dortmund university

Department of Physics

A document submitted in partial fulfillment of the requirements for the degree of
Doctor rerum naturalium

Chasing Prompt Muons: Unfolding the Atmospheric Muon Spectrum to Constrain the Prompt Flux using 12 Years of IceCube Data

Pascal Thomas Gutjahr
born in Dortmund

1995

Astroparticlephysics
Department of Physics
TU Dortmund University

First reviewer:	Prof. Dr. Dr. Wolfgang Rhode
Second reviewer:	Prof. Dr. Kevin Kröninger
Submission date:	February 14, 2026

Abstract

The atmospheric muon flux provides a powerful probe of TeV-to-PeV hadronic interactions in air showers and offers unique access to particle production in the forward kinematic region that is not accessible at collider experiments. While the conventional flux from pion and kaon decays is comparatively well understood, the prompt component, dominated by decays of charmed hadrons, lacks precise experimental constraints at the highest energies. The atmospheric muon flux is also a dominant background for astroparticle analyses, in particular searches for astrophysical neutrino sources. This thesis presents a measurement of the atmospheric muon flux, including its prompt component, using 12.12 years of data from the IceCube Neutrino Observatory. The muon energy spectrum is unfolded to surface level in the range 10 TeV to 15 PeV, extending previous IceCube measurements by roughly an order of magnitude in energy. A key contribution of this work is a machine learning-based reconstruction of the energy of the *leading muon* (the most energetic muon within a bundle), increasing sensitivity to the spectral differences between conventional and prompt components. New CORSIKA simulations were produced to tag muons of prompt or conventional origin, and a dedicated event selection was developed for this analysis. The unfolded flux is compared with MCEq and CORSIKA predictions for several combinations of primary cosmic-ray and hadronic interaction models. Among the models considered here, the MCEq prediction using the GlobalSplineFit primary cosmic-ray model together with SIBYLL 2.3c provides the best agreement with the high-energy unfolded flux. For this prediction, the prompt atmospheric muon flux is observed with a significance of $4.4\,\sigma$. Several models have been tested and prompt and conventional normalizations are determined, with most fits exceeding $5\,\sigma$. The results provide new experimental constraints on forward charm production in air showers and will contribute to ongoing investigations of the muon deficit in air shower experiments.

Kurzfassung

Der atmosphärische Myonenfluss ermöglicht es, die Teilchenproduktion in hochenergetischen hadronischen Wechselwirkungen in Luftschauern im kinematischen Vorwärtsbereich zu untersuchen, der für Kolliderexperimente nicht zugänglich ist. Während der konventionelle Fluss aus Pion- und Kaonzerfällen vergleichsweise gut verstanden ist, fehlt es für die prompte Komponente, die durch Zerfälle von charmhaltigen Hadronen dominiert wird, an präzisen Messungen bei den höchsten Energien. Der atmosphärische Myonenfluss ist zudem der dominierende Untergrund für Analysen in der Astroteilchenphysik, insbesondere für die Suche nach astrophysikalischen Neutrinoquellen. Diese Arbeit präsentiert eine Messung des atmosphärischen Myonenflusses, einschließlich seiner prompten Komponente, unter Verwendung von 12, 12 Jahren an Daten des IceCube Neutrino Observatoriums. Das Myonenenergiespektrum wird an die Oberfläche im Energiebereich von 10 TeV bis 15 PeV entfaltet und erweitert damit frühere IceCube-Myonenflussmessungen um etwa eine Größenordnung in Energie. Ein zentraler Beitrag ist eine neuartige, auf maschinellem Lernen basierende Rekonstruktion zur Bestimmung der Energie des *führenden Myons* (des höchstenergetischen Myons innerhalb eines Bündels), die die Sensitivität auf spektrale Unterschiede zwischen konventioneller und prompter Komponente erhöht. Ergänzend wurden neue CORSIKA-Simulationen zur Zuordnung von Myonen zu promptem bzw. konventionellem Ursprung sowie eine dedizierte Ereignisselektion für diese Analyse entwickelt. Die entfalteten Ergebnisse werden mit MCEq- und CORSIKA-Vorhersagen für mehrere Kombinationen aus Modellen der primären kosmischen Strahlung und der hadronischen Wechselwirkungen verglichen. Unter den in dieser Arbeit betrachteten Modellen zeigt die MCEq-Vorhersage mit dem GlobalSplineFit-Modell in Kombination mit SIBYLL 2.3c die beste Übereinstimmung mit dem hochenergetischen entfalteten Fluss. Für dieses Referenzmodell wird die prompte Komponente mit einer Signifikanz von $4,4\,\sigma$ nachgewiesen. Mehrere Modelle wurden getestet und die Normierungen der prompten und konventionellen Komponenten bestimmt, wobei die meisten Fits $5\,\sigma$ überschreiten. Die Ergebnisse liefern neue experimentelle Einschränkungen für die Charm-Produktion im Vorwärtsbereich bei den höchsten Energien und werden zur Untersuchungen des Myondefizits in Luftschauerexperimenten beitragen.

Contents

1 Introduction **1**

2 The Standard Model: A Selected Overview **3**
2.1 Particle Content . . . 3
2.2 Light Hadron Decays . . . 4
2.3 Charm-Hadron Decays . . . 6

3 Astroparticle Physics **7**
3.1 Multi-Messenger Astronomy . . . 7
3.2 Cosmic Rays and their Mass Composition . . . 9
3.3 Air Showers . . . 13
3.4 Cascade Equations . . . 14
3.5 The Atmospheric Muon Flux: Conventional and Prompt . . . 17
3.5.1 `MCEq` and Hadronic Interaction Models (HIMs) . . . 18
3.5.2 Muon Flux Prediction . . . 18
3.6 The Muon Puzzle . . . 20

4 The IceCube Neutrino Observatory **23**
4.1 Detector Design . . . 23
4.1.1 Digital Optical Modules . . . 23
4.1.2 Arrangement of Digital Optical Modules . . . 24
4.2 Detection Mechanism . . . 25
4.2.1 Neutrino Interactions in Ice . . . 25
4.2.2 Cherenkov Radiation . . . 26
4.2.3 Particle Signatures . . . 26
4.3 IceCube Data Acquisition System . . . 29

5 Simulation **31**
5.1 Air Shower Simulation with CORSIKA 7 . . . 31
5.1.1 Extended History . . . 34
5.1.2 Weight Simulation to Physical Flux . . . 36
5.1.3 Comparison to MCEq Solutions . . . 36
5.2 Lepton Propagation Through the Detector . . . 37
5.3 Ice Systematics – SnowStorm . . . 38
5.3.1 DOM Efficiency . . . 38
5.3.2 Absorption and Scattering . . . 38
5.3.3 Hole Ice Parameters (p0 and p1) . . . 39
5.4 Final Simulation – Processing Framework IceProd . . . 39
5.5 Summary: Simulation . . . 40

6 Reconstruction **41**
6.1 Traditional Reconstructions . . . 41
6.1.1 Energy Reconstruction with TruncatedEnergy . . . 42
6.1.2 Directional Reconstruction with SplineMPE . . . 42

6.2 Machine Learning Reconstructions . . . 43
6.2.1 DNN-reco Framework . . . 46
6.2.2 Uncertainty Estimation . . . 47
6.2.3 Input Features . . . 47
6.2.4 Training . . . 48
6.2.5 Pearson and Spearman Correlation Coefficients . . . 52
6.2.6 Variable and Network Definitions and Notation . . . 53
6.2.7 Energy . . . 54
6.2.8 Direction . . . 58
6.2.9 Track Geometry . . . 61
6.2.10 Traditional vs. Machine Learning Reconstructions . . . 65
6.3 Summary: Reconstruction . . . 68

7 Selection **69**
7.1 Level 3 – Filter . . . 69
7.2 Level 4 – Add Reconstructions and Pre Cut . . . 72
7.3 Level 5 – Data-MC Quality Cuts . . . 72
7.4 Final Level – Leadingness Cut . . . 84
7.5 Summary: Selection . . . 87

8 Unfolding of the atmospheric muon flux **89**
8.1 Unfolding Basics . . . 89
8.1.1 Method . . . 89
8.1.2 Sensitivity correction / Effective Area . . . 91
8.1.3 Python Tool funfolding . . . 94
8.1.4 Visualization of the Unfolding Principle . . . 95
8.2 Unfolding of Small Test Dataset (Burnsample) . . . 97
8.2.1 Ice Systematics . . . 98
8.2.2 Effective Area . . . 101
8.2.3 Regularization . . . 103
8.2.4 Robustness Tests . . . 103
8.2.5 Burnsample Unfolding . . . 103
8.2.6 Data–MC Comparison . . . 110
8.2.7 Impact of Primary and Hadronic Interaction Models . . . 111
8.2.8 Impact of Coincident Primaries . . . 113
8.2.9 Neutrino Background . . . 114
8.3 Full Sample 12 Years of Data Unfolding . . . 115
8.3.1 Combination of MC Datasets . . . 116
8.3.2 Choose Proxy Binning . . . 117
8.3.3 Ice Systematics . . . 119
8.3.4 Effective Area . . . 120
8.3.5 Regularization . . . 123
8.3.6 Robustness Tests . . . 123
8.3.7 Prompt Sensitivity Test . . . 124
8.3.8 12 Years of MC Data Unfolding . . . 126
8.3.9 12 Years of Experimental Data Unfolding . . . 128
8.3.10 Flux Normalization Measurement . . . 129
8.4 Summary: Unfolding . . . 137

9 Conclusion **139**

10 Outlook **141**

A Reconstruction **149**
A.1 Training Data . . . 149
A.2 Direction . . . 152
A.3 Training on Newest Ice Model . . . 152

B Selection **155**
B.1 Level 3 – Filter . . . 155
B.2 Level 4 – Pre Cut . . . 160
B.3 Data–MC: Level 5 . . . 163
B.4 Data–MC: Final Level . . . 171
B.5 Final Level – Potential Leadingness Cuts . . . 179

C Burnsample Unfolding **183**
C.1 Check for Second Minimum in Tau Scan . . . 183
C.2 Re-Weighted Systematics . . . 184

D Impact of Coincident Primaries **187**
D.1 Dataset 20904 . . . 187
D.2 Dataset 22615 . . . 189

E Burnsample Unfolding – Combined MC **191**
E.1 General . . . 191
E.2 MC Unfolding . . . 193
E.3 Data Unfolding . . . 195

F Full Sample Unfolding **199**
F.1 Proxy Binning . . . 200
F.2 Regularization Scan . . . 203
F.3 MC Unfolding . . . 208
F.4 Experimental Data Unfolding . . . 210
F.4.1 Data-MC Pre and Post Fit . . . 216
F.4.2 Unfolded Flux Values . . . 217
F.5 Flux Normalization Fits . . . 223
F.5.1 Results Sorted by Unfolding Primary Model . . . 223
F.5.2 Results Sorted by Primary Model Prediction . . . 227
F.5.3 Results Sorted by Hadronic Interaction Model Prediction . . . 232
F.6 CORSIKA vs MCEq Comparison . . . 237
F.7 Scaling Factors . . . 239
F.8 Stopping Muon Event Display . . . 239

Bibliography **241**

List of Abbreviations **251**

Declaration on the Use of Generative AI **253**

Acknowledgements **255**

1 Introduction

Astroparticle physics uses the Universe to study fundamental particles and forces under conditions that cannot be replicated on Earth by today. At the heart of this field is the study of cosmic rays (CRs), high-energy particles that constantly bombard this planet. When these CRs interact with the Earth's atmosphere, they initiate extensive air showers, cascades of secondary particles that rain down towards the surface. These air showers simultaneously offer a window into the properties of cosmic accelerators and high-energy interactions, while presenting a challenging background for neutrino astronomy.

The focus of this thesis is the atmospheric muon flux produced in these showers. While the majority of these muons originate from the decay of light mesons like pions and kaons, referred to as the *conventional* flux, a subdominant component arises from the decay of heavy, short-lived hadrons. This *prompt* component is characterized by a harder energy spectrum, meaning it becomes increasingly important at high energies.

Understanding this prompt flux is critical for two distinct reasons. First, in the search for astrophysical neutrinos with detectors like the IceCube Neutrino Observatory [1], atmospheric muons and neutrinos constitute the dominant background. As the prompt component closely follows the spectral shape of the astrophysical signal more closely than the conventional flux, it represents an irreducible background that must be precisely modeled and subtracted. Second, the production of charmed hadrons in the forward region of air showers probes Quantum Chromodynamics (QCD) in a phase space that is difficult to access with collider experiments. Discrepancies between observed muon fluxes and theoretical predictions, often referred to as the "Muon Puzzle" [2], suggest that our current understanding of hadronic interactions in air showers may be incomplete.

Measuring the prompt component is, however, an experimental challenge. Over most of the accessible energy range, the prompt flux is several orders of magnitude lower than the conventional component. Furthermore, in the high-energy regime where the prompt fraction becomes significant, the steeply falling cosmic ray spectrum results in very low event rates. Isolating this signal therefore requires a large detector exposure, precise energy reconstruction, and a robust handling of systematic uncertainties.

This work presents a dedicated measurement of the prompt atmospheric muon flux using data from IceCube. By leveraging advanced machine learning reconstruction techniques, a new air shower simulation dataset, and a new event selection, this analysis aims to constrain muon production for muon energies from 10^4 TeV to $1.5 \cdot 10^5$ TeV. For this purpose, an unfolding of the atmospheric muon flux at surface level is performed, and the measured flux is compared to theoretical predictions from state-of-the-art hadronic interaction and cosmic ray models.

The thesis is structured as follows: Chapter 2 outlines the Standard Model of particle physics. Chapter 3 introduces the phenomenology of cosmic rays and the physics of air showers. The IceCube Neutrino Observatory and its detection principles are described in Chapter 4. Chapter 5 details the simulation chain used to model the detector response and air shower development. The reconstruction algorithms, including the machine learning methods developed for this work, are explained in Chapter 6. The selection of the data sample is presented in Chapter 7. The analysis, i.e. a model-independent unfolding of the spectrum is presented in Chapter 8. Finally, Chapter 9 summarizes the results and their implications, followed by an outlook on future improvements in Chapter 10.

2 The Standard Model: A Selected Overview

From the beginning humans aim at finding explanations for phenomena in nature—from the largest scales in cosmos down to the smallest scales at the level of the elementary building blocks. Already in the early days mathematics provided the tools to do so in a consistent, well-understood way. In 1543, Nicolaus Copernicus published a mathematically grounded model of a heliocentric solar system, challenging the long-held geocentric worldview and marking a turning point in our understanding of the universe.

Since then, our view of the cosmos has expanded dramatically: just as the motion of planets could be explained by a few physical laws, it turns out that the immense variety of matter observed today emerges from a remarkably small number of elementary particles. These particles—and the fundamental forces that govern their interactions—are theoretically described by the Standard Model of Particle Physics (SM). It evolved throughout the 20th century following first experimental discoveries of elementary particles: the discovery of the electron by Thompson in 1897 [3] or the β decay in the 1930s which led Pauli to postulate the presence of a neutrino [4]. Over the decades, the SM developed into a modern and testable mathematical description of particle physics capable of predicting the Higgs boson, which has been successfully found in 2012 at the Large Hadron Collider (LHC) by the ATLAS and CMS Collaborations [5, 6]. Its discovery completed the SM as it is known today.

The following section introduces its particle content, organized into fermions and bosons.

2.1 Particle Content

The building blocks of the SM are the fundamental particles, categorized by half-integer and integer-one spin value into fermions and bosons. While the fermions can be distinguished further into leptons and quarks, the gauge bosons mediate the three fundamental forces of the SM: the electromagnetic (EM) force by the photon (γ) [7–9], the weak force by the $W^{\pm}$ and Z^0 bosons [10–12], and the strong force by the gluon (g) [13–15]. The Higgs boson has a special role in the SM as the interaction of particles with the Higgs field gives rise to their mass [16–18].

Leptons There are three identical generations of leptons, each with a charged particle and a corresponding uncharged and massless[1] neutrino. The generations are only distinguishable by their increasing mass: electron (e^-) and electron neutrino (ν_e), muon (μ) and muon neutrino (μ_ν), and the tau (τ) with the tau neutrino (ν_τ). Each lepton has an antiparticle with opposite quantum numbers.

Quarks Quarks, similar to leptons, come in three generations with each an up-type quark (up u, charm c, and top t) and a down-type quark (down d, strange s, and bottom b). They carry fractional charges of $+2/3$ and $-1/3$, respectively. Additionally, quarks have a color charge (red, green, and blue) which allows them to interact via the strong force. However, confinement restricts them to form color-neutral states (hadrons), which are referred to as *baryons* (three (anti)quarks) and *mesons* (quark+antiquark).

[1] In the SM neutrinos are assumed to be massless, however, experimentally confirmed neutrino oscillations require a non-vanishing neutrino mass [19].

In the context of this dissertation, hadron production in the stratosphere is of particular interest as motivated further in Chapter 3. For the relevant energy spectrum of the muons under study, two categories of hadrons are most important: the *light* and *charm-flavored* hadrons. Both are introduced in more detail below.

2.2 Light Hadron Decays

Hadrons are produced in large amounts in interactions of charged particles from outer space with Earth's atmosphere. In the following, a theoretical overview is presented, introducing the hadrons, that are produced most in these interactions and that produce muons as soon as they decay. First, the decays of so-called *light* hadrons are detailed, where the naming corresponds to the quark content they are composed of: up, down, and strange. Owing to their low masses, these hadrons are produced with significantly higher probability than their heavier counterparts. As a result, their decay channels[2] dominate the production of atmospheric muons and are therefore most relevant for the work presented here. Table 2.1 summarizes the decay channels of the most important light hadrons in air showers with information from Ref. [20]. It contains the ground state mesons and baryons as well as the first excited states. As can be seen, by far the largest proportion comes from the relatively long-lived pions and kaons, which live at least two orders of magnitude longer compared to the other light hadrons[3]. This unique property is further in the analysis used to categorize muons originating from these decays as *prompt* and *conventional*.

In the following, a second category of hadrons is introduced, which contain at least one charm quark–the charm hadrons.

[2]It might be mentioned here, that if not stated otherwise, the charge conjugated particle (decay) is always included as well.

[3]Here, only the direct parent particle is of interest for defining the muon production channel, i. e. that e. g. the $\rho^0 \to [\mu^+\nu_\mu]_{\pi^+}[\mu^-\bar{\nu}_\mu]_{\pi^-}$ would lead in approximately 99.9 % of the cases to muons in the final state but this is not considered here.

Table 2.1: Main $\mu^{\pm}$ production channels for different light hadrons h. The quark content of each hadron is written in brackets. Decays with just a cited upper limit in literature are omitted here for clarity. The branching ratios are displayed next to the main decay channels with the total $\mu^{\pm}$ production probability right from the bracket.

Hadron h	Mass [MeV]	Lifetime [s]	Main μ production channel: $h \rightarrow$		
$\boldsymbol{\pi^+}$ $(\lvert u\bar{d}\rangle)$	139.57	$2.60 \cdot 10^{-8}$	$\boldsymbol{\mu^+}\nu_\mu$ $\boldsymbol{\mu^+}\nu_\mu\gamma$	$9.99 \cdot 10^{-1}$ $2.00 \cdot 10^{-4}$	$\approx 100\,\%$
$\boldsymbol{K^+}$ $(\lvert u\bar{s}\rangle)$	493.68	$1.24 \cdot 10^{-8}$	$\boldsymbol{\mu^+}\nu_\mu$ $\pi^0\boldsymbol{\mu^+}\nu_\mu$ $\boldsymbol{\mu^+}\nu_\mu\gamma$ $\pi^+\pi^-\boldsymbol{\mu^+}\nu_\mu$ $\pi^0\boldsymbol{\mu^+}\nu_\mu\gamma$ $\pi^0\pi^0\boldsymbol{\mu^+}\nu_\mu$ $\boldsymbol{\mu^+}\nu_\mu e^+e^-$ $e^+\nu_e\boldsymbol{\mu^+\mu^-}$	$6.36 \cdot 10^{-1}$ $3.35 \cdot 10^{-2}$ $6.20 \cdot 10^{-3}$ $1.40 \cdot 10^{-5}$ $1.25 \cdot 10^{-5}$ $3.45 \cdot 10^{-6}$ $7.06 \cdot 10^{-8}$ $1.70 \cdot 10^{-8}$	$\approx 68\,\%$
$\boldsymbol{K^0_S}$ $(\lvert d\bar{s}\rangle)$	497.61	$8.95 \cdot 10^{-11}$	$\pi^{\pm}\boldsymbol{\mu^{\mp}}\nu_\mu$	$4.56 \cdot 10^{-4}$	
$\boldsymbol{K^0_L}$ $(\lvert d\bar{s}\rangle)$	497.61	$5.10 \cdot 10^{-8}$	$\pi^{\pm}\boldsymbol{\mu^{\mp}}\nu_\mu$ $\pi^{\pm}\boldsymbol{\mu^{\mp}}\nu_\mu\gamma$ $\boldsymbol{\mu^+\mu^-}\gamma$ $\boldsymbol{\mu^+\mu^-}\gamma\gamma$	$2.70 \cdot 10^{-1}$ $5.65 \cdot 10^{-4}$ $3.59 \cdot 10^{-7}$ $1.00 \cdot 10^{-8}$	$\approx 27\,\%$
$\boldsymbol{\eta}$ $\left(\frac{\lvert u\bar{u}\rangle+\lvert d\bar{d}\rangle-2\lvert s\bar{s}\rangle}{\sqrt{6}}\right)$	547.86	$\mathcal{O}(10^{-19})$	$\boldsymbol{\mu^+\mu^-}\gamma$ $\boldsymbol{\mu^+\mu^-}$ $\boldsymbol{\mu^+\mu^-\mu^+\mu^-}$	$3.10 \cdot 10^{-4}$ $5.80 \cdot 10^{-6}$ $5.00 \cdot 10^{-9}$	$< 1\,\%$
$\boldsymbol{\rho^0}$ $\left(\frac{\lvert u\bar{u}\rangle+\lvert d\bar{d}\rangle}{\sqrt{2}}\right)$	775.26	$\mathcal{O}(10^{-24})$	$\boldsymbol{\mu^+\mu^-}$	$4.55 \cdot 10^{-5}$	
$\boldsymbol{\omega}$ $\left(\frac{\lvert u\bar{u}\rangle+\lvert d\bar{d}\rangle}{\sqrt{2}}\right)$	782.66	$\mathcal{O}(10^{-23})$	$\pi^0\boldsymbol{\mu^+\mu^-}$ $\boldsymbol{\mu^+\mu^-}$	$1.34 \cdot 10^{-4}$ $7.40 \cdot 10^{-5}$	$< 1\,\%$
$\boldsymbol{\eta'}$ $\left(\frac{\lvert u\bar{u}\rangle+\lvert d\bar{d}\rangle+\lvert s\bar{s}\rangle}{\sqrt{3}}\right)$	957.78	$\mathcal{O}(10^{-21})$	$\boldsymbol{\mu^+\mu^-}\gamma$ $\pi^+\pi^-\boldsymbol{\mu^+\mu^-}$	$1.13 \cdot 10^{-4}$ $2.13 \cdot 10^{-5}$	$< 1\,\%$
$\boldsymbol{\phi}$ $(\lvert s\bar{s}\rangle)$	1019.46	$\mathcal{O}(10^{-22})$	$\boldsymbol{\mu^+\mu^-}$ $\boldsymbol{\mu^+\mu^-}\gamma$	$2.86 \cdot 10^{-4}$ $1.40 \cdot 10^{-5}$	$< 1\,\%$
$\boldsymbol{\Lambda}$ $(\lvert uds\rangle)$	1115.68	$2.62 \cdot 10^{-10}$	$p\boldsymbol{\mu^-}\bar{\nu}_\mu$	$1.51 \cdot 10^{-4}$	
$\boldsymbol{\Sigma^+}$ $(\lvert uus\rangle)$	1189.37	$8.02 \cdot 10^{-11}$	$p\boldsymbol{\mu^+\mu^-}$	$2.40 \cdot 10^{-8}$	
$\boldsymbol{\Sigma^-}$ $(\lvert dds\rangle)$	1197.45	$1.48 \cdot 10^{-10}$	$n\boldsymbol{\mu^+}\bar{\nu}_\mu$	$4.50 \cdot 10^{-4}$	
$\boldsymbol{\Xi^0}$ $(\lvert uss\rangle)$	1314.82	$2.90 \cdot 10^{-10}$	$\Sigma^+\boldsymbol{\mu^-}\bar{\nu}_\mu$	$2.33 \cdot 10^{-6}$	
$\boldsymbol{\Xi^-}$ $(\lvert dss\rangle)$	1321.70	$1.64 \cdot 10^{-10}$	$\Lambda\boldsymbol{\mu^-}\bar{\nu}_\mu$	$3.50 \cdot 10^{-4}$	

2.3 Charm-Hadron Decays

Charmed hadrons are of particular interest for this thesis, as their contribution in atmospheric interactions has not been measured yet. However, as can be seen in Table 2.2, they decay in a non-negligible number of times into muons and contribute to the atmospheric muon flux. The listed hadrons comprise both charm mesons, such as the D^0, D^+, and D_s^+, each consisting of a charm quark bound with a light antiquark, and charm baryons, like the Λ_c^+ and Ξ_c^0, which are made of three quarks including a charm quark.

Table 2.2: Main $\mu^\pm$ production channels for different charm hadrons h. The quark content of each hadron is written in brackets. Decays with just a cited upper limit are omitted here for clarity. The branching ratios are displayed next to the main decay channels with the total $\mu^\pm$ production probability right from the bracket.

Hadron h	Mass [MeV]	Lifetime [s]	Main μ production channel: $h \to$		
$\boldsymbol{D^0}$ $(\lvert c\bar{u}\rangle)$	1864.84	$4.103 \cdot 10^{-13}$	$K^-\boldsymbol{\mu^+}\nu_\mu$	$3.42 \cdot 10^{-2}$	$\approx 6.8\,\%$
			$K^*(892)^-\boldsymbol{\mu^+}\nu_\mu$	$2.06 \cdot 10^{-2}$	
			$K^-\pi^0\boldsymbol{\mu^+}\nu_\mu$	$7.30 \cdot 10^{-3}$	
			$\pi^-\boldsymbol{\mu^+}\nu_\mu$	$2.67 \cdot 10^{-3}$	
$\boldsymbol{D^+}$ $(\lvert c\bar{d}\rangle)$	1869.5	$1.033 \cdot 10^{-12}$	$\bar{K}^0\boldsymbol{\mu^+}\nu_\mu$	$8.68 \cdot 10^{-2}$	$\approx 17.6\,\%$
			$\bar{K}^*(892)^0\boldsymbol{\mu^+}\nu_\mu$	$5.27 \cdot 10^{-2}$	
			$K^-\pi^+\boldsymbol{\mu^+}\nu_\mu$	$3.62 \cdot 10^{-2}$	
			$\pi^0\boldsymbol{\mu^+}\nu_\mu$	$3.50 \cdot 10^{-3}$	
			$\omega\boldsymbol{\mu^+}\nu_\mu$	$1.77 \cdot 10^{-3}$	
			$\rho^0\boldsymbol{\mu^+}\nu_\mu$	$1.64 \cdot 10^{-3}$	
			$\eta\boldsymbol{\mu^+}\nu_\mu$	$1.04 \cdot 10^{-3}$	
$\boldsymbol{D_s^+}$ $(\lvert c\bar{s}\rangle)$	1968.35	$5.012 \cdot 10^{-13}$	$\phi\boldsymbol{\mu^+}\nu_\mu$	$2.24 \cdot 10^{-2}$	$\approx 5.8\,\%$
			$\eta\boldsymbol{\mu^+}\nu_\mu$	$2.24 \cdot 10^{-2}$	
			$\eta'(958)\boldsymbol{\mu^+}\nu_\mu$	$8.00 \cdot 10^{-3}$	
			$\boldsymbol{\mu^+}\nu_\mu$	$5.37 \cdot 10^{-3}$	
$\boldsymbol{\Lambda_c^+}$ $(\lvert udc\rangle)$	2286.46	$2.026 \cdot 10^{-13}$	$\Lambda\boldsymbol{\mu^+}\nu_\mu$	$3.48 \cdot 10^{-2}$	
$\boldsymbol{\Xi_c^0}$ $(\lvert dsc\rangle)$	2470.44	$1.504 \cdot 10^{-13}$	$\Xi^-\boldsymbol{\mu^+}\nu_\mu$	$1.01 \cdot 10^{-2}$	
$\boldsymbol{J/\psi(1S)}$ $\lvert c\bar{c}\rangle$	3096.90	$7.2 \cdot 10^{-21}$	$\boldsymbol{\mu^+\mu^-}$	$5.96 \cdot 10^{-2}$	$\approx 6.0\,\%$
			$e^+e^-\boldsymbol{\mu^+\mu^-}$	$3.53 \cdot 10^{-5}$	
			$\boldsymbol{\mu^+\mu^-\mu^+\mu^-}$	$1.11 \cdot 10^{-6}$	

These hadrons are relatively short-lived, with lifetimes typically on the order of 10^{-13} to 10^{-12} seconds, and they decay predominantly via semileptonic weak decays, often involving a muon and a muon neutrino in the final state.

As an example, Figure 2.1 illustrates the Feynman diagram for the most probable D^+ decay mode from Table 2.2, $D^+ \to \bar{K}^0\mu^+\nu_\mu$. Due to the weak interaction's flavor-changing nature and the sizable phase space, several decay modes contribute to muon production, with branching fractions ranging from a few per mille to nearly 10 %. Notably, the D^+ meson has the largest total branching ratio to μ^+-producing final states among the listed particles. Other charm hadrons such as Ξ_c^+, Σ_c, or Ω_c^0 are not included here as their decays do not yield muons [20].

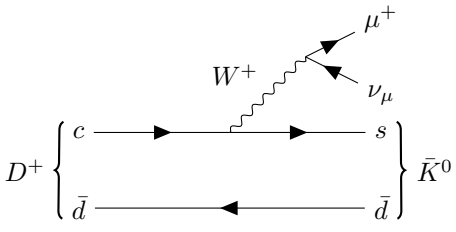


Figure 2.1: Feynman diagram of the weak decay $D^+ \to \bar{K}^0\mu^+\nu_\mu$.

This concludes the chapter about the SM, its particle content and the most relevant unflavored and charm-hadron decays that can give rise to muons. In the following chapter, this theoretical foundation is brought into the context of astroparticle physics.

3 Astroparticle Physics

Astroparticle physics is an interdisciplinary field that explores high-energy particles from space and their interactions with the Earth's atmosphere. This chapter provides an overview of the key concepts and phenomena, including the multi-messenger approach, CRs and their primary fluxes, and the development of air showers; cascades of secondary particles produced when CRs interact with atmospheric nuclei. It also covers the cascade equations used to model these processes, the hadronic interaction models (HIMs), and the muon components of air showers, resulting in a discussion of the ongoing challenge known as the muon puzzle.

3.1 Multi-Messenger Astronomy

Multi-messenger astronomy is an innovative and rapidly evolving field that seeks to understand the universe by observing the same astronomical sources through multiple types of signals, or "messengers". These messengers comprise EM radiation (spanning the spectrum from radio waves to gamma rays), gravitational waves, neutrinos, and CRs. By compiling data from these diverse sources, researchers can achieve a more comprehensive and detailed understanding of astrophysical phenomena, enhancing the ability to probe the universe's most extreme environments and events. This integrative approach allows for cross-validation of findings, thereby increasing the robustness of scientific conclusions and providing insights that are not attainable through single-messenger observations alone.

Traditionally, astronomy began with observations of the universe through visual light, using telescopes like Galileo's in the early 17th century. Galileo's pioneering observations marked a turning point in our understanding of the cosmos. With his telescope, he discovered the moons of Jupiter, demonstrating that not all celestial bodies orbit the Earth, and he resolved the Milky Way into countless individual stars [21]. Over time, astronomy expanded beyond visible light to include the full spectrum of EM radiation, from radio waves to gamma rays, providing unique insights into the physical processes of astronomical objects. For example, the detection of pulsars through radio waves and the study of X-ray emissions from accreting black holes have significantly advanced our understanding of the universe [22]. A landmark discovery was the detection of the cosmic microwave background (CMB) radiation by Penzias and Wilson in 1965, offering compelling evidence for the Big Bang [23, 24]. Despite these advances, EM observations remain limited, since emitted photons can be absorbed by interstellar dust and gas, or attenuated by background radiation fields like the CMB. This motivates the need for additional messengers that can penetrate such blocking matter and reveal hidden processes.

Multi-messenger astronomy addresses these limitations by incorporating other types of signals. Gravitational waves, for instance, are ripples in spacetime caused by accelerating massive objects, such as merging black holes or neutron stars. A landmark event occurred in 2017 with the detection of gravitational waves from the merger of two neutron stars (GW170817) by the Laser Interferometer Gravitational-Wave Observatory (LIGO) and Virgo Collaborations [25]. These observatories detected the gravitational wave signal, which lasted approximately 100 s, providing detailed information about the inspiral phase of the neutron stars and confirming their merger.

This event was also observed across the EM spectrum. The Fermi Gamma-ray Burst Monitor (Fermi-GBM) and the INTEGRAL observatory detected a short gamma-ray burst (GRB) (GRB 170817A) just 1.7 seconds after the gravitational wave signal [26]. This detection established a direct link between

neutron star mergers and short GRBs. Optical follow-up observations began within hours, with the Swope Telescope and Pan-STARRS identifying the optical counterpart [27], later designated as SSS17a, in the galaxy NGC 4993. Additionally, X-ray afterglow emissions were detected by NASA's Chandra X-ray Observatory [28], while radio emissions were observed by the Jansky Very Large Array (VLA) [29], offering insights into the relativistic jets and ejecta dynamics. This event marked the first time a single astrophysical event was detected with both gravitational waves and EM radiation.

Another significant multi-messenger milestone occurred later in 2017 with the detection of a high-energy neutrino, IceCube-170922A, by the IceCube Neutrino Observatory [30]. This neutrino, with an energy of approximately 290 TeV, triggered a real-time alert that led to follow-up observations across the EM spectrum. Fermi-LAT and the MAGIC telescopes identified TXS 0506+056, a blazar located 4 billion light-years away, as the likely source of the neutrino [30]. The blazar was in a flaring state during the detection, emitting gamma rays, X-rays, and optical radiation. This marked the first direct association of a high-energy neutrino with an astrophysical source, confirming that blazars can accelerate CRs and produce neutrinos via hadronic interactions. This event demonstrated the power of multi-messenger astronomy in linking neutrinos and EM observations to pinpoint the origins of high-energy astrophysical phenomena.

Complementing the observations of neutral particles, CRs offer a different perspective on the high-energy universe. These high-energy protons and atomic nuclei, accelerated by sources such as supernova remnants (SNRs), active galactic nuclei (AGNs), and GRBs, provide unique insights into the underlying particle acceleration mechanisms and the interstellar and intergalactic magnetic fields that deflect them during propagation. Beyond their astrophysical importance, CRs are crucial for understanding high-energy particle interactions within Earth's atmosphere, which acts as a target for these cosmic accelerators.

In conclusion, by leveraging the complementary strengths of different observational techniques, multi-messenger observations can probe extreme physical conditions and processes that are otherwise inaccessible. The different messenger types are sketched in Figure 3.1.

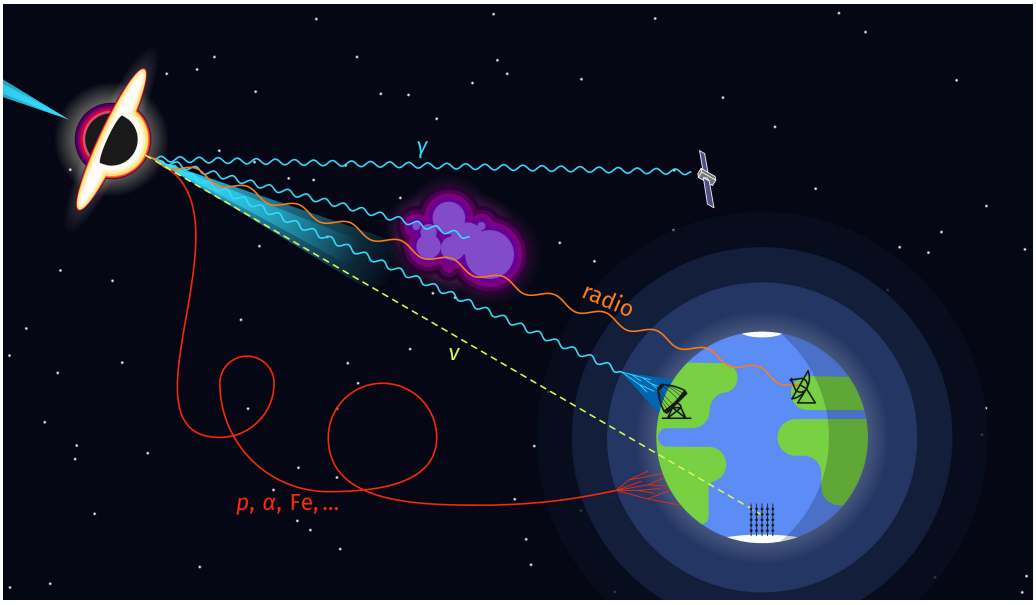


Figure 3.1: This figure illustrates the propagation of different particle types from an astrophysical source to the Earth. Gamma rays (γ, turquoise lines) point to their sources but can be absorbed, unlike radio waves (orange line), which can travel further due to their larger wavelengths. Neutrinos (ν, yellow line) travel straight from their sources and interact rarely. CRs like p, α and Fe (red line) are charged particles deflected by magnetic fields. Figure taken from Ref. [31].

3.2 Cosmic Rays and their Mass Composition

CRs are high-energy, charged particles originating from outer space. These are nuclei, primarily protons ($\approx 90\,\%$), helium nuclei ($\approx 9\,\%$) and heavier nuclei [32]. The discovery of CRs is credited to Victor Hess, an Austrian physicist. In a series of balloon flights from 1911 to 1913, Hess measured ionizing radiation at various altitudes. Contrary to expectations, he found that the radiation strength increased with altitude, indicating an extraterrestrial source. This groundbreaking work earned Hess the Nobel Prize in 1936 and laid the foundation for the field of astroparticle physics, revealing the universe's high-energy processes [33]. However and as mentioned above, EM fields in the Universe deflect charged nuclei and thus provide no direct information about their production origin.

CRs can be classified based on their origin into two main categories: galactic CRs and extragalactic CRs. Galactic CRs originate from our Milky Way galaxy, with SNRs believed to be the primary source. The shock waves from the explosion of massive stars accelerate particles to high energies, ranging from a few GeV to PeV, where they are subject to magnetic confinement in the galactic magnetic field. On the other hand, extragalactic CRs come from outside our galaxy, often from AGNs, GRBs, and other cosmological phenomena. These CRs typically have higher energies, extending up to and beyond the ultra-high-energy cosmic rays (UHECRs) with energies above EeV [34, 35].

The energy spectrum of CRs is presented in Figure 3.2 and follows a power law over a wide range of energies, with a flux that decreases with increasing energy. Here and throughout this thesis, the flux is written in units of $\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$, representing the differential number of particles per unit energy, area (either in m^{-2} or in cm^{-2}), time, and solid angle. This can be described by the differential intensity $\Phi(E)$ as

$$\Phi(E) \propto E^{-\gamma}\,, \tag{3.1}$$

where γ is the spectral index, typically around 2.7 for energies E below the "knee" at around $10^6\,\mathrm{GeV}$. The knee, at energies from $10^6\,\mathrm{GeV}$ to $10^7\,\mathrm{GeV}$, results as the maximum energy that SNRs can accelerate particles to, with an increased spectral index of approximately 3.1. At even higher energies, around $10^8\,\mathrm{GeV}$, the spectrum exhibits a second knee, where the spectral index steepens further to approximately 3.3. This feature marks the maximum energy for heavy galactic CR nuclei, such as iron, as they reach the limits of acceleration and confinement within the galaxy. The spectrum also features the "ankle" at around $10^{9.5}\,\mathrm{GeV}$, where the flux flattens again, suggesting a transition from galactic to extragalactic CRs. This increased flux can be explained with the Hillas energy

$$E = ZeBr\,, \tag{3.2}$$

with the nuclear charge Ze, the magnetic field strength B and the confined accelerator region r [32]. When a charged particle accelerated by a magnetic field exceeds the Hillas energy, it can escape its source and propagate through intergalactic medium. At the highest energies, beyond $10^{10}\,\mathrm{GeV}$, the spectrum steepens sharply due to the Greisen-Zatsepin-Kuzmin (GZK) cutoff [36, 37]. This effect arises from the interaction of UHECRs like protons with photons of the CMB through pion production [20]

$$p + \gamma_{\mathrm{CMB}} \to \Delta^+ \to \begin{cases} p + \pi^0 & (\approx 33\,\%)\,, \\ n + \pi^+ & (\approx 66\,\%)\,. \end{cases} \tag{3.3}$$

In the above reaction Δ^+ marks an excited proton state and is also referred to as Delta resonance [38]. Nuclei heavier than protons instead undergo photodisintegration. These interactions limit the maximum distance over which UHECRs can propagate, leading to a suppression in the observed flux. The GZK cutoff thus marks the effective horizon for CRs with energies above $\approx 50\,\mathrm{EeV}$, providing crucial insight into the interplay between astrophysical acceleration mechanisms and propagation effects.

The CR flux is measured using satellite instruments and ground-based detectors, each offering complementary insights. Satellite-based detectors, such as those aboard the International Space Station

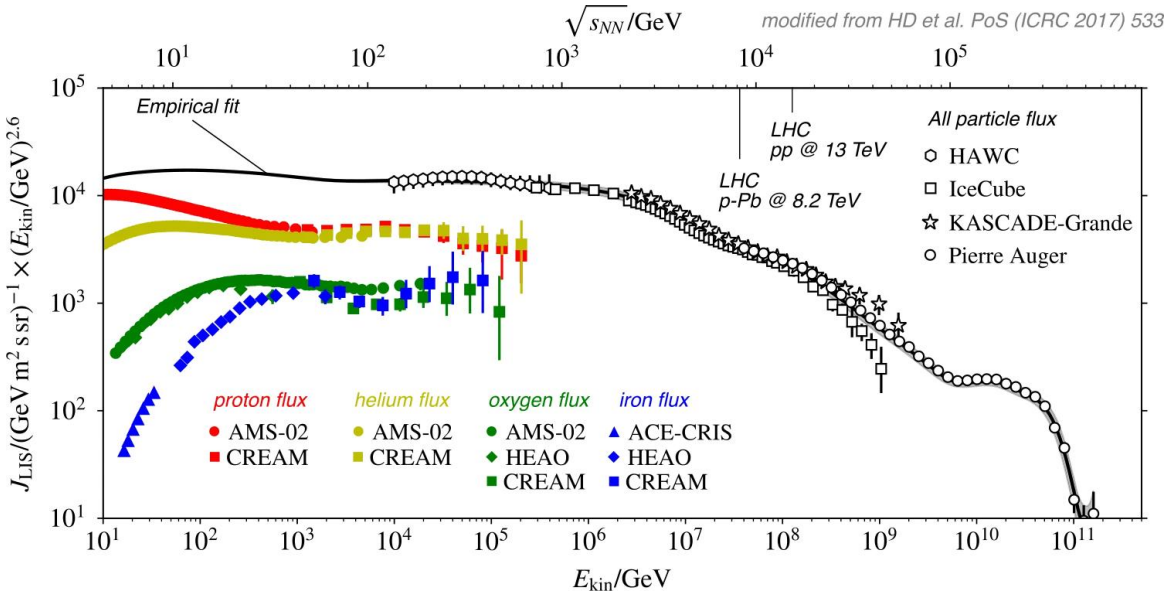


Figure 3.2: The graph depicts the flux of various CR components (proton, helium, oxygen, and iron) as measured by various experiments. The y-axis represents the differential flux, scaled by $E^{2.6}$ in units of $\mathrm{GeV}^{1.6}/\mathrm{m}^2/\mathrm{s}/\mathrm{sr}$. The x-axis shows the kinetic energy per nucleon E_{kin} in GeV. Data from different air-shower experiments (HAWC, IceCube, KASCADE-Grande, and Pierre Auger) are included for all-particle flux, extending the energy range to ultra-high energies. The figure highlights the energy spectrum's behavior across a wide range, from 10 GeV to 10^{11} GeV, with noted comparisons to LHC p-p and p-Pb collision energies [2].

(ISS), measure primary CRs before they interact with Earth's atmosphere. Instruments like the Alpha Magnetic Spectrometer (AMS-02) [39] provide precise measurements of the charge, mass, and energy of CRs, enabling studies of their composition and origins directly in space. Ground-based detectors focus on secondary particles generated when CRs interact with the atmosphere, as detailed in Section 3.3. Large detector arrays like High Altitude Water Cherenkov Experiment (HAWC) [40], KArlsruhe Shower Core and Array DEtector-Grande (KASCADE-Grande) [41], and the Pierre Auger observatory [42] utilize these secondary particles to infer information about the energy and composition of the primary CRs.

In this thesis, different models are used to describe the CR energy spectrum from simulated events. Further details on the simulation are provided in Section 5. Parameterized and pure data-driven models used here are motivated and described below.

Gaisser-Hillas 3 (H3a): This parametrization divides the CR flux into three distinct populations, each associated with different origins and acceleration mechanisms. The first population represents particles accelerated via diffusive shocks in SNRs within the Milky Way, producing CRs with energies up to the knee. The second population, spanning energies from 10^6 GeV to 10^8 GeV, originates from more powerful galactic sources such as pulsars. The third and final population is dominated by extragalactic sources, including AGNs, GRBs, and large-scale structure shocks.

It is assumed that the acceleration depends on the magnetic rigidity

$$R = \frac{E_{\mathrm{tot}}}{Ze}\,, \tag{3.4}$$

with the total energy E_{tot} and charge Ze of the nucleus. Together with the gyroradius, which quantifies how strongly a charged particle is bound to a magnetic field of strength B,

$$r_{\mathrm{L}} = \frac{R}{B}\,, \tag{3.5}$$

the maximum characteristic total energy

$$E_{\mathrm{tot}}^{\mathrm{c}} = ZeBr_{\mathrm{L}} = ZeR_{\mathrm{c}}\,, \tag{3.6}$$

depends on the charge of the nucleus. Thus, a transition from proton to iron is expected towards higher energies.

Then, the total flux per nuclei group $i \in \{\mathrm{H/p},\ \mathrm{He},\ \mathrm{C\text{-}N\text{-}O},\ \mathrm{Mg\text{-}Al\text{-}Si},\ \mathrm{Fe}\}$, ϕ_i, is described by

$$\phi_i(E) = \sum_{j=1}^{3} a_{i,j} E^{-\gamma_{i,j}} \cdot \exp\left(-\frac{E}{Z_i e R_{\mathrm{c},j}}\right) \tag{3.7}$$

for the three populations j, with normalization $a_{i,j}$, and spectral index $\gamma_{i,j}$. Here, ϕ_i is presented as the differential flux $\frac{\mathrm{d}N}{\mathrm{d}\log(E)}$ [43].

Gaisser-Hillas (H4a): This model is an alternative approach to H3a, as described in Equation 3.7, but modifies the third population of extragalactic CRs to consist solely of protons, rather than five components in the H3a model [43].

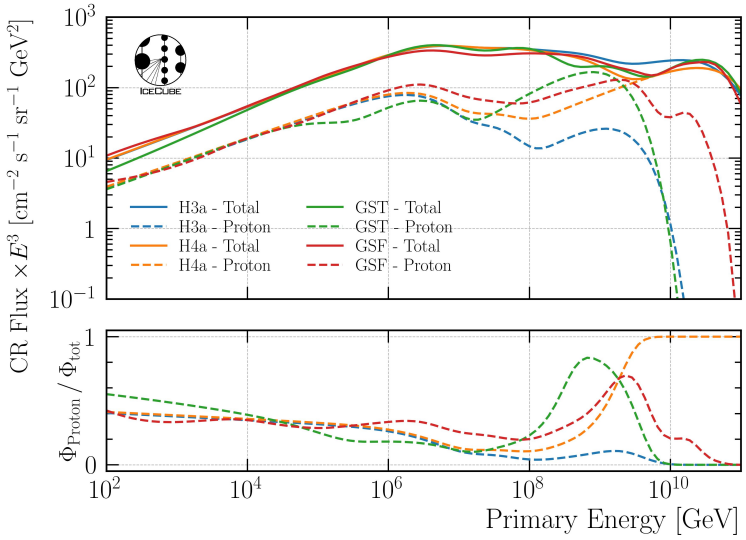
Gaisser-Stanev-Tilav (GST): Similar to the H3a model described in Equation (3.7), the GST model employs three populations but introduces an additional mass group heavier than iron. The model is constructed based on fits to observed data from multiple experiments, providing a refined description of the CR spectrum and composition [44].

Global Spline Fit (GSF): The GSF model is a data-driven framework without underlying theoretical assumptions that describes the CR flux and mass composition by combining direct measurements from satellites and balloons with indirect data from ground-based air-shower experiments. It employs spline functions to create a smooth and flexible fit across the energy spectrum. The model tracks the composition through four primary elemental groups (protons, helium, oxygen, and iron) and explicitly accounts for systematic uncertainties, such as energy scale differences between experiments. The differential flux as a function of rigidity R is given by

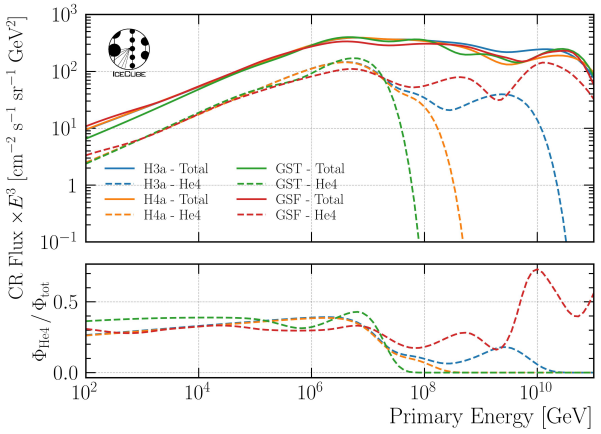
$$J_{\mathrm{L}}(R) = \sum_k a_{\mathrm{L}k} \cdot b_k \left(\ln\left(\frac{R}{\mathrm{GV}}\right)\right) \cdot \left(\frac{R}{\mathrm{GV}}\right)^{-3}\,, \tag{3.8}$$

for each leading component $L \in \{\mathrm{H/p},\ \mathrm{He},\ \mathrm{O},\ \mathrm{Fe}\}$, with $b_k(\ln(R/\mathrm{GV}))$ as standard cubic B-splines, normalization parameters $a_{\mathrm{L}k}$ and shaping parameter $(R/\mathrm{GV})^{-3}$ [45]. A conversion into the differential flux per unit energy as written in Equation (3.7) can be performed by evaluating $J_{\mathrm{L}}(R)$ at $R = E_{\mathrm{tot}}/(Ze)$ and a final division by Ze.

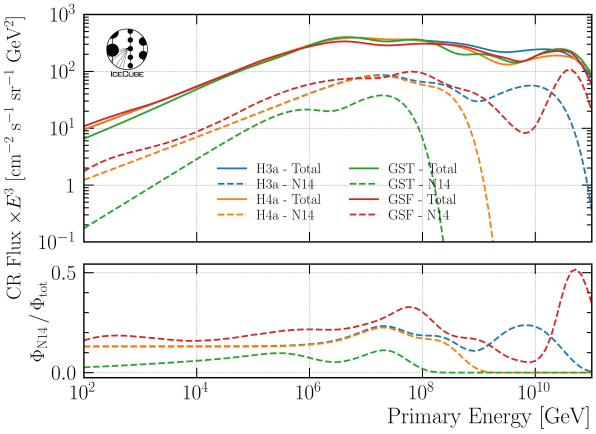
The contribution of the primary particles p, He, N, Al, and Fe used in this thesis to the total flux are presented for all CR models in the Figure 3.3. The flux is weighted with the primary energy E^3 to make spectral features visible. This is common in astroparticle physics because fluxes extend over several orders of magnitude and thus look indistinguishable in an un-scaled representation. For all five particle types, there are differences between all presented models, and the contributions diverge towards higher energies. For H4a, energies above 10^9 GeV are dominated by protons only. For H3a, this energy range is dominated by heavier particles. In GST, the highest energies are mainly modeled by iron, followed by protons. For GSF, the strongest contribution in the highest enegies stems from helium. Overall, even though the total fluxes of all models are similar, the individual contributions differ.



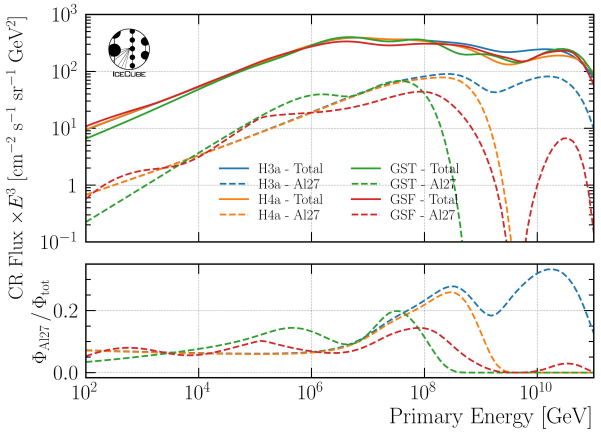
(a) Proton.



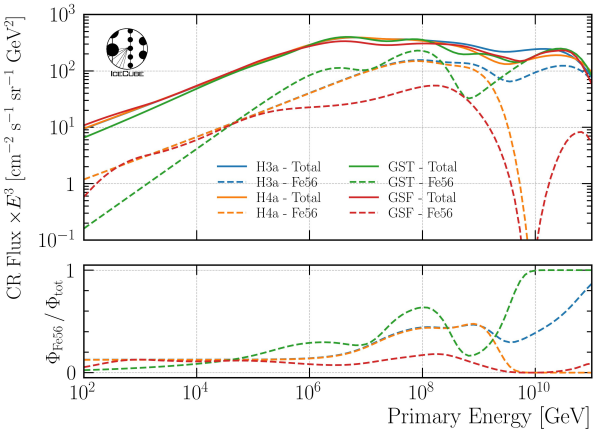
(b) Helium.



(c) Nitrogen.



(d) Aluminum.



(e) Iron.

Figure 3.3: Contribution per component of the particle flux to the entire particle flux scaled with E^3. Dotted lines show the total flux and are the same in all plots. The ratio of each component with respect to the total flux is presented below. Although the flux compositions differ for each model, the total flux is similar for each model presented here.

3.3 Air Showers

For particle physics, CRs serve as natural accelerators, providing access to energies far beyond those achievable by human-made accelerators[1]. This allows to measure and probe particle-production cross sections at highest energies.

When CRs enter the Earth's atmosphere, they interact with atmospheric nuclei, primarily nitrogen and oxygen, producing extensive air showers. In air showers, cascades of secondary particles, primarily the lightest mesons pions and kaons, emerge and propagate through the atmosphere. Especially π^0 decays are of importance, as they dominantly decay into a pair of photons. These photons can then produce electron-positron pairs in the EM field of a nucleus

$$\gamma + N \to e^+ + e^- \,. \tag{3.9}$$

These electrons undergo bremsstrahlung in an EM field of a nucleus or a valence electron, radiating a photon,

$$e^\pm + N \to \gamma + \gamma \,. \tag{3.10}$$

This process happens avalanche-like. Under the assumption that no other processes contribute to the shower development, this leads to the formulation of the *Heitler model* [47]. In the simplified model, the development of an EM shower is described as a series of interactions in which particles repeatedly split their energy while doubling in number. At each step, a high-energy photon undergoes pair production or an electron undergoes bremsstrahlung, dividing its energy equally between two secondary particles. After n steps, the number of particles is

$$N = 2^n \,, \tag{3.11}$$

and the energy per particle is

$$E_n = \frac{E_0}{2^n} \,, \tag{3.12}$$

where E_0 is the energy of the primary particle. The cascade continues until the energy of individual particles drops below the critical energy E_c, where ionization losses dominate, and the shower dissipates. At this point, the number of particles in the shower reaches a maximum, approximately

$$N_{\max} \sim \frac{E_0}{E_c} \,. \tag{3.13}$$

The depth $X_{\max}$ at which this maximum occurs depends logarithmically on the energy of the primary particle,

$$X_{\max} \sim \ln\left(\frac{E_0}{E_c}\right) . \tag{3.14}$$

While the Heitler model neglects physical effects such as scattering, fluctuations, and additional interactions, it provides a simple yet effective framework for understanding the basic structure of EM showers.

Similar to the EM component, the Heitler-Matthews extension also describes cascade-like air shower development for the hadronic component, involving particles such as pions and kaons [48]. Each interaction produces k secondary hadrons, where k is the multiplicity of the interaction. After n generations, the number of particles in the cascade is approximately $N_n = k^n$, and the energy per particle is

$$E_n = \frac{E_0}{k^n} \,. \tag{3.15}$$

[1] Even possible future accelerators like the Future Circular Collider as a proton-proton collider with center-of-mass energies of $\sqrt{s_{\mathrm{NN}}} = 100\,\mathrm{TeV}(\approx 10^7\,\mathrm{TeV}$ fixed target experiment energy) would not exceed the highest possible energies achievable with CRs [46]. See also Figure 3.2.

Of particular relevance to the analysis in this thesis is the muonic component of air showers, since muons are the only particles that penetrate to the Earth's surface and can be directly measured. Hadrons producing muons are listed in Tables 2.1 (unflavored hadrons) and 2.2 (charmed hadrons). By analyzing the muonic content of air showers, the mass and energy of the primary cosmic ray can be inferred since the hadronic component is crucial for shaping the structure of air showers. The relative proportions of neutral and charged pions determine the balance between the EM and muonic components, and the lateral spread of the shower depends on the distribution of secondary hadrons.

In Figure 3.4, an air shower is illustrated. An incoming proton with a spectral index of $\gamma = 2.7$ interacts with the atmosphere and produces a cascade of secondary particles, like charged and neutral pions and kaons. The particles decay further into muons and neutrinos, which then enter the IceCube detector, as explained in Chapter 4. Further details about the different origins of muons by hadronic interactions and decays are explained in the following sections.

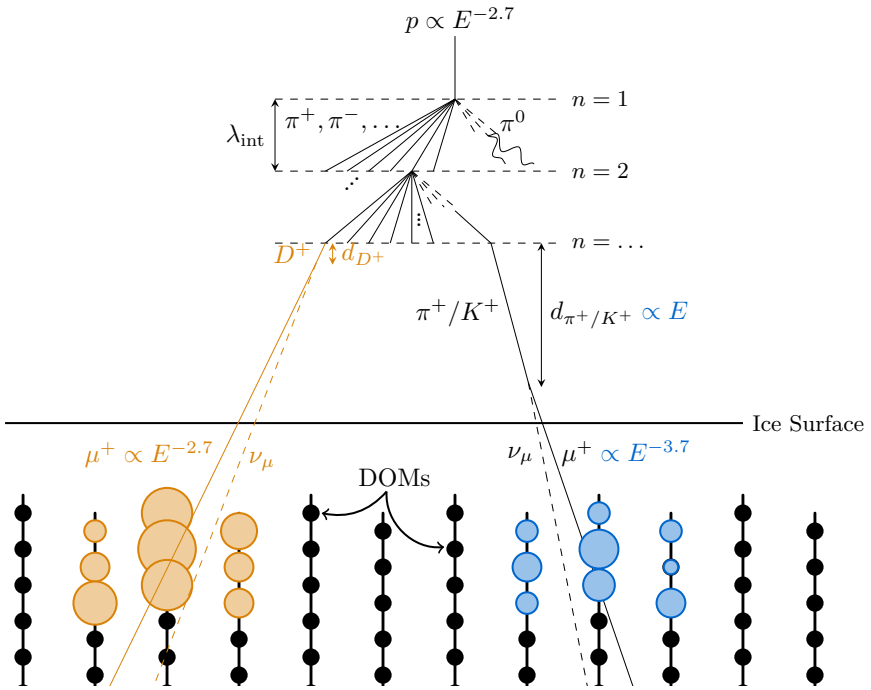


Figure 3.4: Schematic illustration of an air shower initiated by a primary proton. The primary proton interacts with atmospheric nuclei, producing a cascade of secondary particles, including charged and neutral pions and kaons. These secondary particles further decay into muons and neutrinos, which can be detected by instruments like IceCube. Also, heavier particles are produced in the atmosphere, here a D-meson is shown. The figure highlights the different channels prompt and conventional colored in orange and blue, respectively, which are explained in more detail below. Figure adapted from Ref. [49].

3.4 Cascade Equations

While the Heitler model as described above provides a qualitative framework to understand EM and hadronic showers, it remains limited by its simplifying assumptions. In particular, it assumes constant multiplicities, fixed critical energies, and neglects the full energy spectrum and decay channels of secondary particles. However, *cascade equations* provide a semi-analytical way to model Extensive Air Showers (EASs) continuously in the atmosphere, accounting for both interactions and decays. These equations generalize the Heitler model by including energy-dependent interaction cross sections,

inclusive production spectra, decay probabilities, and the full chain of particle species involved in the cascade. Cascade equations are a set of coupled differential equations, for which the *slant depth* X in units of of $\mathrm{kg\,m^{-2}}$ is used as geometrical axis along the shower

$$X(y) = \int_0^y \rho(l)\,\mathrm{d}l\,. \tag{3.16}$$

Here, y is the observation altitude above Earth's surface and $\rho(l)$ the height-dependent density of the atmosphere. For hadrons h the propagation through a medium is described by the set of equations

$$\frac{\mathrm{d}\Phi_h}{\mathrm{d}X} = -\frac{\Phi_h(E,X)}{\lambda_{\mathrm{int},h}(E)} - \frac{\Phi_h(E,X)}{\lambda_{\mathrm{dec},h}(E)} \tag{3.17}$$

$$- \frac{\partial}{\partial E}\left(\mu(E)\Phi_h(E,X)\right) \tag{3.18}$$

$$+ \sum_\ell \int_E^\infty \Phi_\ell(E_\ell,X)\left(\frac{1}{\lambda_{\mathrm{int},\ell}}\frac{\mathrm{d}N_{\ell+\mathrm{air}\to h}}{\mathrm{d}E} + \frac{1}{\lambda_{\mathrm{dec},\ell}}\frac{\mathrm{d}N_{\ell\to h}}{\mathrm{d}E}\right)\mathrm{d}E_\ell\,. \tag{3.19}$$

The *sink* terms (3.17) account for the depletion of a given particle species, either through inelastic interactions with air nuclei or via its decay into secondary particles. Term (3.18) accounts for energy losses from ionization and radiative processes. The *source* term (3.19) is a sum over all possible particle types ℓ. It accounts for either the production of particle h through interactions with the air ($N_{\ell+\mathrm{air}\to h}$) or via direct decay ($N_{\ell\to h}$). The decay length $\lambda_{\mathrm{dec},i}(E)$ is defined as

$$\lambda_{\mathrm{dec},i}(E) = \frac{E\tau_i}{m_i c}\cdot\rho(X)\,, \tag{3.20}$$

with the energy of the particle E, its mass m_i, its proper lifetime in the rest frame τ_i, the speed of light[2] c and the density of the atmosphere at slant depth X. It quantifies in units of $\mathrm{g\,cm^{-2}}$ how far a particle i travels on average before it decays.

In general, solving the full set of cascade equations analytically is not feasible due to the energy-dependent interaction and decay terms, the coupling between particle species, and the energy-integral source terms. However, under certain simplifying assumptions, semi-analytical approximations can be obtained. A widely used approach for hadronic cascades is the Z-moment method (here, only protons are considered first for simplicity[3]), which makes several assumptions:

- The inclusive production spectra of secondary particles depend mainly on the energy fraction $x = E_h/E_\ell$ (ratio of energies between the outgoing and incoming particle with $0 < x < 1$; outgoing particle carries a fraction of the incoming particle energy) and only weakly on the absolute energy of the parent particle. This is known as Feynman scaling and holds approximately at high energies. It allows to rewrite the production term via

$$\frac{\mathrm{d}N_{\ell+\mathrm{air}\to h}(E_h,E_\ell)}{\mathrm{d}E_h} = \frac{1}{E_\ell}\frac{\mathrm{d}N_{\ell+\mathrm{air}\to h}(x)}{\mathrm{d}x} \tag{3.21}$$

- The interaction length

  $$\lambda_{\mathrm{int},h} = \lambda_{\mathrm{int},p} = m_p/\sigma_{p\text{-air}}\,,$$
  $$\text{with } m_p = \text{proton mass and } \sigma_{p\text{-air}} = \text{cross section of proton-air interaction}\,,$$

  is independent of the energy.

- The primary flux follows a power law as function of the energy: $\Phi_p = A_p(X)\cdot E_p^{-\gamma}$.

[2] $c = 299\,792\,458\,\mathrm{m\,s^{-1}}$

[3] Protons do not decay on the timescale of EASs and thus, decay terms are neglected

Inserting the aforementioned assumptions in Eqs. (3.17) and (3.19), it follows for the simplified assumption of $p + \text{air} \to p'$ and a direct division by $E_p^{-\gamma}$:

$$\frac{\mathrm{d}A_p(X)}{\mathrm{d}X} = -\frac{A_p(X)}{\lambda_{\text{int},p}} + \frac{A_p(X)}{\lambda_{\text{int},p}} \int_{E_p}^{\infty} \left(\frac{E_{p'}}{E_p}\right)^{-\gamma} \frac{\mathrm{d}N_{p+\text{air}\to p'}(E_{p'})}{\mathrm{d}E_p}\,\mathrm{d}E_{p'} \tag{3.22}$$

Feynman and substitution of variables: $x = \dfrac{E_p}{E_{p'}} \Rightarrow \mathrm{d}E_{p'} = -\dfrac{E_p}{x^2}\mathrm{d}x$

$$\frac{\mathrm{d}A_p(X)}{\mathrm{d}X} = -\frac{A_p(X)}{\lambda_{\text{int},p}} - \frac{A_p(X)}{\lambda_{\text{int},p}} \int_{E_{p'}\to\infty\Rightarrow x\to 0}^{E_{p'}\to E_p\Rightarrow x\to 1} x^{\gamma}\frac{x}{E_p}\frac{\mathrm{d}N_{p+\text{air}\to p'}(x)}{\mathrm{d}x}\left(-\frac{E_p}{x^2}\right)\mathrm{d}x \tag{3.23}$$

$$= -\frac{A_p(X)}{\lambda_{\text{int},p}} \cdot \underbrace{\left(1 - \int_0^1 x^{\gamma-1}\frac{\mathrm{d}N_{p+\text{air}\to p'}(x)}{\mathrm{d}x}\,\mathrm{d}x\right)}_{1-Z_{pp'}}. \tag{3.24}$$

The integral in Equation (3.24) is called the spectrum weighted Z-factor. From this, the solution for the proton cascade can be calculated as

$$\Phi_p(E_p, X) = A_p(0)E_p^{-\gamma}\exp\left(-\frac{X(1-Z_{pp'})}{\lambda_{\text{int},p}}\right). \tag{3.25}$$

In order to calculate the lepton flux using the formalism outlined above, a more general approach is followed, originally introduced in Ref. [50]. In this formulation, the energy-dependent Z factor is written as

$$Z_{\ell h}(E_h) = \int_{E_h}^{\infty} \left(\frac{E_\ell}{E_h}\right)^{\gamma(E_\ell)-1} \frac{\lambda_{\text{int},\ell}(E_h)}{\lambda_{\text{int},\ell}(E_\ell)}\frac{\mathrm{d}N_{\ell+\text{air}\to h}}{\mathrm{d}E_\ell}\,\mathrm{d}E_\ell\,. \tag{3.26}$$

This ansatz introduces three additional levels of detail [51]. First, the spectral index γ is no longer assumed to be constant but is introduced as an energy-dependent function of any flux Φ, defined by

$$\gamma(E) = -\frac{E}{\Phi(E)}\frac{\mathrm{d}\Phi(E)}{\mathrm{d}E}\,.$$

This allows to vary the spectral index with energy due to the knee or population transition in models like H3a. Second, the energy dependence of the interaction length is now taken into account via $\lambda_{\text{int}}(E) \propto 1/\sigma_{\text{inel}}(E)$. It captures the growth of the inelastic cross section with energy. Third, instead of assuming Feynman scaling with a fixed ratio $x = E_h/E_\ell$, the full energy-dependent yield spectrum $\mathrm{d}N/\mathrm{d}E_\ell$ is considered. This allows the particle production to depend independently on both E_ℓ and E_h.

For a more general description than only for protons, the decay term in the source term in Equation (3.19) needs to be considered in a similar way. It can also be written in terms of a weighted Z factor as

$$Z_{\ell h}^{\text{dec}} = \int_{E_h}^{\infty} \left(\frac{E_\ell}{E_h}\right)^{\gamma(E_\ell)-1} \frac{\lambda_{\text{dec},\ell}(E_h)}{\lambda_{\text{dec},\ell}(E_\ell)}\frac{\mathrm{d}N_{\ell\to h}}{\mathrm{d}E_\ell}\,\mathrm{d}E_\ell\,. \tag{3.27}$$

This allows to rewrite the cascade Eqs. (3.17) and (3.19) as

$$\frac{\mathrm{d}\Phi_h}{\mathrm{d}X} = -\frac{\Phi_h(E,X)}{\lambda_{\text{int},h}(E)} - \frac{\Phi_h(E,X)}{\lambda_{\text{dec},h}(E)} + \sum_\ell \left(\frac{Z_{\ell h}\Phi_\ell(E_\ell,X)}{\lambda_{\text{int},\ell}} + \frac{Z_{\ell h}^{\text{dec}}\Phi_\ell(E_\ell,X)}{\lambda_{\text{dec},\ell}}\right). \tag{3.28}$$

In the following section the cascade equations for the atmospheric muon flux are introduced, which allow to define muons as *conventional* and *prompt*.

Table 3.1: Critical energies for the most relevant parent particles of atmospheric muons. Due to the τ^{-1} dependence of $\varepsilon_{\text{crit}}$ long-lived particles dominate the production at lower energies, while the short-lived heavy hadrons dominate the muon production towards higher energies.

Hadron	Mass [MeV]	Lifetime [s]	$c \cdot \tau$ [m]	Critical energy $\varepsilon_{\text{crit}}$ [GeV]
π^+	139.57	$2.603 \cdot 10^{-8}$	7.80	$1.14 \cdot 10^2$
K^+	493.68	$1.238 \cdot 10^{-8}$	3.71	$8.51 \cdot 10^2$
K^0_{S}	497.61	$8.954 \cdot 10^{-11}$	$2.68 \cdot 10^{-2}$	$1.18 \cdot 10^5$
D^0	1864.84	$4.103 \cdot 10^{-13}$	$1.23 \cdot 10^{-4}$	$9.71 \cdot 10^7$
D^+	1869.50	$1.033 \cdot 10^{-12}$	$3.10 \cdot 10^{-4}$	$3.84 \cdot 10^7$
D^+_s	1968.35	$5.012 \cdot 10^{-13}$	$1.50 \cdot 10^{-4}$	$8.34 \cdot 10^7$
Λ^+_c	2286.46	$2.026 \cdot 10^{-13}$	$6.07 \cdot 10^{-4}$	$2.44 \cdot 10^8$

3.5 The Atmospheric Muon Flux: Conventional and Prompt

Muons reaching Earth's surface are solely produced in interactions of e. g. CRs with the atmosphere, since they would have decayed on their way through space. Cascade equations as introduced above allow to calculate the atmospheric muon flux semi-analytically. For this, the relevant terms read as follows[4]

$$\frac{\mathrm{d}\Phi_\mu(E,X)}{\mathrm{d}X} = \underbrace{-\frac{\Phi_\mu(E,X)}{\lambda_{\text{dec},\mu}(E)}}_{\text{Sink term}} + \underbrace{\frac{Z^{\text{dec}}_{\pi\mu}\Phi_\pi(E,X)}{\lambda_{\text{dec},\pi}(E,X)} + \frac{Z^{\text{dec}}_{K\mu}\Phi_K(E,X)}{\lambda_{\text{dec},K}(E,X)} + \frac{Z^{\text{dec}}_{K^0_{\text{L}}\mu}\Phi_{K^0_{\text{L}}}(E,X)}{\lambda_{\text{dec},K^0_{\text{L}}}(E,X)} + \dots}_{\pi,\, K, \text{ and } K^0_{\text{L}} \text{ source terms}} . \tag{3.29}$$

The source terms written in Equation (3.29) account for the main production channel of atmospheric muons via the decay of charged pions and kaons (see also Table 2.1). The competition between decay and further interaction of a hadron i is determined by its *critial energy* $\varepsilon_{\text{crit},i}$, defined as the energy at which the decay length of it equals its interaction length

$$\varepsilon_{\text{crit},i} = \frac{m_i c h_0}{\tau_i} . \tag{3.30}$$

In Equation (3.30), h_0 refers to the atmospheric scale and describes how atmospheric density decreases with altitude l starting from the density at sea level ρ_0

$$\rho(l) = \rho_0 \mathrm{e}^{-\frac{l}{h_0}} .$$

Assuming an isothermal atmosphere, h_0 can be calculated from the ideal gas law via $h_0 = {}^{k_{\text{B}}T}/_{mg} \approx 6.4\,\text{km}$. This parameter enters into the calculation of critical energies of hadrons because it defines the effective path length over which the density changes and, thus, the probability of hadron decay versus interaction in the atmosphere.

Below $\varepsilon_{\text{crit},i}$, hadron decays dominate, while above $\varepsilon_{\text{crit},i}$ interactions dominate. A summary of the critical energies for the most relevant hadrons of this thesis is given in Table 3.1. This allow to categorize atmospheric muons into conventional and prompt, depending on their parent particles and the corresponding decay characteristics:

Conventional muons They arise predominantly from the decays of long-lived light mesons, namely charged pions and kaons. At low energies ($E \ll \varepsilon_{\text{crit},\pi,K}$) these hadrons decay promptly and the resulting muon flux follows the same spectral index as the primary CR spectrum ($\Phi^{\text{conv}}_\mu \propto E^{-\gamma}$).

[4] Charges are omitted for clarity here, if relevant, neutral charged are written explicitly.

At $E \gg \varepsilon_{\mathrm{crit},\pi,K}$ these hadrons are more likely to interact before decaying, leading to a steeper spectrum ($\Phi_\mu^{\mathrm{conv}} \propto E^{-(\gamma+1)}$). Typically, leptons originating from mesons with $c \cdot \tau(K_{\mathrm{S}}^0) > 2.68\,\mathrm{cm}$ are referred to as *conventional*, while the rest is called *prompt*.

Prompt muons They originate from the decay of short-lived heavy hadrons, primarily charm and unflavored hadrons. Their critical energies ($\mathcal{O}(\varepsilon_{\mathrm{crit},D^0,D^+,D_s^+,\Lambda_c^+}) = (10^7 - 10^8)\,\mathrm{GeV}$) are so high that, for all practical energies (up to the PeV range and beyond), these particles decay immediately after production. As a result, their muon spectrum is not suppressed by interaction effects and retains the same spectral index as the primary CRs ($\Phi_\mu^{\mathrm{prompt}} \propto E^{-\gamma}$).

While the prompt flux continues as $E^{-\gamma}$, the conventional flux steepens by one additional power of energy ($E^{-(\gamma+1)}$) beyond $\varepsilon_{\mathrm{crit},\pi} \sim 114\,\mathrm{GeV}$ and $\varepsilon_{\mathrm{crit},K} \sim 851\,\mathrm{GeV}$. Due to this different spectral behavior, the prompt component decreases more slowly with energy. Consequently, a crossover is expected at energies around 1 PeV, where prompt muons begin to dominate the atmospheric muon flux.

3.5.1 `MCEq` and Hadronic Interaction Models (HIMs)

In order to solve the cascade equations numerically, the interaction and decay Z factors in Eqs. (3.26) and (3.27) must be computed from the energy spectra and decay distributions of secondary particles.

In practice, this is accomplished using the `MCEq` framework. `MCEq` is a modular, one-dimensional cascade-equation solver that evolves particle fluxes through the atmosphere as a function of energy and slant depth. Its core algorithm relies on tabulated or interpolated inclusive differential cross sections and decay spectra to compute the energy-dependent Z-moments numerically. These inputs are provided by HIMs and enable precise, model-dependent predictions of secondary lepton fluxes observed in neutrino telescopes and CR detectors.

This, in turn, requires knowledge of the inclusive particle-production cross sections, which are not accessible from first principles in the non-perturbative regime of Quantum Chromodynamic (QCD). Instead, phenomenological HIMs such as `SIBYLL`[5] are used to extrapolate experimental collider data to higher energies and forward phase-space regions relevant to EASs. These models provide information on particle-production spectra, cross sections, hadronization, and fragmentation fractions, among other quantities.

`SIBYLL` is based on the dual-parton model, in which hadron–hadron interactions proceed via multiple color-string excitations that subsequently fragment into hadrons. At high energies, perturbative-QCD effects are modeled through a minijet component. In version `2.3d`, updated parton distribution functions and an improved tuning of charm-hadron production have been implemented, leading to a more accurate description of the kinematics and spectral shapes of charmed particles.

3.5.2 Muon Flux Prediction

As described above, `MCEq` allows to predict the muon flux as a function of its energy at Earth's surface in GeV. The spectrum is scaled by E^3 to emphasize also small changes in a double-logarithmic representation as shown in Figure 3.5. The total flux as well as its individual components are highlighted together with ratio plots of the different components to the conventional, prompt, and total flux, respectively.

The charged pions and kaons account for approximately 99.5 % of the conventional muon flux over the entire energy range. The remaining contribution of about 0.5 % to the conventional flux is mainly

[5]Version `2.3d` is used throughout this thesis [52]

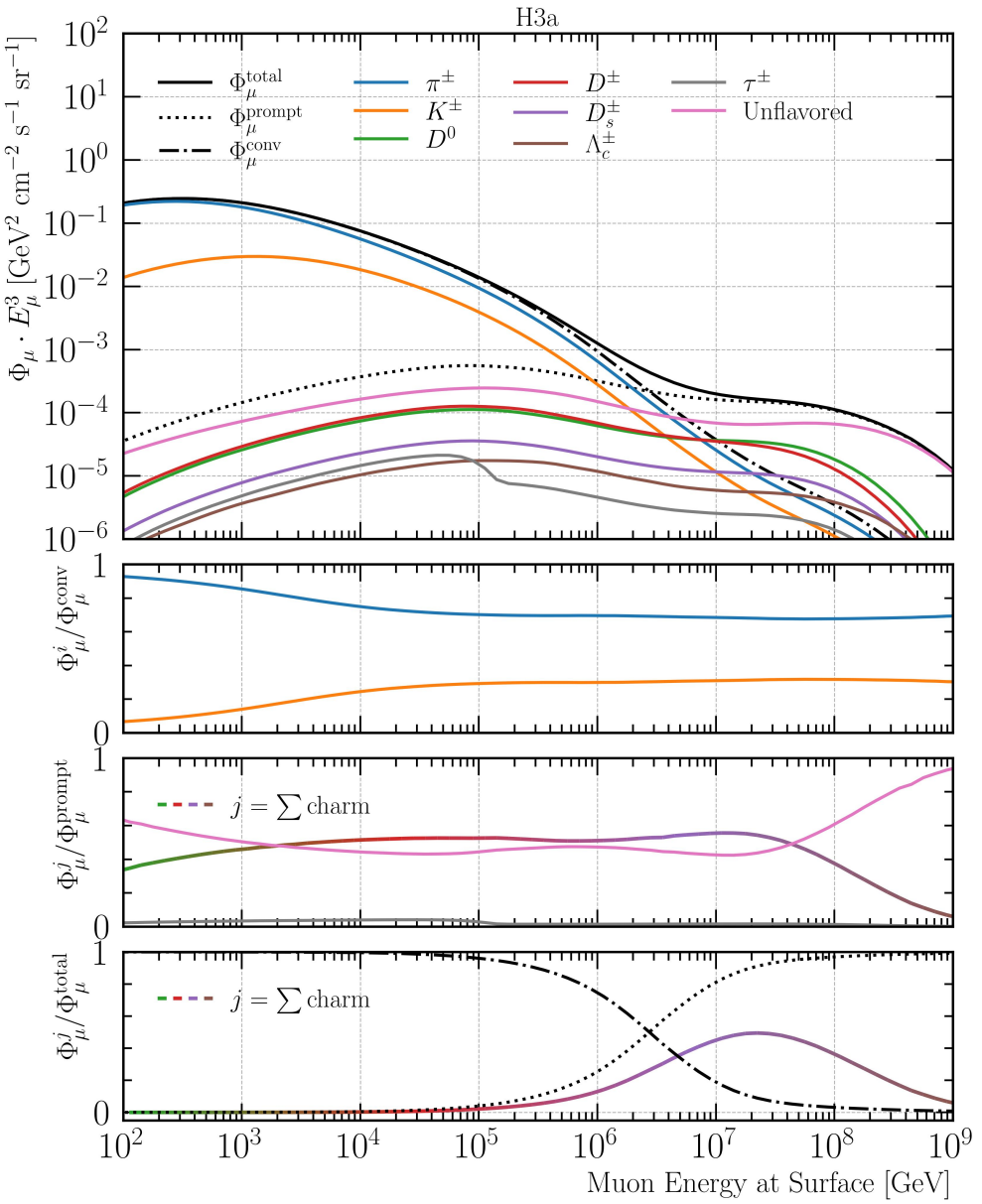


Figure 3.5: Predicted muon flux at the Earth's surface as a function of energy. Shown are the total flux, the conventional and prompt muon fluxes, and their respective parent contributions. Ratio plots emphasize the most important contributors to each component. A transition from a conventional-dominated to a prompt-dominated flux occurs at about $2 \cdot 10^6$ GeV. All fluxes are computed with `MCEq` using `SIBYLL 2.3d` as the HIM and H3a as the primary cosmic-ray model.

composed of K_L^0 decays, which decays in 27 % in muons, but are omitted for clarity in the plot. For the prompt muon flux, the charm component $\in [D^0, D^\pm, D_s^\pm, \Lambda_c^\pm]$ accounts for about half of the entire prompt muon flux, while most of the remaining fraction is composed of decays of the unflavored η, η', ρ^0, ω, ϕ mesons (they account for about 99 % of the unflavored portion to the prompt flux). The leptonic $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$ decay channel with a branching ratio of 17.39 % makes up the remaining contribution to the prompt muon flux[6]. However, towards the highest energies above 10^8 GeV the unflavored component overtakes the charm component and becomes the dominant contribution to the prompt flux.

Overall, the prompt flux is suppressed in the lower energy regime up to 10^6 GeV in contrast to the conventional muon flux. Above $\approx 2 \cdot 10^6$ GeV the prompt flux becomes the dominant contribution, composed to about 50 % of charm-hadron decays in the energy range relevant for this thesis.

3.6 The Muon Puzzle

Despite the advances in modeling particle cascades and predicting the atmospheric muon flux, a persistent discrepancy remains between simulations and observations–referred to as the *Muon Puzzle* [2]. Measurements of extensive air showers by several experiments, including the Pierre Auger Observatory and Telescope Array, consistently report a higher number of muons at ground level than predicted by leading HIMs such as `SIBYLL` [52, 53], `QGSJet-II` [54], or `EPOS-LHC` [55].

This discrepancy becomes particularly large at ultra-high energies of the primary CR particles $\gtrsim 10^9$ GeV, where the muon content in inclined air showers exceeds predictions by up to 30 % to 60 %, depending on the zenith angle and shower reconstruction method. Since the muon number is highly sensitive to the details of hadronic multiparticle production, especially in the forward region of phase space, this excess challenges the current understanding of particle interactions at energies well beyond the reach of ground-based accelerators.

To quantify this tension between data and simulation, the Working group for Hadronic Interactions and Shower Physics (WHISP) introduced the *muon rescaling parameter* z, defined as

$$z = \frac{\log(N_\mu^{\text{data}}) - \log(N_\mu^{\text{p}})}{\log(N_\mu^{\text{Fe}}) - \log(N_\mu^{\text{p}})}\,, \tag{3.31}$$

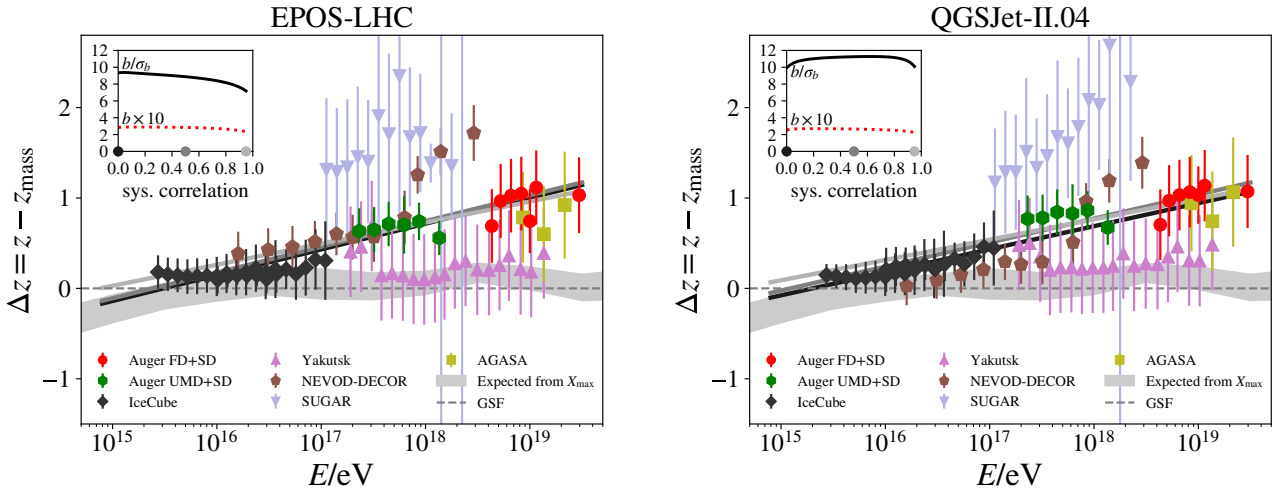
where N_μ^{data} is the observed muon number, and N_μ^{p}, N_μ^{Fe} are the simulated average muon numbers for proton and iron primaries, respectively. This variable interpolates between expectations for pure proton ($z = 0$) and iron-induced ($z = 1$) showers. Values $z > 1$ imply that the observed muon number exceeds even the predictions for iron showers, highlighting the excess of muons in a model-independent way.

To further isolate the discrepancy from uncertainties in the primary mass composition, the difference

$$\Delta z = z - z_{\text{mass}} \tag{3.32}$$

is introduced, where z_{mass} represents the expected value of z corresponding to the composition inferred from shower observables such as the depth of the shower maximum X_{max}, assuming the HIM is correct. In this context, z_{mass} is calculated by simulating the expected muon number N_μ^{mass} for the inferred composition and replacing N_μ^{data} with N_μ^{mass} in Equation (3.31). A nonzero value of Δz, in particular $\Delta z > 0$, indicates that the observed muon number exceeds the predictions even after accounting for the inferred primary mass. This provides a composition-independent measure of the muon excess and is shown in Figure 3.6 for the two HIMs `EPOS-LHC` and `QGSJet-II.04`. In both figures, data from

[6] Due to its lifetime of $2.903 \cdot 10^{-13}$ s, it follows a $c \cdot \tau_\tau \approx 8.7 \cdot 10^{-5}$ m and a critical energy of $\varepsilon_{\text{crit},\tau} = 1.18 \cdot 10^{11}$ GeV.



(a) `EPOS-LHC` used as HIM.

(b) `QGSJet-II.04` used as HIM.

Figure 3.6: The muon rescaling parameter z subtracted by z_{mass} as defined in the text as a function of the primary particle energy on the x-axis. A deviation from $\Delta z = 0$ (no measured muon deficit) has been excluded by a linear fit (solid lines) by at least eight standard deviations. Figures taken from Ref. [2].

several experiments are shown in addition to the expectation. In both cases, a non-zero slope of Δz, presented as black line in Figure 3.6, as a function of the primary particle energy could be excluded with a significance of eight standard deviations.

Several hypotheses have been proposed to resolve the muon deficit, including modified hadronization schemes, enhanced baryon or strange-meson production, and suppressed production of π^0 mesons which would otherwise feed the electromagnetic component of the shower. Recent results from LHC experiments such as LHCb suggest that forward production of strangeness and baryons may indeed be underestimated in current models. Dedicated CR and collider efforts–such as oxygen-oxygen collisions and forward physics experiments at the LHC–are therefore crucial to further constrain and improve the modeling of high-energy air showers. On the other hand, measurements by the IceCube Neutrino

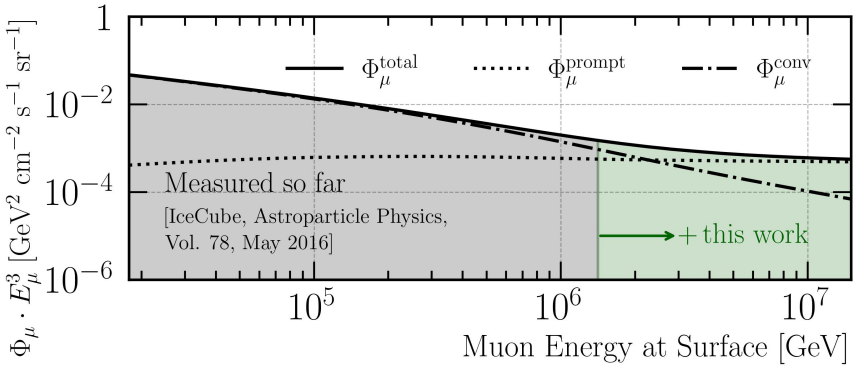


Figure 3.7: A zoomed representation of the muon flux scaled by E_μ^3 predicted by `MCEq`. It illustrates the energy range of the measured muon flux as gray area from Ref. [56] and the additional energy range added in this work. As can be seen, at energies above approximately $2 \cdot 10^6$ GeV the prompt muon flux becomes the dominant contribution.

Observatory probe the high-energy muon content in air showers up to approximately $1.5 \cdot 10^6$ GeV energies. By identifying stochastic bremsstrahlung losses in muon bundles, IceCube can distinguish events containing high-energy muons, providing evidence for a prompt component from decays of

charmed and unflavored hadrons. While the conventional flux from light-hadron decays above 10^4 GeV is well reproduced by `SIBYLL2.1`, discrepancies in the zenith-angle distribution of unresolved muons remain, consistent with the broader muon deficit problem. This upper limit is also illustrated in Figure 3.7, which presents the energy range of the already measured muon flux, but also highlights the additional energy range added by this dissertation. It can clearly be seen that the energy range now extends above the range, where the prompt muon flux component becomes the dominant fraction of the total muon flux.

4 The IceCube Neutrino Observatory

Installed deep down in the antarctic ice as part of the Amundsen-Scott South Pole Station and spanning a surface area of approximately $1\,\mathrm{km}^2$, the IceCube Neutrino Observatory is a neutrino telescope developed and built to measure neutrino signatures from 10^7 to 10^{21} eV. The surrounding ice is used as a transparent detection volume of about $1\,\mathrm{km}^3$ to infer the origin and energy of charged particles as remnants of high-energetic neutrinos and CR decay products from emitted light. In the following chapter, the detector design in described in detail in Section 4.1 before the detection principle is outlined in Section 4.2. Flavor-specific neutrino signatures are described in Section 4.2.3 which also concludes the chapter.

4.1 Detector Design

The unique location of the IceCube telescope has several advantages. Its volume of one cubic kilometer is sufficiently large to account for the small cross section of neutrinos interacting with matter. Secondly, the ice has a long optical attenuation length which allows for a wide spacing of optical modules. Details on the optical instruments are provided below. Finally, technical infrastructural and personnel logistics can be provided by the Amundsen-Scott South Pole Station.

For convenience, the coordinate system used by IceCube is introduced here and shown in Figure 4.1. The IceCube detector is indicated as gray pictogram with a through-going muon shown as black line. Its $(0, 0, 0)^\top$ point is located in the center of the detector with its z-component extending from the surface to the ground. The azimuth and zenith angle $\phi \in [0, 2\pi]$ and $\theta \in [0, \pi/2]$ are indicated as blue and orange lines as well.

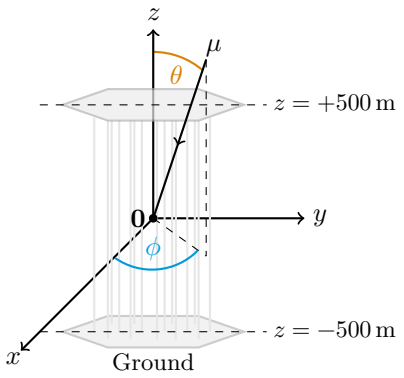
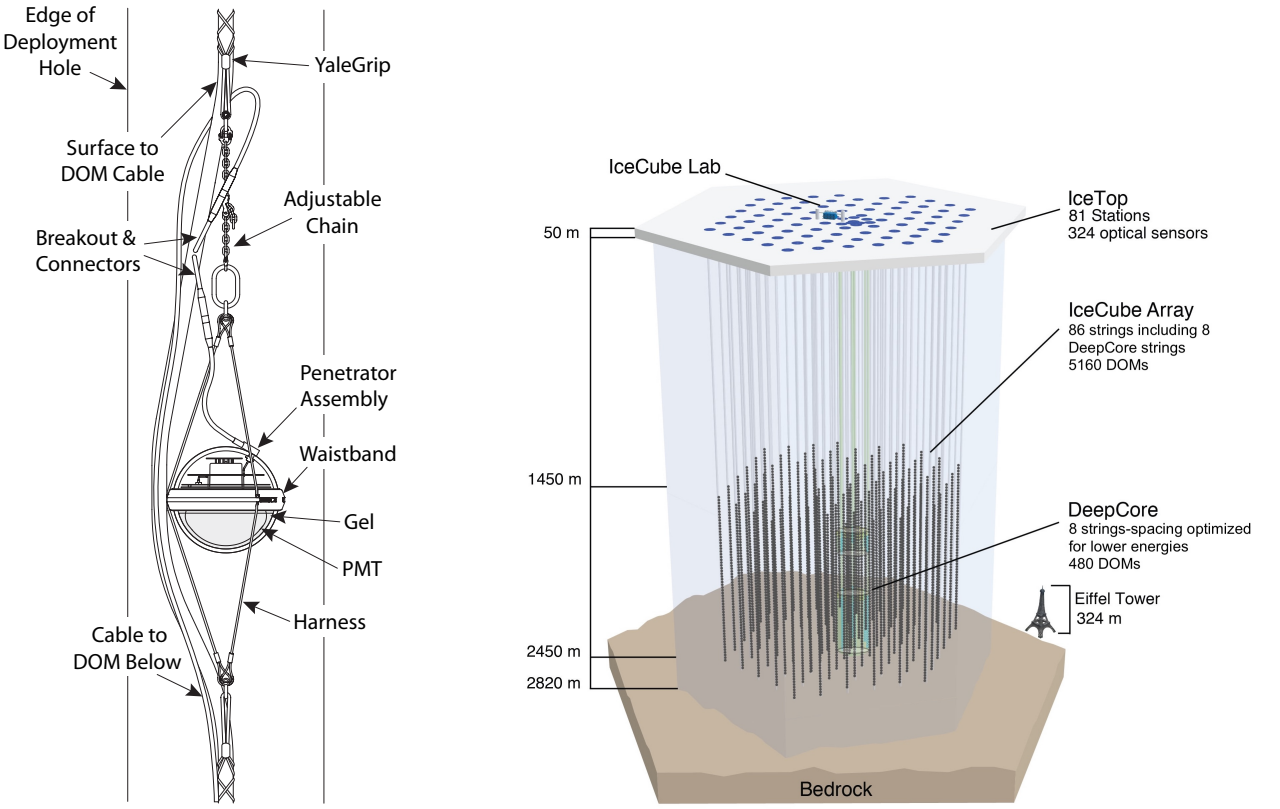


Figure 4.1: The coordinate system of the IceCube Neutrino Observatory with its $(0, 0, 0)^\top$ point located in the center of the detector.

4.1.1 Digital Optical Modules

The general concept of light detection emitted by charged particles traversing the ice is based on digital optical moduledigital optical modules (DOMs). They consist of a spherical glass with an outer diameter of 33 cm, a downward-facing 25 cm photomultiplier tube (PMT) and electronics for power supply, data acquisition, and calibration on top of it [1]. In addition, each DOM has a LED installed for calibration of neighboring DOMs. An illustrative sketch is presented in the left panel of Figure 4.2. It shows the DOM held in place by a harness and a chain together with the cable for power and communication. Photon hit times as well as the corresponding PMT waveform are stored and sent to the IceCube Lab (ICL) in bunches of 1 s recordings.



(a) Installation of a DOM in a deployment hole. The spherical DOM is shown together with additional mounting accessories. Figure taken from Ref. [1].

(b) Arrangement of strings forming the IceCube Neutrino Observatory. Eiffel tower shown for scale comparison. Figure modified from Ref. [1].

Figure 4.2: Sketched representation of a DOM in the left panel and the full arrangement in the right panel. In total, a volume of $1\,\mathrm{km}^3$ is instrumented.

4.1.2 Arrangement of Digital Optical Modules

To date, the main IceCube in-ice array consists of 78 hexagonally arranged strings with a horizontal spacing of 125 m which hold DOMs in depths between 1450 m and 2450 m with a vertical spacing of 17 m. 60 DOMs are attached to each string (4680 in total). The in-ice array has lower and upper energy thresholds of 100 GeV and $\mathcal{O}$(PeV), respectively.

In order to lower the minimum energy threshold by about a factor of ten down to 10 GeV, the DeepCore array comprises eight additional strings [57]. In depths between 2100 m and 2450 m DOMs are installed with vertical spacings of 7 m. Therefore DeepCore allows to further extend the sensitivity in searches for solar weakly interacting massive particles or atmospheric neutrino oscillations [57]. Depths between 2000 m and 2100 m are not monitored by DeepCore due to reduced transparency originating from a layer of dust in the ice which increases optical scattering and absorption in this region [58, 59].

On top of the ice, the IceTop array is installed to detect CR showers and to serve as a coincident event trigger to detect downgoing neutrinos [1, 60]. With an arrangement of 81 stations with each two ice-filled tanks – where each tank contains two DOMs – IceTop is sensitive to CR energy ranges from 10^6 GeV up to 10^9 GeV.

While IceCube has made significant contributions in recent years [61–64], the planned IceCube-Gen2 extension aims to further increase its sensitivity and extend its accessible energy range. The planned upgrade comprises a significantly expanded detector volume through additional in-ice strings and

enhanced surface instrumentation, enabling improved resolution of neutrino sources and more precise measurements of ultra-high-energy cosmic-ray events [65]. Consequently, IceCube-Gen2 will significantly advance the understanding of astrophysical neutrinos and atmospheric particle physics.

4.2 Detection Mechanism

Neutrinos interact only via the weak force of the SM necessitating large target volumes like the cubic kilometer of ice used by IceCube due to their low interaction cross section with matter. Even then, neutrinos cannot be directly detected but via charged secondary particles. These secondary particles then produce light signatures in the detector which can be captured with DOMs. Both, the production of secondary particles and their light signature in the detector are described in the following.

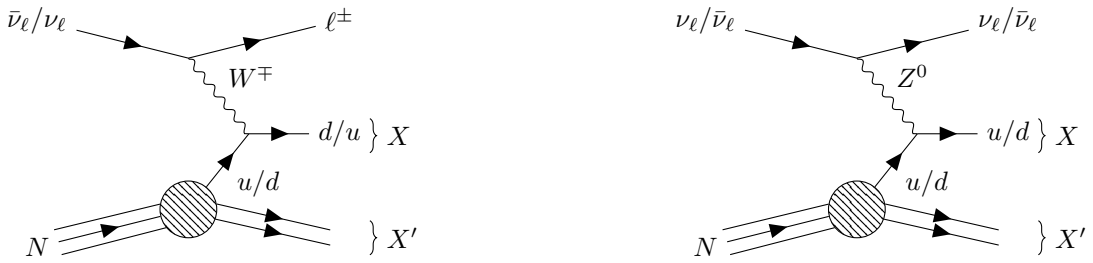
4.2.1 Neutrino Interactions in Ice

Although the main goal of this analysis is to directly measure prompt muons from the decay of heavy hadrons in the atmosphere, neutrinos from atmospheric or stellar sources produce charged leptons in ice as well. They are considered as background in this analysis and therefore explained below.

Neutrinos with energies of the order of $\mathcal{O}(\text{TeV})$ interact with the ice molecules (two hydrogen and one oxygen atom) via deep inelastic scattering. There are two processes that describe the collision of neutrinos with the quarks of the hydrogen and oxygen atoms, the charged current (CC) and neutral current (NC) interactions. Both are shortly described below.

Charged current The CC interaction proceeds via the transformation of the neutrino ν_ℓ into a charged lepton $\ell^\pm$ through the exchange of a $W^\pm$ boson, with $\ell \in \{e, \mu, \tau\}$. In the nucleus N the quark charge changes by one while also changing the flavor primarily between u and d (or vice versa) dictated by the probabilities of the Cabibbo-Kobayashi-Maskawa (CKM) matrix. The process at parton level is shown in the left panel of Figure 4.3. Finally, a cascade from the resulting hadrons X and X' emerges, which usually cannot be resolved due to the sparse detector grid.

Neutral current The NC interaction is mediated by a Z^0 boson which leaves the quark flavor unchanged and can only be seen in the detector by a hadronic shower initiated by the hadrons X and X'.



(a) CC weak interaction mediated by $W^\pm$ boson resulting in a charged lepton $\ell^\pm$.

(b) NC weak interaction mediated by Z^0 boson depositing the visible energy entirely in the hadrons X and X'.

Figure 4.3: Feynman diagrams of the charged (left panel) and neutral (right panel) current interaction.

4.2.2 Cherenkov Radiation

Measurable light signatures from charged particles like leptons arise from the Cherenkov effect [66, 67]. Its origin lies in the transient polarization of the atomic electron shell of an atom like hydrogen or oxygen. This causes EM radiation due to the time-varying electric dipole moment. In case the velocity of the charged particle v_p is above the phase velocity c/n with c being the speed of light in vacuum and n the refractive index of the medium[1], the emitted radiation from individual atoms interferes constructively under a characteristic angle, resulting in the emission of a cone of light. This characteristic angle is given by

$$\cos(\theta_\mathrm{C}) = \frac{c}{n v_\mathrm{p}} \,. \tag{4.1}$$

The intensity of the Cherenkov light emission directly depends on its wavelength λ which scales with $1/\lambda^2$, resulting in the emission of predominantly short wavelengths in the blue spectrum. The emitted light travels through the ice and is eventually captured by the DOMs. Depending on the arrival time and the intensity of the light, track or cascade like signatures as well as the energy can be inferred in the reconstruction process. The different types of signatures are outlined in further detail below.

4.2.3 Particle Signatures

As described above, particles penetrating the ice with higher velocity than the phase velocity of light in ice induce the emission of blue light. However, depending on the outgoing lepton $\ell^\pm$ in the CC interaction different light signatures can be seen in the detector. The hadronic remnants on the other side (e. g. X and X' in Figure 4.3 from both CC and NC interactions) usually produce spherical, locally restrained light patterns, which add light especially at the interaction point. They are, since independent from the lepton flavor, not further discussed below. A schematic sketch is shown in Figure 4.4 for a track-like signature from a muon on the left side and a cascade from e. g. a hadronic interaction, which shows the located light pattern compared to the evolving light cone as the muon travels through the ice.

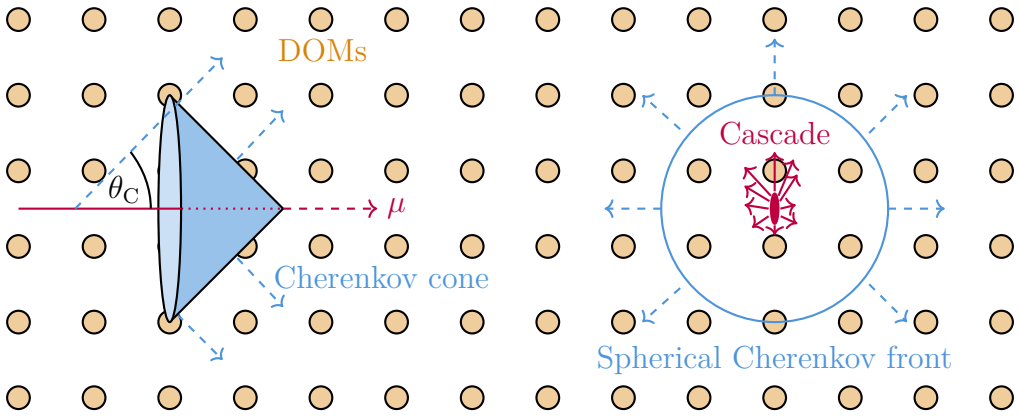


Figure 4.4: The Cherenkov light-cone from a muon as it traverses the ice on the left side and the spherical signature from a hadronic cascade on the right side. The DOMs which capture the light are indicated as orange dots. Figure modified from Ref. [69].

Lepton-Ice Interactions

Leptons interacting with ice molecules deposit their energies through different mechanisms:

[1] $c = 299\,792\,458\,\mathrm{m\,s^{-1}}$ and $n \approx 1.78$ below 300 m depth [68].

- **Bremsstrahlung** Photon radiation when the lepton is deflected by the nucleus, the radiative loss scales via m_ℓ^{-2} [70]
- **Ionization** Shell electron from nucleus is kicked out [71, 72]
- **Pair Production** High-energy lepton produces an $\ell'^{+}\ell'^{-}$ pair near the nucleus, the energy loss scales via m_ℓ^{-2} [73, 74]
- **Photonuclear Interaction** Interaction with the nucleus via a virtual photon, which excites or breaks apart the nucleus [75]

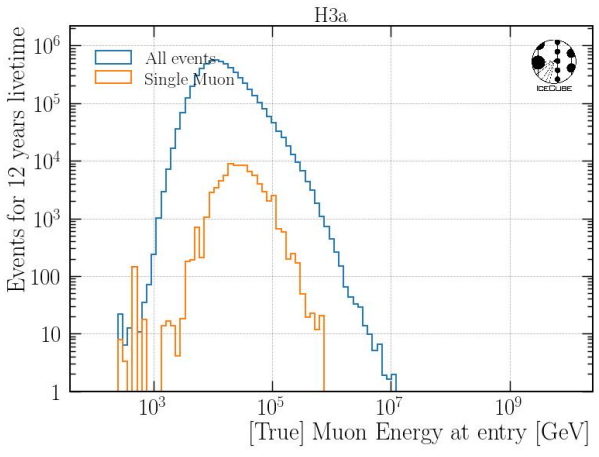
Depending on the lepton flavor and energy, the energy losses contribute in different proportions to the total energy loss.

