cosmic-Ray Showers Reveal Muon Mystery. *Physics* **9** (2016).
40. Lawrence, M. A., Reid, R. J. O. & Watson, A. A. The cosmic ray energy spectrum above 4×10^{17} eV as measured by the Haverah Park array. *Journal of Physics G: Nuclear and Particle Physics* **17**, 733 (1991).
41. Abu-Zayyad, T. *et al.* The prototype high-resolution Fly's Eye cosmic ray detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **450**, 253–269 (2000).
42. Tanabashi, M. *et al.* Review of Particle Physics. *Phys. Rev. D* **98**, 030001 (2018).
43. Fermi, E. On the Origin of the Cosmic Radiation. *Phys. Rev.* **75**, 1169–1174 (1949).
44. Hillas, A. M. The Origin of Ultra-High-Energy Cosmic Rays. *Annual Review of Astronomy and Astrophysics* **22**, 425–444 (1984).
45. Peters, B. Primary cosmic radiation and extensive air showers. *Il Nuovo Cimento* **22**, 800–819 (1961).
46. Unger, M., Farrar, G. R. & Anchordoqui, L. A. Origin of the ankle in the ultrahigh energy cosmic ray spectrum, and of the extragalactic protons below it. *Phys. Rev. D* **92**, 123001 (2015).
47. Veberič, D. https://web.ikp.kit.edu/darko/auger/auger-array/auger_array-pdf/auger_array-ad.pdf. Accessed: 2021.
48. *Pierre Auger Collaboration Website* <https://auger.org/>. Accessed: August/2023.
49. *Photomultiplier XP1805* <http://hzcphotonics.com>. Accessed May-2018.

50. Aglietta, M. *et al.* Response of the Pierre Auger Observatory Water Cherenkov Detectors to Muons in *Proceedings of 29th International Cosmic Ray Conference, Pune 7* (2005), 83–86.
51. Bertou, X. *et al.* Calibration of the surface array of the Pierre Auger Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **568**, 839–846 (2006).
52. Etchegoyen, A. *et al.* Muon-track studies in a water Cherenkov detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **545**, 602–612 (2005).
53. Abraham, J. *et al.* Trigger and aperture of the surface detector array of the Pierre Auger Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **613**, 29–39 (2010).
54. Aab, A. *et al.* Reconstruction of events recorded with the surface detector of the Pierre Auger Observatory. *Journal of Instrumentation* **15**, P10021 (Oct. 2020).
55. Kamata, K. & Nishimura, J. The Lateral and the Angular Structure Functions of Electron Showers. *Progress of Theoretical Physics Supplement* **6**, 93–155 (1958).
56. Newton, D., Knapp, J. & Watson, A. The optimum distance at which to determine the size of a giant air shower. *Astroparticle Physics* **26**, 414–419 (2007).
57. Gaisser, T. K. & Hillas, A. M. Reliability of the method of constant intensity cuts for reconstructing the average development of vertical showers in *15th International Cosmic Ray Conference, Plovdiv, Bulgaria* **8** (1977), 353–357.
58. Hersil, J., Escobar, I., Scott, D., Clark, G. & Olbert, S. Observations of Extensive Air Showers near the Maximum of Their Longitudinal Development. *Physical Review Letters* **6**, 22–23 (1961).
59. Schulz, A. The measurement of the energy spectrum of cosmic rays above 3×10^{17} eV with the Pierre Auger Observatory in *Proceedings of 33rd International Cosmic Ray Conference, Rio de Janeiro* (2013).
60. Argirò, S. *et al.* The Offline software framework of the Pierre Auger Observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **580**, 1485–1496 (2007).