ν_e signature

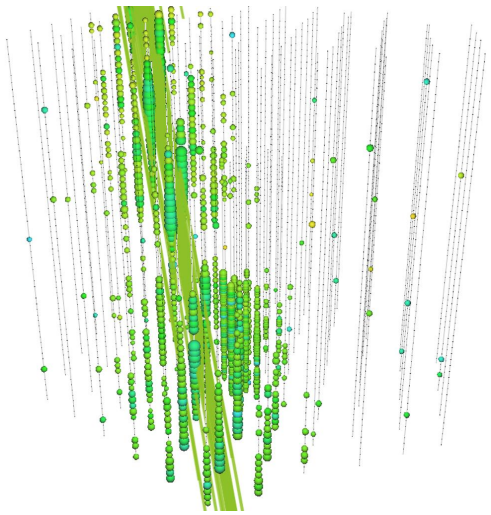
Electron-neutrino interactions via the CC produce both, an electromagnetic and a hadronic cascade. Due to the low mass ($m_e = 511.00\,\text{keV}$ [20]) the main interaction of the outgoing charged electron is via bremsstrahlung when deflected by the nuclei of the ice. The resulting bremsstrahlung photons can then convert into a pair of electron and positron (pair production) which themselves radiate photons. This electromagnetic shower is locally resolved and creates a roughly spherical pattern.

μ and ν_μ signature

Charged muons with energies typically above 1 TeV entering the instrumented detector volume lose their energy mainly via bremsstrahlung, pair production, photonuclear interaction and ionization of the ice atoms' electrons. These processes happen much more gradually for muons than for electrons due to their approximately 200 times heavier mass ($m_\mu = 105.66\,\text{MeV}$ [20]). This allows the muon to travel long distances and producing Cherenkov light along its way. Its pattern is recognized as a track in the detector.



(a) Single-muon contribution to the total event rate in bins of the muon energy.



(b) Event display of a muon bundle.

Figure 4.5: Muon event rate energy distribution separately for single muons and muon bundles. A suppression of at least one order of magnitude over the entire energy spectrum can be seen. The right panel shows an example event display with multiple simulated muons entering the detector, a muon bundle.

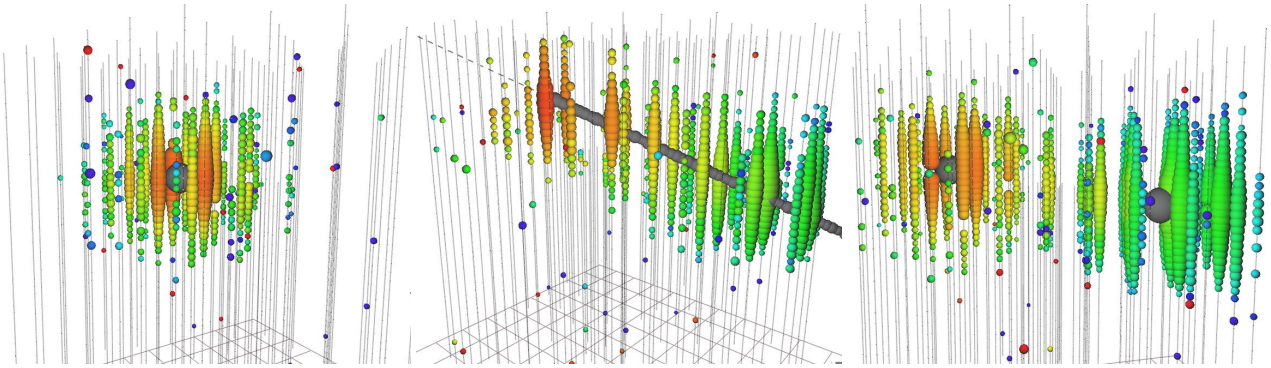
Usually, a bundle of muons from an air shower enters the detector leading to so-called *muon bundles*. While they are spatially difficult to distinguish from single muons, their energy loss signature differs. The superposition of continuous energy losses from multiple muons results in a smoother light deposition profile compared to large stochastic losses of single high-energy muons. These bundles majorly contribute to the muon flux from atmospheric interactions. This is illustrated in Figure 4.5, which presents in the left panel the contributions from single muons and muon bundles to the entire spectrum. The right panel shows a corresponding event display with multiple simulated muons entering the detector.

In contrast, muon neutrinos producing a charged muon in the CC interaction issue an additional hadronic cascade primarily at the beginning of the track's origin.

ν_τ signature

Although the tau lepton is by far the heaviest charged lepton ($m_\tau = 1.78\,\text{GeV}$ [20]) produced in the CC interaction, its short lifetime of the order of $10^{-13}\,\text{s}$ only allows it to travel a few tens of meters depending on the energy. In any case, a localized cascade at the interaction point is visible from the hadronic component. In addition, the energy and subsequent τ^+ decay further define the signature in the detector. With a fraction of approximately 65 % the hadronic decay $\tau^+ \to h^+\nu_\tau$ issues a second localized cascade at the τ^+ decay vertex [20]. In case of a leptonic decay into a muon via $\tau^+ \to \mu^+\nu_\mu\bar{\nu}_\tau$ a track-like signature from the muon arises from the τ^+ decay vertex. Both second signatures appear displaced from the first hadronic cascade depending on the energy of the τ^+. However, searches for this characteristic *double bang* signature are challenging due to the low ν_τ flux from atmospheric interactions, their short decay length which makes them indistinguishable from hadronic interactions and backgrounds from coincident muons. Here, stochastic muon losses can mimic the τ^+ decay. An analysis of 9.7 years of IceCube data revealed seven possible *double bang* candidates, possibly issued by ν_τ interactions [76].

The different signatures induced by the different neutrino flavors are shown in Figure 4.6. It presents a shower signature from an electron neutrino, a track-like signature from a muon, as well as a *double bang* induced by a tau neutrino from left to right, respectively. All pictures have been derived from simulated events. The color code indicates the light-arrival time (early to late in red to green), while the size of the blobs is proportional to the amount of deposited light in the DOMs.



(a) ν_e induced shower. (b) ν_μ induced track. (c) ν_τ induced *double bang*.

Figure 4.6: Three possible signatures in the IceCube detector with color coding explained in the text below. Figures taken from Ref. [77].

4.3 IceCube Data Acquisition System

Light information from traversing particles is recorded by DOMs as a waveform, which corresponds to the collected charge from the PMTs as a function of time in a time window of 6.4 µs (referred to as *pulse* in the following) [1]. Reconstruction algorithms are employed to infer quantities such as the energy or the direction of a track from a series of pulses at the different DOMs [78]. An example of a pulse is given in Figure 6.4. The collected charge in bins of the relative time is shown in the top panel. Here, relative time corresponds to bias in time from an optimized time window of 6 µs which maximizes the total collected charge c_{total}. A Gaussian kernel density estimator (KDE) is used to approximate the charge collection over time as presented in red. Furthermore, the cumulative charge is extracted from the Gaussian KDE and shown in the lower panel of Figure 6.4. Different characteristic quantities can be extracted from both representations, which aim at describing the pulse [79]. In this analysis, they serve as inputs for a multivariate energy-, direction-, and track-geometry estimation and are further discussed in Chapter 6.

Only the timestamp information from a GPS-synchronized precision clock (timing resolution of the order of ns) as well as the maximum voltage are stored if no neighboring DOMs has recognized a light signal within 1 µs. The local coincidence is used to suppress background noise and reduce the amount of data sent to the ICL. This stage of the data processing is referred to as *Level1*. At surface, trigger systems look for time- and space-correlated patterns of DOMs like the Simple Multiplicity Trigger 8 (SMT8) and Simple Multiplicity Trigger 12 (SMT12) trigger, which rejects events where less than eight/twelve DOMs have fired within a sliding 5 µs window. If met, information of all DOMs in a surrounding time window of a few µs are combined into one event and further passed to the event builder. This includes a first pulse shape cleaning to reduce background noise as well as a likelihood-based track fitting procedure [80]. Further information on the reconstructions are given in Chapter 6. This processed data represents the *Level2* stage. The combined and triggered events sum up to approximately 100 $\frac{\text{GB}}{\text{d}}$ and are transmitted via satellite to offline reconstruction sites [81].

5 Simulation

The measurement of the prompt atmospheric muon flux with IceCube requires precise simulation of particle interactions and propagation from the upper atmosphere to the detector array deep within the Antarctic ice. To achieve this, IceCube employs a modular simulation framework known as IceTray [82]. It provides the flexibility to integrate various specialized simulation components into a coherent processing chain. Central to the simulation approach is the Monte Carlo Simulation (MC), which uses random sampling to simulate particle interactions and detector responses realistically. Given IceCube's primary function as a neutrino observatory, neutrino event simulation is a central component. An early neutrino generator utilized was Neutrino Simulation (NuSim) [83], succeeded by All Neutrino Interaction Simulation (ANIS) [84]. The most recent generator, Neutrino Generator (NuGen) [85], is adapted from ANIS. NuGen samples neutrinos from a power law distribution

$$\phi_{\mathrm{s}}(E_{\mathrm{p}}) \propto E_{\mathrm{p}}^{-\gamma_{\mathrm{s}}}\,, \tag{5.1}$$

with the sampled flux ϕ_{s} depending on the primary CR particle energy E_{p}, and the spectral index γ_{s}. In a simulation, typically $\gamma_{\mathrm{s}} = 2$ is used. This generates neutrinos whose primary energy distribution approximately follows a realistic astrophysical spectrum, $\gamma_{\mathrm{astro}} = 2.52 \pm 0.04$, as measured by IceCube [86]. By assuming a given spectral index and normalization, the simulated flux can be re-weighted to a physical flux. After the neutrino generation, the particle is propagated through the Earth accounting for absorption, and finally forced to interact near or within the IceCube in-ice array, to produce events relevant for studies and analyses. This procedure ensures an accurate and computationally feasible simulation of neutrino-induced leptons. Complementary to neutrino simulations, atmospheric muons and neutrinos generated by CR interactions are simulated using the dedicated air shower simulation tool COsmic Ray SImulations for KAscade (CORSIKA) [87], discussed in detail in the following Section 5.1. The propagation of the finally produced leptons is then described in Section 5.2, which also explains the treatment of ice systematics. At the end of this chapter, the processing framework *IceProd* is mentioned in Section 5.4. It has been used to create the simulations utilized for this thesis.

5.1 Air Shower Simulation with CORSIKA 7

To simulate atmospheric muons originating from CR interactions in the upper atmosphere, IceCube relies on the extensive air shower simulation software CORSIKA [87]. CORSIKA is a well-established MC code widely used across leading cosmic-ray, gamma-ray, and neutrino experiments to model the development of particle cascades initiated by high-energy cosmic rays [40–42, 88–97]. These cascades produce a multitude of secondary particles, including muons, which may reach the IceCube detector and constitute a major background in neutrino analyses. Since this analysis targets to measure the atmospheric muon flux, the resulting air shower simulation serves as the signal.

The simulation begins by sampling primary CRs from predefined energy and mass spectra, representing a mixture of elements ranging from protons to iron nuclei. These primaries are injected at the top of the atmosphere and propagated through the air, where they undergo a series of hadronic and electromagnetic interactions. CORSIKA tracks all relevant secondary particles, including hadrons and leptons, accounting for decay processes, energy losses, and deflections in the geomagnetic field.

These particles are propagated to an so-called *Observation level*, which is the ice surface of the IceCube detector in this thesis.

The version employed in this analysis is CORSIKA 7.7500. For high-energy interactions, the `SIBYLL 2.3d` model [52] is used due to its ability to simulate charm production, which is essential for modeling the prompt muon component, as mentioned in Section 3.5.2. For low-energy interactions of GeV, the `Gheisha` [98] model is applied.

The atmospheric density profiles are modeled for each month from January to December [99]. To stop the shower simulation early, certain energy cuts can be applied. This is useful to avoid simulating particles down to energies that would not allow them to reach the detector. These cuts are named `ECUTS1`, `ECUTS2`, `ECUTS3`, and `ECUTS4`, and refer to the energy thresholds for the hadrons, muons, electrons and photons in GeV. When the particle energy drops below the corresponding threshold, the particle is not further simulated. All four cuts are set to 273 GeV (see Figure 5.1). Muons with lower energies are not able to traverse the ice from the surface vertical to the detector. Hence, they would never be detected by the in-ice array and thus they do not need to be simulated further. For this reason, the other cuts are also set to the same threshold.

In order to reject angular-dependent low-energy muons that are unlikely to reach the in-ice detector volume from the surface, a minimum energy threshold can be applied based on the expected range of muons in ice. The energy loss of high-energy muons as they propagate through matter is approximated by the following continuous loss equation

$$\frac{dE}{dx} = -a - bE\,, \tag{5.2}$$

where a accounts for ionization losses (nearly constant) and b represents radiative losses such as bremsstrahlung, pair production, and photonuclear interactions [20]. By integrating this equation from an initial energy E_0 to zero, the total range $R(E_0)$ that a muon can travel before stopping is obtained

$$R(E_0) = \frac{1}{b} \ln\left(1 + \frac{bE_0}{a}\right)\,. \tag{5.3}$$

This range is typically expressed in meters water equivalent (mwe), a standard unit for describing propagation through dense media [20]. The parameters are chosen as $a = 0.212\,\mathrm{GeV\,mwe^{-1}}$ and $b = 0.251 \times 10^{-3}\,\mathrm{mwe^{-1}}$, corresponding to standard energy-loss coefficients in ice [20]. To account for the actual density of the medium, these values are rescaled by a factor of 1.2, assuming a density of $\sim 0.92\,\mathrm{g/cm^3}$ for Antarctic ice [100].

In slanted geometries, the slant depth increases with zenith angle. Therefore, for a muon traveling at a zenith angle θ, the required minimum range becomes $R_{\mathrm{min}}/\cos(\theta)$, with $R_{\mathrm{min}} = 1450\,\mathrm{m}$ for the in-ice array. This cut ensures that only muons that would physically be able to reach the detector are kept. This means that towards the horizon, only higher-energy particles can reach the detector, as they must traverse a greater thickness of ice. This option is called `TrimShower`. The range as a function of the muon energy for several angles is illustrated in Figure 5.1. However, in comparison to the fixed `ECUTS` of 273 GeV, a muon at the horizon would be rejected by this filter, even though it has energies above the TeV scale. When calculating an effective area, as explained in Section 8.1.2, even though these particles would not make it to the detector, they need to be kept at the surface to correctly determine the muon flux at the surface. Hence, the `TrimShower` option is not used here.

Another common option to speed up the simulation is `Thinning`. This option combines a bunch of low-energy muons and treats them as one higher energetic muon. This setting is not used either to avoid any assumptions regarding the muon flux.

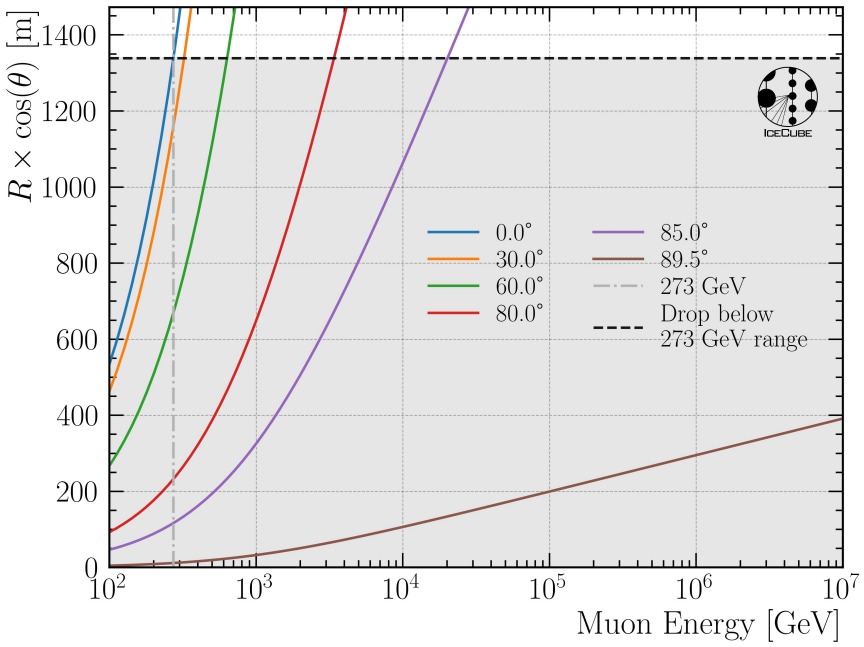


Figure 5.1: The muon range in ice normalized for different zenith angles as a function of the energy is presented, calculated with Equation (5.3).

For the direction, muons with zenith angles $\theta \in [0°, 90°]$ and azimuth angles $\phi \in [0°, 360°]$ are generated with the coordinate system introduced in Figure 4.1. At angles close to 90°, referred to as the horizon, the muons are not able to traverse all the dense medium required to reach the in-ice array. Thus, muons above the horizon do not need to be simulated. For the target volume, a cylindrical detector with radius $r = 700\,\mathrm{m}$ and height $h = 1400\,\mathrm{m}$ is set. A particle must reach this volume, otherwise it is discarded. The volume is chosen to be slightly larger than the detector, since the Cherenkov light is also detected, when a particle passes the detector at a distance of up to a few hundred meters, depending on the energy.

The injection of the primary particle is varied for the different primary energies. This is done to optimize the simulation towards the expectation, as presented in Figure 3.2. For this purpose, test datasets have been created with the same settings used for the final simulation. For the five primary particles proton, helium, nitrogen, aluminum, and iron, the ratio of the total event rate to the event rate of the primary particle is presented in Figures 5.2a, 5.2b, 5.2c, 5.2d and 5.2e. The four different primary weightings H3a, H4a, GST and GSF are used. The mean is calculated for five different energy ranges. These ranges are chosen because of technical reasons. The higher the primary particle energy, the more memory is needed. This follows from the fact that higher energetic primaries produce larger air showers. To avoid requesting too much or too little memory per requested simulation job, distinct energy sets are created. The mean and standard deviation are presented in Table 5.1. The normalization used for the final simulation is presented in Table 5.2. The software requires integers, thus all values are multiplied by 100. The normalization n_s for the sampling is then calculated by

$$n_\mathrm{s} = \frac{n_\mathrm{t}}{\sum_{i \in \mathrm{t}} n_\mathrm{i}} \,. \tag{5.4}$$

Due to the large variations towards very high energies, the primary composition is chosen to be equal. Additionally, even when the contribution per particle is below 1 % in the test simulation, the actual normalization is set to $n_\mathrm{t} = 1$. This ensures that all particle types are simulated over the entire energy range. These normalizations could have also been calculated by using the theory predictions as presented in Section 3.2. However, these refer to the muon flux at surface. To optimize the simulation

for this analysis, the composition in the event sample is determined after two cuts have been applied. One is the so-called *MuonFilter*, and the other cut selects events with an estimated muon bundle energy at surface of at least 200 TeV. These cuts are made to select high-energetic muons, and they are explained in detail in Chapter 7.

Table 5.1: Contribution of each primary particle to the total event rate. The mean and standard deviation are calculated for each distinct simulation energy range in GeV based on the four different weightings H3a, H4a, GST, and GSF.

Particle	10^4–10^6	10^6–10^8	10^8–10^9	10^9–10^{10}	10^{10}–10^{11}
P (mean)	0.80	0.09	0.17	0.48	0.30
P (std)	0.03	0.04	0.10	0.23	0.41
He (mean)	0.18	0.23	0.08	0.10	0.16
He (std)	0.03	0.02	0.08	0.11	0.27
N (mean)	0.02	0.19	0.15	0.08	0.08
N (std)	0.00	0.07	0.09	0.08	0.08
Al (mean)	0.00	0.14	0.17	0.07	0.10
Al (std)	0.00	0.01	0.08	0.08	0.16
He (mean)	0.00	0.35	0.42	0.27	0.36
He (std)	0.00	0.11	0.15	0.13	0.41

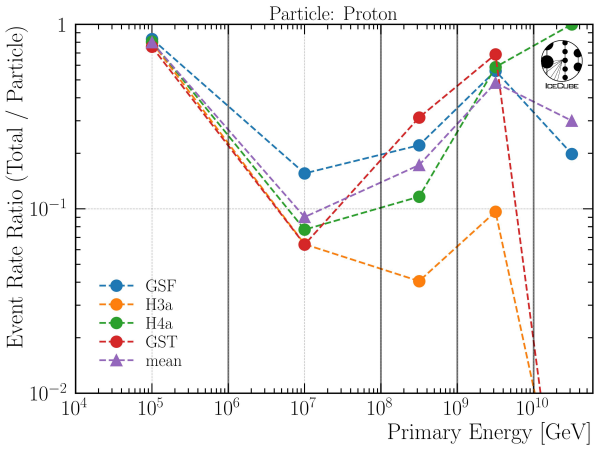
Table 5.2: The simulation is divided into five distinct primary energy E_{prim} ranges, referred to as dataset numbers. The spectral index of the sampling is γ_{s}, the normalizations for the primary particle types are n_{t} with $t \in [\text{P}, \text{He}, \text{N}, \text{Al}, \text{Fe}]$, and $N_{\text{events},\,s}$ is the total number of sampled events.

Dataset	E_{prim} / TeV	γ_{s}	n_{P}	n_{He}	n_{N}	n_{Al}	n_{Fe}	$N_{\text{events, s}}$
22774	10^1–10^3	2.7	80	18	2	1	1	$1.4 \cdot 10^{11}$
22775	10^3–10^5	2.5	9	23	19	14	35	$2.7 \cdot 10^{9}$
22776	10^5–10^6	2.5	17	8	15	17	42	$6.9 \cdot 10^{6}$
22777	10^6–10^7	2.5	5	1	1	1	3	$3.0 \cdot 10^{5}$
22778	10^7–10^8	2.5	1	1	1	1	1	$8.6 \cdot 10^{3}$

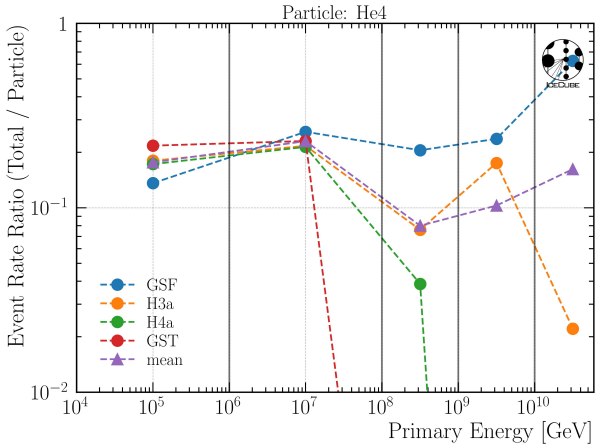
Detailed information about CORSIKA can be found in the user guide [101]. All settings of the simulation used in this thesis are stated in Section 5.4. The resulting particles at the surface serve as input for subsequent lepton propagation through the Antarctic ice, as described in Section 5.2.

5.1.1 Extended History

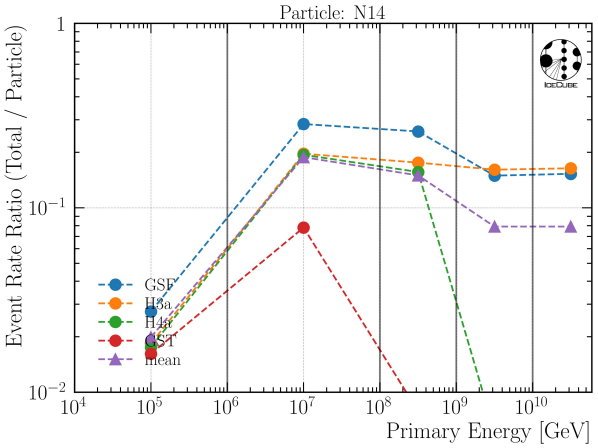
The standard output of CORSIKA 7 includes detailed information about particle interactions and decays within extensive air showers. By enabling the Extended History (`EHISTORY`) option, additional information about the genealogy of the particles are stored in the final simulation output. The data include identifiers for two ancestral generations termed parent and grandparent particles of muons and neutrinos. These ancestor particles are indicated by negative Particle IDs (PIDs) since they are not present at the observation level. Utilizing this information allows the classification of muons into conventional or prompt categories by identifying their true parent particles based on the Hadron Generation Counter (HGC), which increments with each interaction or decay. However, there are some exceptions [102]. These cases have been studied by Ludwig Neste in his masters thesis under my supervision [49]. The entire `EHISTORY` option has been studied in detail, and comparisons to theory



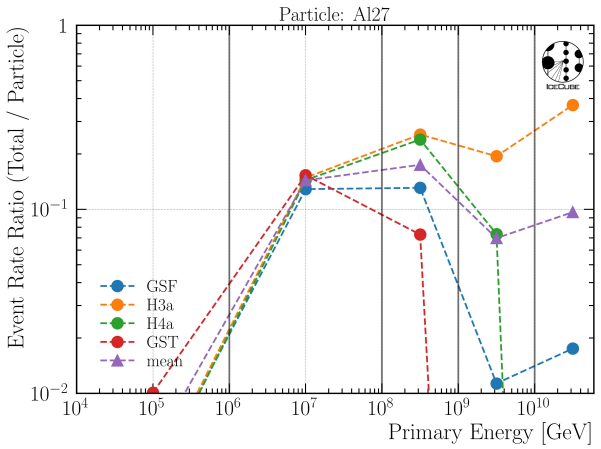
(a) Proton.



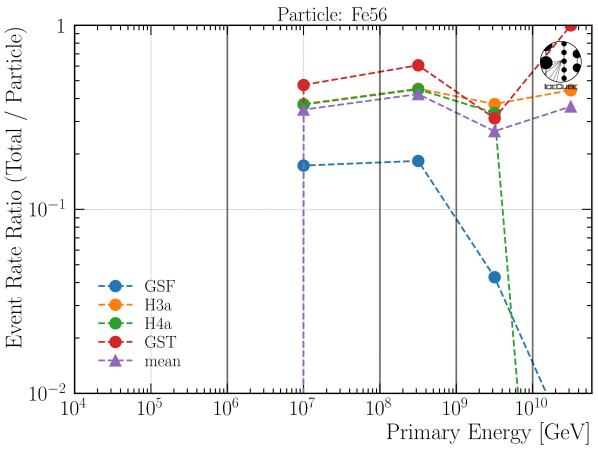
(b) Helium.



(c) Nitrogen.



(d) Aluminum.



(e) Iron.

Figure 5.2: Contribution of each particle type to the total CR flux based on the event rate at *Level 2* after applying the *MuonFilter* and an energy cut of 200 TeV on the reconstructed muon bundle energy at surface (see Chapter 7) for different primary weightings.

predictions haven been performed. Additionally, a python package PANdas And Multicore utils for corsikA 7 (PANAMA)[1] [103] has been developed to perform several tasks. The tool runs CORSIKA 7 on multiple cores, parses the extended history output correctly with all exceptions, stores the data in commonly used hdf5 files and performs the weighting for the four primary compositions mentioned before. The findings haven been successfully published in the *European Physics Journal C* [104]. The next section presents the comparison of `MCEq` predictions to CORSIKA.

5.1.2 Weight Simulation to Physical Flux

To use the simulated events for analysis, they must be weighted to represent a physical cosmic-ray flux. As described in Section 5.1, the primary particles are sampled from a power-law distribution with a spectral index γ_s, which does not necessarily match the physical spectrum of cosmic rays. To correct for this bias and to normalize the simulation to a specific cosmic-ray flux model, a weighting procedure is applied.

This analysis utilizes the `simweights` package [105], a Python library designed for IceCube simulations. `Simweights` calculates the weight w_i for each event i based on the primary particle properties (energy, type, zenith angle) and the generation parameters stored in the simulation files (such as the spectral index, energy range, and generation number). The weight represents the rate of such events expected from a specific cosmic-ray flux model $\Phi(E, t)$:

$$w_i = \frac{\Phi(E_i, t_i)}{\frac{dN}{dE\, d\Omega\, dA\, dt}(E_i, t_i)}\,, \tag{5.5}$$

where the denominator represents the generation probability density function. Since the generation is performed over a total number of generated events N_{gen}, a generation area A_{gen}, and a solid angle Ω_{gen}, the weight can be expressed as:

$$w_i = \frac{\Phi(E_i, t_i)}{P_{\text{gen}}(E_i, t_i) \cdot N_{\text{gen}}}\,. \tag{5.6}$$

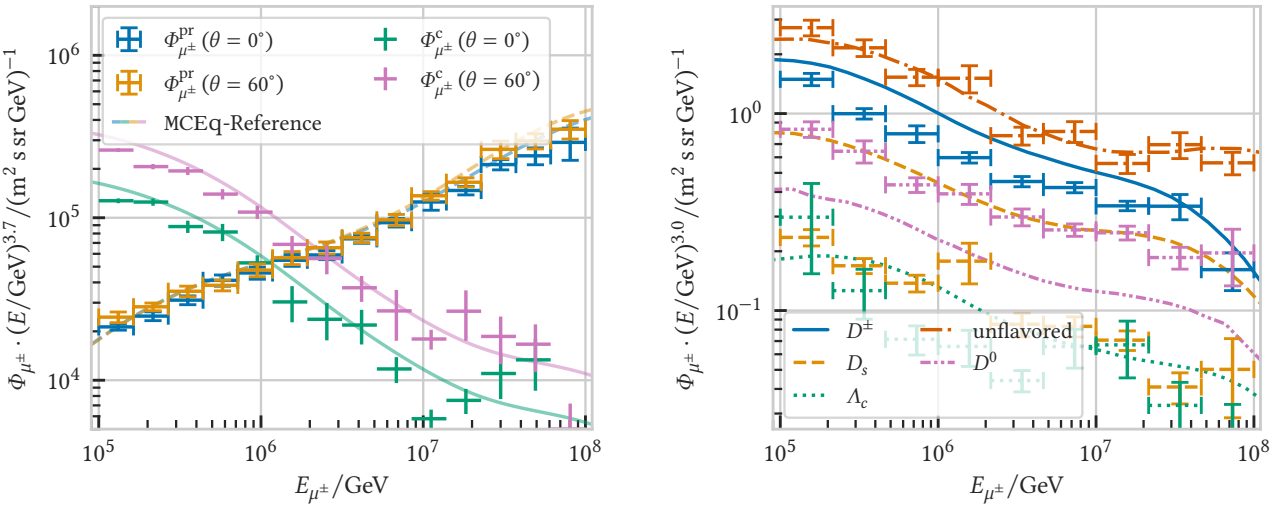
Here, E_i and t_i denote the primary energy and particle type of event i, respectively. The function Φ represents the physical differential flux of cosmic-ray primaries in units of ($\text{GeV}^{-1}\,\text{s}^{-1}\,\text{sr}^{-1}\,\text{m}^{-2}$), while P_{gen} is the probability density function used to generate the event based on the configured simulation settings (spectral index, energy range). Finally, N_{gen} is the total number of generated events in the simulation set. This framework allows utilizing the same simulation dataset to be re-weighted to different cosmic-ray models without running new simulations. For this analysis, the primary models described in Section 3.2 are used.

5.1.3 Comparison to MCEq Solutions

After implementing the `EHISTORY` output, taking care of exceptions and tagging muons as prompt or conventional depending on their ancestor particles, the entire procedure needs to be validated. For this, a comparison to the numerical solutions of the cascade equations, solved by the tool `MCEq`, is performed. The inclusive prompt and conventional muon flux for two different zenith angles of 0° and 60° is presented in Figure 5.3a. The results agree within the uncertainties for the presented energy range from $1 \cdot 10^5$ GeV to $1 \cdot 10^8$ GeV. The sub groups of the prompt component, $D^{\pm}$, D_s, D^0, Λ_c, and unflavored are compared for a zenith angle of 60° in Figure 5.3b. The unflavored CORSIKA simulation agrees well with `MCEq`. However, it appears that D^0 and D_s are swapped. The `MCEq` prediction is taken from Ref. [53], and the authors have confirmed that the lines in their paper are wrong. In

[1] https://github.com/The-Ludwig/PANAMA, `pip install corsika-panama`

Ref. [104], the production probabilities are compared for these two particles, which also indicate a wrong assignment in the MCEq paper. Hence, D^0 also agrees well. However, in CORSIKA, fewer $D^{\pm}$ and D_s particles are produced. One difference in this comparison is the underlying HIM. In MCEq, SIBYLL 2.3c [106] is used, while in CORSIKA, the newer version SIBYLL 2.3d [52] is used. It should be noted that neither MCEq nor CORSIKA predicts the correct muon flux at surface. These are two different ways to solve the same problem. Measurements are needed to confirm which prediction is more accurate. Since there is an overall agreement between the two methods, the implemented tagging of prompt and conventional muons in CORSIKA is validated. However, the goal of this thesis is to measure the inclusive atmospheric muon flux at the surface. Hence, a data-based comparison can be performed after the measurement is completed. It should be noted that for the figures below a separate CORSIKA simulation with PANAMA has been performed. Details can be found in Ref. [104].



(a) Prompt and conventional muon flux is compared for two different zenith angles of 0° and 60°.

(b) Sub groups of prompt muons are compared for a zenith angle of 60°. D^0 and D_s are swapped in MCEq, as discussed in Ref. [104].

Figure 5.3: A comparison of the muon flux at surface simulated with CORSIKA and calculated with MCEq is presented. PANAMA is used to run CORSIKA simulations. H3a is used for weighting. Figures are taken from Ref. [104].

5.2 Lepton Propagation Through the Detector

After generating muons at the Earth's surface using CORSIKA, the propagation of these leptons through the ice to the IceCube detector must be accurately modeled. This propagation accounts for energy loss processes such as ionization, bremsstrahlung, pair production, and photonuclear interactions, which are explained in Section 4.2.3. This task is handled by the software package PRopagator with Optimal Precision and Optimized Speed for All Leptons (PROPOSAL) [107, 108], a MC tool specifically developed for high-energy lepton propagation. PROPOSAL efficiently simulates stochastic energy losses and continuous energy losses experienced by muons traversing dense media. It also accounts for the deflection due to multiple scattering and the stochastic interactions [109]. The tool also simulates the propagation of electrons, muons and taus.

Following lepton propagation, the photons generated by Cherenkov radiation must be accurately simulated as they propagate through the Antarctic ice towards the DOMs. This photon propagation is performed using the software package clsim (Cherenkov Light Simulation Module) [110]. Clsim models the generation and scattering of Cherenkov photons, taking into account the optical properties of the ice, including absorption and scattering coefficients that vary with depth. This detailed modeling

ensures realistic simulation of the arrival time distribution and intensity of photons at the IceCube sensors, which is essential for accurate reconstruction of particle events. The variations of the ice are treated as systematics, as explained in Section 5.3.

After the light, i.e. signal, is detected, IceCube's trigger algorithms search for clusters of hits from High-Level Coincidence (HLC) optical modules in space and time, distinguishing genuine particle-induced events from random detector noise. The primary trigger is the Simple Multiplicity Trigger (SMT), requiring a certain number of coincident hits within a predefined time window. Additional specialized triggers, such as the Volume Trigger and String Trigger, impose spatial constraints to efficiently identify localized or vertically-oriented events, respectively. Once trigger conditions are satisfied, readout windows are defined around the trigger period, capturing relevant photon hits from the detector for subsequent event reconstruction. More detailed information is provided in Ref. [1].

After data acquisition via the trigger system, IceCube employs subsequent data processing stages known as *Level 1* and *Level 2*. The *Level 1* processing primarily involves data cleaning, removing detector artifacts, instrumental noise, and clearly unrelated hits, significantly reducing the data volume. Following this, *Level 2* processing applies initial event reconstructions and basic filtering algorithms designed to select high-quality events for further physics analysis. These refined datasets form the input for advanced event reconstruction techniques discussed in Section [1, 81, 111].

5.3 Ice Systematics – SnowStorm

Understanding and accurately modeling the optical properties of the Antarctic ice surrounding IceCube is essential for precise reconstruction and interpretation of the data. The ice can be divided into bulk ice and hole ice. Bulk ice is between the strings, it is untouched during the drilling operation. Hole ice is melted during the deployment, and then refrozen. This has an impact on the ice structure and thus on its optical parameters [58, 100, 112]. Variations in the ice's optical characteristics influence the detection and reconstruction of Cherenkov photons, which in turn affects the measured properties of particle interactions. To systematically address and quantify these uncertainties, IceCube employs a dedicated simulation framework known as *SnowStorm* [113], implemented in IceCube's software framework IceTray [82]. SnowStorm allows the systematic variations of ice-related parameters such as DOM efficiency, absorption, scattering, and hole ice parameters by generating ensembles with different parameter combinations. This enables to model the impact on the detector's performance, reconstruction and the subsequent physics results.

5.3.1 DOM Efficiency

The DOMs in IceCube detect Cherenkov photons produced by relativistic charged particles traversing the ice. DOM efficiency refers to the effective sensitivity of these optical sensors, accounting for variations in photon detection efficiency due to factors such as photomultiplier tube (PMT) quantum efficiency, glass transparency, and local optical coupling conditions. Variations in DOM efficiency directly influence the number of photons detected and thus affect event reconstruction quality and energy estimations. Typically, DOM efficiencies are varied within physically motivated ranges of $\pm 10\,\%$ from nominal values, and the effects of these variations are studied extensively [1, 114].

5.3.2 Absorption and Scattering

The propagation of Cherenkov photons through the ice is impacted by two optical properties: absorption and scattering. Absorption refers to the loss of photons due to their conversion into heat or fluorescence within the ice, effectively reducing photon counts at the detectors. Scattering, on the other hand,

redirects photons from their initial trajectory without significant loss in photon numbers, altering the photon arrival time distribution and directional information.

The ice around IceCube displays depth-dependent variations in both absorption and scattering coefficients due to changes in ice density, crystal structure, and impurities. SnowStorm accounts for these depth-dependent variations through parameterized ice models, derived from extensive calibration using artificial light sources and naturally occurring events. In SnowStorm simulations, absorption and scattering coefficients are systematically varied within physically plausible bounds, capturing uncertainties arising from ice model parameterizations. Accurate modeling of absorption primarily influences the overall photon count, affecting energy estimations, while precise scattering modeling is critical for reconstructing event directionality and timing [115, 116].

5.3.3 Hole Ice Parameters (p0 and p1)

The installation process of IceCube involves drilling deep holes into the ice, into which the DOM strings are deployed. The ice refreezes after deployment, creating a distinct region known as the hole ice, characterized by optical properties different from those of bulk ice. Hole ice significantly impacts the local photon detection behavior of DOMs due to altered scattering and absorption properties in the immediate vicinity of the optical sensors.

To accurately model the photon propagation within this region, SnowStorm employs two empirical parameters, known as p0 and p1. The parameter p0 characterizes the effective scattering length within the hole ice, influencing how frequently photons scatter within the refrozen column of ice around each DOM. Lower values of p0 indicate higher scattering, causing photons to diffuse more significantly before detection, thus broadening the photon arrival time distribution. Conversely, higher p0 values imply reduced scattering, preserving photon directionality.

The parameter p1 relates to the anisotropy of the photon scattering within the hole ice. Anisotropic scattering affects the angular distribution of photons around the DOMs, influencing their detection probability depending on the direction of incoming photons. Variations in p1 systematically affect the sensitivity and directional response of individual DOMs. Both p0 and p1 parameters are systematically varied in SnowStorm simulations to quantify their impact on event reconstructions and detector performance, with typical variations based on empirical calibration data [58, 117].

5.4 Final Simulation – Processing Framework IceProd

Large-scale MC production for IceCube is coordinated through the *IceProd* framework [118][2]. IceProd provides a unified layer on top of heterogeneous computing resources and is designed to run large simulation and reprocessing in a reproducible and traceable way. It organizes production into datasets consisting of tasks and modules, each specified through a JSON configuration that records executable code, input and output data, software environments, and resource requirements. Because these configurations are stored centrally and versioned, any dataset can be reproduced at a later time, which is essential for validating simulation results and ensuring long-term analysis consistency.

A key feature of IceProd is its ability to distribute work across multiple computing sites. It interfaces primarily with HTCondor-based clusters but can incorporate additional resources through the same framework, enabling the collaboration to utilize computing centers all over the world. IceProd monitors task states, handles transient errors through automatic retries, and manages data movement between sites when necessary. Together, these capabilities allow IceCube to sustain multi-site simulation

[2] `https://iceprod.icecube.aq`

activities consisting of millions of tasks while maintaining full traceability and reproducibility of all generated datasets.

For the simulation used in this analysis, several IceProd datasets have been produced, each with a unique dataset number. The datasets used are 22774, 22775, 22776, 22777 and 22778. Each dataset corresponds to a specific primary energy range, as detailed in Table 5.2. As expained in Section 5.1, these ranges are chosen to optimize memory usage during simulation. Also, different spectral indices for the sampling are applied to enhance simulation efficiency across the energy spectrum. Later on, also an additional dataset 22875 has be simulated for primary energy ranges from $600\,\mathrm{GeV} - 10\,\mathrm{TeV}$ to test whether some of these lower energetic primaries contribute to the final event sample. However, it turned out that none of these events pass the selection cuts, as explained in Chapter 7. Thus, this dataset is not used in the final analysis. For the EAS simulation with CORSIKA version 7.7500, the HIMs `SIBYLL 2.3d` [52] and `Gheisha` [98] are used for high- and low-energy interactions, respectively. The energy cuts `ECUTS1`, `ECUTS2`, `ECUTS3` and `ECUTS4` are all set to 273 GeV. The target volume is a cylinder with a radius of 700 m and a height of 1400 m. The primary particles proton, helium, nitrogen, aluminum and iron are simulated with the normalizations as presented in Table 5.2. After the air shower simulation, the leptons are propagated through the ice using PROPOSAL, followed by photon propagation with clsim. Finally, the IceCube detector response is simulated, including triggering and data acquisition. As explained above, the SnowStorm framework is employed to systematically vary ice properties, ensuring that the simulation accurately reflects the uncertainties associated with the Antarctic ice environment. For the IceCube software framework, IceTray 1.11.0-rc1 is used. Primaries are injected with zenith angles $\theta \in [0°, 90°]$ and azimuth angles $\phi \in [0°, 360°]$. To avoid issues with correct weighting, no coincident primaries are injected. In Section 8.2.8 it has been shown that this does not impact the final result. To mimic seasonal variations, the simulation is performed for 12 different atmospheres corresponding to each month of the year. The total number of simulated events per dataset is presented in Table 5.2.

5.5 Summary: Simulation

This chapter describes the simulation chain used to model atmospheric muon events for IceCube. EAS are generated with CORSIKA 7.7500 using `SIBYLL,2.3d` for high-energy interactions, together with monthly atmospheric profiles. Low-energy particles below 273 GeV are discarded, and options that could bias the spectrum, such as `TrimShower` or thinning, are avoided. The surviving secondaries at the ice surface are passed to the in-ice simulation, where leptons are propagated with PROPOSAL, Cherenkov light is simulated with `clsim`, and detector trigger and filtering (*Level 1/Level 2*) produce analysis-level events. The simulated muon sample is separated into prompt and conventional components using the extended CORSIKA history information. As an external cross-check, inclusive flux predictions from `MCEq` are compared and the results have been validated. This has been published in Ref. [104].

Detector-related systematics are modeled with the SnowStorm framework by varying DOM efficiency, bulk-ice absorption and scattering, and hole-ice parameters within physically motivated ranges. All large-scale production is performed with IceProd, which ensures reproducible, versioned configurations and distributes jobs across multiple sites. For this analysis, five production datasets (22774-22778), covering the primary-energy ranges from $10^4\,\mathrm{GeV} - 10^{11}\,\mathrm{GeV}$, are generated, and an additional low-energy test sample has been validated to have no impact after selection. Studies have been performed to optimize the primary particle normalizations for efficient simulation of high-energy muons relevant to this analysis. These simulations of atmospheric muons with quantified uncertainties are forming the basis for the surface muon flux measurement presented in this thesis. The information about the muon parent particles has never been available before on large scale simulations for IceCube. This allows for the first time to study the prompt muon component in detail. Next, the reconstruction of muon events is described in Chapter 6.

6 Reconstruction

IceCube detects relativistic charged particles indirectly by recording the Cherenkov photons emitted as they traverse the glacial ice. The detector consists of an array of DOMs embedded in the ice. Each event is thus a sparse, noisy space-time pattern of photon charges and arrival times distributed over the detector. The task of reconstruction is to invert this pattern into high-level physics quantities like the event direction, energy, and a topology label indicating whether the event is a muon track or a cascade. These reconstructions are needed to select a clean muon sample, and to unfold the atmospheric muon flux. In this chapter, traditional reconstruction methods in IceCube are introduced in Section 6.1, and Machine Learning (ML) based reconstructions are explained in Section 6.2.

6.1 Traditional Reconstructions

Traditional reconstruction approaches in IceCube rely either on maximum likelihood estimation (MLE), or on calibration-based observables that summarize the light yield. MLE is an established statistical technique used to estimate parameters of a probabilistic model by maximizing the likelihood function. This likelihood function quantifies how probable it is to observe the measured data given specific parameter values. The attractiveness of MLE lies in its theoretical properties, especially under large-sample conditions: consistency, efficiency, and asymptotic normality. Consistency ensures that as the sample size grows, the estimated parameters converge toward their true values. Efficiency guarantees that MLE achieves the lowest possible variance among unbiased estimators for large samples. Asymptotic normality implies that, with increasing sample size, the distribution of the estimated parameters approaches a normal distribution [119].

MLE-based reconstruction in IceCube involves defining a parametric statistical model characterized by unknown parameters, which are then inferred from observed detector data. The event properties of interest, such as energy deposition, interaction vertex, event direction, and particle type, are determined by finding parameter values that maximize the likelihood of observing the detected photon patterns under a given event hypothesis. While MLE is optimal when the underlying model accurately represents the data and the sample size is sufficiently large, it can introduce biases and inaccuracies in parameter estimation if the model assumptions are incorrect or the dataset is limited.

The central challenge of applying MLE in IceCube is accurately modeling the probability density functionss (PDFs) that link hypothesized event parameters to the expected photon signals observed by the detector. Several methodologies exist to construct these PDFs. Analytical approaches provide explicit mathematical descriptions [120], while simulation-based methods involve computationally intensive photon propagation simulations [121–124].

Direct, simulation-based modeling minimizes approximation bias, but it only realizes this advantage when MC statistics are sufficient to make simulation-induced fluctuations subdominant to the data. Due to computational limitations, IceCube primarily employs approximations based on pre-computed MC tables. These tables contain expected photon counts generated from extensive simulations over a grid of potential event scenarios. The discrete simulation results are subsequently interpolated and smoothed using spline functions, which yield continuous, differentiable approximations of photon distributions. This significantly reduces computation time but requires approximations to limit table dimensionality.

Complementary to likelihood fits using individual photons, IceCube employs robust energy observables derived from the underlying physics of light production. Energy estimators map statistics to an energy via a calibration derived from simulations. While not based on a likelihood itself, the method is firmly driven by physics, fast to evaluate, and well suited for tracks traversing the detector [125].

In the following, the standard energy estimator *TruncatedEnergy* for muon tracks is explained, followed by the standard muon direction MLE reconstruction *SplineMPE*.

6.1.1 Energy Reconstruction with TruncatedEnergy

In the energy reconstruction of high-energy muons in IceCube, a key observable is the energy loss per unit length, $\mathrm{d}E/\mathrm{d}x$, as it correlates with the muon's energy. However, due to the stochastic nature of muon energy losses at high energies, especially from processes like bremsstrahlung, pair production, and photonuclear interactions, the raw $\mathrm{d}E/\mathrm{d}x$ profile contains large fluctuations and outliers. These rare but intense energy losses can bias conventional estimators, making the relationship between the observed $\mathrm{d}E/\mathrm{d}x$ and the muon energy highly non-linear and difficult to model.

To address this, a method known as the *truncated mean* of $\mathrm{d}E/\mathrm{d}x$ was introduced [125]. In this approach, the muon track is segmented into multiple track sections (or bins), and for each segment, the local $\mathrm{d}E/\mathrm{d}x$ is calculated based on the detected charge. Instead of averaging over all segments, the highest-energy loss segments — typically those in the upper 40 % percentile – are excluded. The remaining segments are then averaged to compute the *truncated mean*:

$$\langle \mathrm{d}E/\mathrm{d}x \rangle_{\mathrm{trunc}} = \frac{1}{N_{\mathrm{trunc}}} \sum_{i=1}^{N_{\mathrm{trunc}}} \left(\frac{E_i}{L_i} \right), \tag{6.1}$$

where E_i is the reconstructed energy deposition and L_i the length of segment i, and N_{trunc} is the number of remaining segments after truncation.

This method suppresses the impact of large stochastic energy losses and leads to a more stable and approximately linear correlation between $\langle \mathrm{d}E/\mathrm{d}x \rangle_{\mathrm{trunc}}$ and the true muon energy, particularly for through-going tracks. A functional fit or regression model can then be applied to this truncated observable to infer the muon energy based on simulations. The resulting energy proxy is denoted as *TruncatedEnergy*.

The energy loss can be estimated in two different ways: either by calculating it at the position of each hit DOM (referred to as the DOM method), or by projecting the hit DOMs onto the reconstructed muon track and calculating the energy loss along predefined segments of the track (referred to as the BIN method). Both approaches produce comparable outcomes [126].

The truncated mean approach significantly improves energy resolution, especially at energies above several TeV where stochastic losses become dominant. Moreover, the method is computationally efficient and robust to noise and detector artifacts, making it suitable for real-time applications and large-scale MC processing in IceCube [125].

6.1.2 Directional Reconstruction with SplineMPE

The SplineMPE algorithm is a widely used directional reconstruction method in IceCube, designed to estimate the trajectory of muon tracks by maximizing a likelihood function based on photon arrival times. It operates under the assumption of an infinitely long, minimum-ionizing muon traversing the detector, which simplifies the event hypothesis and makes the algorithm both computationally efficient and robust.

SplineMPE relies on a multi-photo-electron (MPE) likelihood that incorporates both the arrival time of the first detected photon at each DOM and the total number of photo-electrons recorded. For a given track hypothesis $\vec{\theta}$, the likelihood function evaluates the probability of observing the measured time residuals $t_{\text{res},i}$ at each DOM, defined as the difference between the measured and expected photon arrival times. The MPE likelihood is expressed as

$$\mathcal{L}_{\text{MPE}}(\vec{x} \mid \vec{\theta}) = \prod_{i}^{\text{1st hits}} n_i \cdot p(t_{\text{res},i} \mid \vec{\theta}) \cdot \left(1 - P(t_{\text{res},i} \mid \vec{\theta})\right)^{n_i - 1} , \tag{6.2}$$

$$P(t_{\text{res},i} \mid \vec{\theta}) = \int_{-\infty}^{t_{\text{res}}} p(t \mid \vec{\theta})\, \mathrm{d}t\,. \tag{6.3}$$

where n_i is the number of detected photons at DOM i, $p(t \mid \vec{\theta})$ is the PDF for the expected photon arrival time, and $P(t \mid \vec{\theta})$ is its corresponding cumulative distribution function.

The expected timing distribution $p(t \mid \vec{\theta})$ is obtained from precomputed spline tables, which are generated using MC simulations of photon propagation in the glacial ice. These splines account for the complex optical properties of the ice, including scattering and absorption, and allow for smooth interpolation across different geometries and positions.

By incorporating both the timing of the first detected photon and the number of photo-electrons, SplineMPE efficiently constrains the geometry of the track. The algorithm yields high-precision directional reconstructions and serves as a core component in many IceCube event selection pipelines, particularly for identifying through-going muon tracks [127, 128].

6.2 Machine Learning Reconstructions

Disclaimer

The reconstructions performed within the scope of this thesis are evaluated for several levels of the data selection stage. In total, there are 4 analysis specific levels: *Level 3* (muon filter), *Level 4* (energy cut), *Level 5* (data-MC quality cuts) and *Final Level* (leadingness cut). These levels are defined in Chapter 7. Since the difference between the weightings is small, only H3a weighting is used in this chapter to avoid overload.

ML has emerged as a powerful methodology in high-energy astrophysics and particle physics due to its ability to effectively handle large datasets and uncover complex, nonlinear relationships within them. This has not only led to the discovery of neutrinos emerging from the galactic plane of our own Galaxy, the Milky Way [61], but also to the Nobel Prize in Physics being awarded jointly to John J. Hopfield and Geoffrey Hinton "for foundational discoveries and inventions that enable machine learning with artificial neural networks" [129].

Traditional reconstruction methods in detectors like IceCube typically rely on detailed physical models and likelihood-based optimization approaches, as mentioned above. While these classical techniques are robust, they often become computationally expensive or less accurate when confronted with highly sparse or noisy data. In addition, the MLE-based reconstructions above have been optimized for neutrino-induced muons. Thus, single muons within the detector [127, 128]. However, in this thesis, muon bundles, i.e. multiple muons within the detector, are utilized as signal to measure the atmospheric muon flux. Hence, the underlying hypotheses for the MLE reconstructions are not accurate. ML offers an alternative, where algorithms extract meaningful patterns directly from the simulated training data, defined as needed [130].

At its core, ML encompasses algorithms that improve their performance on a specific task through experience, explicitly by learning from examples. In supervised ML, a model is provided with labeled input-output pairs, learning to predict the desired output for new, previously unseen inputs. In IceCube event reconstruction, supervised models are trained using simulated detector signals (inputs) with corresponding true particle parameters such as energy, direction, and detector geometries like entry points.

A fundamental and widely utilized ML architecture in reconstruction tasks is the fully connected neural network, also known as a multilayer perceptron (MLP) [131, 132]. An MLP consists of several layers of interconnected nodes (neurons), where each node receives weighted inputs from the previous layer, computes a nonlinear transformation via an activation function, and passes the resulting value to the next layer. A common activation function is the Rectified Linear Units (ReLU) function [133–135], defined as:

$$\mathrm{ReLU}(x) = \max(0, x)\,, \tag{6.4}$$

which introduces nonlinearity and helps the network model complex relationships.

Another frequently used activation function is the exponential linear unit (ELU) [135, 136], which improves gradient flow and allows for small negative outputs

$$\mathrm{ELU}(x) = \begin{cases} x & \text{if } x \geq 0 \\ \alpha(e^x - 1) & \text{if } x < 0 \end{cases}, \tag{6.5}$$

where α is a hyperparameter typically set to 1.

In some architectures, the absolute value (Abs) function [135, 137, 138] is also used as an activation:

$$\mathrm{Abs}(x) = |x|\,. \tag{6.6}$$

It is beneficial in tasks that require sign-invariant transformations or specific symmetry properties.

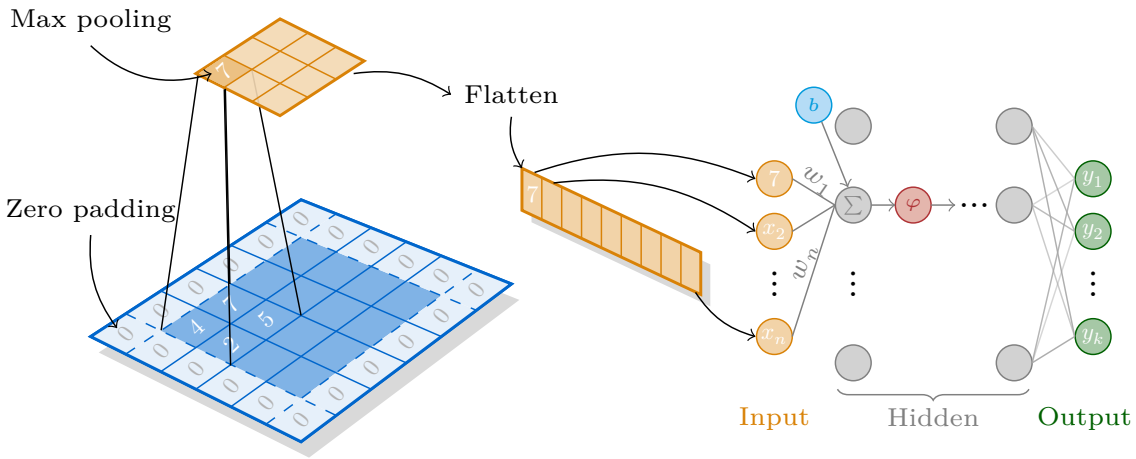


Figure 6.1: Schematic overview of a CNN architecture. A 2×2 kernel (orange) performs max pooling over the input space. Zero padding is illustrated but not applied in this example. The resulting feature map is flattened and passed to a MLP. In each hidden layer, the weighted inputs are summed, a bias term b is added, and the result is transformed by an activation function φ. This process is repeated through successive layers until the final output y_k is produced.

Training neural networks involves iteratively adjusting weights and biases to minimize a specified loss function, quantifying the difference between predicted outputs and true labels. A commonly used loss

function in reconstruction tasks is the mean squared error (MSE) [135, 138], mathematically defined as

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2\,, \tag{6.7}$$

where y_i are the true values, $\hat{y}_i$ the predicted values, and N the total number of samples. Optimization algorithms, such as stochastic gradient descent (SGD) or adaptive algorithms like Adam, efficiently minimize the loss by navigating the high-dimensional parameter space.

In this work, however, a Gaussian likelihood loss function is employed [135, 138]. Instead of predicting a single output value, the network estimates both the mean μ_i and the standard deviation σ_i for each target variable, assuming the target values are normally distributed. The corresponding negative log-likelihood loss is given by:

$$\mathcal{L}_{\text{Gaussian}} = \frac{1}{N}\sum_{i=1}^{N}\left[\log(\sigma_i^2) + \frac{(y_i - \mu_i)^2}{\sigma_i^2}\right]. \tag{6.8}$$

This formulation allows the model not only to predict the expected value of a quantity but also to estimate its associated uncertainty. As a result, the network is encouraged to report higher uncertainties in regions where predictions are more difficult, making this approach particularly well suited for scientific applications involving uncertainty-aware regression.

To mitigate overfitting, where a network performs well on training data but poorly on unseen data, techniques such as dropout are employed [139]. Dropout randomly deactivates a subset of neurons during each training iteration, effectively reducing co-dependencies between neurons and promoting generalization. Formally, dropout can be represented as

$$y = f\left(\sum_j w_j x_j m_j + b\right), \quad \text{with} \quad m_j \sim \text{Bernoulli}(p), \tag{6.9}$$

where m_j represents the dropout mask with probability p, set between 0 and 1.

CNNs represent a specialized class of neural networks efficient at processing structured, high-dimensional data such as images or spatially-arranged detector signals. CNNs identify local patterns and spatial hierarchies through convolutional layers, using small, learnable filters (kernels) to scan across input data. Each kernel generates a feature map by convolving with the input, capturing spatial correlations and local dependencies [133, 140, 141].

Following convolution, CNNs apply nonlinear activation functions to facilitate complex feature extraction. Subsequent pooling layers, commonly max or average pooling, reduce spatial dimensions of feature maps by summarizing local regions into single values. This captures prominent features, reduces network sensitivity to positional variations, and enhances computational efficiency. Stacking multiple convolutional and pooling layers progressively extracts abstract representations, from basic edges to high-level features [141].

The output from convolutional and pooling layers is flattened and passed into fully connected layers, performing final prediction tasks such as event reconstructions in IceCube analyses. The training of CNNs follows the same principles as fully connected networks, using gradient-based optimization algorithms to minimize a chosen loss function. Because CNNs naturally capture spatial structures and require fewer parameters due to weight sharing, they are especially effective for analyzing spatially correlated data, such as particle interactions in large detector arrays. A visualization of convolutions that are flattened and fed into a fully connected layer is presented in Figure 6.1.

6.2.1 DNN-reco Framework

The `dnn_reco`[1] [130] framework is a modular software toolkit specifically developed for training and applying deep learning models within IceCube. The framework `dnn_reco` systematically handles configurations, ensuring consistent training and model usage with a focus on reproducibility. Core components of `dnn_reco` include the *SetupManager*, managing configurations; *NNModel*, defining neural network architectures, loss functions, event weighting, and training routines; *DataTransformer*, responsible for normalizing data inputs; and *DataHandler*, managing data loading, labeling, filtering, and applying biasing strategies. The framework saves detailed metadata, configurations, and training parameters, providing straightforward model sharing and reproducibility. The number of layers, convolution kernel sizes, activation functions, and loss functions are configurable, allowing adjustments based on specific reconstruction or classification tasks. All settings are stored in one configuration file that includes all information mentioned above.

The framework supports any neural network architecture $f : \vec{X} \rightarrow \vec{y}$ that maps a DOM-based feature matrix $\vec{X}$ of constant length to the desired output quantities $\vec{y}$. In this thesis, CNNs are utilized.

The neural network architecture in `dnn_reco` is user-defined and specified through dynamically loaded modules. By default, IceCube data is segmented according to its geometry into three parts: the main IceCube array, and lower and upper DeepCore arrays. These segments are independently processed using three separate CNNs, as presented in Figure 6.2. The outputs of these CNNs are then flattened and combined into fully-connected sub-networks, which predict the quantities of interest along with associated uncertainties. Convolutions applied to the main IceCube array utilize hexagonal kernels, specifically matching the detector's geometric structure. This is sketched in Figure 6.3.

The entire framework is explained in detail in Ref. [79, 142]. The codebase, documentation, and usage examples are publicly available on GitHub[1] and Zenodo [130].

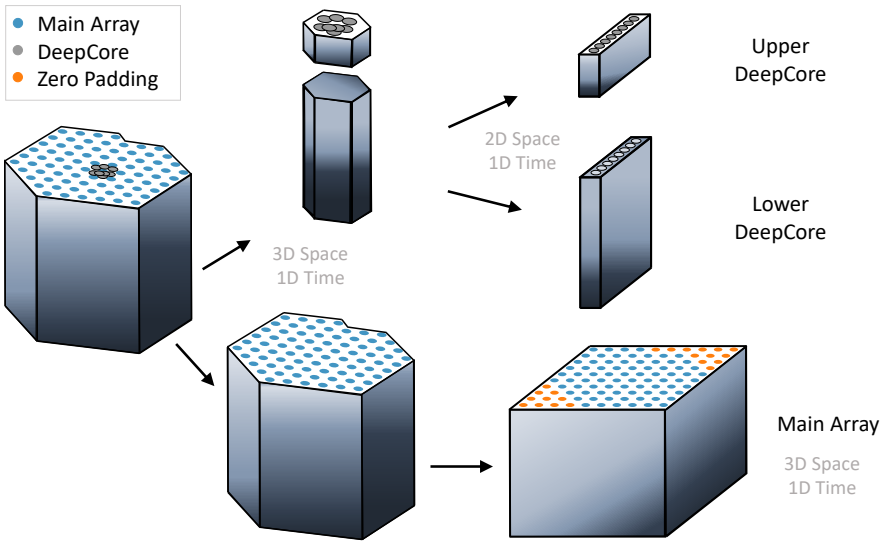


Figure 6.2: The IceCube main array and the DeepCore sub-array are processed separately due to their different geometries. To represent the hexagonal layout of the main array on a regular grid, the detector plane is transformed from axial to Cartesian coordinates. This is achieved by aligning the rows and inserting zeros at appropriate positions (shown as orange dots), resulting in a 10×10 grid in the $x - y$ plane. Figure taken from Ref. [79].

[1] https://github.com/icecube/dnn_reco

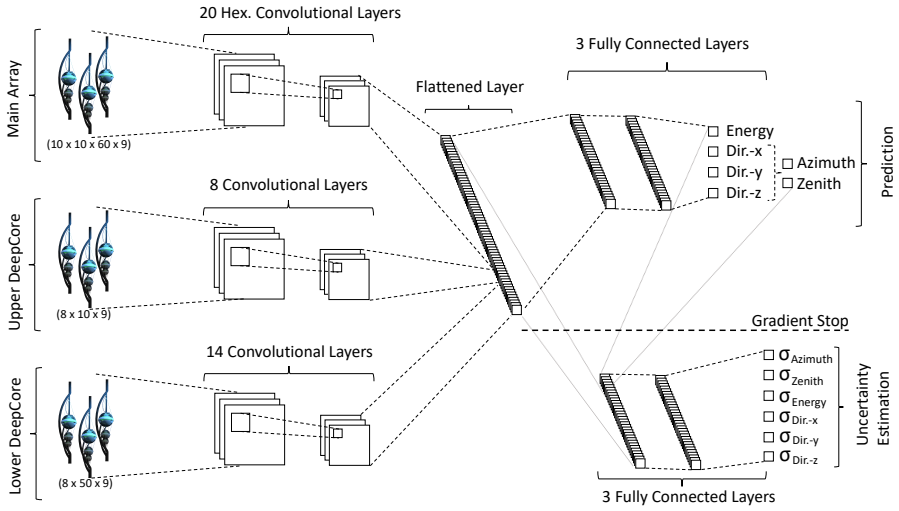


Figure 6.3: Schematic overview of the neural network architecture. Input data from the three detector sub-arrays are first processed independently through convolutional layers. The resulting feature maps are flattened and merged, then passed to two separate fully connected branches: one for the reconstruction task and one for estimating the associated uncertainty. The uncertainty branch additionally receives the reconstruction output as an input feature. Figure taken from Ref. [79].

6.2.2 Uncertainty Estimation

In addition to providing point estimates y_{pred} for physical quantities such as energy or direction, the `dnn_reco` framework is also capable of estimating the associated uncertainty σ_{pred} for each prediction.

The underlying assumption is that the residuals $\Delta y = y_{\text{pred}} - y_{\text{true}}$ between the predicted and true values follow a Gaussian distribution. Under this assumption, the network is trained using a negative log-likelihood loss function derived from the Gaussian distribution

$$\mathcal{L}_{\text{Gaussian}} = \ln(\sigma_{\text{pred}}) + \frac{1}{2}\left(\frac{y_{\text{true}} - y_{\text{pred}}}{\sigma_{\text{pred}}}\right)^2 . \tag{6.10}$$

This loss function is minimized during training, encouraging the network to not only make accurate predictions, but also to assign uncertainty estimates. The formulation penalizes both overconfident (too small uncertainty) and underconfident (too large uncertainty) predictions. If the network predicts a value with a very small uncertainty, but the residual Δy is large, the loss will increase sharply. On the other hand, large uncertainties will be penalized by the $\ln(\sigma_{\text{pred}})$ term.

The uncertainty estimation is applied individually for each predicted quantity, such as energy, zenith, or azimuth angle. In its current implementation, potential correlations between these quantities are ignored, and the Gaussian likelihood is applied separately for each output dimension.

This approach enables uncertainty-aware predictions, allowing downstream analyses to account for per-event prediction confidence. While the assumption of Gaussian-distributed residuals is an approximation, it generally holds to a reasonable degree in practice and serves as a valuable basis for including uncertainties in neural network-based reconstructions [79, 142].

6.2.3 Input Features

Each IceCube event consists of a collection of light pulses recorded by the DOMs distributed throughout the detector, as described in Section 4.3. These pulse series vary in length, both across DOMs within a

single event and between different events, due to differences in event topology, energy, and geometry. However, CNNs and fully connected neural networks require fixed-size inputs. To accommodate this constraint, the pulse information recorded at each DOM must be condensed into a fixed-length feature vector.

The design of the input features is task-specific. In this analysis, which focuses on reconstructing muons and muon bundles, the selected features are optimized to retain as much relevant physical information as possible while ensuring computational efficiency. Domain knowledge is embedded in the feature construction to make the network sensitive to properties such as the energy deposition, timing, and spatial development of the event. Additionally, the features are chosen to be robust with respect to known systematic uncertainties and simulation mismodeling [79, 142].

The default configuration used in most of the networks in this analysis includes the following nine input features per DOM

1. c_{total}: Total charge collected at the DOM.
2. $c_{100\,\text{ns}}$: Charge collected within 100 ns after the first pulse.
3. $c_{500\,\text{ns}}$: Charge collected within 500 ns after the first pulse.
4. t_{first}: Time of the first pulse, relative to a global time offset.
5. t_{last}: Time of the last pulse, relative to the same global offset.
6. $t_{20\,\%}$: Time at which 20 % of the total charge at the DOM has been collected.
7. $t_{50\,\%}$: Time at which 50 % of the total charge at the DOM has been collected.
8. t_{mean}: Mean time of all pulses at the DOM.
9. t_{std}: Standard deviation of the pulse times at the DOM.

A visualization of these nine features is presented in Figure 6.4. All time-related features are computed relative to a global offset, defined per event as the beginning of a 6000 ns time window that maximizes the contained charge. This normalization introduces a form of time-translation invariance, which is beneficial for training neural networks.

One of the networks used in this analysis employs a reduced feature set consisting of only the three most informative inputs: c_{total}, t_{first}, and t_{std}. These features are chosen to capture the essential temporal and energetic structure of the light emission at each DOM while allowing for faster feature extraction and more compact network architectures. The total charge c_{total} correlates with the local energy deposition and is particularly important for energy predictions. The arrival time of the first pulse t_{first} provides valuable geometric information, as the earliest photons are typically the least scattered and thus most informative about the position of the particle track. The standard deviation t_{std} encodes the temporal width of the pulse distribution, which relates to the distance and angular orientation of the light source with respect to the DOM [79, 142].

6.2.4 Training

Training the neural networks follows the standard supervised learning paradigm [135, 138]. For each task, a labeled data set of simulated events is constructed, where the inputs are the per-DOM features described above and the targets are the true physical quantities of interest, such as energies, directions, or track geometry parameters. The data are split into three disjoint subsets: a training set used to update the network weights, a validation set used to monitor performance during training and tune hyperparameters (for example learning rate, depth, or dropout fraction), and an independent test set used only for the final performance evaluation. During training, the optimizer iteratively adjusts

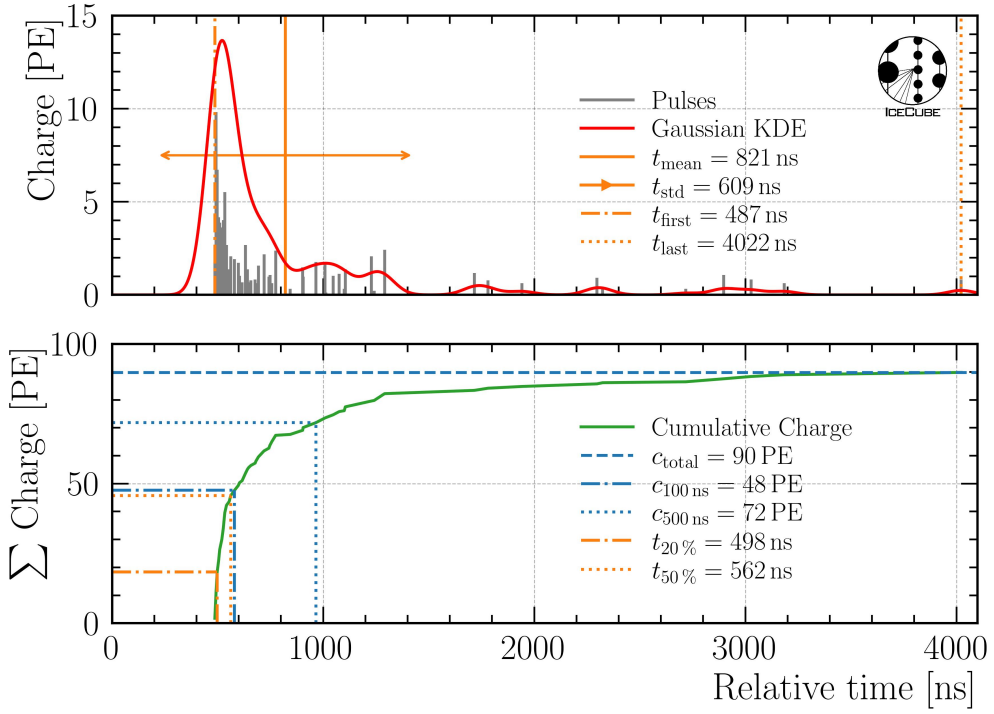


Figure 6.4: Collected charge (individually on top, cumulated in the bottom) as a function of the relative time. A Gaussian KDE is utilized to approximate the charge distribution. In total, nine features serve as inputs for a multivariate estimation. The figure is adapted from Ref. [79].

the network parameters to minimize the chosen loss function, here a Gaussian likelihood loss that allows the network to predict both central values and associated uncertainties. Early stopping and regularization techniques, such as dropout, are used to reduce overfitting and to ensure that the model generalizes well beyond the samples it has seen during optimization [79, 139, 142].

The training samples are based on several dedicated MC datasets that span primary energy ranges from $600\,\text{GeV} - 10^{11}\,\text{GeV}$. Two different ice models are used in the simulations, Spice3.2 and Spice3.2.1 [116] to account for uncertainties in the optical properties of the ice and to be robust against potential mismodeling. The datasets are summarized in Table 6.1. This combination ensures good coverage of the phase space relevant for this analysis, in particular for high-energy muon bundles that dominate the final event sample. To optimize the performance, the networks are trained separately for different data selection levels. The pre-cut networks is trained on *Level 3* events, while the energy, direction and track geometry networks are trained on *Level 4* events with a looser muon bundle energy cut of $E^{\text{bundle}}_{\mu,\,\text{surf}} > 200\,\text{TeV}$ instead of 500 TeV to be sensitive to potential outliers below to the cut value.

Table 6.1: Overview of the training datasets used for the different reconstructions. These are all official IceCube MC datasets. All set are generated with the HIM Sibyll2.3c. Details can be found on IceProd[1]. The last two columns show the number of events used for training at *Level 3* and *Level 4*, respectively.

Dataset	Primary Energy [GeV]	Ice Model	Events L3	Events L4
20904	$6 \cdot 10^2 - 1 \cdot 10^8$	Spice3.2	$1.28 \cdot 10^6$	$1.42 \cdot 10^7$
21962	$1 \cdot 10^9 - 1 \cdot 10^{11}$	Spice3.2.1	$7.00 \cdot 10^3$	$1.72 \cdot 10^4$
22020	$1 \cdot 10^6 - 1 \cdot 10^{10}$	Spice3.2.1		$5.33 \cdot 10^6$
22187	$2 \cdot 10^8 - 2 \cdot 10^{10}$	Spice3.2.1	$7.27 \cdot 10^4$	$1.58 \cdot 10^5$

The training data utilized on *Level 3* to train the pre-cut network are presented in Figure 6.5. Four different distributions are shown: the leading muon energy at surface and entry, the zenith angle distribution and the bundle energy at surface, which is the target variable of the energy reconstruction. Figure 6.5a presents that using only one dataset would not be sufficient to cover the entire energy range of interest. Set 20904 has a high statistics at low energies, but lacks events above approximately 10 PeV. For this, datasets 21962 and 22187 are used to cover the high-energy region, however, they provide lower statistics and no events below a bundle energy at surface of 100 TeV. The zenith distribution is roughly flat, as expected, and looses statistics towards $\cos(\theta) = 0$, which refers to the horizontal direction. The muons are not able to pass the Earth for nearly 90°, hence no events are in this region.

For the training on *Level 4*, a bias is applied to obtain a more balanced training sample. Due to the steeply falling energy spectrum of cosmic rays of about E^{-2} in the simulations, the majority of events are at low energies, while only a small fractions of events are at the highest energies. When training a network, this leads to a bias towards low-energy events, as the loss function is dominated by these abundant samples. To obtain a more balanced training sample, a one-dimensional bias function $p_{\text{keep}}(x)$ is applied to the bundle energy at surface,

$$x = \log_{10}(E^{\text{bundle}}_{\mu,\,\text{surf}}/\text{GeV})\,. \tag{6.11}$$

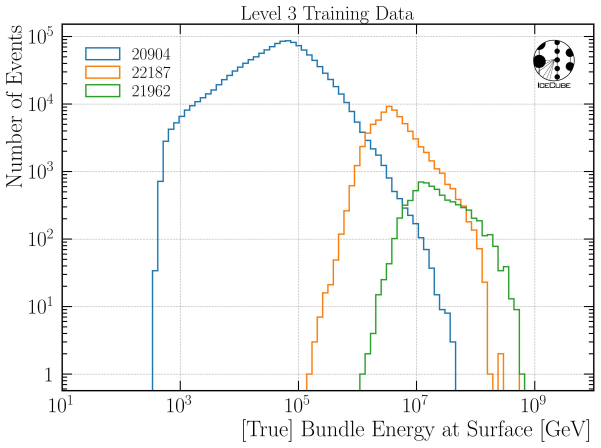
For x values above a chosen upper transition point x_{cut}, all events are kept, while for x values below a lower transition point $x^{\text{low}}_{\text{cut}}$ the keep probability is flattened to a quadratic function to avoid vanishing probabilities at the low-energy end. In the intermediate region, a power law with exponent β is used. The resulting bias function is

$$p_{\text{keep}}(x) = \begin{cases} 1 & \text{for } x \geq x_{\text{cut}}\,, \\ \left(\dfrac{x}{x_{\text{cut}}}\right)^{\beta} & \text{for } x^{\text{low}}_{\text{cut}} < x < x_{\text{cut}}\,, \\ \dfrac{((x^{\text{low}}_{\text{cut}}/x_{\text{cut}})^{\beta} - 1)}{{x^{\text{low}}_{\text{cut}}}^2}\, x^2 + 1 & \text{for } x \leq x^{\text{low}}_{\text{cut}}\,. \end{cases} \tag{6.12}$$

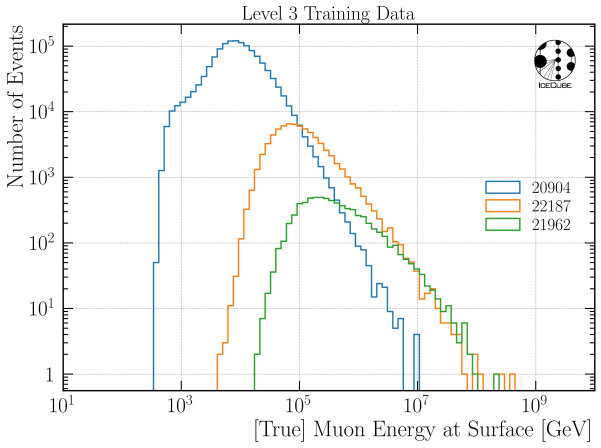
In this analysis, the parameters are set to $\beta = 8.0$, $x_{\text{cut}} = \log_{10}(2 \times 10^6)$ and $x^{\text{low}}_{\text{cut}} = \log_{10}(3 \times 10^5)$. This function is visualized in Figure A.1a.

The bias is realized by drawing, for each event, a uniform random number $u \sim \mathcal{U}(0, 1)$ and keeping the event if $u \leq p_{\text{keep}}(x)$. Events that pass this selection receive an additional weight factor $1/p_{\text{keep}}(x)$, such that the expected contribution of the biased sample remains unchanged while the effective training statistics at high energies are enhanced. This re-weighting scheme is presented in Figures A.1b and A.1c. For dataset 20904 at *Level 3*, this procedure reduces the number of events from $12.8 \cdot 10^5$ to $6.7 \cdot 10^5$, corresponding to a keep fraction of 0.53. The same bias is also applied to dataset 22020 and effectively suppresses the large number of low-energy events. The resulting distributions are still following a power law, but with a flatter spectrum. Since the final sample of this analysis also follows a power low, the training distribution does not need to be completely flat, but the biasing helps to improve the performance at high energies by reducing the low-energy events roughly by a factor of two. The final training distributions on *Level 4* are presented in Figure A.2.

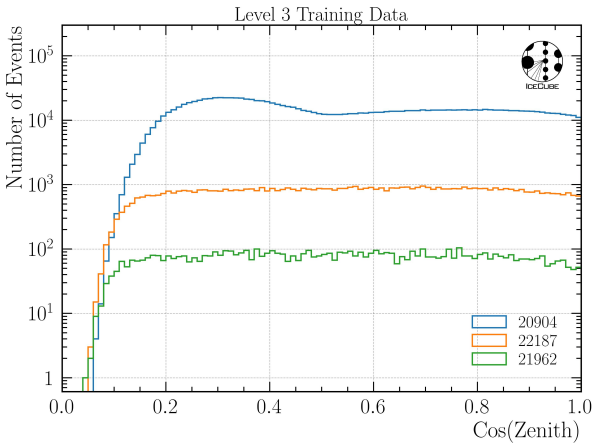
All in all, many reconstructions have been performed, and in principle, all reconstructed variables can be evaluated at all selection levels. However, to avoid an overload not all results are shown for all levels; some are only presented on *Level 4* because these reconstructions have been used on that level to apply quality cuts. The energy and zenith reconstructions and the angular resolution are presented for all levels because these are the potential proxy variables for the final analysis. In the end, only the energy reconstruction of the leading muon energy at the detector entry on *Finel Level* as presented in Figure 6.10 is used in the analysis.



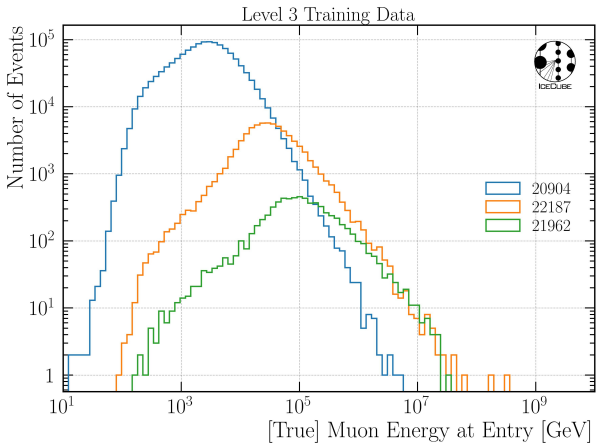
(a) Bundle energy at surface.



(b) Leading muon energy at surface.



(c) $\cos(\theta)$.



(d) Leading muon energy at entry.

Figure 6.5: These figures show the true distributions of the training data used for the different reconstructions on *Level 3*. Each plot presents the individual statistics of three different datasets used for training, as summarized in Table 6.1. Most of the statistics is coming from 20904, however, it is necessary to include the other datasets to cover the high-energy region.

Additionally, only the final networks used in this analysis are presented. However, first investigations have been done in two bachelors theses by Leander Flottau and Benjamin Brandt und my supervision. Different layers, number of nodes, max and average pooling, dropout, 3 and 9 input features have been studied and also with time window cleaning and without. The results of these studies are stated in Ref. [143] and Ref. [144] for energy and directional reconstructions, respectively. Afterwards, further hyperparameter optimizations have been performed to finally obtain the presented networks.

Creating the training data is done with the software framework `ic3_data`[2], and the processing of the jobs is performed with `ic3_processing`[3]. The actual training of the networks is then done with the `dnn_reco` framework[1] [130].

The final models used in this thesis are evaluated on the new MC samples simulated within the scope of this thesis (datasets 22774-22778) and on real data at different stages of the event selection, as discussed in Sections 6.2.7, 6.2.8 and 6.2.9 for the energy, direction and track geometry, respectively.

6.2.5 Pearson and Spearman Correlation Coefficients

To quantify the relationship between two variables, the two correlation coefficients *Pearson* [145] and *Spearman* [146] are widely applied. Both assess the strength and direction of association between two variables, but they differ in sensitivity and interpretation: the Pearson coefficient is typically used when a linear relationship is expected and the variables are approximately normally distributed, while the Spearman coefficient is used when the data are not normally distributed or when the relationship is monotonic but not necessarily linear.

Pearson Correlation Coefficient

The *Pearson* correlation coefficient ρ_{P} measures the linear relationship between two continuous variables. It is defined as the covariance of the two variables divided by the product of their standard deviations

$$\rho_{\mathrm{P}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}\,, \tag{6.13}$$

where x_i and y_i are the individual sample values, $\bar{x}$ and $\bar{y}$ are the sample means of the respective variables, and n is the number of data points. The coefficient ρ_{P} takes values in the range $[-1, 1]$, with $\rho_{\mathrm{P}} = 1$ indicating a perfect positive linear correlation, $\rho_{\mathrm{P}} = -1$ indicating a perfect negative linear correlation, and $\rho_{\mathrm{P}} = 0$ indicating no linear correlation.

Spearman Rank Correlation Coefficient

The *Spearman* correlation coefficient ρ_{S} assesses how well the relationship between two variables can be described by a monotonic function. Unlike the Pearson coefficient, Spearman's method operates on the ranks of the data rather than the raw values. This makes it more robust to outliers and suitable for non-linear but monotonic relationships.

It is computed as the Pearson correlation between the ranked variables. For datasets without tied ranks, it can be expressed as

$$\rho_{\mathrm{S}} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}\,, \tag{6.14}$$

[2] https://github.com/icecube/ic3-data
[3] https://github.com/mhuen/ic3-processing

where $d_i = \text{rank}(x_i) - \text{rank}(y_i)$ is the difference between the ranks of the i-th elements of the two variables, and n is the number of observations. As with the Pearson coefficient, $\rho_S \in [-1, 1]$, with the same interpretation regarding the strength and direction of association, but applied to ranked (ordinal) data.

6.2.6 Variable and Network Definitions and Notation

This subsection provides definitions and notations for all variables and terms used in this thesis. These variables are reconstructed using four different neural networks, as summarized in Table 6.2.

At surface The Earth's surface at a height of 2835 m above sea level, corresponding to the IceTop altitude, i.e. IceCube's surface [60].

At detector entry The entry of the detector. A convex hull is built around the in-ice strings and extended by 200 m in all directions. This extension accounts for light detected from tracks that pass a few meters outside the instrumented volume.

Bundle The collection of all muons produced in the EAS that reach the considered reference surface (surface or detector entry).

Leading muon The most energetic muon within the bundle at the respective reference surface (surface or detector entry).

Bundle energy at surface $E_{\mu,\text{surf}}^{\text{bundle}}$ Sum of the energies of all muons in the bundle at the IceTop surface altitude.

Leading muon energy at surface $E_{\mu,\text{surf}}^{\text{leading}}$ Energy of the leading muon at the IceTop surface altitude.

Bundle energy at entry E_{μ}^{bundle} Sum of the energies of all muons that cross the extended detector entry surface.

Leading muon energy at entry E_{μ}^{leading} Energy of the leading muon at the extended detector entry surface.

Zenith angle θ Angle between the track direction and the vertical downward axis, with $\theta = 0°$ for vertically downgoing tracks and $\theta = 90°$ at the horizon, see Figure 4.1.

Azimuth angle ϕ Angle of the track in the horizontal plane, see Figure 4.1.

Entry point $(x_{\text{entry}}, y_{\text{entry}}, z_{\text{entry}}, t_{\text{entry}})$ Space–time coordinates of the leading muon at detector entry, with spatial coordinates in the IceCube coordinate system and t_{entry} given relative to the event time window.

Center point $(x_{\text{center}}, y_{\text{center}}, z_{\text{center}}, t_{\text{center}})$ Space–time coordinates of the point of closest approach of the leading muon track to the detector center $(0, 0, 0)$.

Track length in detector s Distance traveled by the leading muon between the entry and exit points of the extended convex hull around the in-ice array.

Total track length s_{tot} Distance between the muon production vertex at the surface and the exit point of the extended detector volume.

Table 6.2: Overview of the networks and the reconstructed properties. For all variables, the network also predicts an associated uncertainty.

	`precut_surface_bundle_energy_3inputs_6ms_01`
Energy (pre-cut)	muon bundle energy at surface
`leading_bundle_surface_leading_bundle_energy_OC_inputs9_large_log_02`	
Energy	leading muon energy at the detector entry muon bundle energy at the detector entry
	`direction_9inputs_6ms_medium_02_03`
Direction	zenith angle of leading muon at entry azimuth angle of leading muon at entry direction x, y, z of leading muon at entry
	`track_geometry_9inputs_6ms_medium_01`
Track Geometry	entry point x, y, z, t of leading muon at entry center point x, y, z, t of leading muon at entry propagation length in detector of leading muon total propagation length of leading muon

6.2.7 Energy

First of all, the reconstruction of different energies is evaluated for the different levels of the data selection stage. This includes the bundle energy at surface, the leading muon energy at the detector entry, and the bundle energy at entry.

In all plots, H3a weighting is used. The title of all plots include the Spearman and Pearson correlation coefficients, as defined in Section 6.2.5. The top panel presents a two-dimensional histogram of the reconstructed energy on the x-axis and the true energy on the y-axis with the median and 68 % interval in orange. The bottom panel shows the median of the relative difference $(y - x)/x$ normed on the prediction with propagated uncertainties highlighted as error bars.

Level 3

In Figure 6.6, the bundle energy at surface is evaluated. The median prediction with a 68 % interval agrees with the true energy up to reconstructions of $5 \cdot 10^8$ GeV. The Spearman correlation of $\rho_S = 0.75$ and the Pearson correlation of $\rho_P = 0.67$ indicate a linear correlation. However, in most of the bins, the prediction includes outliers of true energies up to the maximum energy of about 10^9 GeV. For this reconstruction, it needs to be mentioned that the muon bundle travels from a few hundreds of meters up to several kilometers in ice before entering the detector with energy losses of all individual muons. These losses can not be detected by the in-ice array. Additionally, the muon bundles have enough energy to pass through the detector, and thus only a fraction of the deposited energy is detected. Considering these facts, the reconstruction is satisfactory. This prediction is used in Section 7.1 to apply a pre-cut to select events with a bundle energy at surface of at least $E_{\mu,\,\mathrm{surf}}^{\mathrm{bundle}} > 500\,\mathrm{TeV}$.

In Figure 6.7, the bundle energy at entry and the leading muon energy at entry are evaluated. The reconstruction of the bundle energy at entry shows the largest deviations below energies of 20 TeV. This effect is caused by the training. The according network was trained on events that have at least 200 TeV bundle energy at the surface. Hence, the network has never seen any lower energetic events, except of some outliers that have passed the cut even though their actual bundle energy at surface

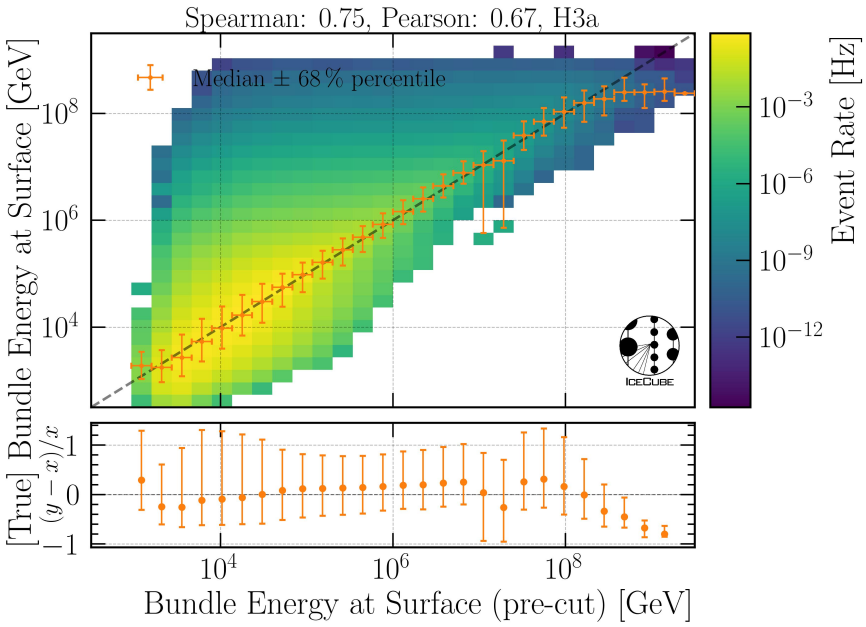
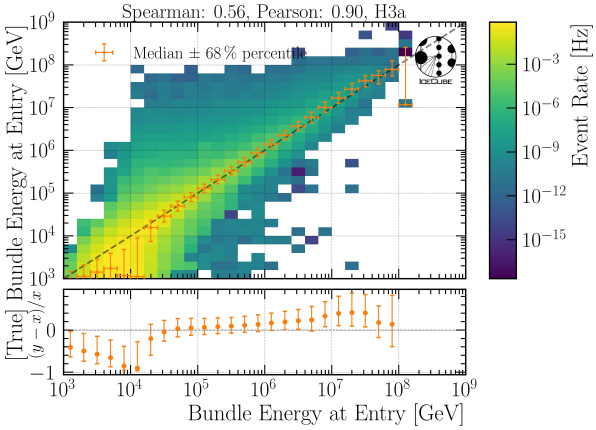


Figure 6.6: The true bundle energy at surface vs. the predicted energy, weighted with H3a ist presented. This is based on *Level 3*. The bottom panel shows the relative difference normed on the prediction. This reconstruction is used to apply the pre-cut, as stated in Section 7.1.

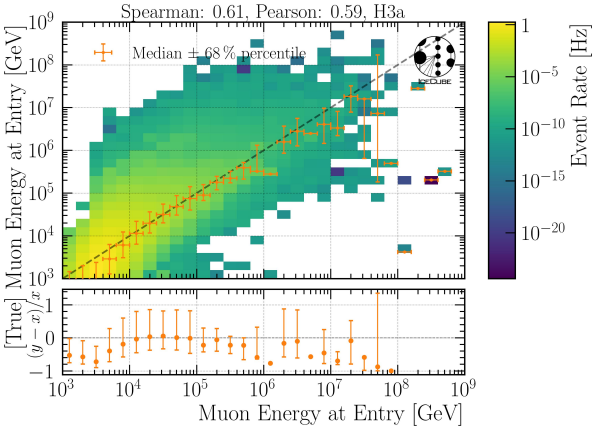
is below this threshold. A similar effect is observed for the leading muon energy at entry, but it is shifted to lower energies of about 4 TeV. Since both energies are reconstructed by the same network, the same explanation as above applies here. The prediction is in agreement with the reconstruction for energies up to 1 PeV. At higher energies, on the one hand, the statistics decreases, on the other hand, the reconstruction smears out. The Pearson coefficient is $\rho_{\mathrm{P}} = 0.59$. For this reconstruction, it needs to be stated that an entire muon bundle enters the detector. The higher the energy, the more muons are included in the bundle [147]. The spatial distance between the muons is on the order of cm up to a few m [104] and the muons enter the detector nearly at the same time. Hence, the light emitted by all muons overlaps in the DOMs, and it is challenging to match which light is coming from the most energetic muon in that bundle. Additionally, no specific quality cuts are applied. With that, the presented reconstructions of the leading muon energy are promising, and these are the first reconstructions of the leading energy ever performed in IceCube.

Level 4

The energy reconstructions are evaluated on *Level 4* in this paragraph. Hence, low energy events with $E^{\mathrm{bundle}}_{\mu,\,\mathrm{surf}} > 500\,\mathrm{TeV}$ have been discarded. Assuming a perfect reconstruction, only events below this threshold would be rejected. However, as presented in Figure 6.6, there are outliers. For example, light from traversing but not entering muons can affect nearby DOMs, which then detect only a part of the emitted Cherenkov light. This can cause an energy prediction below the actual energy. The prediction of the bundle energy at entry is shown in Figure 6.8a. The distribution is less smeared out overall. Also both correlation coefficients are higher with $\rho_{\mathrm{S}} = 0.95$ and $\rho_{\mathrm{P}} = 0.95$. The leading energy at entry is presented in Figure 6.8b. Similar effects as described before can be observed with a narrower distribution. However, the Pearson coefficient is slightly lower with $\rho_{\mathrm{P}} = 0.53$. At this stage, the data–MC quality cuts are performed, resulting in *Level 5*. The exact cuts are described in Section 7.3.

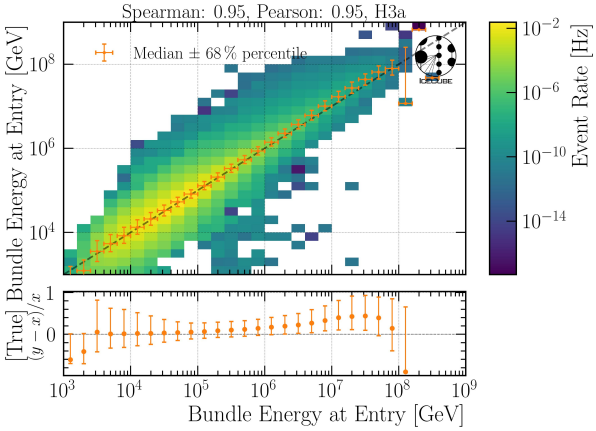


(a) Bundle energy at entry.

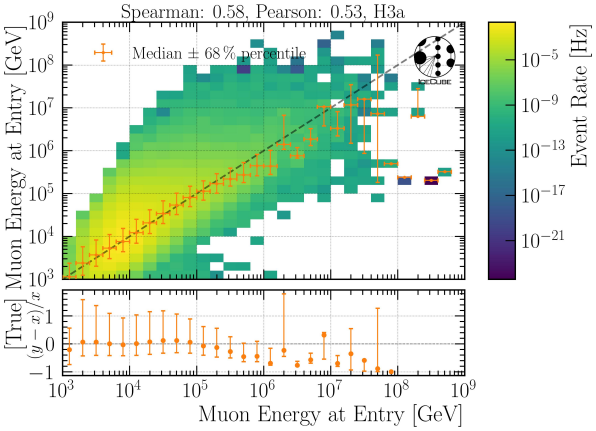


(b) Leading muon energy at entry.

Figure 6.7: Both figures show the true energy vs. the predicted reconstructed energy, weighted with H3a. Left panel is for the bundle energy at detector entry, right panel is for the leading muon energy at detector entry. This is based on *Level 3*. The bottom panel shows the relative difference normed on the prediction.



(a) Bundle energy at entry.

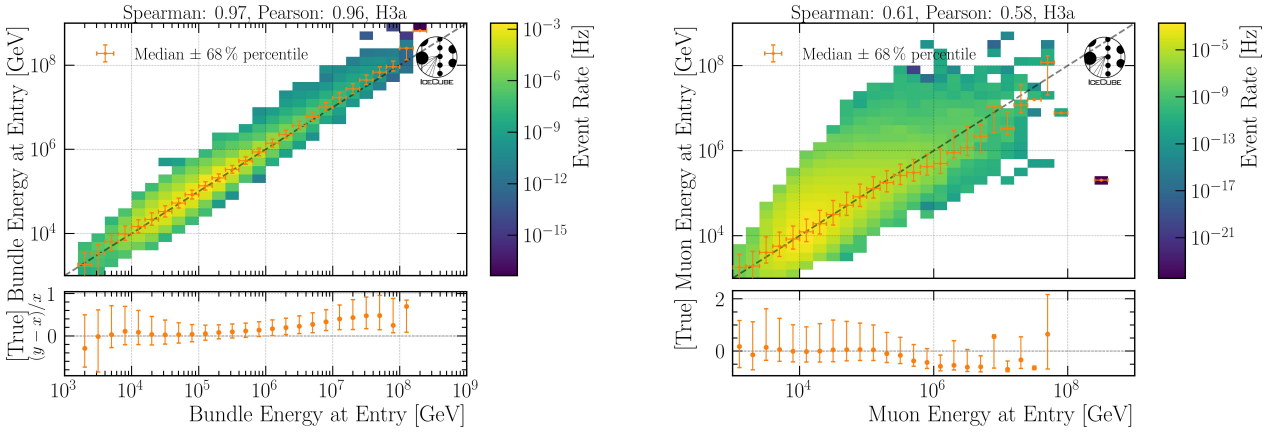


(b) Leading muon energy at entry.

Figure 6.8: Both figures show the true energy vs. the predicted reconstructed energy, weighted with H3a. Left panel is for the bundle energy at detector entry, right panel is for the leading muon energy at detector entry. This is based on *Level 4*. The bottom panel shows the relative difference normed on the prediction.

Level 5

After the application of quality cuts to improve the data–MC agreement (see Table 7.3), the energy reconstructions are evaluated on *Level 5*. The bundle energy at entry is presented in Figure 6.9a. The distribution is less smeared out compared to *Level 4* and the correlation coefficients are higher with $\rho_S = 0.97$ and $\rho_P = 0.96$. However, a slope remains indicating an underestimation of the true bundle energy towards the highest energies. The leading muon energy at entry is presented in Figure 6.9b. The distribution is also less smeared out compared to *Level 4* and the Pearson coefficient is $\rho_P = 0.58$. The 68 % interval agrees with the true energy up to energies of about 1 PeV. Towards higher energies, the prediction overestimates the true energy. As explained above, the light of too many high-energetic muons overlays in the DOMs, which makes it hard to reconstruct the true leading muon energy. The network associates more energy with the leading muon.



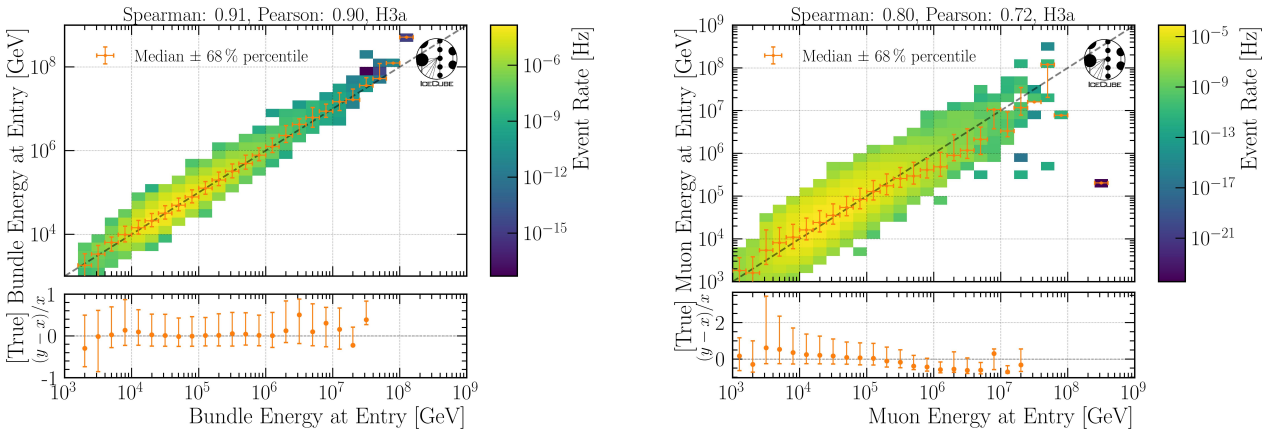
(a) Bundle energy at entry.

(b) Leading muon energy at entry.

Figure 6.9: Both figures show the true energy vs. the predicted reconstructed energy, weighted with H3a. Left panel is for the bundle energy at detector entry, right panel is for the leading muon energy at detector entry. This is based on *Level 5*. The bottom panel shows the relative difference normed on the prediction.

Final Level

The *Final level* is the last stage of the data selection. Here, it is required that the leading muon carries at least 40 % of the energy of the entire muon bundle at entry, referred to as leadingness cut. In Figure 6.10a, the bundle energy at entry reconstructions is evaluated. The distribution is even less smeared out compared to *Level 5*. Additionally, the slope indicating an underestimation towards higher energies vanishes. However, the 68 % intervals increase, with correlations of $\rho_S = 0.91$ and $\rho_P = 0.90$. The lower correlations are caused by outliers due to less statistics, since every additional cut reduces the number of events. The leading muon energy at entry is evaluated in Figure 6.10b. This reconstruction gains the most advantage by the leadingness cut. The correlations go up to $\rho_S = 0.80$ and $\rho_P = 0.72$. Also, this cut removes most of the outliers for this prediction. The better reconstruction is caused by the fact that the leading muon carries at least 40 % of the energy of the entire muon bundle at entry. This means that one muon dominates the bundle, and thus emits most of the detected light which simplifies the reconstruction with the DNN.



(a) Bundle energy at entry.

(b) Leading muon energy at entry.

Figure 6.10: Both figures show the true energy vs. the predicted reconstructed energy, weighted with H3a. Left panel is for the bundle energy at detector entry, right panel is for the leading muon energy at detector entry. This is based on *Final Level*. The bottom panel shows the relative difference normed on the prediction.

6.2.8 Direction

In this section, the directional reconstruction of the leading muon is evaluated. For this, the network predicts not only the zenith and azimuth angle in spherical coordinates, but also independently the cartesian direction vector without using angular relations between the two. For completeness, also the angular resolution is evaluated. The angular resolution is defined as the opening angle $\Delta\Psi$ between the true and predicted direction. In spherical coordinates, the calculation is done via

$$\cos(\Delta\Psi_\mathrm{s}) = \cos(\phi_\mathrm{true} - \phi_\mathrm{pred}) \cdot \sin(\theta_\mathrm{true}) \cdot \sin(\theta_\mathrm{pred}) + \cos(\theta_\mathrm{true}) \cdot \cos(\theta_\mathrm{pred}) \tag{6.15}$$

$$\Delta\Psi_\mathrm{s} = \arccos\left(\cos(\Delta\Psi_\mathrm{s})\right) \tag{6.16}$$

with the azimuth angles $\phi_\mathrm{true}, \phi_\mathrm{pred}$ and zenith angles $\theta_\mathrm{true}, \theta_\mathrm{pred}$, respectively.

In cartesian coordinates, given two direction vectors

$$\vec{v}_\mathrm{true} = (x_\mathrm{true}, y_\mathrm{true}, z_\mathrm{true})^\top, \tag{6.17}$$

$$\vec{v}_\mathrm{reco} = (x_\mathrm{reco}, y_\mathrm{reco}, z_\mathrm{reco})^\top, \tag{6.18}$$

the angular deviation $\Delta\Psi_\mathrm{c}$ is computed via the dot product

$$\Delta\Psi_\mathrm{c} = \arccos\left(\frac{x_\mathrm{true}x_\mathrm{reco} + y_\mathrm{true}y_\mathrm{reco} + z_\mathrm{true}z_\mathrm{reco}}{|\vec{v}_\mathrm{true}| \cdot |\vec{v}_\mathrm{reco}|}\right) \tag{6.19}$$

where the respective vector norms are given by

$$|\vec{v}_i| = \sqrt{x_i^2 + y_i^2 + z_i^2} \tag{6.20}$$

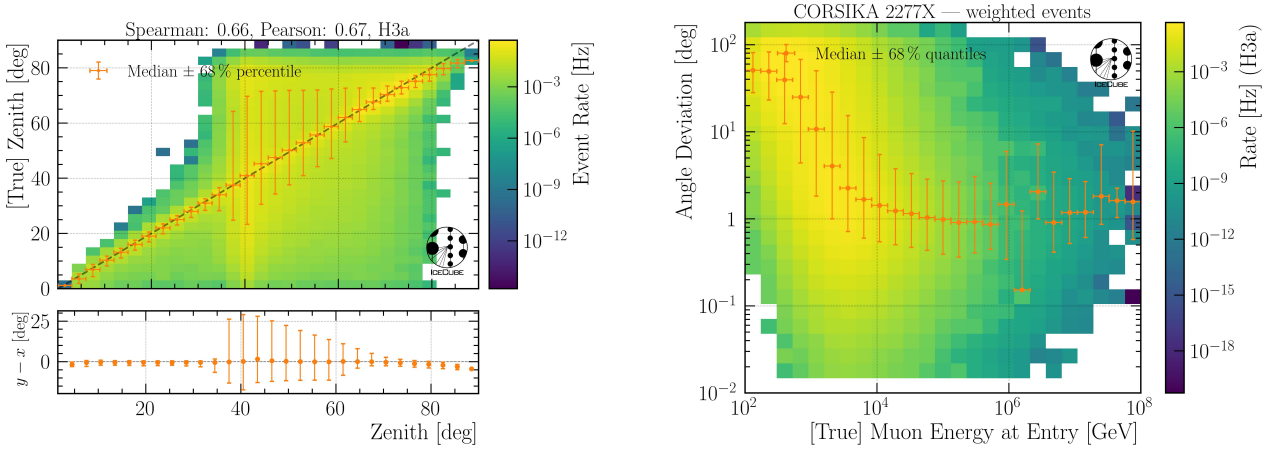
Values outside the domain $[-1, 1]$ due to numerical precision are clipped, and undefined cases (e. g., zero-length vectors) are excluded. A comparison between the two methods is presented in Figure 6.13 on *Level 4*. Since the results are nearly identical, the choice of the coordinate system has no significant impact. In the following, the cartesian coordinates are used. Also, all angular resolution plots have a minimum around 0.015° due to numerical precision.

As a short remark: for the training, the true zenith angle is not taken from the leading muon, but from the primary cosmic ray particle. This was only done because in the preparation of the training data,

the leading muon zenith angle has not been stored. However, the difference between the primary zenith and the leading muon zenith has been investigated afterwards and has been found to be negligible as presented in Figure A.3. In the following, the zenith and azimuth reconstruction as well as the angular resolution are evaluated for the different levels of the data selection stage. Azimuth is only presented on *Level 4*.

Level 3

First of all, the zenith reconstruction and the angular resolution are presented on *Level 3* in Figure 6.11. The median zenith prediction agrees with the truth, however, the 68 % intervals between $\theta_{\text{reco}} \approx 35°$ and $\theta_{\text{reco}} \approx 65°$ are larger in comparison to the rest, and extend up to a deviation of 25°. The median angular resolution is presented in the right panel of Figure 6.11 in bins of the true muon energy at entry. Below $E_\mu^{\text{leading}} < 10\,\text{TeV}$ the deviation is above 2°. Up to 1 PeV, it decreases to about 1°. Similar to the energy reconstruction, this network was also trained on events after the pre-cut, hence it has never seen low energy events. This, in addition to the less track-like signature of low-energy muons, leads to the larger deviations for lower-energetic muons.

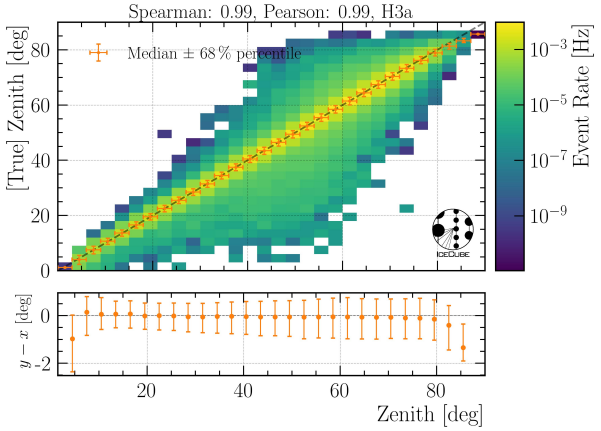


(a) Zenith. **(b)** Angular resolution.

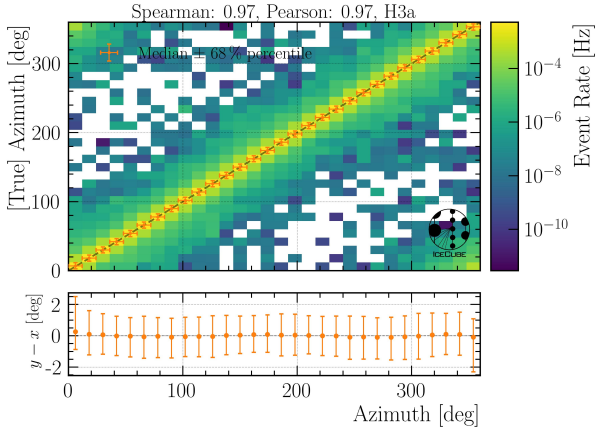
Figure 6.11: The left panel shows the reconstructed zenith angle vs. the true zenith angle. The bottom panel shows the relative difference normed on the prediction. The right panel shows the angular difference between the true and reconstructed direction. Both figures use H3a weighting. This is based on *Level 3*.

Level4

In Figure 6.12a, the zenith reconstruction is evaluated. The removal of low energy events leads to a better reconstruction and removes outliers. The Pearson coefficient is $\rho_{\text{P}} = 0.99$, and the 68 % interval is $\pm(0.5-1)\,°$ between $\theta_{\text{reco}} \approx 5°$ and $\theta_{\text{reco}} \approx 80°$. The azimuth reconstruction is shown in Figure 6.12b. With a Pearson coefficient of $\rho_{\text{P}} = 0.97$ and a 68 % interval of $\pm 1.5°$ over the entire range, the azimuth angle is also well reconstructed. The angular resolution is presented in Figure 6.13 utilizing both the spherical and cartesian prediction. Both predictions lead to nearly identical results. The median angular resolution is about 1° for energies $2\,\text{TeV} < E_\mu^{\text{leading}} < 4\,\text{PeV}$. The angular resolution is better than on *Level 3*, especially in the low energy region.

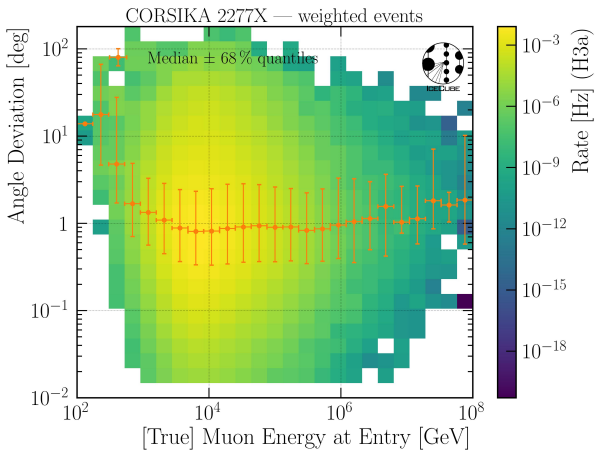


(a) Zenith.

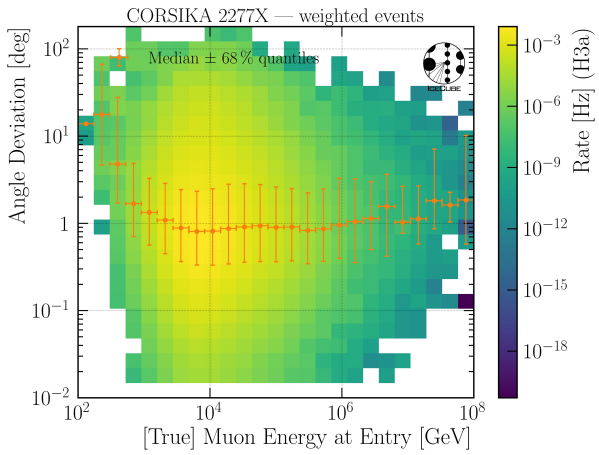


(b) Azimuth.

Figure 6.12: The left panel shows the reconstructed zenith angle vs. the true zenith angle. The right panels shows the reconstructed azimuth angle vs. the true azimuth angle. The bottom panels show the difference. Both figures use H3a weighting. This is based on *Level 4*.



(a) Spherical coordinates.

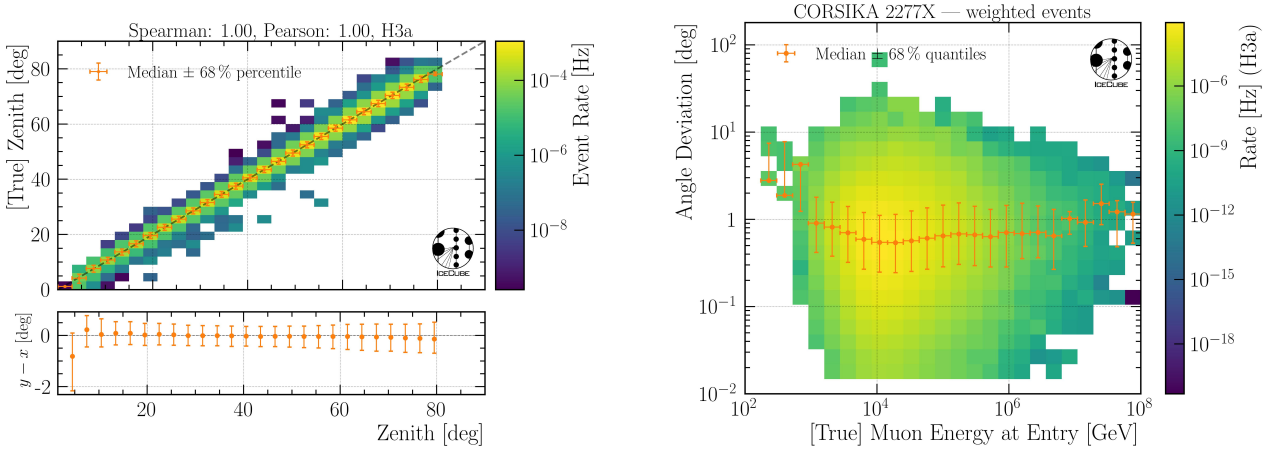


(b) Cartesian coordinates.

Figure 6.13: Both figures present the angular resolution of the leading muon direction utilizing spherical and cartesian coordinates. H3a weighting is used. This is based on *Level 4*. Both methods lead to nearly the same results.

Level 5

On *Level 5* only the zenith angle and the angular resolution are presented, in Figure 6.14. The application of quality cuts (see Table 7.3) lead to a removal of many outliers in the zenith reconstruction. The Pearson coefficient is $\rho_{\mathrm{P}} = 1.00$ and the 68 % interval is $\pm 0.5°$ for all angles above $\theta_{\mathrm{reco}} \approx 5°$. Also the angular resolution is improved. The median angular resolution is about 0.5° at $E_{\mu}^{\mathrm{leading}} \approx 10\,\mathrm{TeV}$, and it is about 0.7° for $500\,\mathrm{GeV} < E_{\mu}^{\mathrm{leading}} < 5\,\mathrm{PeV}$.



(a) Zenith. **(b)** Angular resolution.

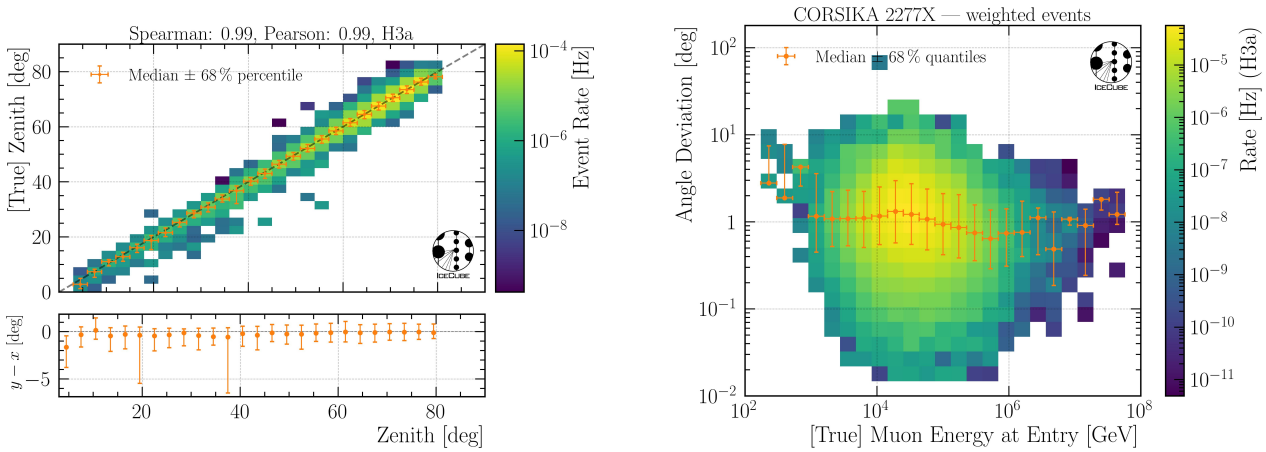
Figure 6.14: The left panel shows the reconstructed zenith angle vs. the true zenith angle. The bottom panel shows the relative difference normed on the prediction. The right panel shows the angular difference between the true and reconstructed direction. Both figures use H3a weighting. This is based on *Level 5*.

Final Level

On the *Final Level*, also the zenith reconstruction and angular resolution are evaluated, as presented in Figure 6.15b. The 68 % intervals of the zenith reconstruction shrink, except of two outliers at about 20° and at about 38°. These can be explained by the reduced statistics. The Pearson correlation is $\rho_{\mathrm{P}} = 0.99$. The right panel of the figure shows the angular resolution. For energies up to $E_{\mu}^{\mathrm{leading}} \approx 100\,\mathrm{TeV}$, the median angular resolution is above 1°. Up to 2 PeV, it is below 1°. However, the resolution is worse than before the additional cut. This worsening can be explained by the training data. The network was trained on events on *Level 4*. Due to the selection of events with a leading muon that carries at least 40 % of the energy of the entire muon bundle at entry, the leading muon is more dominant in the bundle. Hence, these events deposit their energy more stochastically, than bundles where no dominant muon is present. These stochastic deposits lead to a worse angular resolution. To improve this, a separate network needs to be trained on events speficically with a leading muon that carries at least 40 % of the energy of the entire muon bundle at entry. However, this goes beyond the scope of this thesis since angular information is not relevant for this analysis, but only for data-MC quality cuts, as stated in Section 7.3.

6.2.9 Track Geometry

In this section, several properties of the track geometry are evaluated. First, there is the entry point that is defined as the point where the muon enters the detector. Similar to the definition of the entry energy, the IceCube in-ice array is extended by a convex hull of 200 m. This coordinate is given by



(a) Zenith.

(b) Angular resolution.

Figure 6.15: The left panel shows the reconstructed zenith angle vs. the true zenith angle. The bottom panel shows the relative difference normed on the prediction. The right panel shows the angular difference between the true and reconstructed direction. Both figures use H3a weighting. This is based on *Final Level*.

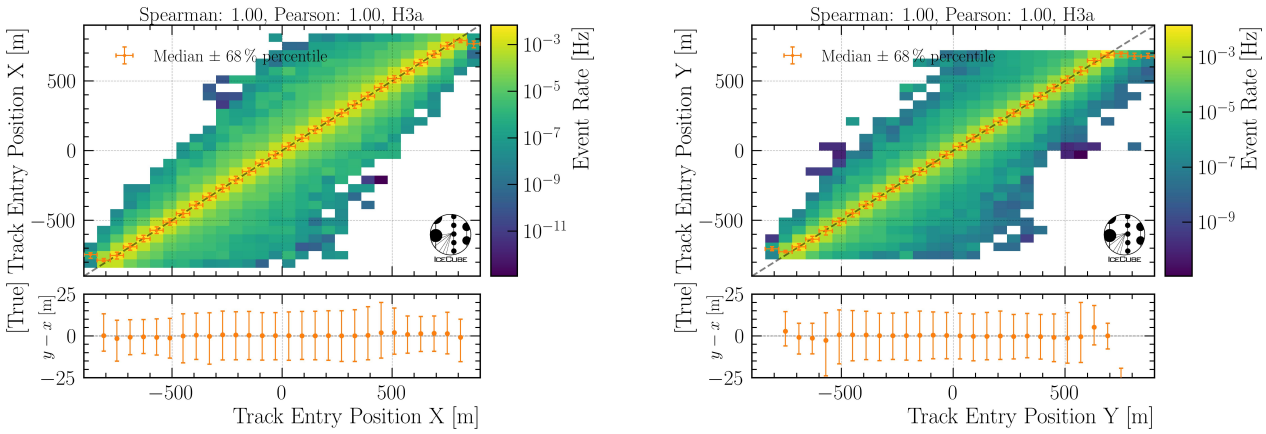
x_{entry}, y_{entry} and z_{entry}. As the fourth component, the relative time t_{entry} is reconstructed. Analogue to the entry point, the center vertex is defined as the closest approach point to the center of the detector, which is defined as $(0, 0, 0)$. The coordinates are given by x_{center}, y_{center} and z_{center}. The relative time is given as t_{center}. For this, the position along the muon track with the shortest distance perpendicular to the center of the detector is calculated.

The total track length s_{tot} is defined as the distance between the vertex at the surface and the vertex where the muon stops. The track length in the detector s is defined as the distance between the entry point and the exit point of the muon in the detector. The exit point is defined as the point where the muon leaves the extended convex hull around the in-ice array.

The performance of all track geometry reconstructions is shown only on *Level 4* because the predictions are only used to choose appropriate quality cuts on that level. They have no further use for the analysis.

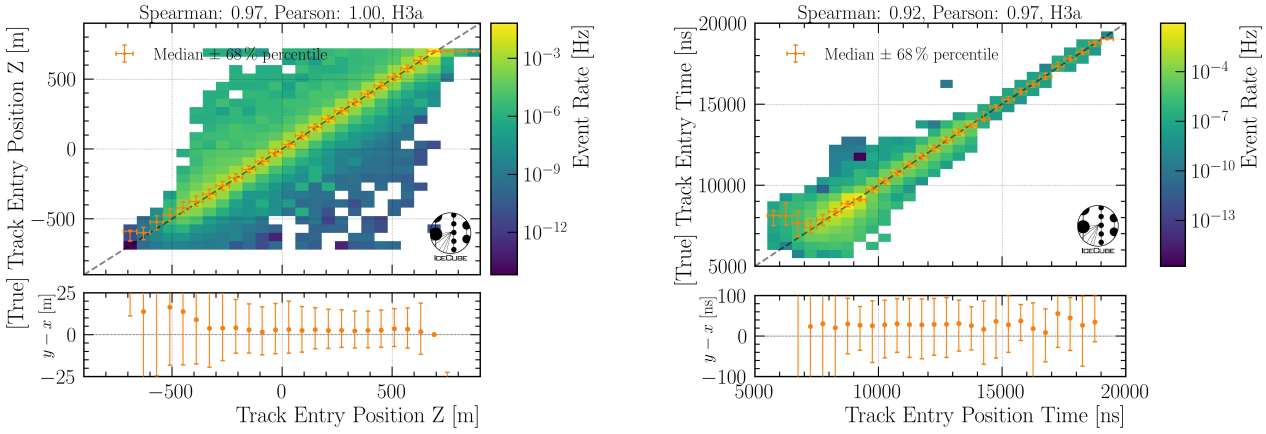
Level 4

At first, the entry vertex is evaluated in the two Figures 6.16 and 6.17. For all three spatial coordinates, the Pearson coefficients are $\rho_{\text{P}} = 1.00$. For the x- and y-vertex, the 68 % interval is approximately $\pm 15\,\text{m}$. For the z-vertex (depth), below 100 m the 68 % interval broadens from approximately $\pm 15\,\text{m}$ with increasing depth. In this region, the network shows a bias toward greater depth (i.e., it predicts z values that are too deep). The properties of the ice change with depth [148], and the network has been trained on an ice model that is not depth dependent. Analyzing the exact properties of the ice is beyond the scope of this thesis, but it is an important and on-going research topic for the IceCube collaboration [114]. However, a network has also been trained on a MC dataset that has been generated with the latest ice model [148, 149]. The reconstruction is presented in Figure A.4. Since not enough training data in the high energy region is available, for this thesis, the networks trained on the old ice model are utilized. Also, the z-vertex is only used for quality cuts and it is of no importance for the further analysis. The prediction of the relative time of the entry position has a Pearson coefficient of $\rho_{\text{P}} = 0.97$. Overall, the prediction is about $(20 - 40)\,\text{ns}$ below the truth. This variable is also less important for the analysis, since only one data–MC quality cut is performed on its uncertainty prediction.



(a) Entry x-vertex. **(b)** Entry y-vertex.

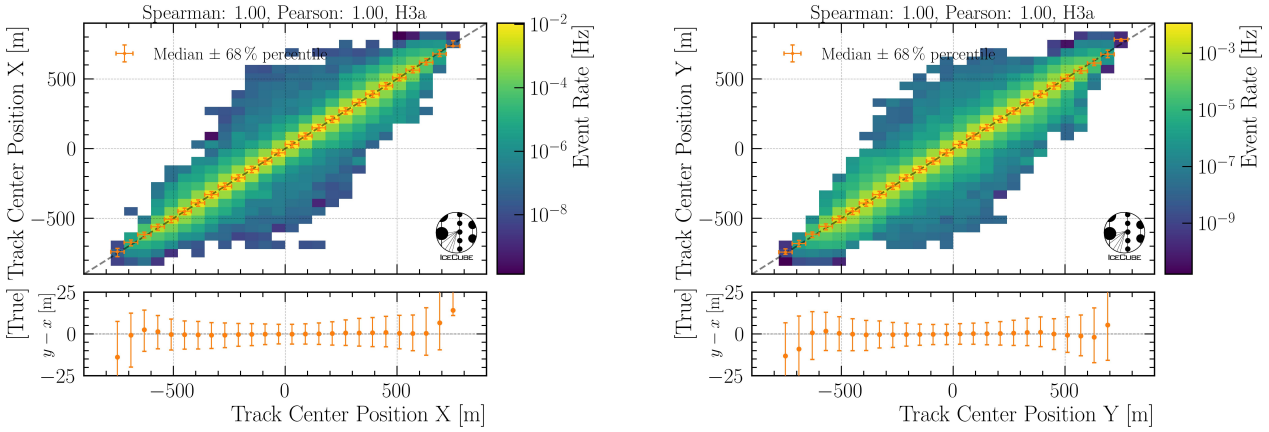
Figure 6.16: The left panel shows the reconstructed x-vertex of the entry point to the detector. The right panel shows the entry y-vertex reconstruction. The bottom panels show the difference. Both figures use H3a weighting. This is based on *Level 4*.



(a) Entry z-vertex. **(b)** Entry time.

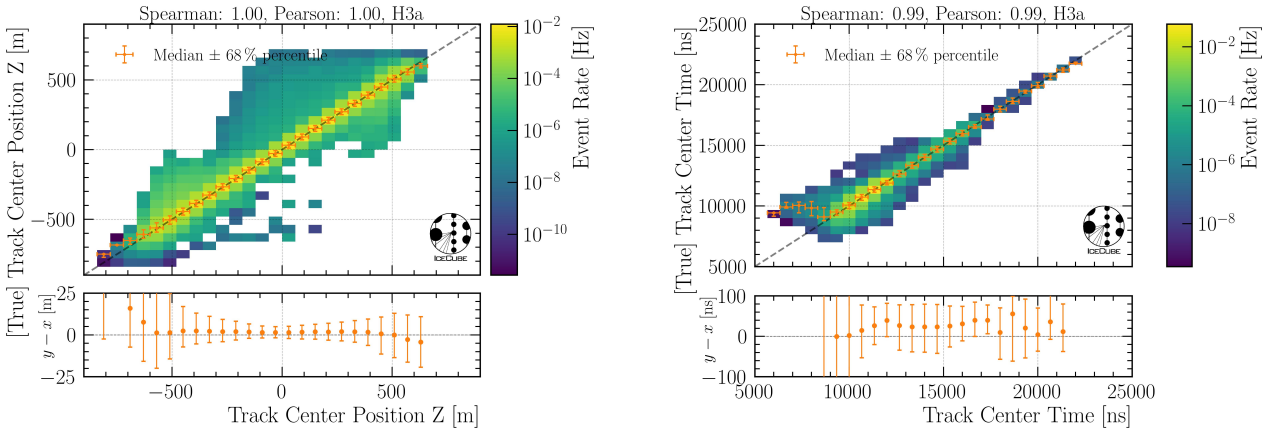
Figure 6.17: The left panel shows the reconstructed z–vertex of the entry point to the detector. The right panel shows the relative time reconstruction of the entry point. The bottom panels show the difference. Both figures use H3a weighting. This is based on *Level 4*.

Similar to the entry vertex, the center vertex is evaluated in Figures 6.18 and 6.19. The Pearson coefficients for the x- and y-vertex are $\rho_{\mathrm{P}} = 1.00$. The central 68 % interval is approximately $\pm 5\,\mathrm{m}$ near the detector center and widens to approximately $\pm 15\,\mathrm{m}$ toward the detector edges for both horizontal coordinates (x and y). Also, the z-vertex has a Pearson coefficient of $\rho_{\mathrm{P}} = 1.00$. The 68 % intervals are similar with even smaller ranges close to the detector center. However, for $z_{\mathrm{center}} < -500\,\mathrm{m}$ the intervals increase. This is expected since the lower border of the detector is at that depth. Hence, a muon whose point of closest approach to the detector center lies at this depth enters the detector nearly horizontally and only grazes the in-ice array rather than fully traversing it. A similar behavior is observed at the top of the detector at $z_{\mathrm{center}} > 500\,\mathrm{m}$. The relative time prediction has a Pearson coefficient of $\rho_{\mathrm{P}} = 0.99$. It is similar to the entry time prediction, but shifted towards higher times since time passes from the entry point to the center point.



(a) Center x-vertex. **(b)** Center y-vertex.

Figure 6.18: The left panel shows the reconstructed x-vertex of the closest approach point to the center of the detector. The right panel shows the center y-vertex reconstruction. The bottom panels show the difference. Both figures use H3a weighting. This is based on *Level 4*.

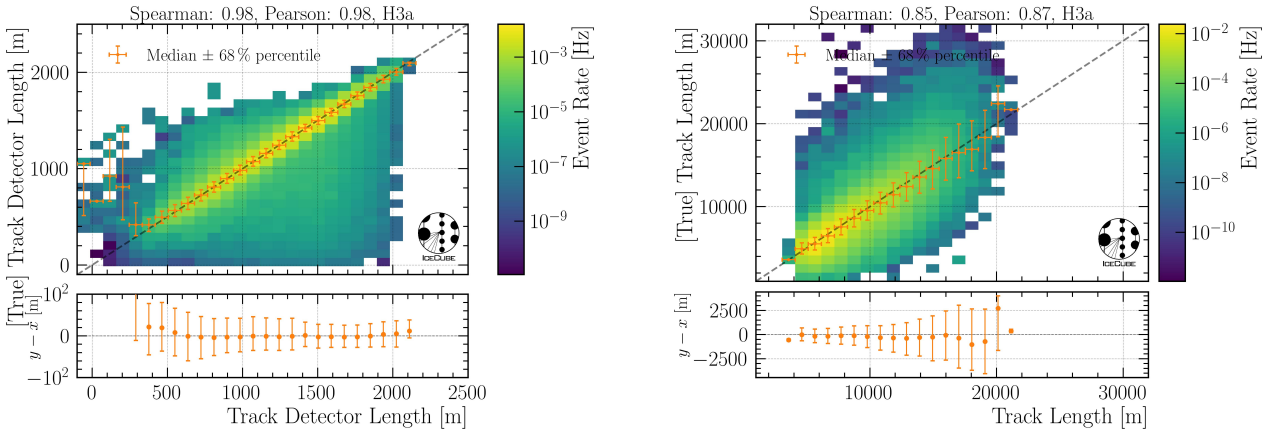


(a) Center z-vertex. **(b)** Center time.

Figure 6.19: The left panel shows the reconstructed z-vertex of the closest approach point to the center of the detector. The right panel shows the relative time reconstruction of the center point. The bottom panels show the difference. Both figures use H3a weighting. This is based on *Level 4*.

At last, the propagation lengths are evaluated in Figure 6.20. The left panel shows the propagation length within the detector. The Pearson coefficient is $\rho_{\mathrm{P}} = 0.98$ and the 68 % interval is about

$\pm(20-40)$ m, increasing towards lower track lengths. With a detector size approximately 1 km × 1 km, track lengths below 1 km are caused by muons that traverse only a part of the in-ice array. Hence, less Cherenkov light can be detected resulting in worse reconstructions. The right panel shows the total propagation length. The Pearson coefficient is $\rho_P = 0.87$ with 68 % intervals increasing from about ±750 m to ±3000 m. This prediction is still sufficient considering that the detector detects the muons' light for only up to 2 km by traversing the in-ice array diagonally, but it needs to extend this to the surface. For this, the network has to put more importance to the muons' direction.



(a) Length in detector.

(b) Total propagation length.

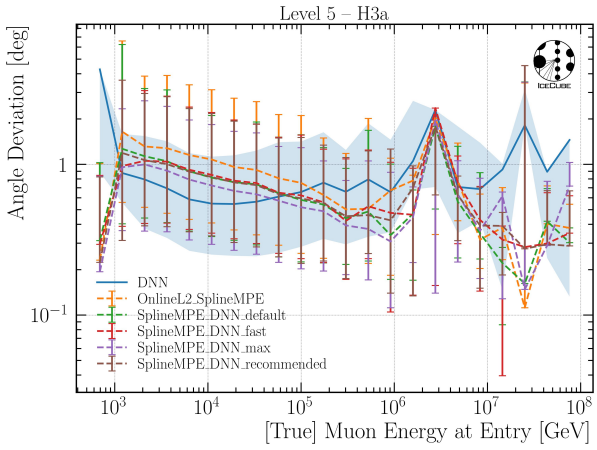
Figure 6.20: The left panel shows the reconstructed propagation length within the detector. The right panel shows the total propagation length reconstruction. The bottom panels show the difference. Both figures use H3a weighting. This is based on *Level 4*.

6.2.10 Traditional vs. Machine Learning Reconstructions

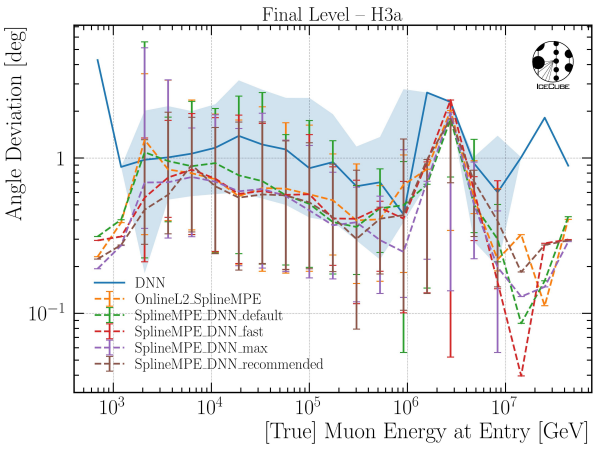
In this section, a comparison between the traditional and the ML reconstructions is presented to establish the eligibility of the ML reconstructions. This is done on *Level 5* and *Final Level* for the directional, leading, and bundle energy at entry reconstructions. All comparisons are presented in Figure 6.21. The angular resolution is evaluated as a function of the true leading muon energy at entry. For the application of the standard *SplineMPE* reconstruction, four different settings are available: *default, fast, max,* and *recommended*. Here, all four settings are tested. Additionally, a specific configuration named *OnlineL2_SplineMPE* is tested.

On *Level 5* (see Figure 6.21a) all methods achieve similar performances for the angular resolution within the 68 % intervals of the median, highlighted as blue shaded area or as errorbars in case of the traditional reconstructions. However, in most bins the ML-based reconstruction leads to narrower 68 % intervals but seems to achieve a worse performance for the highest energetic muons at entry with energies above 10 PeV. At about 3 PeV there is a significant decrease in angular resolution up to approximately 2° for both, the MC-based and the traditional reconstruction, for which no explanation like limited sample size in this bin could be found. Overall, an angular resolution around and mostly below 1° is achieved.

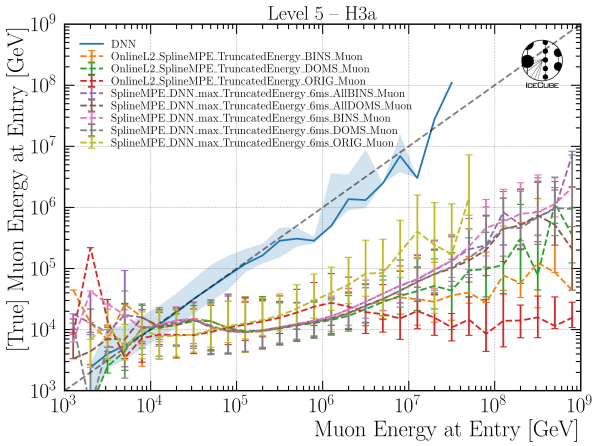
On *Final Level* (see Figure 6.21b), the ML reconstruction is outperformed by the traditional reconstructions for nearly all energies. The ML resolution is around 1°, while the traditional reconstructions are below 1° for almost the entire energy range. However, the artifact at 3 PeV remains. While the standard method has an improved resolution, the ML resolution decreases in comparison to *Level 5*. This can be explained by the additional cut that selects events with a leading muon that carries at least 40 % of the energy of the entire muon bundle at entry. This causes the event to be more like



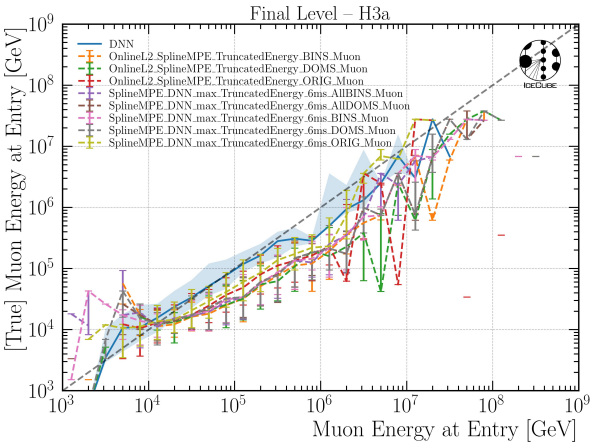
(a) Angular resolution on *Level 5*.



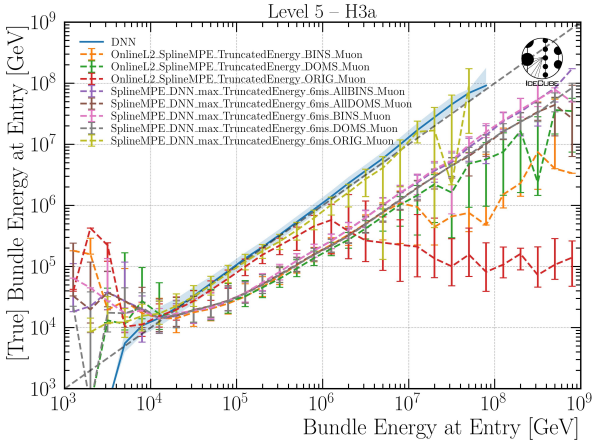
(b) Angular resolution on *Final Level*.



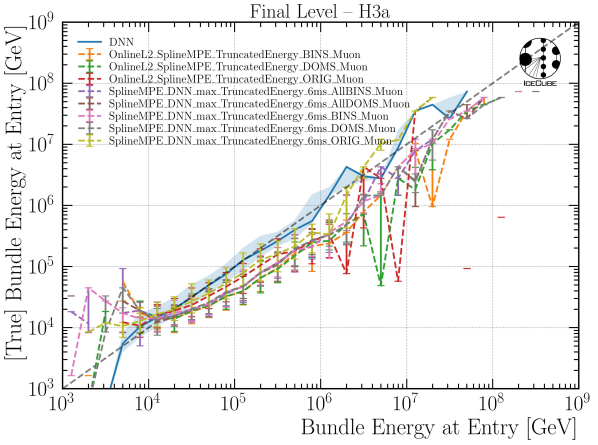
(c) Leading energy on *Level 5*.



(d) Leading energy on *Final Level*.



(e) Bundle energy on *Level 5*.



(f) Bundle energy on *Final Level*.

Figure 6.21: This figure shows the comparison of traditional IceCube reconstruction methods and the ML reconstructions performed in this thesis. The left column shows the results on *Level 5*, the right column on *Final Level*. The top row shows the angular resolution, the middle row the leading muon energy at entry, and the bottom row the bundle energy at entry. All figures use H3a weighting. The method *SplineMPE* is used for the angular reconstruction, *TruncatedEnergy* is compared to both the leading and bundle energy at entry. In all plots, the blue line represents the ML reconstruction with a 68 % interval, and the colored dashed lines with error bars represent the standard reconstructions. Only a subset of the entire MC dataset is used for this comparison.

a single muon event, and the standard methods have been optimized to reconstruct a single muon track. The ML reconstruction has not been trained on this specific event class, and hence, it is not able to adapt to this change in the event topology. However, in this analysis, the angular resolution is not relevant because it is not needed to search for specific event directions in the sky. Additionally, the ML reconstruction is performed in (0.06 ± 0.06) s, while the fastest *SplineMPE* reconstruction takes (0.46 ± 0.75) s, and the recommended version needs (0.83 ± 1.16) s on median with standard deviation. The distribution of the computation time for all ML reconstructions and the *SplineMPE* reconstructions is presented in Figure 6.22. Considering the sufficient angular MC resolution where the standard reconstructions are within the 68 % of the ML reconstruction, and the faster computation time, the ML reconstruction is used for this analysis. It also must be stated that the *SplineMPE* method requires a seed track, thus a prior reconstruction, and in this scenario, the ML prediction has been used as a seed.

The leading muon energy at entry reconstruction is compared in the middle panel of Figure 6.21. For *TruncatedEnergy*, several options for how the muon energy loss is projected are available, like *BINS, ALLBINS, DOMS, ALLDOMS* and *ORIG*. On *Level 5*, the ML reconstruction outperforms all standard reconstructions over the entire energy range, which can be seen by the accurate agreement with the dashed gray line indicating a perfect reconstruction. This is expected since the deep neural network (DNN) is trained to predict the energy of the most energetic muon within a bundle of muons, while the standard reconstructions just take the energy losses of the entire muon bundle. On *Final Level*, the ML reconstruction is still better than all standard reconstructions, but as expected for more single muon-like events, the prediction is improved, while still overestimating the energy. The

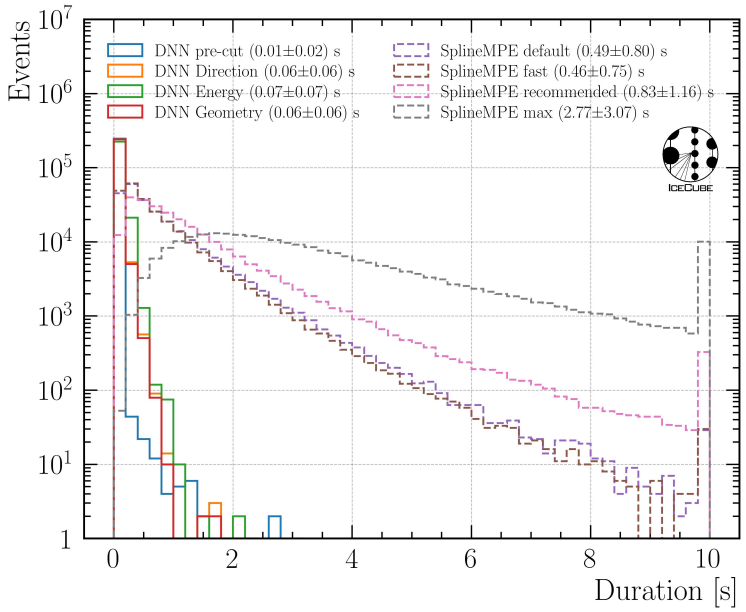


Figure 6.22: This figure shows the computation time of the ML reconstruction (in solid lines) in comparison to the traditional *SplineMPE* reconstruction (in dashed lines) on *Level 5*. The distribution includes an overflow bin for all times above 10 s. In the legend, the median time with standard deviation is presented for all eight reconstructions. The ML reconstructions are significantly faster. However, the comupation time of *TruncatedEnergy* is only (0.009 ± 0.006) s. It needs to be mentioned that the DNN reconstructions have been run on CPUs and not on GPUs, which would be faster, and all predictions have been made on a computer cluster with several different CPU models.

bundle energy at entry reconstruction is presented in the bottom panel of Figure 6.21. On *Level 5*, the ML reconstruction is better than all standard reconstructions, except for the *SplineMPE_ORIG* method that performs almost equally well. However, the ML prediction has fewer outliers expressed

by a tighter 68 % interval. Since *SplineMPE* rejects large stochastic energy losses, the prediction also underestimates the true energy, as explained in Section 6.1. On *Final Level*, the ML reconstruction is still better than all other standard reconstructions. For the standard methods, almost all methods have an improved prediction, except for the previous best prediction *SplineMPE_ORIG*, which is worse than on *Level 5*. Regarding computation time (median±standard deviation of the median), *TruncatedEnergy* is performed in (0.009 ± 0.006) s, while the energy DNN needs (0.07 ± 0.07) s. This timing evaluation has been done on CPUs instead of GPUs, which would be faster.

All comparisons are performed only on a subset of the entire MC dataset to reduce the computation time.

6.3 Summary: Reconstruction

In this chapter, the reconstruction of several muon properties utilizing CNNs has been presented. The reconstructions have been evaluated on several selection levels used in this analysis. Overall, the reconstructions show good performance for all variables. The angular resolution of the leading muon direction is about 0.5° at 10 TeV at *Level 5*, see Figure 6.14b. This is comparable to IceCube's state-of-the-art neutrino directional reconstructions [150]. However, it must be noted that this is a muon reconstruction and not a neutrino reconstruction. The leading muon energy at entry is reconstructed sufficiently on *Final Level* as presented in Figure 6.10b. Correlations of $\rho_S = 0.80$ and $\rho_P = 0.72$ are achieved. This is the first reconstruction of the leading muon energy at entry ever performed in IceCube. The bundle energy at entry is shown in Figure 6.10a on *Final Level*, with correlations of $\rho_S = 0.91$ and $\rho_P = 0.90$. These new ML-based reconstructions are compared to traditional IceCube reconstruction methods in Figure 6.21. The angular resolution is comparable to the traditional *SplineMPE* reconstruction on *Level 5*. The leading muon energy reconstruction outperforms the traditional *TruncatedEnergy* reconstruction on both, *Level 5* and *Final Level*, as expected, since these traditional methods are not able to reconstruct the leading muon energy specifically. The bundle energy reconstruction also outperforms the traditional *TruncatedEnergy* reconstruction on both levels.

In addition to the reconstructions presented above, networks have also been trained on MC datasets that have been generated with the latest ice model Spice FTP-v3m [148, 149], as presented in Appendix A.3. These reconstructions show even better performance, however, due to limited training data in the high energy region, they are not used in this analysis. This is an important improvement for future analyses in IceCube.

In conclusion, the ML-based reconstructions presented in this chapter provide faster and more accurate predictions of several muon properties than traditional reconstruction methods. They form the basis for the subsequent analysis steps, including event selection and the final physics analysis. Furthermore, these reconstructions can also be utilized in other IceCube analyses that require accurate muon property reconstructions.

7 Selection

In this chapter, the selection of leading muons is presented. It starts at the official IceCube *Level 2*. All further levels are designed to perform the selection used in this thesis. The naming is not standardized, but given by the author. In *Level 3*, the IceCube *MuonFilter* is applied. This is followed by *Level 4* in which a machine learning based pre-cut is performed to select high-energetic muons. Furthermore, machine learning based reconstructions of energies, directions and track geometries are performed. In *Level 5*, quality cuts are applied to improve the data–MC agreement and to reduce the neutrino background. At last, a minimum leadingness is required to improve the data–MC agreement of the leading muon energy reconstruction, referred to as *Final Level*.

7.1 Level 3 – Filter

The goal of this thesis is to unfold the atmospheric muon flux at the surface of the IceCube detector. Hence, muons need to be selected. There are several standard filters in IceCube that are usually performed at an early stage of the selection because they are easy and fast to apply. In this section, several filters have been studied to select the highest energetic muons with the largest efficiency. They are described below with a focus on the *MuonFilter*, which has been found to be the most efficient one and therefore further used in this thesis. In this thesis, filters with a focus on high energy muons are investigated. Here, the *MuonFilter* is explained since it is used for this analysis. A detailed description of the other filters can be found in Appendix B.1.

MuonFilter The *MuonFilter* selects events based on reconstructed track quality and charge brightness, using criteria that depend on the reconstructed zenith angle. It requires three inputs: a cleaned pulse series map, a reconstructed track, and the log-likelihood fit parameters associated with the track. From the pulse series map, the number of hit DOMs N_{ch} and the total charge Q_{tot}, defined as the sum of the pulse charges over all DOMs, is determined. Then, the logarithmic charge $\ell = \log_{10}(Q_{\mathrm{tot}})$ is calculated. The cosine of the reconstructed zenith angle is computed as $\cos(\theta)$, and the per-DOM log-likelihood value is retrieved from the fit parameters.

Events are classified into three angular zones based on the value of $\cos(\theta)$:

- Zone 1 from $-1.0 < \cos(\theta) \leq 0.2$: the filter computes a scaled log-likelihood value $\tilde{\lambda} = \mathrm{rlogl} \cdot \frac{N_{\mathrm{ch}}-5}{N_{\mathrm{ch}}-3}$, and accepts the event if $\tilde{\lambda} \leq 8.7$.
- Zone 2 from $0.2 < \cos(\theta) \leq 0.5$: the filter applies a brightness-based cut and accepts the event if $\ell > 3.9 \cdot \cos\theta + 0.65$.
- Zone 3 from $0.5 < \cos(\theta) \leq 1.0$: a similar brightness cut is used with a shallower slope $\ell > 0.6 \cdot \cos\theta + 2.3$.

The event is accepted if it passes the selection criteria for any of the three zones. This design ensures that events are selected based on fit quality in the near-horizontal region and on brightness in the vertical direction, balancing background suppression with high signal efficiency across the full zenith-angle range.

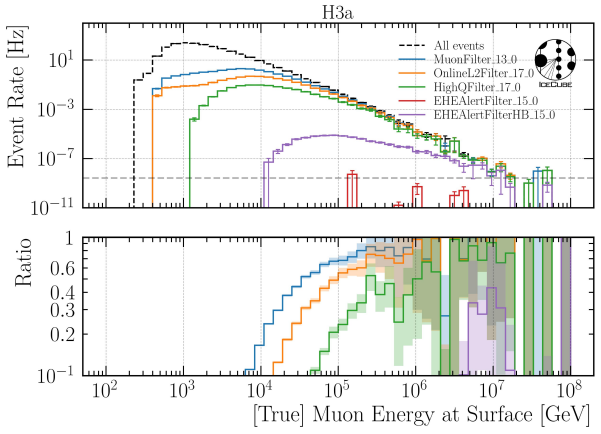
OnlineL2Filter The *OnlineL2Filter* is a higher-level, more nuanced filter tuned for rapid selection of high-quality muon tracks in real time. It also divides events into three $\cos(\theta)$ regions, but within each region, it combines up to three different variables: total charge, a normalized log-likelihood, a direct-hit "elliptical" combination of the number of direct hits and the direct-track length into a logical OR. In the most horizontal region it will keep any event that is sufficiently bright, or has a low normalized likelihood, or lies outside an ellipsoidal acceptance in (direct hits, direct length) space. In the more vertical regions it replaces the direct-hit ellipsis with a $\cos(\theta)$-dependent charge threshold, still retaining the log-likelihood test as a second path. This multi-pronged approach allows to retain a larger fraction of well-reconstructed muons, especially those with ambiguous brightness or fit quality, while still suppressing poorly reconstructed or low-energy backgrounds.

HighQFilter The *HighQFilter* is a simple charge-threshold filter: it retrieves the total event charge Q_{tot} and accepts the event if $Q_{\text{tot}} \geq Q_{\text{min}}$, where Q_{min} is a configurable parameter (default 1000).

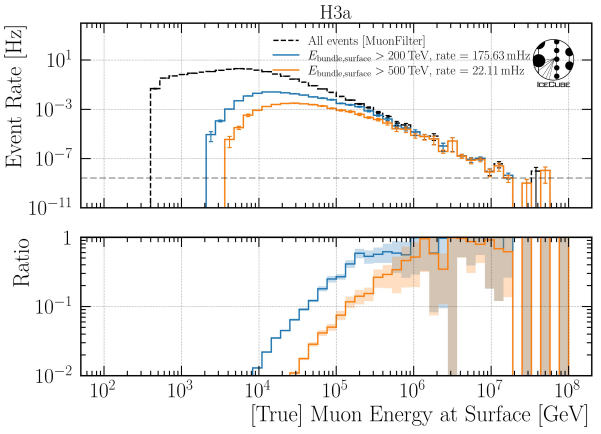
EHEAlertFilter/-HB The *EHEAlertFilter* first ensures that an event is both very bright (high total pulse charge) and well reconstructed (fit-quality within a set range and a minimum number of hit DOMs). Those surviving cuts are then tested in a two-dimensional brightness vs. zenith-angle space: only tracks whose pulse height lies above a direction-dependent boundary are kept. In the heartbeat instantiation (*EHEAlertFilterHB*), this boundary is deliberately relaxed–events that pass the looser criterion but would fail the nominal one are selected–producing a low-rate stream of roughly four alerts per day.

The impact of these filters is studied in Figure 7.1. The event rate of the leading muon energy at surface and the entry of the detector as well as the $\cos(\theta)$ distribution are investigated for the H3a weighting. All other weightings are presented in Figure B.1, Figure B.2 and Figure B.3. For all three observables, the filters behave similar. The two filters *EHEAlertFilter* and *EHEAlertFilterHB* remove the most events over the entire energy and zenith range, followed by the *HighQFilter*. The *OnlineL2Filter* and *MuonFilter* keep the most events. In Figure 7.1e, the shape after the application of the filter is different between the *HighQFilter* and *MuonFilter*. The *HighQFilter* rejects more events towards the horizon. This can be explained due to the requirement of a high charge in the in-ice array. When a muon enters the detector horizontally, it has to pass more medium compared to a vertical entry angle. This causes more energy losses which leads to muons entering the detector below the charge threshold required by the *HighQFilter*. However, the interest lies in the muon flux at surface. Hence, this filter would reject high energy muons at surface which are signal for this analysis. Since this first filter is only used to perform a quick cut and the goal is to keep as many high energetic muons as possible, the *MuonFilter* is chosen.

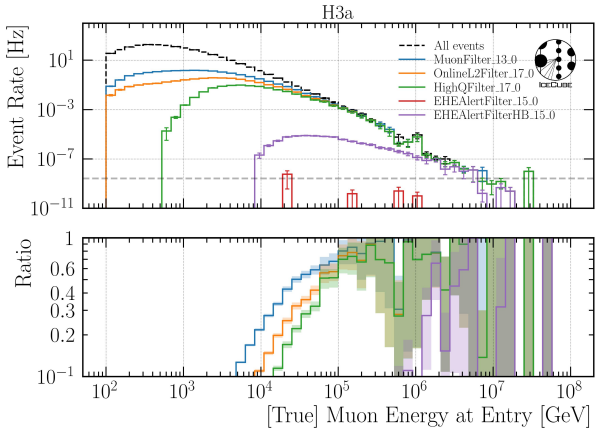
The event rate before the filter is 1341.61 Hz (H3a), after the filter it is 16.26 Hz (H3a). For a livetime of 12 years, this would correspond to $5.1 \cdot 10^{11}$ events before and $6.2 \cdot 10^{9}$ events after the cut. In IceCube, one data-taking process, called *Run*, is 8 h. This would correspond to $3.9 \cdot 10^{7}$ events before and $4.7 \cdot 10^{5}$ events after the cut, per Run. Assuming a processing time of 1 s per event on a single core this processing would take 10 733 h (before filter) and 130 h (after filter), respectively. This is technically not feasible, especially not when 12 years of IceCube data can be utilized. The rates for all primary compositions are presented in Table 7.1. For the rates, it needs to be stated that only primary particles with energies above 10 TeV are simulated and thus considered here.



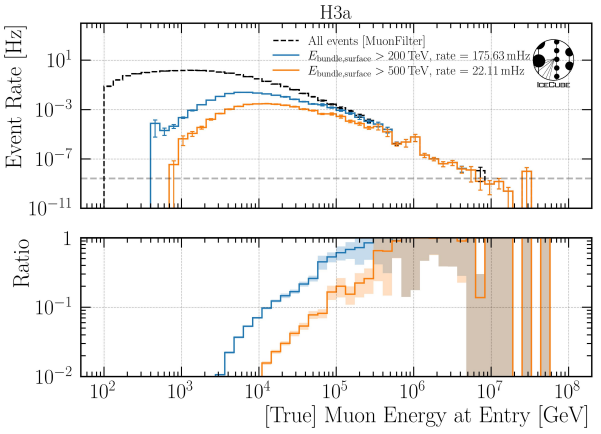
(a) Filter impact on the leading muon energy at surface.



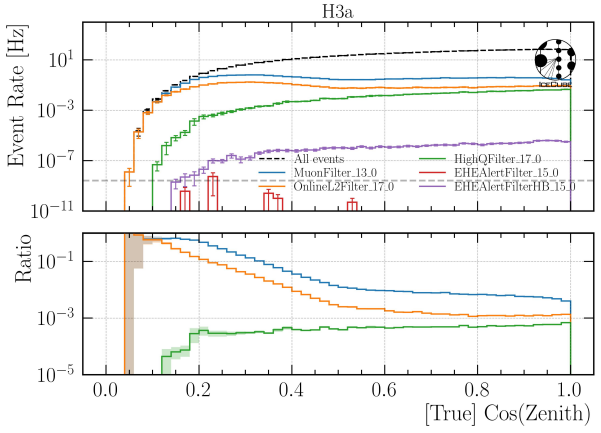
(b) Impact of bundle cut on leading muon energy at surface.



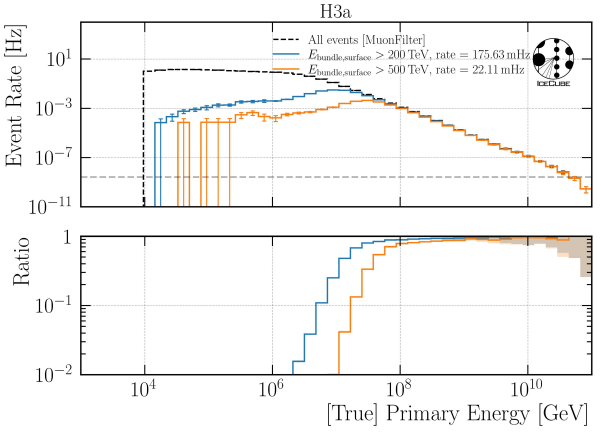
(c) Filter impact on the leading muon energy at entry.



(d) Impact of bundle cut on leading muon energy at entry.



(e) Filter impact on the leading muon zenith distribution.



(f) Impact of bundle cut on primary energy.

Figure 7.1: This figure presents the impact of several IceCube filters and an energy cut on the estimated bundle energy at surface. The distributions are shown for the leading muon energy at surface and entry, its cos(zenith) distribution and the primary energy. Figures are presented for H3a, all other weightings are shown in Section B.1. The ratio is the event rate of selected events divided by the total event rate. Only a subset of the entire MC statistics is used. For this selection, the *MuonFilter* and a 500 TeV bundle energy cut at surface are applied.

Table 7.1: The event rates before and after applying the *MuonFilter* and 500 TeV bundle energy at surface cut (pre-cut) are presented for the four primary weightings. The total number of expected events is presented for 12 years of IceCube data taking after both cuts have been applied. For GSF, $(30-33)$ % less events are expected. Only events with a primary energy above 10 TeV are considered.

CR model	Total [Hz]	MuonFilter [Hz]	Pre-cut [mHz]	Events for 12 years
H3a	1341.61	16.26	22.11	$8.4 \cdot 10^6$
H4a	1370.05	16.58	21.61	$8.2 \cdot 10^6$
GST	1314.99	15.05	21.21	$8.0 \cdot 10^6$
GSF	1357.72	15.72	14.71	$5.6 \cdot 10^6$

7.2 Level 4 – Add Reconstructions and Pre Cut

In *Level 4*, the focus lies on further reducing the event statistics. Since the analysis goal is to detect prompt muons which become the dominant part at energies above PeV, an energy cut to remove low-energetic muons is investigated. For this, two scenarios with the network **precut_surface_bundle_energy_3inputs_6ms_01** are investigated to reject events with a reconstructed muon bundle energy at surface: below 200 TeV and 500 TeV, after the *MuonFilter* is applied. In Figure 7.1, the passing fractions for the leading muon at surface and entry and the primary energy are presented. As expected, the 500 TeV cut removes more events, with a final rate of 22.11 mHz for H3a weighting. The cut keeps most of the events with a leading muon energy at surface of above 1 PeV and the rate at the minimum target energy of 10 TeV is still above 1 mHz. For 12 years, this results in $8.4 \cdot 10^6$ events (H3a).

A complete table with all event rates before and after the filter and bundle cut are presented for all four primary weightings in Table 7.1. It also must be stated that these rates are calculated based on the simulation performed for this thesis. The minimum primary energy starts at 10 TeV, as stated in Section 5.4. Hence, when also muons below this energy threshold would be considered, the rates are obviously even higher. Since the unfolding of the muon flux starts above 10 TeV, this does not lead to any assumptions here, because primaries with lower energies can not generate muons with energies above 10 TeV.

Additionally, the energy, direction and track geometry reconstructions are performed at this stage, as stated in Section 6.2.

7.3 Level 5 – Data-MC Quality Cuts

In *Level 5*, the reconstructions are performed on parts of actual IceCube data and MC simulation. The goal is to investigate whether the reconstructions perform similar on data and MC. In case they do, it can be assumed that the simulations describe the data accurately. Typically, there are some phase spaces where the data and MC do not agree, for example for outlier regions due to assumptions in the simulations. To improve the overall agreement, so-called quality cuts are performed to remove these phase spaces. In the following, several data–MC comparisons are presented for energy, direction, and track geometry reconstructions, and their corresponding uncertainty predictions. In case a quality cut is applied, this is marked by a vertical black line in the plots. The exact cut values are stated in the legend of the bottom panel, and Table 7.3 presents all cuts. These cuts have been performed by eye. In all plots, all four primary weighting models are presented. For the data, 2486.97 h of IceCube data are chosen for the years 2012, 2016 and 2020, referred to as *Burnsample*. For each of the twelve months, the 5th, 15th and 25th day of the month are selected. This is done to have a representative sample of

the data over several years and seasons, since seasonal variations of the muon flux are expected [126, 151, 152]. The presented MC distributions also include the expected atmospheric and astrophysical neutrino fluxes which are assumed to be background in this analysis. The atmospheric neutrino fluxes are weighted respectively for all four primary weightings using `MCEq`. The astrophysical neutrino flux is weighted using a single powerlaw

$$\frac{\mathrm{d}N_\nu}{\mathrm{d}E_\nu} = n_\nu \cdot \left(\frac{E_\nu}{100\,\mathrm{TeV}}\right)^{-\gamma_\nu}, \tag{7.1}$$

with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86]. The lines are presented in pink with a dashed linestyle. However, since the impact of the different primary compositions on the atmospheric neutrino flux is negligible here, the lines are all plotted in the same color.

By comparing the total number of simulated events with the number of events in the data, a total offset in the normalization is observed. To match the offset between data and MC to compare the shape of the distributions, a scaling factor $n_{\mathrm{s},\,i}$ for the four primary weightings $i \in \{\mathrm{H3a}, \mathrm{H4a}, \mathrm{GST}, \mathrm{GSF}\}$ is calculated and applied to the MC distributions. This scaling factor is defined as

$$n_{\mathrm{s},\,i} = \frac{N_{\mathrm{data}}}{N_{\mathrm{MC},\,i}}, \tag{7.2}$$

where N_{data} is the number of events in the data, and $N_{\mathrm{MC},\,i} = \sum_j w_j \cdot t$ is the sum of the weights over all simulated events j multiplied by the lifetime t, here 2486.97 h for the Burnsample statistics. The scaling factors are presented in Table 7.2 and are calculated for each level. As stated in Section 8, the normalization has no impact on the unfolding results.

Table 7.2: Scaling factors $n_{\mathrm{s},\,i}$ calculated with Equation (7.2) are presented for the four primary weightings and selection stages. The factor is applied to MC to match a normalization offset to the data. The *Level 4* factor is applied after the *MuonFilter* and a 500 TeV bundle energy cut at surface, the *Level 5* factor is applied after quality cuts, and the *Final Level* factor is applied after the leadingness cut.

Primary weighting	Scaling factor $n_{\mathrm{s},\,i}$		
	Level 4	Level 5	Final Level
H3a	0.86	0.82	1.12
H4a	0.88	0.84	1.09
GST	0.89	0.84	1.32
GSF	1.32	1.28	1.09

In the following, the data–MC comparisons are presented for the direction, track geometry and energy reconstructions in the Figures 7.2, 7.3, 7.4, 7.5, 7.6, 7.7, 7.8, and 7.9.

At first, the directional reconstructions presented in Figure 7.2 are discussed. The azimuth distribution is flat, as expected, with a good data–MC agreement for all primary compositions. The $\cos(\theta)$ distributions agrees best with data for GST, showing a slight slope. Towards the horizon, i. e. $\cos(\theta) \to 0$, the distribution is dominated by neutrinos. This is expected because a particle has to traverse the Earth and thus lots of dense medium, until it can reach the detector. For the high energy muon selection, almost no muon makes it through the Earth while surviving the pre-cut. To reduce the neutrino background in the sample, a cut on the zenith angle is applied at $\cos(\theta) > 0.2$. The uncertainty predictions for both the zenith and azimuth prediction seem to agree for low uncertainties, but in the data, there are more outliers with higher uncertainties, which are cut away.

Second, the center track geometry reconstructions presented in Figure 7.3 and 7.4 are discussed. The center position x and y distributions show a good data–MC agreement for $x, y \in [-550\,\mathrm{m}, 550\,\mathrm{m}]$.

Table 7.3: Overview of all applied cuts to build the leading muon sample. Cuts are categorized into classification, uncertainty, containment and neutrino veto to select leading muons with sufficient data-MC agreement. These cuts are applied after the *MuonFilter* and the 500 TeV bundle energy cut at surface.

Cut Type	Feature	$>$	$<$
Containment	length in detector	1000 m	2000 m
	entry pos x, y	−750 m	750 m
	entry pos z	−500 m	750 m
	center pos x, y	−550 m	550 m
	center pos z	−650 m	650 m
Neutrino Veto	$\cos(\theta)$	0.2	
	length	5000 m	15000 m
Classification	leadingness	0.4	
Uncertainty	bundle energy at entry		$0.9 \log_{10}(\text{GeV})$
	bundle energy at surface		$2.0 \log_{10}(\text{GeV})$
	zenith		0.1 rad
	azimuth		0.2 rad
	entry pos x, y, z		42 m
	center pos x, y, z		50 m
	entry pos time		200 ns
	center pos time		600 ns
	length in detector		160 m
	length		2000 m

This corresponds to the actual size of the detector, which is about 1 km in diameter. However, the light is already measured by the detector, when the particle is a few hundred meters away from the detector. The uncertainty predictions agree up to 50 m. Overall, all primary compositions lead to similar distributions. The reconstruction of the center position z shows a slope for all four compositions. At low depths around −500 m, there are more data events than expected from MC, and at higher depths around 500 m it becomes vice versa. Several correlations have been investigated between the center position z and the bundle energy at surface, and the zenith angle. However, no correlation has been found. It is known that the ice properties change with the depth [148]. The reconstructions utilized in this analysis have been trained on MC data with several ice models, except of the latest one. This is caused by the lack of MC statistics for the latest ice model at the time of the network training. The latest ice model, which is also used for the MC in this thesis, includes a better description of the ice properties. For testing this impact, another model has been trained on newer ice models, and the same data–MC plot for the center position z has been performed, as presented in Figure A.4c. The agreement is better, which confirms the impact of the ice model on the z position reconstruction. Since there are not enough high energy events to train all networks on the latest ice model, in this analysis, the networks trained on the older ice models are utilized. However, since the z position is never used as an observable for the unfolding, this does not impact the unfolding results.

The uncertainty predictions also become worse above 50 m. The center time agrees worse than the spatial coordinates. This can be caused by a not exactly simulated time in the MC events. Hence, no cut is applied on this feature. Instead, the uncertainty prediction works reliably, and a cut is applied at 600 ns.

Third, the same track geometry reconstructions are evaluated for the entry positions, as presented in Figure 7.5 and 7.6. The entry position x and y distributions show a good data–MC agreement for $x, y \in [-750\,\text{m}, 750\,\text{m}]$. The uncertainty predictions agree up to 42 m. The entry position z shows a slope for all four compositions. At low depths around −500 m, there are more data than MC, and at higher depths around 500 m, there are more MC than data. The same ice model impact as for the

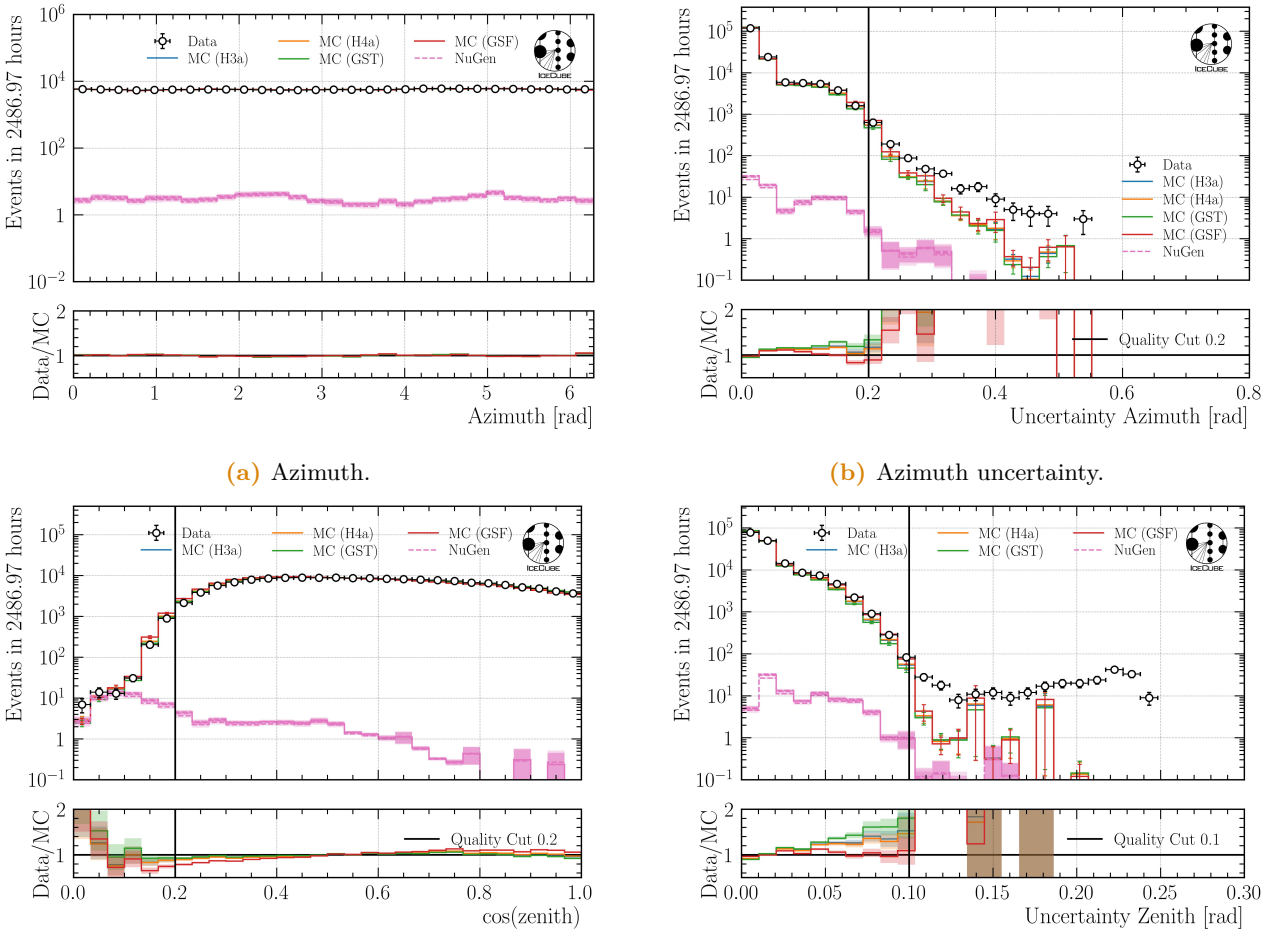
center position z is observed here. The uncertainty predictions also become worse above 42 m. Similar to the center time, for the entry time it follows that the agreement to the data is worse than for the spatial coordinates. A quality cut is applied at 200 ns, since the uncertainty predictions work reliably. The entry and center times are not used as an observable for the unfolding, hence this does not impact the unfolding results.

Fourth, the track length reconstructions are evaluated in Figure 7.7. The track length within the detector agrees best for values from $[1000\,\mathrm{m}, 2000\,\mathrm{m}]$. This corresponds to the simulated cylinder size with height $h = 1400\,\mathrm{m}$ and radius $r = 700\,\mathrm{m}$. The diagonal for this cylinder results to $\sqrt{(1400\,\mathrm{m})^2 + (2 \cdot 700\,\mathrm{m})^2} \approx 1980\,\mathrm{m}$. With the lower bound of 1000 m, the best data–MC agreement is achieved by through-going tracks that deposit energy losses along the track over a minimum range. The uncertainty cut is applied at 160 m. Additionally, the total track length is evaluated. The agreement is sufficient for values from $[5000\,\mathrm{m}, 15\,000\,\mathrm{m}]$. The uncertainty predictions agree up to 2000 m. The prediction is worse for the total track length than for the track length within the detector. This is expected, since there is no possibility to measure or detect the total track length of the particle, caused by the finite size of the detector. However, the reconstruction is still able to predict the total track length with a good agreement. This cut also reduces the neutrino background, caused by horizontal neutrino tracks passing lots of dense medium.

Fifth, the energy reconstructions at surface are evaluated in Figure 7.8. In total, the bundle energy at surface, reconstructed by the pre-cut network and the energy network, and the leading muon energy at surface prediction are presented. In all figures, the data–MC agreement depends on the primary CR composition. For the pre-cut network, H3a, H4a and GSF agree better than GST, which starts to disagree for energies above 4 PeV. The same observable, the bundle energy at surface, reconstructed by the energy network, shows a good data–MC agreement only for GSF, and worse agreements for the other three compositions. The energy network is trained only on events that have passed the pre-cut, hence the training data impact these results. For the leading muon energy at surface, none of the four models agrees well with the data. For the uncertainty predictions, the results are similar. Only one cut is applied here. The uncertainty prediction for the bundle energy at surface is cut at $2.0 \log_{10}(\mathrm{GeV})$, to remove outliers with an uncertain energy prediction, since this network is used to apply the pre-cut. In the data–MC comparisons for the energy network, the neutrino background starts to dominate for energies above $\sim 10\,\mathrm{PeV}$. This can be explained by the fact that these neutrinos are mainly coming from the horizon. Hence, when a particle enters the detector horizontally with a certain energy, the network predicts its surface bundle energy very high, since the networks have never been trained on neutrino events, but only on atmospheric air showers. This is validated in Figure B.13c, where the same prediction is presented after horizontal events with $\cos(\theta) < 0.2$ have been removed. However, for the pre-cut network, the neutrinos are never predicted to be the dominant component of the presented energy range. Hence, the training on low energy events is helpful for the neutrino energy prediction. Since none of the surface energy estimators are used for the actual analysis, a data–MC mismatch has no impact on the results.

At last, the energy reconstructions at the detector entry are evaluated in Figure 7.9. For the leading muon energy at entry, none of the primary models agrees with data for energies above $\sim 20\,\mathrm{TeV}$, similar to the surface reconstruction. For the reconstruction of the bundle energy at entry, also only GSF agrees with the data. However, no quality cut is applied on these reconstructions.

All data–MC comparisons on *Level 5* after the application of the quality cuts are presented in Appendix B.3.



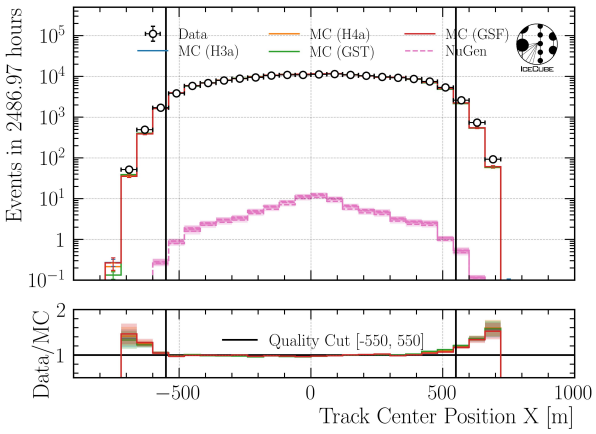
(a) Azimuth.

(b) Azimuth uncertainty.

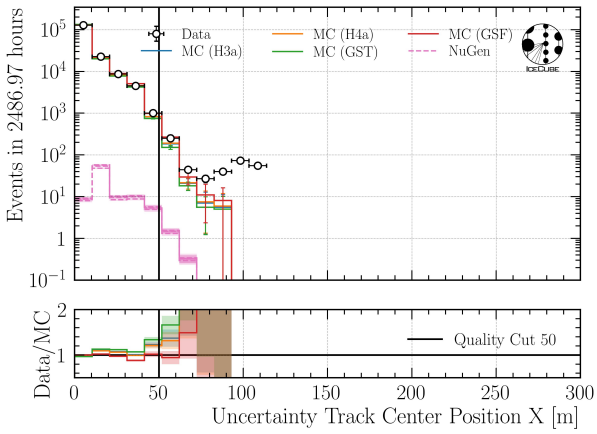
(c) $\cos(\theta)$.

(d) $\cos(\theta)$ uncertainty.

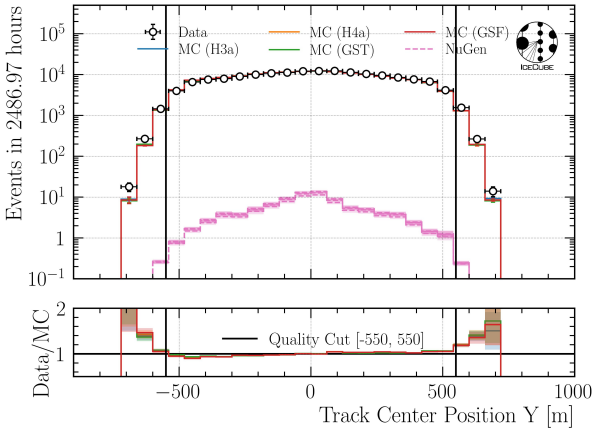
Figure 7.2: This figure presents the agreement between 2486.97 h of IceCube data and MC for the directional reconstructions. Figures include four MC predictions, one for each primary weighting. The ratio presents the number of events on data divided by the expected MC events. These comparisons are performed after the *MuonFilter* and a 500 TeV bundle energy cut at surface are applied. Vertical black lines indicate the cuts applied to reach *Level 5* to select well reconstructed events. The neutrino background (NuGen) is weighted with an atmospheric weighting respectively, and the astrophysical component is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86]. The MC distributions include both the muon and neutrino flux.



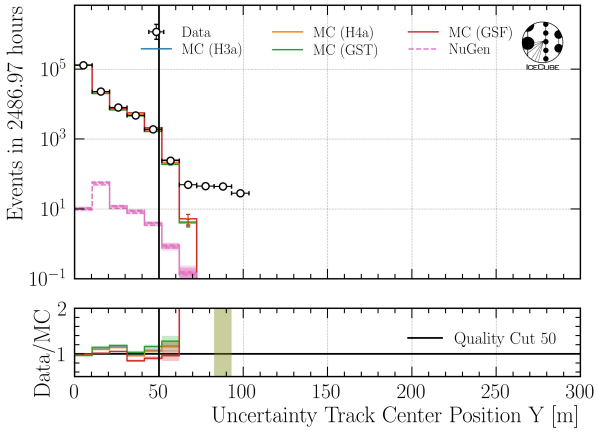
(a) Center position x.



(b) Center position x uncertainty.

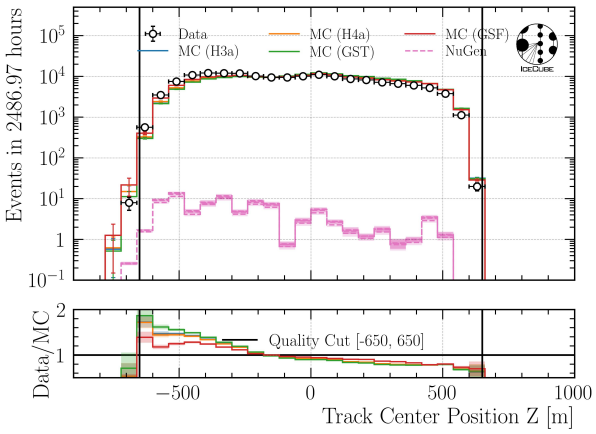


(c) Center position y.

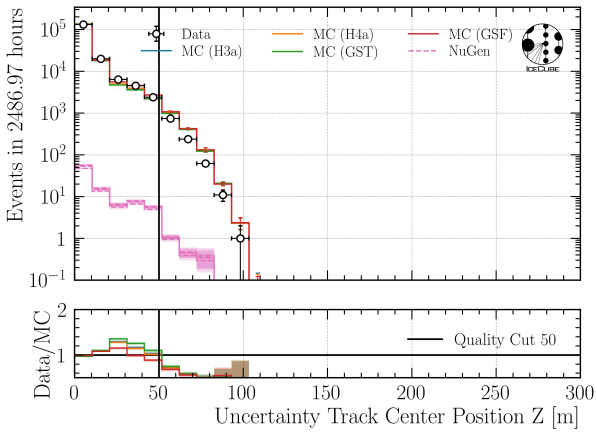


(d) Center position y uncertainty.

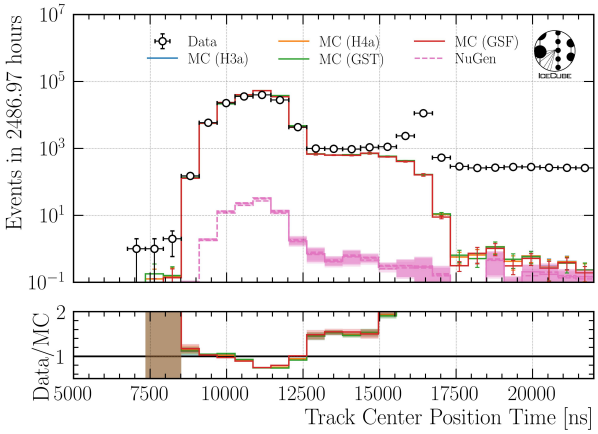
Figure 7.3: This figure presents the agreement between 2486.97 h of IceCube data and MC for the center position x and y. Figures include four MC predictions, one for each primary weighting. The ratio presents the number of events on data divided by the expected MC events. These comparisons are performed after the *MuonFilter* and a 500 TeV bundle energy cut at surface are applied. Vertical black lines indicate the cuts applied to reach *Level 5* to select well reconstructed events. The neutrino background (NuGen) is weighted with an atmospheric weighting respectively, and the astrophysical component is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86]. The MC distributions include both the muon and neutrino flux.



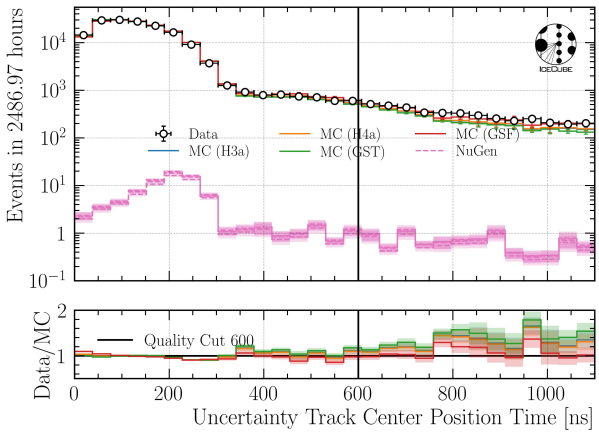
(a) Center position z.



(b) Center position z uncertainty.

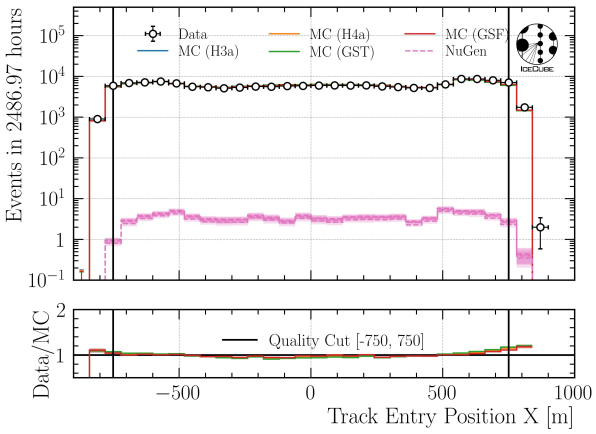


(c) Center position t.

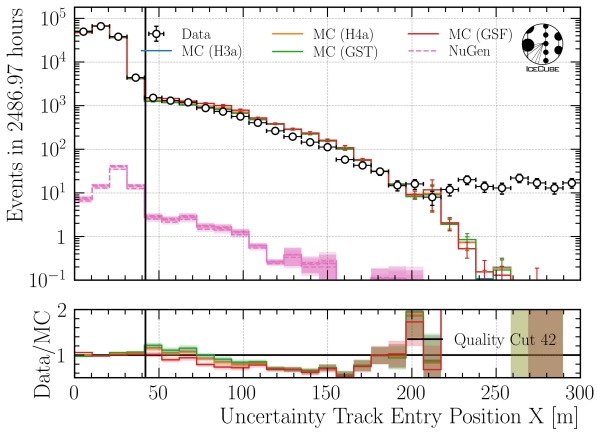


(d) Center position t uncertainty.

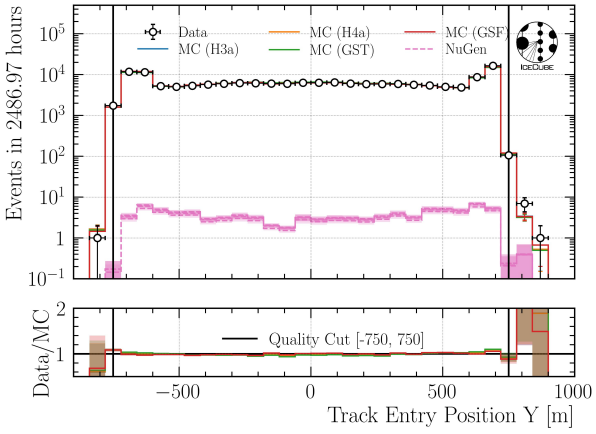
Figure 7.4: This figure presents the agreement between 2486.97 h of IceCube data and MC for the center position z and t. Figures include four MC predictions, one for each primary weighting. The ratio presents the number of events on data divided by the expected MC events. These comparisons are performed after the *MuonFilter* and a 500 TeV bundle energy cut at surface are applied. Vertical black lines indicate the cuts applied to reach *Level 5* to select well reconstructed events. The neutrino background (NuGen) is weighted with an atmospheric weighting respectively, and the astrophysical component is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86]. The MC distributions include both the muon and neutrino flux.



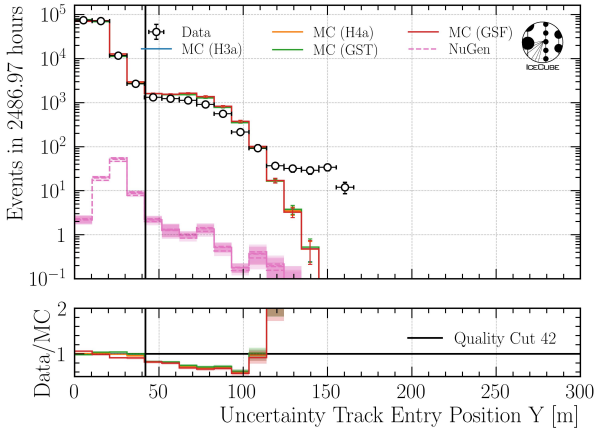
(a) Entry position x.



(b) Entry position x uncertainty.

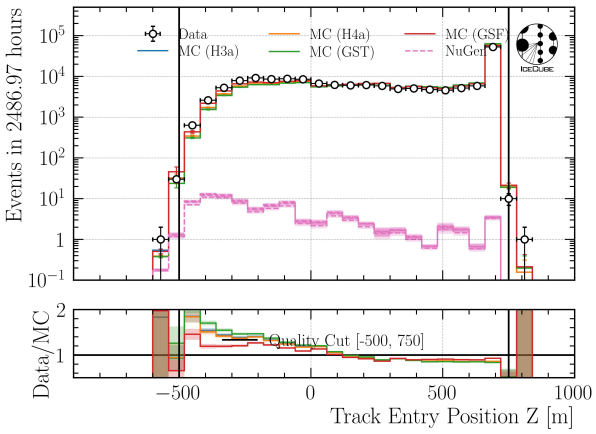


(c) Entry position y.

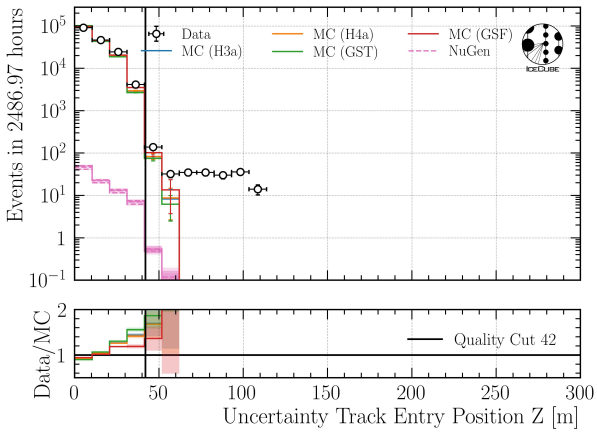


(d) Entry position y uncertainty.

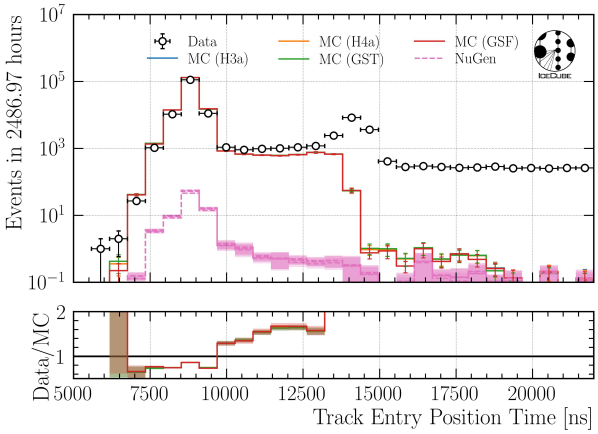
Figure 7.5: This figure presents the agreement between 2486.97 h of IceCube data and MC for the entry position x and y. Figures include four MC predictions, one for each primary weighting. The ratio presents the number of events on data divided by the expected MC events. These comparisons are performed after the *MuonFilter* and a 500 TeV bundle energy cut at surface are applied. Vertical black lines indicate the cuts applied to reach *Level 5* to select well reconstructed events. The neutrino background (NuGen) is weighted with an atmospheric weighting respectively, and the astrophysical component is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86]. The MC distributions include both the muon and neutrino flux.



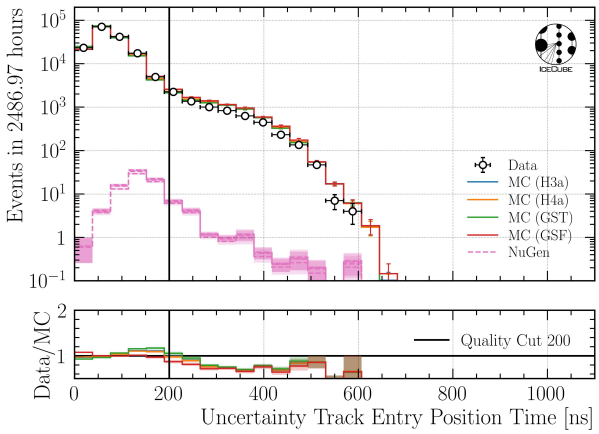
(a) Entry position z.



(b) Entry position z uncertainty.

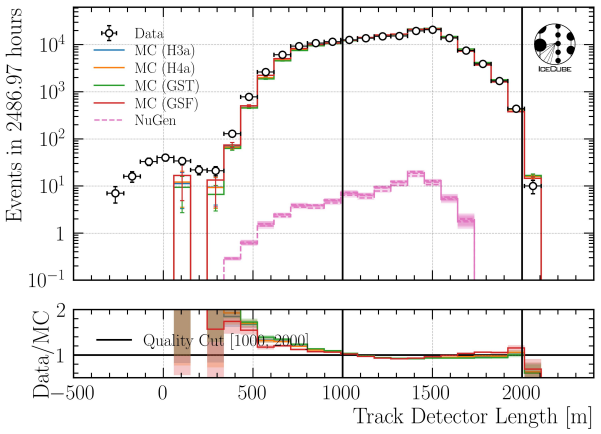


(c) Entry position t.

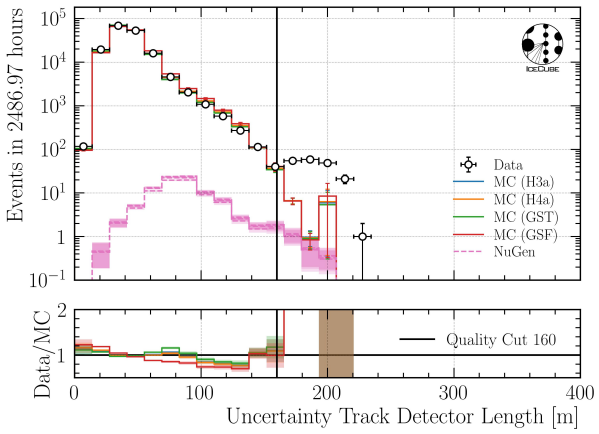


(d) Entry position t uncertainty.

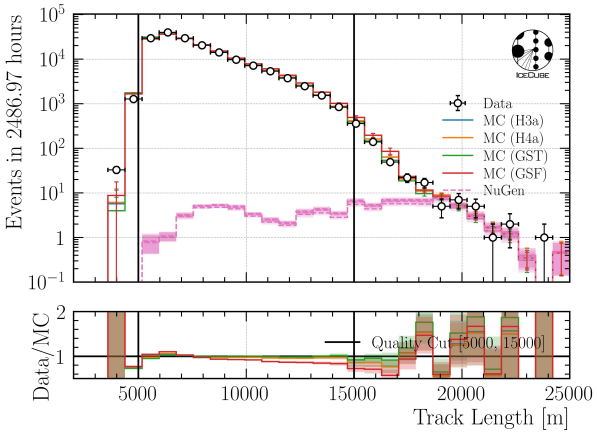
Figure 7.6: This figure presents the agreement between 2486.97 h of IceCube data and MC for the entry position z and t. Figures include four MC predictions, one for each primary weighting. The ratio presents the number of events on data divided by the expected MC events. These comparisons are performed after the *MuonFilter* and a 500 TeV bundle energy cut at surface are applied. Vertical black lines indicate the cuts applied to reach *Level 5* to select well reconstructed events. The neutrino background (NuGen) is weighted with an atmospheric weighting respectively, and the astrophysical component is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86]. The MC distributions include both the muon and neutrino flux.



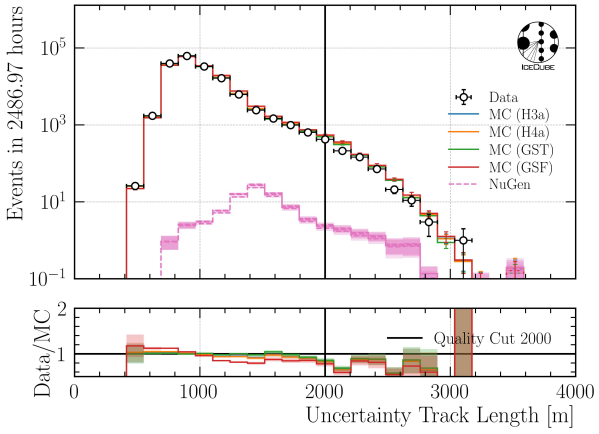
(a) Track length in detector.



(b) Track length in detector uncertainty.

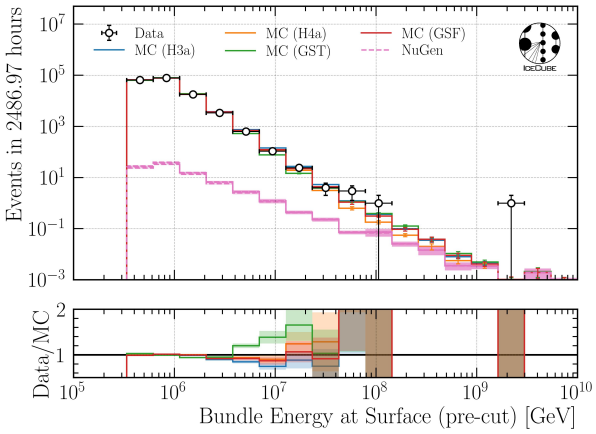


(c) Track length.

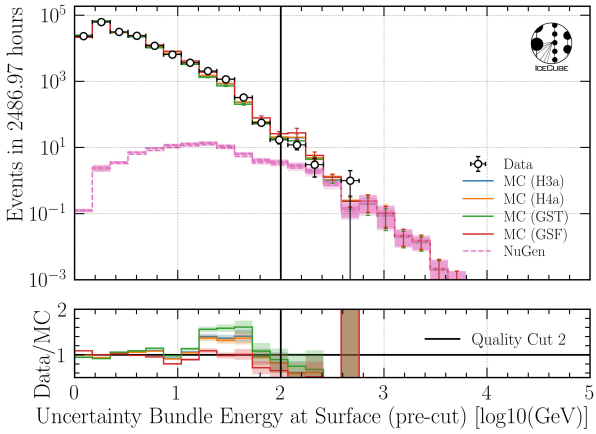


(d) Track length uncertainty.

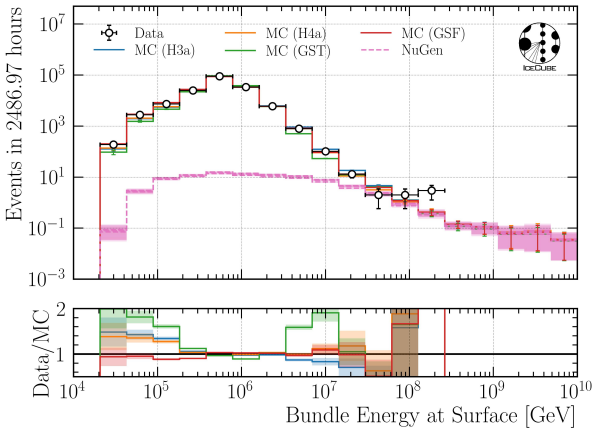
Figure 7.7: This figure presents the agreement between 2486.97 h of IceCube data and MC for the track lengths. Figures include four MC predictions, one for each primary weighting. The ratio presents the number of events on data divided by the expected MC events. These comparisons are performed after the *MuonFilter* and a 500 TeV bundle energy cut at surface are applied. Vertical black lines indicate the cuts applied to reach *Level 5* to select well reconstructed events. The neutrino background (NuGen) is weighted with an atmospheric weighting respectively, and the astrophysical component is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86]. The MC distributions include both the muon and neutrino flux.



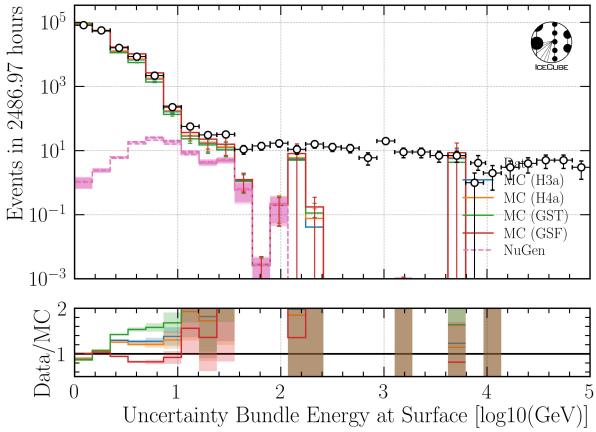
(a) Pre-cut bundle energy at surface.



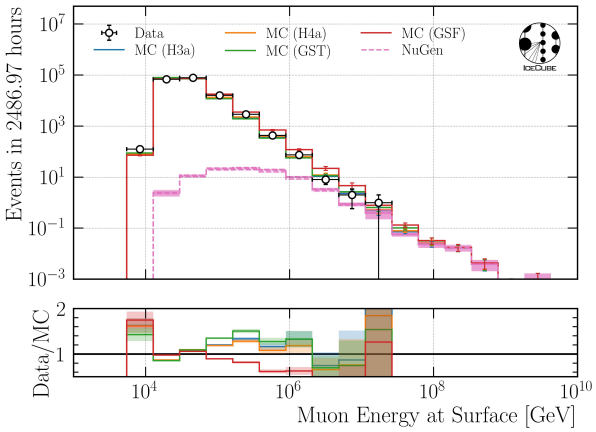
(b) Pre-cut bundle energy at surface uncertainty.



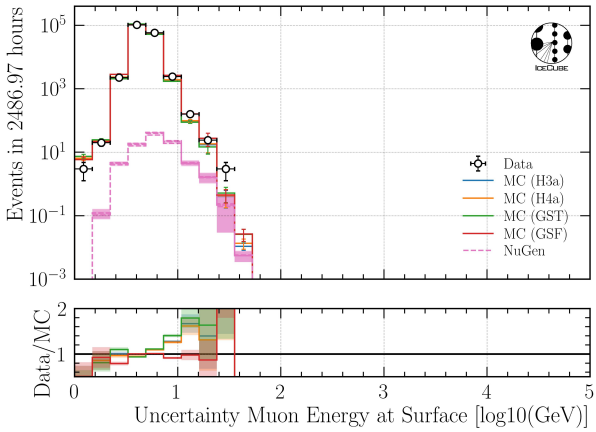
(c) Bundle energy at surface.



(d) Bundle energy at surface uncertainty.

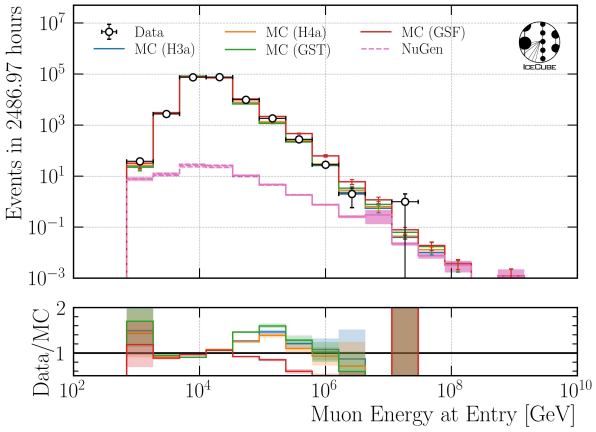


(e) Leading muon energy at surface.

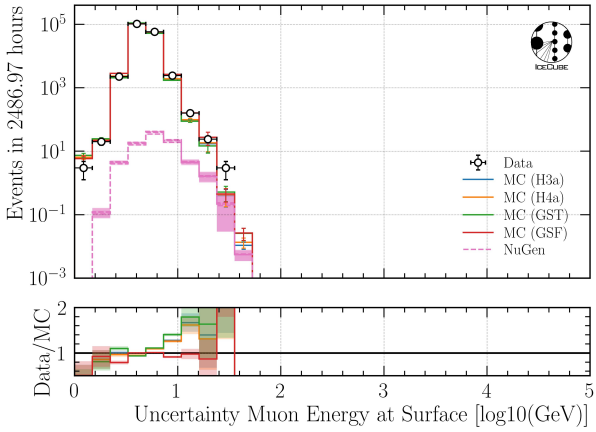


(f) Leading muon energy at surface uncertainty.

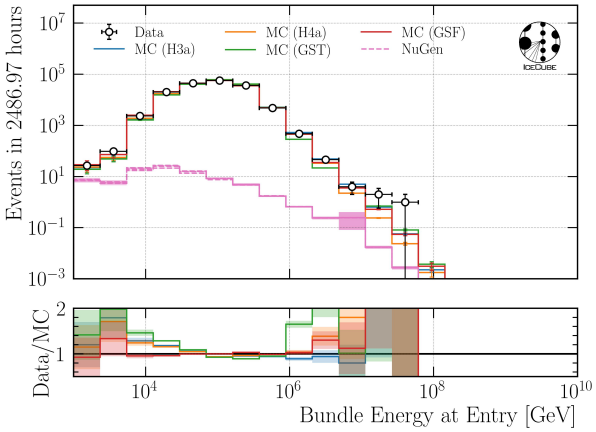
Figure 7.8: This figure presents the agreement between 2486.97 h of IceCube data and MC for the energz reconstructions at surface. Figures include four MC predictions, one for each primary weighting. The ratio presents the number of events on data divided by the expected MC events. These comparisons are performed after the *MuonFilter* and a 500 TeV bundle energy cut at surface are applied. Vertical black lines indicate the cuts applied to reach *Level 5* to select well reconstructed events. The neutrino background (NuGen) is weighted with an atmospheric weighting respectively, and the astrophysical component is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86]. The MC distributions include both the muon and neutrino flux.



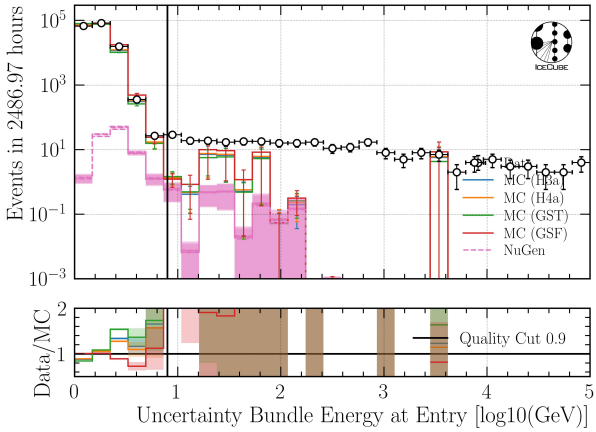
(a) Leading muon energy at entry.



(b) Leading muon energy at entry uncertainty.



(c) Bundle energy at entry.



(d) Bundle energy at entry uncertainty.

Figure 7.9: This figure presents the agreement between 2486.97 h of IceCube data and MC for the energy reconstructions at entry. Figures include four MC predictions, one for each primary weighting. The ratio presents the number of events on data divided by the expected MC events. These comparisons are performed after the *MuonFilter* and a 500 TeV bundle energy cut at surface are applied. Vertical black lines indicate the cuts applied to reach *Level 5* to select well reconstructed events. The neutrino background (NuGen) is weighted with an atmospheric weighting respectively, and the astrophysical component is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86]. The MC distributions include both the muon and neutrino flux.

7.4 Final Level – Leadingness Cut

The *Final Level* is the last selection stage of this analysis. An additional cut on the leadingness L defined as

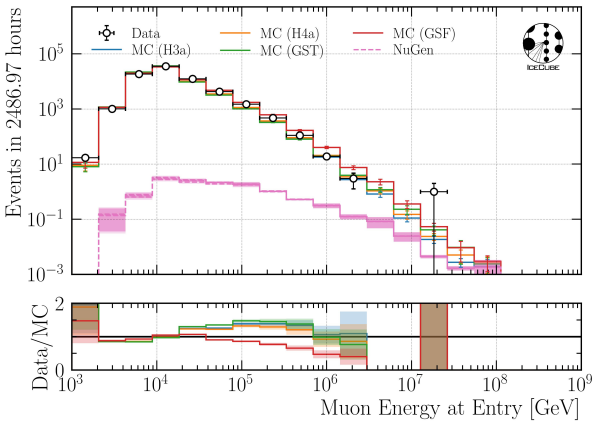
$$L = \frac{E_{\text{leading}}}{E_{\text{bundle}}}\,, \tag{7.3}$$

at the detector entry is applied. This cut is introduced to improve the data–MC agreement of the reconstructed leading muon energy at the detector entry since this observable is used for the unfolding. The cut is chosen to optimize the data–MC agreement of the leading muon energy at entry, while keeping as many events as possible. In Figure 7.10, the impact of the leadingness cuts $L \in [0.1, 0.6]$ are presented for the leading muon energy at entry. A low cut of $L > 0.1$ leads to a good data–MC agreement for energies from 2 TeV to 20 TeV. Towards higher energies, there are less MC events than data for the predictions of H3a, H4a and GST, and more MC events than data for GSF. With a cut of $L > 0.3$, the predictions for H3a, H4a and GST agree with data for the entire energy spectrum covered by the Burnsample statistics, within the uncertainties. However, GSF still shows a divergence starting around 200 TeV. The distributions for the cuts $L \in \{0.4, 0.5, 0.6\}$ are similar within the uncertainties, but the number of events decreases. Thus, for the *Final Level*, a leadingness cut of $L > 0.4$ is chosen, evaluated by eye. The cuts $L \in [0.7, 0.9]$ for the leading muon energy at entry and $L \in [0.3, 0.5]$ for the bundle energy at entry are presented in Figure B.23. For tighter cuts, there is no significant change in the distributions. Also, the bundle energy at entry distributions are similar for the three cuts $L \in \{0.3, 0.4, 0.5\}$.