61. Brun, R. & Rademakers, F. ROOT — An object oriented data analysis framework. *Nucl. Instrum. Methods Phys. Res., Sect. A* **389**. New Computing Techniques in Physics Research V, 81–86 (1997).
62. Aab, A. *et al.* Observation of a large-scale anisotropy in the arrival directions of cosmic rays above 8×10^{18} eV. *Science* **357**, 1266–1270 (2017).
63. Abraham, J. *et al.* Correlation of the Highest-Energy Cosmic Rays with Nearby Extragalactic Objects. *Science* **318**, 938–943 (2007).
64. Abreu, P. *et al.* Update on the correlation of the highest energy cosmic rays with nearby extragalactic matter. *Astroparticle Physics* **34**, 314–326 (2010).
65. Aab, A. *et al.* Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events. *Phys. Rev. D* **91**, 032003 (2015).
66. Aab, A. *et al.* Measurement of the Fluctuations in the Number of Muons in Extensive Air Showers with the Pierre Auger Observatory. *Phys. Rev. Lett.* **126**, 152002 (2021).
67. Kalashev, O. E., Rubtsov, G. I. & Troitsky, S. V. Sensitivity of cosmic-ray experiments to ultrahigh-energy photons: Reconstruction of the spectrum and limits on the superheavy dark matter. *Phys. Rev. D* **80**, 103006 (2009).
68. Colladay, D. & Kostelecký, V. A. Lorentz-violating extension of the standard model. *Phys. Rev. D* **58**, 116002 (1998).
69. Huege, T. *et al.* The Radio Detector of the Pierre Auger Observatory – status and expected performance. *EPJ Web Conf.* **283**, 06002 (2023).
70. Streich, A. *Performance of the Upgraded Surface Detector of the Pierre Auger Observatory* (PhD Thesis) (Karlsruhe Institute of Technology, 2022).
71. Anastasi, G. A. *et al.* *The Calibration of the Small PMT Auger* internal note GAP-2022-018 (2022).
72. Hiller, R. & Roth, M. *An update on the signal accuracy using the infill array Auger* internal note GAP-2012-012 (2012).
73. Coleman, A., Ghia, P. & Mostafá, M. *The Station Triggering Efficiency Using the New Triggers Auger* internal note GAP-2018-017 (2018).

74. *Scipy minimize method* <https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html>. Accessed: April/2023.
75. Schulz, T., Roth, M., Schmidt, D. & Veberič, D. *New Baseline Algorithm for UB Traces* Auger internal note GAP-2022-045 (2022).
76. Schulz, T., Roth, M., Schmidt, D. & Veberič, D. *New Baseline Algorithm for UUB Traces* Auger internal note GAP-2023-007 (2023).
77. Convenga, F., Sato, R., Nitz, D. & Schmidt, D. *Impact of downsampling for VEM calibration of compatibility triggers* Auger internal note GAP-2023-006 (2023).
78. Antoni, T. *et al.* The cosmic-ray experiment KASCADE. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **513**, 490–510 (2003).
79. Werner, K., Liu, F.-M. & Pierog, T. Parton ladder splitting and the rapidity dependence of transverse momentum spectra in deuteron-gold collisions at the BNL Relativistic Heavy Ion Collider. *Phys. Rev. C* **74**, 044902 (2006).
80. Pierog, T. & Werner, K. EPOS Model and Ultra High Energy Cosmic Rays. *Nuclear Physics B - Proceedings Supplements* **196**. Proceedings of the XV International Symposium on Very High Energy Cosmic Ray Interactions (ISVHECRI 2008), 102–105 (2009).
81. Ferrari, A., Sala, P. R., Fasso, A. & Ranft, J. FLUKA: A multi-particle transport code. Report number CERN-2005-010, SLAC-R-773, INFN-TC-05-11 (2005).
82. Battistoni, G. *et al.* Overview of the FLUKA code. *Annals Nucl. Energy* **82**, 10–18 (2015).
83. Fasso, A. *et al.* The Physics models of FLUKA: Status and recent developments. *eConf* **C0303241**, MOMT005. arXiv: [hep-ph/0306267](https://arxiv.org/abs/hep-ph/0306267) (2003).
84. Nelson, W. & Namito, Y. The EGS4 code system: solution of gamma-ray and electron transport problems. *SLAC PUB* **5193** (1990).
85. Heck, D. & Pierog, T. *Extensive Air Shower Simulation with CORSIKA: A User's Guide* (2019).
86. Agostinelli, S. *et al.* Geant4—a simulation toolkit. *Nucl. Instrum. Methods Phys. Res., Sect. A* **506**, 250–303 (2003).