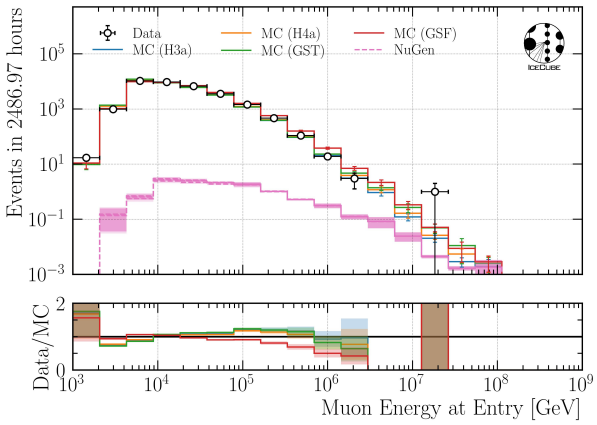
The scaling factors for all leadingness cuts are stated in Table 7.4. Then, Figure B.25 shows the scaling factors as a function of the leadingness cut. The scaling factor for GST increases towards higher leadingness cuts, while the scaling factors for the other three models behave similar above $L > 0.4$. For lower cuts, GSF behaves contrary to the other three models. Again, primary energies below 10 TeV are not considered here.

In Figure B.26, the impact of three different leadingness cuts $L \in \{0.2, 0.4, 0.6\}$ on the reconstructed leading muon energy at entry and the true leading muon energy at entry is presented. These variables are used as proxy and target variable in the unfolding. The left column of the figures shows a cut applied to the true leadingness, while the right column shows a cut applied to the reconstructed leadingness. With increasing leadingness cut, the low energy muons are removed mainly. With the fact of a power law distribution of cosmic rays and thus muon energies, this explains the impact of the leadingness cut on the scaling factor. Furthermore, with a higher leadingness cut, the reconstruction resolution improves because the leading muon becomes more dominant within the muon bundle. Thus, a larger part of the light detected by the DOMs is coming from the leading muon. The proxy distribution shows a similar behavior for the true and reconstructed leadingness cut. However, for the target distribution, the reconstructed leadingness cut removes more high energy muons compared to the true leadingness cut. This is a consequence of the reconstruction resolution of the leading and bundle energy at entry. As presented in Figure 6.9a, the bundle energy underestimates the true energy towards higher energies, while the leading energy overestimates the true energy, see Figure 6.9b. In principle, this effect cancels out in the leadingness, however, the reconstructed leadingness underestimates the true leadingness, as shown in Figure B.24.

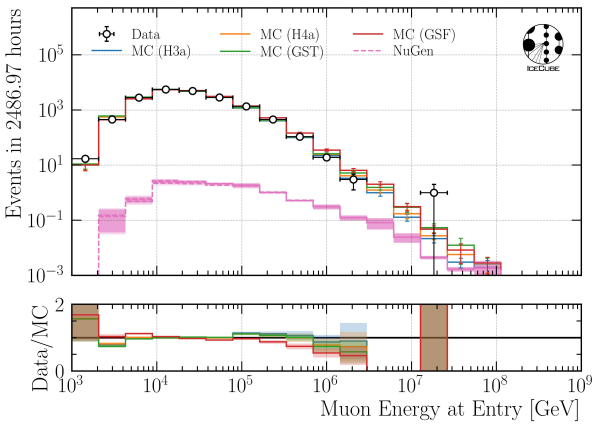
A list of all data–MC comparisons on the *Final Level* after the leadingness cut of $L > 0.4$ is applied is presented in Appendix B.4.



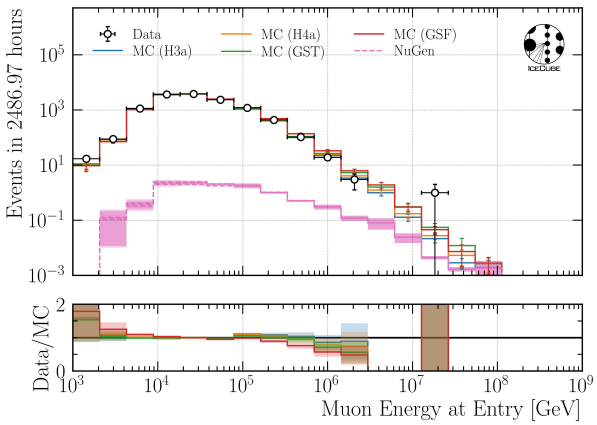
(a) $L > 0.1$.



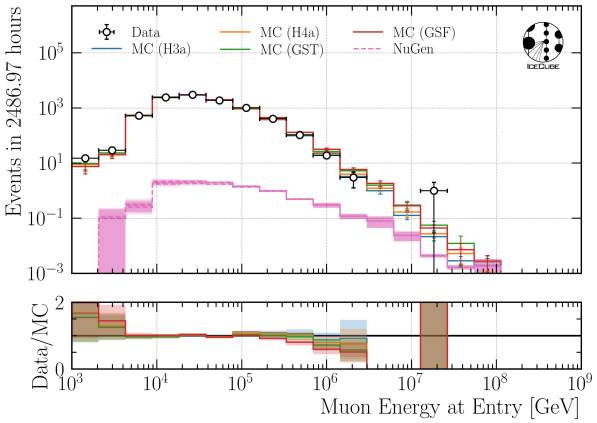
(b) $L > 0.2$.



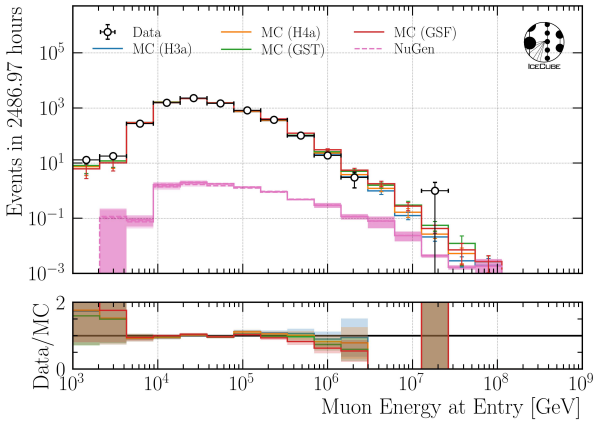
(c) $L > 0.3$.



(d) $L > 0.4$.

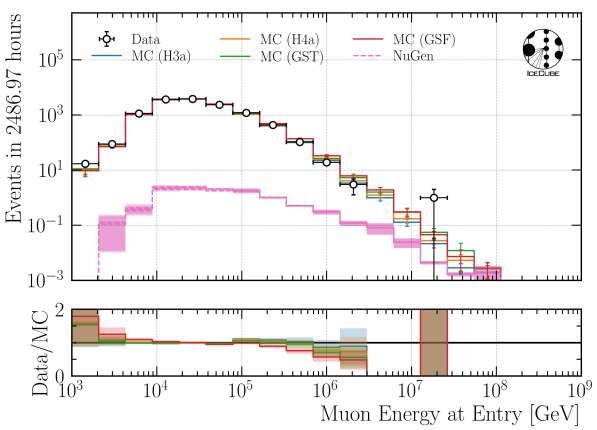


(e) $L > 0.5$.

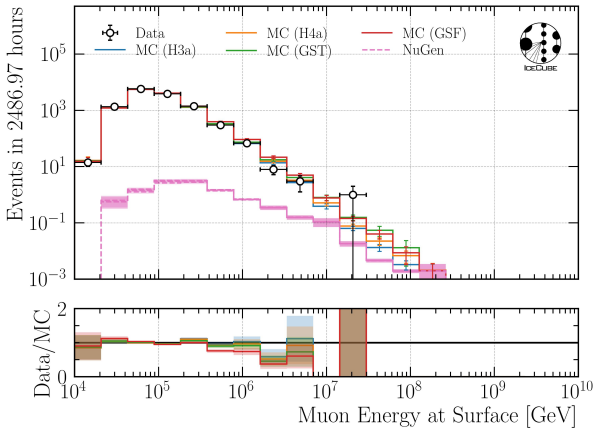


(f) $L > 0.6$.

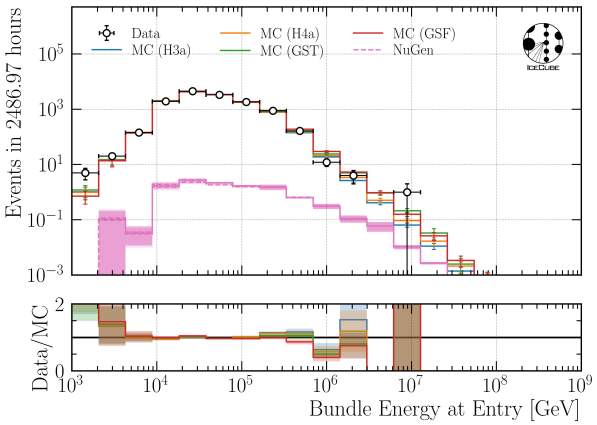
Figure 7.10: This figure presents the agreement between 2486.97 h of IceCube data and MC for the leading muon energy reconstructions at entry. Figures include four MC predictions, one for each primary weighting. The ratio presents the number of events on data divided by the expected MC events. These comparisons are performed on potential final levels after all cuts have been applied for several leadingness cuts. The neutrino background (NuGen) is weighted with an atmospheric weighting respectively, and the astrophysical component is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86]. The MC distributions include both the muon and neutrino flux.



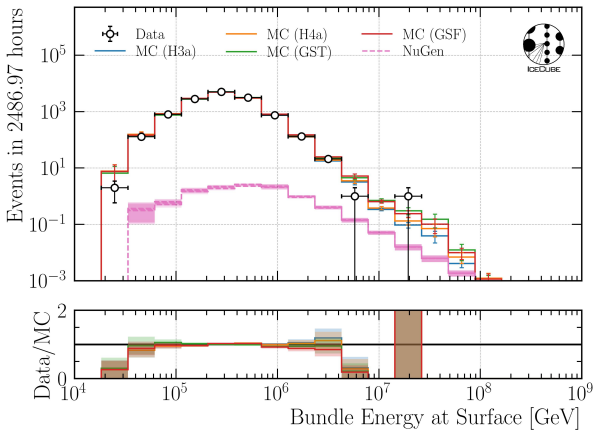
(a) Leading muon energy at entry.



(b) Leading muon energy at surface.



(c) Bundle energy at entry.



(d) Bundle energy at surface.

Figure 7.11: This figure presents the agreement between 2486.97 h of IceCube data and MC for the energy reconstructions at entry and at surface. Figures include four MC predictions, one for each primary weighting. The ratio presents the number of events on data divided by the expected MC events. These comparisons are performed on the *Final Level* after all cuts have been applied. The neutrino background (NuGen) is weighted with an atmospheric weighting respectively, and the astrophysical component is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86]. The MC distributions include both the muon and neutrino flux.

Table 7.4: Scaling factors $n_{\mathrm{s},\,i}$ calculated with Equation (7.2) are presented for the four primary weightings and different leadingness cuts L. The factor is applied to MC to match a normalization offset to the data. The factors are applied to *Level 5* after quality cuts, and the application of the leadingness cut.

Primary weighting	Scaling factor $n_{\mathrm{s},\,i}$								
Leadingness	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
H3a	0.89	1.01	1.10	1.12	1.12	1.11	1.11	1.11	1.13
H4a	0.90	1.01	1.08	1.09	1.08	1.07	1.07	1.07	1.08
GST	0.94	1.13	1.27	1.32	1.34	1.35	1.38	1.39	1.42
GSF	1.30	1.22	1.13	1.09	1.06	1.03	1.04	1.01	1.02

7.5 Summary: Selection

In this chapter, a new selection for high-energy leading muons is presented, which is designed specifically for this analysis. It utilizes the new machine learning-based reconstructions introduced in Section 6 and is based on the new CORSIKA simulations described in Section 5. The selection proceeds through several levels, starting from the standard IceCube *Level 2* filter. First, the *MuonFilter* is applied, followed by a bundle energy cut $E_{\mu,\mathrm{surf}}^{\mathrm{bundle}} > 500\,\mathrm{TeV}$ to select high-energy events. Subsequently, a set of 23 quality cuts is applied to ensure a high purity of the sample and a good agreement between data and simulation. These cuts include requirements on containment, neutrino veto, and reconstruction uncertainty, as listed in Table 7.3. Finally, a cut on the leadingness is performed to improve the reconstruction of the leading muon energy at entry, which serves as the proxy variable for the unfolding. This selection chain reduces the initial event rate of approximately 2.89 kHz to a final rate of 1.29 mHz for H3a. The selection utilizes a burnsample comprising 2486.97 h of data, distributed equally over the years 2012, 2016, and 2020 to account for seasonal variations, to evaluate the agreement between data and MC. An offset in the overall normalization is observed, which depends on the primary cosmic ray model and the leadingness cut, as studied in Figure B.25. However, since the unfolding method relies on the shape of the distribution rather than its absolute normalization, this offset does not affect the final results. The final sample achieves a good agreement between data and simulation for the proxy variable used in the unfolding, as shown in Figure 7.11a.

8 Unfolding of the atmospheric muon flux

This chapter presents the measurement of the atmospheric muon energy spectrum at the surface using 12.12 years of IceCube data. The measurement employs a regularized likelihood unfolding technique to reconstruct the true physical distribution from the measured observables by inverting the detector response.

The chapter is structured as follows: Section 8.1 introduces the mathematical principles of the unfolding method. Section 8.2 details the validation of the procedure on a subset of the data, the *burnsample*. Section 8.3 presents the application to the full dataset. This analysis utilizes an expanded MC dataset and a refined treatment of systematic uncertainties compared to the initial validation. A consistency check using the burnsample with this updated configuration is presented, followed by the unfolding of the complete 12.12-year dataset. Finally, the findings are summarized in Section 8.4.

8.1 Unfolding Basics

This section outlines the fundamental concepts and the mathematical framework of the unfolding method used in this analysis. The core principle, a regularized matrix inversion, is explained in Section 8.1.1. Subsequently, Section 8.1.2 describes the effective area calculation, which serves as a sensitivity correction to translate the unfolded event counts into a physical flux at the surface. The technical implementation is performed using the Python package `funfolding`, introduced in Section 8.1.3. Finally, the unfolding procedure is visually demonstrated in Section 8.1.4, illustrating the correlation between variables, the conversion to a flux, and the impact of regularization.

8.1.1 Method

Experiments aim to determine the distribution of a physical quantity $f(x)$, referred to as the *target* distribution. However, this distribution is often not directly accessible. Instead, experiments measure a set of variables y, which are correlated with the quantity of interest. These variables are referred to as *observables* and represent the response of the detector to the underlying physical truth. *Unfolding* describes the method of inverting this measurement process to estimate the most probable true distribution $f(x)$ from the measured distribution $g(y)$, referred to as the *proxy* distribution.

The true distribution $f(x)$ is related to the observable distribution $g(y)$ via the *Fredholm integral equation* [153]:

$$g(y) = \int A(x, y) f(x) \, \mathrm{d}x + b(y). \tag{8.1}$$

The detector response function $A(x, y)$ describes the specific detection process and is typically obtained from MC simulations. It incorporates the probability of observing a value y given a true value x. In the discrete case, the matrix $\mathbf{A}$ is constructed by binning the simulated events into a two-dimensional histogram of the observable y versus the true variable x. Each column j is then normalized to the total number of generated events in that true energy bin, such that A_{ij} represents the probability for an event with true parameter x_j to be reconstructed in the observable bin y_i. The term $b(y)$ accounts for background contributions. For sufficiently pure samples, such as those used in this analysis (see

Sections 8.2.8 and 8.2.9), the background term can be neglected. In practice, the continuous variables x and y are discretized, transforming the integral into a sum:

$$g_i = \sum_j A_{ij} f_j \qquad \text{or} \qquad \mathbf{g} = \mathbf{A} \cdot \mathbf{f}. \tag{8.2}$$

Assuming a Poisson distribution for the event counts in each bin and taking the negative logarithm allows for the construction of a likelihood function:

$$\ell(\mathbf{f}|\mathbf{g}) = -\ln \mathcal{L}(\mathbf{f}|\mathbf{g}) = \sum_i \Big((\mathbf{A} \cdot \mathbf{f})_i - g_i \ln \left((\mathbf{A} \cdot \mathbf{f})_i \right) \Big). \tag{8.3}$$

The most probable distribution $\mathbf{f}$, given the measurements $\mathbf{g}$, is found by minimizing Equation 8.3.

Unfolding is an *ill-posed problem*. Small statistical fluctuations in the measurement $\mathbf{g}$ can be amplified during the inversion, often resulting in unphysical oscillations in the solution. To mitigate this, a common approach is to apply *Tikhonov regularization* [154]. This method favors a smooth target distribution by introducing a penalty term that suppresses large second derivatives of the solution vector $\mathbf{f}$. The second derivative is approximated via the *finite difference method* as $\mathbf{f}'' = \mathbf{C}\mathbf{f}$, with the regularization matrix $\mathbf{C}$ defined as:

$$\mathbf{C} = \begin{bmatrix} 1 & -2 & 1 & & & & \\ 0 & 1 & -2 & & & & \vdots \\ & & & \ddots & & & \\ \vdots & & & & -2 & 1 & 0 \\ & & & & 1 & -2 & 1 \end{bmatrix}. \tag{8.4}$$

The complete negative log-likelihood function then reads:

$$\ell(\mathbf{f}|\mathbf{g}) = \sum_i \Big((\mathbf{A} \cdot \mathbf{f})_i - g_i \ln \left((\mathbf{A} \cdot \mathbf{f})_i \right) \Big) + \mathcal{R}(\mathbf{f}), \tag{8.5}$$

with the *Tikhonov regularization* term:

$$\mathcal{R}(\mathbf{f}) = \frac{1}{2} \tau^{-1} \, \mathbf{f}^\top \, \mathbf{C}^\top \, \mathbf{C} \, \mathbf{f}. \tag{8.6}$$

The parameter τ determines the regularization strength. A small τ imposes a strong penalty, enforcing a smoother spectrum, whereas a large τ leaves the initial likelihood predominantly unchanged. Following Ref. [153], the optimal τ is chosen to minimize the total correlation ρ between the components of the solution vector $\mathbf{f}$. This correlation is defined as the sum of the off-diagonal elements V_{ij} of the covariance matrix $\mathbf{V}$:

$$\rho = \sum_{i>j} V_{ij}. \tag{8.7}$$

The minimization of ρ serves as a robust criterion to balance the bias-variance trade-off. In the regime of weak regularization (large τ), the solution is dominated by statistical fluctuations. Due to the smearing of the detector response, a statistical excess in one bin is often compensated by a deficit in the neighboring bin during the inversion, leading to strong negative correlations (oscillations). Conversely, strong regularization (small τ) suppresses the curvature of the solution, forcing adjacent bins to be similar and inducing positive correlations. The optimal τ is therefore found where the global average correlation is minimized, effectively decoupling the bins as much as the detector resolution allows. It is determined via a grid search over a predefined range of τ. The unfolding is performed for each τ value in the grid, and the corresponding correlation ρ is evaluated. Crucially, this optimization is performed directly on the experimental data. This distinguishes the method from traditional approaches where the regularization strength is often tuned on MC simulations to minimize the deviation from the true generator distribution. Optimizing on MC implicitly assumes that the simulation accurately describes

the data, potentially biasing the result towards the theoretical prediction. In contrast, minimizing the bin-to-bin correlation is a data-driven approach that reduces model dependency by selecting the regularization strength based on the internal properties of the measured dataset itself.

The uncertainties of the unfolded spectrum are derived from the inverse Hessian matrix evaluated at the minimum $\hat{\mathbf{f}}$. Since the minimization is performed on the negative log-likelihood $\ell(\mathbf{f})$, the Hessian provides a quadratic approximation of ℓ around the minimum, serving as an estimator for the covariance matrix of the unfolded bin counts:

$$V(\mathbf{f}) = \mathbf{H}^{-1}(\hat{\mathbf{f}}) = \left(\frac{\partial^2 \ell(\mathbf{f})}{\partial f_i \, \partial f_j} \right)^{-1} \Bigg|_{\mathbf{f}=\hat{\mathbf{f}}} . \tag{8.8}$$

This approximation is strictly valid only if the negative log-likelihood is well-described by a quadratic (Gaussian) shape near the minimum. This requires sufficiently high statistics per bin and may break down in low-count regimes or when numerous systematic nuisance parameters are fitted simultaneously. In such cases, the likelihood curvature becomes non-Gaussian, and the Hessian may underestimate uncertainties or become numerically unstable [155, 156].

Systematic uncertainties are incorporated into the solution by modeling them as nuisance parameters. Using the SnowStorm method employed in IceCube [113], the bin counts in observable space are parameterized by linear weighting functions $w_{j,i}$, which quantify the impact of varying a systematic parameter $j \in \{\text{Absorption}, \text{Scattering}, \text{DOM efficiency}, \text{HoleIce p0}, \text{HoleIce p1}\}$ on proxy bin i. These are integrated into the detector response matrix:

$$\mathbf{A}' = \mathbf{w}_j^{\top} \cdot \mathbf{A}, \tag{8.9}$$

allowing for the simultaneous minimization of the likelihood with respect to these additional parameters.

Figure 8.1 illustrates the smearing of the true distribution f into the observable space g, described by the detector response matrix A. Unfolding inverts this process to recover the true distribution. The unphysical oscillations, visualized below in Figure 8.2, which are often observed in unregularized unfolding, arise from the ill-posed nature of the problem. This instability is increased by the discretization of the variables: events from a single target bin are distributed across multiple proxy bins due to the detector's finite resolution. Consequently, inverting this process amplifies small statistical fluctuations in the measured data into large, unphysical variations in the target space.

8.1.2 Sensitivity correction / Effective Area

The sensitivity of the detector is quantified by the effective area. For a simple one-dimensional toy model with an energy-independent detector area, the effective area can be written as

$$A_{\text{eff}} = A_{\text{det}} \frac{N_{\text{det}}}{N_{\text{gen}}}, \tag{8.10}$$

where A_{det} is the geometric detection area and N_{det} (N_{gen}) denote the number of detected (generated) events in a given true-energy bin obtained from MC simulation.

In this analysis, the situation is more involved, because the projected detection area of the IceCube in-ice array depends on the zenith angle θ of the incoming muon. Approximating the fiducial volume as a cylinder with radius r and height l, the projected area is

$$A_{\text{det}}(\theta) = \pi r^2 \cos\theta + 2rl \sin\theta \,, \tag{8.11}$$

with $r = 700\,\text{m}$ and $l = 1400\,\text{m}$ in the simulation.

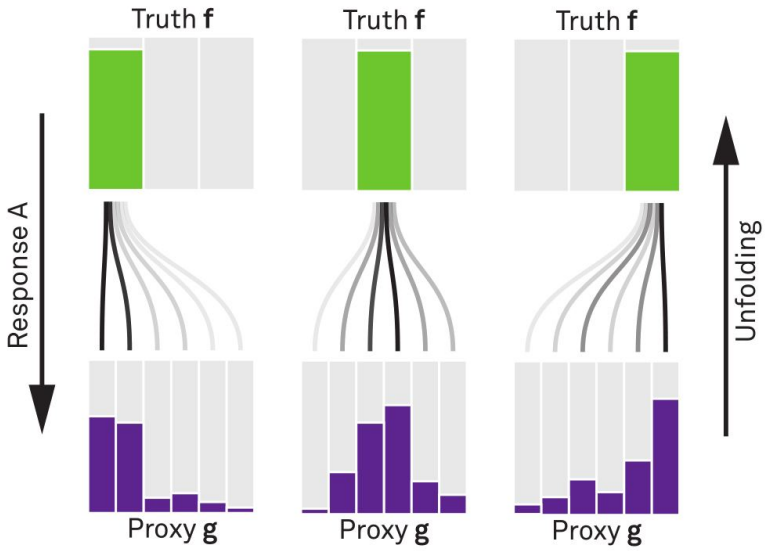


Figure 8.1: This sketch visualizes the distribution of true values f into several proxy bins g via the detector response matrix A. The unfolding inverts this process to determine the true distribution based on the measured property. In this thesis, the true distribution is the individual muon energy at the Earth's surface, and the measured proxy is the reconstructed energy of the leading muon at the entry of the detector.

The effective area is therefore constructed in two dimensions, binned in true muon energy at surface $E_{\mu,i}$ and true $\cos\theta_j$. For each bin (i, j), the MC provides *weighted* numbers of generated and selected events,

$$N_{ij}^{\text{gen}} = \sum_{k\in(i,j)} w_k^{\text{gen}}, \tag{8.12}$$

$$N_{ij}^{\text{sel}} = \sum_{k\in(i,j)} w_k^{\text{sel}}, \tag{8.13}$$

where w_k^{gen} and w_k^{sel} are the generator and selection weights of event k, respectively. The corresponding variances of the weighted sums are estimated as

$$\text{Var}\left(N_{ij}^{\text{gen}}\right) = \sum_{k\in(i,j)} \left(w_k^{\text{gen}}\right)^2, \tag{8.14}$$

$$\text{Var}\left(N_{ij}^{\text{sel}}\right) = \sum_{k\in(i,j)} \left(w_k^{\text{sel}}\right)^2. \tag{8.15}$$

The selection efficiency in each (i, j) bin is then given by

$$\varepsilon_{ij} = \frac{N_{ij}^{\text{sel}}}{N_{ij}^{\text{gen}}}, \tag{8.16}$$

with variance obtained by standard error propagation,

$$\text{Var}(\varepsilon_{ij}) = \frac{\text{Var}\left(N_{ij}^{\text{sel}}\right)}{\left(N_{ij}^{\text{gen}}\right)^2} + \frac{\left(N_{ij}^{\text{sel}}\right)^2 \text{Var}\left(N_{ij}^{\text{gen}}\right)}{\left(N_{ij}^{\text{gen}}\right)^4}. \tag{8.17}$$

Multiplying the selection efficiency by the projected area yields the 2D effective area per bin, denoted as $A_{\text{eff},ij}$:

$$A_{\text{eff},ij} = A_{\text{det}}(\theta_j)\,\varepsilon_{ij}, \qquad \text{Var}(A_{\text{eff},ij}) = A_{\text{det}}^2(\theta_j)\,\text{Var}(\varepsilon_{ij}). \tag{8.18}$$

For the unfolding, an energy-dependent effective area $A_{\text{eff},i}$ is required. It is obtained by averaging $A_{\text{eff},ij}$ over the zenith angle, with weights proportional to the generated MC distribution:

$$N_i^{\text{gen}} = \sum_j N_{ij}^{\text{gen}}, \tag{8.19}$$

$$p_{ij} = \frac{N_{ij}^{\text{gen}}}{N_i^{\text{gen}}}, \tag{8.20}$$

$$A_{\text{eff},i} = \sum_j p_{ij}\, A_{\text{eff},ij}, \tag{8.21}$$

and the statistical uncertainty on the effective area is propagated as

$$\text{Var}(A_{\text{eff},i}) = \sum_j p_{ij}^2 \,\text{Var}(A_{\text{eff},ij}), \qquad \sigma_{A_{\text{eff},i}} = \sqrt{\text{Var}(A_{\text{eff},i})}. \tag{8.22}$$

The unfolded event rate in bin i is finally converted to a differential flux via

$$\Phi_i = \frac{N_i}{T\,\Omega_i\, A_{\text{eff},i}\, \Delta E_i}, \tag{8.23}$$

where N_i is the unfolded number of events in the i-th true-energy bin, T is the livetime, Ω_i is the solid angle corresponding to the bin, and ΔE_i is the energy-bin width.

Systematic uncertainties on the effective area

In addition to the statistical uncertainties, the effective area is subject to systematic uncertainties related to the detector and ice properties (e.g. optical efficiency, absorption and scattering in the ice). These effects are modeled using the SnowStorm parameterization, as introduced in Section 5.3, where each event carries values of the underlying systematic parameters s_k together with an associated weight.

For each SnowStorm parameter s_k, with central (nominal) value $s_{k,0}$, the MC sample is split into two subsamples:

$$\text{“low” variation:} \quad s_k < s_{k,0}, \tag{8.24}$$

$$\text{“high” variation:} \quad s_k > s_{k,0}. \tag{8.25}$$

To preserve the overall normalization, the selection weights of the events in the subsamples are renormalized by a factor of 2. This renormalization is necessary because the simulation is split into two independent subsets based on the systematic parameter range. For each subsample, a separate effective area is computed using exactly the same procedure as for the baseline:

$$A_{\text{eff},i}^{\text{low},k},\ \sigma_{A_{\text{eff},i}}^{\text{low},k} = \text{from events with } s_k < s_{k,0}, \tag{8.26}$$

$$A_{\text{eff},i}^{\text{high},k},\ \sigma_{A_{\text{eff},i}}^{\text{high},k} = \text{from events with } s_k > s_{k,0}. \tag{8.27}$$

Here, i labels the true-energy bin and the statistical uncertainties $\sigma_{A_{\text{eff},i}}^{\text{low/high},k}$ are obtained from the MC weights as described in Eqs. (8.17)–(8.22).

Let $A_{\text{eff},i}^{\text{nom}}$ and $\sigma_{A_{\text{eff},i}}^{\text{nom}}$ denote the baseline effective area and its statistical uncertainty. For each systematic parameter k and energy bin i, the effect of the systematic is quantified by the envelope of the low and

high variations around the nominal value,

$$\Delta_i^{(k,\mathrm{low})} = A_{\mathrm{eff},i}^{\mathrm{nom}} - A_{\mathrm{eff},i}^{\mathrm{low},k}, \tag{8.28}$$

$$\Delta_i^{(k,\mathrm{high})} = A_{\mathrm{eff},i}^{\mathrm{high},k} - A_{\mathrm{eff},i}^{\mathrm{nom}}, \tag{8.29}$$

$$\Delta_i^{(k)} = \max\left(\left|\Delta_i^{(k,\mathrm{low})}\right|, \left|\Delta_i^{(k,\mathrm{high})}\right|\right). \tag{8.30}$$

Assuming the different systematic parameters are independent, the total systematic uncertainty on the effective area in bin i is obtained by adding the individual contributions in quadrature,

$$\sigma_{\mathrm{sys},i}^2 = \sum_k \left(\Delta_i^{(k)}\right)^2. \tag{8.31}$$

Finally, the statistical baseline uncertainty and the systematic contributions are combined as

$$\sigma_{A_{\mathrm{eff},i}}^{\mathrm{total}} = \sqrt{\left(\sigma_{A_{\mathrm{eff},i}}^{\mathrm{nom}}\right)^2 + \sigma_{\mathrm{sys},i}^2}, \tag{8.32}$$

where only the statistical uncertainty of the baseline effective area is included explicitly. Since all SnowStorm variations are derived from the same underlying MC sample, the additional statistical fluctuations in the low/high subsamples are highly correlated with the baseline and are therefore not added as an independent contribution.

8.1.3 Python Tool funfolding

The unfolding in this work is performed using the Python package `funfolding`, a modular framework for Poisson-likelihood unfolding developed at TU Dortmund University [156, 157]. It implements the regularized Poisson likelihood discussed in Section 8.1.1, specifically the negative log-likelihood defined in Eq. (8.5) augmented with a second-derivative Tikhonov penalty term [154]. Beyond the core likelihood minimization, the package provides tools for handling detector systematic uncertainties via nuisance parameters and supports flexible binning schemes.

In this analysis, the minimization is performed deterministically using `Minuit`, as provided by the `iminuit` interface [158, 159]. The optimization procedure begins with the robust, derivative-free SIMPLEX algorithm to locate a reliable starting point. Subsequently, the solution is refined using MIGRAD, a variable-metric method that ensures rapid convergence near the minimum. If asymmetric parameter uncertainties are required, profile likelihood scans can be performed using MINOS.

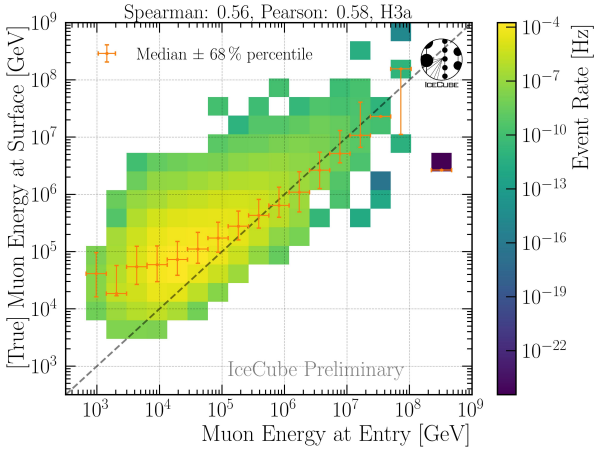
The regularization strength τ is determined by scanning over a predefined range and selecting the value that minimizes the global bin-to-bin correlation of the unfolded spectrum, as described in Section 8.1. This criterion balances the suppression of statistical fluctuations (oscillations) against the bias introduced by smoothing.

Systematic uncertainties are incorporated using the SnowStorm method [113] (see Section 5.3). `funfolding` implements this by modifying the response matrix with per-bin weighting functions that depend on the nuisance parameters. The minimization is then performed simultaneously over the physics parameters $\mathbf{f}$ and the nuisance parameters, ensuring that systematic correlations are correctly propagated into the covariance of the result. Physical constraints, such as non-negativity ($f_j \geq 0$), can be enforced to guarantee meaningful solutions.

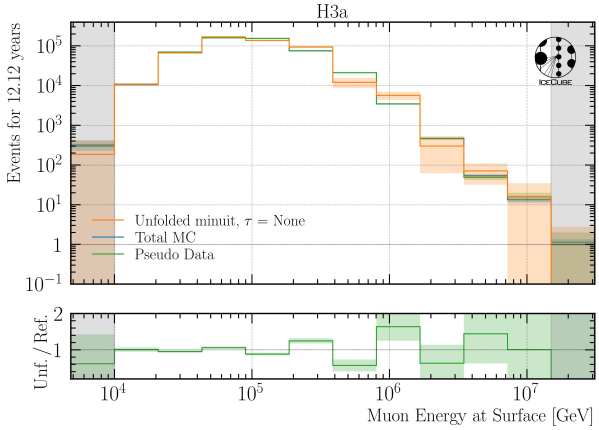
The primary inputs to `funfolding` are the detector response matrix $\mathbf{A}$, the observed data vector $\mathbf{g}$, the regularization matrix $\mathbf{C}$, and the set of systematic templates. The software returns the unfolded spectrum $\hat{\mathbf{f}}$ and its covariance matrix $\mathbf{V}(\hat{\mathbf{f}})$, derived from the inverse Hessian at the minimum (see Eq. (8.8)). Methodologically, `funfolding` adheres to the classical unfolding formulation by Blobel [155] while offering a modern, extensible Python interface tailored for the inclusion of continuous systematic variations.

8.1.4 Visualization of the Unfolding Principle

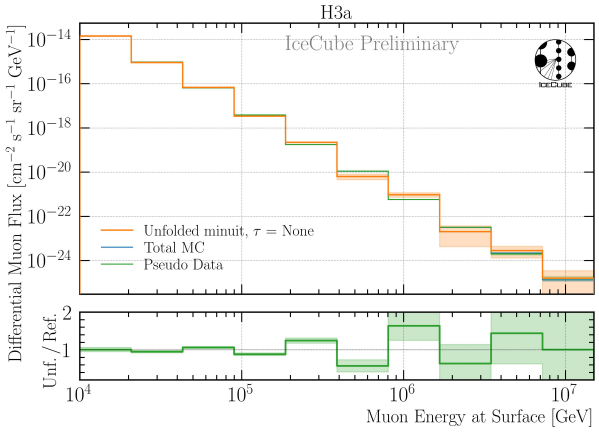
The unfolding procedure is summarized in Figure 8.2. First, the correlation between the target variable, being the true muon energy at the surface, and the proxy variable, which is the reconstructed leading muon energy at the detector entry, is shown in Figure 8.2a. The Pearson and Spearman correlation coefficients are $\rho_{\mathrm{P}} = 0.58$ and $\rho_{\mathrm{S}} = 0.56$, respectively, indicating a positive but moderate correlation due to the stochastic energy losses of muons in the ice. Figure 8.2b presents an example unfolding of a so-called *pseudo dataset*. A pseudo dataset is created by sampling events from the weighted MC simulations to mimic a specific physical distribution (e.g., H3a) for a defined livetime. This approach allows for the validation of the unfolding procedure using known truth information before applying it to experimental data. In this example, the distribution corresponds to a livetime of 12 years. The top panel displays the expected event count, while the bottom panel shows the ratio of the unfolded result to the true distribution of the pseudo dataset. Without regularization, the result exhibits significant unphysical oscillations. In Figure 8.2c, the unfolded event rate is converted into a physical flux using the effective area. Since no regularization is applied, the flux inherits the oscillatory behavior. Finally, Figure 8.2d demonstrates the effect of Tikhonov regularization with a strength of $\tau = 0.0005$. The resulting unfolded flux is smooth and reproduces the true underlying distribution.



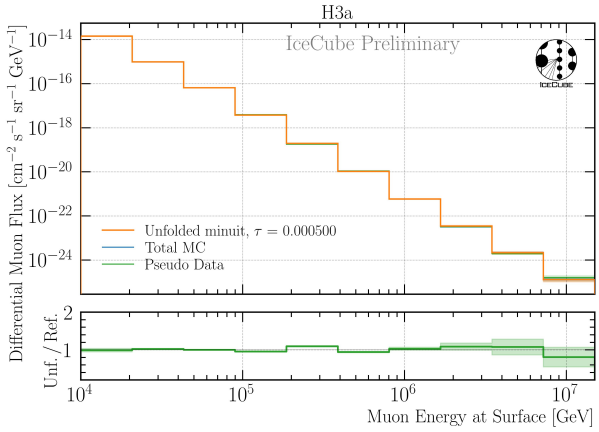
(a) Correlation between the true muon energy at the surface and the reconstructed leading muon energy at the detector entry for an H3a-weighted MC set. The correlation coefficients are $\rho_S = 0.56$ and $\rho_P = 0.58$.



(b) Example unfolding of a pseudo dataset (sampled from H3a, 12-year livetime). The bottom panel shows the ratio of the unfolded result to the truth. Without regularization, strong unphysical oscillations are observed.



(c) Conversion of the unregularized event rate from (b) into a flux using the effective area. The result remains oscillatory and unphysical.



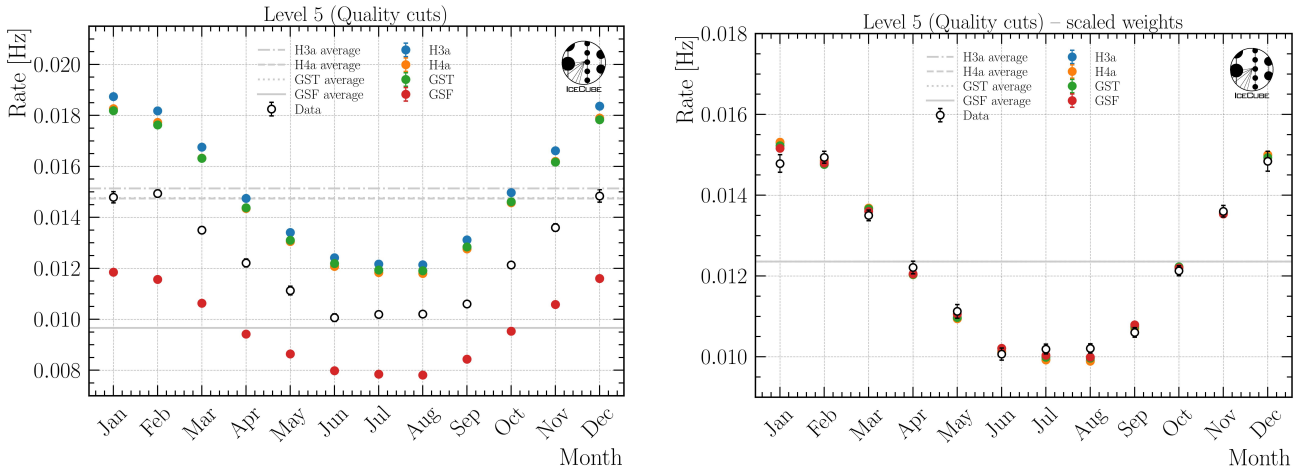
(d) The same unfolding as in (c), but with Tikhonov regularization applied ($\tau = 0.0005$). The unfolded flux is smooth and matches the true distribution.

Figure 8.2: Illustration of the unfolding principle. (a) shows the correlation between the target (surface energy) and proxy (entry energy) variables. (b) demonstrates the unstable result of unregularized unfolding on event counts. (c) shows the corresponding physical flux. (d) shows the successful recovery of the true spectrum after applying regularization.

8.2 Unfolding of Small Test Dataset (Burnsample)

Physics analyses are typically set up *blind*. This means that the entire analysis chain is based on MC simulations. When everything is validated and cross-checked, the MC data are exchanged with measured experimental data. However, to test whether the MC properly describes the data, a small amount of data is used to verify that the entire analysis chain is working as expected. This test dataset should not exceed statistics that allow significant measurements. Typical test dataset sizes are between 1 % and 10 % of the total experimental data in IceCube. In this analysis, about 2487 h of data are used for testing. It refers to about 2 % of the 12 years livetime used for the full analysis. This set is also referred to as the *burnsample*.

To cover different seasons, days and years, data from the years 2012, 2016 and 2020 are utilized for the 5th, 15th and 25th day of each month. The resulting event rate per month on *Level 5* is shown in Figure 8.3. The left plot shows the events per month weighted according to the 4 primary models. The right plot shows the same but with weights scaled to match the total number of events in the data. As expected, the event rate is different for each month due to seasonal effects in the atmosphere. It is higher in the austral summer months December to February and lower in the winter months June to August [151, 152]. After scaling the weights as described in Eq. (7.2), the MC describes the data well within statistical uncertainties for all 4 primary models.



(a) Unscaled weights.

(b) Scaled weights.

Figure 8.3: This figure shows the event rate per month for the burnsample of about 2487 h of data on *Level 5*. The left plot shows the events per month weighted with the 4 primary models. The right plot shows the same but with scaled weights to match the total number of events in data.

Whenever an unfolding of the burnsample is presented in the following sections, the binning is kept the same. For the unfolding, 6 bins in true leading muon energy at surface between 20 TeV and 2 PeV are used. For the proxy variable, the reconstructed leading muon energy at entry is binned in 25 bins between 5 TeV and 2 PeV. Both binnings are logarithmic. For the test dataset, a pseudo dataset is created by sampling from the H3a-weighted MC distribution for the burnsample livetime of 2487 h.

The burnsample results presented below have been presented at the International Cosmic Ray Conference (ICRC) 2025 in Geneva [160].

In the following, the impact of the ice systematics on the reconstructed leading muon energy at entry is presented in Section 8.2.1. Then, the effective area is calculated including systematic uncertainties in Section 8.2.2. In Section 8.2.3, the regularization strength is determined. To test the robustness of the unfolding against different primary model assumptions, unfoldings of pseudo datasets sampled

from different primary models are presented in Section 8.9a. Finally, the unfolding is applied to the burnsample data in Section 8.11. This is followed by studies of the impact of different HIMs in Section 8.2.7, and the estimation of coincident background events in Section 8.2.8.

8.2.1 Ice Systematics

As described in Section 5.3, the ice systematics are taken into account via the SnowStorm method [113]. The systematic parameters are varied in the MC simulations, and the effect on the reconstructed observable distribution is quantified via linear weight functions. These functions are then used in the unfolding to incorporate the ice systematics as nuisance parameters.

In the following, the impact of the ice systematics on the reconstructed leading muon energy is presented and also compared to the burnsample data. In Figure 8.4, the effect of absorption, DOM efficiency and scattering is shown. The left column presents the data–MC ratio of the reconstructed leading muon energy at entry for variations of each systematic parameter. The simulated systematic range is divided into 5 equidistant subsets. The dotted–dashed line indicates the nominal simulation, based on the entire systematic range. The impact of the two hole ice parameters $p0$ and $p1$ is shown in Figure 8.5. Again, the left column shows the data–MC ratio. The largest effect is observed for the absorption systematic, followed by the DOM efficiency and scattering. The hole ice parameters have a smaller effect on the reconstructed energy distribution. This is also quantified in the right column of both figures. The right column shows the corresponding χ^2 values between the lowest and highest MC subset for each variation. The p–values are presented in the title of each plot. The two hole ice parameters have a p–value of 0.012 and 0.563 for $p0$ and $p1$, respectively. The other three systematics have p-values smaller than 0.0005. Regarding the data–MC ratio, adjusting only the hole ice models in MC does not cover the true data distribution. Around 10 TeV, there are fewer data events than predicted, above 20 TeV, there are more data events than predicted. Also, variations of scattering cannot cover the data distribution. Absorption covers the data points over the entire energy range, and also DOM efficiency shows a similar behavior. In all plots, the MC is weighted to H3a.

To model the impact of detector systematics on the proxy distribution, a linear weight is derived per proxy bin and per systematic parameter. For each systematic parameter, the systematic range is divided into three equidistant bins $s \in -1, 0, 1$. The baseline is represented by the entire systematic range $s_\mathrm{b} = \sum s$. For each proxy bin i, the ratios

$$r_i(s) = \frac{N_i(s)}{N_i(s_\mathrm{b})} \tag{8.33}$$

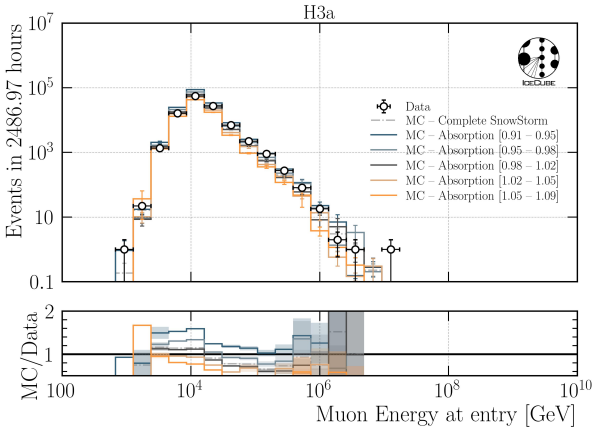
are computed, with $N_i(s)$ the bin content at setting s and $N_i(s_\mathrm{b})$ the baseline bin content.

As a first step, it is tested whether a constant model $r_i(s) = 1$ describes the three ratios within the statistical uncertainties. This is done with a χ^2 goodness-of-fit test using the uncertainties of the ratios resulting from the weights. If the p–value is ≥ 0.05, the bin is treated as insensitive to that systematic and the slope is set to zero.

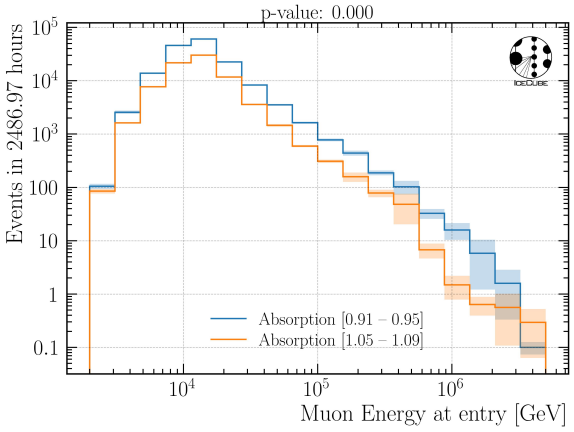
If the constant model is rejected ($p < 0.05$), a straight line constrained to pass through the baseline point is fitted,

$$r_i(s) = 1 + a_i\, s, \tag{8.34}$$

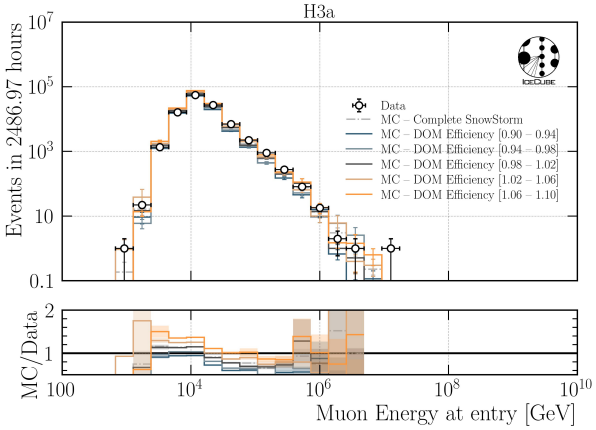
and the smallest slope $|a_i|$ is chosen that is still compatible with the three ratios at the 5 % level ($p \geq 0.05$). To find this conservative slope, a_i is scanned in steps of 0.005 and the solution with minimal absolute value that satisfies the goodness-of-fit requirement is selected. Enforcing the line to go through $(s, r) = (0, 1)$ (for example for absorption) guarantees that the baseline configuration remains unbiased. Selecting the smallest compatible slope avoids unphysical behavior that can occur when a direct fit



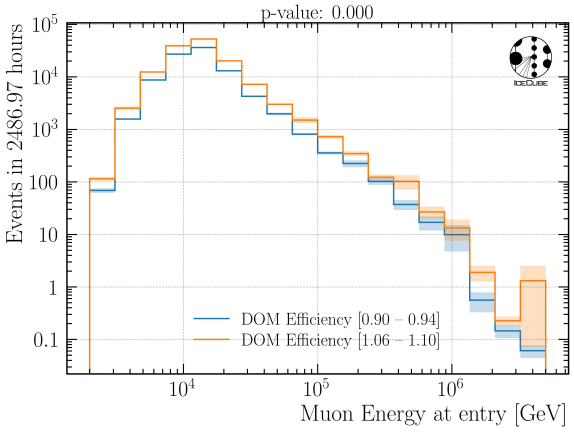
(a) Data–MC absorption.



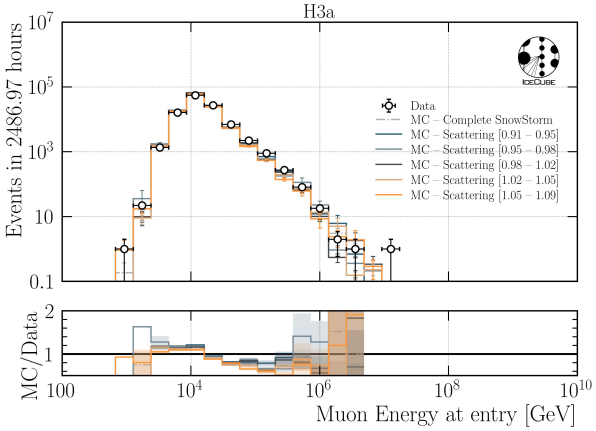
(b) χ^2 absorption.



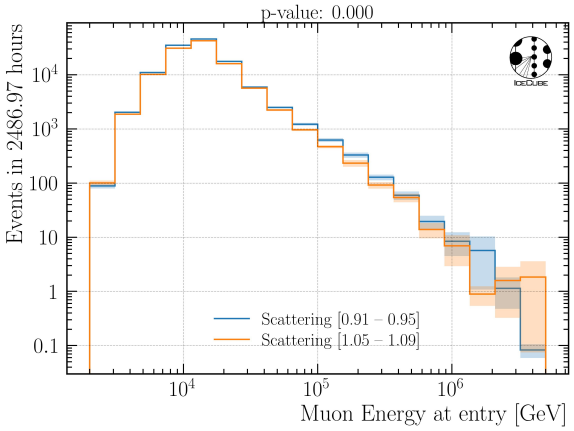
(c) Data–MC DOM efficiency.



(d) χ^2 DOM efficiency.

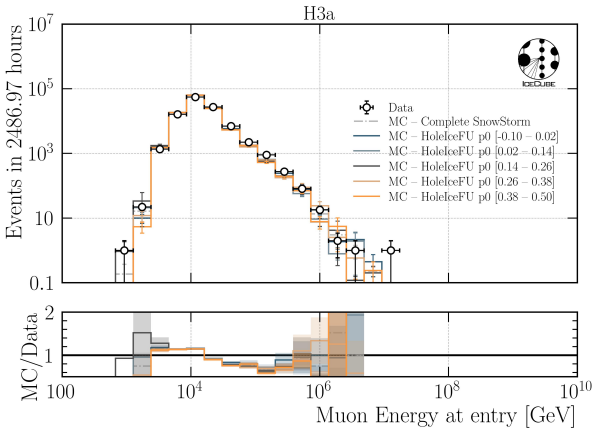


(e) Data–MC scattering

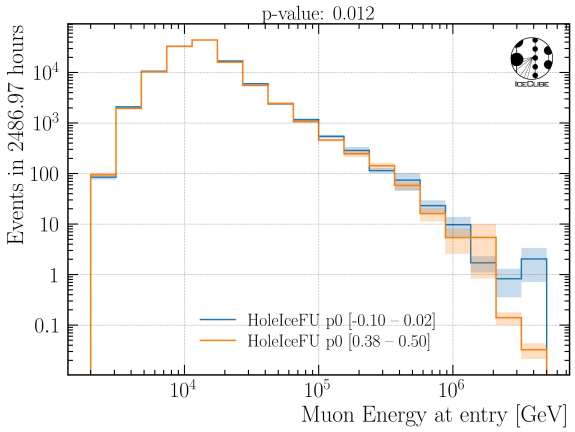


(f) χ^2 scattering.

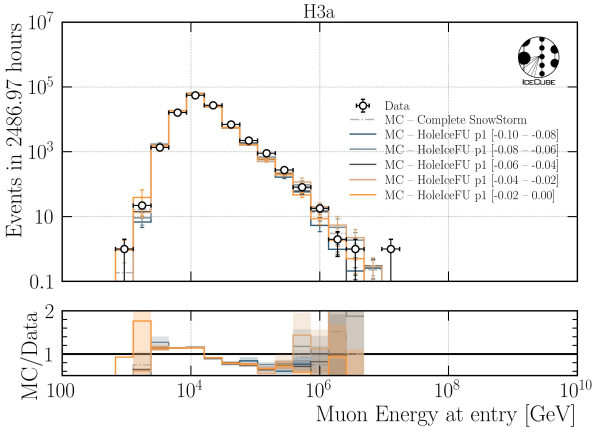
Figure 8.4: This figure shows the effect of absorption, DOM efficiency and scattering on the reconstructed leading muon energy at entry. The left column presents the data–MC ratio for variations of each systematic parameter. The simulated systematic range is divided into 5 equidistant subsets. The dotted–dashed line indicates the nominal simulation, based on the entire systematic range. The right column shows the corresponding χ^2 values between the lowest and highest MC subset for each variation. The p–values are presented in the title of each plot. The MC is weighted to H3a.



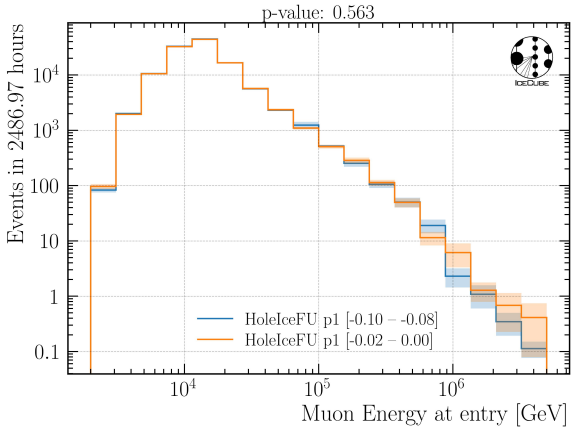
(a) Data–MC hole ice p0.



(b) χ^2 hole ice p0.



(c) Data–MC hole ice p1.



(d) χ^2 hole ice p1.

Figure 8.5: This figure shows the effect of the two hole ice parameters $p0$ and $p1$ on the reconstructed leading muon energy at entry. The left column presents the data–MC ratio for variations of each systematic parameter. The simulated systematic range is divided into 5 equidistant subsets. The dotted–dashed line indicates the nominal simulation, based on the entire systematic range. The right column shows the corresponding χ^2 values between the lowest and highest MC subset for each variation. The p–values are presented in the title of each plot. The MC is weighted to H3a.

to three points is dominated by bins with very small or very large statistical uncertainties in the sub-datasets.

This procedure yields a per-bin, per-systematic linear weight of the form $w_i(s) = 1 + a_i s$. It is simple, robust against limited MC statistics in the systematic subsets, and provides a stable input for the unfolding with nuisance parameters.

In Figure 8.6, a selection of the resulting linear fits is presented. In Figure 8.6a, the fit for absorption in proxy bin 5 is shown. Increasing the absorption leads to a decrease in the number of events in this bin. The bin has sufficient statistics and a linear fit describes the ratios well. In Figure 8.6b the effect of DOM efficiency is shown. The number of counts per bin decreases with increasing DOM efficiency. Also, a linear fit describes the ratios well. In Figure 8.6c and Figure 8.6d, the fits for the two hole ice parameters are presented for bin 0 and bin 10, respectively. The uncertainties are larger due to less statistics in these bins. Both distributions are compatible with a constant model, thus, the slope is set to zero. In Figure 8.6e, scattering is shown for bin 0. The first bin has large uncertainties, while the second bin has small uncertainties. A linear fit can describe the ratios, even though a parabolic behavior is visible. However, this is physically not expected. In Figure 8.6f, scattering is shown for bin 5. Due to the larger statistics, a linear fit can describe the ratios. It is visible that this fitting procedure results in rather conservative slopes. This could lead to an underestimation of the impact of the ice systematics in the unfolding. However, it avoids unphysical behavior in some bins. The unfolding has also been tested with direct linear fits to the three points, but this resulted in instabilities in the unfolding due to unphysical parameterizations in some bins.

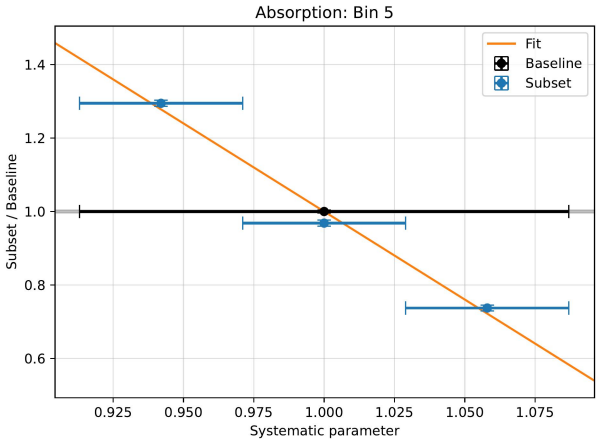
The burnsample unfolding uses 25 proxy bins, thus, 25 such fits are performed per systematic parameter. This results in 125 linear weight functions that are used in the unfolding to incorporate the ice systematics as nuisance parameters.

8.2.2 Effective Area

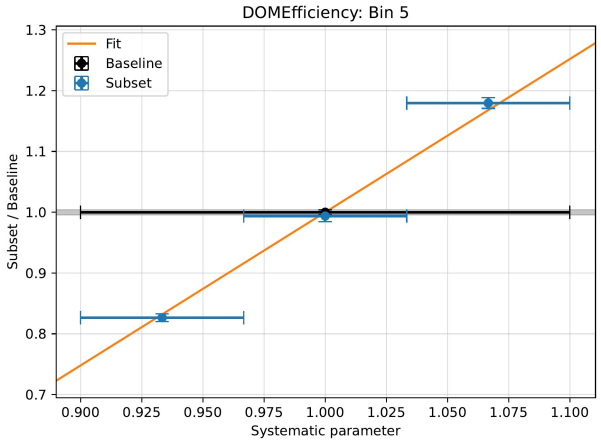
In the burnsample, the effective area is constructed exactly as outlined above: a two-dimensional efficiency in true energy and zenith is combined with the projected detector area and then averaged over zenith using the generated MC as weights, yielding an energy-dependent effective area $A_{\text{eff},i}$ with statistical uncertainties propagated from the weighted sums as shown in Eqs. (8.21)–(8.22). The energy binning of the effective area matches the unfolding binning, as required by the flux conversion. Here, 25 bin are used, logarithmically spaced between 50 TeV and 2 PeV.

Panel (a) of Figure 8.7 summarizes the resulting baseline A_{eff} together with its statistical band and the total systematic band. The relative statistical uncertainty is largest at the lowest energies, decreases towards the region of high selection efficiency, and increases again at the highest energies. This is expected because the pre-cut removes events at low energies, and the expensive CORSIKA simulations have limited statistics towards higher energies. The total systematic uncertainty, constructed by adding the individual systematic envelopes in quadrature as described in Eq. (8.32), has the largest impact on the entire distribution.

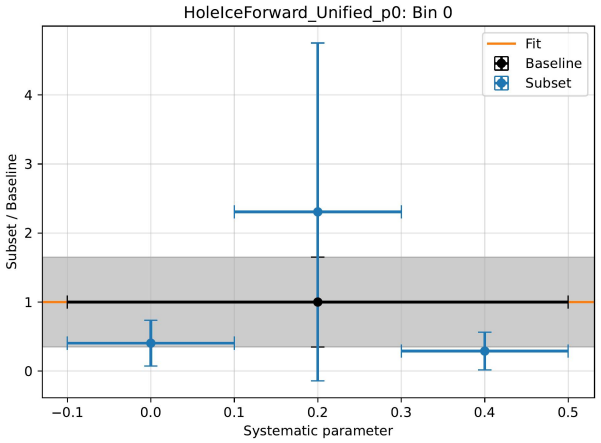
Panels (b)–(f) show the effect of each SnowStorm parameter when selecting only events with parameter values below or above the nominal center. The shifts are coherent across energy and bracket the baseline. The absorption of light in the bulk ice induces the largest excursion of A_{eff} over the full range with about 25 % impact in the lowest energy bin. As physically expected, a stronger absorption reduces the acceptance relative to the baseline. Variations of the DOM optical efficiency are the second-most relevant effect with about 20 % in the lowest energy bin, yielding an energy-dependent change of the acceptance that remains smaller than for absorption. Changes in scattering and the two hole-ice parameters $p0$ and $p1$ are subdominant in this selection and mostly within the baseline statistical band.



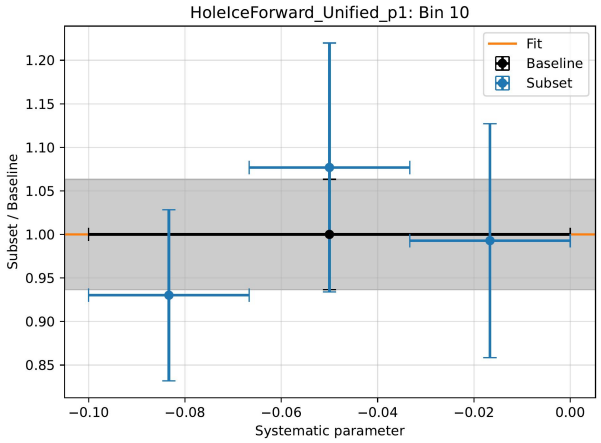
(a) Absorption.



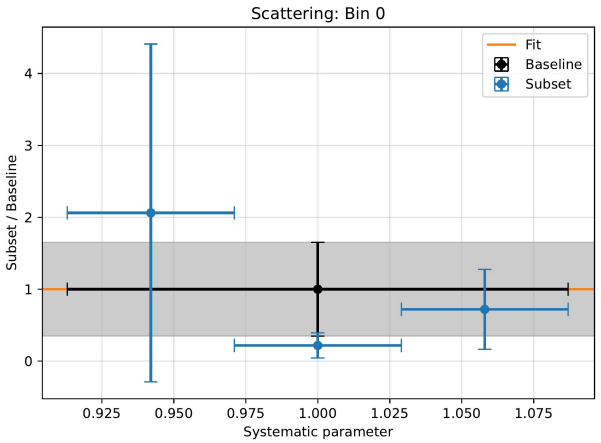
(b) DOM efficiency.



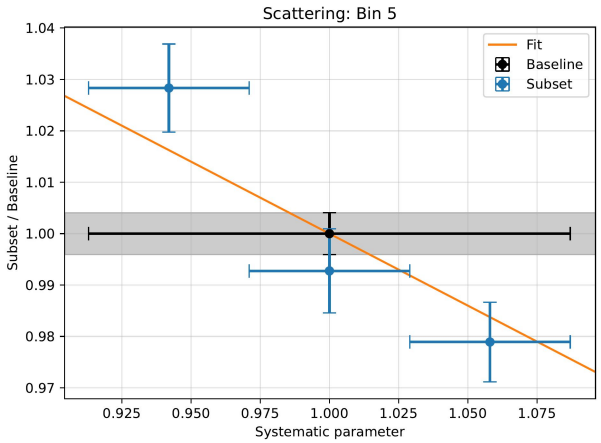
(c) Hole ice p0.



(d) Hole ice p1.



(e) Scattering.



(f) Scattering.

Figure 8.6: This figure shows a selection of the resulting linear fits for the ice systematics. These linear weight functions are used in the unfolding to incorporate the ice systematics as nuisance parameters. In total, 125 such linear weight functions are used in the burnsample unfolding.

In the following, the baseline $A_{\text{eff}}(E_\mu)$ and its combined uncertainty band are used to convert the unfolded event rates to a flux via Eq. (8.23). The systematic variations serve as input to the flux-uncertainty propagation.

8.2.3 Regularization

The strength of the Tikhonov regularization is chosen by minimizing the global correlation of the unfolded spectrum, as introduced in Eq. (8.7). For the burnsample, this optimization is first performed on MC. For each candidate value of the regularization parameter τ, an unfolding is performed and the Hessian-based covariance matrix of the solution is evaluated. The total bin-to-bin correlation $\rho(\tau)$ is then computed as the sum of the off-diagonal elements of this covariance matrix. The scan covers 280 values of τ logarithmically spaced in the range $[10^{-6}, 1]$.

To mitigate possible fluctuations in $\rho(\tau)$, a smoothed estimator, calculated as a rolling mean with a window size of five neighboring τ points, is also evaluated. Both the raw curve and the smoothed estimator yield the same optimal regularization strength of $\tau = 0.038075$. For this optimization, the unfolding is based on the H3a weighting, and the pseudo dataset is sampled from H3a scaled to the burnsample livetime.

8.2.4 Robustness Tests

Two tests are performed to evaluate the robustness of the unfolding procedure against model assumptions. In the first test, the response matrix is constructed using different primary models (H3a, H4a, GST, and GSF), while the pseudo dataset is always sampled from H3a. The resulting differential fluxes are shown in Figure 8.9a. All results are compatible with the truth within uncertainties. However, a systematic overshoot in the first bin is observed for all models. This deviation is likely driven by the limited statistics in the lowest energy bin.

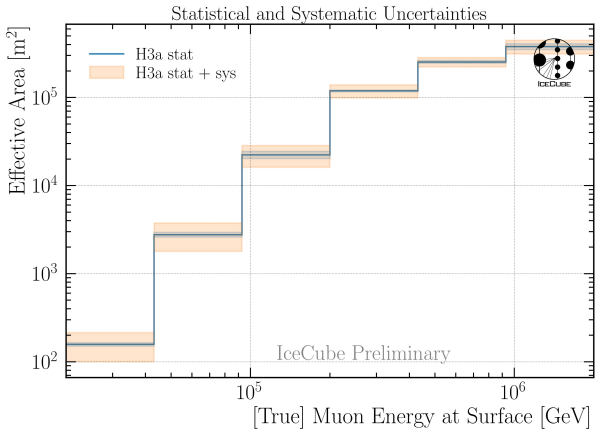
In the second test, the response matrix is constructed from H3a simulation weighted with a spectral index shifted by $\Delta\gamma = \pm 0.1$, while the pseudo data remains H3a. The spectral index shift is applied by modifying the event weights in the MC simulations according to

$$w_{k,\Delta\gamma}(E) = w_k(E) \cdot \left(\frac{E}{E_0}\right)^{\Delta\gamma}, \tag{8.35}$$

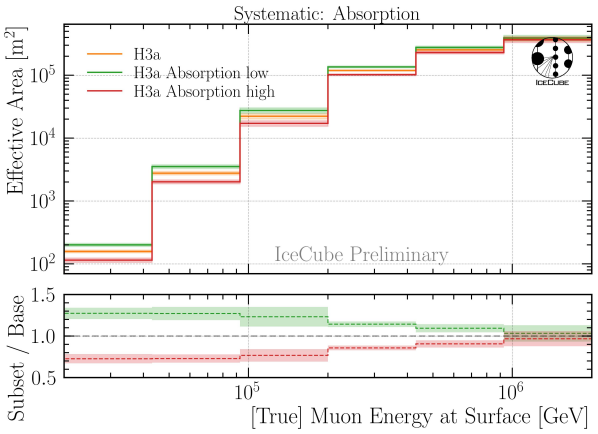
where $w_k(E)$ is the original weight of event k with primary energy E, and $E_0 = 1\,\text{GeV}$ is the reference energy. The results are presented in Figure 8.9b. Again, the unfolded spectra are compatible with the true distribution within uncertainties. Similar to the first test, the first bin consistently overshoots the true distribution. This indicates that the unfolding cannot fully correct for differences in the lowest energy bin, likely due to resolution effects or statistics. Overall, both tests demonstrate that the unfolding procedure is robust against reasonable variations in the primary model and spectral shape.

8.2.5 Burnsample Unfolding

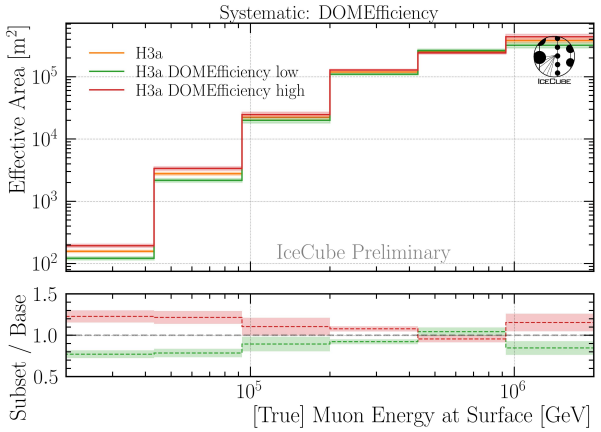
After all selection criteria on the *Final Level*, the burnsample contains 12,754 events. The unfolding is performed with the setup detailed above. A Poisson likelihood with second-derivative Tikhonov regularization and systematics included via linear SnowStorm weights in the response is minimized to obtain the unfolded events. These events are then converted to a differential flux using the energy-dependent effective area and its uncertainty band.



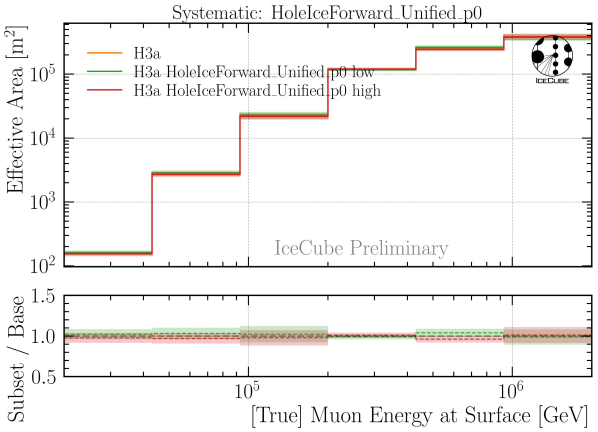
(a) Statistical and systematic uncertainties.



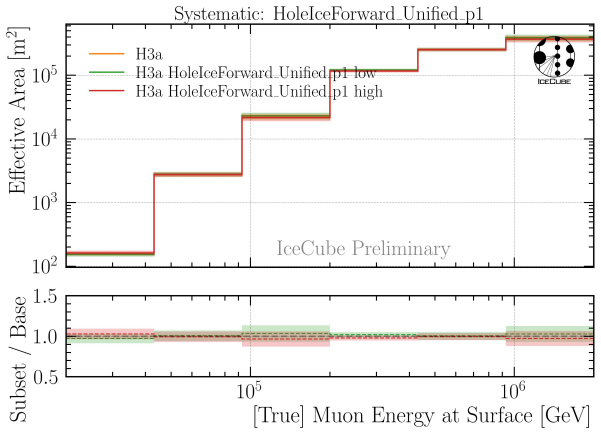
(b) Absorption.



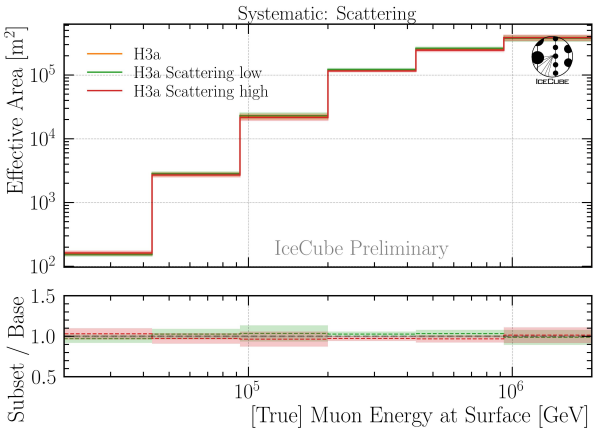
(c) DOM efficiency.



(d) Hole ice p0.



(e) Hole ice p1.



(f) Scattering.

Figure 8.7: This figure shows the effective area used in the burnsample unfolding. The different plots present the total systematic uncertainty as well as the individual contributions from the five ice systematics. The orange line presents the baseline, i.e. the entire snowstorm dataset. Green and red present the effect if only systematics lower and larger than the center value of the systematic range are used, respectively. Figure 8.7a shows the statistical uncertainty in blue and the sum of the systematic uncertainties via Eq. (8.32). These effective areas are calculated using H3a as primary model; however, since the weighting is done at the generation and detection level, the effective area is effectively independent of the primary model because it cancels out in Eq. (8.10).

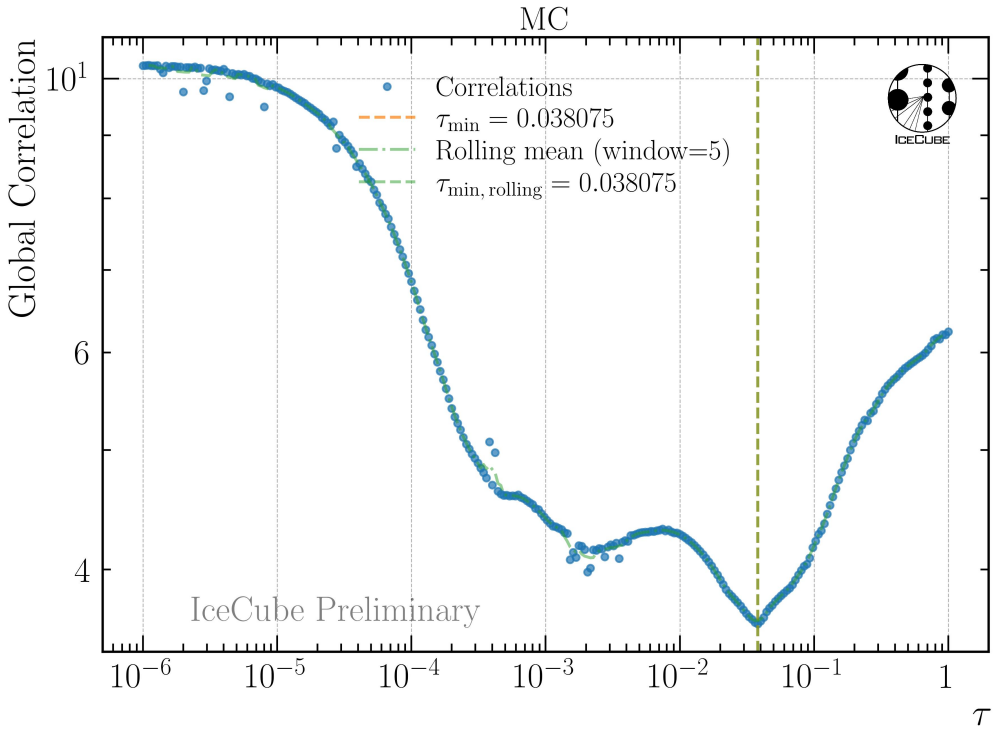
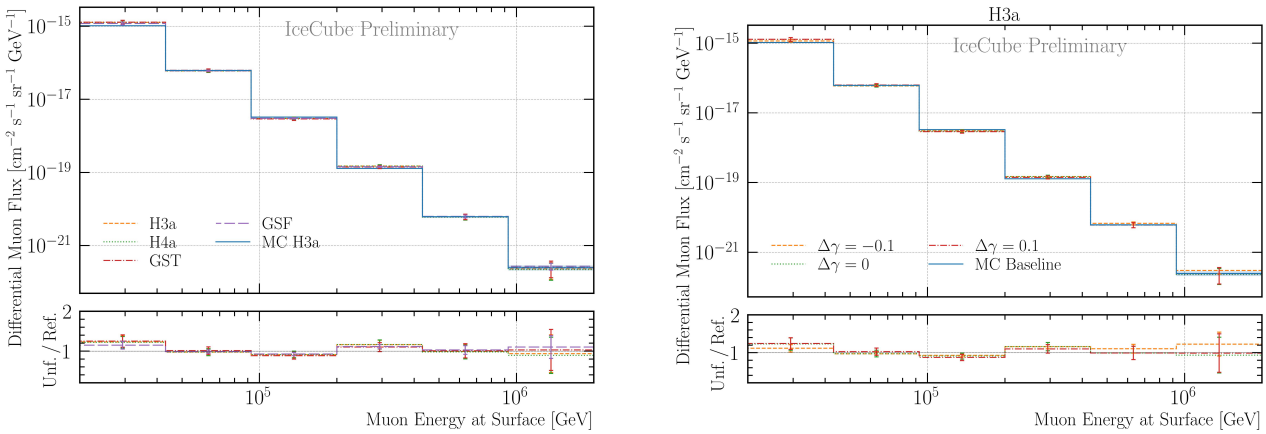


Figure 8.8: Global correlation $\rho(\tau)$ obtained from MC for 280 scanned values of τ between 10^{-6} and 1. The minimum of the curve occurs at $\tau = 0.038075$. A rolling-mean smoothing with a window of five points yields the same optimum.



(a) Unfolding robustness for different primary models.

(b) Unfolding robustness for different spectral indices.

Figure 8.9: Robustness tests of the unfolding procedure. In (a), the response matrix is constructed from different primary models. In (b), the response matrix is based on H3a with a spectral index shifted by $\Delta\gamma = \pm 0.1$. In both cases, the pseudo dataset is sampled from default H3a. The unfolding results are compatible with the true distribution within uncertainties. The persistent overshoot in the first bin indicates a limitation in correcting for differences in the lowest energy region.

Figure 8.10 shows the global correlation $\rho(\tau)$ evaluated on the experimental burnsample data. The curve exhibits a similar qualitative behavior as the MC-based scan. The regularization strength is determined to be $\tau = 0.00276$.

After the unfolding, the differential flux Φ_i in each true-energy bin i is obtained by dividing the unfolded event rate by the effective area $A_{\text{eff},i}$ according to Eq. (8.23). The unfolding procedure, including all statistical uncertainties and the detector and ice systematics treated as nuisance parameters, yields an uncertainty on the unfolded flux, $\sigma^{\text{unf}}_{\Phi_i}$, which is derived from the covariance matrix returned by the minimizer.

The systematic uncertainty on the effective area introduces an additional uncertainty on the flux. The corresponding relative uncertainty matches that of the effective area:

$$\sigma^{A_{\text{eff}}}_{\Phi_i} = \Phi_i \, \frac{\sigma^{\text{total}}_{A_{\text{eff},i}}}{A^{\text{nom}}_{\text{eff},i}}, \tag{8.36}$$

where $\sigma^{\text{total}}_{A_{\text{eff},i}}$ is the combined statistical and systematic uncertainty on the effective area defined in Eq. (8.32).

The total uncertainty on the flux is calculated by combining the unfolding uncertainty and the effective-area uncertainty in quadrature,

$$\sigma_{\Phi_i} = \sqrt{\left(\sigma^{\text{unf}}_{\Phi_i}\right)^2 + \left(\sigma^{A_{\text{eff}}}_{\Phi_i}\right)^2}. \tag{8.37}$$

The first term accounts for the measurement statistics and the effect of detector and ice systematics treated as nuisance parameters. The second term incorporates the statistical and systematic uncertainties of the effective area.

In this analysis, the flux is calculated using the baseline effective area $A^{\text{nom}}_{\text{eff},i}$, meaning $A_{\text{eff},i} \equiv A^{\text{nom}}_{\text{eff},i}$ in Eq. (8.23). The superscript "nom" emphasizes that the uncertainty $\sigma^{\text{total}}_{A_{\text{eff},i}}$ quantifies potential deviations of $A_{\text{eff},i}$ from its central value due to systematic effects.

The resulting unfolded differential flux for the burnsample is presented in Figure 8.11. The spectrum is smooth and physically plausible across the entire energy range. At low energies, the error bars are dominated by the effective-area systematics, while at high energies, they are dominated by limited statistics. For comparison, the flux is shown alongside predictions from `MCEq` using the H3a primary model and the SIBYLL 2.3c HIM. The bottom panel displays the ratio of the unfolded flux to the `MCEq` prediction. Within uncertainties, the result is compatible with the prediction for the first four bins, up to a surface muon energy of approximately 400 TeV. In the two highest energy bins, extending to 2 PeV, the result lies above the `MCEq` prediction. However, given larger uncertainties, particularly from limited statistics, the result remains compatible with the model. The top panel also shows the separate predictions for the prompt and conventional muon flux from `MCEq`. In the highest energy bin, the prompt flux is comparable in magnitude to the conventional flux. Nevertheless, due to the large uncertainties, no conclusion regarding a prompt contribution can be drawn from this burnsample unfolding. A quantitative physics interpretation is reserved for the full-sample analysis.

In Figure 8.8, the global correlation has a second minimum around $\tau = 0.04$. To test the stability of the unfolding result against the choice of the regularization parameter, the unfolding is also performed with this second minimum. Within uncertainties, the result is compatible with the unfolding of the minimum at $\tau = 0.00276$. The result is shown in Figure C.1.

The presented uncertainties are further divided into per-bin contributions from the individual sources and presented in Figure 8.12. The total uncertainty includes the statistical uncertainty from the effective area, and the so-called minuit error. The minuit error presents the uncertainty obtained from

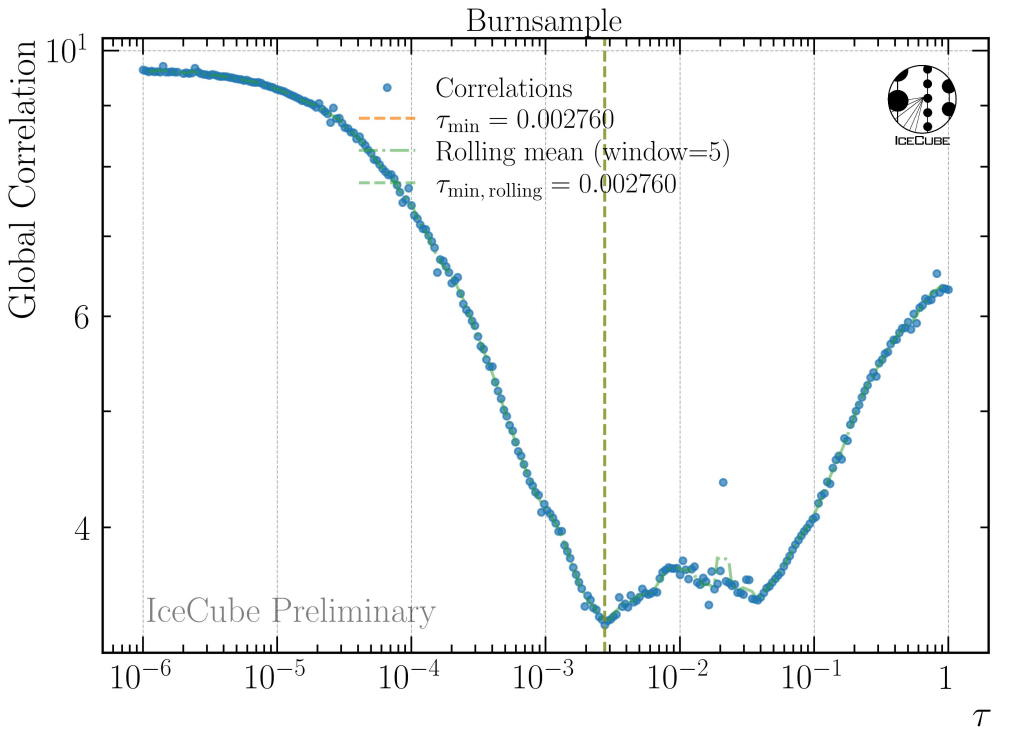


Figure 8.10: Global correlation $\rho(\tau)$ evaluated on the experimental burnsample data. The values of τ are scanned in the same range as for MC, i.e., 280 values between 10^{-6} and 1. The minimum of the curve occurs at $\tau = 0.00276$.

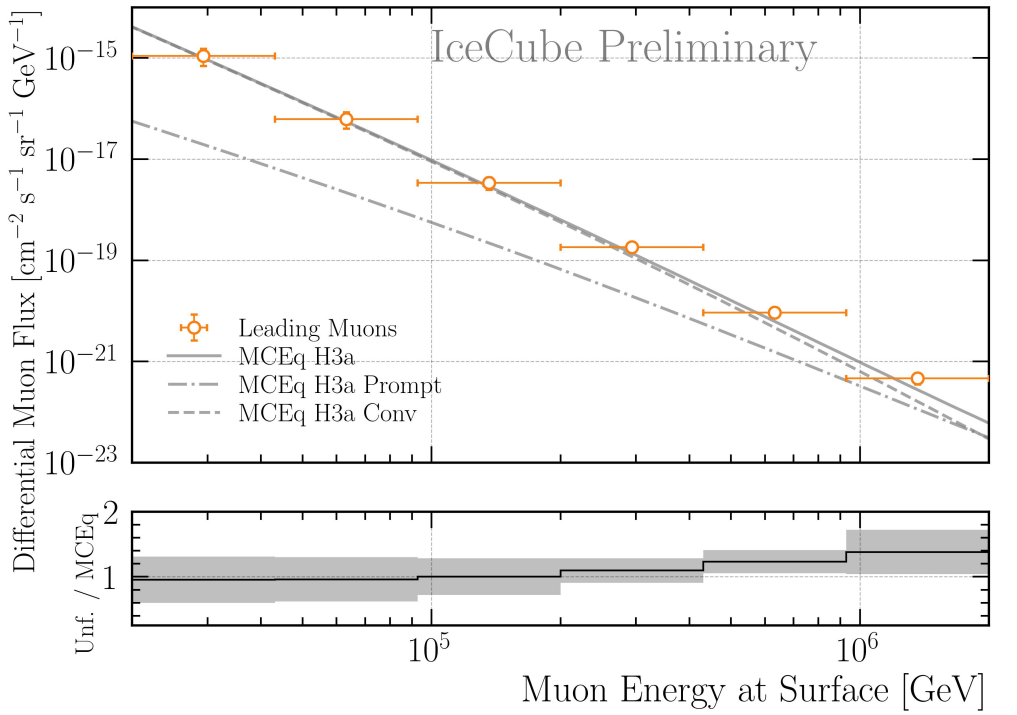


Figure 8.11: Unfolded differential atmospheric muon flux for the burnsample using $\tau = 0.00276$. Error bars represent statistical uncertainties from the inverse Hessian and the effective-area systematics and statistics propagated as in Section 8.1.2.

the inverse Hessian matrix after the unfolding fit. This uncertainty includes the statistical uncertainty from the data as well as the effect of the ice systematics treated as nuisance parameters in the unfolding fit. To quantify how the total uncertainty of the unfolded muon flux decomposes into statistical and systematic components, the full covariance matrix C_{full} returned by the minimizer is utilized. It contains correlations between the unfolded flux bins x_i (here 6 bins, later 10 bins for the full sample unfolding), the five systematic nuisance parameters s_j, and two additional parameters associated with underflow (UF) and overflow (OF) bins. After the UF/OF entries are removed, a reduced covariance C_{red} is obtained, and its precision matrix is used for all subsequent manipulations,

$$H_{\text{red}} = C_{\text{red}}^{-1} = \begin{pmatrix} H_{xx} & H_{xs} \\ H_{sx} & H_{ss} \end{pmatrix}, \tag{8.38}$$

where H_{xx} couples the 6 burnsample unfolded bins, H_{ss} couples the five systematics, and H_{xs} encodes their correlations. Fixing all systematic parameters corresponds to removing their degrees of freedom, so the purely statistical covariance is obtained simply as

$$C_{\text{stat}} = H_{xx}^{-1}. \tag{8.39}$$

To include the effect of all five systematics while keeping UF/OF fixed, marginalization over the nuisance block is performed using the Schur complement,

$$S = H_{xx} - H_{xs}\, H_{ss}^{-1}\, H_{sx}, \tag{8.40}$$

by which the covariance with all systematics free is obtained as

$$C_{x,\text{sysfree}} = S^{-1}. \tag{8.41}$$

The additional uncertainty introduced by the five nuisance parameters is therefore given by

$$C_{\text{sys}} = C_{x,\text{sysfree}} - C_{\text{stat}}. \tag{8.42}$$

The fully marginalized covariance of the unfolded bins, including UF/OF variations, is expressed as

$$C_{x,\text{full}} = C_{\text{full}}\big|_{\{x_1,\ldots,x_6\}}, \tag{8.43}$$

and the contribution of the UF/OF parameters is identified through

$$C_{\text{UF/OF}} = C_{x,\text{full}} - C_{x,\text{sysfree}}. \tag{8.44}$$

To extract the uncertainty contribution of an individual systematic parameter s_j, the Schur complement procedure is repeated while only the j-th nuisance parameter is freed. Denoting the j-th column of H_{xs} by $H_{xs}^{(j)}$ and the corresponding diagonal element of H_{ss} by $(H_{ss})_{jj}$, the precision matrix with only s_j free is given by

$$S_j = H_{xx} - H_{xs}^{(j)}\, (H_{ss})_{jj}^{-1}\, (H_{sx})^{(j)}, \tag{8.45}$$

and the corresponding covariance becomes

$$C_x^{(j)} = S_j^{-1}. \tag{8.46}$$

The incremental uncertainty generated by systematic s_j is then obtained from

$$\Delta C^{(j)} = C_x^{(j)} - C_{\text{stat}}, \tag{8.47}$$

which represents an exact but non-additive contribution. For comparison, a linearized and additive approximation is derived by first-order error propagation around the statistical covariance. With

$$K = -C_{\text{stat}}\, H_{xs}, \qquad K_j = K[:, j], \tag{8.48}$$

the linearized covariance contribution of s_j is expressed as

$$\Delta C_{\text{lin}}^{(j)} = (K_j \otimes K_j)\, C_{ss,jj}, \tag{8.49}$$

which becomes exact in the limit of a diagonal systematic sub-block C_{ss}. Finally, each covariance contribution C is converted into a relative uncertainty for bin i according to

$$\sigma_i^{\text{rel}} = \frac{\sqrt{C_{ii}}}{x_i} \times 100\ \%. \tag{8.50}$$

Through this procedure a complete decomposition of the uncertainty is obtained, including statistical contributions, the combined effect of all systematics, the impact of UF/OF parameters, and the individual contribution of each systematic nuisance parameter.

The systematic uncertainties resulting from the effective area are not presented here, and thus they are not included in the total uncertainty shown in Figure 8.12. Although they dominate the total uncertainty (as visualized in Figure 8.7a), they are excluded to avoid double counting, since the ice systematics are already incorporated as nuisance parameters in the unfolding fit. Note that in the final burnsample unfolding result (Figure 8.11), these effective area systematics were included. This resulted in a conservative overestimation of the uncertainty, which was deemed appropriate for the preliminary results presented at the International Cosmic Ray Conference (ICRC) 2025 [160]. For the final full-sample analysis, this double counting is avoided.

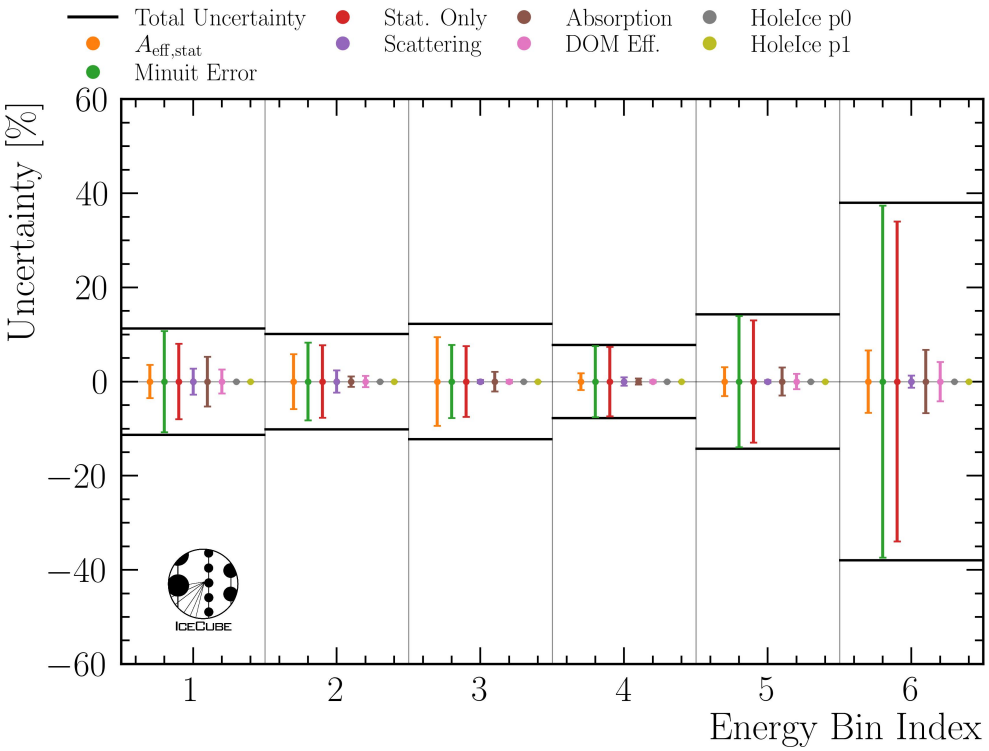


Figure 8.12: Per-bin uncertainties of the burnsample unfolding result. The total uncertainty (black) comprises the statistical uncertainty from the effective area (orange) and the minuit error from the unfolding fit (green). The latter includes the data statistical uncertainty (red) and the effect of the five ice systematics treated as nuisance parameters. The systematic uncertainties from the effective area are excluded here to prevent double counting of ice systematics, as they are already incorporated via nuisance parameters in the unfolding fit.

8.2.6 Data–MC Comparison

To incorporate the best-fit detector and ice systematics obtained from the unfolding fit, the SnowStorm MC sample is re-weighted to reflect the posterior distribution of the nuisance parameters. In the SnowStorm approach, each systematic parameter s_i is sampled uniformly within predefined bounds $[s_i^{\min}, s_i^{\max}]$. For an N-dimensional vector of systematics $\mathbf{s} = (s_1, \dots, s_N)$, the simulation prior therefore factorizes as

$$p_{\text{sim}}(\mathbf{s}) = \prod_{i=1}^{N} p_{\text{sim},i}(s_i) = \prod_{i=1}^{N} \frac{1}{s_i^{\max} - s_i^{\min}}. \tag{8.51}$$

During the unfolding procedure, the nuisance parameters corresponding to these systematics are fitted simultaneously with the flux parameters. This yields best-fit values $\hat{s}_i$ and a covariance matrix C_{ij}, from which the one-sigma widths

$$\sigma_i = \sqrt{C_{ii}} \tag{8.52}$$

are extracted.

To propagate the fitted systematics to the simulation, the posterior distribution of the nuisance parameters is approximated by a product of independent Gaussian functions centered at the best-fit point,

$$p_{\text{target}}(\mathbf{s}) \propto \prod_{i=1}^{N} \exp\left[-\frac{1}{2}\left(\frac{s_i - \hat{s}_i}{\sigma_i}\right)^2\right]. \tag{8.53}$$

Because the SnowStorm simulation contains events only within the finite interval $[s_i^{\min}, s_i^{\max}]$, the Gaussian priors are truncated to this range. The unnormalized Gaussian density in dimension i is given by

$$g_i(s_i) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{s_i - \hat{s}_i}{\sigma_i}\right)^2\right]. \tag{8.54}$$

Its normalization within the simulation bounds is

$$\mathcal{N}_i = \int_{s_i^{\min}}^{s_i^{\max}} g_i(s)\,\mathrm{d}s \tag{8.55}$$

$$= \Phi\left(\frac{s_i^{\max} - \hat{s}_i}{\sigma_i}\right) - \Phi\left(\frac{s_i^{\min} - \hat{s}_i}{\sigma_i}\right), \tag{8.56}$$

where Φ denotes the standard normal cumulative distribution function. The corresponding truncated Gaussian prior becomes

$$p_{\text{target},i}(s_i) = \frac{g_i(s_i)}{\mathcal{N}_i}. \tag{8.57}$$

Since the SnowStorm sampling prior in each dimension is uniform,

$$p_{\text{sim},i}(s_i) = \frac{1}{s_i^{\max} - s_i^{\min}}, \tag{8.58}$$

the importance re-weighting factor in dimension i is obtained as

$$w_i(s_i) = \frac{p_{\text{target},i}(s_i)}{p_{\text{sim},i}(s_i)} = g_i(s_i)\,\frac{s_i^{\max} - s_i^{\min}}{\mathcal{N}_i}. \tag{8.59}$$

Assuming independence between the systematic dimensions, the full N-dimensional re-weighting factor is given by

$$w(\mathbf{s}) = \prod_{i=1}^{N} w_i(s_i) = \frac{p_{\text{target}}(\mathbf{s})}{p_{\text{sim}}(\mathbf{s})}. \tag{8.60}$$

Table 8.1: Best-fit values and 1σ uncertainties of the five detector and ice systematics obtained from the burnsample unfolding fit.

Systematic Parameter	Best-fit Value with Uncertainty (1σ)
Scattering	0.95 ± 0.06
Absorption	1.00 ± 0.01
DOM efficiency	0.99 ± 0.02
HoleIce p_0	0.20 ± 0.30
HoleIce p_1	-0.05 ± 0.05

For each MC event, the values of the corresponding SnowStorm parameters s_i stored in the metadata are inserted into the expression above, resulting in one multiplicative weight per event. This procedure is implemented in the `MultiGaussReweight` class, which evaluates the Gaussian densities, the truncated normalizations, and the uniform simulation densities for all systematic parameters.

As the re-weighting modifies only the relative contributions of the events but not the overall simulated statistics, the sum of event weights after re-weighting is renormalized to match the original total. Denoting the baseline weights by w_k^{base} and the systematic re-weighting factors by w_k^{sys}, the normalization is chosen such that

$$\sum_k w_k^{\text{base}} = \sum_k w_k^{\text{base}}\, w_k^{\text{sys}}. \tag{8.61}$$

This procedure prevents an artificial shift in the absolute event rate and ensures that only the distribution in the systematic-parameter space is modified.

For the purpose of data–MC comparison, the final, systematic-adjusted simulation weights are therefore calculated as

$$w_k^{\text{final}} = w_k^{\text{base}} \cdot w_k^{\text{sys}}, \tag{8.62}$$

where w_k^{sys} encodes the transformation from the uniform SnowStorm prior to the Gaussian posterior implied by the unfolding fit. This allows simulation-based predictions to be evaluated at the best-fit values of the detector and ice systematics.

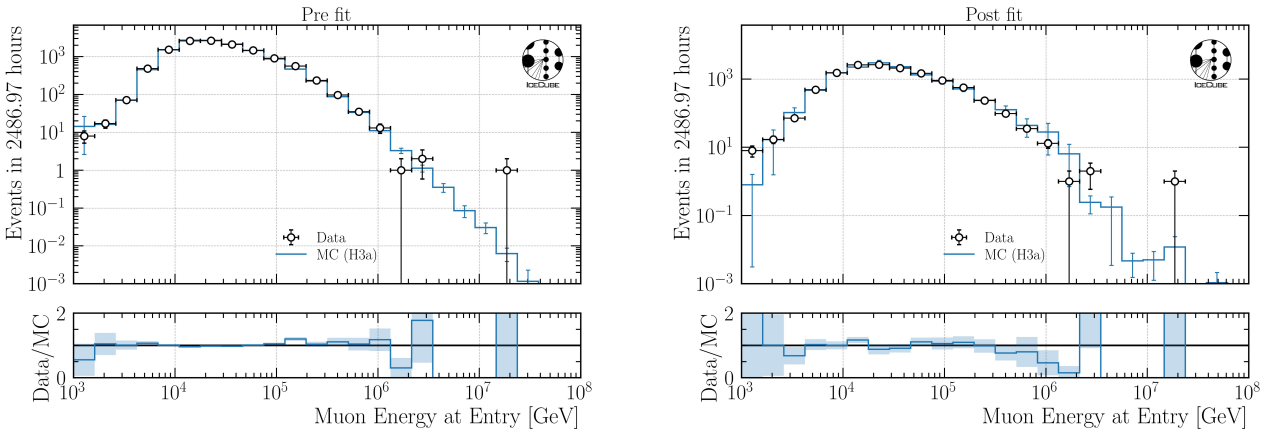
The comparison of the data–MC agreement before and after the unfolding fit is shown in Figure 8.13. The pre-fit MC uses the entire SnowStorm dataset with all simulated systematics, while the post-fit MC is re-weighted to the best-fit parameters resulting from the unfolding. Both the pre-fit and post-fit MC describe the data well within uncertainties. A disagreement appears around 1 PeV, but this is caused by the little statistics of one and two events in the experimental data.

The best-fit values and 1σ uncertainties of the five detector and ice systematics obtained from the burnsample unfolding fit are summarized in Table 8.1. The re-weighted distributions of the five detector and ice systematics are shown in Figure C.2.

8.2.7 Impact of Primary and Hadronic Interaction Models

The atmospheric muon flux at the surface depends on the assumed primary cosmic ray flux model and the HIM determining the air showers. To estimate the magnitude of these systematic uncertainties, the muon flux is predicted using `MCEq` for different model combinations.

Figure 8.14a illustrates the impact of the primary cosmic ray model on the predicted muon flux. The flux is calculated using `MCEq` with SIBYLL 2.3c as the fixed HIM, while varying the primary model between H3a, H4a, GST, and GSF. Below an energy of approximately 300 TeV, the differences between the models appear primarily as a constant offset in the normalization. Since the unfolding procedure relies on the relative relation between the proxy and the target variable rather than the absolute



(a) Pre fit. **(b)** Post fit.

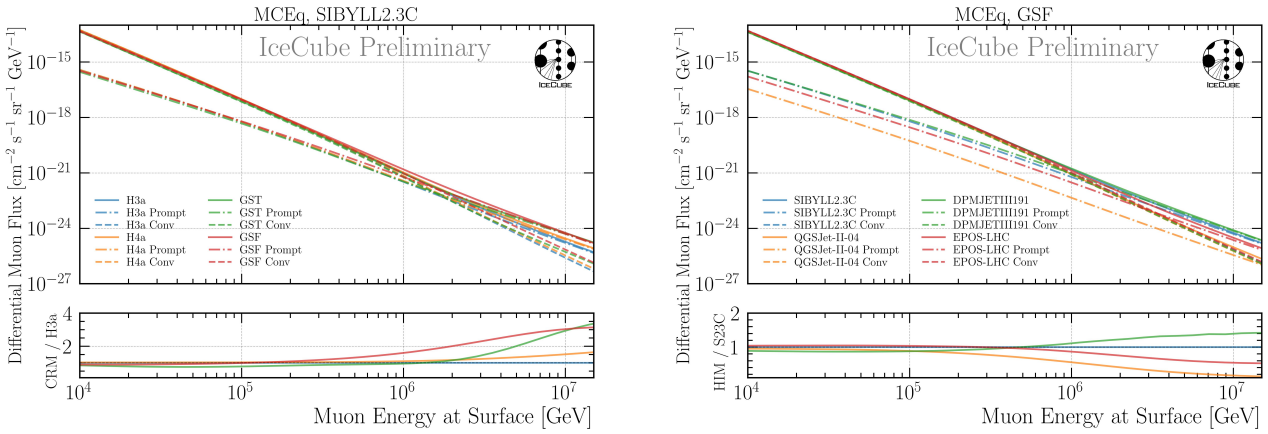
Figure 8.13: This figure presents the data–MC comparison of the reconstructed leading muon energy at entry before and after the unfolding fit for the burnsample. For the pre-fit, the entire SnowStorm dataset with all simulated systematics is used. For the post-fit, the SnowStorm dataset is re-weighted to the best-fit parameters resulting from the unfolding. The MC is weighted to H3a. The pre-fit as well as the post-fit MC both describe the data well within uncertainties.

normalization, these constant offsets do not affect the result. However, towards higher energies, the shape of the spectrum changes noticeably. For instance, at 10 PeV, the GSF and GST models predict a flux that is approximately 250 % higher than that of H3a.

Figure 8.14b shows the influence of the HIM. Here, the primary model is fixed to GSF, and the HIM is varied between SIBYLL 2.3c [53], QGSJET-II-04 [54], DPMJET-III 19.1 [161], and EPOS-LHC [55]. Among these models, only SIBYLL and DPMJET include a charm component. Consequently, the predictions from QGSJET and EPOS-LHC are expected to be lower than the SIBYLL prediction, which is observed in the figure. DPMJET predicts an even higher flux, exceeding SIBYLL by about 50 % at the highest energies. Similar to the primary models, the differences below approximately 200 TeV appear as a constant offset, which is negligible for the unfolding analysis.

For the burnsample analysis, which is limited to energies up to 2 PeV, the impact of the primary model and HIM is subdominant. As shown in the robustness tests in Section 8.2.4 (Figure 8.9a), unfolding with response matrices based on different primary models yields results consistent within uncertainties. Since the variations induced by different HIMs are generally smaller than those from primary models (as seen in Figure 8.14), their impact is also considered negligible for this energy range.

However, extending the analysis to 15 PeV for the full sample increases the model dependence. As illustrated in Figure 8.14a, the divergence between primary models grows significantly at higher energies. This leads to a larger potential bias in the unfolding, which is quantified in Section 8.3.6. Figure 8.22 demonstrates a residual mismatch between the injected and unfolded spectra when using different primary models. Although this bias remains below 40 % at the highest energies, much smaller than the 250 % variation between model predictions, it represents a relevant systematic uncertainty. Consequently, for the full sample analysis, the unfolding is performed independently using all four primary models to explicitly estimate this systematic uncertainty. The smaller impact of the HIM is neglected in comparison. Ideally, additional CORSIKA datasets with different HIMs would be simulated to study this effect directly. However, given the large amount of resources needed, this is not feasible within the order of years, even with access to large computer clusters.



(a) Impact of primary models.

(b) Impact of HIMs.

Figure 8.14: This figure shows the impact of different (a) primary models and (b) HIMs on the predicted atmospheric muon flux from `MCEq`. The baseline model used in this analysis is SIBYLL 2.3d for the HIM. For the primary model, H3a is used in the burnsample unfolding, while the full sample unfolding is performed four times using all four primary models.

8.2.8 Impact of Coincident Primaries

The MC dataset utilized in this analysis does not include simulated coincident primaries. This means that every MC event contains only one primary cosmic ray that generates one EAS and then the resulting muon bundle. However, in real experimental data, coincident primaries can occur. This means, that in addition to the signal primary cosmic ray, other cosmic rays can enter the atmosphere within the same time window, generating additional EASs and muon bundles. These coincident primaries can contribute additional muons to the detected muon bundle, potentially affecting the reconstructed properties of the leading muon, such as its energy. To study this effect, the MC dataset 20904 is used, which contains coincident background primaries in addition to the signal primary. In Figure 8.15a, the target distribution of the leading muon energy at surface is shown for this dataset, weighted to H3a for a livetime of 12 years. The blue distribution presents the leading muon energy at surface considering only the signal primary, while the orange distribution shows the bundle energy of the background tree. This potentially contributes additional muons to the detected leading muon. The background distribution has a similar shape as the target distribution, but with a lower rate. To further investigate this impact, the asymmetry between the signal leading muon energy at surface and the background bundle energy at surface is studied. This asymmetry is defined as

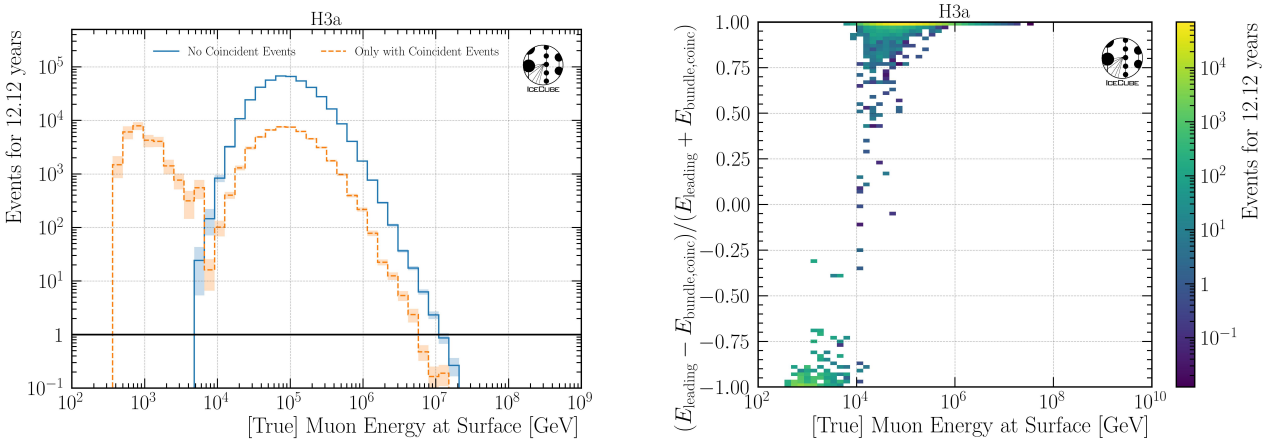
$$X_{\text{asym}} = \frac{E_{\text{leading}} - E_{\text{bundle, coinc}}}{E_{\text{leading}} + E_{\text{bundle, coinc}}}\,, \tag{8.63}$$

where E_{leading} is the leading muon energy at surface from the signal primary, and $E_{\text{bundle, coinc}}$ is the bundle energy at surface from the coincident background primary. The resulting 2D distribution of the leading muon energy at surface versus this asymmetry is shown in Figure 8.15b. When this asymmetry is close to -1, it indicates that the background event has a much higher bundle energy compared to the leading muon energy at surface from the signal primary. When the asymmetry is close to 1, it indicates that the leading muon energy at surface from the signal primary is much higher than the background bundle energy, i.e. the impact of the coincident primary is negligible. The presented 2D distribution shows that below a leading muon energy of about 10 TeV, the background dominates the event sample. For all energies above 10 TeV, the signal rate dominates the distribution. A transition phase is observed between 10 TeV and roughly 1 PeV, where the signal becomes increasingly dominant. In this region, the background rate mostly contributes less than 30 %, with only a few outliers showing

a higher contribution. However, for most events, the signal events dominates the leading muon energy at surface. A 1D projection of this asymmetry is shown in Figure D.1a in Appendix D. This plot in logarithmic scale shows that most events have an asymmetry $X_{\text{asym}} \approx 1$. This can also be expressed in terms of the event rate. In total, the signal rate for this dataset is 1.12 mHz, while the background rate is 0.20 mHz, weighted to H3a. However, for events with a leading muon energy at surface above 10 TeV, the signal rate is 1.11 mHz, while the background rate is only 0.13 mHz. With an additional requirement that the background muon bundle energy contributes at least 10% to the leading muon energy at surface, the background rate reduces to < 0.01 mHz. The light of the background events overlaps with the signal light, and thus a 10 % contribution could shift the reconstructed energy up by a similar amount. Comparing this to the uncertainties of the reconstruction, the effect is within the given resolution.

Detailed plots of the reconstruction performance for this dataset are shown in Appendix D. Specifically, Figure D.1b shows the performance without coincident events, which can be compared to the performance with coincident events included (Figures D.1c and D.1d). The comparison shows that the reconstruction performance is not significantly affected by coincident primaries within the given resolution. Overall, the impact of coincident primaries on the leading muon energy at surface is small compared to the reconstruction uncertainties of this analysis for the target energy range of events with at least 10 TeV. Therefore, the effect of coincident primaries is neglected in this analysis.

In addition, the impact of coincident primaries was also tested for another dataset, 22615. The event rates for this dataset are listed in Table 8.2. The results are consistent with the findings for dataset 20904. Detailed plots and a discussion for dataset 22615 can be found in Appendix D.



(a) 1D target distribution.

(b) 2D target distribution.

Figure 8.15: This figure presents the impact of coincident primaries on the target distribution, the leading muon energy at surface. For this, the dataset 20904 is used. In addition to the signal primary, this dataset contains coincident background primaries. The left panel shows the 1D target distribution, while the right panel shows the 2D distribution as introduced in Eq. (8.63). For leading muon energies above 10 TeV, the signal primary dominates the event sample, with only a few events where the background contributes significantly to the leading muon energy at surface. Since the unfolding energy range starts at 10 TeV, the effect of coincident primaries is neglected in this analysis.

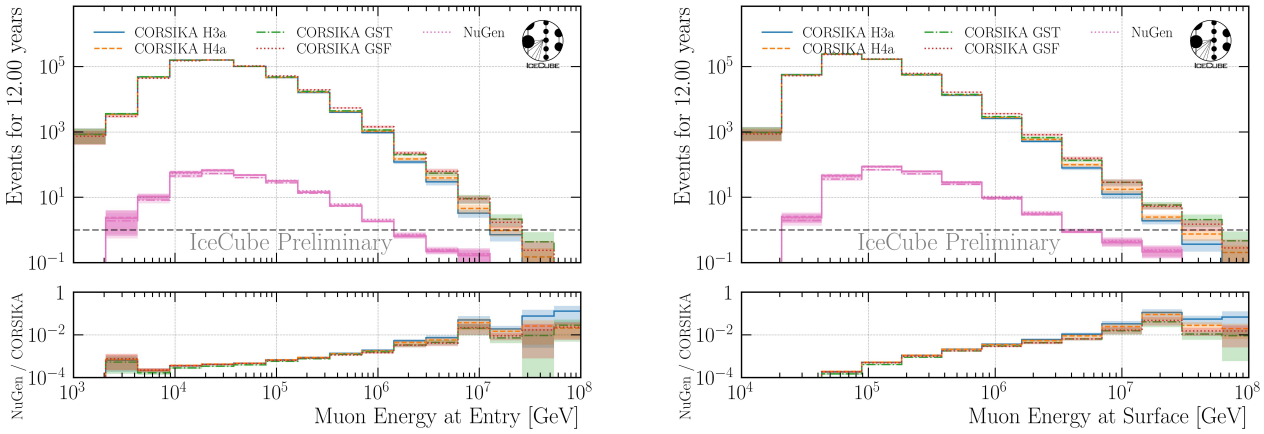
8.2.9 Neutrino Background

Neutrinos produced in EAS, as well as astrophysical neutrinos, constitute a potential background for this measurement. The atmospheric neutrinos are weighted to `MCEq` fluxes using the four primary

Table 8.2: This table presents the event rates for the datasets 20904 and 22615 on *Final Level* for different cuts. The signal rate is defined as the rate of events without any coincident primaries, $N_{\text{conic}} = 0$, while the background rate is defined as the rate of events with at least one coincident primary, $N_{\text{conic}} \geq 1$. The unfolding starts at a minimum muon energy at surface larger than 10 TeV. Given the resolution of the energy estimation, the rate is also calculated for events where the background muon energy contributes to more than 10 % to the signal energy. The rates are weighted to H3a.

Cut	Rate [mHz] 20904	22615
$N_{\text{conic}} = 0$	1.12	0.84
$N_{\text{conic}} \geq 1$	0.20	0.68
$N_{\text{conic}} = 0, E_{\text{lead}} > 10\,\text{TeV}$	1.11	0.74
$N_{\text{conic}} \geq 1, E_{\text{lead}} > 10\,\text{TeV}$	0.13	0.58
$N_{\text{conic}} = 0, E_{\text{lead}} > 10\,\text{TeV}, \frac{E_{\text{Bkg}}}{E_{\text{lead}}} > 0.1$	0.00	0.02

models H3a, H4a, GST, and GSF. The astrophysical neutrinos are weighted to the latest IceCube result [162], which follows a broken power-law distribution. The investigations outruled a single power-law spectrum by more than $4\,\sigma$. Figure 8.16 shows the distributions of the proxy and target variables, visualizing the weighted event rates for 12 years of livetime. A dashed grey line marks the threshold of one expected event. For both variables, the expected neutrino contribution remains below 1 %. Given the larger statistical and systematic uncertainties of the measurement, the neutrino background is negligible and is therefore neglected in the unfolding.



(a) Proxy distribution.

(b) Target distribution.

Figure 8.16: Comparison of the atmospheric muon signal (CORSIKA) and the neutrino background (NuGen) for (a) the proxy distribution and (b) the target distribution. The atmospheric muons are shown for all four primary models (H3a, H4a, GST, GSF). The neutrino background includes atmospheric neutrinos weighted to `MCEq` for the four primary models, as well as astrophysical neutrinos weighted to a broken power-law spectrum [162]. The dashed grey line marks the expectation of one event for a livetime of 12 years. The neutrino background is negligible across the entire energy range.

8.3 Full Sample 12 Years of Data Unfolding

The full sample unfolding utilizing 12.12 years of IceCube data is performed similarly to the burnsample unfolding presented above. However, some modifications are done. To increase the MC statistics,

multiple MC datasets are combined, as explained in Section 8.3.1. An optimized binning for the unfolding proxy variable is used, as explained in Section 8.3.2. In addition, the larger MC statistics enables are better description of the ice systematics, see Section 8.3.3. The effective area for the full sample unfolding is presented in Section 8.3.4. The regularization strength is re-optimized for the full sample unfolding, as explained in Section 8.3.5. Section 8.3.6 presents robustness tests performed for the full sample unfolding, leading to an estimation of the systematic uncertainty caused by the choice of the primary model. Then, a sensitivity study is presented in Section 8.3.7 to estimate whether the unfolding procedure is sensitive to a prompt muon component in the MC dataset, followed by an unfolding of the MC dataset for the full sample in Section 8.3.8. Finally, the results of the full sample unfolding on experimental data are presented in Section 8.3.9. Given the unfolded muon spectrum, the normalization of the prompt muon component is estimated in Section 8.3.10 using template fits. Furthermore, significances of the prompt muon component are calculated, where also preliminary results of another stopping muon IceCube analysis are included, to stabilize the fit in the lower energy region. The chapter concludes with a summary of the unfolding in Section 8.4.

8.3.1 Combination of MC Datasets

Within the scope of this thesis, new MC datasets have been produced with the "Extended History" option enabled in CORSIKA to tag individual muons as either prompt or conventional origin. Since these simulations are computationally expensive, it is not feasible to produce a single MC dataset with sufficient statistics for the entire livetime of 12.12 years. Therefore, multiple MC datasets are combined to increase the overall statistics. The combined MC dataset used for the full sample unfolding consists of the datasets listed in Table 8.3. Dataset label 2353X refers to the two sets 23530 and 23531. Datasets 2277X and 2353X were simulated for this thesis with a cylinder radius $r = 700\,\mathrm{m}$ and length $h = 1400\,\mathrm{m}$, while the other two datasets (23111 and 20904) were simulated with $r = 600\,\mathrm{m}$ and $h = 1200\,\mathrm{m}$. The cylinder corresponds to the simulated detector volume of the IceCube in-ice array. Sets 2277X, 2353X, and 23111 use the Spice FTP-v3m ice model, while dataset 20904 uses the older Spice 3.2 ice model. Since the ML-based reconstructions are trained on datasets with the older ice model, this does not introduce a bias in the reconstruction performance. The HIM used is SIBYLL 2.3d for all datasets except 20904, which utilizes SIBYLL 2.3c. SIBYLL 2.3d builds upon the improved charm production model of 2.3c by introducing a new treatment of baryon pairs and leading particle production to better describe EAS [52, 53]. The primary energy ranges from $10^4\,\mathrm{GeV}$ to $10^{11}\,\mathrm{GeV}$ for datasets 2277X and 2353X, while it ranges from $600\,\mathrm{GeV}$ to $10^8\,\mathrm{GeV}$ for datasets 23111 and 20904. For datasets 2277X and 2353X, the primary energy is injected per nucleus, while for datasets 23111 and 20904, it is injected per nucleon. This means that for iron primaries, for example, the injected energy is multiplied by 56 to obtain the energy per nucleus.

The simulated and weighted target and proxy distributions of the combined MC dataset are shown in Figure 8.17. The total number of events after all selection cuts is approximately 449559, in comparison to 145292 events utilizing only 2277X. This corresponds to an increase by roughly a factor of three. The weighted distributions are shown for a livetime of 12.12 years and weighted to H3a. The agreement between the different datasets in the combined spectrum demonstrates the consistency of the weighting procedure. The weighting of the combined dataset is performed using the `simweights` package [105], consistent with the single dataset weighting described in Section 5.1.2.

For the full sample unfolding, a total of 540774 experimental data events from the 12.12 years of IceCube data pass all selection cuts. The comparison of the MC and experimental data statistics is shown in Figure F.1. While the combined MC dataset provides more events than the experimental data above 100 TeV, at lower energies more data events than MC events are available. However, it is not necessary to have more MC events than data events for a successful unfolding. Rather, it is essential that the simulation covers the entire phase space of physical parameters relevant to the measurement, including the full range of incoming directions (zenith and azimuth angles), different primary particle

types (from proton to iron nuclei), and the complete energy range. Adequate MC statistics in each bin of the response matrix is crucial for a robust estimation of the detector response and to avoid biases in the unfolding. For clarity, the following dataset groupings are used in the analysis: *Combined MC* refers to all MC sets together, *Systematics MC* refers to the SnowStorm datasets that include systematic ice and detector variations (2277X, 2353X, and 23111), and *EHist MC* refers to the datasets simulated with the extended history option for prompt and conventional tagging (2277X and 2353X).

The entire burnsample analysis has been repeated using the combined MC dataset to ensure consistency. The data–MC agreement for the combined set (see Figure E.1b) is comparable to that of the single dataset 2277X (see Figure E.1a). The effective area and response matrix are shown in Figure E.1. Dataset 20904 accounts for coincident primary events, which explains the population of events with a true leading muon surface energy below 10^4 GeV (marked by the shaded area). These events are expected to fall into the underflow bin of the unfolding.

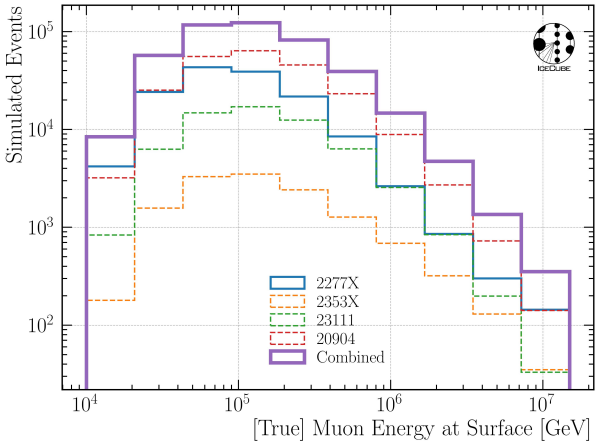
The burnsample analysis presented in Appendix E serves as a consistency check and is performed using the exact procedure of the full sample unfolding. This procedure introduces several refinements compared to the preliminary burnsample analysis presented in Section 8.2. Specifically, the proxy binning is optimized based on the energy resolution, and the fitting of ice systematics is improved by the enhanced statistics. The effective area is recalculated using the combined MC dataset. Furthermore, the unfolding is performed for all four primary model assumptions to account for model dependencies in the detector response, with the final result being the average of these four realizations. All these modifications are explained in detail in the following sections. Despite these changes, the unfolded burnsample remains consistent with the previous result shown in Figure 8.11, while showing reduced uncertainties and less bias in the recovery of the injected spectrum (see Figure E.2b). The exact values of this consistency check are listed in Table F.2. In summary, the combined MC dataset is suitable for the full sample unfolding.

Table 8.3: This table lists the MC datasets combined for the full sample unfolding. The total number of events N_{tot} after all selection cuts is 449559 for the combined dataset. Datasets 2277X and 2353X were simulated with primary energy defined per nucleus, while datasets 23111 and 20904 used primary energy defined per nucleon. Two different simulation volumes with cylinder radius r and length h, and two different Spice ice models are used. Additionally, the table indicates whether the "Extended History" option was enabled (EHist) and if the dataset includes systematic ice and detector variations (SnowStorm). In the following, *Combined MC* refers to all datasets, *Systematics MC* refers to the SnowStorm datasets (2277X, 2353X, 23111), and *EHist MC* refers to the extended history datasets (2277X, 2353X). Detailed settings of the individual datasets are stored at *IceProd*[1] [118].

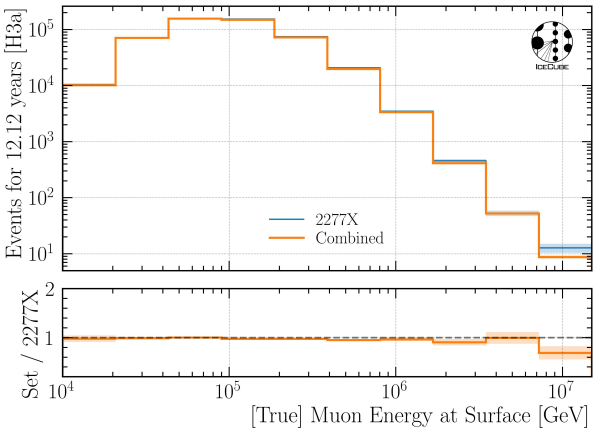
Dataset	N_{tot}	r / m	h / m	Ice Model	HIM	E_{prim} / GeV	EHist	SnowStorm
2277X	145292	700	1400	FTP-v3m	SIBYLL 2.3d	10^4–10^{11}	yes	yes
2353X	13411	700	1400	FTP-v3m	SIBYLL 2.3d	10^6–10^9	yes	yes
23111	61399	600	1200	FTP-v3m	SIBYLL 2.3d	600–10^8	no	yes
20904	229460	600	1200	3.2	SIBYLL 2.3c	600–10^8	no	no

8.3.2 Choose Proxy Binning

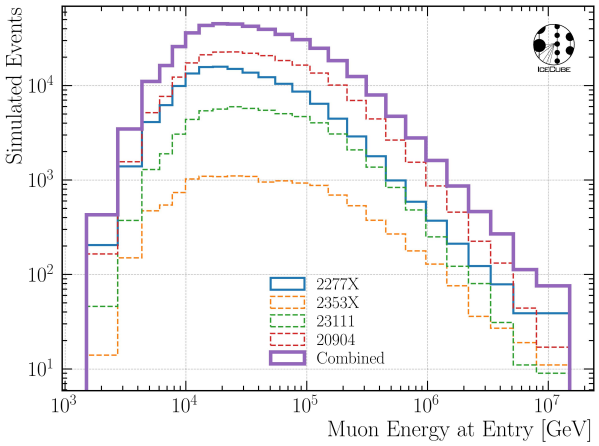
The binning of the proxy variable, the reconstructed leading muon energy, is optimized to balance the energy resolution of the detector with the stability of the unfolding result. A bin width significantly smaller than the detector resolution would lead to large migrations between bins, resulting in high correlations and an unstable unfolding result. Conversely, bins that are too wide would degrade the available spectral information [155].



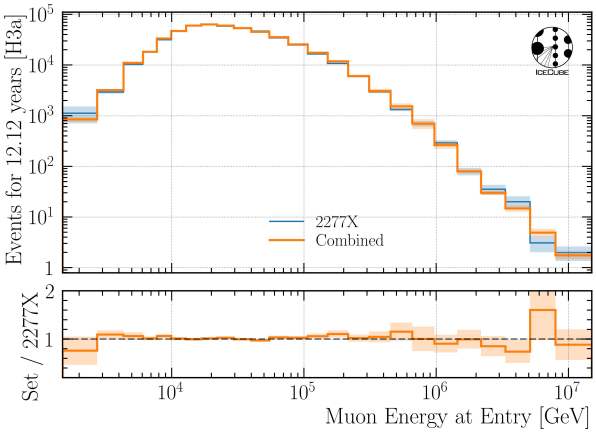
(a) Target distribution simulated statistics.



(b) Target distribution weighted statistics.



(c) Proxy distribution simulated statistics.



(d) Proxy distribution weighted statistics.

Figure 8.17: This figure presents the target and proxy distributions of the combined MC datasets used for the full sample unfolding. The left column shows the simulated statistics, while the right column shows the distributions weighted to H3a for a livetime of 12.12 years. The top row shows the target distribution of the leading muon energy at surface, while the bottom row shows the proxy distribution of the reconstructed leading muon energy at entry.

To determine the optimal binning, the energy resolution is first evaluated as a function of the true muon energy. The resolution is quantified by the standard deviation, $\sigma_{\ln E}$, of the logarithmic energy deviation, defined as $\Delta \ln E = \ln(E_{\text{reco}}/E_{\text{true}})$. This metric is chosen because the energy reconstruction uncertainty scales approximately logarithmically with energy. The distribution of $\Delta \ln E$ is analyzed in ten logarithmically spaced bins of true energy, ranging from $1.5 \cdot 10^3$ GeV to $1.5 \cdot 10^7$ GeV. This coarse binning is chosen to ensure sufficient statistics in each bin for a robust estimation of the resolution parameters, as the detector resolution varies slowly with energy. The distributions are weighted according to the H3a primary cosmic ray model. For each true energy bin, the weighted mean and standard deviation of $\Delta \ln E$ are calculated. A continuous resolution function, $\sigma(E_{\text{true}})$, is then obtained by interpolating these discrete measurements.

The bin edges for the proxy variable are defined iteratively starting from the threshold energy. The width of each bin is set to be proportional to the local energy resolution. Specifically, the step size ΔE at a given energy E is determined such that it corresponds to a fraction k of the resolution:

$$\Delta E \approx k \cdot \sigma(E) \cdot E \tag{8.64}$$

where k is a scaling factor. Several fractions k have been tested. A factor of $k = 0.7$ is found to provide a stable unfolding result for this analysis. This method ensures that the bin width is consistently adapted to the detector performance across the entire energy range, maintaining a constant ratio between the bin size and the resolution. This approach maximizes the number of bins while keeping bin-to-bin migrations and correlations at a manageable level.

The derived energy resolution, which serves as the basis for the proxy binning, is presented in Figure 8.18a. The mean bias and the standard deviation of the logarithmic energy deviation, $\Delta \ln E$, are shown for the ten initial logarithmic energy bins used to estimate the resolution function. The correlation between the true muon energy at the surface (target variable) and the reconstructed muon energy at entry (proxy variable) is depicted in Figure 8.18b. The dashed horizontal lines indicate the binning of the target distribution, while the dashed vertical lines represent the optimized proxy binning resulting from the resolution-based algorithm. The proxy bins adapt to the energy-dependent resolution, ensuring that the migration between bins is well-controlled.

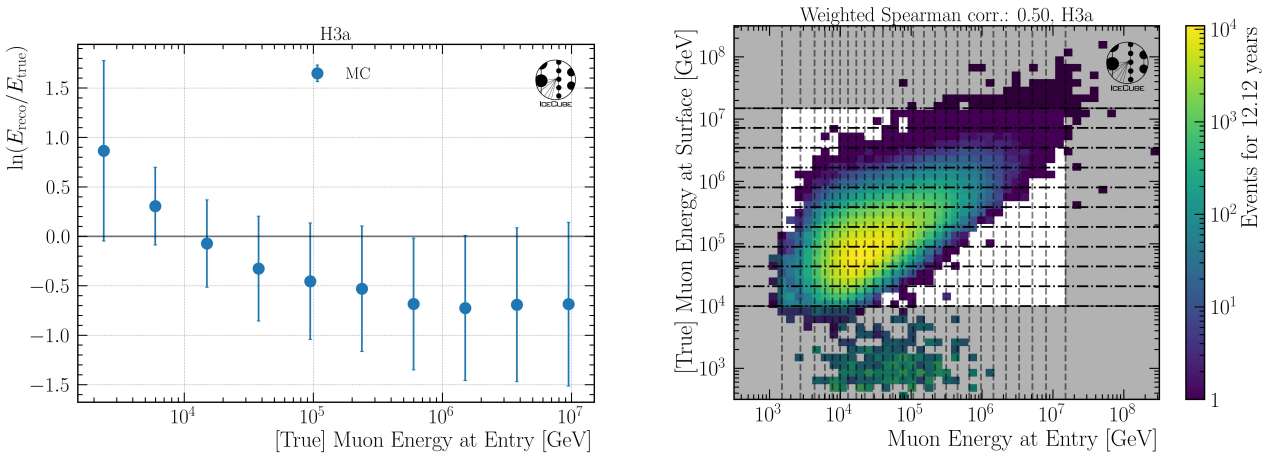
In Section F.1, the resolution estimation to determine the proxy binning is presented.

8.3.3 Ice Systematics

The treatment of ice systematics in the full sample unfolding builds upon the methodology established for the burnsample analysis as presented in Section 8.2.1 but benefits from the significantly increased statistics of the combined MC dataset. This allows for a more refined parameterization of the systematic effects, leading to a more stable unfolding result. The systematic uncertainties are parameterized by modeling the relative change in the expected event counts in each proxy bin as a function of the systematic parameters. The *Systematics MC* dataset (2277X, 2353X, and 23111) is used for this purpose, as it is generated using the SnowStorm method, which includes event-by-event variations of the ice model parameters.

For each systematic parameter s, the Monte Carlo simulations provide support points s_j within the range $[s_{\min}, s_{\max}]$. The relative change in event counts for each analysis bin i is calculated as the ratio $R_{ij} = N_i(s_j)/N_i(s_{\text{base}})$ with respect to the baseline expectation. A continuous re-weighting function $f_i(s)$ is then derived for each bin using a robust fitting procedure designed to suppress statistical fluctuations while preserving physical trends.

First, a global linear fit is performed across all bins simultaneously to determine the global sign of the correlation, $\sigma_{\text{global}} = \text{sgn}(dR/ds)$, between the systematic parameter and the overall event rate. This ensures physical consistency across the phase space. Subsequently, a linear function $f_i(s) = a_i s + b_i$ is



(a) Energy resolution.

(b) Correlation matrix.

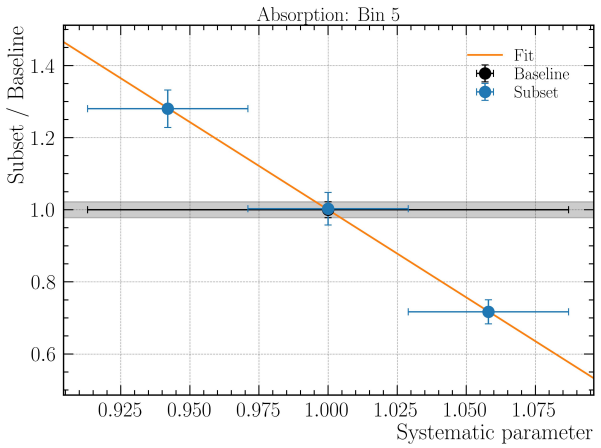
Figure 8.18: The left panel shows the energy resolution (bias and standard deviation of $\Delta \ln E$) as a function of true muon energy. The right panel displays the correlation between the true surface energy and the reconstructed leading muon entry energy. Dashed lines indicate the chosen binning for the unfolding (horizontal) and the proxy variable (vertical). The shaded areas indicate the under and overflow bins. The second distribution below 10 TeV target energy is caused by coincident primaries, see Section 8.2.8.

fitted to the ratios R_{ij} for each individual bin i, constrained to pass exactly through the baseline point ($f_i(s_{\mathrm{base}}) = 1$).

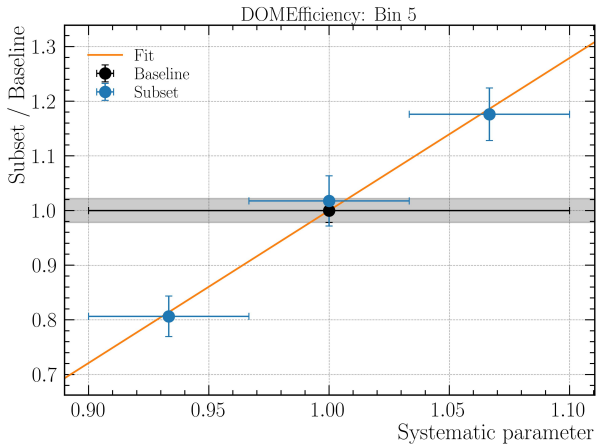
To ensure robustness, the slope a_i of the local fit is compared to the global trend. If the local slope agrees with the global sign, the fit is accepted. However, if the local slope contradicts the global trend, likely due to statistical fluctuations, the fit is rejected and replaced by a fallback linear function. This fallback function respects the global sign and assumes a conservative magnitude of a 5% deviation at the parameter boundary. Finally, a χ^2 test is performed to assess the statistical significance of the systematic effect. If the data is compatible with the null hypothesis of no systematic effect ($f_i(s) = 1$) with a p-value greater than 0.99, the re-weighting function is set to unity for that bin. This limit is set high to ensure that the fitted parameterization is retained for the majority of bins, thus modeling at least a minimum systematic impact. This rigorous selection process ensures that physically consistent systematic effects are included in the unfolding model.

8.3.4 Effective Area

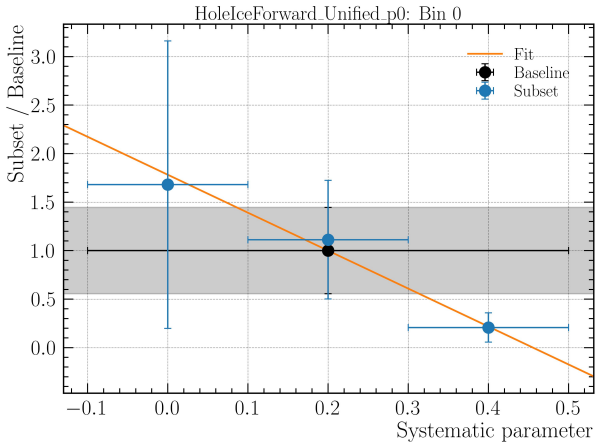
The effective area for the full sample unfolding is calculated following the methodology described in Section 8.2.2. This calculation utilizes the *EHist MC*, since it is generated with the larger detector volume. The impact of the systematic uncertainties on the effective area is presented in Figure 8.20. Figure 8.20a displays the statistical uncertainty alongside the total systematic uncertainty, which is derived from the quadratic sum of the individual contributions of the five ice systematics (see Eq. (8.32)). The individual panels illustrate the isolated impact of each systematic effect. In the final unfolding, only the statistical uncertainty of the effective area is propagated. The systematic uncertainties are excluded at this stage because the ice systematics are directly incorporated into the unfolding likelihood via nuisance parameters (see Section 8.1.1). Including them here would result in double counting.



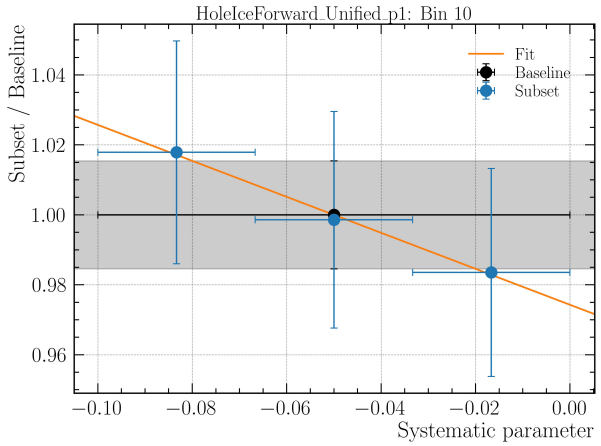
(a) Absorption.



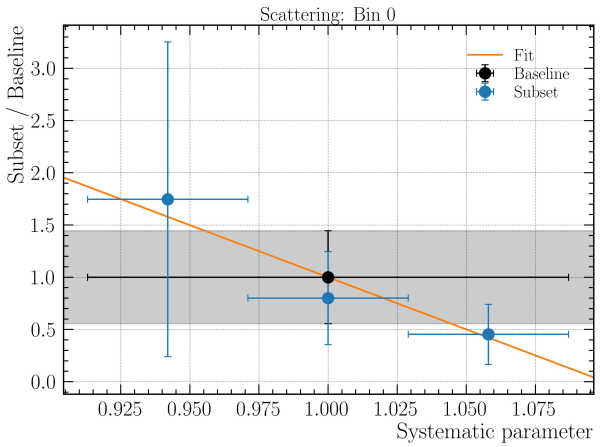
(b) DOM efficiency.



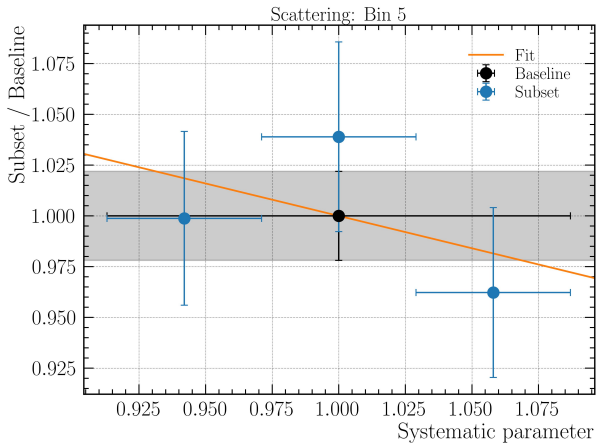
(c) Hole ice p0.



(d) Hole ice p1.

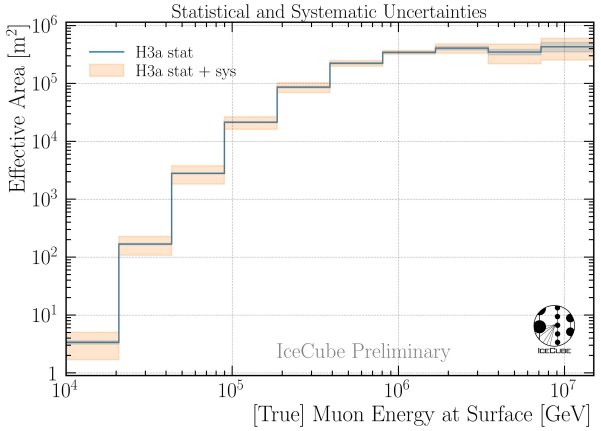


(e) Scattering.

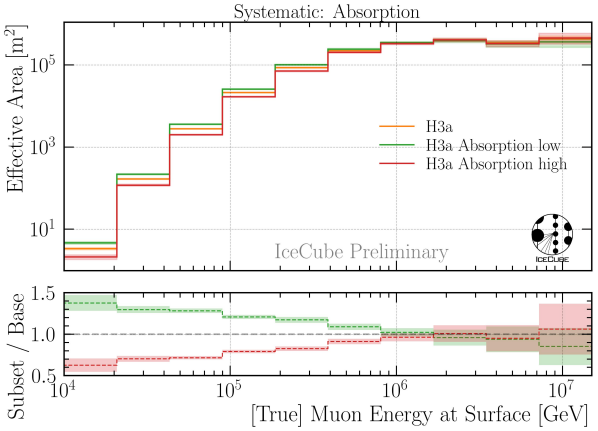


(f) Scattering.

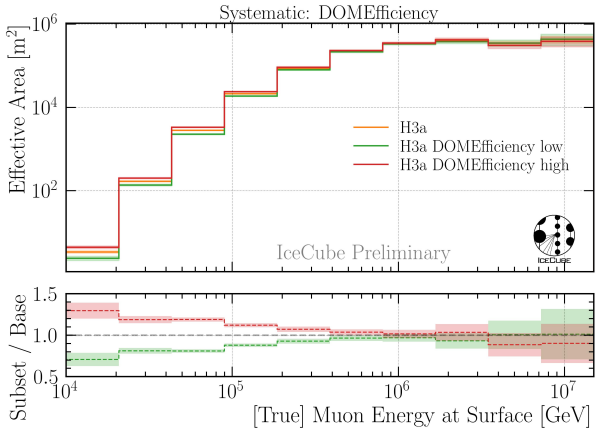
Figure 8.19: Selection of linear fits parameterizing the ice model systematics. These weight functions describe the relative change in event rates and are incorporated into the unfolding as nuisance parameters. In total, 125 such functions are utilized in the full sample analysis. The fits are derived from the *Systematics MC* dataset, which increases the statistics in comparison to 2277X only.



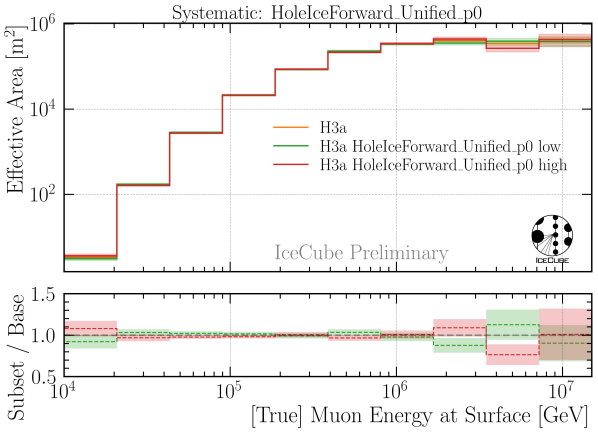
(a) Statistical and systematic uncertainties.



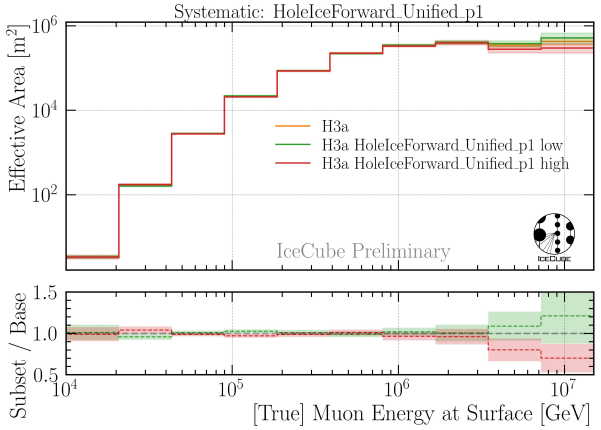
(b) Absorption.



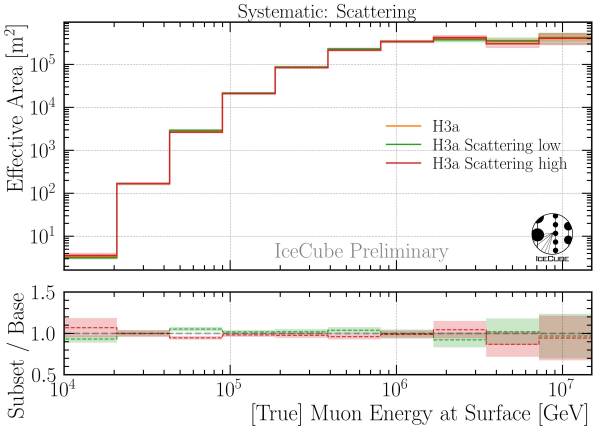
(c) DOM efficiency.



(d) Hole ice p0.



(e) Hole ice p1.



(f) Scattering.

Figure 8.20: Effective area used in the full sample unfolding. The panels display the individual contributions of the five ice systematics alongside the total uncertainty. The orange line represents the baseline expectation from the full SnowStorm dataset. Green and red lines indicate the deviations resulting from systematic variations towards the lower and upper bounds of the parameter range, respectively. Panel (a) compares the statistical uncertainty (blue) with the total systematic uncertainty calculated via Eq. (8.32). Although calculated using the H3a primary model, the effective area is independent of this choice as the flux normalization cancels out in Eq. (8.10). The calculation utilizes the *EHist MC* dataset.

8.3.5 Regularization

In this section, the regularization strength τ for the full sample unfolding is determined using the global correlation method, as introduced in Section 8.2.3. Figure 8.21a shows the correlation profile for an MC test where both the response matrix and the pseudo-data are derived from the H3a primary model. A polynomial function is fitted to the profile to identify the optimal regularization strength at the minimum correlation. However, the profile exhibits some roughness near the minimum, likely due to statistical fluctuations. The absolute minimum of the scan is found at $\tau = 0.00135$, while the minimum of the polynomial fit lies at $\tau = 0.0018$.

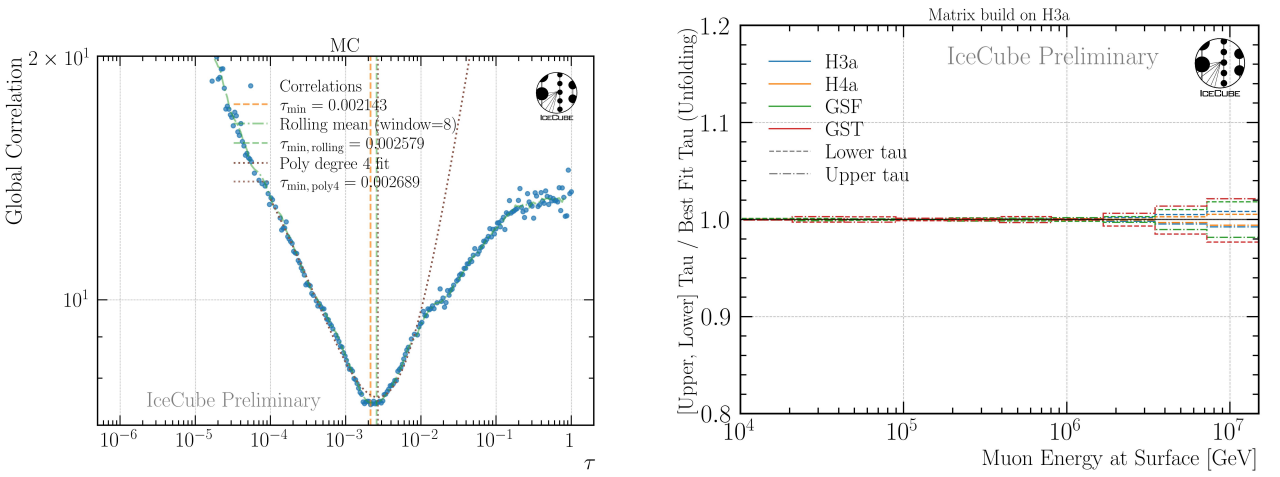
To quantify the systematic uncertainty associated with the choice of τ, the unfolding is repeated with varied regularization strengths. The upper and lower bounds for τ are defined by the points where the polynomial fit increases by 0.5 % relative to the minimum. These bounds are chosen to encompass the typical fluctuations observed between the absolute minimum and the polynomial minimum, as visualized in Figure F.4a. The impact of these variations on the unfolded flux is shown in Figure 8.21b. The deviation increases towards higher energies but remains below 3 %. This optimization and stability test is performed for all 16 combinations of response matrices and injection models (see Section F.2). In all cases, the impact remains below 5 %. compared to the other systematic uncertainties (see Figure 8.12), the uncertainty due to the choice of regularization strength is subdominant and is therefore neglected in this analysis.

Furthermore, Tikhonov regularization minimizes the curvature of the solution vector, effectively biasing it towards a smooth or flat shape. Since the atmospheric muon flux follows a steeply falling power law, applying regularization directly to the flux would introduce a bias. To efficiently use the regularization, the unfolding should be performed on a quantity that is approximately flat. In the burnsample analysis, the input distribution was scaled by the inverse of the effective area. This operation converts the event counts to a flux-like quantity but does not remove the steep spectral slope. Consequently, the regularization still acted on a steeply falling distribution, which is not optimal. For the full sample analysis, this is improved by using a physical reference model to flatten the distribution. The true MC distribution weighted to the H4a model is used as a prior. The regularization is then applied to the ratio of the unfolded flux to this reference model. Since this ratio is expected to be close to unity and spectrally flat, the regularization performs without introducing spectral bias. H4a is chosen as the reference because it predicts an intermediate flux between the lower bound (H3a) and the upper bound (GSF) at high energies (see Figure 8.14a). Any residual bias regarding the choice of this reference model is covered by the robustness tests detailed in the following section.

8.3.6 Robustness Tests

To validate the stability of the unfolding result against assumptions on the primary cosmic ray flux, a dedicated injection test is performed. The *Combined MC* dataset is weighted according to the four primary models H3a, H4a, GST, and GSF. This yields four distinct response matrices, each reflecting the detector response under a specific primary model assumption. Similarly, four pseudo-data distributions are constructed, representing the expected observations for each model.

A complete cross-check is conducted by unfolding each of the four pseudo-data distributions with each of the four response matrices, resulting in a total of 16 unfolding realizations. For each combination, the unfolded flux is compared to the true injected flux to quantify the bias. The corresponding bias curves for the full sample analysis using the *Combined MC* dataset are presented in Figure 8.22. In the first eight bins, the deviations are approximately symmetric and range from 2 % to 10 %. The penultimate bin exhibits an overshoot of up to 24 %, depending on the model combination. In the highest energy bin, the deviation reaches −41 % when GST is used for both the response matrix and the injection. For most other combinations, the deviation remains below 30 %. In contrast, the H3a



(a) Matrix H3a, injection H3a.

(b) Impact of regularization strength variation.

Figure 8.21: This figure presents the determination of the regularization strength for the full sample unfolding on the *Combined MC* for a matrix and injection test set both weighted to H3a. The left panel shows the global correlation profile. The optimal regularization strength is chosen at the minimum of this profile. The right panel shows the impact of varying the regularization strength on the unfolded flux. The nominal unfolding is performed at $\tau = 0.0018$, while the lower and upper values correspond to an increase of 0.5 % of the polynomial fit at the minimum, as presented in Figure F.4a. The impact of the regularization strength on the unfolded flux is below 2 % across the entire energy range.

injection generally shows a positive shift of approximately 10 %, except when unfolded with the GST matrix.

The maximum deviation observed in each energy bin across all model combinations is interpreted as the systematic uncertainty arising from the model dependence of the unfolding. This relative uncertainty, $b_{\text{unf,bias}}$, is converted into an absolute uncertainty via

$$\sigma_{\text{unf,bias}} = b_{\text{unf,bias}} \cdot \Phi_{\text{unf}} \,, \tag{8.65}$$

which is added in quadrature to the total uncertainty. The resulting values for $b_{\text{unf,bias}}$ are listed in Table 8.4.

This robustness test is also performed for the burnsample analysis using the *Combined MC* dataset. The results are presented in Figure E.2b, and the corresponding values are listed in Table E.1. The deviations are largely symmetric and range from 2 % to 6 % for most bins, increasing up to 12 % in the highest energy bin.

8.3.7 Prompt Sensitivity Test

To assess the sensitivity of the unfolding procedure to the prompt muon component, a dedicated test is performed using datasets 2277X and 2353X. These datasets retain the ancestry of individual muons, allowing for a distinction between prompt and conventional origins. Figure 8.23 presents the results for the H3a (top row) and GSF (bottom row) primary models. The left column illustrates the scenario where prompt muons are included in the injected MC sample. Here, the unfolding is expected to recover the total flux, indicated by the agreement of the ratio with unity (grey solid line) within uncertainties. The right column displays the results when prompt muons are explicitly excluded from the injected sample. In this case, the unfolded flux is expected to match the conventional-only flux

Table 8.4: This table lists the unfolding bias values $b_{\text{unf,bias}}$ for all 16 combinations of primary models used to build the response matrix and to weight the test MC set. The rows present the maximum downward (Bias low), upward (Bias up), and symmetric (Bias symmetric) deviations in each energy bin of the unfolding result compared to the true injected flux. The symmetric values correspond to the maximum deviation, and they are visualized in Figure 8.22. These values are determined for the full sample unfolding using the *Combined MC* dataset.

Uncertainty per bin [%]										
Bin	1	2	3	4	5	6	7	8	9	10
Bias low ↓	2.0	0.7	7.0	4.2	5.8	4.0	5.7	7.0	24.2	9.5
Bias up ↑	3.5	1.4	6.6	3.7	6.3	4.5	10.2	8.4	2.9	41.1
Bias symmetric ↕	3.5	1.4	7.0	4.2	6.3	4.5	10.2	8.4	24.2	41.1

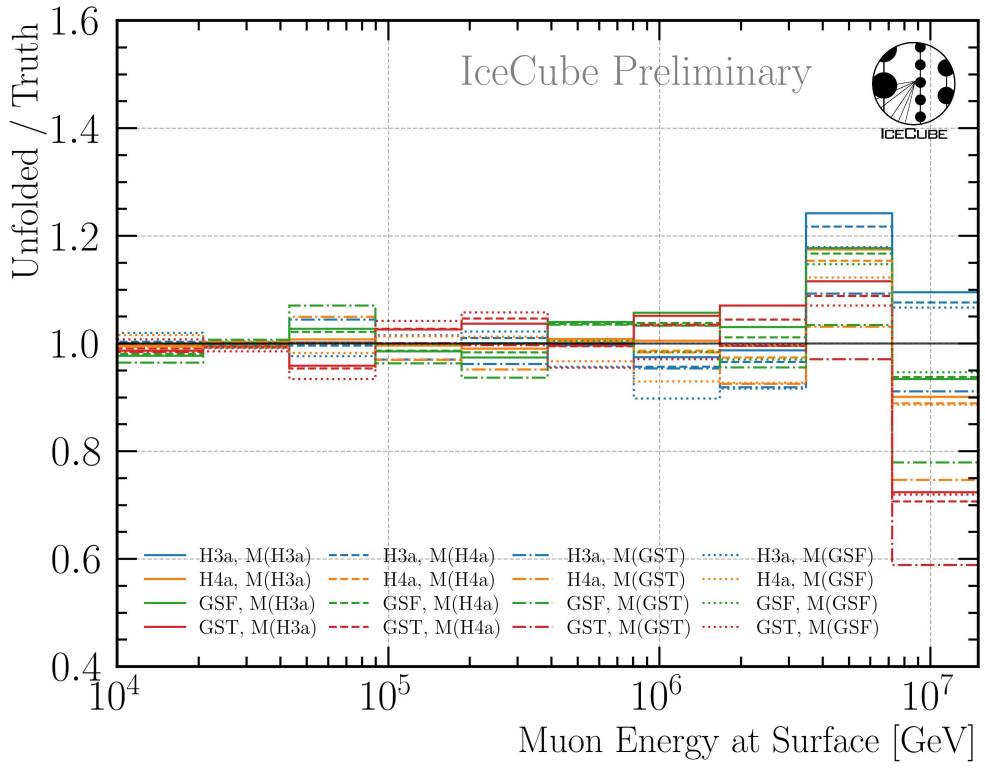
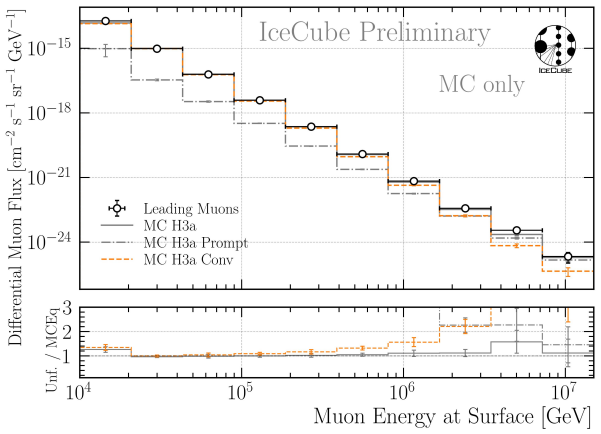
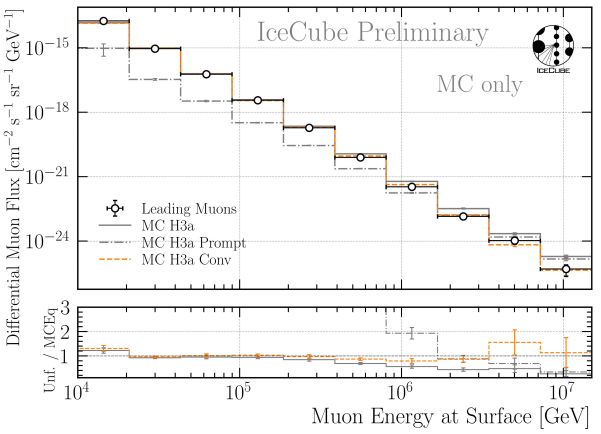


Figure 8.22: Unfolding bias for 16 combinations of primary models. The four primary models are used to build the response matrix (H3a, H4a, GST, GSF), and the *Combined MC* set weighted to these models is then unfolded. The curves show the ratio of the unfolded flux to the true flux. This test quantifies the systematic uncertainty associated with the primary model dependence of the response matrix.

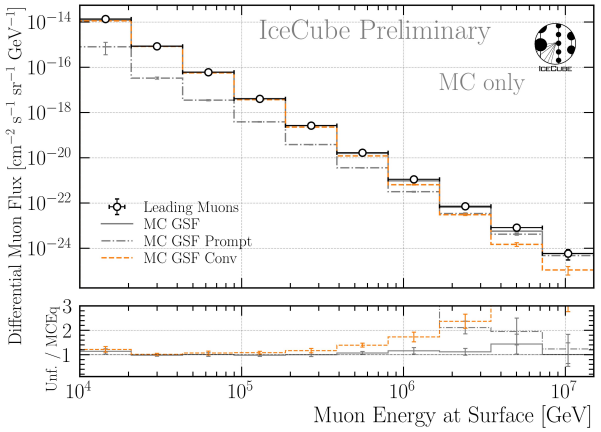
(orange dashed line). This test confirms that the unfolding procedure accurately distinguishes between scenarios with and without prompt muon contributions and reliably recovers the respective injected spectra under different primary model assumptions. For this test, the *EHist MC* is used.



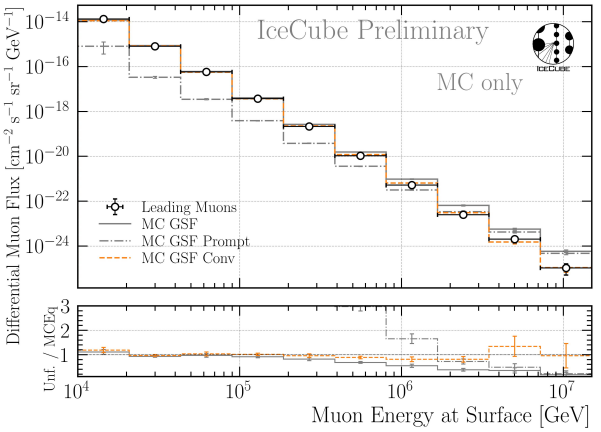
(a) H3a with prompt.



(b) H3a without prompt.



(c) GSF with prompt.



(d) GSF without prompt.

Figure 8.23: Results of the prompt sensitivity test for the full sample unfolding using the *EHist MC* is shown. The top row corresponds to the H3a primary model, and the bottom row to GSF. The left column shows the unfolding of the complete muon flux (including prompt), where the result (grey solid line) is expected to agree with the total injected flux (unity ratio). The right column shows the unfolding of an injected sample with prompt muons excluded. Here, the result is expected to agree with the conventional-only flux (orange dashed line). The agreement within uncertainties in both scenarios demonstrates the capability of the procedure to correctly identify the presence or absence of the prompt component.

8.3.8 12 Years of MC Data Unfolding

Prior to the application to experimental data, the unfolding procedure is validated using the *Combined MC* dataset. Following the methodology of the robustness tests (see Section 8.3.6), four response matrices are generated, each based on a different primary model (H3a, H4a, GST, and GSF). These matrices are used to unfold MC test distributions weighted to each of the four primary models, resulting in a total of 16 unfolding realizations. All statistical and systematic uncertainties, as well as the unfolding bias, are incorporated into the result. The unfolded spectra are compared to the true injected fluxes from the CORSIKA simulations in Figure F.9. The procedure successfully recovers the true

injected flux within uncertainties for all model combinations. Additionally, an average result is derived for each primary model by combining the four unfoldings obtained with different response matrices.

Averaging of multiple unfoldings and uncertainty propagation. To obtain a robust estimate of the muon flux, the unfolding is performed using four different response matrices, corresponding to the primary models H3a, H4a, GST, and GSF. While keeping the input data identical, this yields four unfolded flux spectra $\hat{\phi}_k$ $(k = 1, \ldots, 4)$, each with its associated bin-wise uncertainty $\sigma_{\mathrm{unf},k}$.

In addition to the unfolding uncertainty $\sigma_{\mathrm{unf},k}$ derived from the likelihood scan (see Section 8.1.1), the relative symmetric unfolding bias $b_{\mathrm{unf,bias}}$ and the relative effective area uncertainty $\delta_{A_{\mathrm{eff}}}$ are taken into account. The latter is calculated from the effective area A_{eff} and its statistical uncertainty $\sigma_{A_{\mathrm{eff}}}$ as

$$\delta_{A_{\mathrm{eff}}} = \frac{\sigma_{A_{\mathrm{eff}}}}{A_{\mathrm{eff}}} \,. \tag{8.66}$$

The total uncertainty for each unfolding variant k is obtained by adding the individual contributions in quadrature:

$$\sigma_k = \sqrt{\sigma_{\mathrm{unf},k}^2 + \left(b_{\mathrm{unf,bias}}\,\hat{\phi}_k\right)^2 + \left(\delta_{A_{\mathrm{eff}}}\,\hat{\phi}_k\right)^2} \,. \tag{8.67}$$

The final unfolded flux is determined as the arithmetic mean of the four variants:

$$\bar{\Phi}_{\mathrm{unf}} = \frac{1}{4}\sum_{k=1}^{4} \hat{\phi}_k \,. \tag{8.68}$$

Since all variants are derived from the same experimental data, their statistical uncertainties are assumed to be fully correlated. Consequently, the uncertainty of the averaged spectrum is given by the average of the individual uncertainties:

$$\sigma_{\mathrm{corr}} = \frac{1}{4}\sum_{k=1}^{4} \sigma_k \,. \tag{8.69}$$

To quantify the systematic uncertainty arising from the choice of the primary model in the response matrix, the sample standard deviation of the four unfolded spectra is calculated:

$$\sigma_{\mathrm{spread}} = \sqrt{\frac{1}{4-1}\sum_{k=1}^{4} \left(\hat{\phi}_k - \bar{\Phi}_{\mathrm{unf}}\right)^2} \,. \tag{8.70}$$

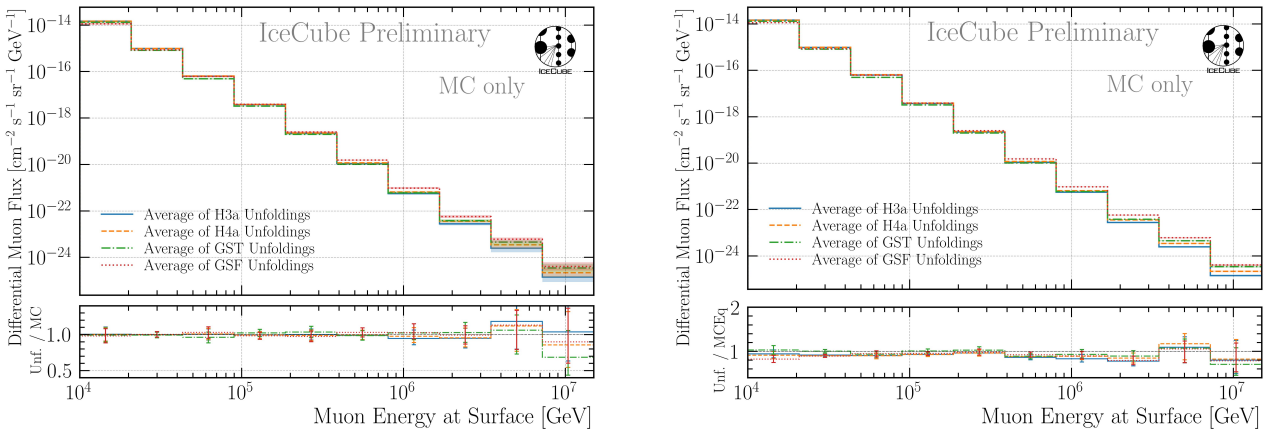
Note that $\bar{\Phi}_{\mathrm{unf}}$ is used here instead of Φ_{unf}.

Finally, the total uncertainty on the averaged unfolded flux is obtained by adding the correlated uncertainty component and the spread in quadrature:

$$\sigma_{\mathrm{tot}} = \sqrt{\sigma_{\mathrm{corr}}^2 + \sigma_{\mathrm{spread}}^2} \,. \tag{8.71}$$

The averaged unfolding result is compared to the true injected flux from CORSIKA simulations and to `MCEq` calculations in Figure 8.24. Some deviations are observed at the highest energies. However, within the uncertainties, the average unfolded flux is in agreement with both CORSIKA and `MCEq` predictions for all primary models.

The uncertainties per bin are presented in Figure F.10 for the unfolding with the H3a matrix and H3a injection. Similar to the burnsample unfolding, the total uncertainty is dominated by the unfolding bias, followed by statistical uncertainties.



(a) Compare to CORSIKA.

(b) Compare to MCEq.

Figure 8.24: This figure shows the comparison of the average unfolding result on the *Combined MC* for all four primary models used to build the response matrix. For example, for *Average of H3a Unfoldings*, all four unfoldings build on different primary models to unfold an H3a dataset are averaged. The left panel presents the comparison to the true injected fluxes from CORSIKA simulations, while the right panel shows the comparison to `MCEq` calculations. Within the uncertainties, the average unfolded flux is in agreement with both CORSIKA and `MCEq` predictions.

8.3.9 12 Years of Experimental Data Unfolding

The experimental data is unfolded following the same procedure as the MC validation. To account for model dependencies, the data is unfolded four times, each time using a response matrix derived from a different primary model (H3a, H4a, GST, and GSF). These four realizations are then averaged to obtain the final result, including all systematic uncertainties as described in Paragraph 8.3.8.

The stability of the result is evaluated by comparing the four individual unfoldings to their average, as shown in Figure 8.25a. In the lower energy range, the spread between the models is small ($< 5\,\%$). At the highest energies, the spread increases to roughly $10\,\%$. However, this variation is significantly smaller than the total uncertainty (see Figure F.13), confirming that the choice of the primary model for the response matrix has only a minor impact on the final result.

The averaged unfolded flux is compared to CORSIKA simulations in Figure 8.25b. The experimental result generally exceeds the predictions. Compared to H3a and H4a, the measured flux is about $20\,\%$ higher at the threshold energy of $10\,\mathrm{TeV}$. At energies above $200\,\mathrm{TeV}$, the deviation increases significantly, reaching a factor of two for H3a. The GST model predicts a lower flux overall, resulting in a constant offset of roughly $30\,\%$ up to $200\,\mathrm{TeV}$. The GSF model shows the best agreement at high energies ($\geq 100\,\mathrm{TeV}$), although it underestimates the flux by approximately $50\,\%$ in the lower energy region. All simulations presented here are based on the SIBYLL 2.3d HIM.

Comparisons to `MCEq` calculations using SIBYLL 2.3c and DPMJET-III 19.1 are shown in Figure 8.25c and Figure 8.25d. The predictions using SIBYLL 2.3c show similar trends to the CORSIKA results, which is expected due to the similarity of the interaction models. The GSF model combined with SIBYLL 2.3c provides the best description of the data at high energies. In contrast, calculations using DPMJET-III 19.1 do not properly describe the measured flux for any of the primary models. They generally underestimate the flux in the lower energy region and do not match the spectral shape at higher energies.

To quantify the agreement, a χ^2 value is computed for the default model normalization assuming the

Table 8.5: χ^2 values for comparisons of the unfolded flux to model predictions. The MCEq block lists all combinations of four primary models with four hadronic interaction models (SIBYLL 2.3c, DPMJET-III 19.1, EPOS-LHC, QGSJet-II-04). The CORSIKA columns show results for SIBYLL 2.3d using the nominal weights and the scaled weights from Table F.18. *EHist MC* is used for this test.

	MCEq				CORSIKA	CORSIKA (scaled)
Primary Model	S2.3c	DPMJET	EPOS-LHC	QGSJET	S2.3d	S2.3d
H3a	52.88	56.12	69.99	182.09	68.51	44.61
H4a	36.67	34.75	60.28	157.47	41.63	27.06
GST-3gen	131.17	198.84	134.42	313.74	127.91	18.27
GSF	13.50	68.75	12.26	108.13	38.99	25.19

total muon flux prediction,

$$\chi^2 = \sum_i \frac{(y_i - \mu_i)^2}{\sigma_i^2}\,, \tag{8.72}$$

with y_i the unfolded flux in bin i, μ_i the corresponding model prediction, and σ_i the total uncertainty in that bin. This test is applied for each combination of primary model and HIM using `MCEq` and CORSIKA predictions (Table 8.5). The CORSIKA comparison is performed with the standard weights and with weights scaled to the detected event number. As expected, the scaled weights yield lower χ^2 values and thus better agreement. The two best agreements are obtained for GSF with `MCEq` EPOS-LHC ($\chi^2 = 12.26$) and GSF with `MCEq` SIBYLL 2.3c ($\chi^2 = 13.50$). EPOS-LHC does not include a charm component, so part of the prompt contribution is absent. The good agreement is driven by the low-energy region, in particular the first two energy bins, where SIBYLL shows a similar level of agreement. At high energies, SIBYLL describes the data better, but because the uncertainties are larger at high energy than in the low-energy region, the χ^2 test prefers EPOS-LHC overall. The predictions for the three best agreements with `MCEq` are presented in Figure F.16.

The uncertainties per bin for the four unfoldiungs on experimental data are shown in Figure F.13 and listed in Table F.3. Consistent with the MC validation, the total uncertainty is dominated by the systematic unfolding bias, followed by statistical uncertainties.

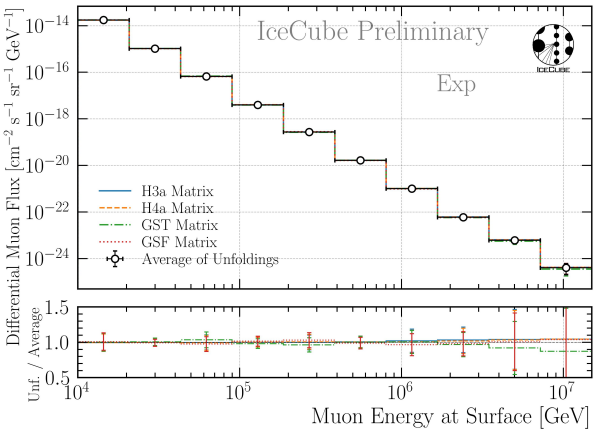
The agreement between experimental data and the simulation is evaluated in detail in Section F.4.1 by comparing the leading muon energy distributions before and after the unfolding fit. Initially, the nominal simulation shows reasonable consistency with the data, over almost the entire energy range. To assess the post-fit agreement, the simulation is re-weighted using the best-fit systematic parameters derived from the unfolding. While a direct Gaussian re-weighting is limited by finite simulation statistics, alternative linear and spectral re-weighting methods demonstrate improved compatibility. Specifically, the spectral re-weighting, which adjusts the simulation to the unfolded flux, achieves the best agreement across the entire energy range, confirming the self-consistency of the result.

8.3.10 Flux Normalization Measurement

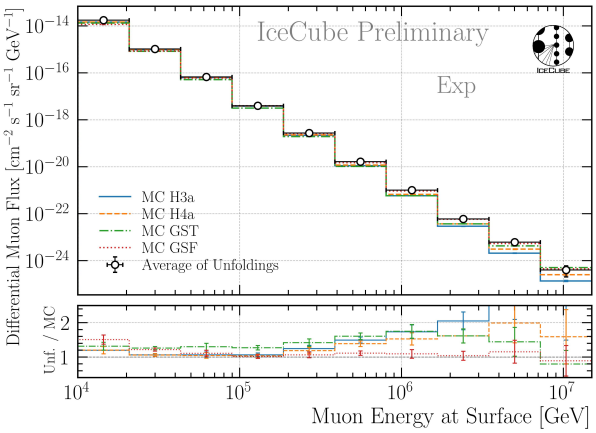
To measure the flux normalization of the prompt component, a χ^2 minimization method is employed. A custom Python class, `LikelihoodFitter`, is developed to perform this analysis. This fitter compares the measured data against theory predictions (e.g. MC simulations or `MCEq` calculations). The predictions are split into a conventional component (non-prompt) and a prompt component.

The model expectation in each bin i, μ_i, is defined as the sum of the conventional and prompt components

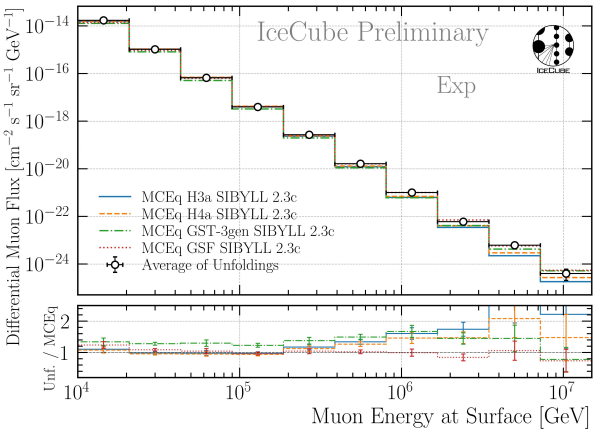
$$\mu_i(N_{\text{prompt}}, N_{\text{conv}}) = \Phi_{\text{conv},i}(N_{\text{conv}}) + \Phi_{\text{prompt},i}(N_{\text{prompt}})\,, \tag{8.73}$$



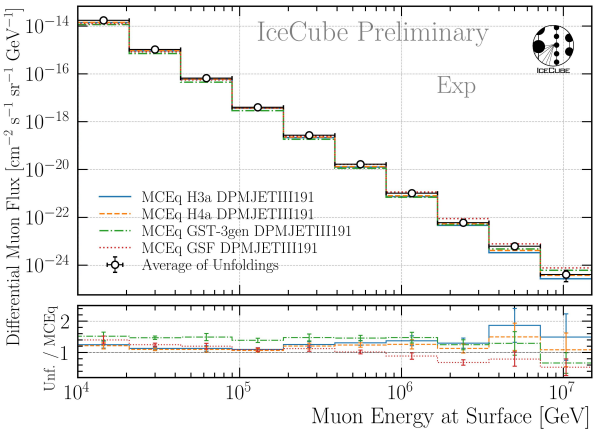
(a) Averaged unfolding compared to individual unfoldings.



(b) Averaged unfolding compared to CORSIKA.



(c) Average unfolding compared to `MCEq` SIBYLL 2.3c.



(d) Average unfolding compared to `MCEq` DPMJET-III 19.1.

Figure 8.25: Results of the experimental data unfolding. The four unfoldings, performed with different primary models, are averaged with propagated uncertainties as described in Section 8.3.8. The top left panel allows for a comparison of the individual unfoldings with their mean. The spread is shown in the ratio plot and is small compared to the total uncertainty. The top right panel compares the average result to the *Combined MC* CORSIKA simulations. The two bottom panels present comparisons to `MCEq` calculations using two different HIMs: SIBYLL 2.3c (left) and DPMJET-III 19.1 (right). Among the SIBYLL-based predictions, the GSF model shows the best agreement with the unfolded flux, while none of the DPMJET-based models describe the data well.

where $\Phi_{\text{type},i}(N_{\text{type}}) = N_{\text{type}} \cdot \Phi^{\text{baseline}}_{\text{type},i}$ is the flux of the respective component in bin i, scaled by its normalization factor N_{type}. This results in

$$\mu_i(N_{\text{prompt}}, N_{\text{conv}}) = N_{\text{conv}} \cdot \Phi^{\text{baseline}}_{\text{conv},i} + N_{\text{prompt}} \cdot \Phi^{\text{baseline}}_{\text{prompt},i} \,. \tag{8.74}$$

These normalization factors are the free parameters in the fit.

The fitting procedure minimizes the χ^2 function [163]. The fitter supports different uncertainty specifications for the measurement. If a full covariance matrix C is provided, the generalized χ^2 is computed as:

$$\chi^2(N_{\text{prompt}}, N_{\text{conv}}) = (\vec{y} - \vec{\mu})^T C^{-1} (\vec{y} - \vec{\mu}) \,, \tag{8.75}$$

where $\vec{y}$ is the vector of measured values and $\vec{\mu}$ is the vector of model expectations. If only diagonal uncertainties σ_i are provided, the standard χ^2 definition is used:

$$\chi^2(N_{\text{prompt}}, N_{\text{conv}}) = \sum_i \frac{(y_i - \mu_i)^2}{\sigma_i^2} \,. \tag{8.76}$$

The minimization is performed using the `scipy.optimize.minimize`[2] [164] function with the L-BFGS-B algorithm, allowing for boundary constraints on the normalization parameters (typically $N \geq 0$). To quantify the significance of the prompt component, a hypothesis test is performed. The null hypothesis H_0 assumes no prompt component ($N_{\text{prompt}} = 0$), while the alternative hypothesis H_1 allows N_{prompt} to float subject to the physical constraint $N_{\text{prompt}} \geq 0$. The test statistic is defined as the difference in χ^2:

$$\text{TS} = \chi^2_{H_0} - \chi^2_{H_1} \,. \tag{8.77}$$

Here, $\chi^2 \equiv -2 \ln L$ denotes the (profile) Gaussian deviance, i.e. the generalized chi-square constructed from the measurement covariance (or diagonal uncertainties).

Since the parameter N_{prompt} is constrained to be non-negative and thus lies on the boundary of the parameter space under the null hypothesis, the asymptotic null distribution of the test statistic is a mixture distribution with a point mass at zero,

$$\text{TS} \;\sim\; \frac{1}{2}\,\delta(\text{TS}) \;+\; \frac{1}{2}\chi^2_1 \,, \tag{8.78}$$

according to Chernoff's theorem [165]. The corresponding (one-sided) p-value is:

$$p = \frac{1}{2} \operatorname{erfc}\left(\sqrt{\frac{\text{TS}}{2}}\right) \,, \tag{8.79}$$

where $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$ is the complementary error function.

The significance of the result is usually reported in units of standard deviations (σ) of a Gaussian distribution. It is defined via the one-sided tail probability as:

$$\sigma = \sqrt{2}\, \operatorname{erfc}^{-1}(2p) \,. \tag{8.80}$$

By substituting the p-value implied by Chernoff's theorem into this definition, the significance simplifies to the square root of the test statistic:

$$\frac{\sigma}{\sqrt{2}} = \operatorname{erfc}^{-1}\left(\operatorname{erfc}\left(\sqrt{\frac{\text{TS}}{2}}\right)\right) = \sqrt{\frac{\text{TS}}{2}} \quad \Rightarrow \quad \sigma = \sqrt{\text{TS}} \,. \tag{8.81}$$

[2] `https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html`

This relationship holds under the asymptotic assumptions of Chernoff's theorem. It is valid for sufficiently large statistics and requires a well-constrained fit with an adequate model description of the data.

With this method, the flux normalization of the prompt and conventional components is measured, along with the significance of the prompt component in the data. This is done for several model assumptions. Both the CORSIKA simulations and `MCEq` calculations are used as theory predictions, with different primary models and HIMs. The *EHist MC* is used for the CORSIKA-based fits. In addition, the individual unfoldings on different primary models come with the full covariance matrix, which allows to perform the fit with the diagonal uncertainties as well as with the full covariance matrix. Also, the average unfolding result is fitted using only diagonal uncertainties. This results in five unfolding variants, that can be fitted to the four different primary models and two different HIMs (SIBYLL-2.3c and DPMJET-III-19.1) for `MCEq`, and one HIM (SIBYLL-2.3d) for CORSIKA. Furthermore, the energy range can be extended towards lower energies by including a flux unfolding from a stopping muon sample, as described below. In total, 100 different model combinations have been evaluated to determine the flux normalization.

Inclusion of Stopping Muon Sample Stopping muons in IceCube are atmospheric muons from extended air showers that enter the detector and decay after coming to rest within the instrumented volume. Unlike through going muons, stopping muons deposit their entire remaining energy inside the detector. This enables a precise reconstruction of the stopping vertex. Combined with the directional reconstruction, the propagation length through the ice can be determined, which is directly correlated with the muon energy at the surface. This technique provides sensitivity to atmospheric muon energies in the range from 600 GeV to 10 TeV [160]. It complements the leading muon sample by extending the measurement to lower energies and seamlessly connecting to the lowest energy bin of the leading muon unfolding presented in this work. An event display of a stopping muon is shown in Figure F.21.

The stopping muon analysis was developed by Lucas Witthaus at TU Dortmund University. A joint measurement combining the stopping and leading muon fluxes using a representative data subset (burnsample) was presented at ICRC 2025 [160]. Since then, the stopping muon analysis has undergone further refinements and has been approved by the IceCube Collaboration. The results presented in this thesis incorporate these improvements. However, as the full sample stopping muon analysis is ongoing, the combined fits in this work utilize the stopping muon burnsample. Note that the stopping muon unfolding is averaged over the full zenith range $0 < \cos(\theta) < 1$, whereas the leading muon unfolding is averaged over $0.2 < \cos(\theta) < 1$.

The normalization fits using the averaged unfolding are summarized in Table 8.6. First, the results of the **leading muon unfolding alone** are evaluated against the CORSIKA simulations. The prompt normalization varies significantly with the primary model assumption, ranging from $N_{\text{prompt}} = 3.481 \pm 0.365$ for H3a to $N_{\text{prompt}} = 0.548 \pm 0.204$ for GSF. The corresponding significances lie between $1.90\,\sigma$ and $6.75\,\sigma$.

When using `MCEq` calculations as the theoretical prediction, the results shift. Assuming SIBYLL 2.3c as the HIM, N_{prompt} ranges from 2.638 ± 0.238 (H3a) to 0.758 ± 0.123 (GSF), with significances between $4.37\,\sigma$ and $7.84\,\sigma$. For DPMJET-III 19.1, values range from $N_{\text{prompt}} = 1.551 \pm 0.138$ (H3a) to 0.510 ± 0.082 (GSF), yielding significances between $4.41\,\sigma$ and $7.95\,\sigma$.

The difference between the two HIMs, SIBYLL 2.3d in CORSIKA and SIBYLL 2.3c in `MCEq`, leads to shifts in the prompt normalization. For the H3a model, the prompt normalization is approximately 32 % higher in CORSIKA, whereas for GSF, it is about 28 % lower. This discrepancy stems from differences in the predicted total muon flux, which can vary by up to 20 % depending on the primary model (see Figure F.20).

An unexpected observation is the fitted conventional normalization for the leading muon sample alone, which is found to be below unity for H3a in both CORSIKA ($N_{\text{conv}} = 0.963 \pm 0.034$) and `MCEq` ($N_{\text{conv}} = 0.942 \pm 0.028$). This contradicts the expectation from the scaling factors derived in the burnsample and full sample selections (Tables 7.2 and F.18), which require the simulation to be scaled *up* by factors of 1.08 to 1.12. Since the atmospheric muon flux follows a steep power law, the total event count is strongly dominated by the low energy part of the spectrum, which consists almost entirely of conventional muons. Consequently, the fitted conventional normalization is expected to be close to the scaling factor. While it should be slightly lower than the overall scaling factor due to the prompt component's contribution to the total rate, obtaining a value below unity is inconsistent with the overall data excess. Crucially, when the **stopping muon sample** is included to constrain the low energy flux, this discrepancy is resolved. For the combination of `MCEq` SIBYLL 2.3c and H3a, the conventional normalization is fitted to $N_{\text{conv}} = 1.059 \pm 0.006$. Although this comparison is against `MCEq` and covers a lower energy range, it strongly suggests that the stopping muon sample is essential for correctly anchoring the conventional component. Consequently, the prompt normalization decreases to $N_{\text{prompt}} = 1.424 \pm 0.154$. The resulting significances for the four primary models range from $5.78\,\sigma$ to $7.18\,\sigma$.

The same fits are performed for `MCEq` calculations using DPMJET-III 19.1 regarding the inclusion of the stopping muon sample. Here, the prompt normalization for H3a results in $N_{\text{prompt}} = 1.179 \pm 0.092$, with a conventional normalization of $N_{\text{conv}} = 1.103 \pm 0.006$. For GSF, the prompt normalization is significantly lower at $N_{\text{prompt}} = 0.610 \pm 0.055$, while the conventional normalization is $N_{\text{conv}} = 1.166 \pm 0.007$. The significances for the prompt component across the four primary models range from $7.77\,\sigma$ to $9.44\,\sigma$.

Since CORSIKA simulations for the stopping muon sample are unavailable, fits including the stopping muon unfolding rely exclusively on `MCEq` predictions.

Detailed fit results for the individual unfoldings are provided in Appendix F.5, sorted by unfolding variant (Section F.5.1), primary model (Section F.5.2), and HIM (Section F.5.3). As no single primary model is currently preferred, the results from the averaged unfolding (Table 8.6) constitute the final result of this analysis.

In Figure 8.26, the final average unfolding result is presented alongside the stopping muon burnsample flux. The results are compared to `MCEq` calculations assuming SIBYLL 2.3c as the HIM, with one panel for each of the four primary models. The top part of each panel displays the fluxes, including the `MCEq` prediction for leading muons. The bottom part shows the ratio of the data to the prediction for the respective zenith ranges of the stopping ($0 < \cos\theta < 1$) and leading ($0.2 < \cos\theta < 1$) muon samples.

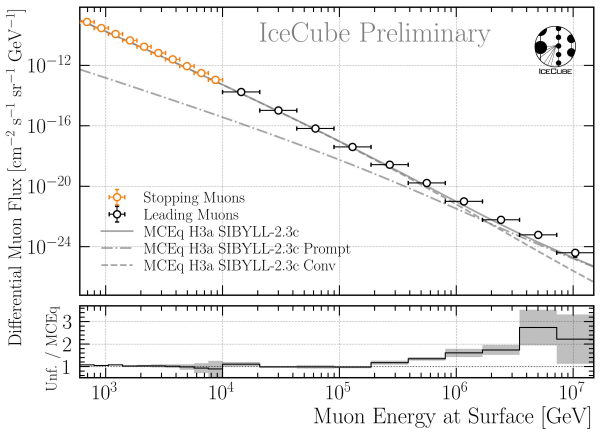
In the overlap region at 10 TeV, the fluxes agree within uncertainties; however, the nominal value of the leading muon flux is higher than that of the stopping muons. Several factors may contribute to this offset. First, the two samples cover different zenith ranges. Second, they utilize different event selections and reconstruction methods. Specifically, the leading muon unfolding relies on the reconstructed energy of the leading muon within a bundle, whereas the stopping muon unfolding uses the reconstructed propagation length of the muon track to estimate the surface energy. Finally, the stopping muon result is currently based on a burnsample, and including the full statistics could potentially shift the flux.

Regarding the high-energy tail of the spectrum (4 PeV – 15 PeV), the second-to-last bin exhibits an overshoot, while the final bin indicates a drop in flux. This behavior likely arises from the unfolding bias discussed in Section 8.3.6 (see Figure 8.22), where similar trends are observed for most primary model combinations. However, given the uncertainties, these fluctuations do not significantly impact the overall flux measurement.

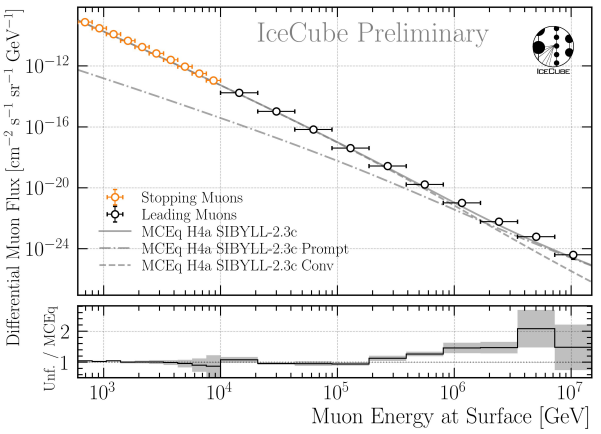
To visually demonstrate that the conventional component alone is insufficient to describe the unfolded muon flux, the result is compared to the conventional-only predictions from 16 different `MCEq` calculations in Figure 8.27. Each panel represents one of the four primary models, showing predictions for four different HIMs: SIBYLL 2.3c, DPMJET-III 19.1, EPOS-LHC, and QGSJet-II-04.

Table 8.6: Fit results for the average unfolded spectrum, as described in Paragraph 8.3.8, are compared to theoretical predictions from CORSIKA simulations based on the *Ehist MC* and `MCEq` calculations. Predictions are calculated for three different HIM: SIBYLL 2.3c, SIBYLL 2.3d and DPMJet-III-19.1, each containing prompt muon production. All predictions are weighted to the four primary composition models H3a, H4a, GST and GSF. Fits are performed for the leading muon unfolding result only, and also with the preliminary stopping muon unfolding result to extend the low-energy region. The χ^2 values quantify the agreement between the measured flux and the model prediction of the total muon flux, with the prompt and conventional components as free parameters. The degree of freedom is nDoF = 8 for the fits without the stopping muons, and nDoF = 18 including stopping muons. Due to averaging the four individual unfoldings, these fits are performed without the covariance matrix, but with the propagated uncertainties. Significance in σ and p-values represent the detection significance for the prompt component, as explained in Section 8.3.10.

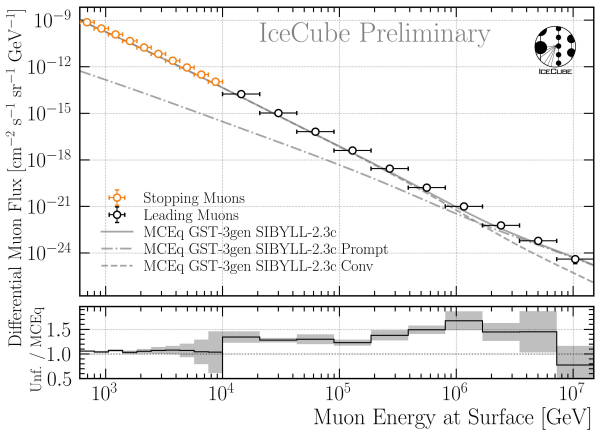
Prediction	Primary Model	Prompt Norm	Conv Norm	χ^2/nDoF	p-value	Significance/σ
CORSIKA SIBYLL-2.3d						
	H3a	3.481 ± 0.365	0.963 ± 0.034	0.38	7.38e-12	6.75
	H4a	2.392 ± 0.293	0.998 ± 0.032	0.54	4.03e-09	5.77
	GST	1.700 ± 0.263	1.266 ± 0.036	1.10	2.41e-06	4.57
	GSF	0.548 ± 0.204	1.159 ± 0.033	1.29	2.88e-02	1.90
MCEq SIBYLL-2.3c						
	H3a	2.638 ± 0.238	0.942 ± 0.028	0.47	2.34e-15	7.84
	H4a	2.146 ± 0.202	0.934 ± 0.026	0.50	2.95e-14	7.51
	GST	1.588 ± 0.179	1.289 ± 0.033	0.74	1.61e-10	6.29
	GSF	0.758 ± 0.123	1.075 ± 0.028	0.48	6.21e-06	4.37
MCEq SIBYLL-2.3c (with stopping muons)						
	H3a	1.424 ± 0.154	1.059 ± 0.006	0.82	3.27e-11	6.53
	H4a	1.352 ± 0.133	1.036 ± 0.006	0.73	3.37e-13	7.18
	GST	1.170 ± 0.123	1.058 ± 0.006	0.58	7.58e-12	6.75
	GSF	0.669 ± 0.082	1.118 ± 0.006	0.69	3.84e-09	5.78
MCEq DPMJET-III-19.1						
	H3a	1.551 ± 0.138	1.121 ± 0.031	0.31	8.98e-16	7.95
	H4a	1.326 ± 0.123	1.101 ± 0.030	0.35	1.25e-14	7.62
	GST	1.200 ± 0.129	1.485 ± 0.038	0.38	2.75e-11	6.56
	GSF	0.510 ± 0.082	1.253 ± 0.031	0.36	5.27e-06	4.41
MCEq DPMJET-III-19.1 (with stopping muons)						
	H3a	1.179 ± 0.092	1.103 ± 0.006	1.16	9.32e-20	9.02
	H4a	1.108 ± 0.083	1.078 ± 0.006	1.51	1.89e-21	9.44
	GST	1.113 ± 0.088	1.099 ± 0.006	1.59	3.00e-19	8.89
	GSF	0.610 ± 0.055	1.166 ± 0.007	0.78	3.84e-15	7.77



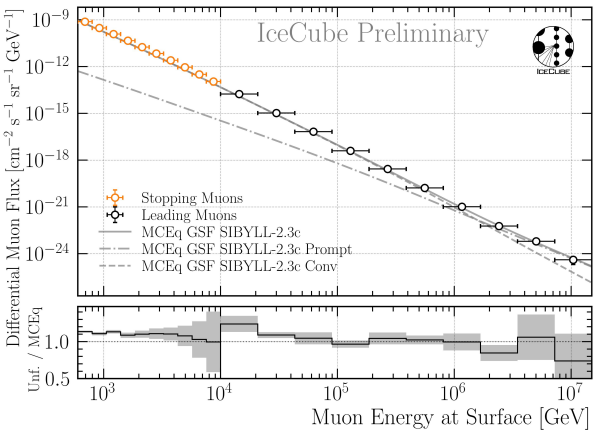
(a) `MCEq` SIBYLL 2.3c with H3a.



(b) `MCEq` SIBYLL 2.3c with H4a.



(c) `MCEq` SIBYLL 2.3c with GST.

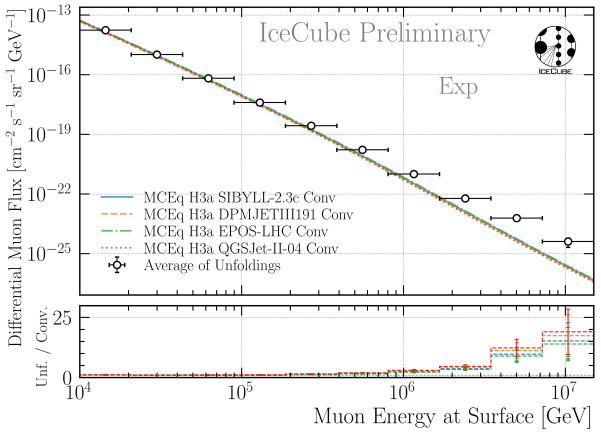


(d) `MCEq` SIBYLL 2.3c with GSF.

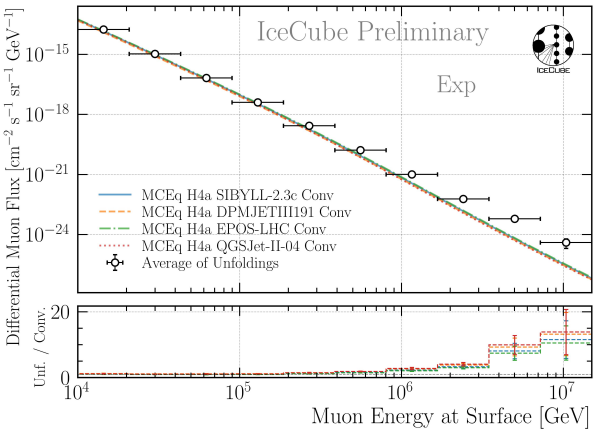
Figure 8.26: Final results of the unfolding of the atmospheric muon flux at the surface. The four unfoldings based on different primary models are averaged with propagated uncertainties as described in Section 8.3.8. Additionally, the burnsample unfolding of the stopping muon sample is included for the low-energy region. While the leading muon sample is averaged over the zenith range $0.2 < \cos\theta < 1$, the stopping muon sample covers $0 < \cos\theta < 1$. The results are compared to `MCEq` calculations assuming SIBYLL 2.3c as the HIM. The top panel of each plot displays the fluxes and the `MCEq` prediction for leading muons, while the bottom panel shows the ratio for the stopping and leading muon zenith ranges, respectively. The fluxes agree within uncertainties at 10 TeV; however, the nominal value of the leading muon flux is higher than that of the stopping muons.

Across all model combinations, the conventional component is unable to reproduce the measured flux at high energies, indicating the presence of a significant prompt muon contribution. This observation aligns with the calculated significances for the prompt component obtained from the average unfolding (see Table 8.6), which range from $4.37\,\sigma$ to $9.44\,\sigma$ for the MCEq predictions. The highest significance of $10.99\,\sigma$ is obtained for the individual unfolding on H3a fitted against MCEq with SIBYLL 2.3c and weighted to H3a.

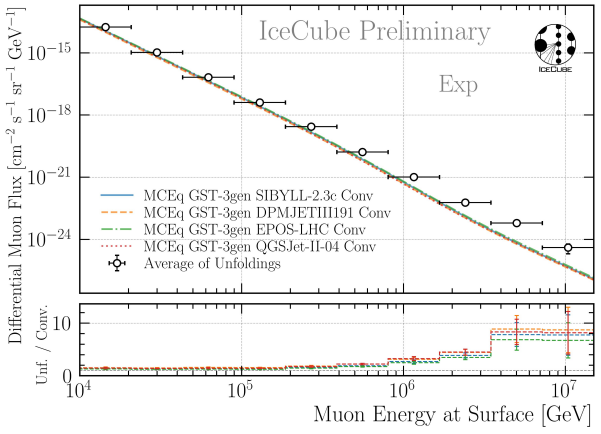
For an alternative perspective, the same comparisons are organized by HIM in Figure F.15, where each plot displays the predictions for all four primary models.



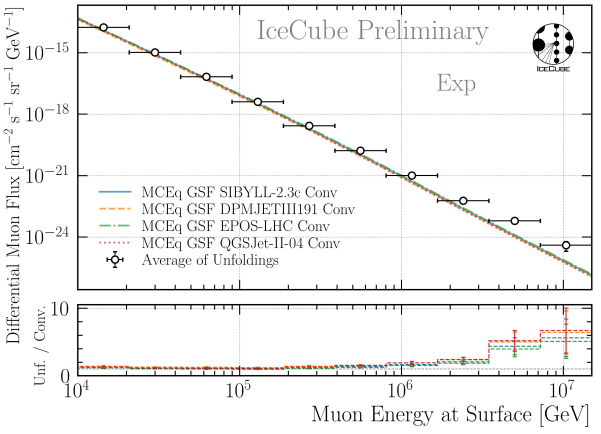
(a) MCEq with H3a conventional only.



(b) MCEq with H4a conventional only.



(c) MCEq with GST conventional only.



(d) MCEq with GSF conventional only.

Figure 8.27: This figure presents the final results of the unfolding of the atmospheric muon flux at the surface. The four unfoldings based on different primary models are averaged with propagated uncertainties as described in Section 8.3.8. The results are compared to MCEq calculations assuming the four HIM SIBYLL 2.3c, DPMJET-III 19.1, EPOS-LHC, and QGSJet-II-04 for all four primary models. Only the conventional muon flux of the MCEq calculations is shown here. The ratio panel in each plot shows that the conventional component alone is not sufficient to describe the measured flux at high energies, indicating a significant contribution of prompt muons. The same predictions are also shown in Figure F.15, where each plot shows all four primary models for a single HIM.

Finally, the unfolded flux is compared to the previous measurement by IceCube [56] in Figure 8.28. The previous analysis relied on different simulations, selections, and reconstructions. Furthermore, no unfolding was performed, and the zenith range differs. The measurement is shown for two zenith ranges: the full range $0° < \theta < 90°$, including correlated uncertainties, and a restricted range of $0° < \theta < 60°$,

presented here without uncertainties. Due to a shorter livetime of less than two years, the previous measurement extends only up to 1.5 PeV.

The nominal values of the all-sky measurement are higher than the unfolded flux, while the measurement for $\theta < 60°$ is lower. Since the angular range of the unfolded flux ($\theta < 78°$) lies between these two intervals, and considering the larger uncertainties of the previous measurement, the results are in agreement.

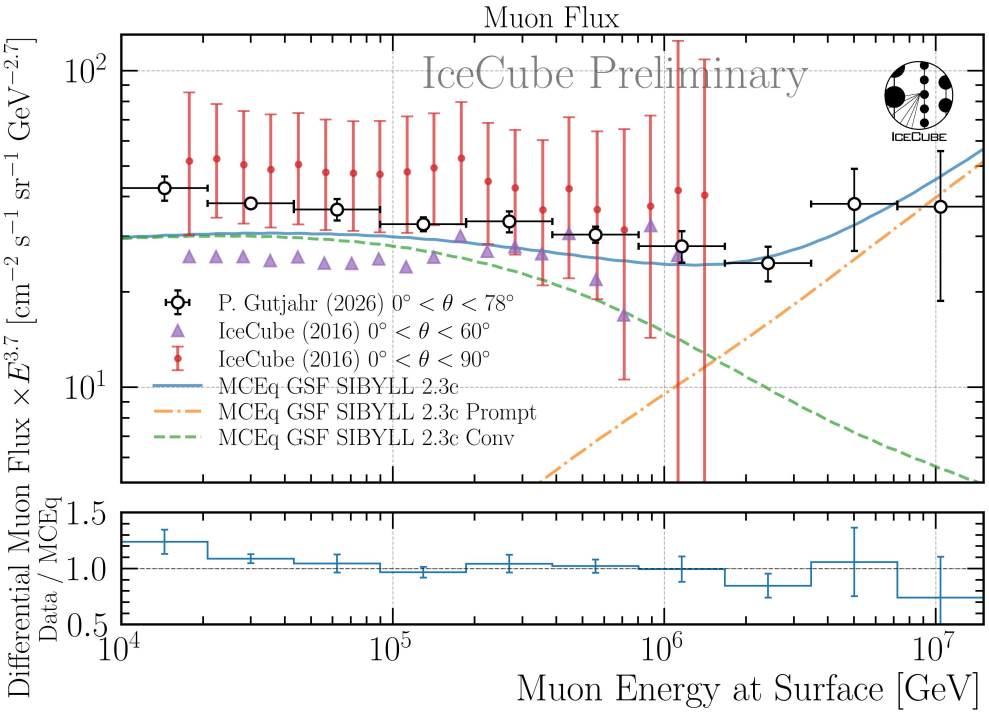


Figure 8.28: Comparison of the unfolded muon flux at the surface to the previous measurement by IceCube [56]. The bottom panel presents the ratio of the unfolded flux to the total `MCEq` prediction. The flux is weighted by $E^{3.7}$ at the logarithmic bin center to enhance the visibility of the spectral shape in the high-energy region. The unfolded flux covers zenith angles $0° < \theta < 78°$, while the previous measurement is shown for $0° < \theta < 90°$ (including correlated uncertainties) and $0° < \theta < 60°$ (without uncertainties). Given the larger uncertainties of the previous measurement and the differing zenith ranges, the two results are consistent.

8.4 Summary: Unfolding

In this chapter, the atmospheric muon energy spectrum at the surface is derived using 12.12 years of IceCube data. New simulation data created specifically for this thesis is used to model the detector response. The systematic impact of ice properties and detector response is parametrized and incorporated into the unfolding. An effective area is calculated specifically for this analysis. A new approach is used to determine the regularization strength by finding the minimum global correlation coefficient.

Initially, a burnsample with a livetime of approximately 2487 h is unfolded, covering an energy range from 20 TeV to 2 PeV (see Figure 8.11). This analysis was published at the ICRC 2025 [160]. The analysis is further refined, and to increase statistics, multiple MC datasets are combined. This results in three different set: *Combined MC*, *Systematics MC*, and *EHist MC*, referring to all sets, sets

with systematic SnowStorm variations, and sets with the information whether a muon is prompt or conventional, respectively. The burnsample analysis is repeated to verify consistency, showing good agreement (see Figure E.7b).

Subsequently, 12.12 years of experimental data are processed and unfolded from 10 TeV to 2 PeV (Figure 8.25). The normalizations for CORSIKA with H3a and SIBYLL 2.3d are fitted to $N_{\text{prompt}} = 3.48 \pm 0.37$ and $N_{\text{conv}} = 0.963 \pm 0.034$. Given the scaling factors in Table F.18, the conventional component is expected to be close to 1.10. Hence, the low energy range is extended by a stopping muon unfolding, as described in [160]. This helps to constrain the conventional component to $N_{\text{conv}} = 1.059 \pm 0.006$ (for `MCEq` SIBYLL 2.3c), leading to a prompt normalization of $N_{\text{prompt}} = 1.424 \pm 0.154$.

To confirm the existence of the prompt component, Figure 8.27 shows the measured flux against several `MCEq` predictions for the conventional component only. In the high-energy region, none of the predictions is able to describe the measured flux. This excess corresponds to a significance ranging from $4.4\,\sigma$ to $9.4\,\sigma$ for `MCEq` predictions, and $1.9\,\sigma$ to $6.8\,\sigma$ for CORSIKA predictions. Further normalization fits and significances are stated in Table 8.6.

9 Conclusion

In this thesis, a measurement of the atmospheric muon flux and its prompt component is presented using 12.12 years of data from the IceCube Neutrino Observatory. The analysis aims to constrain the prompt muon flux, which arises from the decay of heavy-quark hadrons and constitutes a background for astrophysical neutrino searches as well as a sensitive probe for QCD in the forward region.

To achieve this, new air shower simulations are performed using CORSIKA with the `EHISTORY` option, which enables storing the information about the muon parent particles. A novel method is implemented to tag muons as either prompt or conventional on an event-by-event basis. Conventional muons arise from pion and kaon decays, prompt muons from the rest. This method, published in Ref. [104], facilitates a precise separation of the flux components in the simulation. As demonstrated in Section 5.1.3, the resulting CORSIKA simulations are in agreement with theoretical predictions from MCEq, providing a robust basis for the analysis.

Based on these simulations, the first-ever reconstruction of the leading muon energy in IceCube is developed using machine learning techniques, specifically CNNs, as detailed in Section 6.2. In addition to the leading muon energy, other networks reconstruct the muon bundle energy (both in-ice and at the surface), the particle direction, the entry and closest approach to center point, and the propagation length. An uncertainty is provided for all predictions as well. While no comparable leading muon reconstruction currently exists within the IceCube analysis framework, the machine learning-based bundle energy reconstruction developed here outperforms classical reconstruction methods, as presented in Figure 6.21. Furthermore, the angular resolution achieved is compatible with state-of-the-art IceCube reconstructions, even without dedicated optimization for directionality. This angular precision proves sufficient for the unfolding analysis performed in this work.

A dedicated event selection is designed to isolate leading muons solely based on these new reconstructions, as described in Chapter 7. The selection process comprises four sequential steps. At the final selection level, good agreement between data and MC simulations is observed for all four considered primary cosmic ray models. This agreement is demonstrated for the burnsample (a small data subset used for testing) in Figure 7.11 and for the full dataset in Figure F.17.

The atmospheric muon flux at the surface is determined through an unfolding procedure. This relies on a newly calculated effective area, detailed in Section 8.1.2, and a novel, data-driven determination of the regularization strength, which is independent of prior assumptions (see Eq. (8.7)). The unfolding method is validated on the burnsample with a livetime of about 2487 h, covering a muon energy at surface range from 20 TeV to 2 PeV, and is presented at the ICRC [160]. Subsequently, the full 12.12-year dataset is processed to the final selection level using the new machine learning reconstructions. The resulting atmospheric muon flux is unfolded using a customized binning over an extended energy range from 10 TeV to 15 PeV, as presented in Figure 8.25.

At first, a χ^2 test quantifies which model best describes the unfolded flux. Among the model predictions that include charm production, the best agreement is obtained for GSF with SIBYLL 2.3c calculated with `MCEq`. The corresponding χ^2 values are listed in Table 8.5.

To characterize the prompt and conventional components, a χ^2 likelihood fit is performed in Section 8.3.10. It determines the conventional and prompt normalizations simultaneously. The unfolded flux is compared to CORSIKA simulations (Sibyll 2.3d weighted to the four cosmic ray primary models H3a, H4a, GST and GSF) as well as MCEq predictions (four primary models, Sibyll 2.3c, and

DPMJET-III 19.1). Using the unfolded muon flux only above 10 TeV yields conventional normalizations of approximately 0.95 for H3a and SIBYLL. Since the total number of events is dominated by lower energies due to the power-law spectrum, the ratio of the simulated and measured events should roughly correspond to the conventional normalization. Given these so-called scaling factors in Table F.18, for H3a and SIBYLL 2.3d a normalization of 1.09 is expected. To better constrain this normalization, the energy range is extended down to 600 GeV by incorporating a complementary analysis of stopping muons, low-energy muons that stop within the detector volume. This study, described in Section 8.3.10 and performed on the burnsample by Lucas Witthaus using a similar analysis chain, allows for a precise determination of the conventional normalization of 1.059 ± 0.006 for H3a. Constraining the conventional component using these low-energy measurements results in a prompt normalization of 1.424 ± 0.154 for H3a and SIBYLL 2.3c predicted by `MCEq`. All fitted normalizations are listed in Table 8.6 for the respective models.