87. Allison, J. *et al.* Recent developments in Geant4. *Nucl. Instrum. Methods Phys. Res., Sect. A* **835**, 186–225 (2016).
88. Schmidt, D. AugerPrime implementation in the $\overline{\text{Offline}}$ simulation and reconstruction framework. *PoS ICRC2017* (2017).
89. Hahn, S. T. *Methods for Estimating Mass-Sensitive Observables of Ultra-High Energy Cosmic Rays using Artificial Neural Networks* (PhD Thesis) (Karlsruhe Institute of Technology, 2022).
90. Schmidt, D. *Sensitivity of AugerPrime to the masses of ultra-high-energy cosmic rays* (PhD Thesis) GAP-2019-030 (Karlsruhe Institute of Technology, 2018).
91. Lettesier-Selvon, A. & Billoir, P. *Separating the muonic and electromagnetic components of a shower through the matrix formalism for a two-fold detector* Auger internal note GAP-2015-045 (2015).
92. Martello, D. *Considerations about the method of the matrix inversion* Auger internal note GAP-2019-033 (2019).
93. Pothast, M., Timmermans, C. & de Jong, S. *SSD and WCD signal model* Auger internal note GAP-2021-58 (2021).
94. Stadelmaier, M. *On Air-Shower Universality and the Mass Composition of Ultra-high-energy Cosmic Rays* (PhD Thesis) GAP-2022-009 (Karlsruhe Institute of Technology, 2022).
95. Aab, A. *et al.* Muons in air showers at the Pierre Auger Observatory: Measurement of atmospheric production depth. *Phys. Rev. D* **90**, 012012 (2014).
96. Heizmann, F. *Deep Learning Reconstruction of the Muon Signal Fraction for Mass Composition Studies with AugerPrime* (Master's Thesis) (Karlsruhe Institute of Technology, 2023).

APPENDIX A

Model for the reconstruction parameters

A model is here provided to understand better the underlying physical processes that determine the behaviour of the reconstruction parameters α and β . These parameters are defined in Eq. (8.4) as a function of the matrix coefficients of Eq. (8.1). Since these matrix coefficients relate to the signal components in the detectors through Eq. (8.2), modelling such components as a function of the muonic and electromagnetic fluxes offers an interpretation of the matrix coefficients and, consequently, of the reconstruction parameters α and β .

Fortunately, a model for the signal components in the WCD and SSD is provided in Ref. [93]. For convenience, we write down the expressions obtained therein that are relevant to our study. The muonic and electromagnetic signals in the detectors are given by

$$S_w^\mu = \mathcal{A}_w(\theta') \left[\frac{1}{E_{\text{VEM}}^2} \int_{E_{\text{th}}}^{E_{\text{VEM}}} \frac{d\mathcal{F}_\mu}{dE} E^2 dE + \frac{\mathcal{A}_w(0)}{\mathcal{A}_w(\theta')} \int_{E_{\text{VEM}}}^{\infty} \frac{d\mathcal{F}_\mu}{dE} dE \right] \text{VEM}, \quad (\text{A.1})$$

$$S_w^{\text{em}} = \mathcal{A}_w(\theta') \left[\frac{\langle E^e \rangle}{E_{\text{VEM}}} \mathcal{F}_e + \frac{\langle E^\gamma \rangle}{E_{\text{VEM}}} \mathcal{F}_\gamma \right] \text{VEM}, \quad (\text{A.2})$$

$$S_s^\mu = \mathcal{A}_s(\theta') \sec \theta' \mathcal{F}_\mu \text{MIP}, \quad (\text{A.3})$$

$$S_s^{\text{em}} = \mathcal{A}_s(\theta') \sec \theta' [\mathcal{F}_e + \epsilon_s^\gamma \mathcal{F}_\gamma] \text{MIP}, \quad (\text{A.4})$$

where θ' is the angle of the incoming fluxes of particles with respect to the vertical direction. $\mathcal{A}_{w(s)}(\theta')$ is the perpendicular area of the WCD (SSD) to the particle fluxes,

whereas $\mathcal{A}_w(0)$ is simply the surface area of the top of WCD. The flux of electromagnetic particles \mathcal{F}_{em} is decomposed into the fluxes \mathcal{F}_e of electrons and \mathcal{F}_γ of photons. E_{VEM} is the kinetic energy necessary to produce a signal of 1 VEM in the WCD. It is ~ 400 MeV for muons, electrons and photons [93]. E_{th} is the threshold energy for the production of Cherenkov radiation in water, and its value is ~ 54 MeV.

When photons cross the SSD, some of them are converted into electron-positron pairs which produce a signal. Vertical photons yield a signal of $\epsilon_s^\gamma \approx 0.03$ MIP. The factors $\langle E^e \rangle$ and $\langle E^\gamma \rangle$ in the Eq. (A.2) are defined as

$$\langle E^{e(\gamma)} \rangle = \frac{1}{\mathcal{F}_{e(\gamma)}} \int \frac{d\mathcal{F}_{e(\gamma)}}{dE} E dE, \quad (\text{A.5})$$

where the integral, as in the case of Eq. (A.1), runs over the energy spectrum of the corresponding particle fluxes.