Finally, the existence of a prompt component is rigorously tested. As shown in Figure 8.27, the measured flux is compared against 16 different `MCEq` predictions assuming only a conventional component. In none of these cases can the conventional flux alone sufficiently describe the data. The presence of a prompt component is observed with significances ranging from $4.4\,\sigma$ to $9.4\,\sigma$ for different MCEq predictions, and from $1.9\,\sigma$ to $6.8\,\sigma$ for CORSIKA predictions (see Table 8.6). While the CORSIKA and `MCEq` predictions are generally consistent within uncertainties, small deviations are observed (see Section F.6). Since the fit does not account for uncertainties in the model predictions, these differences between CORSIKA and `MCEq` translate into different quoted significances.

The measurement is also compared with previous IceCube results on the atmospheric muon flux, as shown in Figure 8.28, and the results are consistent.

In conclusion, this analysis utilizes 12.12 years of data, the largest amount of data ever used for an atmospheric muon flux measurement in IceCube. Together with the new simulations, reconstructions, and the unique selection, this enables the extension of earlier IceCube muon-flux measurements by approximately one order of magnitude in energy, providing the most precise determination of the atmospheric muon flux achieved so far. These results provide strong evidence for the prompt muon flux and offer critical constraints for hadronic interaction models in the high-energy forward region.

10 Outlook

The results presented in this work demonstrate the capability of IceCube to measure the prompt atmospheric muon flux and constrain hadronic interaction models. To further enhance the precision of this measurement and broaden its physics reach, several improvements can be pursued in future analyses.

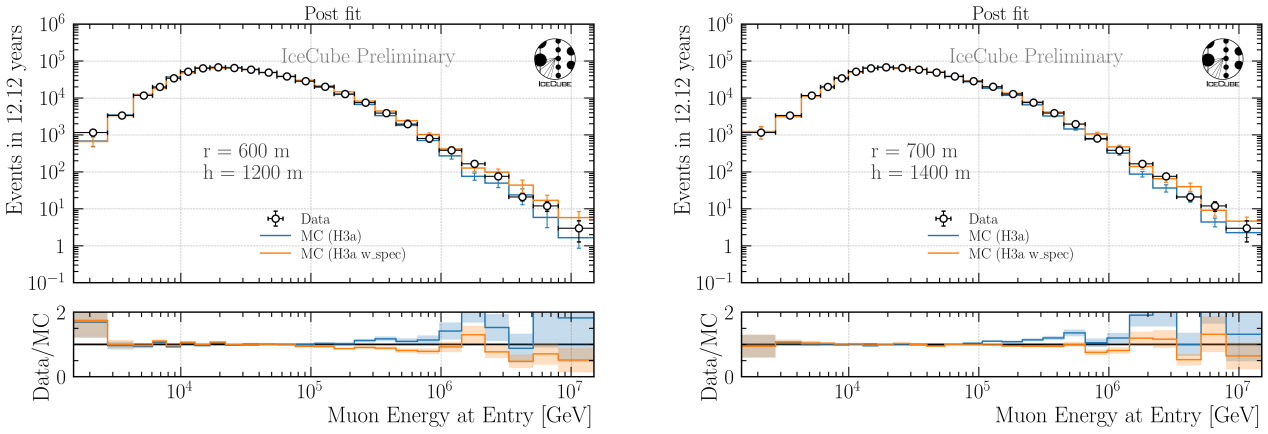
Increase MC statistics with biased simulations: Beginning with the existing simulation framework, increasing the MC statistics, particularly at the highest energies where the prompt component is most significant, helps to reduce statistical uncertainties. This is not only important to specify the detector response more accurately, but also to better parameterize the ice and detector systematics. In the current analysis, several MC datasets have been combined to achieve sufficient statistics. However, this approach introduces additional complexities because the different set haven been simulated with varying ice models, two different hadronic interactions models (SIBYLL 2.3c and 2.3d), different detector cylinder sizes, and not all sets include the information whether a muon is prompt or conventional. A unified simulation set with consistent settings would simplify the analysis and reduce potential systematic biases. To achieve this, a so-called biased simulation approach could be employed, where events are not only generated with a flatter energy spectrum, but with a probability distribution that favors high-energy muons. A similar approach has been used to select more high-energy events for the training of the neural networks used in this thesis, where the Eq. (6.12) has been applied. The idea is that a weight is assigned to each event during the simulation process, which is inversely proportional to the desired bias function. Events below a certain energy threshold are generated less frequently, while those above the threshold are generated more often. Using the weight, the final distribution can be corrected to reflect the true spectrum. This method ensures that the simulation focuses on the energy range of interest without the need for generating more low-energy events that are less relevant for the analysis or already well covered. This function will focus on the target of the unfolding, the leading muon energy at surface. Hence, it can be applied after the CORSIKA simulation in the first step of the simulation chain. Since the bottle-neck of the simulation is the detector simulation, caused by limited GPU resources, this approach could significantly increase the number of high-energy events.

Increase energy range towards lower energies and investigate potential impact of coincident events and the simulated detector volume: Investigating and resolving the potential normalization offset between the low-energy stopping muon analysis and the high-energy leading muon analysis, as shown in Figure 8.26, the leading muon sample could be extended to lower energies. In this work, a pre-cut on the reconstructed muon bundle energy at surface of 500 TeV has been applied to remove low-energy events and to shift the focus of the event sample to higher energies, where the prompt component is more pronounced. This is also necessary to reduce the computation time since 12.12 years of data have been processed, and of course the MC datasets needed to be processed as well. By lowering the bundle energy cut, more low-energy events would be included in the analysis, allowing to move the lower-energy boundary of the unfolding to energies below 10 TeV. On the one hand, this would increase the statistics in the 10 TeV region, on the other hand, it would create an overlap region with the stopping muon analysis. The cut can be changed easily in the processing chain, but the main challenge lies in ensuring that the MC dataset on the final selection level is in good agreement with the data. This requires a dedicated check of all selection variables and steps. Probably, some cuts need to be adjusted and re-optimized. Additionally, the impact of coincident primaries at these lower energies

needs to be studied carefully. With the current setup, the coincident primary rate is low above 10 TeV, but it increases significantly at lower energies, see Figure 8.15b.

Another source for the normalization offset could be the contamination of coincident primary events in the leading muon sample. As shown in Figure 8.18b, the coincident primary events should be located in the underflow bin of the unfolding, since the true leading muon energy is expected to be lower than the reconstructed energy. The unfolding can be performed again with a detector response that does not include coincident primaries to study the impact on the unfolded spectrum.

At last, two further data–MC comparisons are performed in Figure 10.1. The left panel presents the distribution for the MC set 23111, which has been simulated with the standard detector cylinder size of $r = 600\,\mathrm{m}$ and $h = 1200\,\mathrm{m}$, while the right panel shows the distribution for the *EHist* set, which has been simulated with a larger cylinder of $r = 700\,\mathrm{m}$ and $h = 1400\,\mathrm{m}$. The smaller cylinder is more computationally efficient to simulate and process. However, given this comparison, the larger cylinder seems to provide a better agreement between data and MC at low energies. This could be due to the fact that more low-energy events are included in the larger cylinder. Although low-energy muons emit less light, they trigger the DOMs even when passing far from the instrumented volume. The comparison indicates that for surface energies between 10 TeV and 20 TeV, the standard simulation radius excludes detectable events for this selection. The larger cylinder captures these distant tracks, improving the data agreement. At higher energies, the light yield is sufficient to ensure robust detection also for the smaller cylinder. The unfolding should be performed again based on the MC set with the larger cylinder to study the impact on the unfolded spectrum, especially at low energies. If the larger cylinder provides a better agreement between data and MC, it should be used for future analyses, even though it is more computationally expensive. Both sets in this comparison do not include coincident primaries.



(a) MC set 23111: $r = 600\,\mathrm{m}$, $h = 1200\,\mathrm{m}$.

(b) MC set *EHist*: $r = 700\,\mathrm{m}$, $h = 1400\,\mathrm{m}$.

Figure 10.1: This figure shows the data–MC comparison for the leading muon energy at entry distribution for two different MC sets. The left panel presents the distribution for the MC set 23111, which has been simulated with the standard detector cylinder size of $r = 600\,\mathrm{m}$ and $h = 1200\,\mathrm{m}$, while the right panel shows the distribution for the *EHist* set, which has been simulated with a larger cylinder of $r = 700\,\mathrm{m}$ and $h = 1400\,\mathrm{m}$. Both sets do not include coincident primaries. The larger cylinder seems to provide a better agreement between data and MC at low energies, which could be due to the fact that more low-energy events are included in the larger cylinder.

Future simulation framework CORSIKA 8 and improved hadronic interaction and cosmic ray models: The air-shower simulations performed and utilized within this analysis are based on CORSIKA version 7, which has been the standard tool in the field for many years. However, the source code of CORSIKA

7 is outdated and difficult to maintain and to extend. The new prompt and conventional tagging method developed in this thesis, which relies on the `EHISTORY` option of CORSIKA 7, has shown its limitations and assumption. In some cases, the parent particle information is incomplete or missing, leading to potential misclassifications of muons. In Ref. [104], a detailed discussion of these limitations is provided, demonstrating that their impact on this analysis is negligible. However, for future analyses aiming for higher precision, the currently developed simulation framework CORSIKA 8 [166–168] offers significant advantages. CORSIKA 8 is built on modern programming principles, making it more modular, extensible, and easier to maintain. One of its key features is the ability to track particle histories in a more detailed and flexible manner, allowing for an exact tagging of prompt muons based on their production history without relying on probabilistic assumptions. This will enable future analyses to classify muons with higher accuracy and confidence.

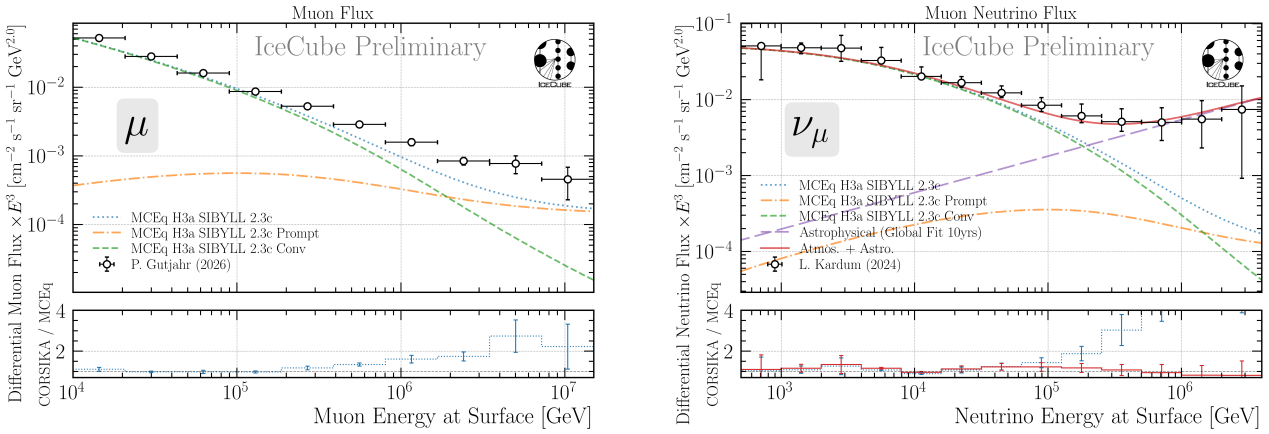
Moreover, different HIM can be fit against the unfolded muon flux. Recently, updated HIMs SIBYLL$^{\bigstar}$ [169], EPOS-LHC-R [170, 171], QGSJet III [172] and PYTHIA 8 [173, 174] have been released. These models incorporate the latest experimental data from collider experiments and cosmic ray observations. However, some of theses models are not yet implemented in CORSIKA or not available in `MCEq`. Once they are, the measured muon flux should be compared against these new predictions to further constrain hadronic interactions in the forward region.

In addition, a new primary model GSF2 was recently presented at the ICRC 2025 [175]. It includes the latest measurements of the cosmic ray spectrum and composition from various experiments. Once available, the measured muon flux should be fitted against predictions based on GSF2. Furthermore, the unfolding matrix itself should be re-weighted to this new primary model, and the unfolding should be repeated. In the current approach, the unfolding is performed four times, once for each primary model, and then the average is built. As presented in Figure 8.22, the choice of the primary model has an impact on the unfolded MC spectrum, even though, the impact on the final data unfolding is below 15 %, see Figure 8.25a. By using a more accurate primary model, and neglecting primary models that violate the latest measurements, the model-dependence of the unfolding can be reduced further.

Constrain the charm component further to improve astrophysical neutrino measurements: In astrophysical neutrino measurements, the atmospheric neutrino flux constitutes a background that must be accurately modeled and subtracted to isolate the astrophysical signal. Due to challenges in the measurement of the prompt atmospheric neutrino flux, current analyses often rely on theoretical predictions with large uncertainties. The latest measurement by IceCube shows large uncertainties, and the existence of the prompt neutrino flux was determined with a significance of $\leq 0.94\,\sigma$ [176]. To reduce these uncertainties, the constraints on the prompt muon flux obtained in this work can be translated into constraints on the prompt neutrino flux. As presented in Chapter 2, prompt muons are produced by hadrons heavier than pions and kaons, while prompt neutrinos originate from charmed hadron decays. Thus, the prompt muon flux measurement provides an upper limit on charm production in air showers.

Given the fit results presented in Table 8.6, the low-energy unfolding utilizing stopping muons provides constraints on the conventional component. In this thesis, only preliminary stopping muon results based on a small burnsample have been used. Note that this preliminary unfolding covers the full zenith range $0 < \cos\theta < 1$, while the leading muon analysis is restricted to $0.2 < \cos\theta < 1$. However, it has been discussed and tested that the stopping muon analysis can also be performed for the same restricted zenith range. While writing this thesis, the full stopping muon analysis is being finalized by Lucas Witthaus. Once available, these results can be used to fix the conventional normalization more precisely. With the conventional component fixed, it will be possible to fit the charm and non-charm components separately, not only a prompt and conventional normalization. This will provide direct constraints on the neutrino flux from charmed hadrons, which can be used in astrophysical neutrino analyses to reduce the uncertainties associated with the atmospheric background.

The connection between the atmospheric muon and neutrino flux measurements is illustrated in Figure 10.2. The atmospheric muon flux measurement from this work is shown in the left panel, while the muon neutrino flux measured by IceCube is presented in the right panel. Both fluxes are multiplied by E^3 to enhance the visibility of spectral features. By improving the understanding of the prompt atmospheric component through muon flux measurements, the precision of astrophysical neutrino analyses can be enhanced, and the sensitivity to detect and characterize astrophysical sources can be enhanced.



(a) Atmospheric muon flux measurement from this thesis.

(b) Muon neutrino flux measurement by IceCube [177].

Figure 10.2: This figure presents the atmospheric muon flux measurement from this thesis (left) and the muon neutrino flux measurement by IceCube (right). Both fluxes are scaled with E^3 to enhance the visibility of features in the spectrum. The muon flux measurement provides constraints on the prompt component, which is directly related to the charm production in air showers. These constraints can be translated into limits on the prompt neutrino flux, which is a crucial background for astrophysical neutrino measurements. By improving the understanding of the prompt atmospheric component, uncertainties in astrophysical neutrino analyses can be reduced, and the ability to detect and characterize astrophysical sources can be enhanced. Here, the astrophysical neutrino flux is presented as a single power-law distribution and taken from Ref. [86].

Forward folding approach to measure the prompt and conventional muon flux: An alternative approach to the unfolding method presented in this thesis is a forward folding technique. Instead of unfolding the measured spectrum to obtain the true flux, a parametric model of the muon flux can be defined, incorporating both conventional and prompt components with adjustable normalizations. This model can then be folded through the detector response, including all relevant systematic uncertainties, to generate predicted distributions of observables such as the reconstructed leading muon energy. This has been investigated in the masterthesis of Leander Flottau [178] under the supervision of the author of this thesis. The results are summarized in the following.

Forward folding is a maximum-likelihood estimation method where counting experiments in n energy bins are compared against simulations. Since the number of events d_j in each bin j follows a Poisson distribution[1], the probability is given by:

$$f(d_j, \mu_j) = \frac{\mu_j^{d_j} e^{-\mu_j}}{d_j!}\,, \tag{10.1}$$

[1]For the analysis, a likelihood incorporating limited MC statistics is used, but for simplicity, only the Poisson formulation is explained here [179].

where μ_j is the expected number of events in bin j. The overall likelihood $L(\vec{\mu})$ for the spectrum is constructed as the product of the likelihoods in the individual bins:

$$L(\vec{\mu}) = \prod_j f(d_j, \mu_j)\,. \tag{10.2}$$

The bin expectations μ_j are derived from weighted Monte Carlo events. Instead of generating new simulations for every hypothesis, events are re-weighted to match the underlying physical parameters $\vec{\Theta}$. The expectation in each bin is calculated as the sum of weights:

$$\mu_j(\vec{\Theta}) = \sum_{i \in \text{bin } j} w_i(\vec{\Theta})\,. \tag{10.3}$$

This allows for the minimization of the likelihood to find the parameter set $\vec{\Theta}$ that best describes the observed data.

To test the sensitivity of this method, Asimov data sets, representing the exact expectation values, and pseudoexperiments (Poisson-sampled event sets) can be used. The significance of a prompt signal can then be estimated using a likelihood ratio test, comparing a free fit (where the prompt normalization is a free parameter, Θ_{free}) against a null hypothesis (where the prompt flux is fixed to zero, Θ_{null}). The test statistic Λ is defined as:

$$\Lambda = -2 \ln \frac{L(\Theta_{\text{null}})}{L(\Theta_{\text{free}})} = -2\left(l(\Theta_{\text{null}}) - l(\Theta_{\text{free}})\right)\,, \tag{10.4}$$

where l denotes the log-likelihood. According to Wilks' theorem, this test statistic follows a χ^2 distribution, allowing for a statistical quantification of the prompt signal carried out on the reconstructed level directly.

Beyond detector-related parameters, additional systematic uncertainties arise from the primary cosmic-ray spectrum, which is not precisely known in either shape or composition. To address this, two parameters are introduced. First, a cosmic-ray model interpolation allows for a linear transition between two flux models (e.g., H3a and GSF):

$$W_{\text{CR}} = W_{\text{H3a}} + \alpha_{\text{CR}}\left(W_{\text{GSF}} - W_{\text{H3a}}\right), \tag{10.5}$$

where $\alpha_{\text{CR}} = 0$ corresponds to H3a and $\alpha_{\text{CR}} = 1$ to GSF. This allows the fit to continuously explore differences between the two models. Due to this interpolation, only one normalization is fitted, in contrast to the unfolding approach where a separate normalization is determined for each assumed primary model. Second, a spectral index shift modifies the spectral slope of the primary flux:

$$W_{\Delta\gamma_{\text{CR}}} = \left(\frac{E_{\text{primary}}}{E_{\text{ref}}}\right)^{-\Delta\gamma_{\text{CR}}}\,, \tag{10.6}$$

where E_{primary} is the energy of the primary cosmic ray and E_{ref} is a reference energy (e.g., the median primary energy). A positive $\Delta\gamma_{\text{CR}}$ corresponds to a steeper spectrum, while a negative value corresponds to a flatter spectrum. Together, these parameters capture uncertainties in both the model choice and the spectral slope. Finally, in addition to the cosmic-ray systematics, the fit also incorporates detector and ice properties related systematic uncertainties, which are modeled using the SnowStorm framework.

Predictions for 12.12 years of data to measure the prompt component are presented in Figure 10.3a. Since the fit is performed on the simulation, the normalizations are expected to be unity by definition. The measurement yields $n_{\text{prompt}} = 0.93^{+0.23}_{-0.20}$ over the null hypothesis $n_{\text{prompt}} = 0$, exceeding 5σ at a p-value of 3.137×10^{-13}.

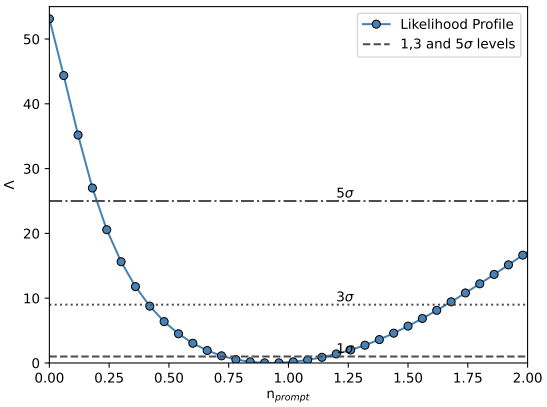
Furthermore, a likelihood scan for 12 years is performed to measure charm and unflavored components separately, as shown in Figure 10.3b. This involves one-dimensional profile likelihood scans of n_{charmed} and $n_{\text{unflavored}}$, where each acts as a nuisance parameter for the other. The best-fit values are found to be $n_{\text{charmed}} = 1.03^{+0.41}_{-0.33}$ and $n_{\text{unflavored}} = 0.84^{+0.35}_{-0.30}$. The results show that n_{charmed} has a flatter profile with a p-value of 5.96×10^{-4} compared to the null hypothesis. The unflavored component is more similar in shape to the prompt component, with a steeper profile and higher significance. Its p-value is 2.03×10^{-4}.

Finally, the method is tested on the burnsample, which is the same dataset used for the unfolding approach in this thesis. The resulting likelihood profile for the prompt component is shown in Figure 10.3c. The scan is performed on the final data level with 10 energy bins and two zenith bins. The best-fit value is $n_{\text{prompt}} = 0.87^{+0.44}_{-0.41}$, with a significance below $2\,\sigma$.

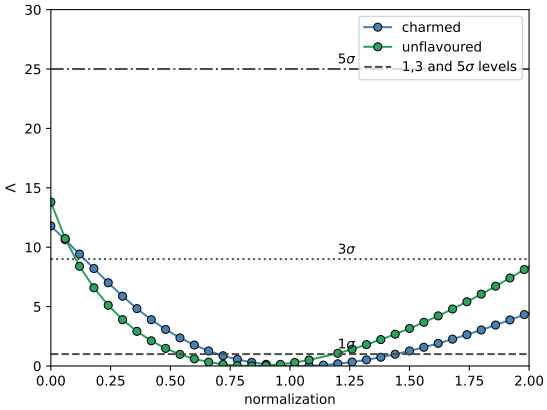
Additionally, the scan for charmed and unflavored components was performed, as shown in Figure 10.3d. These fits are less significant than the total prompt fit, and neither component reaches even $1\,\sigma$ significance. The best-fit value for the charmed component is $n_{\text{charmed}} = 1.07$ with a p-value of 0.49. The best-fit value for the unflavored component is $n_{\text{unflavored}} = 0.75$, with a slightly lower p-value at 0.45. This is despite the lower fitted value, because it has a steeper profile due to being more dominant at high energies.

Due to the limited statistics of the burnsample, the results are not significant. This is expected and even intended by the design of a blind analysis. However, the resulting normalizations are in a reasonable range, providing confidence in the method.

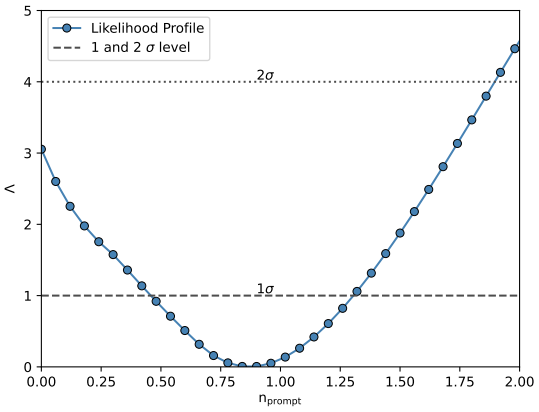
In summary, this work has proven the capability of the forward folding approach to achieve two key objectives. On the one hand, it provides an alternative method to measure the normalization of the prompt component and determine the significance of its existence. On the other hand, it enables the separate measurement of charm and unflavored components. These capabilities will help to further improve hadronic interaction models and constrain the neutrino production in air showers, ultimately improving the measurement of astrophysical neutrino sources. While this work is not final yet, the results indicate that it is a promising approach worth continuing.



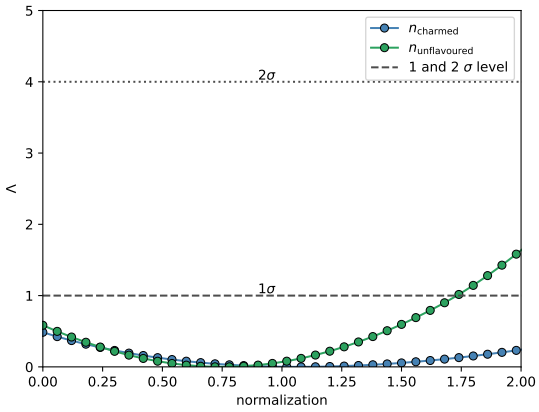
(a) Prediction for 12.12 years of data to detect prompt.



(b) Prediction for 12.12 years of data to detect charm and unflavoured.



(c) Burnsample result to detect prompt.



(d) Burnsample result to detect charm and unflavoured.

Figure 10.3: This figure shows the likelihood scans for the forward folding approach to measure the prompt muon flux. The top row shows the expected sensitivity for 12.12 years of data, while the bottom row shows the results obtained from the burnsample. The left column focuses on detecting the prompt component, while the right column aims to separate charm and unflavoured contributions. Results are based on a masterthesis by Leander Flottau under the authors supervision [178].

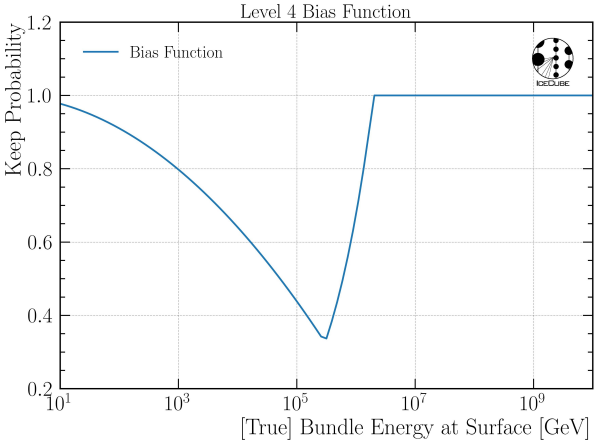
A Reconstruction

This chapter presents additional details on the reconstruction methods. First, the training data for the DNNs is shown in Section A.1, including the effect of the bias function and the true variable distributions. Second, the direction of the primary cosmic ray is compared to the direction of the leading muon in Section A.2. Finally, the performance of the reconstruction networks is tested on simulations using the newest ice model, Spice FTP-v3m, in Section A.3.

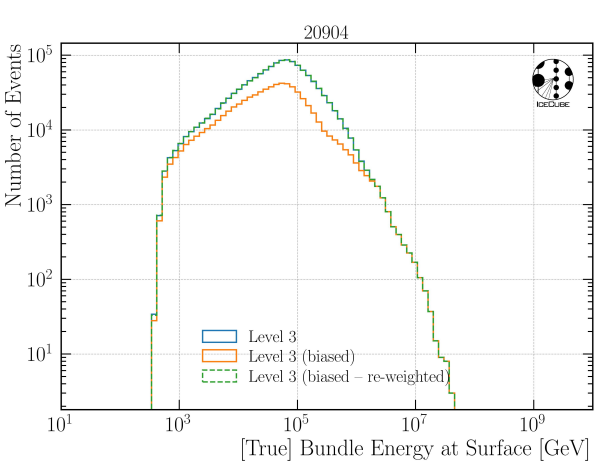
A.1 Training Data

Figure A.1 shows the bias function from Eq. (6.12) and its impact on the training data for two variables on *Level 3*. The effect is shown for dataset 20904, but it is also applied to dataset 22020 with similar results. The biasing reduces the statistics by roughly a factor of two while preserving the high-energy tail. A re-weighting allows for the recovery of the original distribution. Although this re-weighting is not used here, it is presented to demonstrate that recovering the original distribution is possible if needed.

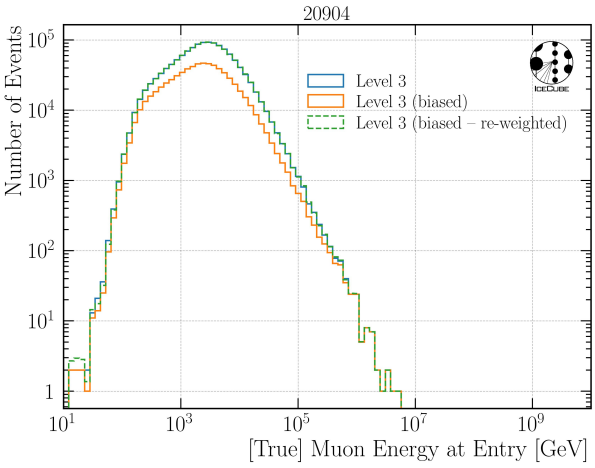
Figure A.2 displays the true distributions of the training data used for the various reconstructions on *Level 4*. The bias function described in Eq. (6.12) and Figure A.1 is applied to datasets 20904 and 22020. For the leading muon energy at the surface, which is the target variable for the unfolding in this thesis, combining the four datasets provides sufficient coverage of the high-energy region around 10 PeV.



(a) Keep probability.

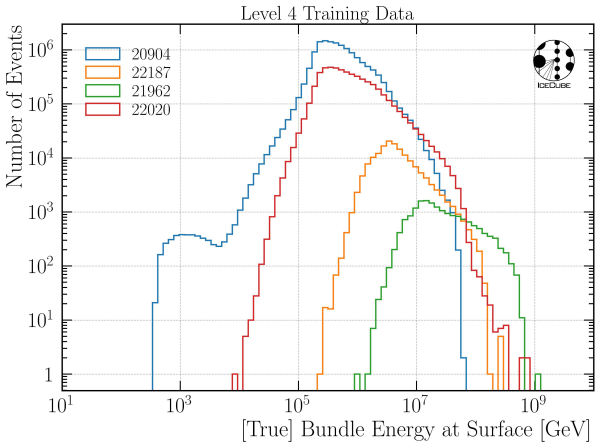


(b) Bundle energy at surface.

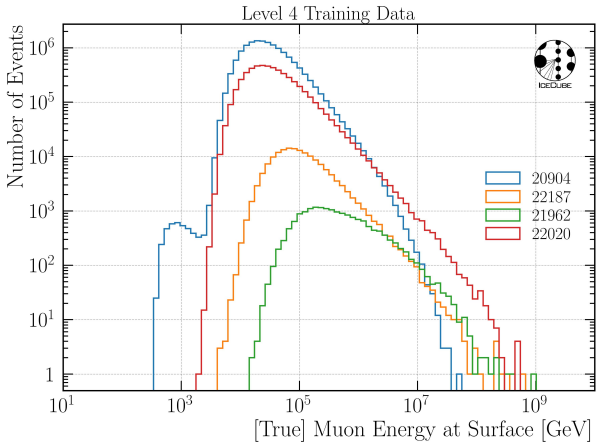


(c) Leading muon energy at entry.

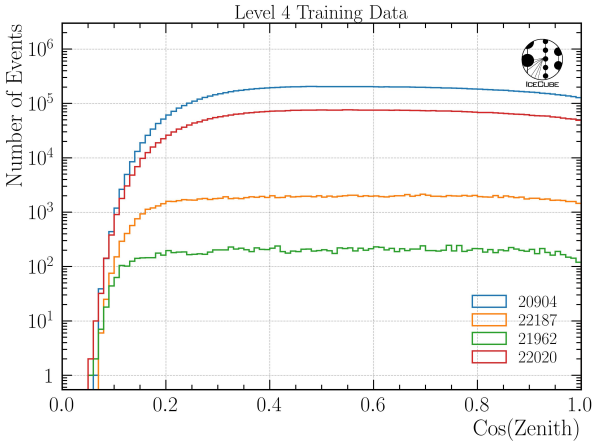
Figure A.1: These figures show the bias function from Eq. (6.12) and its impact on the training data for two variables on *Level 3*. The effect is shown for dataset 20904, but it is also applied similarly to dataset 22020. The bias reduces the statistics by roughly a factor of two while preserving the high-energy tail.



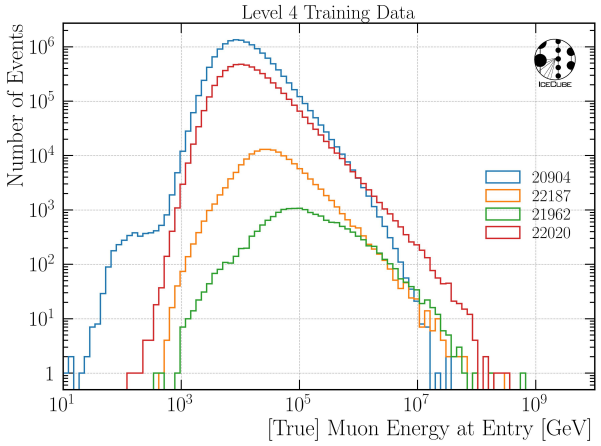
(a) Bundle energy at surface.



(b) Leading muon energy at surface.



(c) $\cos(\theta)$.

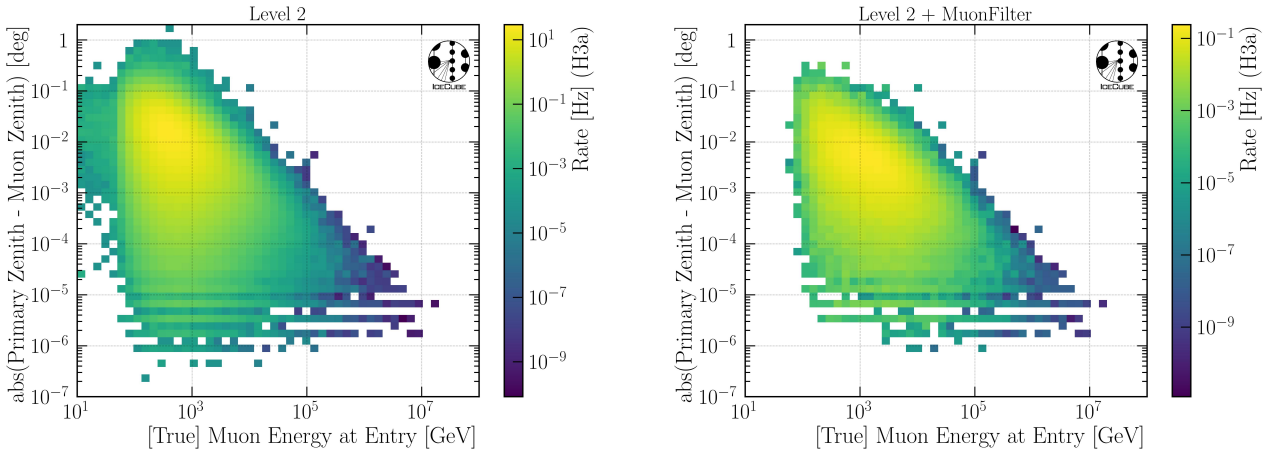


(d) Leading muon energy at entry.

Figure A.2: True distributions of the training data used for the reconstructions on *Level 4*. To ensure network robustness against outliers below the 500 TeV pre-cut, a lower cut of 200 TeV is applied to the training data. The biasing from Eq. (6.12) is applied to datasets 20904 and 22020.

A.2 Direction

In this section, the direction of the primary cosmic ray is compared to the direction of the leading muon. Since the muon is produced in the air shower and carries a significant fraction of the primary's momentum, its direction is expected to be nearly collinear with that of the primary particle. Figure A.3 checks this assumption by showing the absolute difference in zenith angle between the primary and the leading muon.



(a) *Level 2*.

(b) *Level 2* and *MuonFilter*.

Figure A.3: Absolute difference between the true primary zenith angle and the true muon zenith angle as a function of the true leading muon energy at detector entry, weighted with the H3a model. The left panel shows the distribution for *Level 2*, while the right panel shows the result after applying the *MuonFilter*. The angular difference is generally below 1°, which is smaller than the angular resolution of IceCube [150] for this energy range.

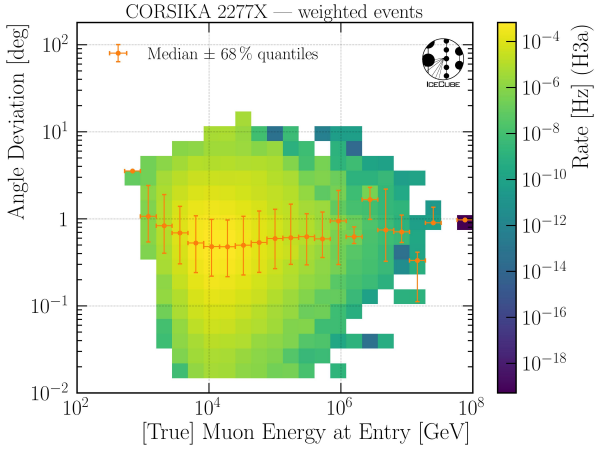
A.3 Training on Newest Ice Model

The networks used in this thesis have been trained on simulations with the ice models Spice3.2 and Spice3.2.1, as stated in Table 6.1. However, within the scope of this thesis, simulations have been produced with a new ice model named Spice FTP-v3m [148, 149]. Since the statistics of these new simulations are limited, it is not possible to split them into training, validation, and test sets. Later on, a similar dataset with the new ice model has been created, 23111. The same networks have been trained on this dataset, and the performance has been evaluated on the new datasets 22774–22778. This is presented in Figure A.4.

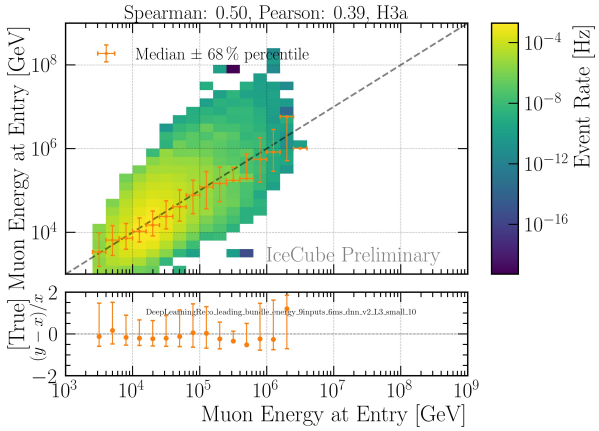
The top left panel shows the angular resolution, where even the 68 % intervals are partly below 1° for leading muon energies between 10 TeV and 1 PeV. In the bottom left panel, the data–MC agreement for the entry z-position is shown. This variable exhibited the largest mismatch between data and MC in this thesis, as presented in Figure 7.6a. Assuming the GSF model, a good agreement is visible. This indicates that the new ice model indeed improves the description of the ice compared to Spice3.2 and Spice3.2.1, and that it has a depth-dependent improvement.

The leading muon energy reconstruction at entry is shown in the top right panel. The reconstruction performs well; however, the network does not predict energies above approximately 3 PeV, even though the true energies go up to 100 PeV. This is caused by the limited statistics of high-energy events in dataset 23111, which is presented in the bottom right panel. Above 10 PeV, only very few events are

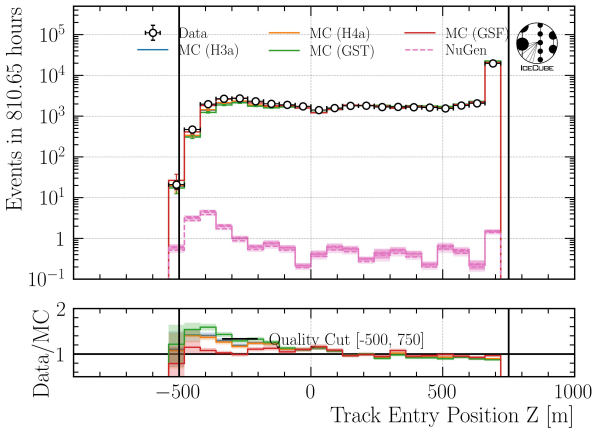
available for training, which makes it nearly impossible for the network to learn to predict these high energies. It needs to be mentioned that these distributions are shown for the *Final Level*, instead of the actual *Level 4* where the training is performed.



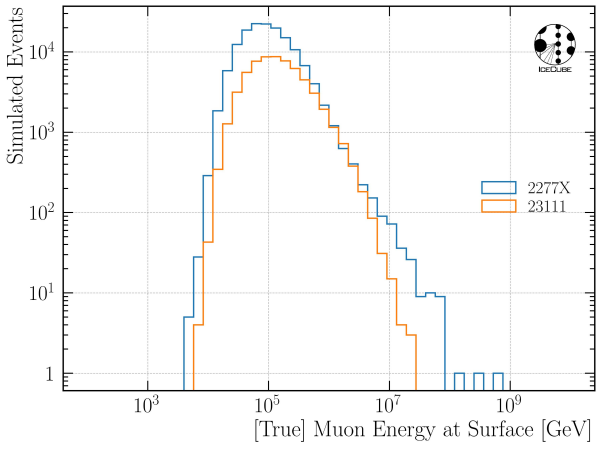
(a) Angular resolution on *Level 5*.



(b) Muon energy at entry on *Level 5*.



(c) Entry z-position on *Level 4*.



(d) MC statistics on *Final Level*.

Figure A.4: These figures present the performance of networks trained on dataset 23111 simulated with the newest ice model Spice FTP-v3m. All in all, the performance is improved in comparison to the models used in this thesis. However, due to the limited MC statistics at high energies, the energy reconstruction is limited to approximately 3 PeV, and this variable is the final variable for the unfolding in this thesis.

B Selection

This chapter provides additional information on the event selection. First, the filter logic of the *Level 3* filter is described in Section B.1. Second, the impact of the *Level 4* pre-cut is shown in Section B.2. Then, the data–MC agreement is presented for *Level 5* (Section B.3) and the *Final Level* (Section B.4). Finally, the impact of different leadingness cuts on the data–MC agreement is shown in Section B.5.

B.1 Level 3 – Filter

OnlineL2Filter

The `OnlineL2Filter` applies a sequence of cuts on reconstructed track quality and event topology, divided into three zenith-angle zones ($i = 1, 2, 3$). The decision is based on the reconstructed zenith angle θ, the number of hit DOMs N_{ch}, the logarithm of the total charge Q_{tot}, the fit log-likelihood ℓ, the number of direct hits N_{dir}, and the total direct-track length L_{dir}. A normalized log-likelihood is defined for each zone as

$$\mathrm{PLogL}_i = \frac{\ell}{N_{\mathrm{ch}} - p_i}, \tag{B.1}$$

where p_i is a zone-dependent offset parameter. The zones are defined as:

$$\text{Zone 1:} \quad -1 \leq \cos\theta \leq \cos(82^\circ), \tag{B.2}$$
$$\text{Zone 2:} \quad \cos(82^\circ) < \cos\theta \leq \cos(66^\circ), \tag{B.3}$$
$$\text{Zone 3:} \quad \cos(66^\circ) < \cos\theta \leq 1. \tag{B.4}$$

Zone 1 Cuts: Using the direct-hit ellipsis $E = (L_{\mathrm{dir}}/160)^2 + (N_{\mathrm{dir}}/9)^2$ and $p_1 = 4.5$, an event is kept if

$$\log_{10} Q_{\mathrm{tot}} \geq 2.7 \quad \vee \quad E \geq 1 \quad \vee \quad \mathrm{PLogL}_1 \leq 8.3. \tag{B.5}$$

Zone 2 Cuts: With a linear charge cut $[\log_{10} Q_{\mathrm{tot}}]_{\mathrm{cut}}(c) = 3.3c + 1.13$ (where $c = \cos\theta$) and $p_2 = 4.5$, an event is kept if

$$\log_{10} Q_{\mathrm{tot}} \geq 2.5 \quad \vee \quad (\mathrm{PLogL}_2 \leq 8.3 \wedge \log_{10} Q_{\mathrm{tot}} \geq [\log_{10} Q]_{\mathrm{cut}}(c)). \tag{B.6}$$

Zone 3 Cuts: For the third zone ($p_3 = 4.5$), a kinked linear charge cut is introduced at $c_{\mathrm{kink}} = 0.5$:

$$[\log_{10} Q]_{\mathrm{cut}}(c) = \begin{cases} 3.3\,c + 1.13 + \Delta b, & c \leq 0.5, \\ 0.6\,c + 2.53, & c > 0.5, \end{cases} \tag{B.7}$$

with $\Delta b \approx -0.065$. An event is kept if

$$\log_{10} Q_{\mathrm{tot}} \geq 3.0 \quad \vee \quad (\mathrm{PLogL}_3 \leq 10.5 \wedge \log_{10} Q_{\mathrm{tot}} \geq [\log_{10} Q]_{\mathrm{cut}}(c)). \tag{B.8}$$

HighQFilter

The `HighQFilter` selects events based on a simple total-charge threshold. Using the homogenized total charge Q_{tot}, an event is kept if

$$Q_{\text{tot}} \;\geq\; 1000\,. \tag{B.9}$$

EHEAlertFilter

The `EHEAlertFilter` targets Extremely High Energy (EHE) events using a two-stage selection process based on charge (NPE), multiplicity (N_{ch}), fit quality (χ^2_r), and a 2D pulse-height vs. zenith requirement. First, a set of basic quality cuts (*Level 3*) must be satisfied simultaneously:

$$\log_{10}\text{NPE} > 3.6\,, \quad N_{\text{ch}} > 300\,, \quad 25 < \chi^2_r < 80\,. \tag{B.10}$$

Subsequently, a *Level 4* requirement is applied, which by default uses the standard 2D cut Pass2D($\log_{10}$ NPE, $\cos\theta$):

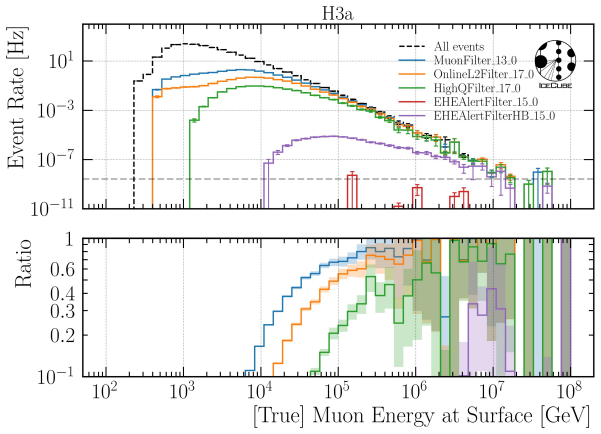
$$\text{Pass2D}: \quad \begin{cases} \log_{10}\text{NPE} \geq 3.6, & \cos\theta < 0.1, \\ \log_{10}\text{NPE} \geq 3.6 + 2.99\sqrt{1-\left(\frac{\cos\theta-0.93}{0.83}\right)^2}\,, & \cos\theta \geq 0.1. \end{cases} \tag{B.11}$$

Alternative configurations allow for looser cuts (`Looser`, `Loosest`) or skipping the 2D test entirely.

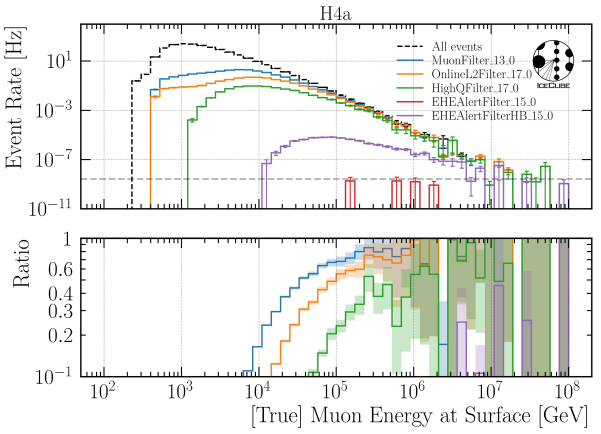
EHEAlertFilterHB

The `EHEAlertFilterHB` is a "heartbeat" version of the EHE filter. It differs from the standard filter by using modified Ophelia fit parameters ("ImpLF") and enforcing a looser 2D cut condition. Events are accepted if they pass the `PassLooser2D` condition but fail the standard `Pass2D` condition, effectively collecting events in a specific phase space region:

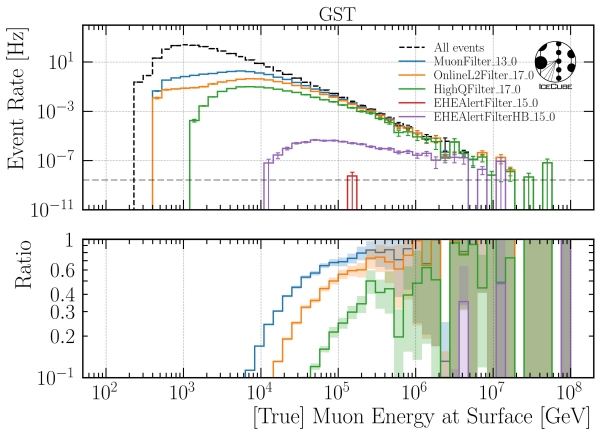
$$\log_{10}\text{NPE} \;\geq\; 3.6 + 1.0\sqrt{1-\left(\frac{\cos\theta-0.85}{0.75}\right)^2} \quad (\cos\theta \geq 0.1)\,. \tag{B.12}$$



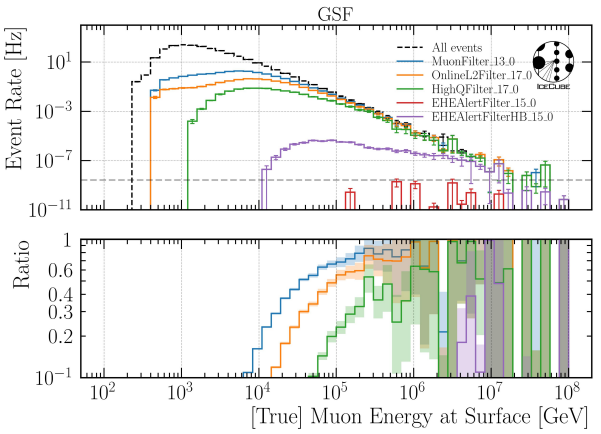
(a) H3a.



(b) H4a.

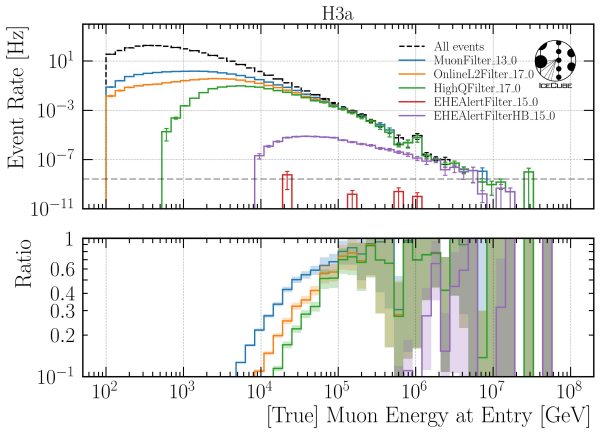


(c) GST.

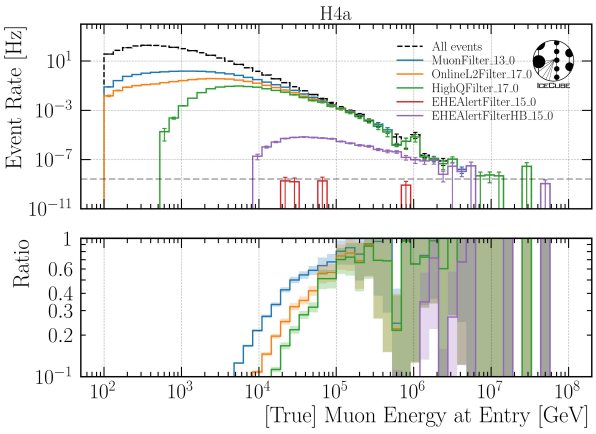


(d) GSF

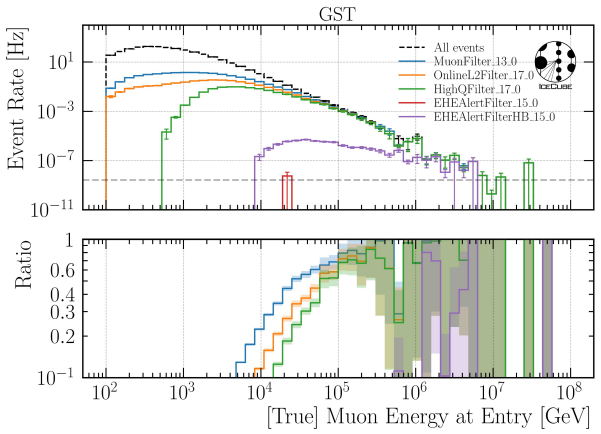
Figure B.1: Impact of various IceCube filters for the four primary weightings. The distributions show the leading muon energy at the surface. The ratio represents the selection efficiency, defined as the event rate of selected events divided by the total event rate. Only a subset of the full MC statistics is used. For this selection, the *MuonFilter* is chosen.



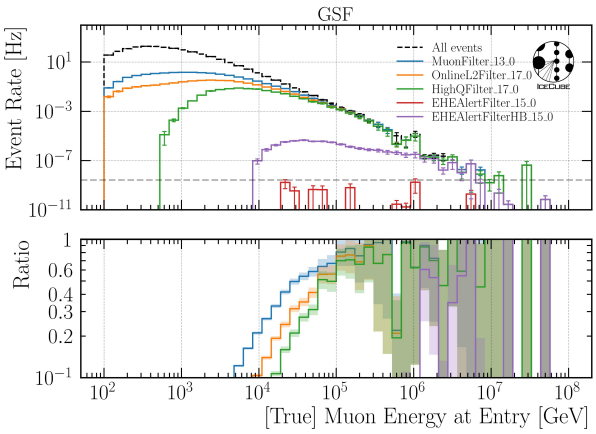
(a) H3a.



(b) H4a.

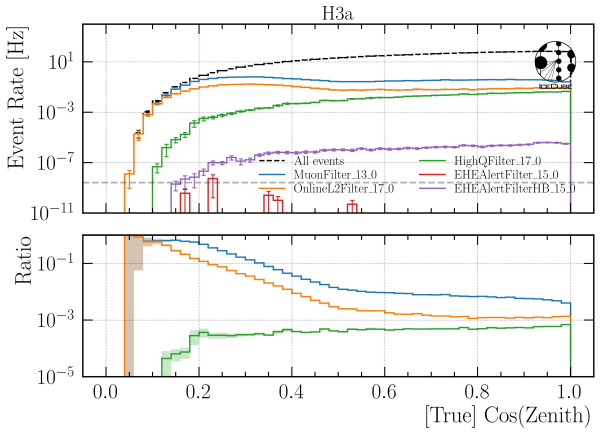


(c) GST.

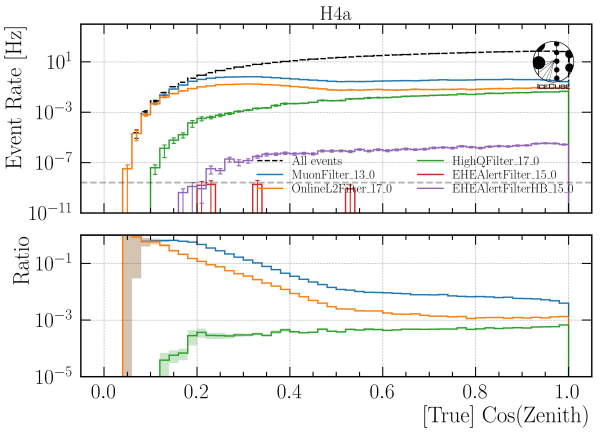


(d) GSF.

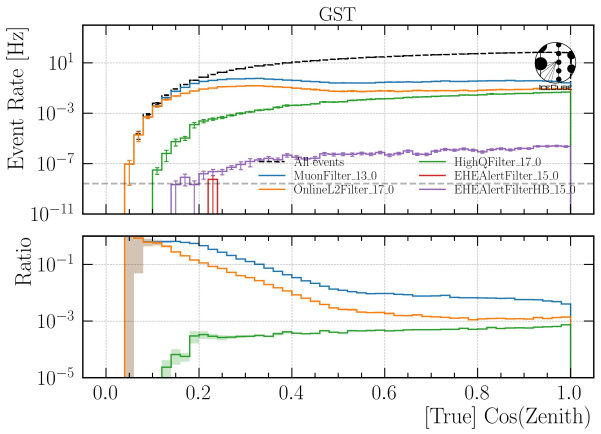
Figure B.2: Impact of various IceCube filters for the four primary weightings. The distributions show the leading muon energy at detector entry. The ratio represents the selection efficiency, defined as the event rate of selected events divided by the total event rate. Only a subset of the full MC statistics is used. For this selection, the *MuonFilter* is chosen.



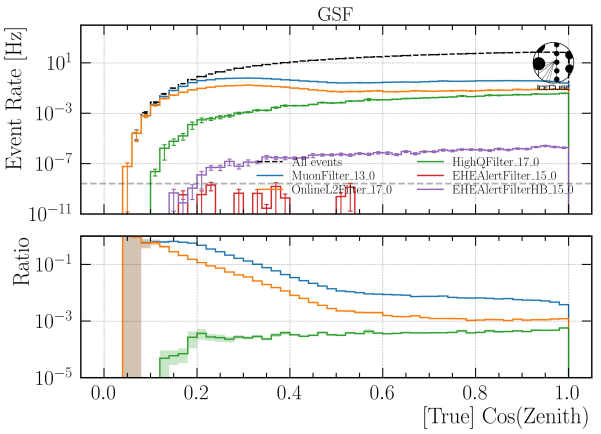
(a) H3a.



(b) H4a.



(c) GST.

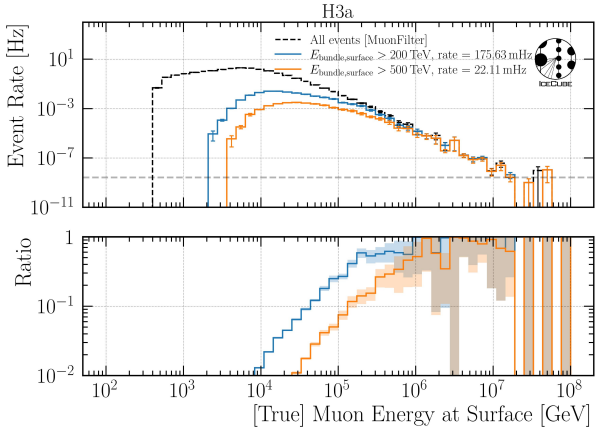


(d) GSF.

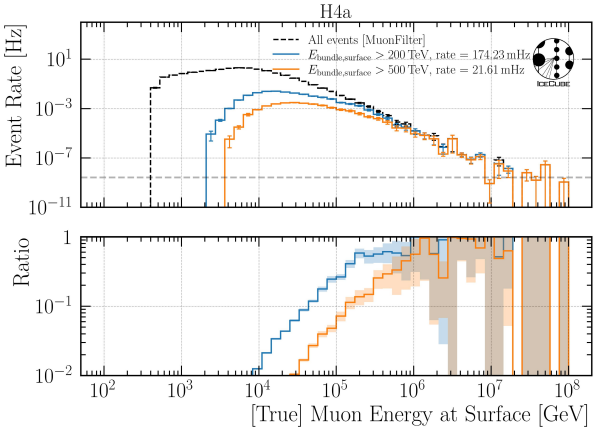
Figure B.3: Impact of various IceCube filters for the four primary weightings. The distributions show the cosine of the zenith angle of the leading muon at detector entry. The ratio represents the selection efficiency, defined as the event rate of selected events divided by the total event rate. Only a subset of the full MC statistics is used. For this selection, the *MuonFilter* is chosen.

B.2 Level 4 – Pre Cut

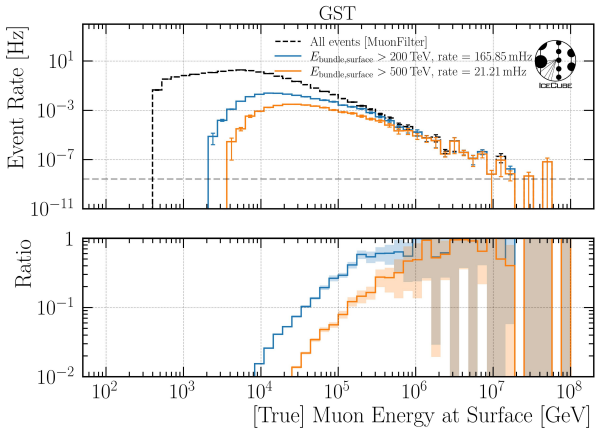
In this section, the impact of the Level 4 pre-cut is shown. This cut is based on a reconstructed muon bundle energy at surface. Besides the chosen cut value of 500 TeV, a cut of 200 TeV is presented. Both cuts show a similar behavior in the high-energy region. However, the 200 TeV cut retains about 8 times more events, which would increase the processing time by a factor of 8. Since the high-energy region is the focus of this analysis, the 500 TeV cut is chosen. The impact on the leading muon energy at surface, entry, and the primary energy is presented for all CR flux models.



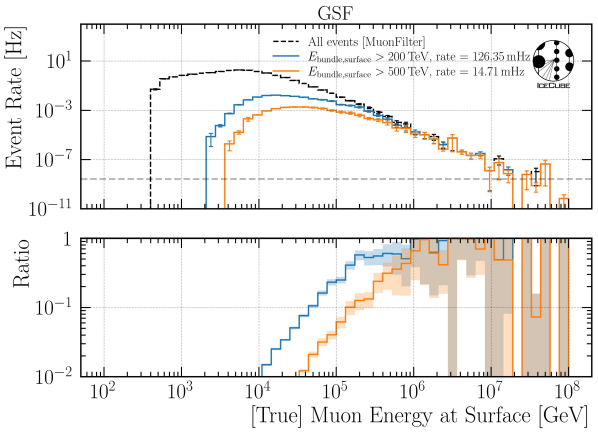
(a) H3a.



(b) H4a.

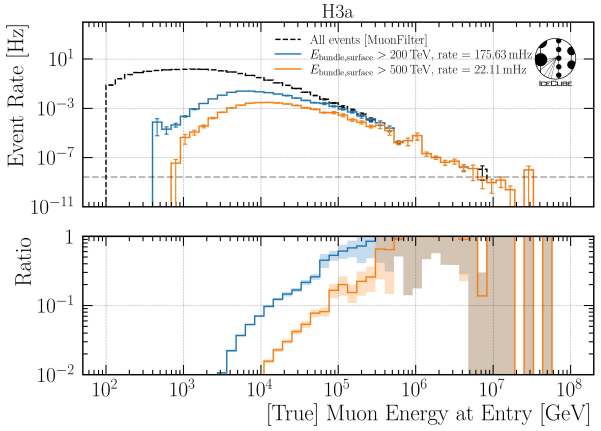


(c) GST.

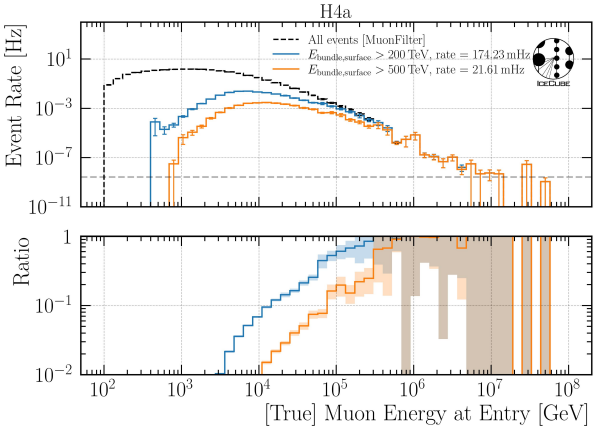


(d) GSF.

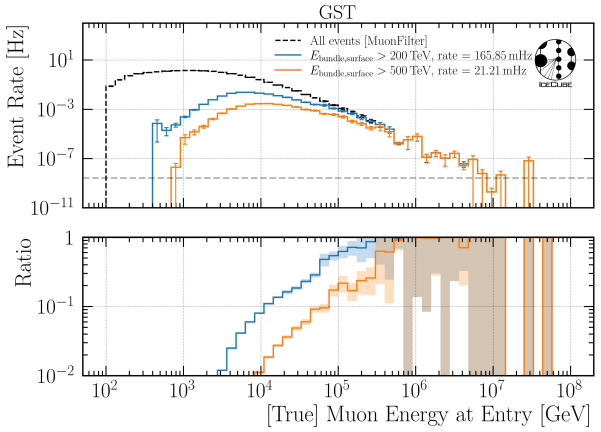
Figure B.4: Impact of the bundle energy cuts on the leading muon energy at surface. The ratio represents the selection efficiency. Results for cuts of 200 TeV and 500 TeV are shown. Only a subset of the entire MC statistics is used. For this selection, the 500 TeV cut is chosen.



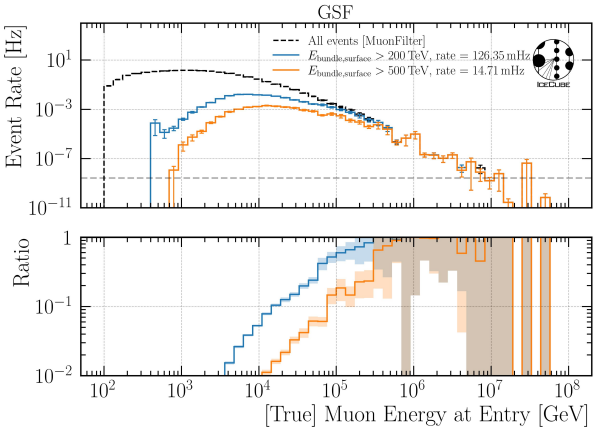
(a) H3a.



(b) H4a.

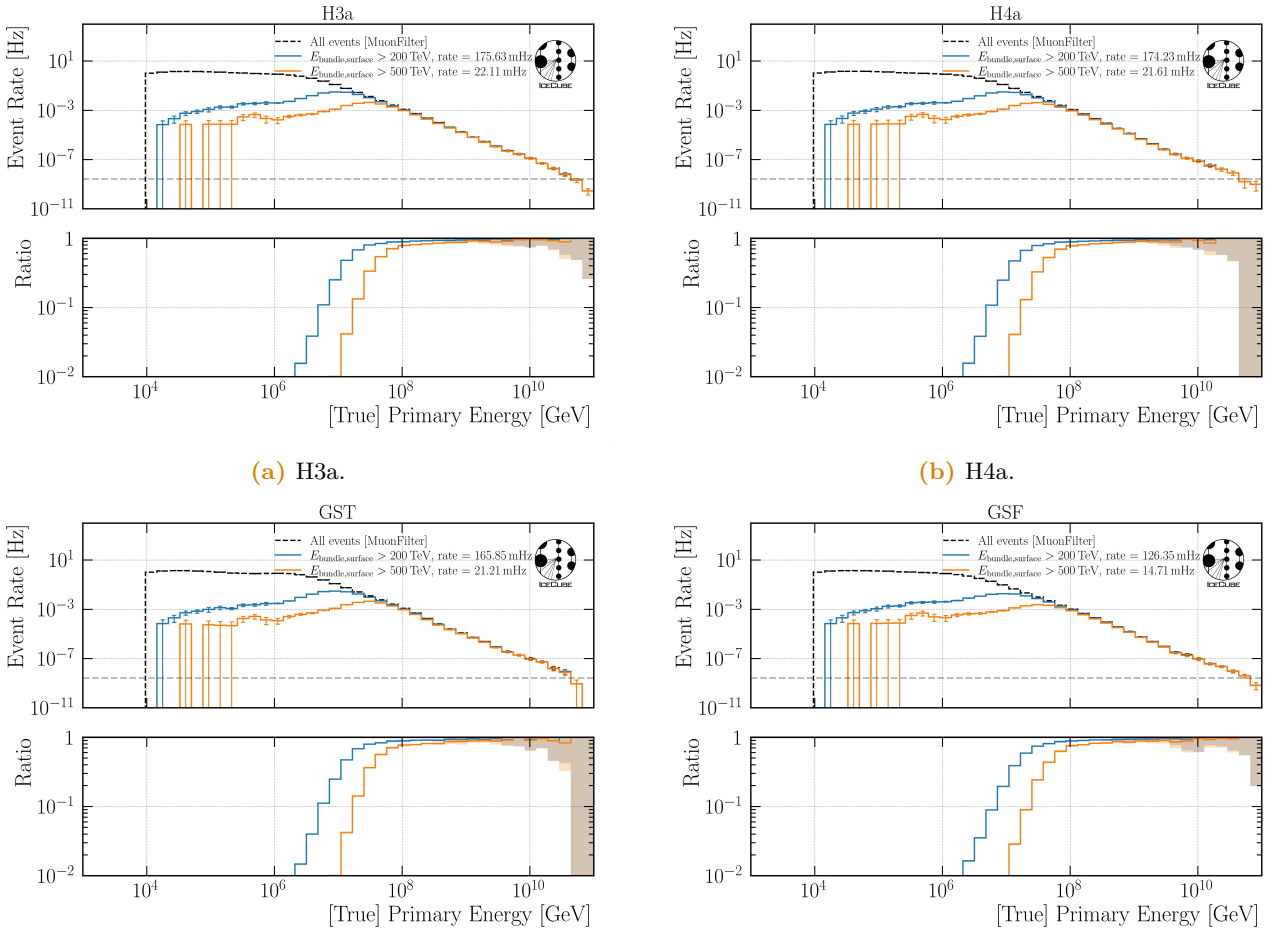


(c) GST.



(d) GSF.

Figure B.5: Impact of the bundle energy cuts on the leading muon energy at entry. The ratio represents the selection efficiency. Results for cuts of 200 TeV and 500 TeV are shown. Only a subset of the entire MC statistics is used. For this selection, the 500 TeV cut is chosen.



(a) H3a.

(b) H4a.

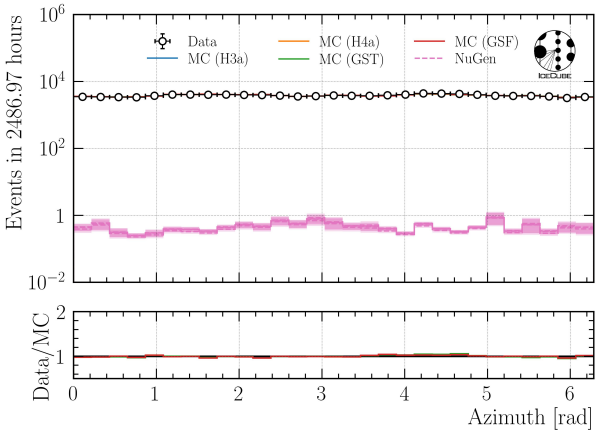
(c) GST.

(d) GSF.

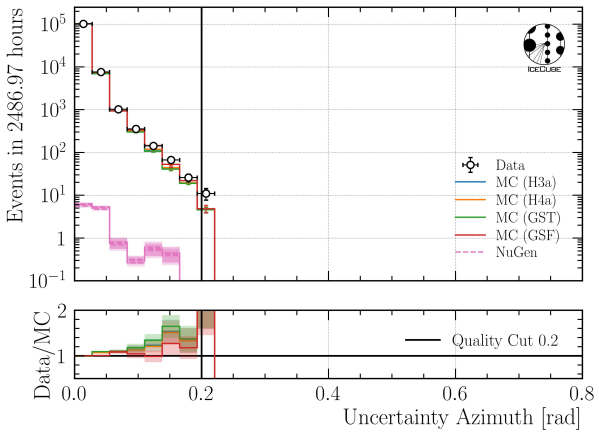
Figure B.6: Impact of the bundle energy cuts on the primary energy. The ratio represents the selection efficiency. Results for cuts of 200 TeV and 500 TeV are shown. Only a subset of the entire MC statistics is used. For this selection, the 500 TeV cut is chosen.

B.3 Data–MC: Level 5

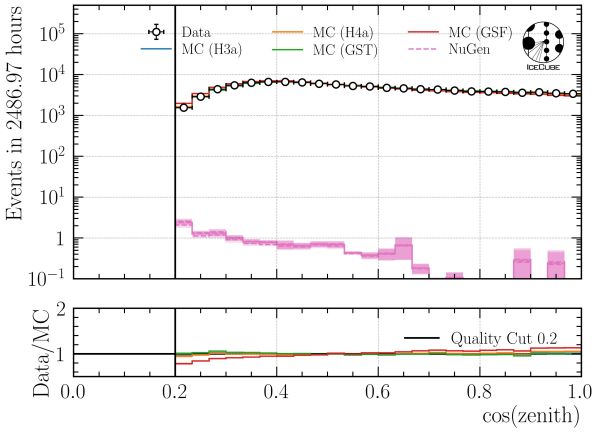
In this section, the agreement between experimental data and MC simulation is presented for Level 5. The distributions of the reconstructed direction, track geometry, and their corresponding uncertainties are shown. The data corresponds to the *burnsample*, which accounts for approximately 2 % of the total data set. The MC simulation includes both the atmospheric muon flux and the neutrino background. The vertical black lines indicate the cuts applied in this level.



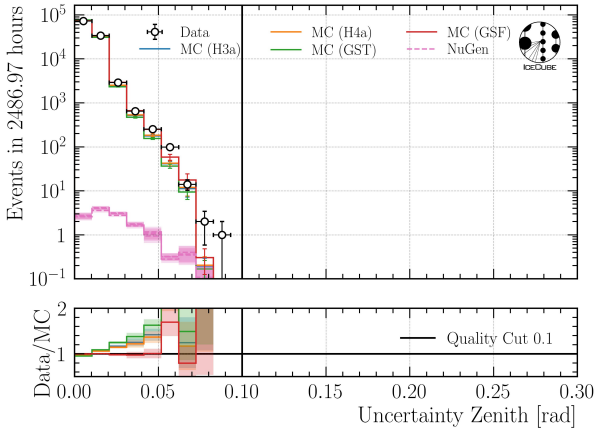
(a) Azimuth.



(b) Azimuth uncertainty.

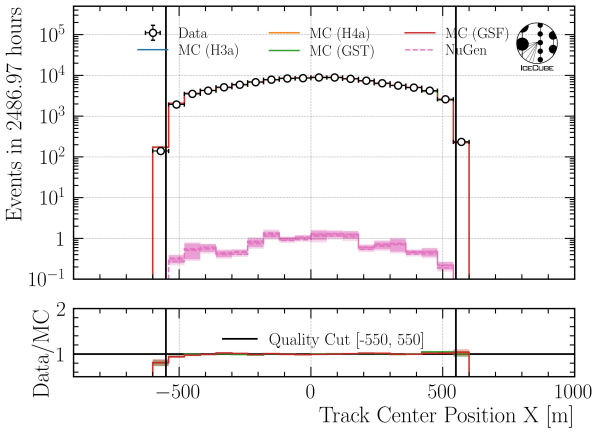


(c) $\cos(\theta)$.

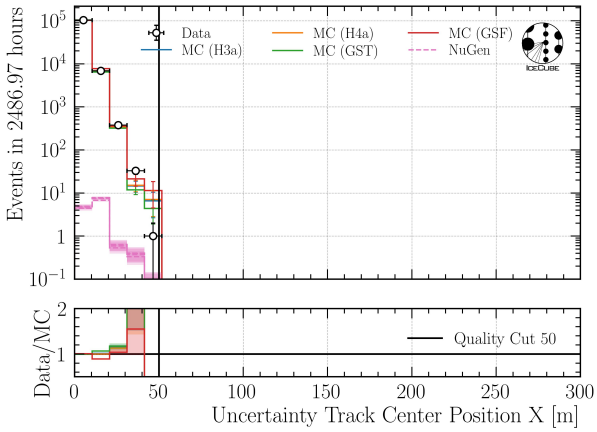


(d) $\cos(\theta)$ uncertainty.

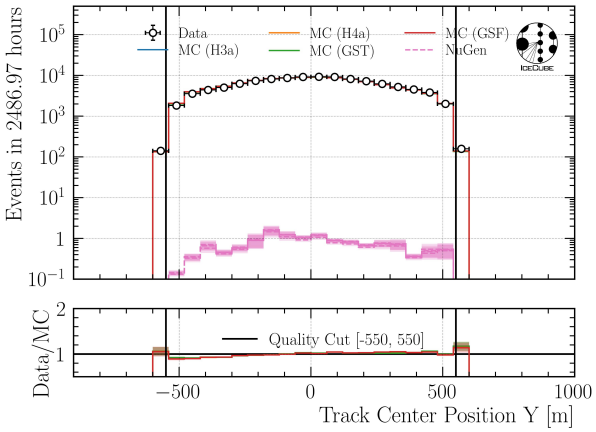
Figure B.7: Data–MC agreement for the directional reconstructions at Level 5. The figures include four MC predictions, one for each primary weighting. The ratio shows the number of data events (2486.97 h) divided by the expected MC events. Vertical black lines indicate the applied quality cuts. The MC predictions include the atmospheric muon flux and the neutrino background. The neutrino background is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86].



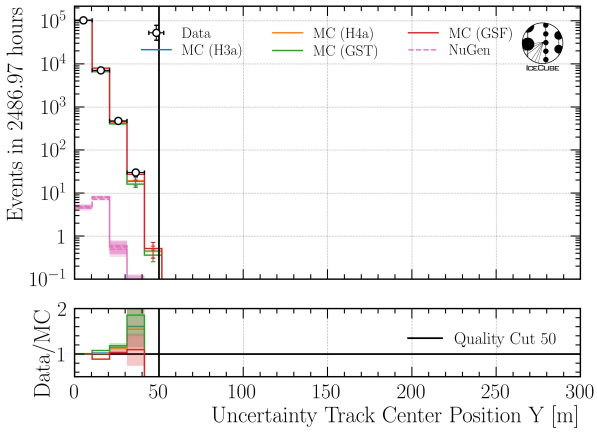
(a) Center position x.



(b) Center position x uncertainty.

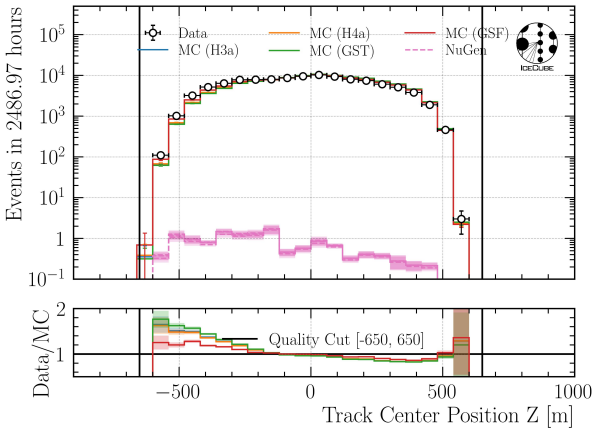


(c) Center position y.

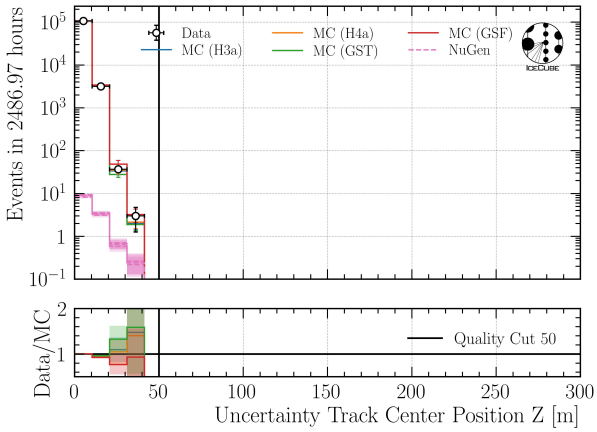


(d) Center position y uncertainty.

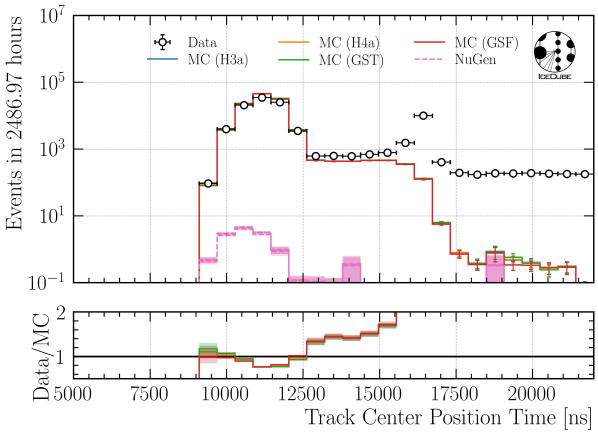
Figure B.8: Data–MC agreement for the center position x and y at Level 5. The figures include four MC predictions, one for each primary weighting. The ratio shows the number of data events (2486.97 h) divided by the expected MC events. Vertical black lines indicate the applied quality cuts. The MC predictions include the atmospheric muon flux and the neutrino background. The neutrino background is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86].



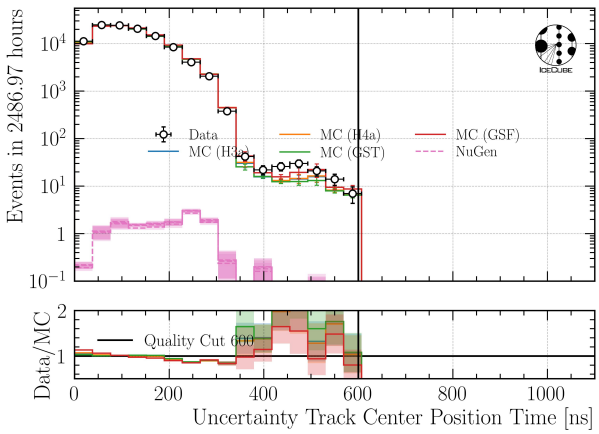
(a) Center position z.



(b) Center position z uncertainty.

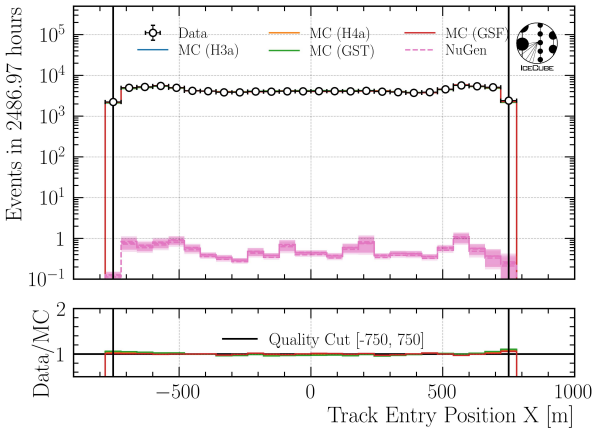


(c) Center position t.

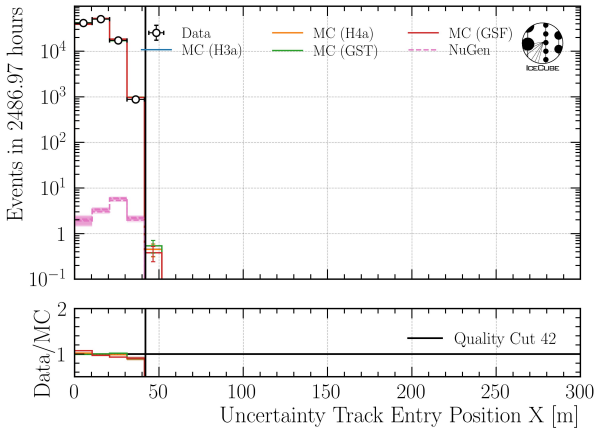


(d) Center position t uncertainty.

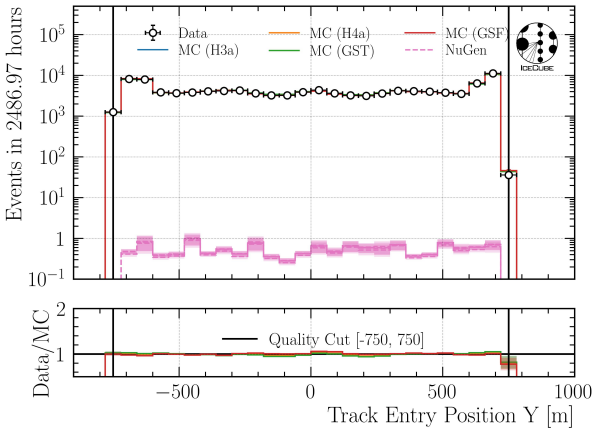
Figure B.9: Data–MC agreement for the center position z and t at Level 5. The figures include four MC predictions, one for each primary weighting. The ratio shows the number of data events (2486.97 h) divided by the expected MC events. Vertical black lines indicate the applied quality cuts. The MC predictions include the atmospheric muon flux and the neutrino background. The neutrino background is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86].



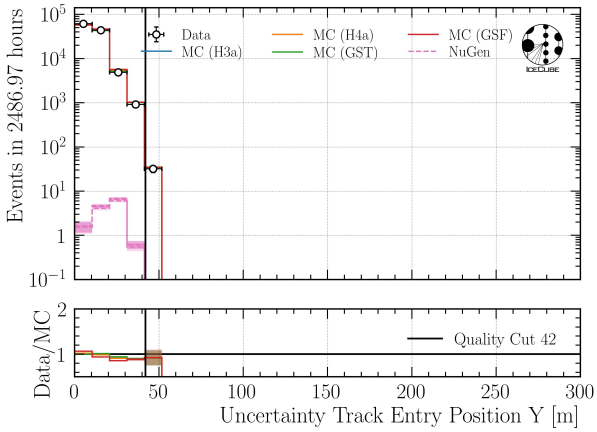
(a) Entry position x.



(b) Entry position x uncertainty.

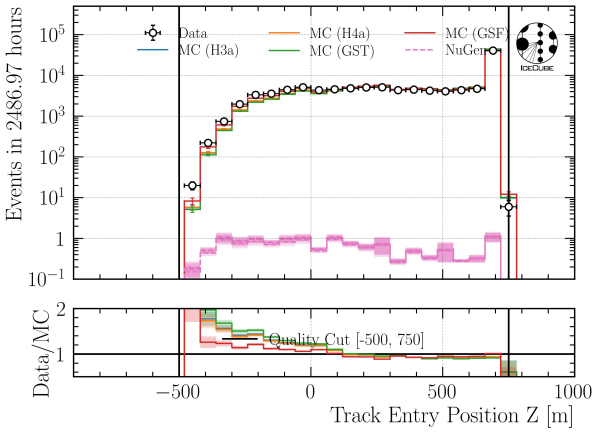


(c) Entry position y.

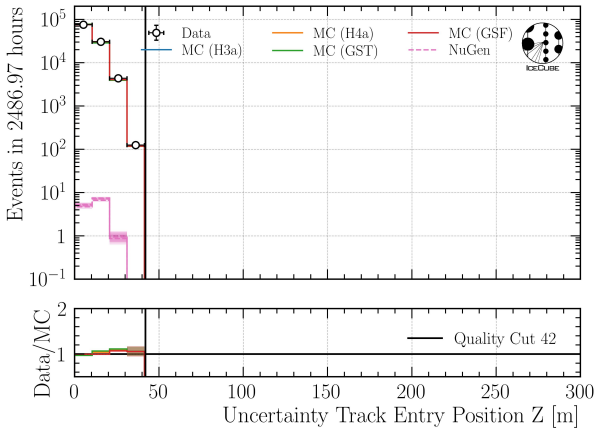


(d) Entry position y uncertainty.

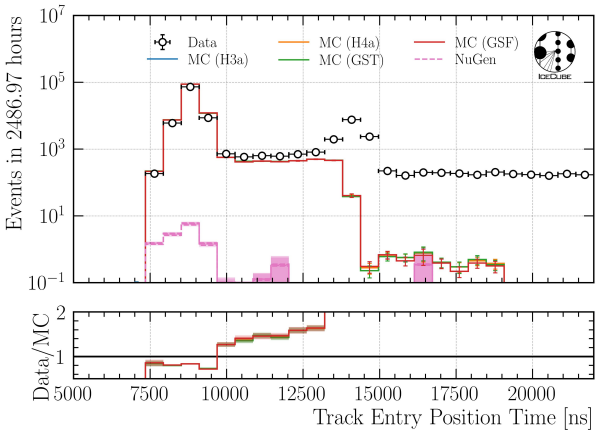
Figure B.10: Data–MC agreement for the entry position x and y at Level 5. The figures include four MC predictions, one for each primary weighting. The ratio shows the number of data events (2486.97 h) divided by the expected MC events. Vertical black lines indicate the applied quality cuts. The MC predictions include the atmospheric muon flux and the neutrino background. The neutrino background is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86].



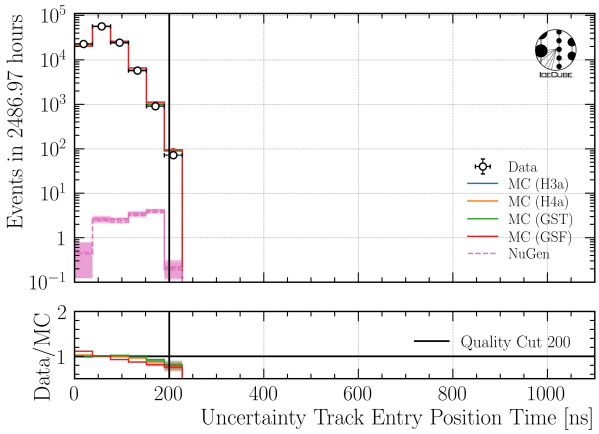
(a) Entry position z.



(b) Entry position z uncertainty.

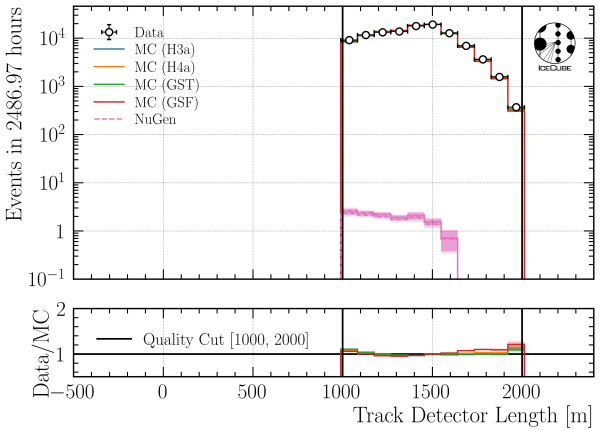


(c) Entry position t.

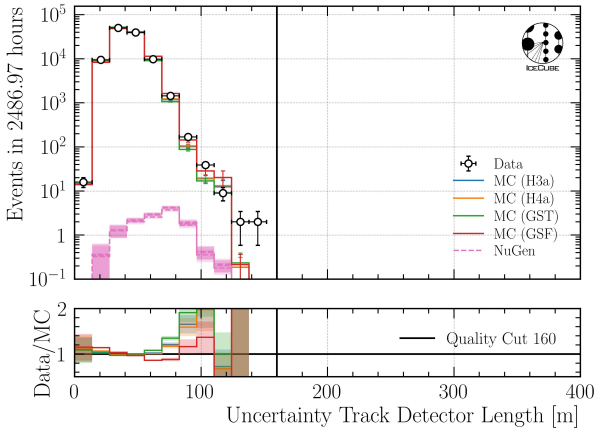


(d) Entry position t uncertainty.

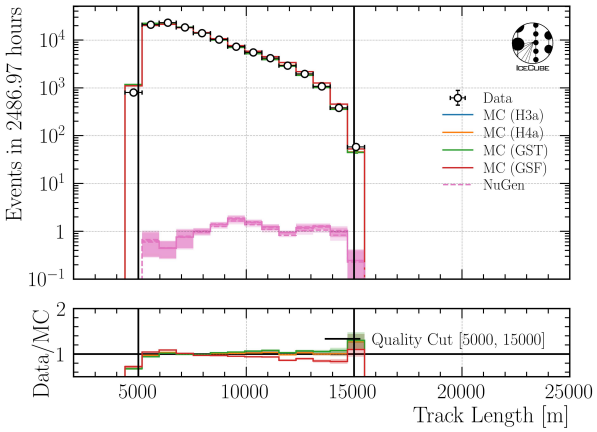
Figure B.11: Data–MC agreement for the entry position z and t at Level 5. The figures include four MC predictions, one for each primary weighting. The ratio shows the number of data events (2486.97 h) divided by the expected MC events. Vertical black lines indicate the applied quality cuts. The MC predictions include the atmospheric muon flux and the neutrino background. The neutrino background is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86].



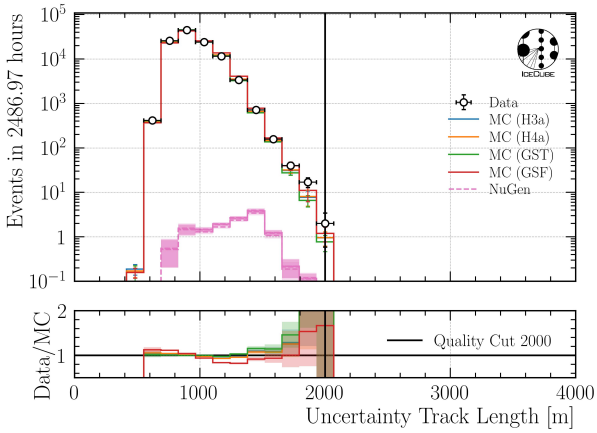
(a) Track length in detector.



(b) Track length in detector uncertainty.

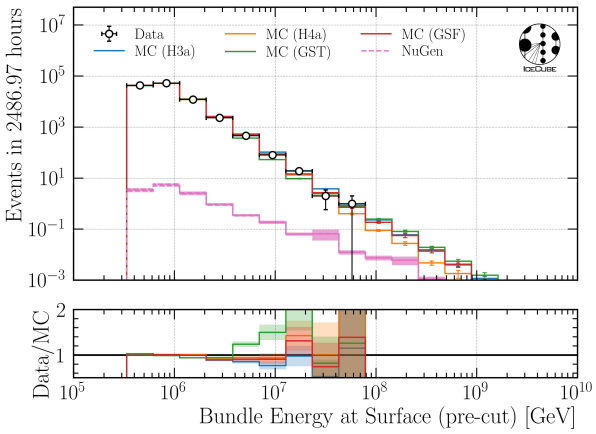


(c) Track length.

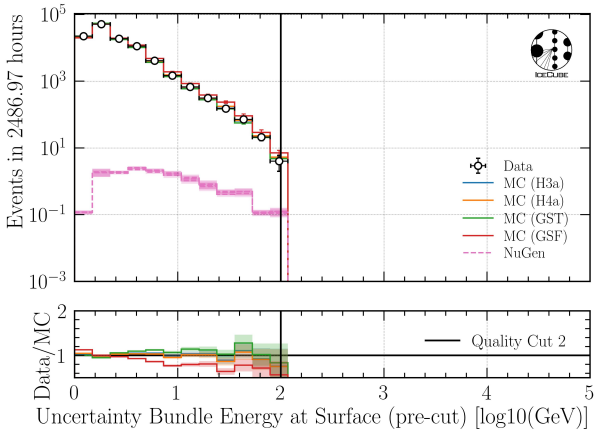


(d) Track length uncertainty.

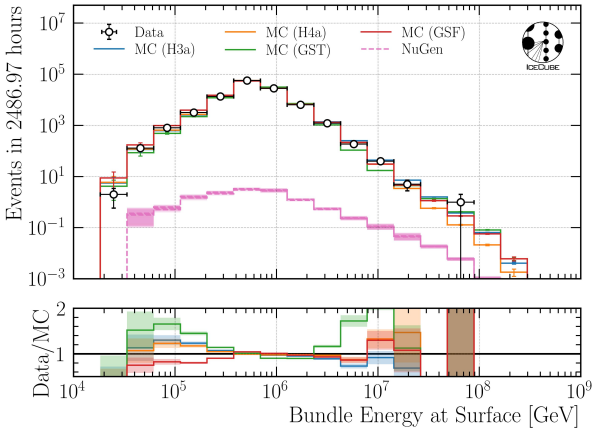
Figure B.12: Data–MC agreement for the track lengths at Level 5. The figures include four MC predictions, one for each primary weighting. The ratio shows the number of data events (2486.97 h) divided by the expected MC events. Vertical black lines indicate the applied quality cuts. The MC predictions include the atmospheric muon flux and the neutrino background. The neutrino background is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86].



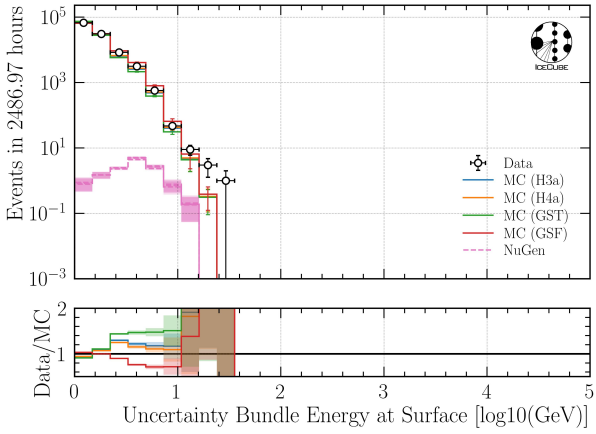
(a) Pre-cut bundle energy at surface.



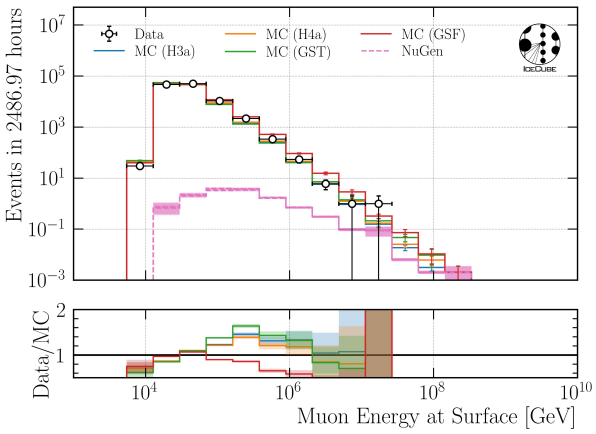
(b) Pre-cut bundle energy at surface uncertainty.



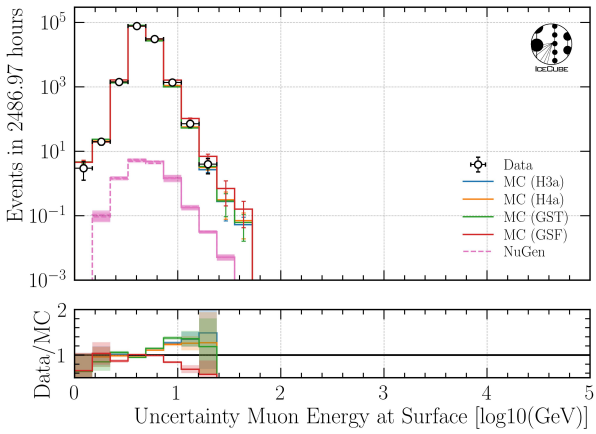
(c) Bundle energy at surface.



(d) Bundle energy at surface uncertainty.

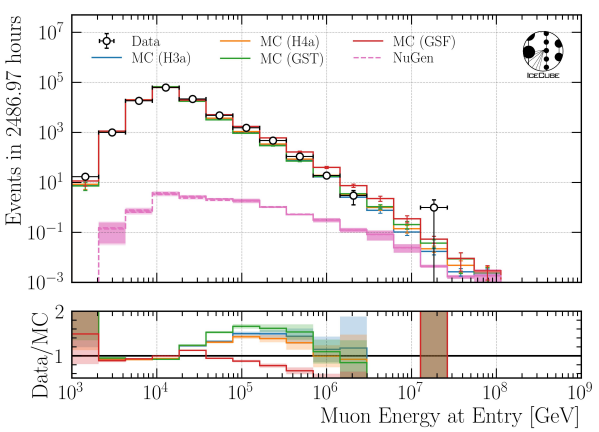


(e) Leading muon energy at surface.

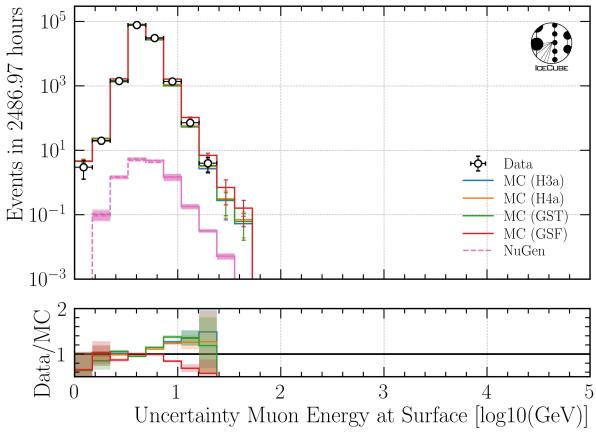


(f) Leading muon energy at surface uncertainty.

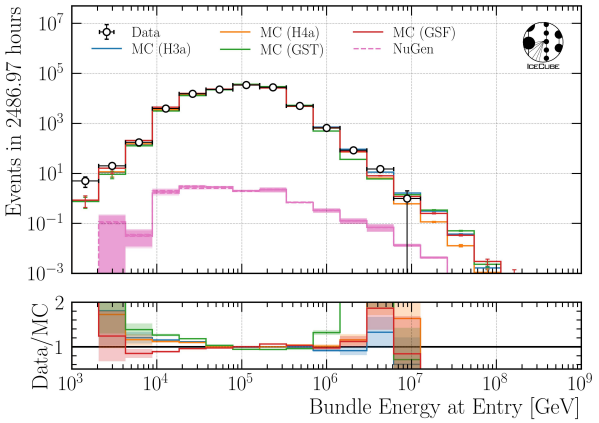
Figure B.13: Data–MC agreement for the energy reconstructions at surface at Level 5. The figure includes four MC predictions, one for each primary weighting. The ratio shows the number of data events (2486.97 h) divided by the expected MC events. Vertical black lines indicate the applied quality cuts. The MC predictions include the atmospheric muon flux and the neutrino background. The neutrino background is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86].



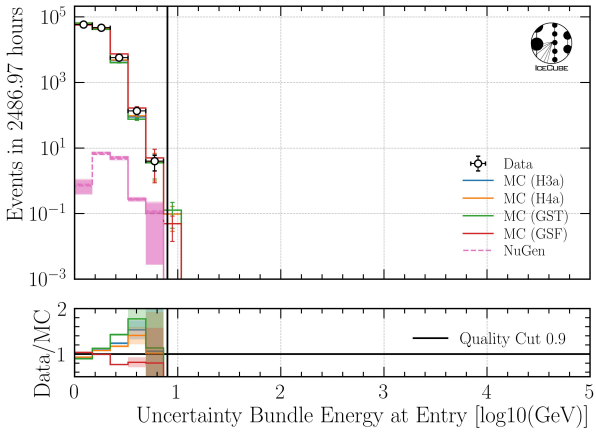
(a) Leading muon energy at entry.



(b) Leading muon energy at entry uncertainty.



(c) Bundle energy at entry.

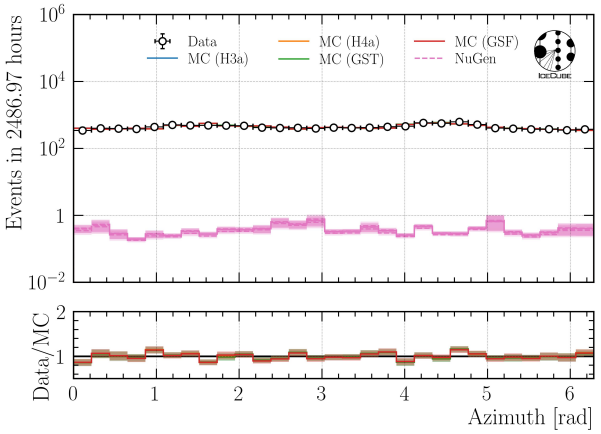


(d) Bundle energy at entry uncertainty.

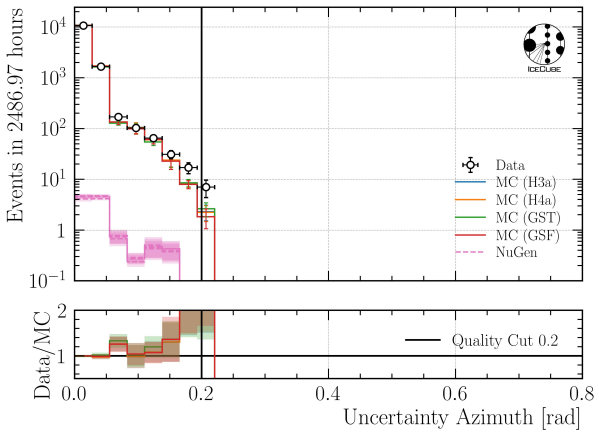
Figure B.14: Data–MC agreement for the energy reconstructions at entry at Level 5. The figures include four MC predictions, one for each primary weighting. The ratio shows the number of data events (2486.97 h) divided by the expected MC events. These comparisons are performed on *Level 5* after the quality cuts have been applied. The MC predictions include the atmospheric muon flux and the neutrino background. The neutrino background is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86].

B.4 Data–MC: Final Level

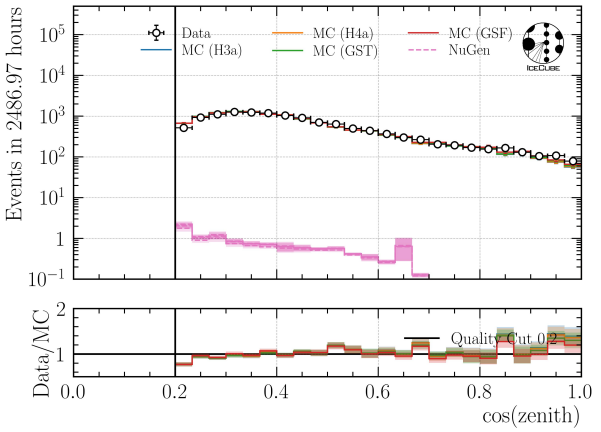
In this section, the agreement between experimental data and MC simulation is presented for the *Final Level*. The distributions of the reconstructed direction, track geometry, and their corresponding uncertainties are shown. The data corresponds to the *burnsample*, which accounts for approximately 2 % of the total data set. The MC simulation includes both the atmospheric muon flux and the neutrino background. In this level, all quality cuts and the leadingness cut of 0.4 are applied.



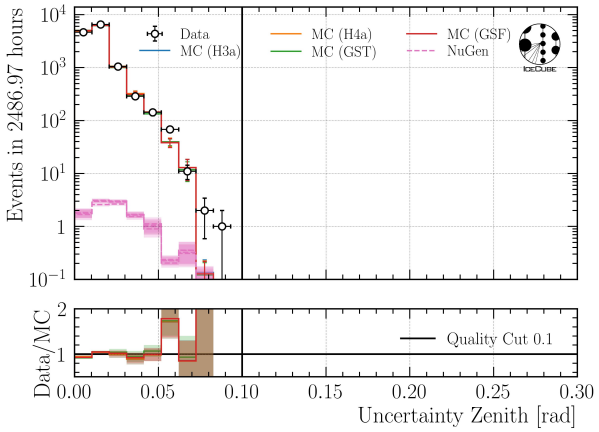
(a) Azimuth.



(b) Azimuth uncertainty.

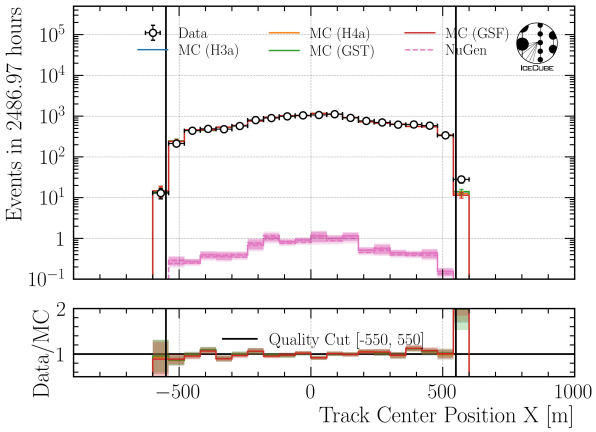


(c) $\cos(\theta)$.

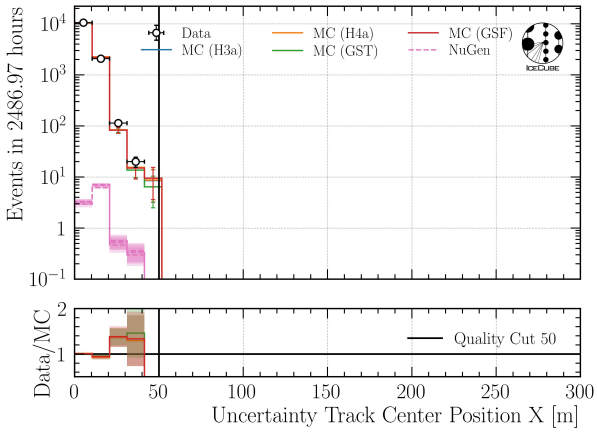


(d) $\cos(\theta)$ uncertainty.

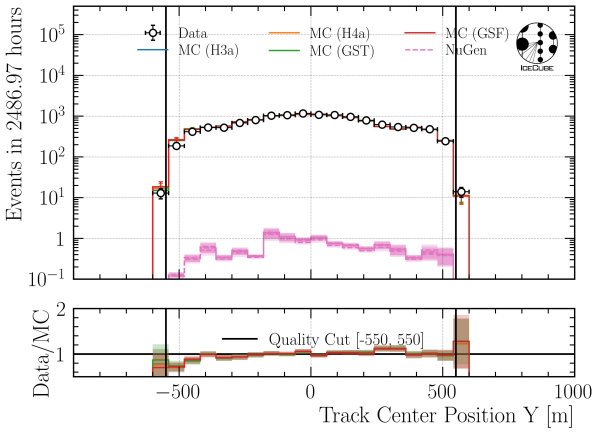
Figure B.15: Data–MC agreement for the directional reconstructions at the Final Level. The figures include four MC predictions, one for each primary weighting. The ratio shows the number of data events (2486.97 h) divided by the expected MC events. These comparisons are performed on the *Final Level*, where all quality cuts have been applied. The MC predictions include the atmospheric muon flux and the neutrino background. The neutrino background is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86].



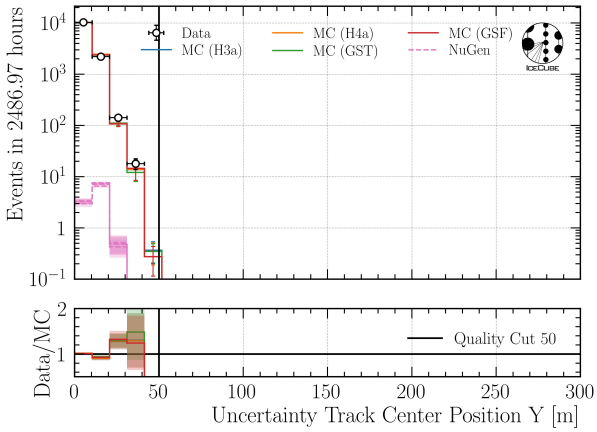
(a) Center position x.



(b) Center position x uncertainty.

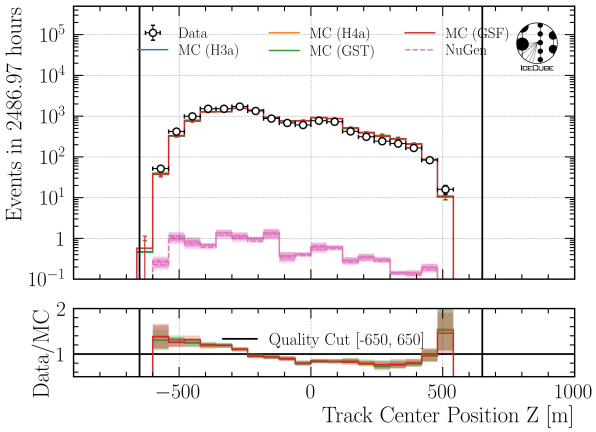


(c) Center position y.

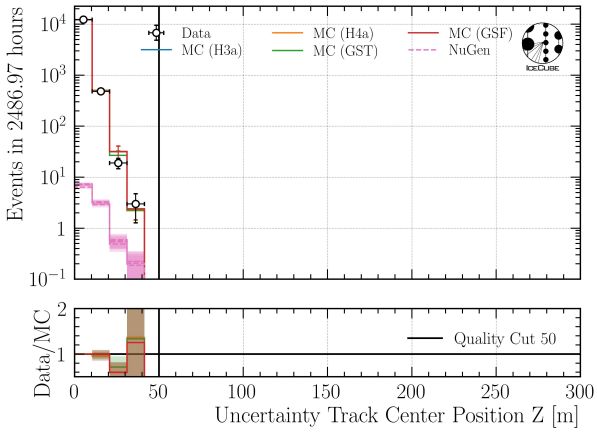


(d) Center position y uncertainty.

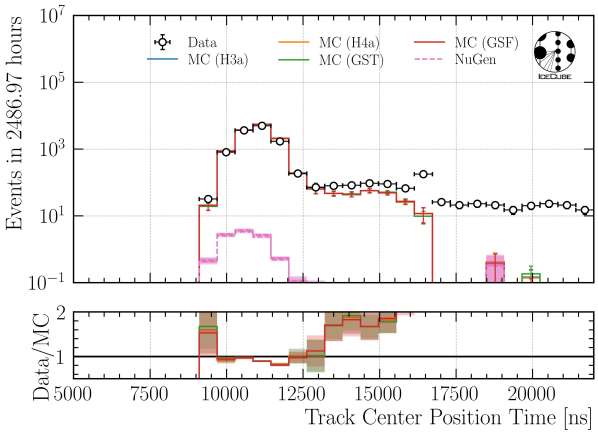
Figure B.16: Data–MC agreement for the center position x and y at the Final Level. The figures include four MC predictions, one for each primary weighting. The ratio shows the number of data events (2486.97 h) divided by the expected MC events. These comparisons are performed on the *Final Level*, where all quality cuts have been applied. The MC predictions include the atmospheric muon flux and the neutrino background. The neutrino background is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86].



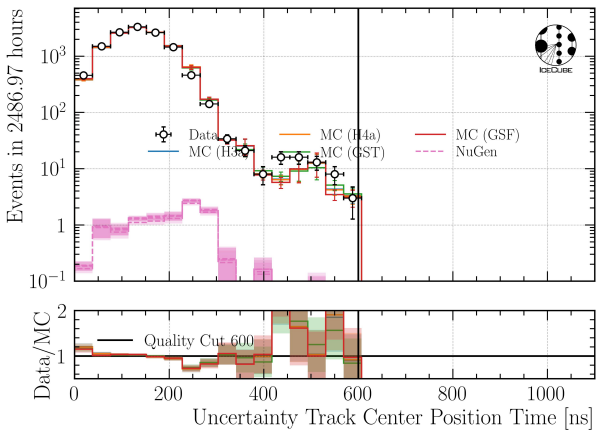
(a) Center position z.



(b) Center position z uncertainty.

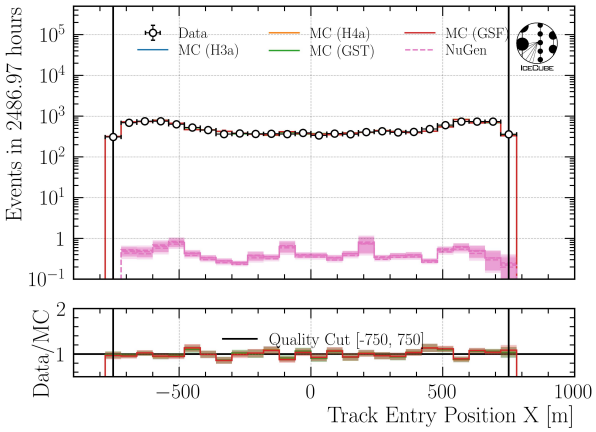


(c) Center position t.

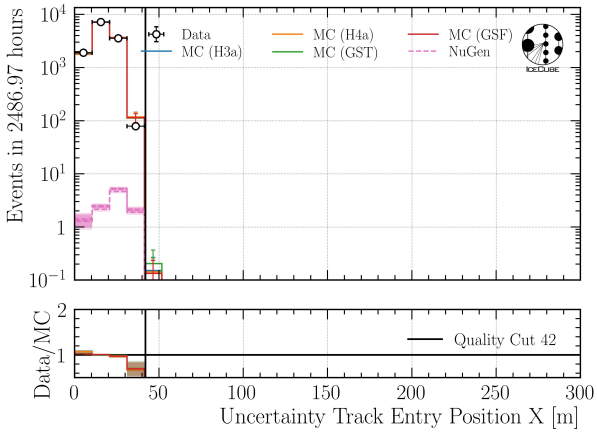


(d) Center position t uncertainty.

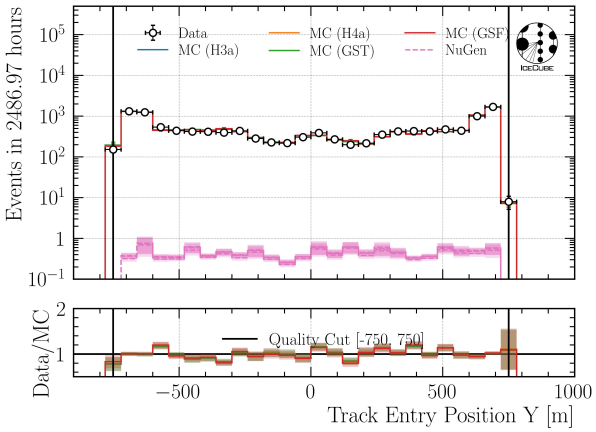
Figure B.17: Data–MC agreement for the center position z and t at the Final Level. The figures include four MC predictions, one for each primary weighting. The ratio shows the number of data events (2486.97 h) divided by the expected MC events. These comparisons are performed on the *Final Level*, where all quality cuts have been applied. The MC predictions include the atmospheric muon flux and the neutrino background. The neutrino background is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86].



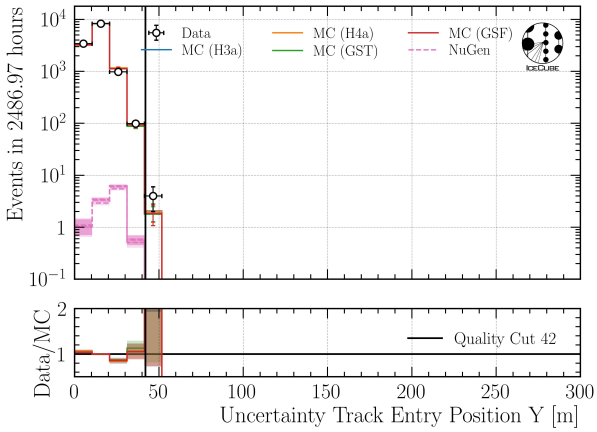
(a) Entry position x.



(b) Entry position x uncertainty.

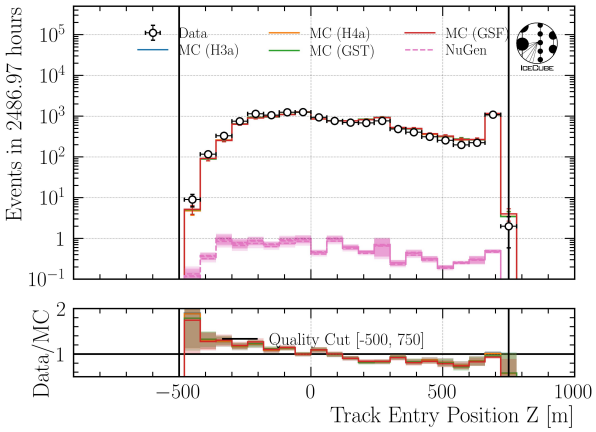


(c) Entry position y.

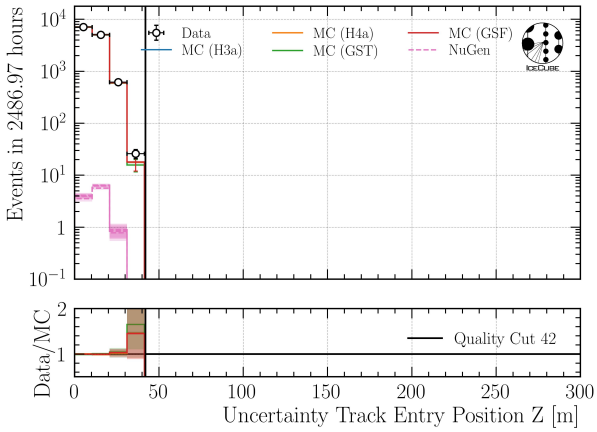


(d) Entry position y uncertainty.

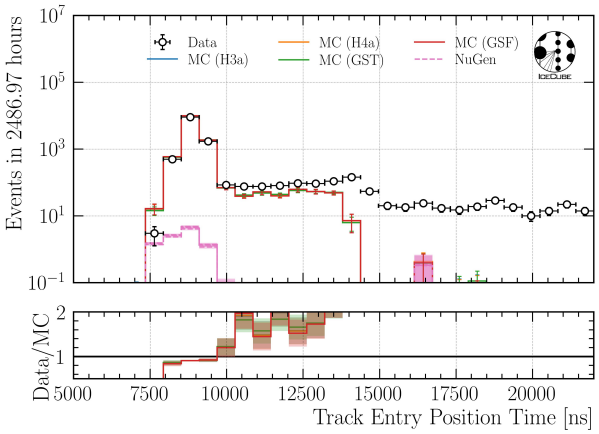
Figure B.18: Data–MC agreement for the entry position x and y at the Final Level. The figures include four MC predictions, one for each primary weighting. The ratio shows the number of data events (2486.97 h) divided by the expected MC events. These comparisons are performed on the *Final Level*, where all quality cuts have been applied. The MC predictions include the atmospheric muon flux and the neutrino background. The neutrino background is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86].



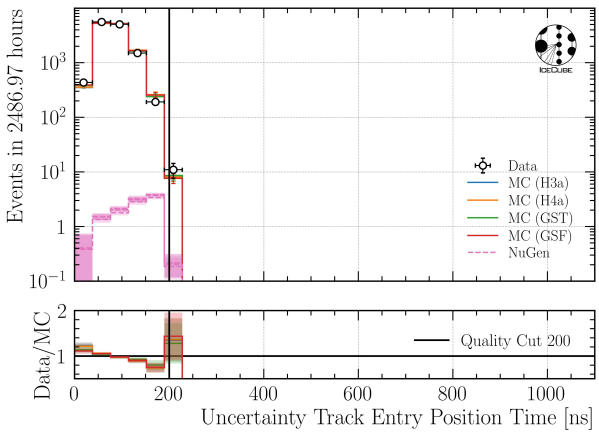
(a) Entry position z.



(b) Entry position z uncertainty.

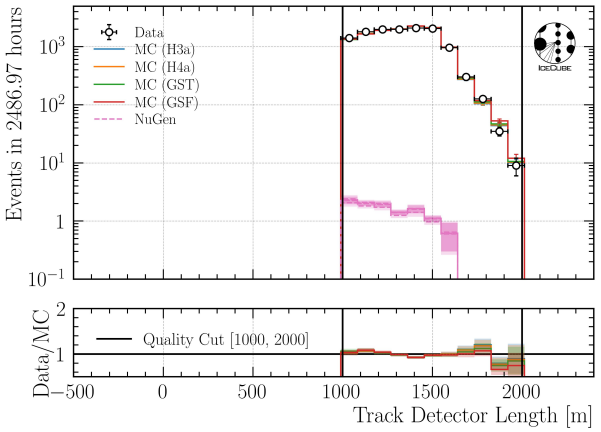


(c) Entry position t.

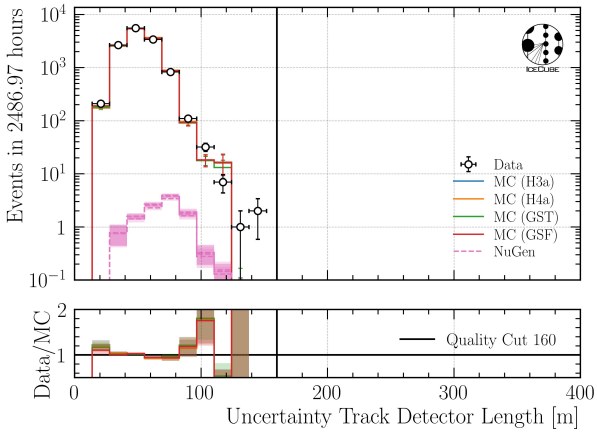


(d) Entry position t uncertainty.

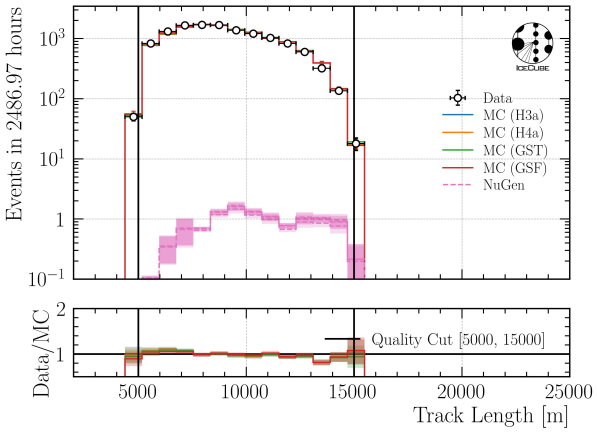
Figure B.19: Data–MC agreement for the entry position z and t at the Final Level. The figures include four MC predictions, one for each primary weighting. The ratio shows the number of data events (2486.97 h) divided by the expected MC events. These comparisons are performed on the *Final Level*, where all quality cuts have been applied. The MC predictions include the atmospheric muon flux and the neutrino background. The neutrino background is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86].



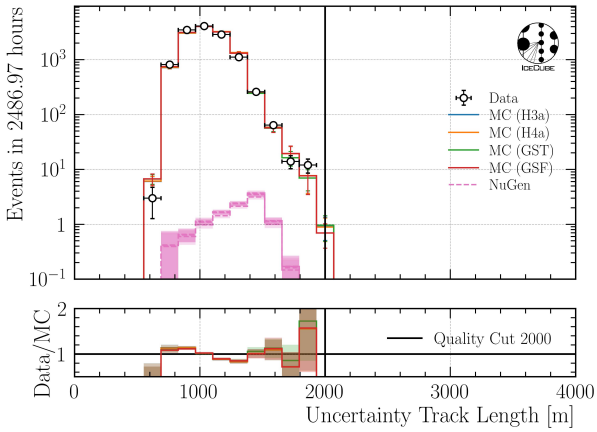
(a) Track length in detector.



(b) Track length in detector uncertainty.

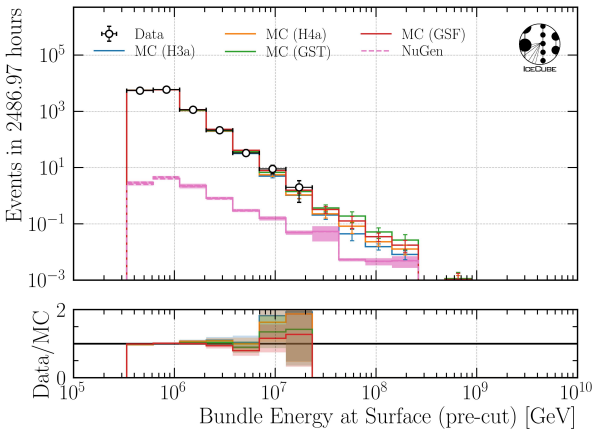


(c) Track length.

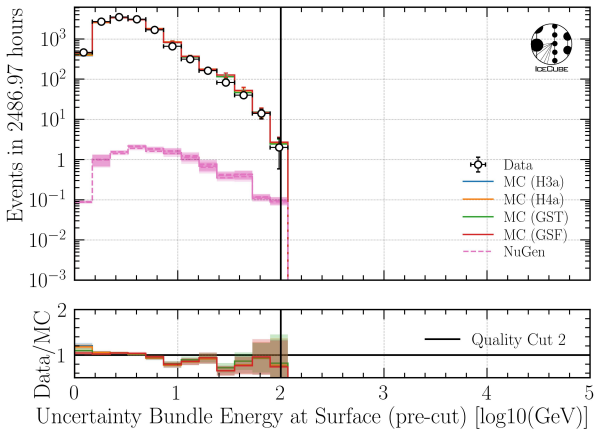


(d) Track length uncertainty.

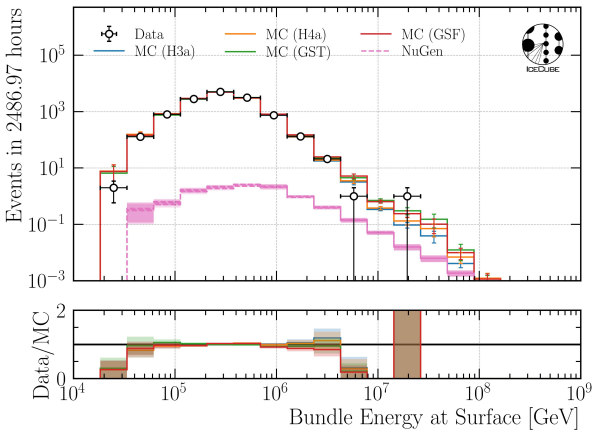
Figure B.20: Data–MC agreement for the track lengths at the Final Level. The figures include four MC predictions, one for each primary weighting. The ratio shows the number of data events (2486.97 h) divided by the expected MC events. These comparisons are performed on the *Final Level*, where all quality cuts have been applied. The MC predictions include the atmospheric muon flux and the neutrino background. The neutrino background is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86].



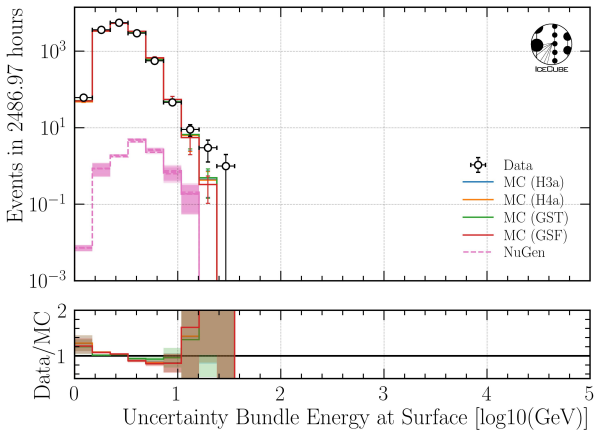
(a) Pre-cut bundle energy at surface.



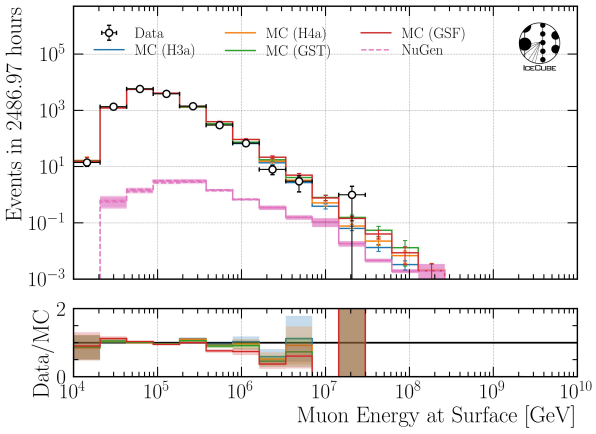
(b) Pre-cut bundle energy at surface uncertainty.



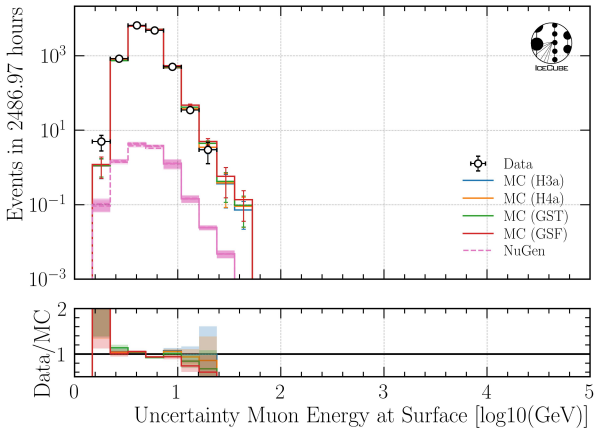
(c) Bundle energy at surface.



(d) Bundle energy at surface uncertainty.

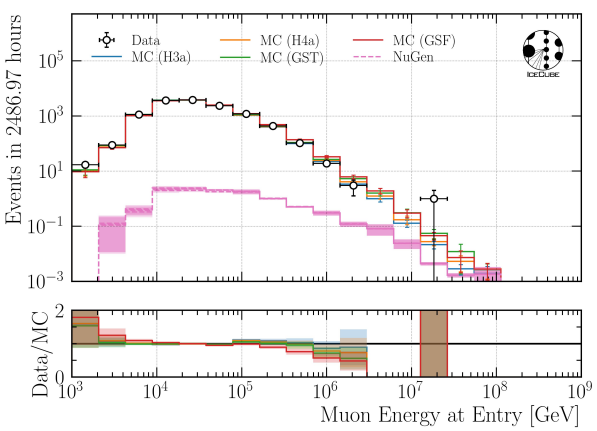


(e) Leading muon energy at surface.

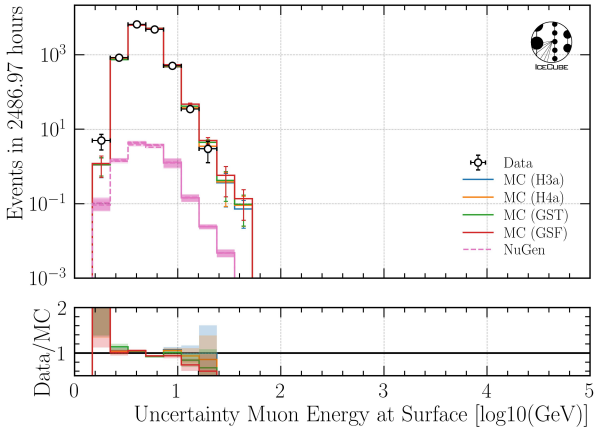


(f) Leading muon energy at surface uncertainty.

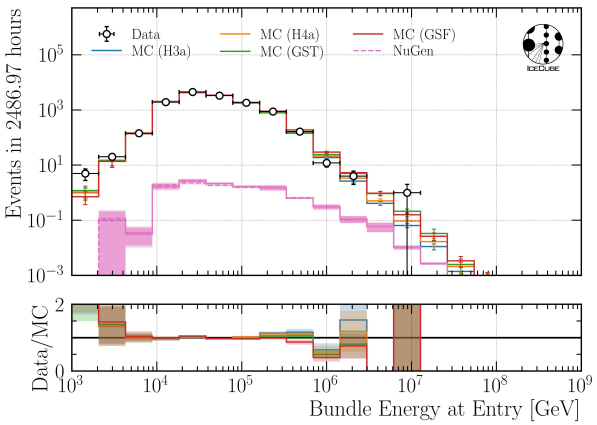
Figure B.21: Data–MC agreement for the energy reconstructions at surface at the Final Level. The figures include four MC predictions, one for each primary weighting. The ratio shows the number of data events (2486.97 h) divided by the expected MC events. These comparisons are performed on the *Final Level*, where all quality cuts have been applied. The MC predictions include the atmospheric muon flux and the neutrino background. The neutrino background is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86].



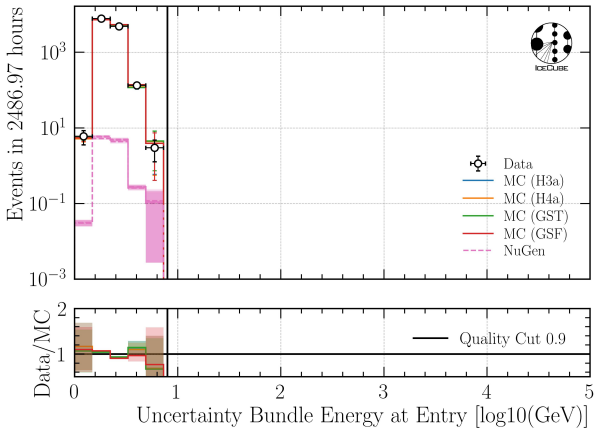
(a) Leading muon energy at entry.



(b) Leading muon energy at entry uncertainty.



(c) Bundle energy at entry.

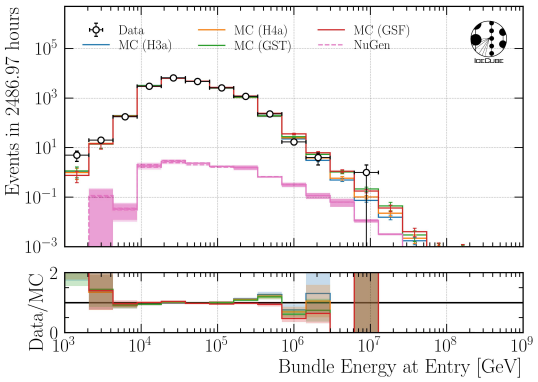


(d) Bundle energy at entry uncertainty.

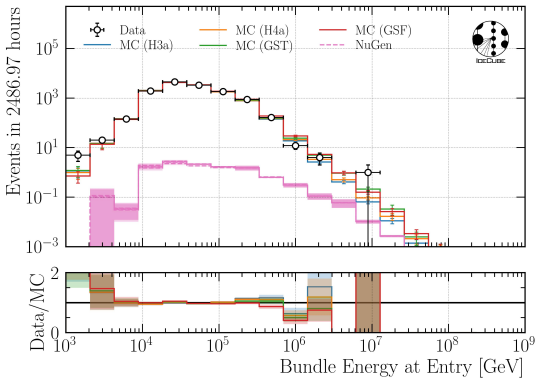
Figure B.22: Data–MC agreement for the energy reconstructions at entry at the Final Level. The figures include four MC predictions, one for each primary weighting. The ratio shows the number of data events (2486.97 h) divided by the expected MC events. These comparisons are performed on the *Final Level*, where all quality cuts have been applied. The MC predictions include the atmospheric muon flux and the neutrino background. The neutrino background is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86].

B.5 Final Level – Potential Leadingness Cuts

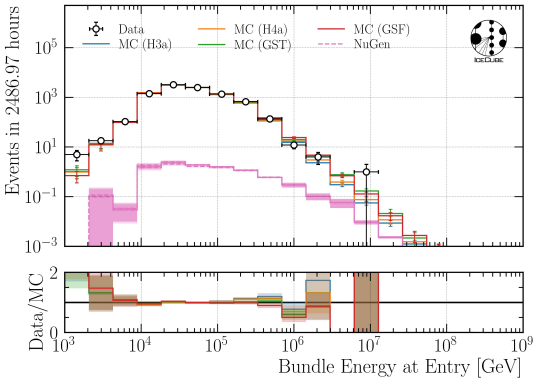
In this section, the impact of different leadingness cuts on the Data–MC agreement is investigated. The leadingness L is a variable to estimate if the leading muon in a bundle carries the main fraction of the bundle energy. It is defined in Eq. (7.3). Figure B.24 shows the resolution of the reconstructed leadingness. In Figure B.23, the agreement between data and MC is shown for different leadingness cuts ranging from $L > 0.3$ to $L > 0.9$. A cut of $L > 0.4$ is chosen for the final selection. In Figure B.25, the scaling factors for the different leadingness cuts are shown. In Figure B.26, the impact of the leadingness cut on the expected number of muon events for 12 years of data is shown.



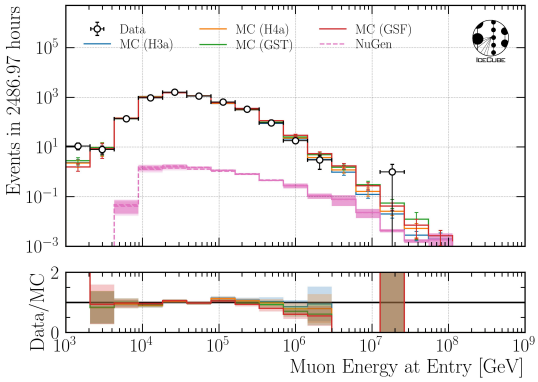
(a) $L > 0.3$.



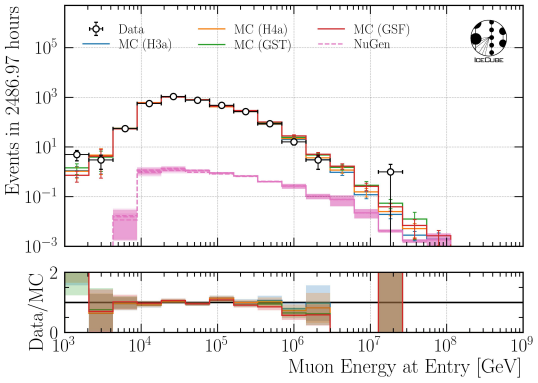
(b) $L > 0.4$.



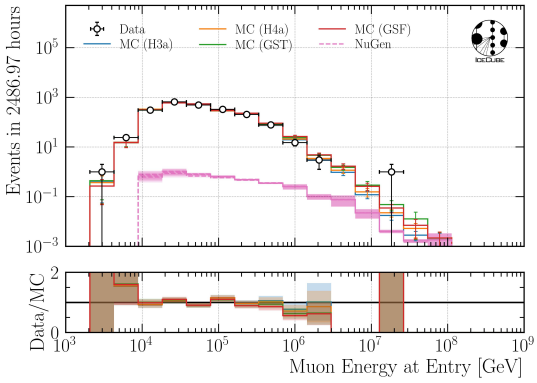
(c) $L > 0.5$.



(d) $L > 0.7$.

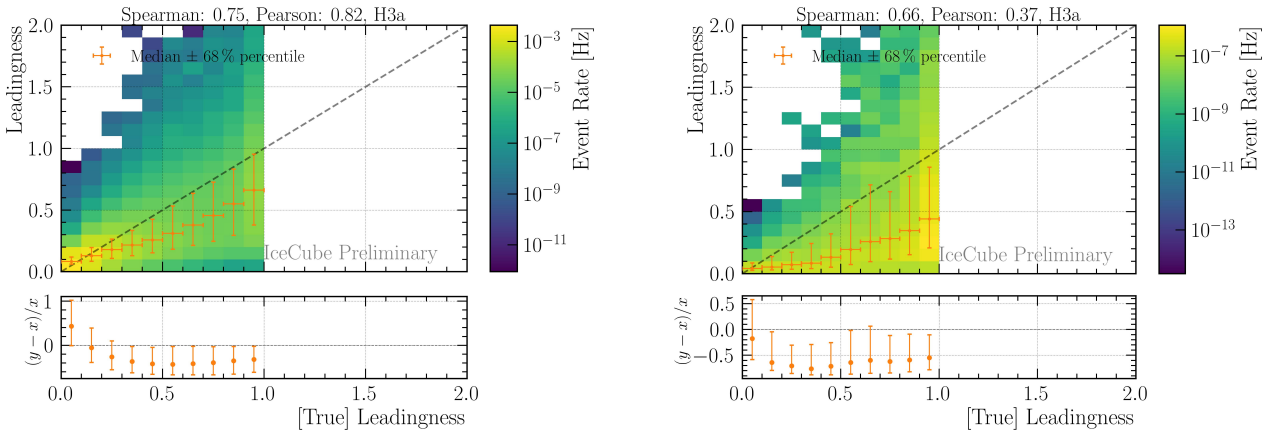


(e) $L > 0.8$.



(f) $L > 0.9$.

Figure B.23: Data–MC agreement for the energy reconstructions at entry for different leadingness cuts (L). The figures include four MC predictions, one for each primary weighting. The ratio shows the number of data events (2486.97 h) divided by the expected MC events. These comparisons are performed on potential final levels after all cuts have been applied for several leadingness cuts. The neutrino background is weighted using a single powerlaw with a normalization of $n_\nu = 1.8 \cdot 10^{-18}\,\mathrm{GeV}^{-1}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$ and a spectral index of $\gamma_\nu = 2.52$ [86].



(a) Leadingness on all events.

(b) Leadingness for true muon energy at surface > 1 PeV.

Figure B.24: Reconstruction performance of the leadingness variable at Level 5. The leadingness is shown for (a) all events and (b) only for events with a true leading muon energy at surface greater than 1 PeV. In both figures, the leadingness reconstruction underestimates the true leadingness. Leadingness is defined as in Equation 7.3.

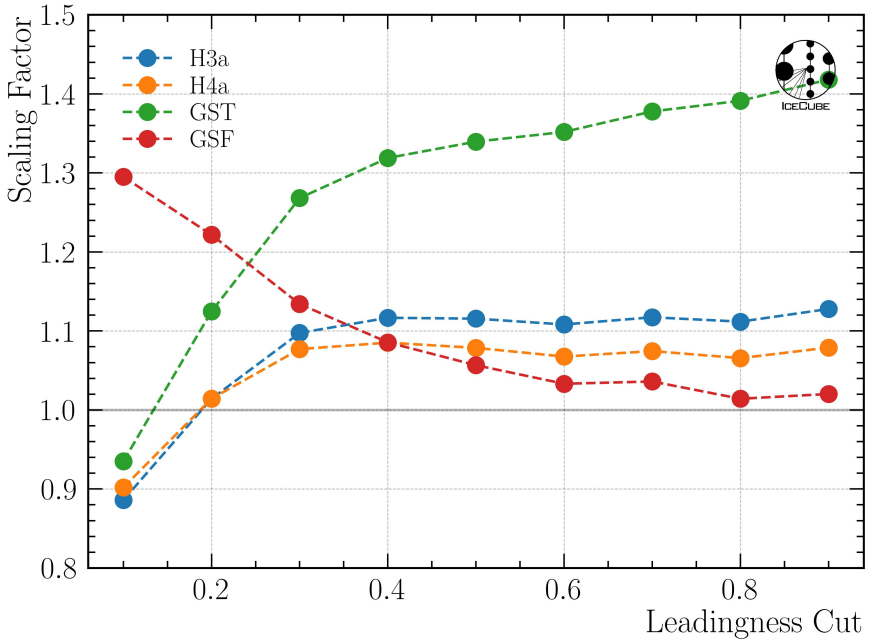
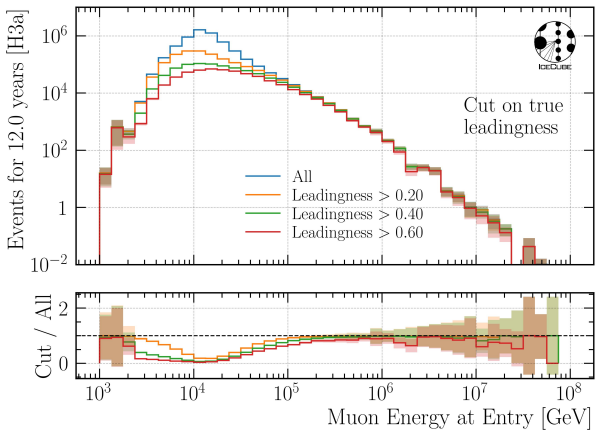
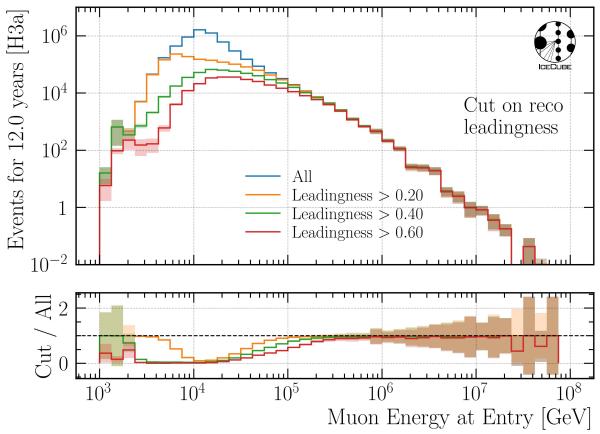


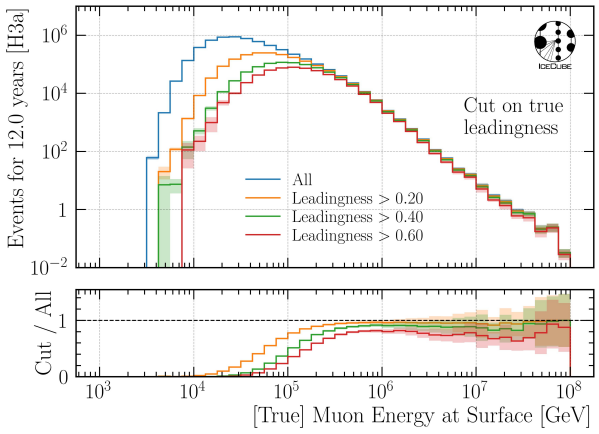
Figure B.25: Scaling factors $n_{s,i}$ calculated with Equation (7.2) are presented as a function of the leadingness cut L.



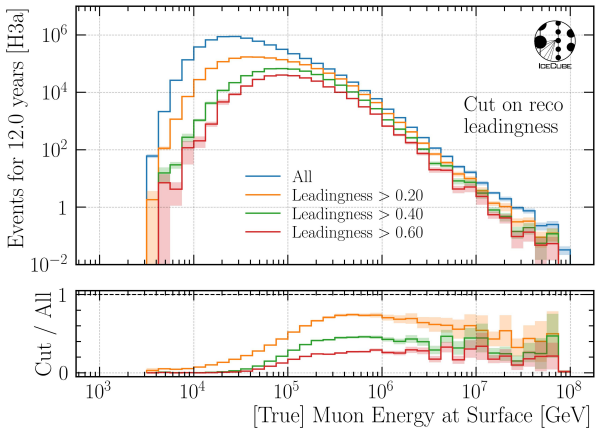
(a) Leading muon energy at entry, cut on true leadingness.



(b) Leading muon energy at surface, cut on reconstructed leadingness.



(c) True muon energy at surface, cut on true leadingness.



(d) Bundle energy at surface, cut on reconstructed leadingness.

Figure B.26: This figure presents the impact of three different leadingness cuts $L \in \{0.2, 0.4, 0.6\}$ on the leading muon energy at entry and at surface, the proxy and target variable of the unfolding, respectively. Distributions are weighted according to the H3a primary model. The leadingness cut is applied either on the true leadingness (left column) or on the reconstructed leadingness (right column). The ratio panel presents the effective survival fraction of events after the leadingness cut with respect to no leadingness cut.

C Burnsample Unfolding

In this chapter, supplementary material for the unfolding of the burnsample is presented. First, the regularization scan exhibits a second minimum, and the unfolding result is checked for that minimum. Then, the Data–MC agreement is evaluated for the re-weighted systematic uncertainties. Finally, the impact of coincident air showers is estimated.

C.1 Check for Second Minimum in Tau Scan

The regularization strength τ is determined using the global correlation scan, as shown in Chapter 8. The optimal value is found to be $\tau = 0.00276$. However, a second local minimum is observed at $\tau \approx 0.04$. The unfolding result using this second minimum is shown in Figure C.1. The best-fit flux values are compared to the nominal unfolding result (optimal τ) via the ratio:

$$\frac{\Phi_{\text{2nd } \tau \text{ min.}}}{\Phi_{\text{optimal } \tau}} = [1.04, 0.97, 1.02, 0.97, 1.01, 1.25]\,. \tag{C.1}$$

For the first five bins, the difference is within 5 %, while the last bin shows a larger deviation of 25 %. This is compatible within the uncertainties of the unfolding result.

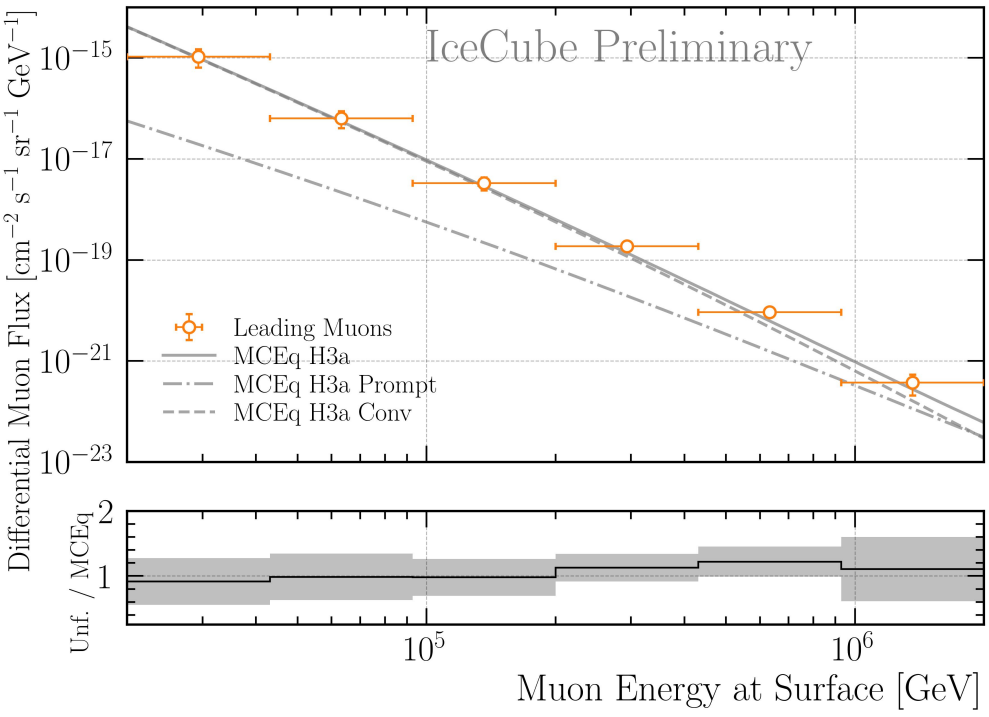
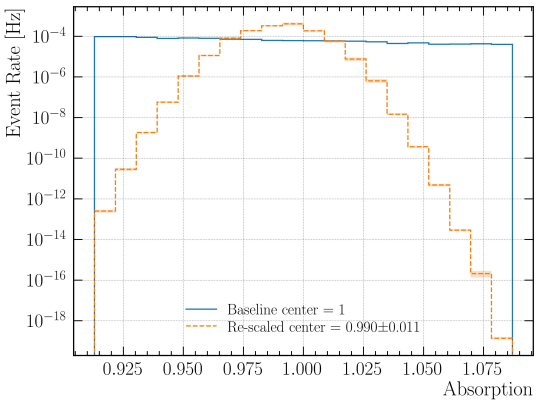


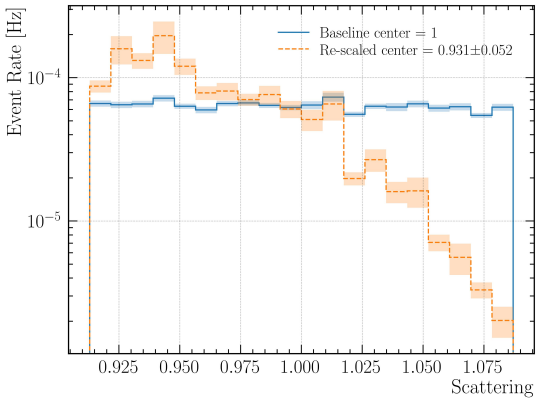
Figure C.1: Unfolding result using the regularization strength of the second minimum ($\tau \approx 0.04$). The result is compatible with the unfolding results using the minimum τ within uncertainties.

C.2 Re-Weighted Systematics

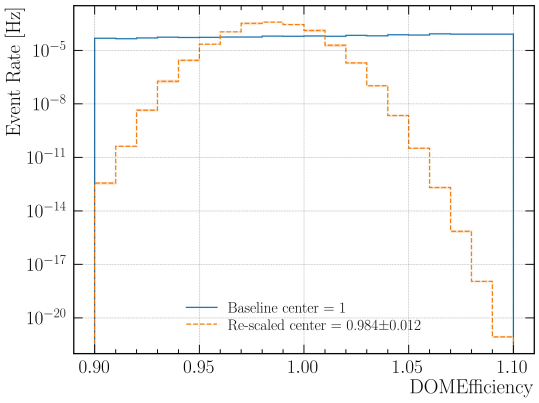
In this section, the agreement between the burnsample data and the MC simulation is presented. The MC is re-weighted using the best-fit nuisance parameters obtained from the unfolding. The agreement is shown for the five systematic uncertainties: absorption, scattering, DOM efficiency, and the hole ice parameters p_0 and p_1.



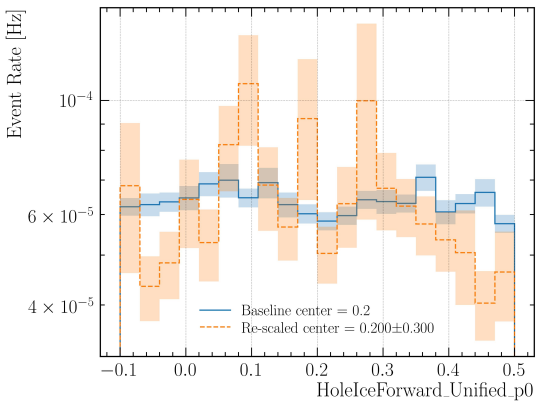
(a) Absorption



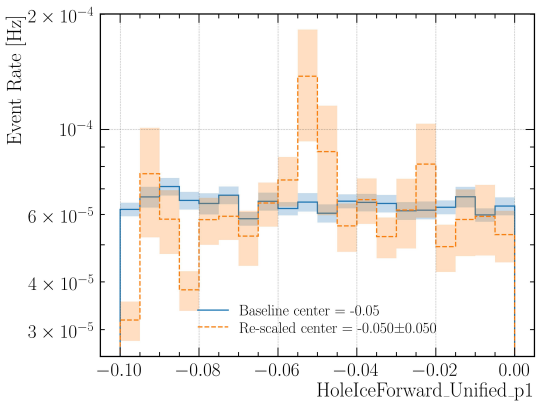
(b) Scattering.



(c) DOM Efficiency.



(d) HoleIce p0.



(e) HoleIce p1.

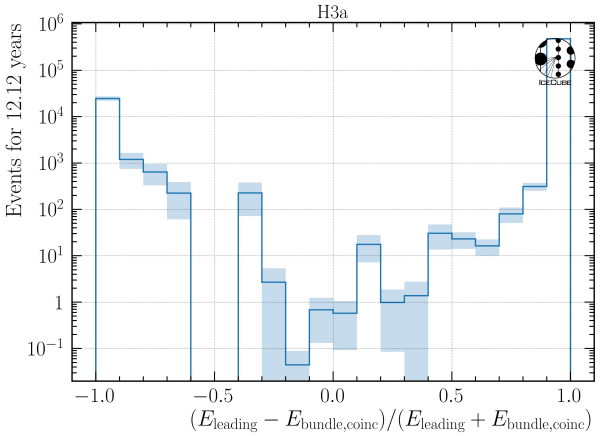
Figure C.2: Data–MC agreement for the five systematic uncertainties after re-weighting. The MC is re-weighted using the best-fit systematics parameters obtained from the unfolding (see Table 8.1). The figures show the absorption, scattering, DOM efficiency, and the hole ice parameters p_0 and p_1.

D Impact of Coincident Primaries

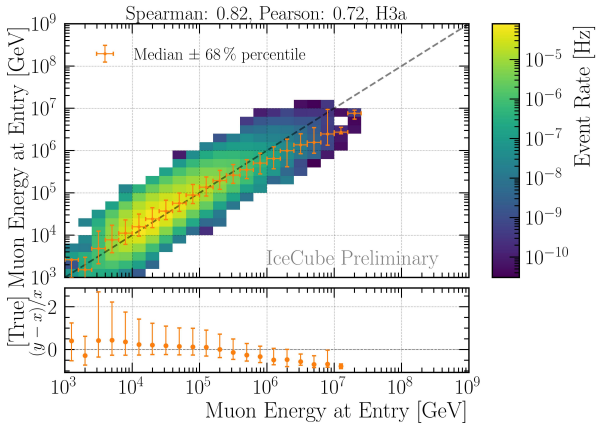
This appendix provides additional material supporting the study of the impact of coincident air showers presented in Section 8.2.8. The study is performed on two independent simulation datasets, 20904 and 22615, to ensure the robustness of the results.

D.1 Dataset 20904

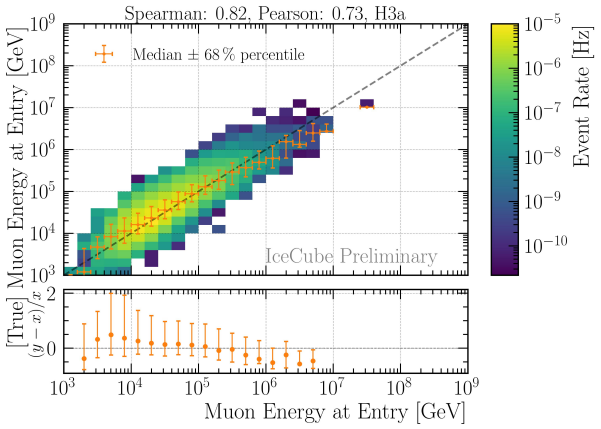
The results on the impact of coincident air showers presented in Chapter 8 rely on dataset 20904. The reconstruction performance for this dataset is shown in Figure D.1. The performance is investigated for three scenarios: without coincident primaries (top-right), with a contribution from the background bundle energy of less than 10 % (bottom-left), and with a contribution from the background bundle energy of at least 10 % (bottom-right). The 1D asymmetry distribution, defined in Equation (8.63), is shown in the top-left panel for reference. A comparison of the resolution matrices confirms that the reconstruction remains robust against coincident primaries in this dataset.



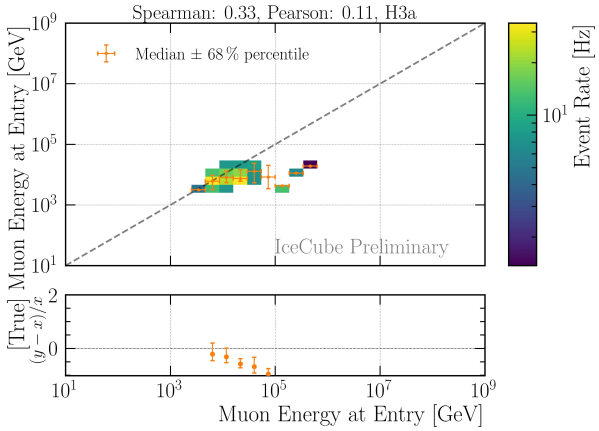
(a) 1D asymmetry as defined in Eq. (8.63).



(b) Reconstruction performance without coincident primaries.



(c) Reconstruction performance with $E_{\text{leading}}/E_{\text{bundle, coinc}} < 0.1$.



(d) Reconstruction performance with $E_{\text{leading}}/E_{\text{bundle, coinc}} \geq 0.1$.

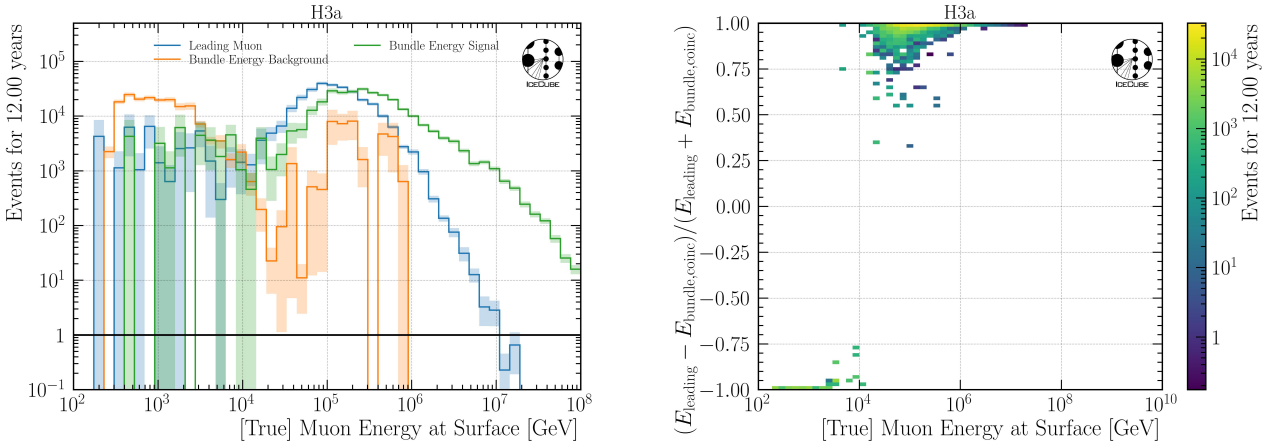
Figure D.1: Impact of coincident primaries on the reconstruction performance of the leading muon energy at entry for dataset 20904. The top-right panel shows the performance without coincident primaries. The bottom-left panel includes coincident primaries with less than 10 % background contribution, while the bottom-right panel includes contributions of at least 10 %. The top-left panel displays the 1D asymmetry (see Equation (8.63)), indicating that most events are dominated by the signal primary. The reconstruction performance is robust against coincident primaries.

D.2 Dataset 22615

In addition, the impact of coincident primaries is studied using dataset 22615. The observed differences specifically the higher rate of coincident primaries in dataset 22615 originate from the simulation setup. First, dataset 22615 was generated without the hit filter module in `polyplopia` that was used for dataset 20904. This filter removes events where the primary particle does not produce any photoelectrons in the detector. Its absence results in an overestimation of the coincident primary rate and the inclusion of noise-triggered events in dataset 22615. Second, different simulation softwares and HIM are used: dataset 22615 utilizes CORSIKA with SIBYLL 2.3d, while dataset 20904 uses MuonGun with SIBYLL 2.1.

Figure D.2a shows the 1D distribution of the leading muon energy at surface (target variable), weighted to H3a. The 2D distribution of the leading muon energy versus the asymmetry X_{asym} (see Equation (8.63)) is presented in Figure D.2b.

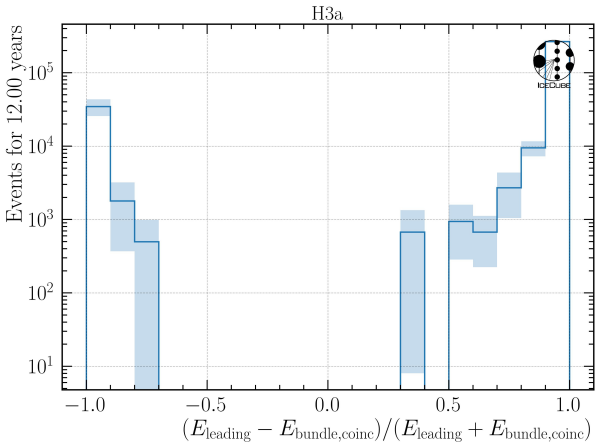
The reconstruction performance is shown in Figure D.3. Consistent with dataset 20904, deviations due to coincident primaries are small and within reconstruction uncertainties. This confirms that the impact of coincident primaries is negligible for the analysis threshold of 10 TeV.



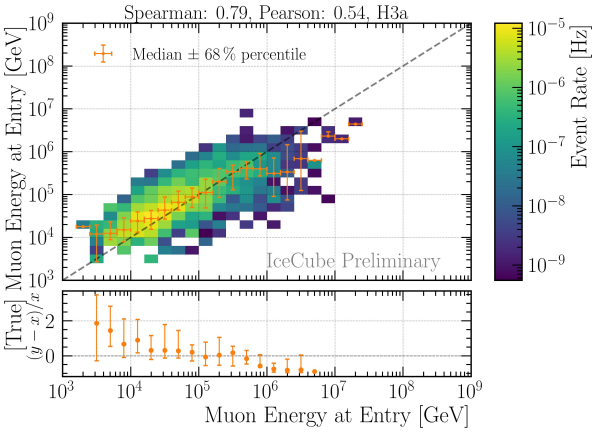
(a) 1D target distribution.

(b) 2D target distribution.

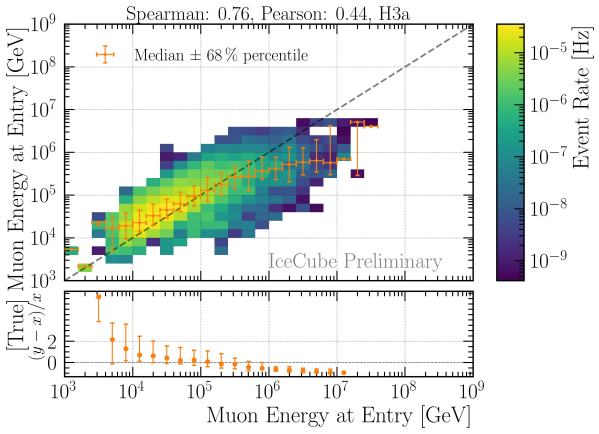
Figure D.2: Impact of coincident primaries on the leading muon energy at surface for dataset 22615. The left panel shows the 1D distribution of the leading muon energy. The right panel displays the 2D distribution versus the asymmetry X_{asym} (see Equation (8.63)). For energies above 10 TeV, the signal primary dominates, with limited significant background contribution.



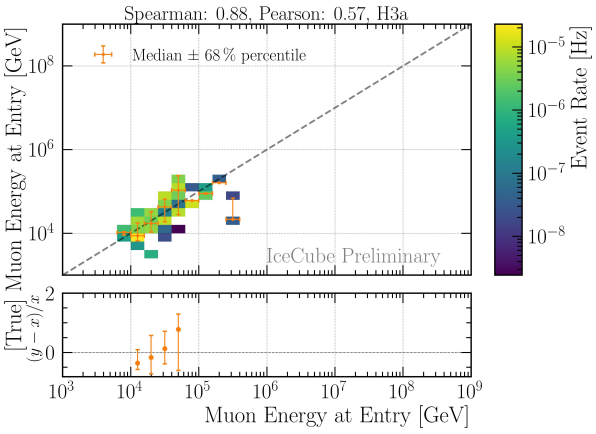
(a) 1D asymmetry as defined in Eq. (8.63).



(b) Reconstruction performance without coincident primaries.



(c) Reconstruction performance with $E_{\text{leading}}/E_{\text{bundle, coinc}} < 0.1$.



(d) Reconstruction performance with $E_{\text{leading}}/E_{\text{bundle, coinc}} \geq 0.1$.

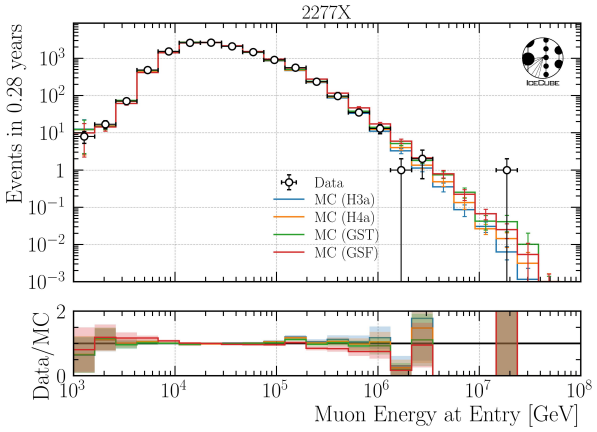
Figure D.3: Impact of coincident primaries on the reconstruction performance of the leading muon energy at entry for dataset 22615. The top-right panel shows the performance without coincident primaries. The bottom-left panel includes coincident primaries with less than 10 % background contribution, while the bottom-right panel includes contributions of at least 10 %. The top-left panel displays the 1D asymmetry (see Equation (8.63)), indicating that most events are dominated by the signal primary. The reconstruction performance remains stable.

E Burnsample Unfolding – Combined MC

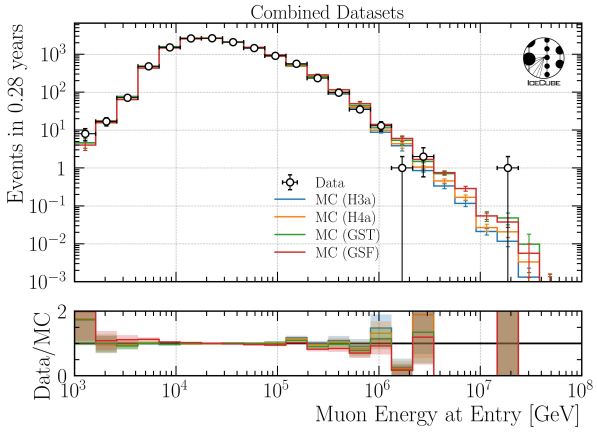
This chapter presents the burnsample unfolding using a combined MC simulation set. Section E.1 covers general properties, including data–MC agreement and the effective area. The unfolding performance on MC simulations is evaluated in Section E.2, and the unfolding of the experimental burnsample data is presented in Section E.3.

E.1 General

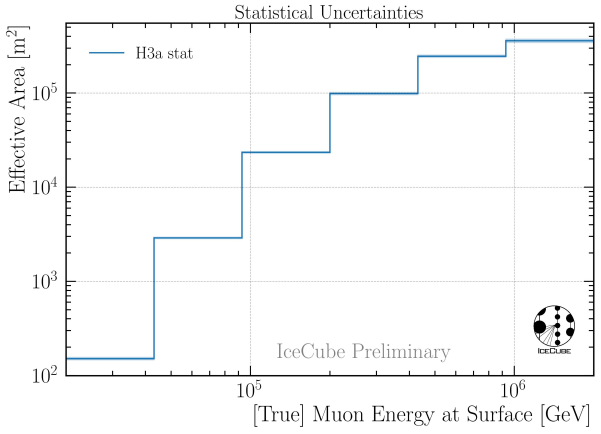
This sections presents the data–MC agreement for 2277X and the combined MC datasets, which are in agreement within uncertainties. The effective area is calculated for the burnsample unfolding with the combined set, and the correlation between the reconstructed leading muon energy at entry and the true leading muon energy at surface is shown.



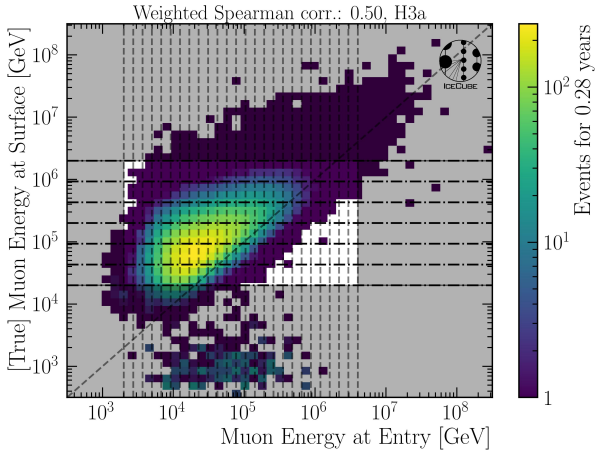
(a) data–MC comparison for burnsample with only 2277X.



(b) data–MC comparison for burnsample with all datasets combined.



(c) Effective area for burnsample with combined MC.

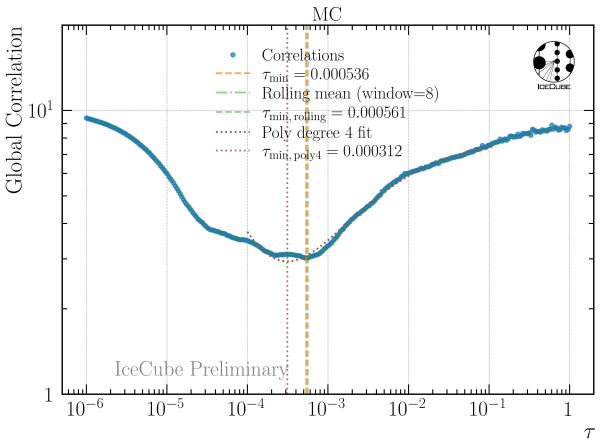


(d) Response matrix for burnsample with combined MC.

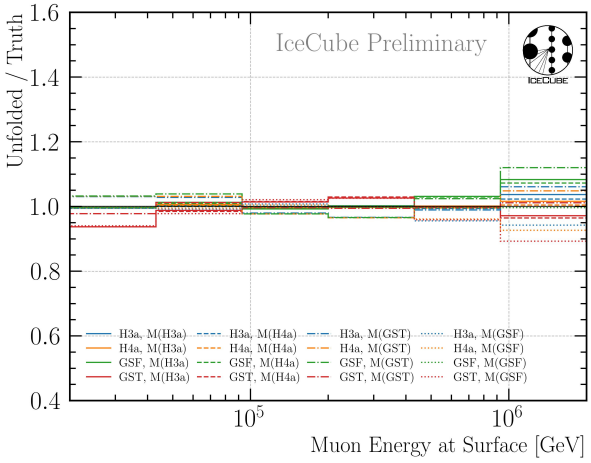
Figure E.1: The data–MC comparison in (a) shows only dataset 2277X, while (b) shows all datasets combined. Both data–MC comparisons are compatible within uncertainties. Figure (c) presents the effective area, and (d) the response matrix on *Final Level* for the burnsample unfolding with combined MC. Shaded area indicates underflow and overflow bins. The second distribution below 10 TeV target energy is caused by coincident primaries, see Section 8.2.8.

E.2 MC Unfolding

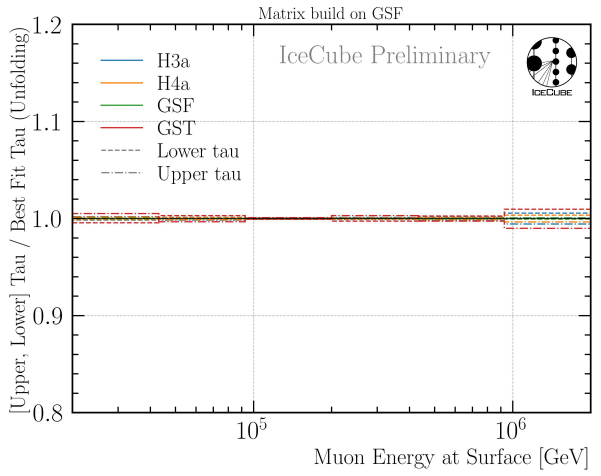
This section presents the unfolding performance on the combined MC set using the same unfolding configuration as for the final full sample unfolding. The global correlation, unfolding bias, regularization impact, and per bin uncertainties are shown for the burnsample unfolding with combined MC. The unfolding bias is below 10 % for all models in all bins. The regularization impact is below 2 %. The per bin uncertainties are dominated by the unfolding bias and the statistical uncertainty.



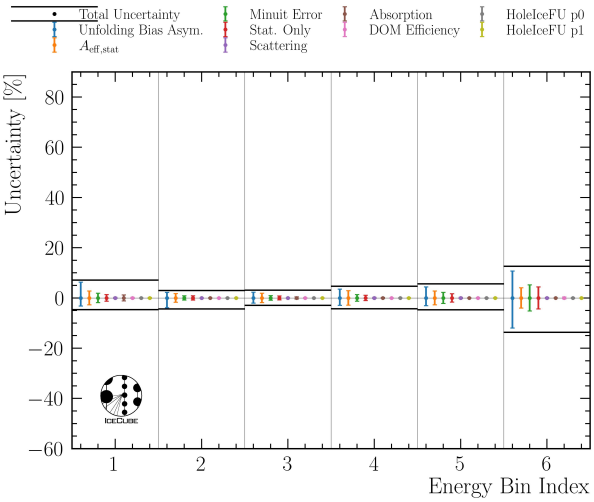
(a) Global correlation for burnsample with combined MC for matrix and test set weighted to H3a.



(b) Unfolding bias for burnsample with combined MC for all models.



(c) Regularization impact on burnsample with combined MC for matrix weighted to GSF.

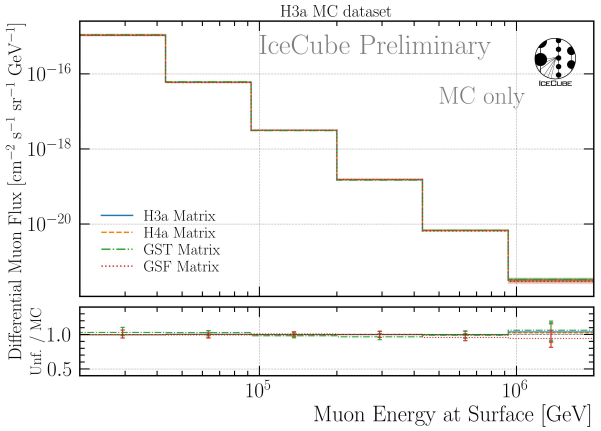


(d) Per bin uncertainties for burnsample with combined MC.

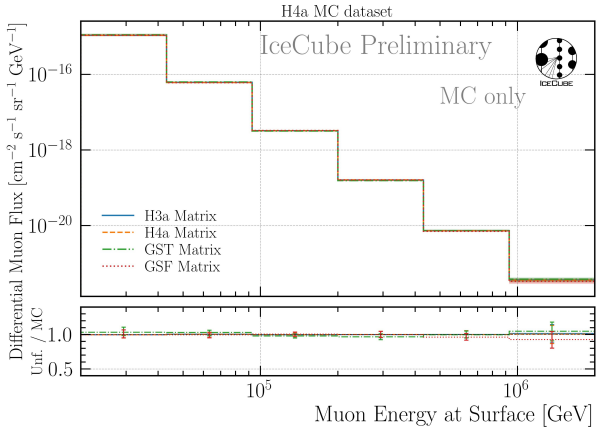
Figure E.2: This figure presents (a) the global correlation, (b) the unfolding bias, (c) the regularization impact, and (d) the per bin uncertainties for the burnsample unfolding with combined MC. The unfolding bias is below 10 % for all models in all bins. The regularization impact is below 2 %. The per bin uncertainties are dominated by the unfolding bias and the statistical uncertainty.

Table E.1: This table lists the unfolding bias values $b_{\text{unf,bias}}$ for all 16 combinations of primary models used to build the response matrix and to weight the test MC set. The rows present the maximum downward (bias low), upward (bias up), and symmetric (bias symmetric) deviations in each energy bin of the unfolding result compared to the true injected flux. The symmetric values correspond to the maximum deviation, and they are visualized in Figure E.2b. These values are determined for the burnsample unfolding using the combined MC dataset.

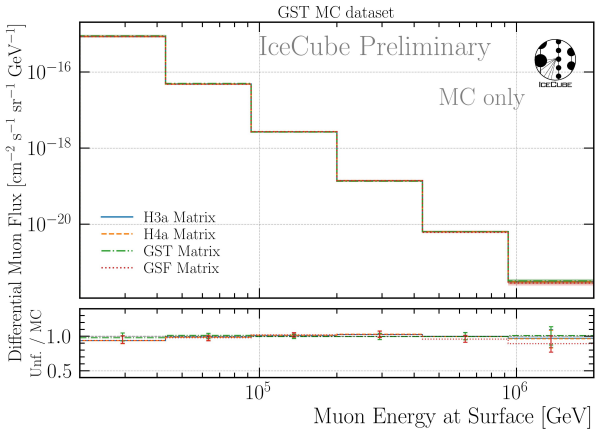
Uncertainty per bin [%]						
Bin	1	2	3	4	5	6
Bias low ↓	3.2	3.9	2.1	2.9	3.1	12.0
Bias up ↑	6.3	2.2	2.3	3.5	4.4	10.7
Bias symmetric ↕	6.3	3.9	2.3	3.5	4.4	12.0



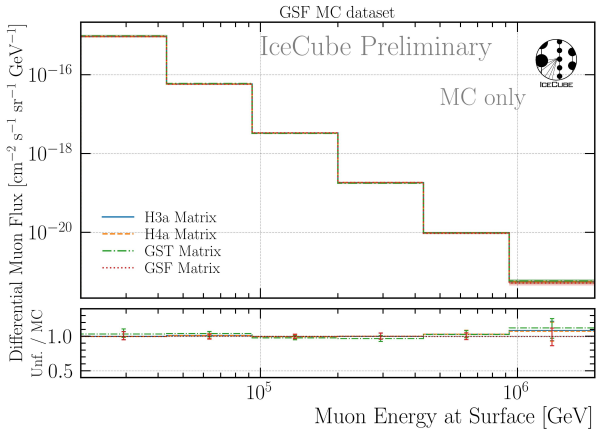
(a) Test set weighted to H3a.



(b) Test set weighted to H4a.

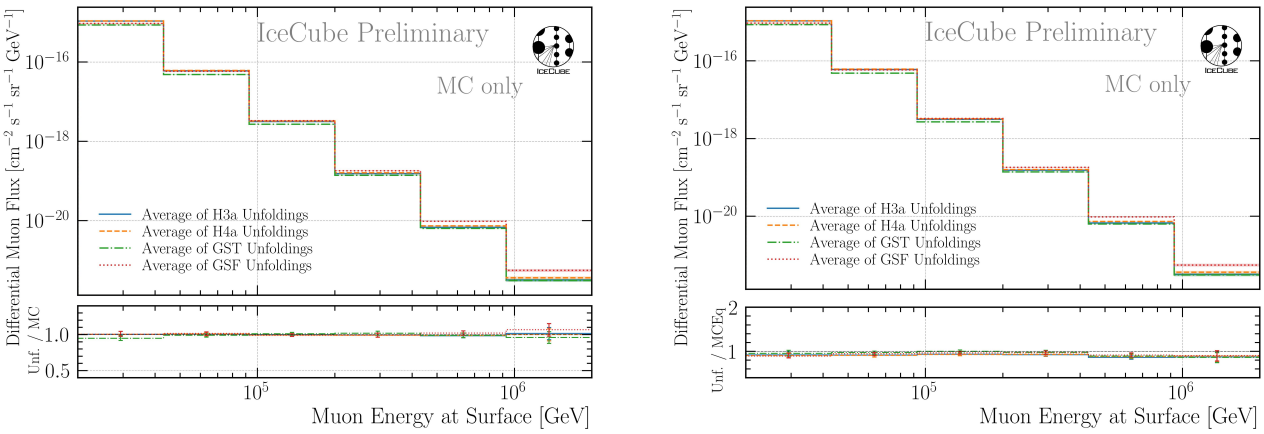


(c) Test set weighted to GST.



(d) Test set weighted to GSF.

Figure E.3: This figure presents the unfolding results for burnsample with combined MC for different test set weightings. The unfolding is performed 16 times in total. For example, in (a), a test set weighted to H3a is unfolded with a response matrix weighted to the four models. In the other three panels, also one test set is unfolded four times with the four different models. The ratios show the unfolded flux divided by the true flux for each model, and it is expected that all ratios align with one. This is the case within uncertainties for all test set weightings. Hence, the unfolding with combined MC is stable against different test set weightings.



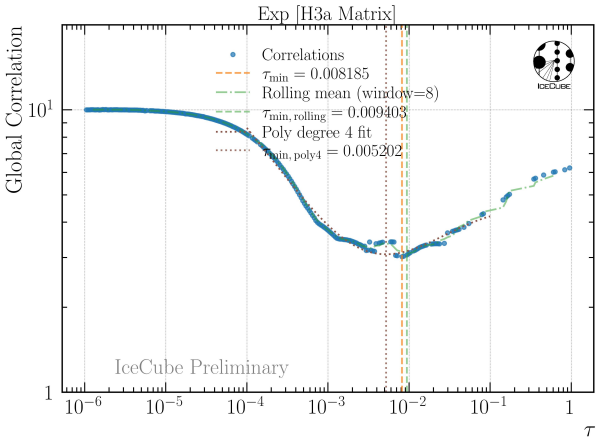
(a) Ratio to CORSIKA.

(b) Ratio to MCEq.

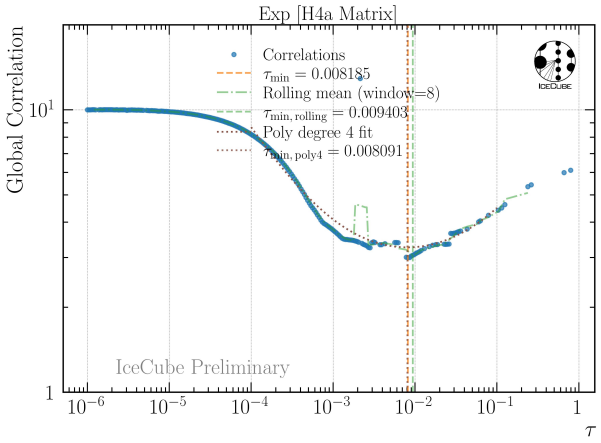
Figure E.4: In this figure, the four unfolding results performed on one test set weighted to the same primary model are averaged. *Average of H3a Unfoldings* means the average of the four unfoldings performed on the H3a test set, for example. In (a), the average unfoldings are compared to the CORSIKA true fluxes, while in (b), they are compared to the MCEq true fluxes. The CORSIKA predictions align well with the unfoldings. The MCEq predictions show small deviations in the highest energy bins, which is caused by the differences between CORSIKA and MCEq.

E.3 Data Unfolding

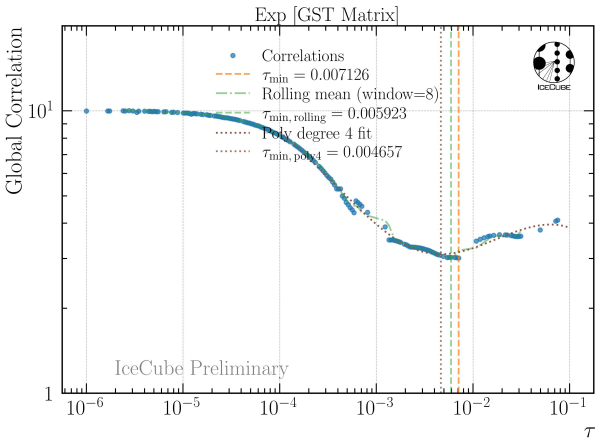
This section presents the unfolding of the burnsample data using the combined MC set. The unfolding is performed for different test set weightings, and the results are compared to the true fluxes of CORSIKA and MCEq. The unfolding results are stable against different test set weightings, and the results are compatible with the final unfolding results presented in Section 8.3.9.



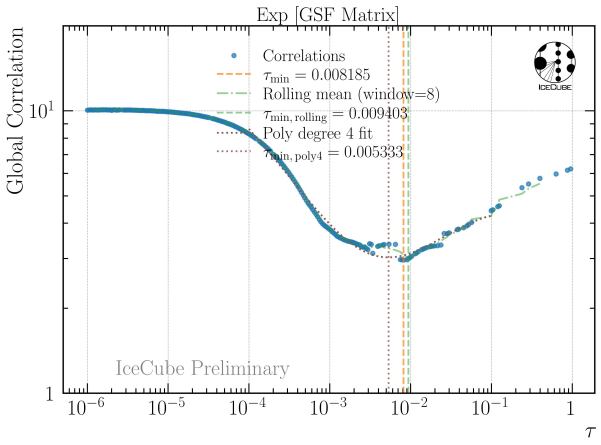
(a) H3a matrix.



(b) H4a matrix.

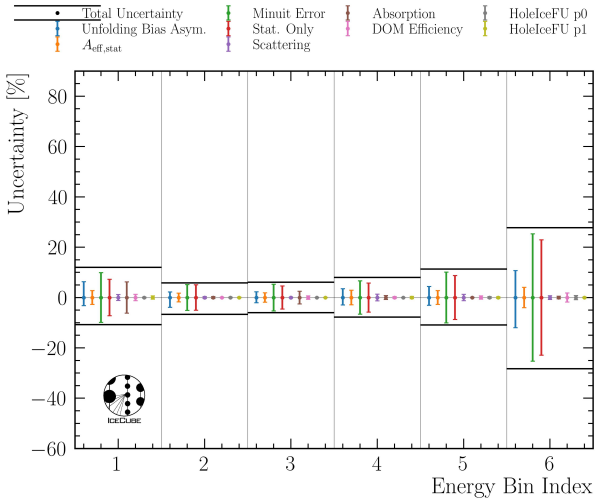


(c) GST matrix.

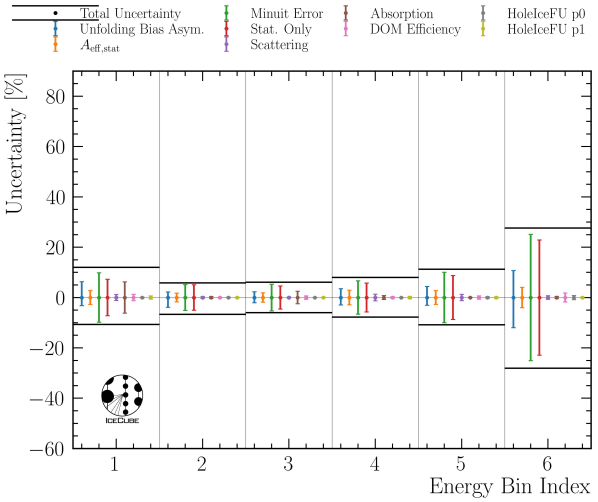


(d) GSF matrix.

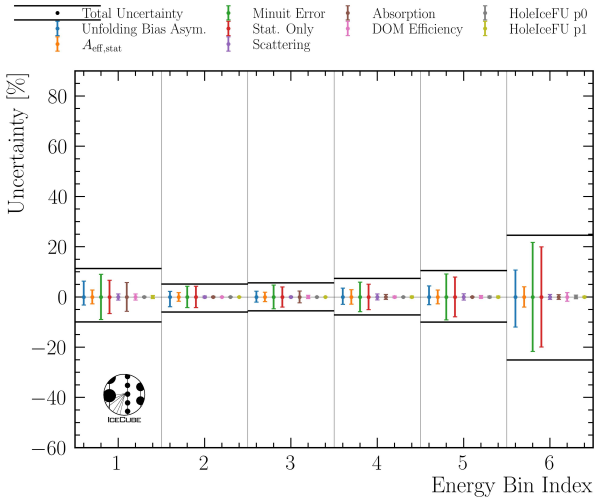
Figure E.5: This figure presents the global correlation for the data unfolding with combined MC for the four different response matrices. All four correlation profiles show a similar behavior, with a minimum around $\tau = 0.01$. The minimum is not smooth; however, given the small impact of regularization strength changes as presented in Figure E.2c, the choice of the optimal regularization strength does not significantly affect the unfolding result.



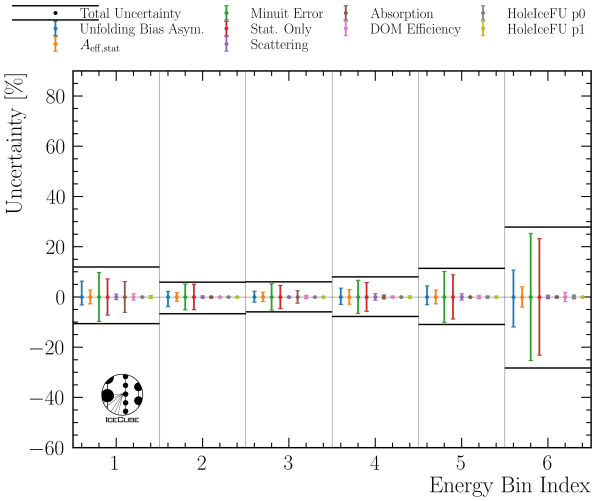
(a) H3a matrix.



(b) H4a matrix.

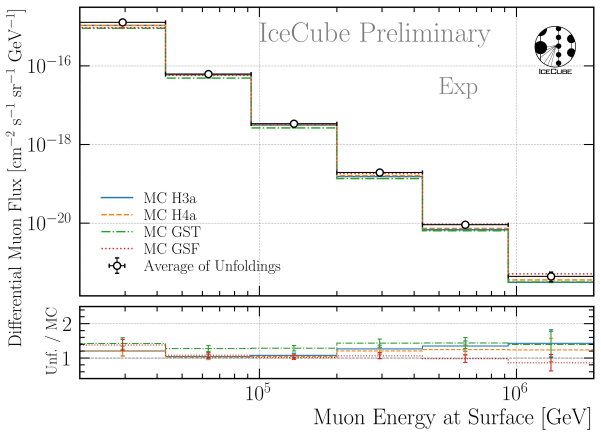


(c) GST matrix.

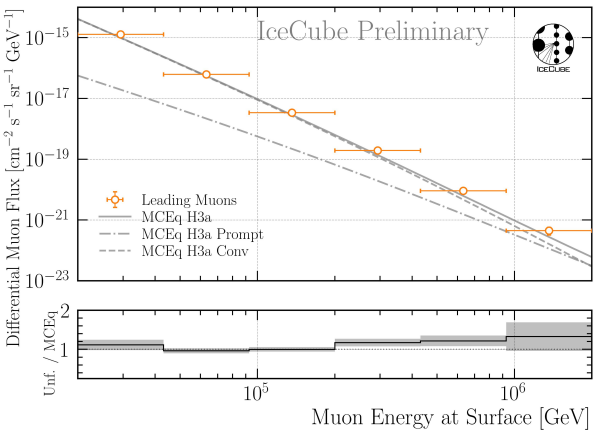


(d) GSF matrix.

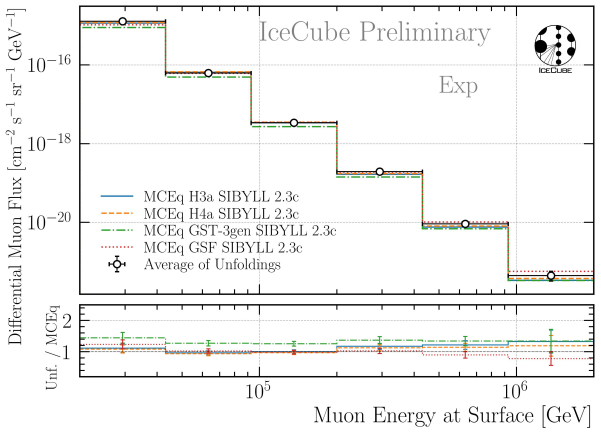
Figure E.6: This figure presents the per bin uncertainties for the data unfolding with the combined MC for the four different response matrices. The uncertainties are similar for all four matrices, with statistical uncertainty dominating, as expected for the small burnsample.



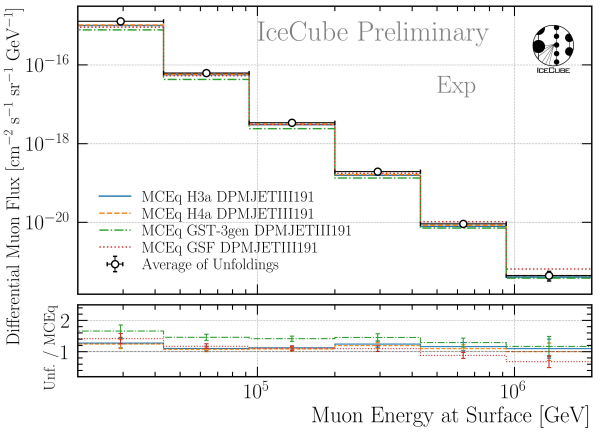
(a) Average burnsample data unfolding compared to CORSIKA.



(b) Average burnsample data unfolding compared to MCEq H3a with SIBYLL 2.3c.



(c) Average burnsample data unfolding compared to MCEq SIBYLL 2.3c.



(d) Average burnsample data unfolding compared to MCEq DPMJET-III 19.1.

Figure E.7: This figure presents the average burnsample data unfoldings with combined MC compared to different predictions. In (a), the average unfolding is compared to the CORSIKA predictions, while in (b), it is compared to the MCEq H3a prediction with SIBYLL 2.3c. In (c) and (d), the average unfolding is compared to the MCEq predictions for all four primary models with SIBYLL 2.3c and DPMJET-III 19.1, respectively. To check whether the combined MC unfolding has an influence on the burnsample unfolding result, panel (b) can be compared to Figure 8.11. Within the uncertainties, the results are compatible.

F Full Sample Unfolding

This chapter contains supporting material for the unfolding of the full data sample of 12.12 years of IceCube data. It starts with the determination of the proxy binning in Section F.1. The regularization scan is shown in Section F.2. Then, the MC unfolding performance is evaluated in Section F.3, followed by the experimental data unfolding in Section F.4. Additionally, flux normalization fits (Section F.5), a comparison between CORSIKA and `MCEq` (Section F.6), and the scaling factors (Section F.7) are presented.

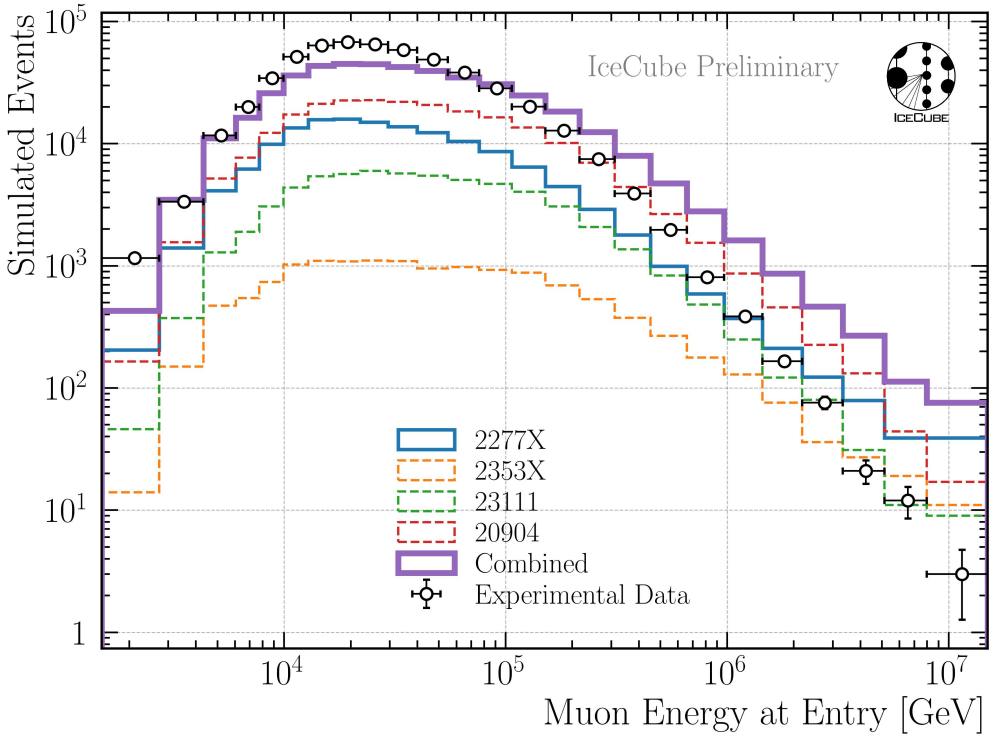
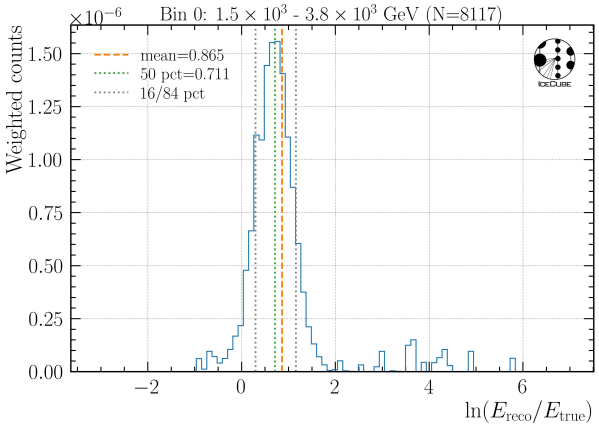


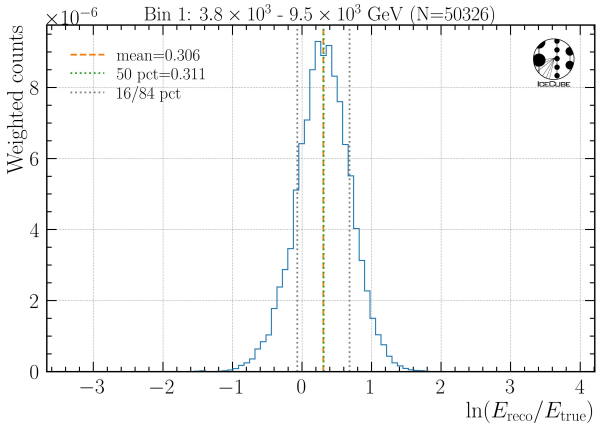
Figure F.1: This figure presents the distribution of the 540774 experimental events for the 12.12 years of IceCube data used for the full sample unfolding. The reconstructed leading muon energy at entry is shown which serves as the proxy for the unfolding. Combining several MC sets helps to increase the statistics to have more simulated events than data events for energies above 100 TeV. For lower energies, more data events than MC events are available. While it is not necessary to have more MC events than data events, it is essential that the simulation covers the entire phase space of physical parameters relevant to the measurement. This includes the full range of incoming directions (zenith and azimuth angles), different primary particle types (from proton to iron nuclei), and the complete energy range. Adequate MC statistics in each bin of the response matrix is crucial for a robust estimation of the detector response and to avoid biases in the unfolding. In total, 449559 events are utilized for the unfolding with the combined MC.

F.1 Proxy Binning

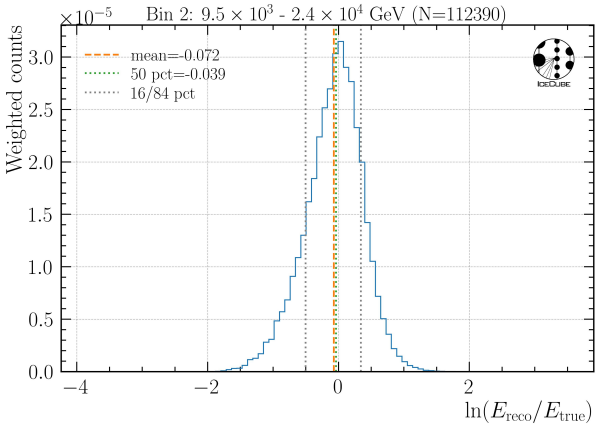
This appendix details the estimation of the energy resolution used to define the proxy binning for the full sample unfolding. The distributions of the logarithmic energy deviation, $\Delta \ln E = \ln(E_{\text{reco}}/E_{\text{true}})$, are shown for the ten logarithmic true energy bins used to sample the resolution function. For each bin, the weighted mean (μ) and standard deviation (σ) are calculated and displayed. These values are used to interpolate the continuous resolution function $\sigma(E_{\text{true}})$ described in Section 8.3.2. Additionally, the 16th, 50th (median), and 84th percentiles are indicated to illustrate the shape of the distribution.



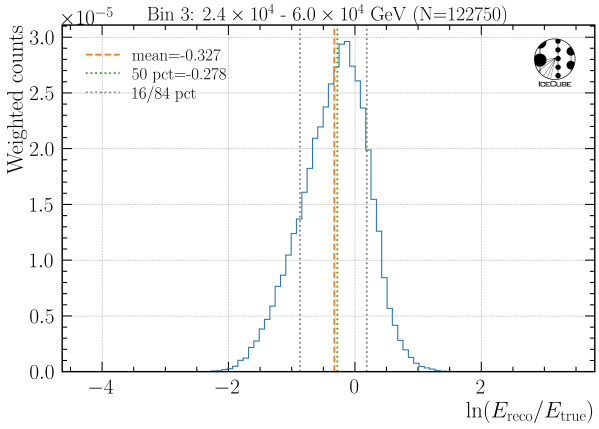
(a) $1.5 \cdot 10^3$ GeV – $3.8 \cdot 10^3$ GeV



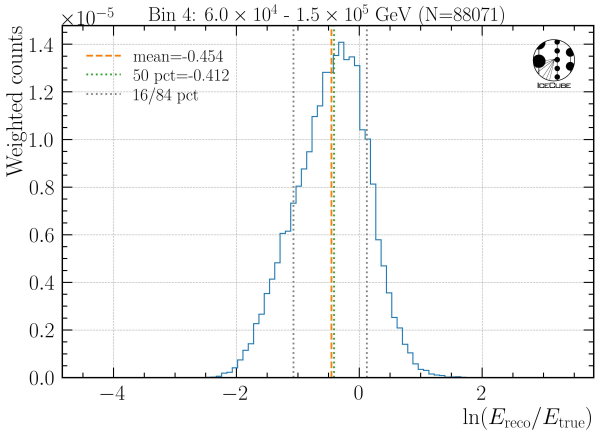
(b) $3.8 \cdot 10^3$ GeV – $9.5 \cdot 10^3$ GeV



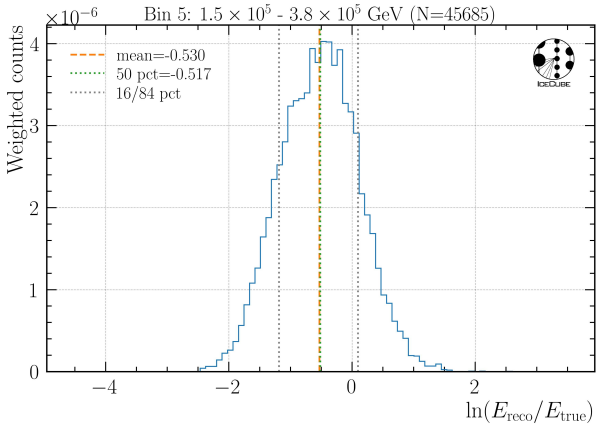
(c) $9.5 \cdot 10^3$ GeV – $2.4 \cdot 10^4$ GeV



(d) $2.4 \cdot 10^4$ GeV – $6.0 \cdot 10^4$ GeV

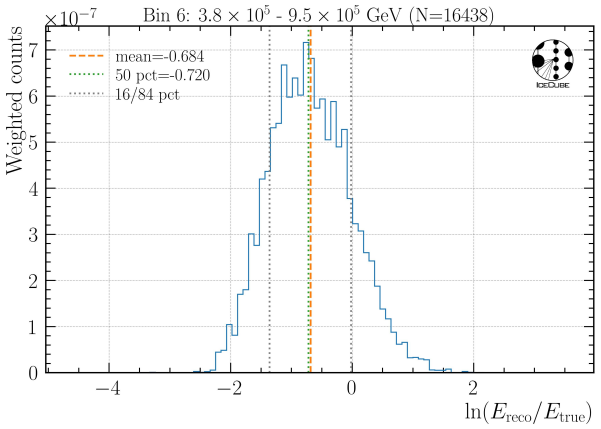


(e) $6.0 \cdot 10^4$ GeV – $1.5 \cdot 10^5$ GeV

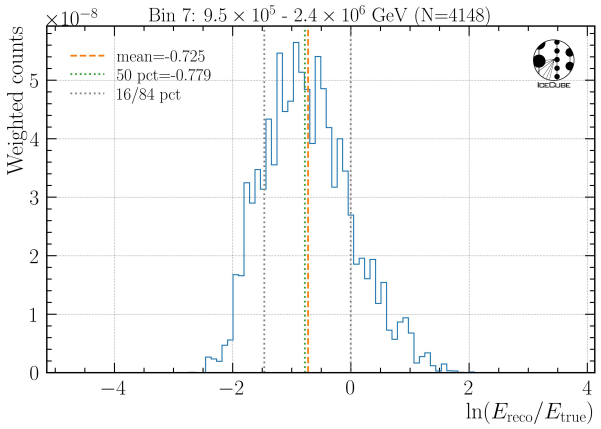


(f) $1.5 \cdot 10^5$ GeV – $3.8 \cdot 10^5$ GeV

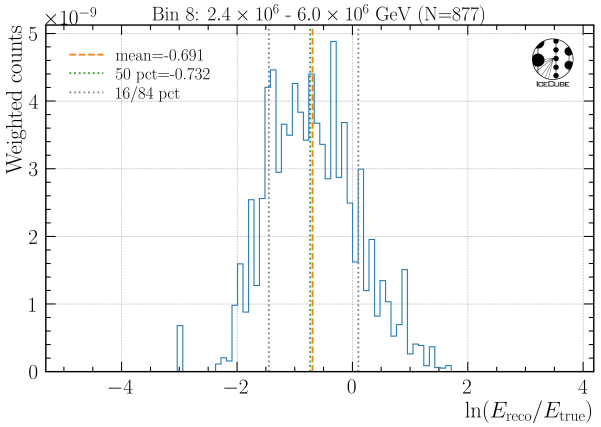
Figure F.2: Distributions of the logarithmic energy deviation $\Delta \ln E$ for the first six true energy bins. The vertical lines indicate the mean (dashed orange), median (dotted green), and the 16th/84th percentiles (dotted gray).