The matrix coefficients of Eq. (8.1) are obtained when comparing the Eqs. (A.1) to (A.4) with Eq. (8.2), which gives the following result

$$a = \left[\mathcal{A}_w(\theta') \frac{\langle (E_1^\mu)^2 \rangle}{E_{VEM}^2} + \mathcal{A}_w(0) f_2^\mu \right] \text{VEM} \quad (\text{A.6})$$

$$b = \mathcal{A}_w(\theta') \left[\frac{\langle E^e \rangle}{E_{VEM}} f_e + \frac{\langle E^\gamma \rangle}{E_{VEM}} f_\gamma \right] \text{VEM} \quad (\text{A.7})$$

$$c = \mathcal{A}_s(\theta') \sec \theta' \text{MIP} \quad (\text{A.8})$$

$$d = \mathcal{A}_s(\theta') \sec \theta' [f_e + \epsilon_s^\gamma f_\gamma] \text{MIP}, \quad (\text{A.9})$$

where

$$\langle (E_1^\mu)^2 \rangle = \frac{1}{\mathcal{F}_\mu} \int_{E_{th}}^{E_{VEM}} \frac{d\mathcal{F}_\mu}{dE} E^2 dE \quad \text{and} \quad f_2^\mu = \frac{1}{\mathcal{F}_\mu} \int_{E_{VEM}}^{\infty} \frac{d\mathcal{F}_\mu}{dE} dE. \quad (\text{A.10})$$

The factors f_e and f_γ are the ratio of electron and photon flux, respectively, to the total electromagnetic flux.

We use the definition of the reconstruction parameters α and β , given in

Eq. (8.4), together with the Eqs. (A.6) to (A.9) to write

$$\alpha = \frac{ad}{ad - bc} = \left(1 - \frac{\langle E^e \rangle + \langle E^\gamma \rangle r^\gamma}{\left[\frac{\langle (E_1^\mu)^2 \rangle}{E_{\text{VEM}}} + E_{\text{VEM}} \frac{\mathcal{A}_w(0)}{\mathcal{A}_w(\theta')} f_2^\mu \right] [1 + \epsilon_s^\gamma r^\gamma]} \right)^{-1}, \quad (\text{A.11})$$

$$\beta = \frac{ab}{ad - bc} = \frac{\text{VEM}}{\text{MIP}} \left(E_{\text{VEM}} \mathcal{A}_s(\theta') \sec \theta' \left[\frac{1 + \epsilon_s^\gamma r^\gamma}{\langle E^e \rangle + \langle E^\gamma \rangle r^\gamma} - \frac{1}{\frac{\langle (E_1^\mu)^2 \rangle}{E_{\text{VEM}}} + E_{\text{VEM}} \frac{\mathcal{A}_w(0)}{\mathcal{A}_w(\theta')} f_2^\mu} \right] \right)^{-1}, \quad (\text{A.12})$$

where $r^\gamma = f_\gamma/f_e$ is simply the ratio between the number of photons and electrons that hit the detectors.

Therefore, the model predicts a dependency of the reconstruction parameters on the incidence angle θ' of the particles, which is correlated with the zenith angle θ of the shower. Besides, these parameters depend on the factors $\langle E^e \rangle$, $\langle E^\gamma \rangle$, r^γ , $\langle (E_1^\mu)^2 \rangle$, and f_2^μ . It has been shown in [93], using Monte-Carlo data, that $\langle (E_1^\mu)^2 \rangle$ and f_2^μ present a dependency on both zenith angle and distance to the shower core. $\langle E^e \rangle$, $\langle E^\gamma \rangle$, and r^γ also depend on the distance to the shower core, however they are fairly constant with zenith angle for showers with inclination smaller than 60° .

The exact dependency of α and β on zenith angle and distance to the shower core is not obvious from Eqs. (A.11) and (A.12). More detailed simulations, with information about the energy distribution of the particles in the detectors would be necessary. However, we see that this result goes in the same direction as that of studying the behaviour of muonic fractions and total signal ratios in Section 8.2, namely the need for a parameterisation of the reconstruction coefficients with shower properties.