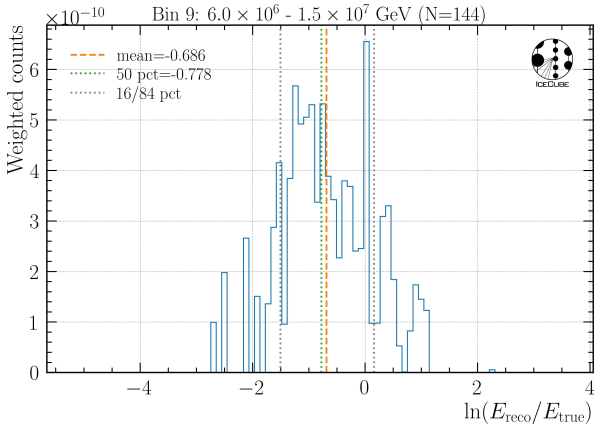
(a) $3.8 \cdot 10^5$ GeV – $9.5 \cdot 10^5$ GeV



(b) $9.5 \cdot 10^5$ GeV – $2.4 \cdot 10^6$ GeV



(c) $2.4 \cdot 10^6$ GeV – $6.0 \cdot 10^6$ GeV

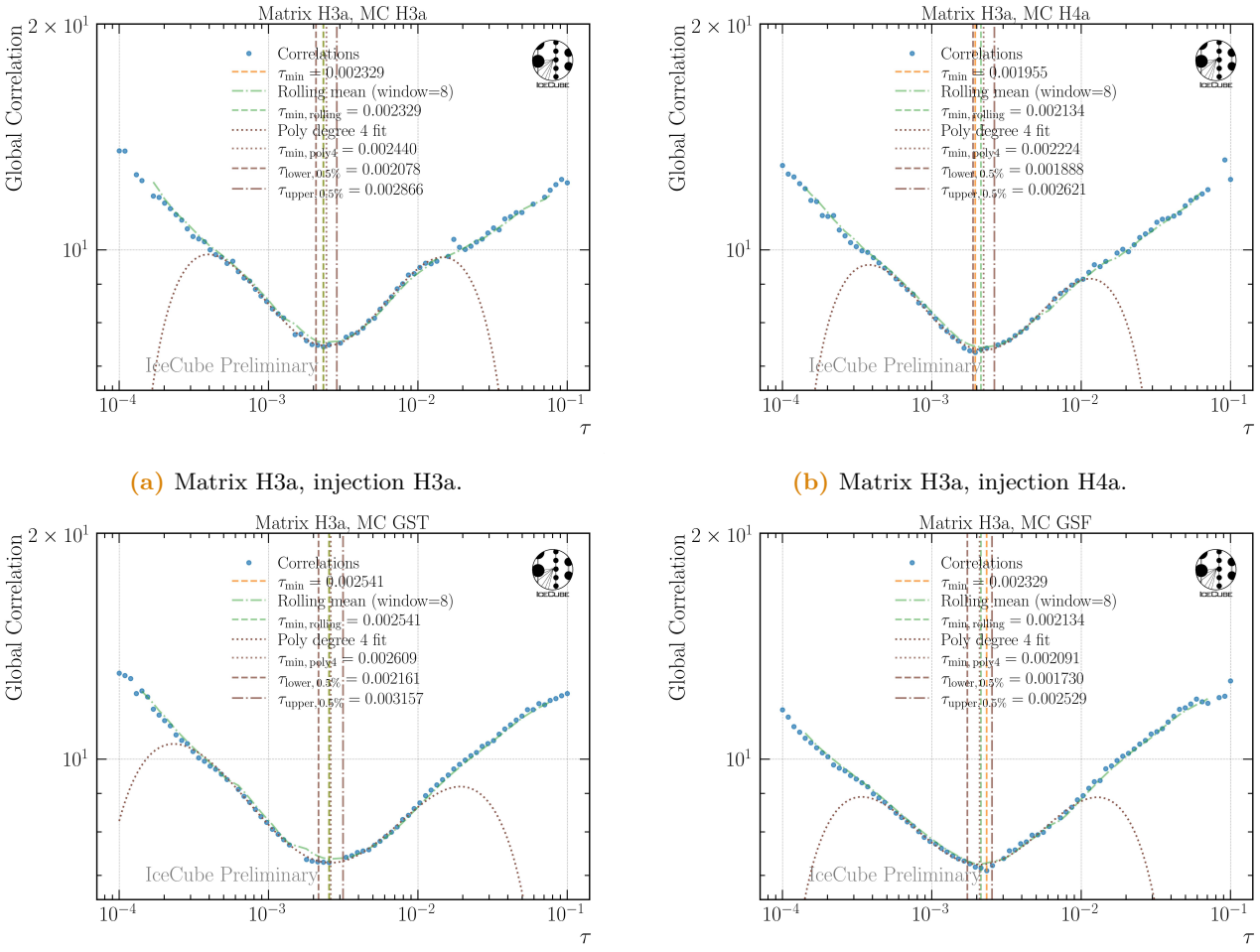


(d) $6.0 \cdot 10^6$ GeV – $1.5 \cdot 10^7$ GeV

Figure F.3: Distributions of the logarithmic energy deviation $\Delta \ln E$ for the last four true energy bins. The vertical lines indicate the mean (dashed orange), median (dotted green), and the 16th/84th percentiles (dotted gray).

F.2 Regularization Scan

Here, the regularization scan for the full sample unfolding is presented as discussed in Section 8.3.5. Overall, the impact of the tested variations is below 4 % for all energy bins, as shown in Figure F.8. Hence, the regularization strength is not a dominant uncertainty in this analysis, and thus neglected.



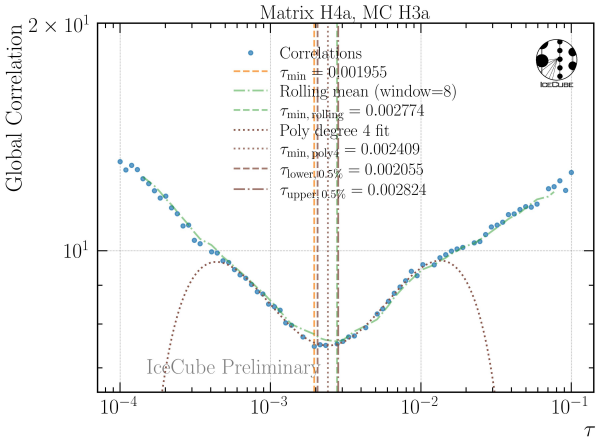
(a) Matrix H3a, injection H3a.

(b) Matrix H3a, injection H4a.

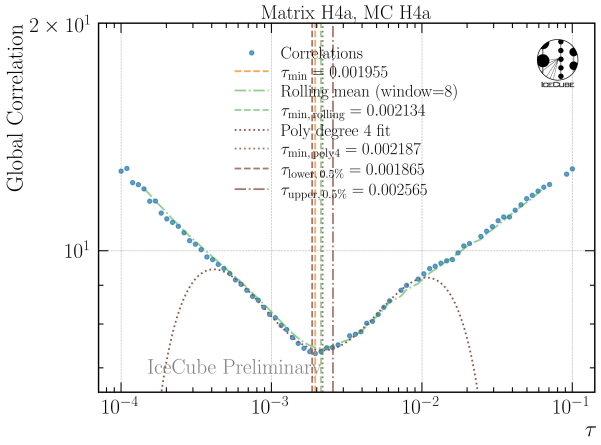
(c) Matrix H3a, injection GST.

(d) Matrix H3a, injection GSF.

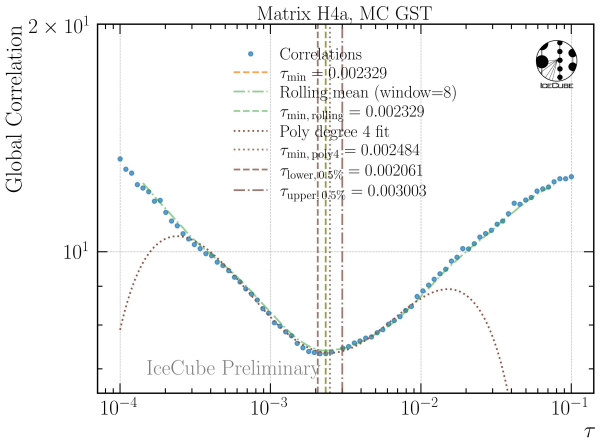
Figure F.4: This figure presents the regularization scan for the full sample unfolding. The four panels show the global correlation as a function of the regularization strength τ for different injection models. The unfolding is trained on the H3a model in all cases. The optimal regularization strength is determined by the minimum of the global correlation. Two test the impact of a slightly different regularization, the lower and upper τ values where the minimum of the global correlation is increased by 0.5 % is determined. These values are indicated by the vertical dashed lines.



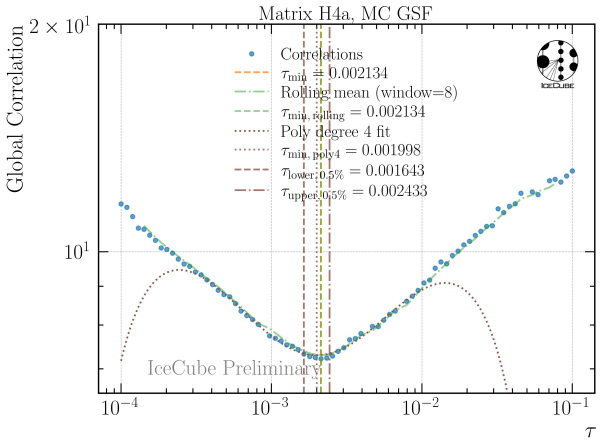
(a) Matrix H4a, injection H3a.



(b) Matrix H4a, injection H4a.

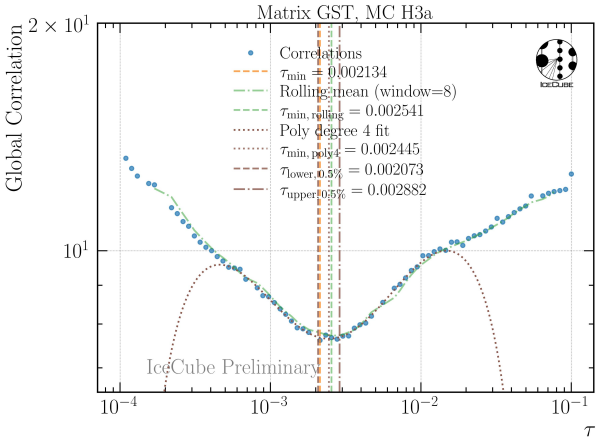


(c) Matrix H4a, injection GST.

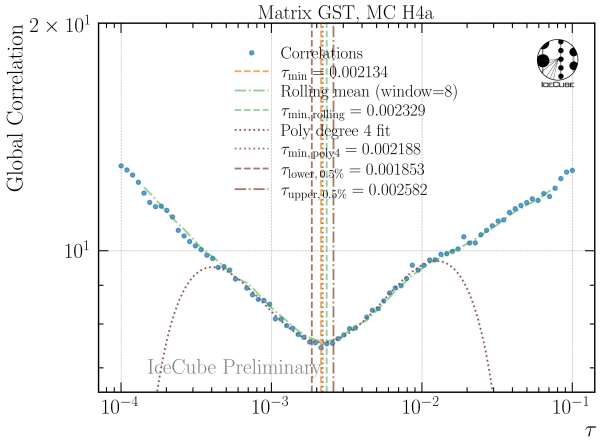


(d) Matrix H4a, injection GSF.

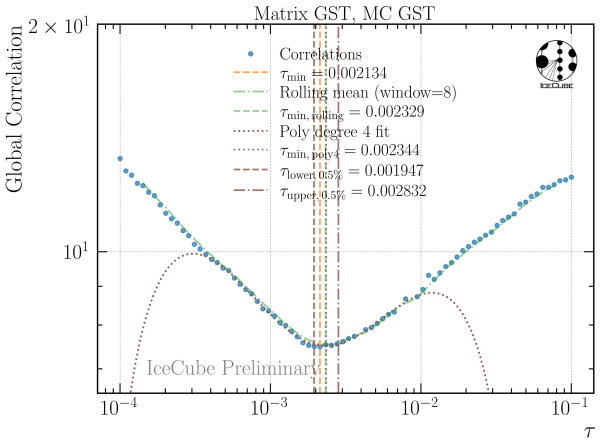
Figure F.5: This figure presents the regularization scan for the full sample unfolding. The four panels show the global correlation as a function of the regularization strength τ for different injection models. The unfolding is trained on the H4a model in all cases. The optimal regularization strength is determined by the minimum of the global correlation. Two test the impact of a slightly different regularization, the lower and upper τ values where the minimum of the global correlation is increased by 0.5 % is determined. These values are indicated by the vertical dashed lines.



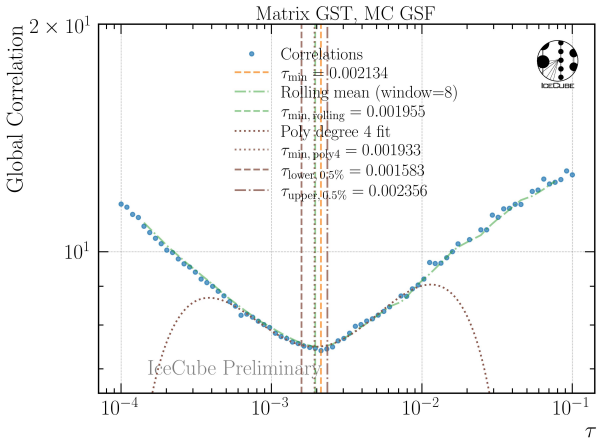
(a) Matrix GST, injection H3a.



(b) Matrix GST, injection H4a.

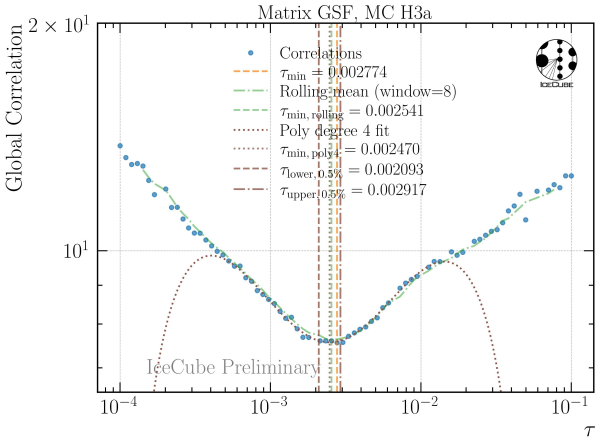


(c) Matrix GST, injection GST.

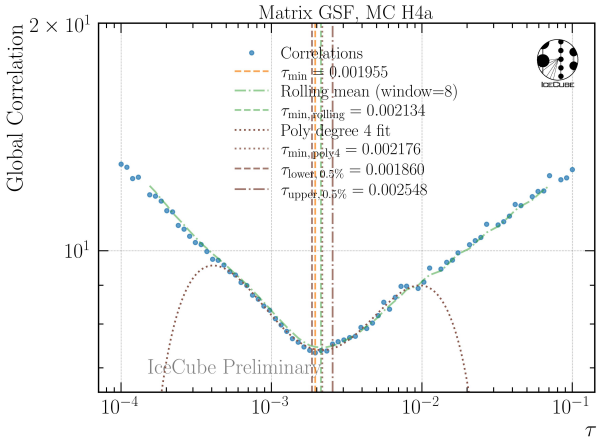


(d) Matrix GST, injection GSF.

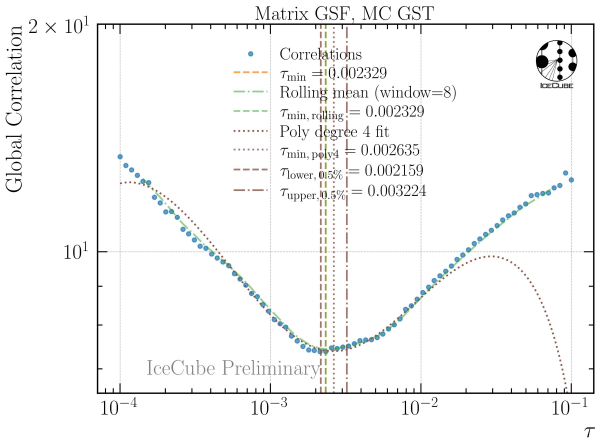
Figure F.6: This figure presents the regularization scan for the full sample unfolding. The four panels show the global correlation as a function of the regularization strength τ for different injection models. The unfolding is trained on the GST model in all cases. The optimal regularization strength is determined by the minimum of the global correlation. Two test the impact of a slightly different regularization, the lower and upper τ values where the minimum of the global correlation is increased by 0.5 % is determined. These values are indicated by the vertical dashed lines.



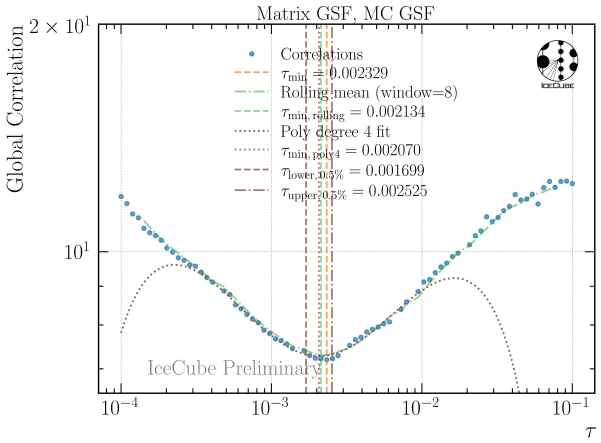
(a) Matrix GSF, injection H3a.



(b) Matrix GSF, injection H4a.

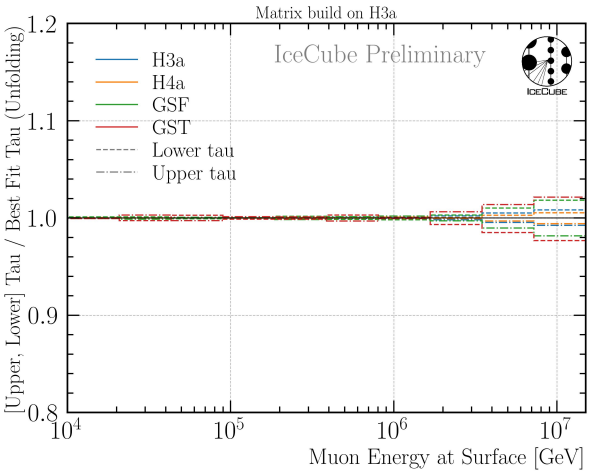


(c) Matrix GSF, injection GST.

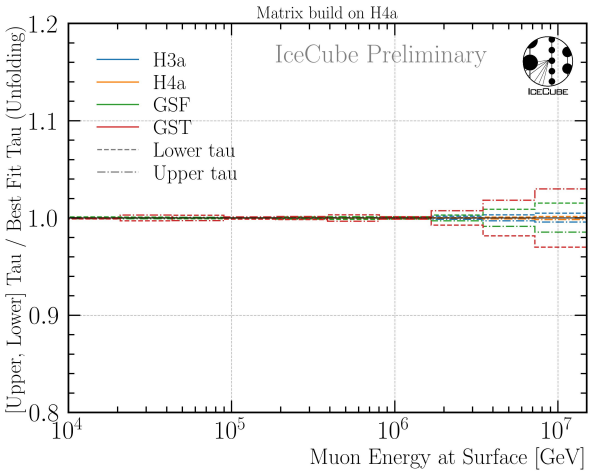


(d) Matrix GSF, injection GSF.

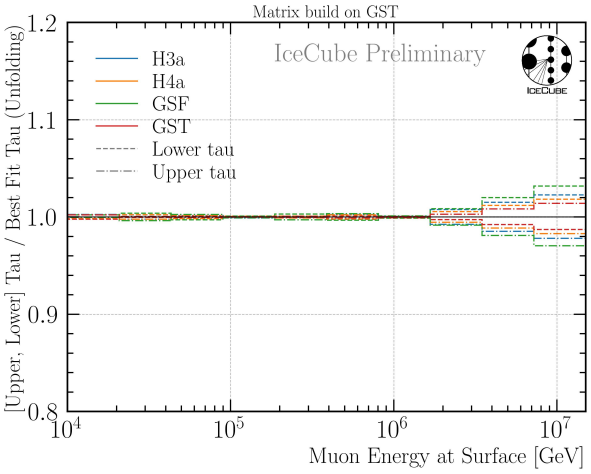
Figure F.7: This figure presents the regularization scan for the full sample unfolding. The four panels show the global correlation as a function of the regularization strength τ for different injection models. The unfolding is trained on the GSF model in all cases. The optimal regularization strength is determined by the minimum of the global correlation. Two test the impact of a slightly different regularization, the lower and upper τ values where the minimum of the global correlation is increased by 0.5 % is determined. These values are indicated by the vertical dashed lines.



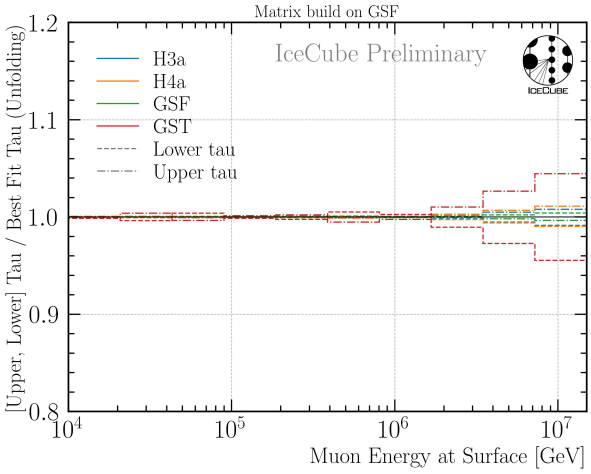
(a) Matrix H3a.



(b) Matrix H4a.



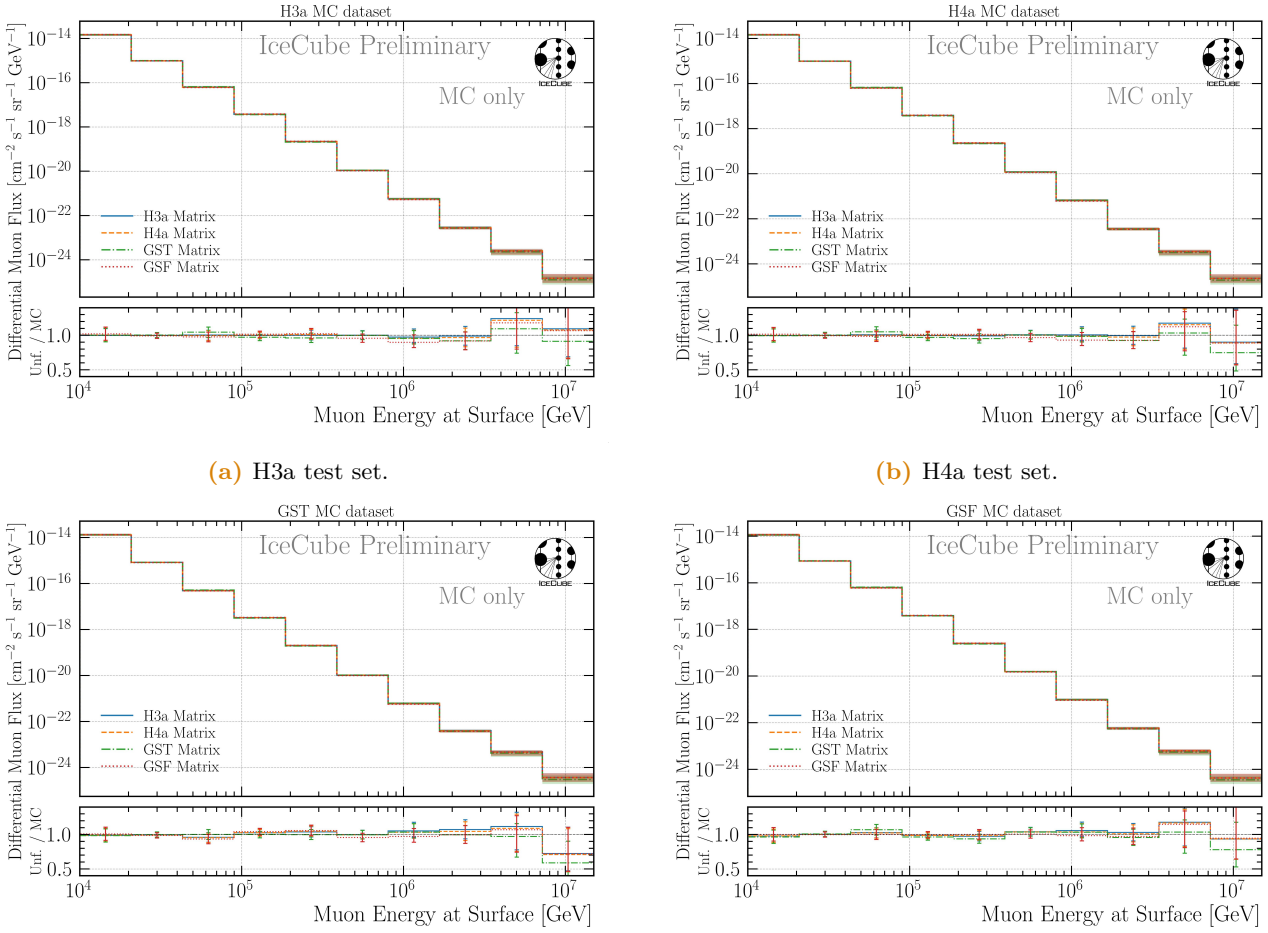
(c) Matrix GST.



(d) Matrix GSF.

Figure F.8: This figure presents the unfolding results for different regularization strengths τ for the full sample unfolding. Each panel shows the unfolding result using a different response matrix. For the variations in τ, the lower and upper values are compared to the τ at the minimum global correlation, as determined in the figures above. For all variations, the impact on the unfolding results is below 4 %.

F.3 MC Unfolding



(a) H3a test set.

(b) H4a test set.

(c) GST test set.

(d) GSF test set.

Figure F.9: This figure shows the unfolding results on MC for all four primary models used to build the response matrix and to weight the test MC set. The unfolded fluxes are compared to the true injected fluxes. The errorbars present the total uncertainty, including statistical, effective area, ice systematics, and unfolding bias uncertainties. Within the uncertainties, the unfolding procedure is able to recover the true injected flux for all model combinations.

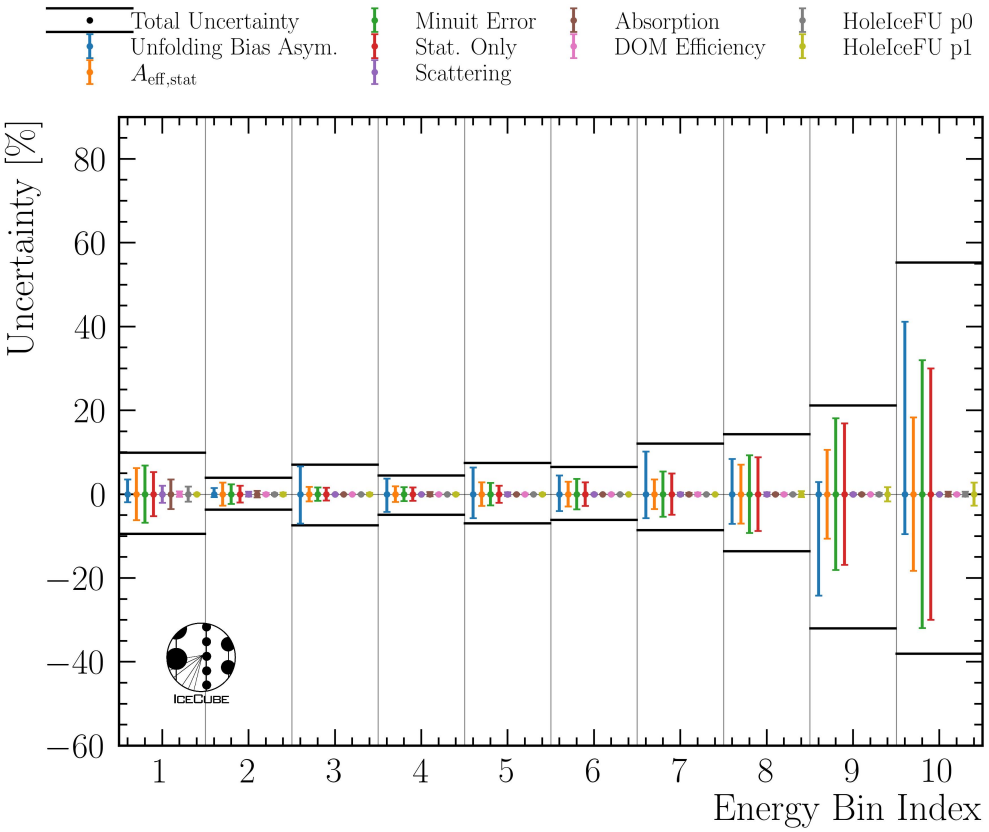
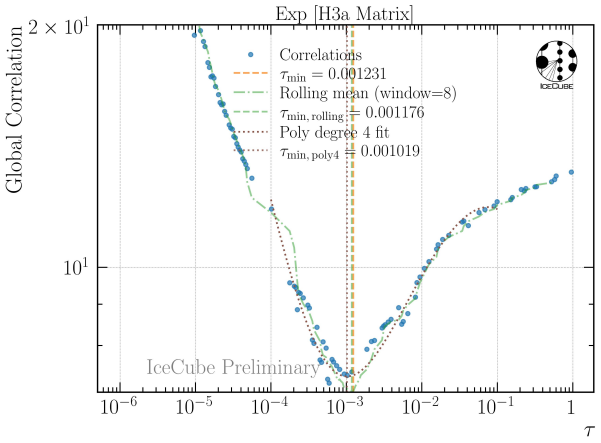
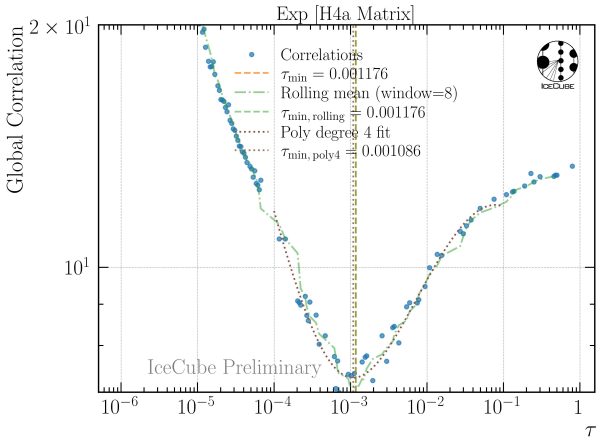


Figure F.10: This figure shows the per bin uncertainties for the full sample unfolding on MC for a matrix build with the combined MC weighted to H3a to unfold a test set also weighted to H3a. The different uncertainty contributions are shown, including statistical (red), effective area (orange), ice systematics (purple, brown, pink, grey and light green), and unfolding bias (blue) uncertainties. The total uncertainty is shown in black. The unfolding bias is the dominant uncertainty contribution in most energy bins, followed by the statistical uncertainties.

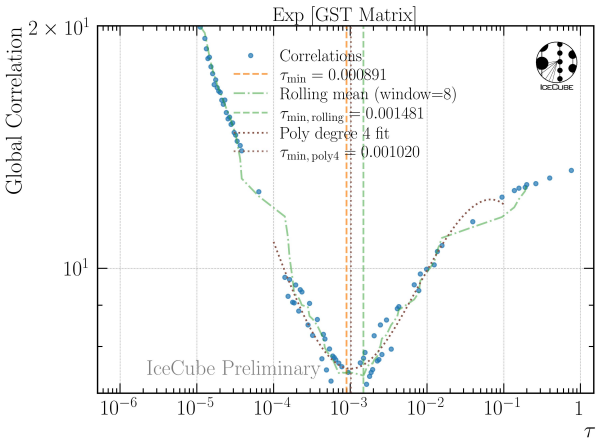
F.4 Experimental Data Unfolding



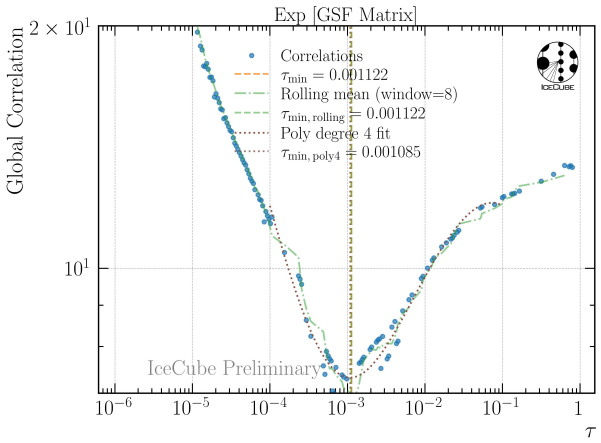
(a) H3a matrix.



(b) H4a matrix.

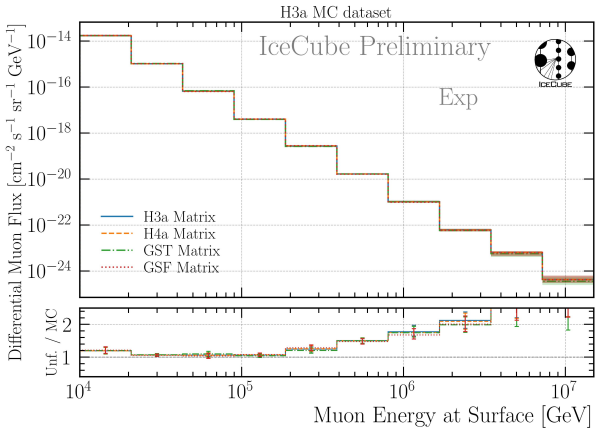


(c) GST matrix.

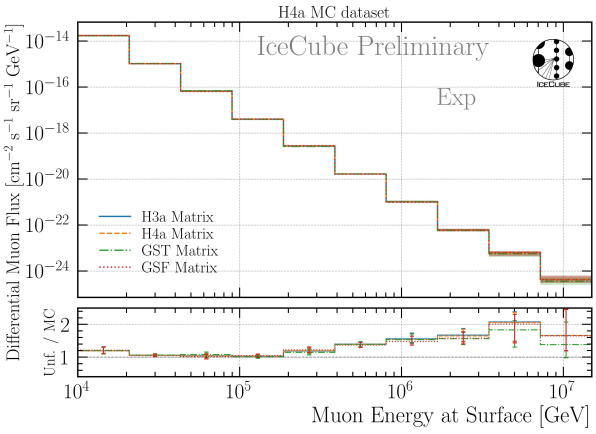


(d) GSF matrix.

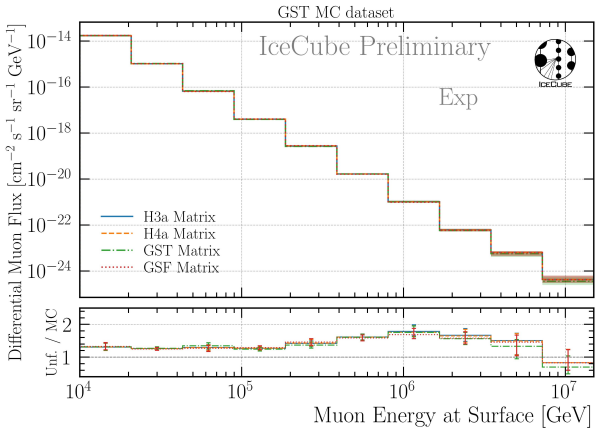
Figure F.11: This figures shows the global correlations for the full sample unfolding on experimental data for all four primary models used to build the response matrix. The global correlation profiles are used to determine the optimal regularization strength for each unfolding. The different profiles show a similar behavior, with a minimum around $\tau \approx 0.001$ for all four primary model unfoldings.



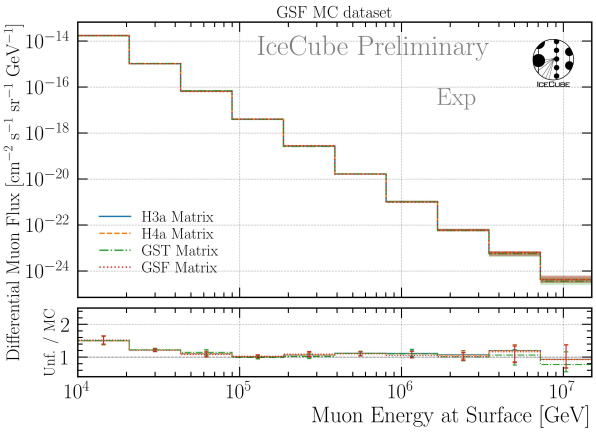
(a) Compare to H3a true CORSIKA.



(b) Compare to H4a true CORSIKA.

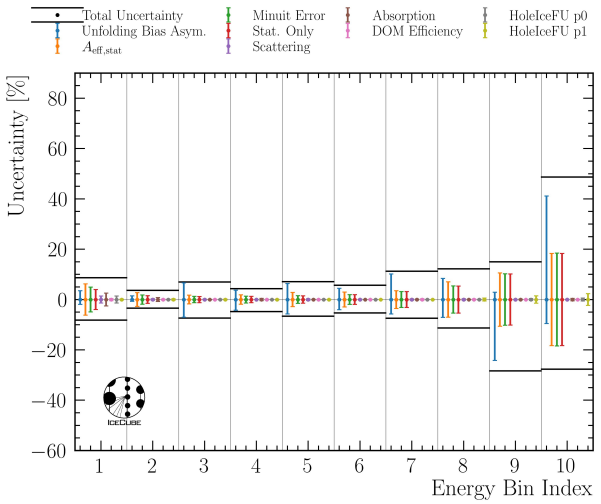


(c) Compare to GST true CORSIKA.

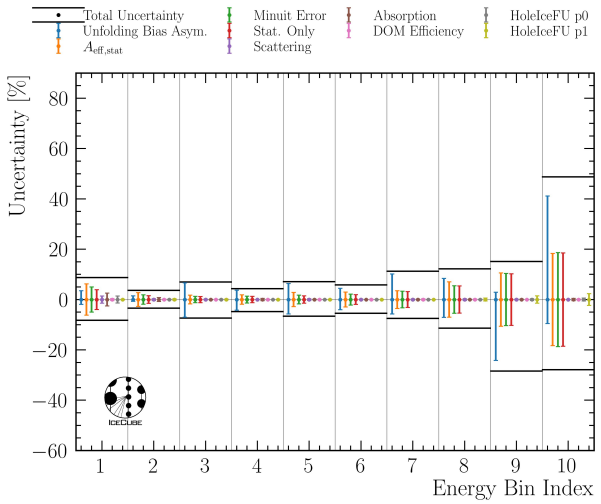


(d) Compare to GSF true CORSIKA.

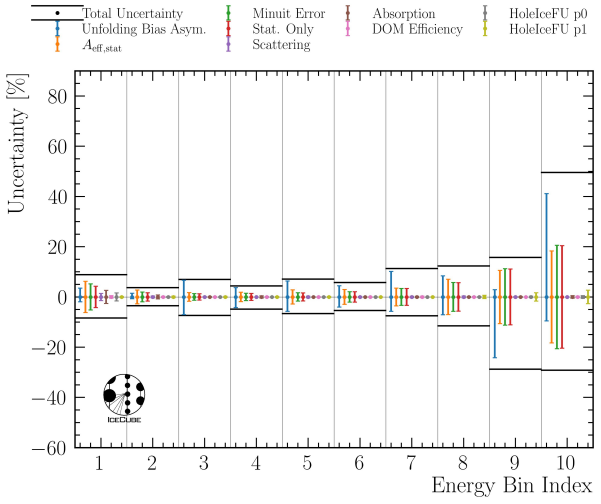
Figure F.12: This figure shows the comparison of the unfolding results on experimental data for all four primary models used to build the response matrix. The unfolded fluxes are compared to the true injected fluxes from CORSIKA simulations. Within the uncertainties, all four unfoldings lead to a similar result. In the low energy region, the unfolded flux is higher than the four CORSIKA predictions, while in the high energy region, the unfolded flux is in agreement with GSF.



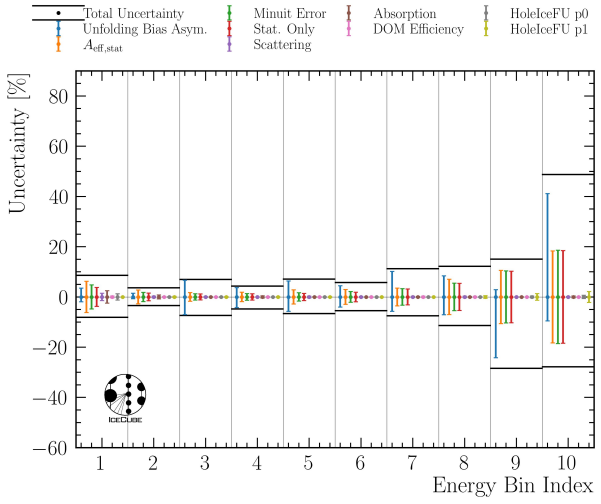
(a) H3a matrix.



(b) H4a matrix.

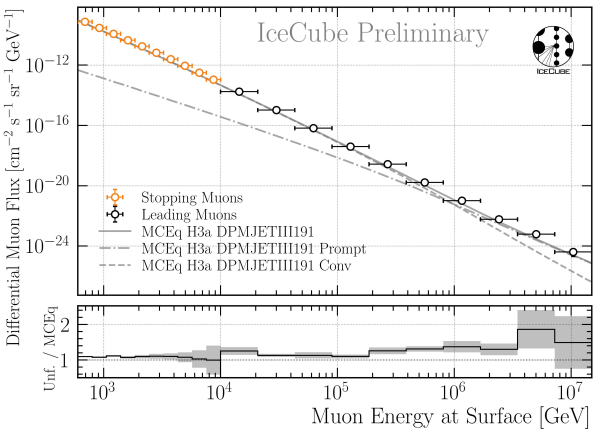


(c) GST matrix.

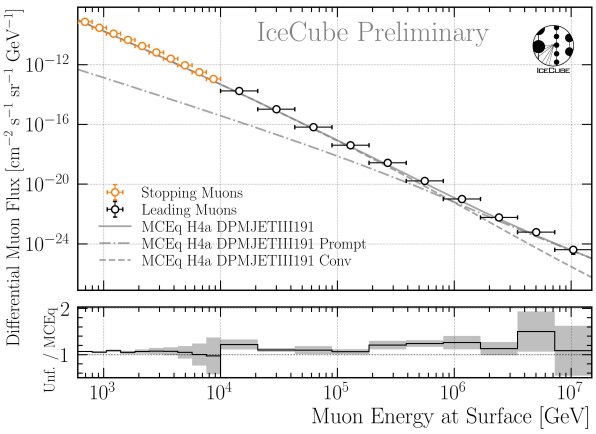


(d) GSF matrix.

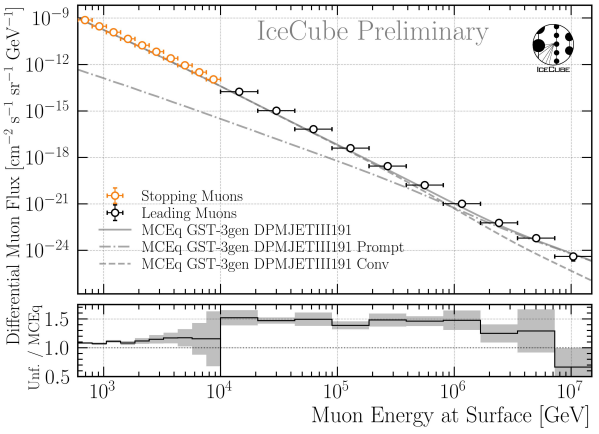
Figure F.13: This figures shows the per bin uncertainties for the full sample unfolding on experimental data for all four primary models used to build the response matrix. The different colors present the different uncertainty contributions: statistical (red), effective area (orange), ice systematics (purple, brown, pink, grey and light green), and unfolding bias (blue). For all four primary model unfoldings, the uncertainty per bin is similar. Overall, the unfolding bias dominates the total uncertainty.



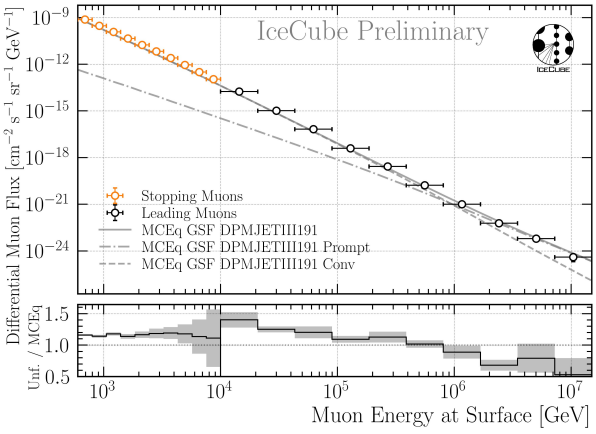
(a) `MCEq` DPMJET-III 19.1 with H3a.



(b) `MCEq` DPMJET-III 19.1 with H4a.

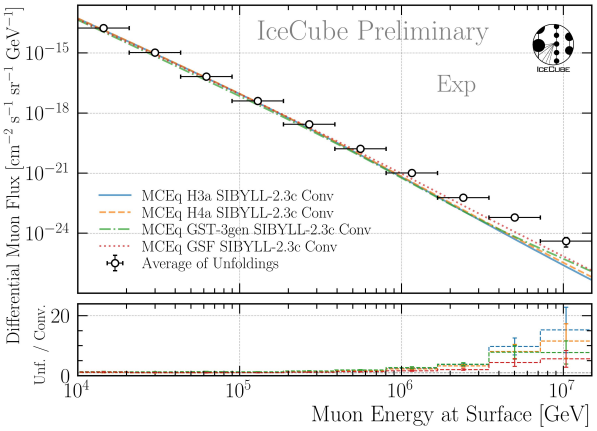


(c) `MCEq` DPMJET-III 19.1 with GST.

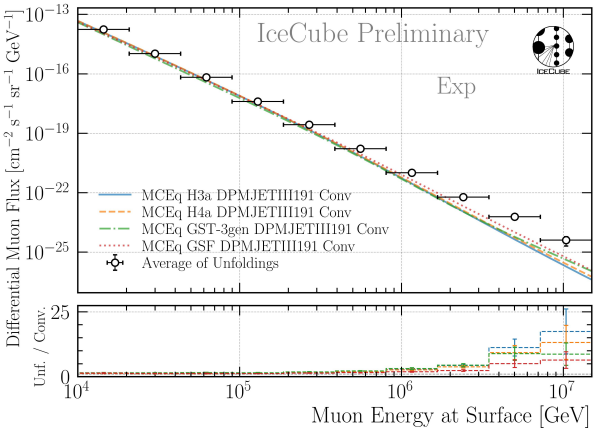


(d) `MCEq` DPMJET-III 19.1 with GSF.

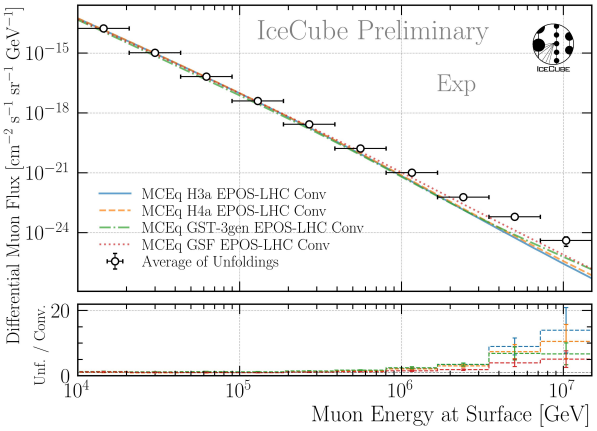
Figure F.14: This figure presents the final results of the unfolding of the atmospheric muon flux at surface. The four unfoldings on different primary models are averaged with propagated uncertainties as described in Paragraph 8.3.8. Additionally, the burnsample unfolding of the stopping muon sample is included for the low-energy region. While the leading muon sample is averaged for zenith angles from $0.2 < \cos(\theta) < 1$, the stopping muon sample is averaged from $0 < \cos(\theta) < 1$. The results are compared to `MCEq` calculations assuming DPMJET-III 19.1 as HIM. The top panel includes the `MCEq` prediction for leading muons, while the bottom panel presents the ratio for the stopping and muon zenith range, respectively. The fluxes agree within the uncertainties at 10 TeV, however, the nominal value of the leading muon flux is higher than the stopping muon flux.



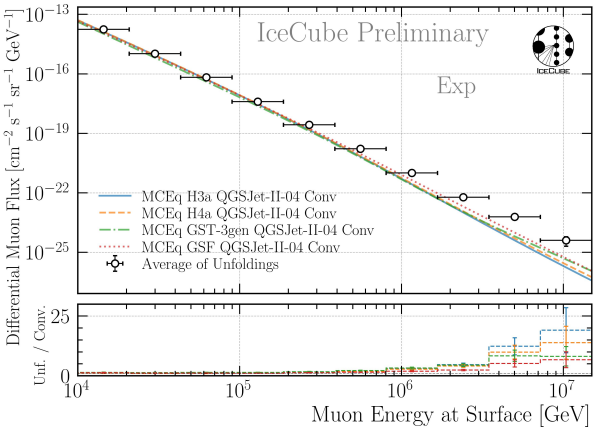
(a) `MCEq` with SIBYLL 2.3c conventional only.



(b) `MCEq` with DPMJET-III 19.1 conventional only.



(c) `MCEq` with EPOS-LHC conventional only.



(d) `MCEq` with QGSJET-II-04 conventional only.

Figure F.15: This figure presents the final results of the unfolding of the atmospheric muon flux at surface. The four unfoldings on different primary models are averaged with propagated uncertainties as described in Paragraph 8.3.8. The results are compared to `MCEq` calculations assuming the four HIM SIBYLL 2.3c, DPMJET-III 19.1, EPOS-LHC and QGSJet-II-04 four all four primary models. Only the conventional muon flux of the `MCEq` calculations is shown here. The ratio panel in each plot shows that the conventional component alone is not sufficient to describe the measured flux at high energies, indicating a significant contribution of prompt muons. The same predictions are also shown in Figure 8.27, where each plot shows all four HIM for a single primary model.

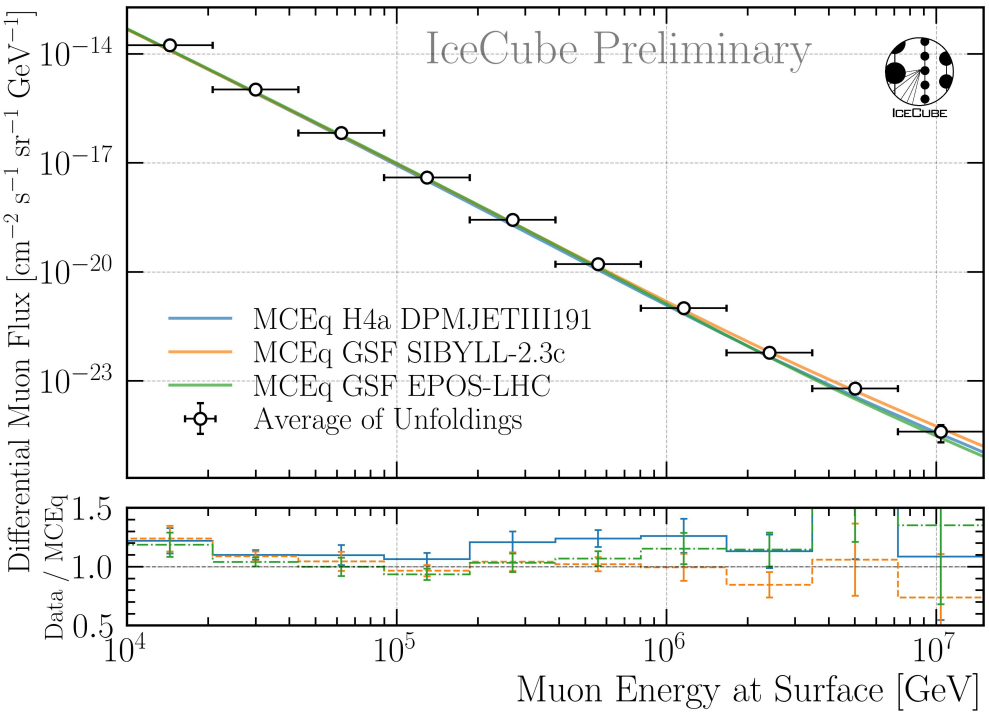


Figure F.16: This figure shows the comparison of the unfolded flux to `MCEq` predictions for the three best-fit combinations of HIM and hadronic interaction model based on the χ^2 test described in Section 8.3.9. The predictions are shown in the top panel, while the ratio to the unfolded flux is shown in the bottom panel. EPOS-LHC does not include a charm production component.

F.4.1 Data-MC Pre and Post Fit

In this section, the data–MC comparisons for the leading muon energy at entry distribution are shown for the full sample unfolding. The pre-fit comparison is shown in Figure F.17. The data distribution is compared to the combined MC distribution weighted to the four primary models. Since the unfolding procedure incorporates the impact of the ice and detector systematics, the MC distribution can be re-weighted to the best-fit systematics. This can be done in different ways. First of all, the re-weighting described in Section 8.2.6 is employed, and presented in Figure F.18a. The distribution is based on the full sample unfolding utilizing the H3a weighting, and the flux is also weighted to H3a. In principle, the distribution should be re-weighted to all 4 primary models. However, to avoid an overload, in the following, only the re-weighted distributions for the weighting used for the unfolding are shown. The re-weighted distribution shown in orange does not match the data distribution at, since it is on oscillating curve that does not reflect the true distribution anymore. This is caused by the fact that the uncertainty on the best-fit systematics is estimated to be small. Since the MC dataset is finite and some events have larger weights than others, the re-weighting leads to large fluctuations. The increased uncertainties on the distribution are another indicator that the re-weighted distribution is unrealistic. To test the impact of the small uncertainties, the re-weighting is also performed where the uncertainty is increased by a factor of 10, which is shown in green. This leads to a smoother distribution that matches the data distribution better.

In addition to the Gaussian re-weighting, also a linear and spectral re-weighting is performed. For the linear approach, each bin in the proxy distribution (leading muon energy at entry) is parameterized by a linear function for each of the five ice and detector systematics, as it is described in Section 8.3.3. These linear functions describe how the bin content changes as a function of the systematic parameter. The parameterization is obtained by evaluating the bin content at multiple values of each systematic parameter and fitting a linear function to the resulting ratio relative to a baseline value. For each systematic s_k ($k = 1, \dots, 5$), the correction factor for bin i is given by

$$f_i^{(k)}(s_k) = a_i^{(k)} + b_i^{(k)} \cdot s_k \,, \tag{F.1}$$

where $a_i^{(k)}$ and $b_i^{(k)}$ are the fitted linear coefficients for that bin and systematic. The total correction factor is obtained by multiplying the contributions from all five systematics,

$$f_i(\mathbf{s}_{\text{fit}}) = \prod_{k=1}^{5} f_i^{(k)}(s_{k,\text{fit}}) \,, \tag{F.2}$$

where $\mathbf{s}_{\text{fit}} = (s_{1,\text{fit}}, \dots, s_{5,\text{fit}})$ are the best-fit systematic parameters from the unfolding. To preserve the total normalization, an auto-normalization factor is applied:

$$\mathcal{N}_{\text{auto}} = \frac{\sum_i N_i^{\text{MC}}}{\sum_i N_i^{\text{MC}} \cdot f_i(\mathbf{s}_{\text{fit}})} \,. \tag{F.3}$$

The re-weighted bin content is then computed as

$$N_i^{\text{reweighted}} = N_i^{\text{MC}} \cdot f_i(\mathbf{s}_{\text{fit}}) \cdot \mathcal{N}_{\text{auto}} \cdot \mathcal{L} \,, \tag{F.4}$$

where N_i^{MC} is the original MC bin content and $\mathcal{L}$ is the experimental livetime. Statistical uncertainties are propagated as $\sigma_i^{\text{reweighted}} = \sigma_i^{\text{MC}} \cdot |f_i(\mathbf{s}_{\text{fit}})| \cdot \mathcal{N}_{\text{auto}} \cdot \mathcal{L}$. In the figures below, this is referred to as w_lin. In Figure F.18b for H3a, the linear re-weighting leads to a distribution that is similar to the pre-fit distribution without re-weighting. At the lower and higher energies, the agreement is slightly better. However, since the non re-weighted distribution already matches the data distribution well in the region with the highest statistics, the impact of the region with lower statistics is smaller on the fit result.

Table F.1: Best-fit values and 1σ uncertainties of the five detector and ice systematics obtained from the full sample unfolding fit for each of the four individual unfoldings.

Systematic Parameter	Baseline	H3a	H4a	GST	GSF
Scattering	1.00	0.982 ± 0.004	0.984 ± 0.004	0.978 ± 0.004	0.991 ± 0.004
Absorption	1.00	1.012 ± 0.003	1.014 ± 0.003	1.006 ± 0.003	1.020 ± 0.003
DOM efficiency	1.00	1.016 ± 0.005	1.015 ± 0.005	1.019 ± 0.005	1.009 ± 0.004
HoleIce p_0	0.20	0.235 ± 0.018	0.227 ± 0.019	0.262 ± 0.017	0.179 ± 0.020
HoleIce p_1	-0.05	-0.053 ± 0.005	-0.050 ± 0.005	-0.060 ± 0.005	-0.040 ± 0.004

For the spectral re-weighting, a different approach is used that does not rely on the ice and detector systematics. Instead, the MC is re-weighted to directly match the unfolded flux result from the data. The unfolded result $\hat{f}_i$ represents the number of events per bin in the true energy space. The re-weighting ratio for each truth-level bin i is computed as

$$r_i = \frac{\hat{f}_i / \mathcal{L}}{N_i^{\mathrm{MC,truth}}}\,, \tag{F.5}$$

where $N_i^{\mathrm{MC,truth}}$ is the MC event count in truth bin i computed using the original MC weights. Each MC event is assigned to its corresponding truth bin based on its true energy value. The event weight is then re-scaled by the bin-specific ratio $w_j^{\mathrm{spec}} = w_j \cdot r_{i(j)}$, where $i(j)$ denotes the truth bin index of event j. This method directly adjusts the MC truth spectrum to match the data-driven unfolded result while preserving event-by-event correlations between truth and reconstruction. In the figures below, this is referred to as w_spec. This is referred to as w_spec in the figures below. In Figure F.18b for H3a, the spectral re-weighting leads to a distribution that matches the data distribution across the entire energy range, including the low and high energy regions with lower statistics, except of the first low energy bin. The good agreement is expected since this re-weighting method directly adjusts the MC truth spectrum to match the data-driven unfolded result. While Figure F.18 shows the post-fit data–MC comparisons for H3a and H4a, the same re-weighting approaches are applied to the unfoldings with the GST and GSF matrices, which are shown in Figure F.19. For all four primary model unfoldings, the Gaussian re-weighting leads to a distribution that does not match the data distribution, while the linear and spectral re-weighting lead to a distribution that matches the data distribution better. This indicates that the small uncertainties on the best-fit systematics lead to an unrealistic re-weighted distribution, while increasing the uncertainties or using a different re-weighting approach leads to a more realistic distribution that matches the data better. The method for w_spec can be performed on the entire combined MC dataset, while the other two methods require the systematic SnowStorm variations, which are only available in a subset of the entire MC set.

The best-fit systematics and their uncertainties are presented in Table F.1 for the four individual unfoldings. All values are within the simulation range. As described above, the uncertainties on the best-fit systematics are small.

F.4.2 Unfolded Flux Values

In this section, the final results are presented for the burnsample and full sample analysis. Table F.2 lists the average and individual unfolded atmospheric muon flux at the surface for the burnsample analysis performed on the combined MC dataset. The average flux values $\Phi(E)$ and their symmetric uncertainties σ_Φ are obtained by averaging the results of the four individual unfoldings (H3a, H4a, GST, GSF), which are also shown individually. The table lists the bin edges E_{min} and E_{max}. Two bin center definitions are provided: the geometric mean E_{log} which is defined as

$$E_{\mathrm{log}} = \sqrt{E_{\mathrm{min}} \cdot E_{\mathrm{max}}}\,, \tag{F.6}$$

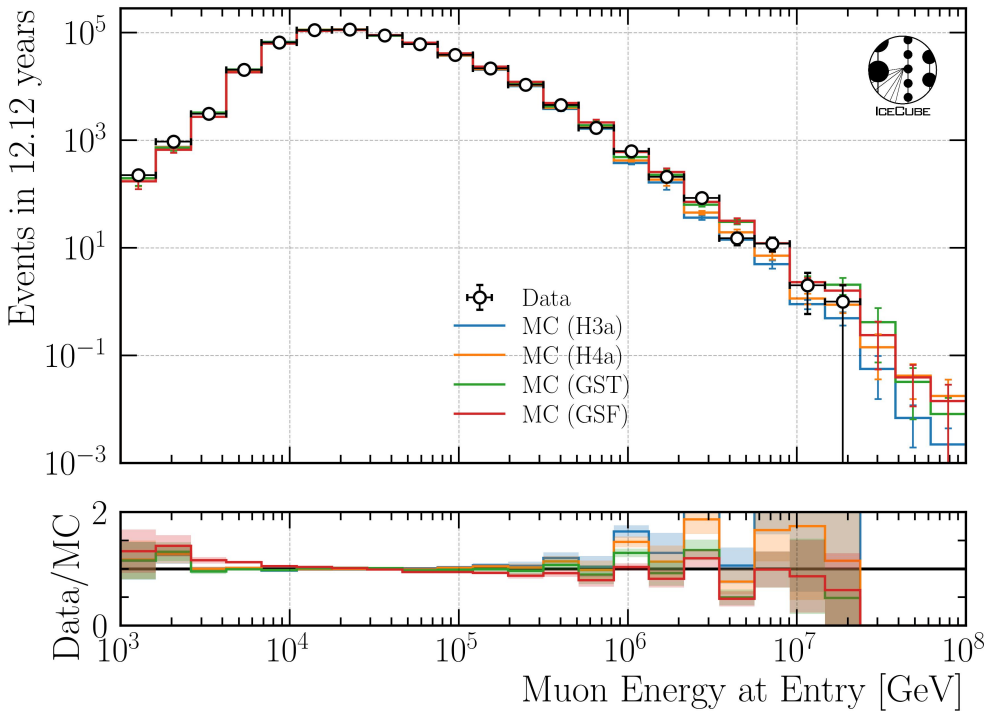
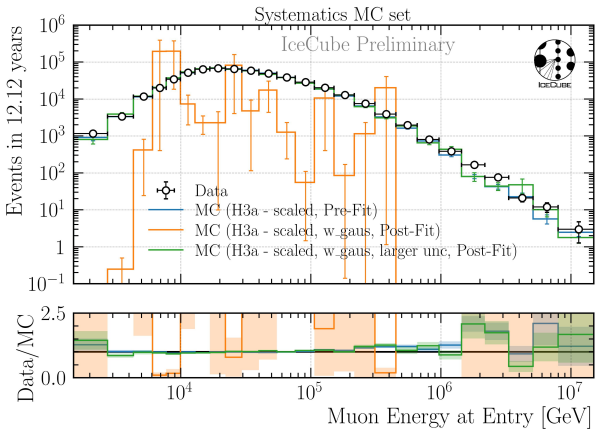


Figure F.17: Data–MC comparison for the leading muon energy at entry distribution for the full sample unfolding. The data distribution is compared to the combined MC distribution weighted to the four primary models. The MC distribution is based on the original MC weights without any re-weighting to the best-fit systematics.

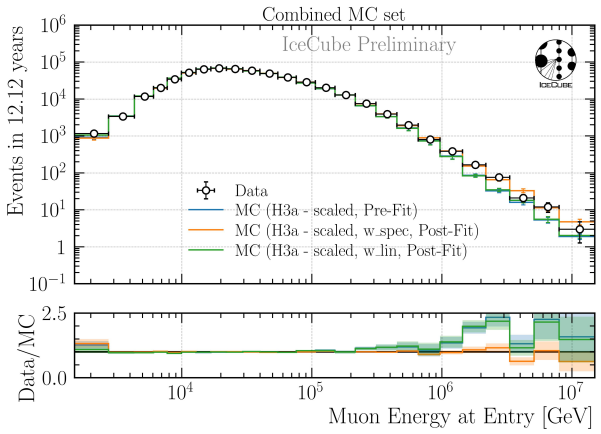
and the weighted expectation value E_{mean} calculated using the H3a-weighted simulation, which is defined as

$$E_{\text{mean}} = \frac{\sum_i w_i E_i}{\sum_i w_i}, \tag{F.7}$$

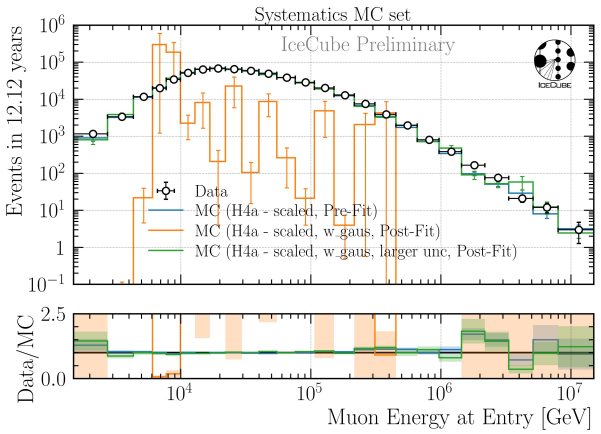
where w_i is the weight of event i and E_i is its energy. Table F.3 lists the results for the full sample analysis.



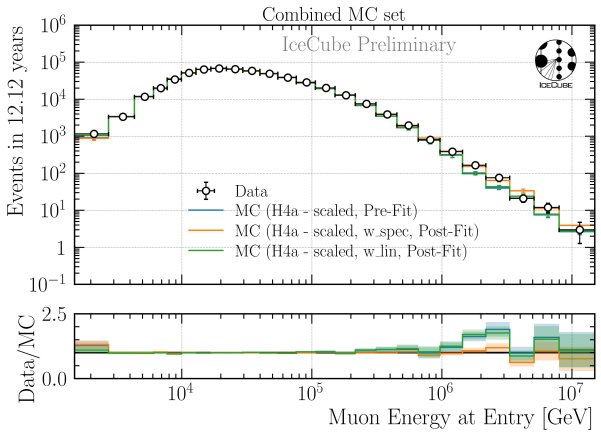
(a) Data–MC Gaussian re-weighting for H3a.



(b) Data–MC linear and spectral re-weighting for H3a.

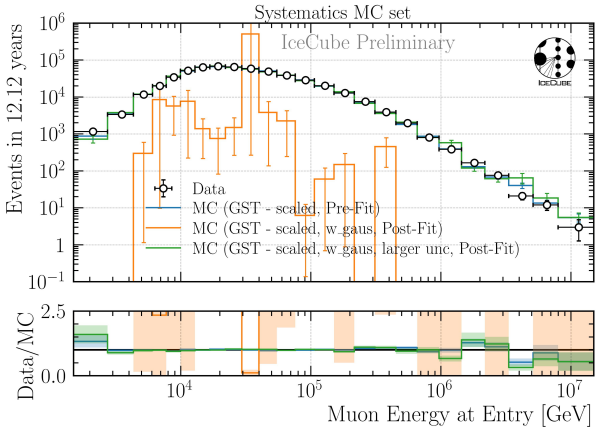


(c) Data–MC Gaussian re-weighting for H4a.

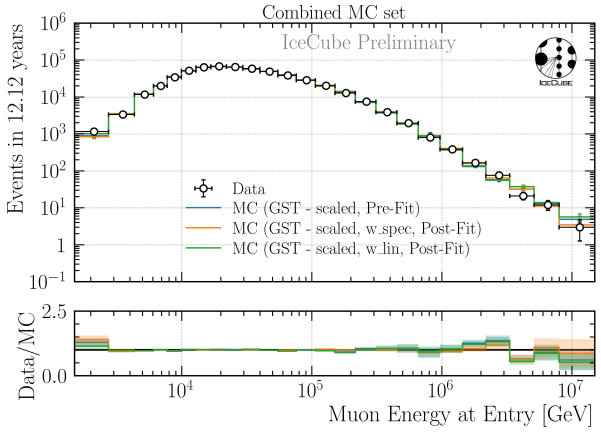


(d) Data–MC linear and spectral re-weighting for H4a.

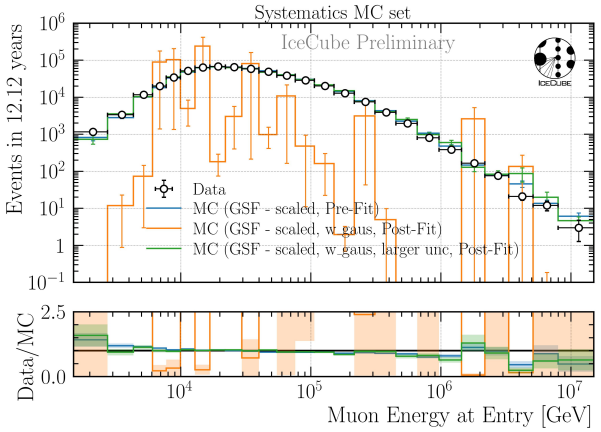
Figure F.18: This figure shows the data–MC comparisons for the leading muon energy at entry distribution for the full sample unfolding for the H3a and H4a matrices. The data distribution is compared to the combined MC distribution weighted to the four primary models. The MC distribution is re-weighted to the best-fit systematics using different approaches, including Gaussian re-weighting and linear and spectral re-weighting. The Gaussian re-weighting leads to a distribution that does not match the data distribution, while the linear and spectral re-weighting lead to a distribution that matches the data distribution better. This indicates that the small uncertainties on the best-fit systematics lead to an unrealistic re-weighted distribution, while increasing the uncertainties or using a different re-weighting approach leads to a more realistic distribution that matches the data better.



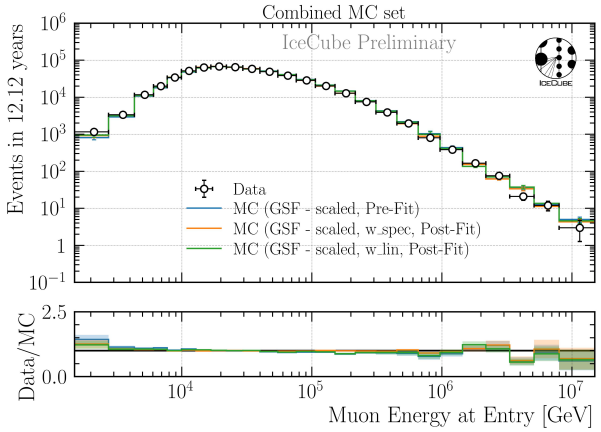
(a) Data–MC Gaussian re-weighting for GST.



(b) Data–MC linear and spectral re-weighting for GST.



(c) Data–MC Gaussian re-weighting for GSF.



(d) Data–MC linear and spectral re-weighting for GSF.

Figure F.19: This figure shows the data–MC comparisons for the leading muon energy at entry distribution for the full sample unfolding for the GST and GSF matrices. The data distribution is compared to the combined MC distribution weighted to the four primary models. The MC distribution is re-weighted to the best-fit systematics using different approaches, including Gaussian re-weighting and linear and spectral re-weighting. The Gaussian re-weighting leads to a distribution that does not match the data distribution, while the linear and spectral re-weighting lead to a distribution that matches the data distribution better. This indicates that the small uncertainties on the best-fit systematics lead to an unrealistic re-weighted distribution, while increasing the uncertainties or using a different re-weighting approach leads to a more realistic distribution that matches the data better.

Table F.2: Average and individual unfolded atmospheric muon flux at the surface for the burnsample analysis performed on the combined MC dataset. The average flux values $\Phi(E)$ and their symmetric uncertainties σ_Φ are obtained by averaging the results of the four individual unfoldings (H3a, H4a, GST, GSF), which are also shown individually. The table lists the bin edges E_{min} and E_{max}. Two bin center definitions are provided: the geometric mean E_{log} and the weighted expectation value E_{mean} calculated using the H3a-weighted simulation.

E_{min} / GeV	E_{max} / GeV	E_{log} / GeV	E_{mean} / GeV	Average	H3a	H4a	GST	GSF
				$\Phi(E)$ / $\text{GeV}^{-1}\,\text{m}^{-2}\,\text{s}^{-1}\,\text{sr}^{-1}$				
$2.00 \cdot 10^4$	$4.31 \cdot 10^4$	$2.94 \cdot 10^4$	$2.64 \cdot 10^4$	$(1.29 \pm 0.15) \times 10^{-15}$	$(1.29 \pm 0.13) \times 10^{-15}$	$(1.29 \pm 0.13) \times 10^{-15}$	$(1.29 \pm 0.12) \times 10^{-15}$	$(1.29 \pm 0.13) \times 10^{-15}$
$4.31 \cdot 10^4$	$9.28 \cdot 10^4$	$6.32 \cdot 10^4$	$5.68 \cdot 10^4$	$(6.23 \pm 0.42) \times 10^{-17}$	$(6.20 \pm 0.32) \times 10^{-17}$	$(6.18 \pm 0.32) \times 10^{-17}$	$(6.39 \pm 0.27) \times 10^{-17}$	$(6.13 \pm 0.32) \times 10^{-17}$
$9.28 \cdot 10^4$	$2.00 \cdot 10^5$	$1.36 \cdot 10^5$	$1.22 \cdot 10^5$	$(3.40 \pm 0.20) \times 10^{-18}$	$(3.39 \pm 0.18) \times 10^{-18}$	$(3.40 \pm 0.18) \times 10^{-18}$	$(3.37 \pm 0.16) \times 10^{-18}$	$(3.43 \pm 0.18) \times 10^{-18}$
$2.00 \cdot 10^5$	$4.31 \cdot 10^5$	$2.94 \cdot 10^5$	$2.62 \cdot 10^5$	$(1.94 \pm 0.16) \times 10^{-19}$	$(1.96 \pm 0.13) \times 10^{-19}$	$(1.97 \pm 0.13) \times 10^{-19}$	$(1.87 \pm 0.11) \times 10^{-19}$	$(1.98 \pm 0.13) \times 10^{-19}$
$4.31 \cdot 10^5$	$9.28 \cdot 10^5$	$6.32 \cdot 10^5$	$5.59 \cdot 10^5$	$(9.19 \pm 1.03) \times 10^{-21}$	$(9.28 \pm 0.94) \times 10^{-21}$	$(9.28 \pm 0.93) \times 10^{-21}$	$(9.13 \pm 0.83) \times 10^{-21}$	$(9.09 \pm 0.92) \times 10^{-21}$
$9.28 \cdot 10^5$	$2.00 \cdot 10^6$	$1.36 \cdot 10^6$	$1.21 \cdot 10^6$	$(4.46 \pm 1.23) \times 10^{-22}$	$(4.50 \pm 1.14) \times 10^{-22}$	$(4.46 \pm 1.12) \times 10^{-22}$	$(4.65 \pm 1.01) \times 10^{-22}$	$(4.23 \pm 1.07) \times 10^{-22}$

Table F.3: Average and individual unfolded atmospheric muon flux at the surface. The average flux values $\Phi(E)$ and their symmetric uncertainties σ_Φ are obtained by averaging the results of the four individual unfoldings (H3a, H4a, GST, GSF), which are also shown individually. The table lists the bin edges E_{min} and E_{max}. Two bin center definitions are provided: the geometric mean E_{log} and the weighted expectation value E_{mean} calculated using the H3a-weighted simulation.

E_{min} / GeV	E_{max} / GeV	E_{log} / GeV	E_{mean} / GeV	Average	H3a	H4a	GST	GSF
				$\Phi(E)$ / $\text{GeV}^{-1}\,\text{m}^{-2}\,\text{s}^{-1}\,\text{sr}^{-1}$				
$1.00 \cdot 10^4$	$2.08 \cdot 10^4$	$1.44 \cdot 10^4$	$1.32 \cdot 10^4$	$(1.74 \pm 0.15) \times 10^{-14}$	$(1.73 \pm 0.09) \times 10^{-14}$	$(1.73 \pm 0.09) \times 10^{-14}$	$(1.75 \pm 0.09) \times 10^{-14}$	$(1.76 \pm 0.08) \times 10^{-14}$
$2.08 \cdot 10^4$	$4.32 \cdot 10^4$	$3.00 \cdot 10^4$	$2.72 \cdot 10^4$	$(1.04 \pm 0.04) \times 10^{-15}$	$(1.04 \pm 0.02) \times 10^{-15}$	$(1.04 \pm 0.02) \times 10^{-15}$	$(1.05 \pm 0.02) \times 10^{-15}$	$(1.03 \pm 0.02) \times 10^{-15}$
$4.32 \cdot 10^4$	$8.97 \cdot 10^4$	$6.22 \cdot 10^4$	$5.65 \cdot 10^4$	$(6.67 \pm 0.52) \times 10^{-17}$	$(6.65 \pm 0.08) \times 10^{-17}$	$(6.62 \pm 0.08) \times 10^{-17}$	$(6.90 \pm 0.09) \times 10^{-17}$	$(6.50 \pm 0.08) \times 10^{-17}$
$8.97 \cdot 10^4$	$1.86 \cdot 10^5$	$1.29 \cdot 10^5$	$1.17 \cdot 10^5$	$(4.00 \pm 0.20) \times 10^{-18}$	$(4.01 \pm 0.05) \times 10^{-18}$	$(4.01 \pm 0.05) \times 10^{-18}$	$(3.93 \pm 0.06) \times 10^{-18}$	$(4.07 \pm 0.05) \times 10^{-18}$
$1.86 \cdot 10^5$	$3.87 \cdot 10^5$	$2.69 \cdot 10^5$	$2.41 \cdot 10^5$	$(2.73 \pm 0.21) \times 10^{-19}$	$(2.72 \pm 0.04) \times 10^{-19}$	$(2.75 \pm 0.05) \times 10^{-19}$	$(2.63 \pm 0.04) \times 10^{-19}$	$(2.81 \pm 0.05) \times 10^{-19}$
$3.87 \cdot 10^5$	$8.05 \cdot 10^5$	$5.58 \cdot 10^5$	$4.97 \cdot 10^5$	$(1.66 \pm 0.10) \times 10^{-20}$	$(1.67 \pm 0.03) \times 10^{-20}$	$(1.66 \pm 0.04) \times 10^{-20}$	$(1.66 \pm 0.03) \times 10^{-20}$	$(1.63 \pm 0.04) \times 10^{-20}$
$8.05 \cdot 10^5$	$1.67 \cdot 10^6$	$1.16 \cdot 10^6$	$1.04 \cdot 10^6$	$(1.02 \pm 0.12) \times 10^{-21}$	$(1.04 \pm 0.03) \times 10^{-21}$	$(1.03 \pm 0.03) \times 10^{-21}$	$(1.02 \pm 0.03) \times 10^{-21}$	$(0.98 \pm 0.03) \times 10^{-21}$
$1.67 \cdot 10^6$	$3.47 \cdot 10^6$	$2.41 \cdot 10^6$	$2.16 \cdot 10^6$	$(6.01 \pm 0.76) \times 10^{-23}$	$(6.22 \pm 0.34) \times 10^{-23}$	$(6.12 \pm 0.33) \times 10^{-23}$	$(5.82 \pm 0.33) \times 10^{-23}$	$(5.89 \pm 0.32) \times 10^{-23}$
$3.47 \cdot 10^6$	$7.22 \cdot 10^6$	$5.01 \cdot 10^6$	$4.53 \cdot 10^6$	$(6.17 \pm 1.79) \times 10^{-24}$	$(6.43 \pm 0.66) \times 10^{-24}$	$(6.37 \pm 0.66) \times 10^{-24}$	$(5.68 \pm 0.64) \times 10^{-24}$	$(6.22 \pm 0.64) \times 10^{-24}$
$7.22 \cdot 10^6$	$1.50 \cdot 10^7$	$1.04 \cdot 10^7$	$9.58 \cdot 10^6$	$(4.04 \pm 2.01) \times 10^{-25}$	$(4.23 \pm 0.78) \times 10^{-25}$	$(4.23 \pm 0.79) \times 10^{-25}$	$(3.52 \pm 0.73) \times 10^{-25}$	$(4.19 \pm 0.78) \times 10^{-25}$

F.5 Flux Normalization Fits

This section is divided into three subsections presenting the complete results of the flux normalization fits discussed in Section 8.3.10. All three subsections contain the same results, but sorted by different criteria for easier reference. Section F.5.1 presents the results sorted by the primary model used in the unfolding procedure, Section F.5.2 shows the results sorted by the primary model used in the flux prediction, and Section F.5.3 contains the results sorted by the HIM used in the flux prediction.

F.5.1 Results Sorted by Unfolding Primary Model

The tables below present the complete results sorted by the primary model used in the unfolding procedure.

Table F.4: Fit results for the H3a unfolded spectrum are compared to theoretical predictions from CORSIKA simulations and `MCEq` calculations. Predictions are calculated for two different HIM: SIBYLL 2.3c and SIBYLL 2.3d, each containing prompt muon production. All predictions are weighted to the four primary composition models H3a, H4a, GST and GSF. Fits are performed for the leading muon unfolding result only, and also with the preliminary stopping muon unfolding result to extend the low-energy region. The χ^2 values quantify the agreement between the measured flux and the model prediction of the total muon flux, with the prompt and conventional components as free parameters. The degree of freedom is nDoF = 8 for the fits without the stopping muons, and nDoF = 18 including stopping muons. These fits are performed with and without the covariance matrix. Significance in σ and p-values represent the detection significance for the prompt component, as explained in Section 8.3.10.

Prediction	Primary Model	Prompt Norm	Conv Norm	χ^2/nDoF	p-value	Significance/σ
CORSIKA SIBYLL-2.3d (no cov. matrix, no stopping muons)						
	H3a	3.605 ± 0.366	0.956 ± 0.034	0.39	1.76e-12	6.96
	H4a	2.507 ± 0.296	0.991 ± 0.032	0.53	1.04e-09	5.99
	GST	1.823 ± 0.267	1.259 ± 0.035	1.11	6.90e-07	4.83
	GSF	0.619 ± 0.207	1.153 ± 0.032	1.33	1.70e-02	2.12
CORSIKA SIBYLL-2.3d (with cov. matrix, no stopping muons)						
	H3a	3.630 ± 0.271	0.952 ± 0.025	0.76	1.40e-21	9.47
	H4a	2.564 ± 0.222	0.986 ± 0.023	1.01	1.45e-16	8.18
	GST	1.915 ± 0.204	1.255 ± 0.026	2.04	1.76e-11	6.62
	GSF	0.628 ± 0.157	1.148 ± 0.024	2.43	2.32e-03	2.83
MCEq SIBYLL-2.3c (no cov. matrix, no stopping muons)						
	H3a	2.734 ± 0.176	0.935 ± 0.021	0.87	2.16e-28	10.99
	H4a	2.236 ± 0.150	0.927 ± 0.019	0.90	2.37e-26	10.56
	GST	1.675 ± 0.134	1.283 ± 0.024	1.39	3.71e-19	8.87
	GSF	0.811 ± 0.091	1.069 ± 0.020	0.88	1.79e-10	6.27
MCEq SIBYLL-2.3c (with cov. matrix, no stopping muons)						
	H3a	2.740 ± 0.176	0.935 ± 0.020	0.89	2.06e-28	10.99
	H4a	2.257 ± 0.151	0.926 ± 0.019	0.91	2.32e-26	10.56
	GST	1.723 ± 0.138	1.280 ± 0.024	1.37	6.22e-19	8.81
	GSF	0.813 ± 0.094	1.069 ± 0.020	0.90	4.04e-10	6.14
MCEq SIBYLL-2.3c (no cov. matrix, with stopping muons)						
	H3a	1.476 ± 0.152	1.058 ± 0.006	0.89	3.47e-12	6.86
	H4a	1.406 ± 0.132	1.035 ± 0.006	0.76	2.16e-14	7.55
	GST	1.233 ± 0.122	1.057 ± 0.006	0.60	5.31e-13	7.12
	GSF	0.705 ± 0.082	1.117 ± 0.006	0.74	4.81e-10	6.12

Table F.5: Fit results for the H4a unfolded spectrum are compared to theoretical predictions from CORSIKA simulations and `MCEq` calculations. Predictions are calculated for two different HIM: SIBYLL 2.3c and SIBYLL 2.3d, each containing prompt muon production. All predictions are weighted to the four primary composition models H3a, H4a, GST and GSF. Fits are performed for the leading muon unfolding result only, and also with the preliminary stopping muon unfolding result to extend the low-energy region. The χ^2 values quantify the agreement between the measured flux and the model prediction of the total muon flux, with the prompt and conventional components as free parameters. The degree of freedom is nDoF = 8 for the fits without the stopping muons, and nDoF = 18 including stopping muons. These fits are performed with and without the covariance matrix. Significance in σ and p-values represent the detection significance for the prompt component, as explained in Section 8.3.10.

Prediction	Primary Model	Prompt Norm	Conv Norm	χ^2/nDoF	p-value	Significance/σ
CORSIKA SIBYLL-2.3d (no cov. matrix, no stopping muons)						
	H3a	3.564 ± 0.365	0.958 ± 0.034	0.37	2.40e-12	6.91
	H4a	2.471 ± 0.294	0.993 ± 0.032	0.52	1.43e-09	5.94
	GST	1.793 ± 0.265	1.261 ± 0.035	1.08	8.86e-07	4.78
	GSF	0.596 ± 0.205	1.155 ± 0.032	1.29	1.99e-02	2.06
CORSIKA SIBYLL-2.3d (with cov. matrix, no stopping muons)						
	H3a	3.579 ± 0.271	0.954 ± 0.025	0.72	5.06e-21	9.33
	H4a	2.517 ± 0.221	0.987 ± 0.023	0.96	4.45e-16	8.04
	GST	1.881 ± 0.204	1.254 ± 0.026	1.96	3.42e-11	6.52
	GSF	0.596 ± 0.156	1.148 ± 0.024	2.32	3.46e-03	2.70
MCEq SIBYLL-2.3c (no cov. matrix, no stopping muons)						
	H3a	2.705 ± 0.176	0.936 ± 0.021	0.85	5.81e-28	10.90
	H4a	2.208 ± 0.149	0.928 ± 0.020	0.89	6.50e-26	10.46
	GST	1.651 ± 0.133	1.284 ± 0.024	1.34	7.53e-19	8.79
	GSF	0.794 ± 0.091	1.070 ± 0.020	0.86	3.27e-10	6.18
MCEq SIBYLL-2.3c (no cov. matrix, with stopping muons)						
	H3a	1.458 ± 0.152	1.058 ± 0.006	0.86	5.99e-12	6.78
	H4a	1.387 ± 0.131	1.035 ± 0.006	0.74	4.30e-14	7.46
	GST	1.213 ± 0.122	1.058 ± 0.006	0.58	1.01e-12	7.03
	GSF	0.692 ± 0.081	1.117 ± 0.006	0.73	8.51e-10	6.02

Table F.6: Fit results for the GST unfolded spectrum are compared to theoretical predictions from CORSIKA simulations and `MCEq` calculations. Predictions are calculated for two different HIM: SIBYLL 2.3c and SIBYLL 2.3d, each containing prompt muon production. All predictions are weighted to the four primary composition models H3a, H4a, GST and GSF. Fits are performed for the leading muon unfolding result only, and also with the preliminary stopping muon unfolding result to extend the low-energy region. The χ^2 values quantify the agreement between the measured flux and the model prediction of the total muon flux, with the prompt and conventional components as free parameters. The degree of freedom is nDoF = 8 for the fits without the stopping muons, and nDoF = 18 including stopping muons. These fits are performed with and without the covariance matrix. Significance in σ and p-values represent the detection significance for the prompt component, as explained in Section 8.3.10.

Prediction	Prediction	Prompt Norm	Conv Norm	χ^2/nDoF	p-value	Significance/σ
CORSIKA SIBYLL-2.3d (no cov. matrix, no stopping muons)						
	H3a	3.287 ± 0.353	0.971 ± 0.034	0.58	2.31e-11	6.58
	H4a	2.203 ± 0.281	1.007 ± 0.032	0.78	1.49e-08	5.54
	GST	1.457 ± 0.246	1.281 ± 0.035	1.41	1.40e-05	4.19
	GSF	0.460 ± 0.194	1.159 ± 0.032	1.59	4.69e-02	1.68
CORSIKA SIBYLL-2.3d (with cov. matrix, no stopping muons)						
	H3a	3.323 ± 0.263	0.968 ± 0.025	1.11	1.88e-19	8.94
	H4a	2.273 ± 0.213	1.002 ± 0.023	1.46	1.98e-14	7.56
	GST	1.560 ± 0.191	1.277 ± 0.026	2.60	3.74e-09	5.78
	GSF	0.469 ± 0.149	1.155 ± 0.023	2.96	1.31e-02	2.22
MCEq SIBYLL-2.3c (no cov. matrix, no stopping muons)						
	H3a	2.536 ± 0.171	0.943 ± 0.021	1.09	4.27e-26	10.50
	H4a	2.046 ± 0.144	0.936 ± 0.020	1.19	5.56e-24	10.03
	GST	1.454 ± 0.125	1.295 ± 0.024	1.88	8.68e-17	8.24
	GSF	0.710 ± 0.087	1.073 ± 0.020	1.15	4.07e-09	5.77
MCEq SIBYLL-2.3c (no cov. matrix, with stopping muons)						
	H3a	1.334 ± 0.148	1.059 ± 0.006	0.96	8.99e-11	6.38
	H4a	1.267 ± 0.127	1.036 ± 0.006	0.87	9.29e-13	7.04
	GST	1.064 ± 0.115	1.058 ± 0.006	0.76	3.09e-11	6.54
	GSF	0.618 ± 0.078	1.118 ± 0.006	0.83	9.99e-09	5.61

Table F.7: Fit results for the GSF unfolded spectrum are compared to theoretical predictions from CORSIKA simulations and `MCEq` calculations. Predictions are calculated for two different HIM: SIBYLL 2.3c and SIBYLL 2.3d, each containing prompt muon production. All predictions are weighted to the four primary composition models H3a, H4a, GST and GSF. Fits are performed for the leading muon unfolding result only, and also with the preliminary stopping muon unfolding result to extend the low-energy region. The χ^2 values quantify the agreement between the measured flux and the model prediction of the total muon flux, with the prompt and conventional components as free parameters. The degree of freedom is nDoF = 8 for the fits without the stopping muons, and nDoF = 18 including stopping muons. These fits are performed with and without the covariance matrix. Significance in σ and p-values represent the detection significance for the prompt component, as explained in Section 8.3.10.

Prediction	Primary Model	Prompt Norm	Conv Norm	χ^2/nDoF	p-value	Significance/σ
CORSIKA SIBYLL-2.3d (no cov. matrix, no stopping muons)						
	H3a	3.443 ± 0.355	0.964 ± 0.034	0.30	3.64e-12	6.85
	H4a	2.367 ± 0.286	0.998 ± 0.031	0.47	2.55e-09	5.84
	GST	1.714 ± 0.259	1.263 ± 0.035	1.00	1.44e-06	4.68
	GSF	0.523 ± 0.200	1.159 ± 0.032	1.18	3.20e-02	1.85
CORSIKA SIBYLL-2.3d (with cov. matrix, no stopping muons)						
	H3a	3.459 ± 0.265	0.960 ± 0.025	0.58	1.27e-20	9.24
	H4a	2.412 ± 0.216	0.992 ± 0.023	0.85	1.44e-15	7.90
	GST	1.798 ± 0.199	1.257 ± 0.026	1.81	8.92e-11	6.38
	GSF	0.520 ± 0.152	1.154 ± 0.024	2.10	7.91e-03	2.41
MCEq SIBYLL-2.3c (no cov. matrix, no stopping muons)						
	H3a	2.612 ± 0.171	0.940 ± 0.020	0.84	1.53e-27	10.81
	H4a	2.125 ± 0.145	0.932 ± 0.019	0.92	2.10e-25	10.35
	GST	1.585 ± 0.129	1.287 ± 0.024	1.21	2.19e-18	8.67
	GSF	0.741 ± 0.088	1.073 ± 0.020	0.83	1.47e-09	5.93
MCEq SIBYLL-2.3c (no cov. matrix, with stopping muons)						
	H3a	1.390 ± 0.149	1.059 ± 0.006	0.79	1.96e-11	6.61
	H4a	1.321 ± 0.128	1.036 ± 0.006	0.71	1.80e-13	7.27
	GST	1.151 ± 0.119	1.058 ± 0.006	0.54	3.84e-12	6.84
	GSF	0.647 ± 0.079	1.118 ± 0.006	0.70	3.44e-09	5.79

F.5.2 Results Sorted by Primary Model Prediction

The tables below present the complete results sorted by the primary model used in the flux prediction.

Table F.8: Fit results for all unfolded spectra compared to H3a theoretical predictions from CORSIKA simulations and `MCEq` calculations. Predictions are calculated for three different HIM: SIBYLL 2.3c, SIBYLL 2.3d and DPMJet-III-19.1, each containing prompt muon production. Fits are performed for the four individual unfoldings utilizing the H3a, H4a, GST and GSF primary composition models, as well as their average, as described in Paragraph 8.3.8. For CORSIKA SIBYLL 2.3d and MCEq SIBYLL 2.3c, fits are shown both with and without the covariance matrix. Fits with stopping muons are also included to extend the low-energy region. The χ^2 values quantify the agreement between the measured flux and the model prediction of the total muon flux, with the prompt and conventional components as free parameters. The degree of freedom is nDoF = 8 for fits without stopping muons and nDoF = 18 including stopping muons. For MCEq DPMJet-III-19.1, only the average unfolding is performed. Significance in σ represents the detection significance for the prompt component, as explained in Section 8.3.10.

Prediction	Unfolding	Prompt Norm	Conv Norm	χ^2/nDoF	Significance/σ
CORSIKA SIBYLL-2.3d (no cov. matrix, no stopping muons)					
	H3a	3.605 ± 0.366	0.956 ± 0.034	0.39	6.96
	H4a	3.564 ± 0.365	0.958 ± 0.034	0.37	6.91
	GST	3.287 ± 0.353	0.971 ± 0.034	0.58	6.58
	GSF	3.443 ± 0.355	0.964 ± 0.034	0.30	6.85
	Average	3.481 ± 0.365	0.963 ± 0.034	0.38	6.75
CORSIKA SIBYLL-2.3d (with cov. matrix, no stopping muons)					
	H3a	3.630 ± 0.271	0.952 ± 0.025	0.76	9.47
	H4a	3.579 ± 0.271	0.954 ± 0.025	0.72	9.33
	GST	3.323 ± 0.263	0.968 ± 0.025	1.11	8.94
	GSF	3.459 ± 0.265	0.960 ± 0.025	0.58	9.24
MCEq SIBYLL-2.3c (no cov. matrix, no stopping muons)					
	H3a	2.734 ± 0.176	0.935 ± 0.021	0.87	10.99
	H4a	2.705 ± 0.176	0.936 ± 0.021	0.85	10.90
	GST	2.536 ± 0.171	0.943 ± 0.021	1.09	10.50
	GSF	2.612 ± 0.171	0.940 ± 0.020	0.84	10.81
	Average	2.638 ± 0.238	0.942 ± 0.028	0.47	7.84
MCEq SIBYLL-2.3c (with cov. matrix, no stopping muons)					
	H3a	2.740 ± 0.176	0.935 ± 0.020	0.89	10.99
	H4a	2.702 ± 0.176	0.935 ± 0.020	0.86	10.84
	GST	2.546 ± 0.171	0.944 ± 0.020	1.13	10.53
	GSF	2.611 ± 0.172	0.939 ± 0.020	0.84	10.73
MCEq SIBYLL-2.3c (no cov. matrix, with stopping muons)					
	H3a	1.476 ± 0.152	1.058 ± 0.006	0.89	6.86
	H4a	1.458 ± 0.152	1.058 ± 0.006	0.86	6.78
	GST	1.334 ± 0.148	1.059 ± 0.006	0.96	6.38
	GSF	1.390 ± 0.149	1.059 ± 0.006	0.79	6.61
	Average	1.424 ± 0.154	1.059 ± 0.006	0.82	6.53
MCEq DPMJET-III-19.1 (no cov. matrix, no stopping muons)					
	Average	1.551 ± 0.138	1.121 ± 0.031	0.31	7.95
MCEq DPMJET-III-19.1 (no cov. matrix, with stopping muons)					
	Average	1.179 ± 0.092	1.103 ± 0.006	1.16	9.02

Table F.9: Fit results for all unfolded spectra compared to H4a theoretical predictions from CORSIKA simulations and `MCEq` calculations. Predictions are calculated for three different HIM: SIBYLL 2.3c, SIBYLL 2.3d and DPMJet-III-19.1, each containing prompt muon production. Fits are performed for the four individual unfoldings utilizing the H3a, H4a, GST and GSF primary composition models, as well as their average, as described in Paragraph 8.3.8. For CORSIKA SIBYLL 2.3d and MCEq SIBYLL 2.3c, fits are shown both with and without the covariance matrix. Fits with stopping muons are also included to extend the low-energy region. The χ^2 values quantify the agreement between the measured flux and the model prediction of the total muon flux, with the prompt and conventional components as free parameters. The degree of freedom is nDoF = 8 for fits without stopping muons and nDoF = 18 including stopping muons. For MCEq DPMJet-III-19.1, only the average unfolding is available. Significance in σ represents the detection significance for the prompt component, as explained in Section 8.3.10.

Prediction	Unfolding	Prompt Norm	Conv Norm	χ^2/nDoF	Significance/σ
CORSIKA SIBYLL-2.3d (no cov. matrix, no stopping muons)					
	H3a	2.507 ± 0.296	0.991 ± 0.032	0.53	5.99
	H4a	2.471 ± 0.294	0.993 ± 0.032	0.52	5.94
	GST	2.203 ± 0.281	1.007 ± 0.032	0.78	5.54
	GSF	2.367 ± 0.286	0.998 ± 0.031	0.47	5.84
	Average	2.392 ± 0.293	0.998 ± 0.032	0.54	5.77
CORSIKA SIBYLL-2.3d (with cov. matrix, no stopping muons)					
	H3a	2.564 ± 0.222	0.986 ± 0.023	1.01	8.18
	H4a	2.517 ± 0.221	0.987 ± 0.023	0.96	8.04
	GST	2.273 ± 0.213	1.002 ± 0.023	1.46	7.56
	GSF	2.412 ± 0.216	0.992 ± 0.023	0.85	7.90
MCEq SIBYLL-2.3c (no cov. matrix, no stopping muons)					
	H3a	2.236 ± 0.150	0.927 ± 0.019	0.90	10.56
	H4a	2.208 ± 0.149	0.928 ± 0.020	0.89	10.46
	GST	2.046 ± 0.144	0.936 ± 0.020	1.19	10.03
	GSF	2.125 ± 0.145	0.932 ± 0.019	0.92	10.35
	Average	2.146 ± 0.202	0.934 ± 0.026	0.50	7.51
MCEq SIBYLL-2.3c (with cov. matrix, no stopping muons)					
	H3a	2.257 ± 0.151	0.926 ± 0.019	0.91	10.56
	H4a	2.222 ± 0.151	0.926 ± 0.019	0.89	10.41
	GST	2.074 ± 0.146	0.935 ± 0.019	1.20	10.07
	GSF	2.139 ± 0.147	0.929 ± 0.019	0.91	10.27
MCEq SIBYLL-2.3c (no cov. matrix, with stopping muons)					
	H3a	1.406 ± 0.132	1.035 ± 0.006	0.76	7.55
	H4a	1.387 ± 0.131	1.035 ± 0.006	0.74	7.46
	GST	1.267 ± 0.127	1.036 ± 0.006	0.87	7.04
	GSF	1.321 ± 0.128	1.036 ± 0.006	0.71	7.27
	Average	1.352 ± 0.133	1.036 ± 0.006	0.73	7.18
MCEq DPMJET-III-19.1 (no cov. matrix, no stopping muons)					
	Average	1.326 ± 0.123	1.101 ± 0.030	0.35	7.62
MCEq DPMJET-III-19.1 (no cov. matrix, with stopping muons)					
	Average	1.108 ± 0.083	1.078 ± 0.006	1.51	9.44

Table F.10: Fit results for all unfolded spectra compared to GST theoretical predictions from CORSIKA simulations and `MCEq` calculations. Predictions are calculated for three different HIM: SIBYLL 2.3c, SIBYLL 2.3d and DPMJet-III-19.1, each containing prompt muon production. Fits are performed for the four individual unfoldings utilizing the H3a, H4a, GST and GSF primary composition models, as well as their average, as described in Paragraph 8.3.8. For CORSIKA SIBYLL 2.3d and MCEq SIBYLL 2.3c, fits are shown both with and without the covariance matrix. Fits with stopping muons are also included to extend the low-energy region. The χ^2 values quantify the agreement between the measured flux and the model prediction of the total muon flux, with the prompt and conventional components as free parameters. The degree of freedom is nDoF = 8 for fits without stopping muons and nDoF = 18 including stopping muons. For MCEq DPMJet-III-19.1, only the average unfolding is available. Significance in σ represents the detection significance for the prompt component, as explained in Section 8.3.10.

Prediction	Unfolding	Prompt Norm	Conv Norm	χ^2/nDoF	Significance/σ
CORSIKA SIBYLL-2.3d (no cov. matrix, no stopping muons)					
	H3a	1.823 ± 0.267	1.259 ± 0.035	1.11	4.83
	H4a	1.793 ± 0.265	1.261 ± 0.035	1.08	4.78
	GST	1.457 ± 0.246	1.281 ± 0.035	1.41	4.19
	GSF	1.714 ± 0.259	1.263 ± 0.035	1.00	4.68
	Average	1.700 ± 0.263	1.266 ± 0.036	1.10	4.57
CORSIKA SIBYLL-2.3d (with cov. matrix, no stopping muons)					
	H3a	1.915 ± 0.204	1.255 ± 0.026	2.04	6.62
	H4a	1.881 ± 0.204	1.254 ± 0.026	1.96	6.52
	GST	1.560 ± 0.191	1.277 ± 0.026	2.60	5.78
	GSF	1.798 ± 0.199	1.257 ± 0.026	1.81	6.38
MCEq SIBYLL-2.3c (no cov. matrix, no stopping muons)					
	H3a	1.675 ± 0.134	1.283 ± 0.024	1.39	8.87
	H4a	1.651 ± 0.133	1.284 ± 0.024	1.34	8.79
	GST	1.454 ± 0.125	1.295 ± 0.024	1.88	8.24
	GSF	1.585 ± 0.129	1.287 ± 0.024	1.21	8.67
	Average	1.588 ± 0.179	1.289 ± 0.033	0.74	6.29
MCEq SIBYLL-2.3c (with cov. matrix, no stopping muons)					
	H3a	1.723 ± 0.138	1.280 ± 0.024	1.37	8.81
	H4a	1.696 ± 0.138	1.279 ± 0.024	1.30	8.70
	GST	1.511 ± 0.130	1.292 ± 0.024	1.86	8.21
	GSF	1.627 ± 0.134	1.282 ± 0.024	1.18	8.56
MCEq SIBYLL-2.3c (no cov. matrix, with stopping muons)					
	H3a	1.233 ± 0.122	1.057 ± 0.006	0.60	7.12
	H4a	1.213 ± 0.122	1.058 ± 0.006	0.58	7.03
	GST	1.064 ± 0.115	1.058 ± 0.006	0.76	6.54
	GSF	1.151 ± 0.119	1.058 ± 0.006	0.54	6.84
	Average	1.170 ± 0.123	1.058 ± 0.006	0.58	6.75
MCEq DPMJET-III-19.1 (no cov. matrix, no stopping muons)					
	Average	1.200 ± 0.129	1.485 ± 0.038	0.38	6.56
MCEq DPMJET-III-19.1 (no cov. matrix, with stopping muons)					
	Average	1.113 ± 0.088	1.099 ± 0.006	1.59	8.89

Table F.11: Fit results for all unfolded spectra compared to GSF theoretical predictions from CORSIKA simulations and `MCEq` calculations. Predictions are calculated for three different HIM: SIBYLL 2.3c, SIBYLL 2.3d and DPMJet-III-19.1, each containing prompt muon production. Fits are performed for the four individual unfoldings utilizing the H3a, H4a, GST and GSF primary composition models, as well as their average, as described in Paragraph 8.3.8. For CORSIKA SIBYLL 2.3d and MCEq SIBYLL 2.3c, fits are shown both with and without the covariance matrix. Fits with stopping muons are also included to extend the low-energy region. The χ^2 values quantify the agreement between the measured flux and the model prediction of the total muon flux, with the prompt and conventional components as free parameters. The degree of freedom is nDoF = 8 for fits without stopping muons and nDoF = 18 including stopping muons. For MCEq DPMJet-III-19.1, only the average unfolding is available. Significance in σ represents the detection significance for the prompt component, as explained in Section 8.3.10.

Prediction	Unfolding	Prompt Norm	Conv Norm	χ^2/nDoF	Significance/σ
CORSIKA SIBYLL-2.3d (no cov. matrix, no stopping muons)					
	H3a	0.619 ± 0.207	1.153 ± 0.032	1.33	2.12
	H4a	0.596 ± 0.205	1.155 ± 0.032	1.29	2.06
	GST	0.460 ± 0.194	1.159 ± 0.032	1.59	1.68
	GSF	0.523 ± 0.200	1.159 ± 0.032	1.18	1.85
	Average	0.548 ± 0.204	1.159 ± 0.033	1.29	1.90
CORSIKA SIBYLL-2.3d (with cov. matrix, no stopping muons)					
	H3a	0.628 ± 0.157	1.148 ± 0.024	2.43	2.83
	H4a	0.596 ± 0.156	1.148 ± 0.024	2.32	2.70
	GST	0.469 ± 0.149	1.155 ± 0.023	2.96	2.22
	GSF	0.520 ± 0.152	1.154 ± 0.024	2.10	2.41
MCEq SIBYLL-2.3c (no cov. matrix, no stopping muons)					
	H3a	0.811 ± 0.091	1.069 ± 0.020	0.88	6.27
	H4a	0.794 ± 0.091	1.070 ± 0.020	0.86	6.18
	GST	0.710 ± 0.087	1.073 ± 0.020	1.15	5.77
	GSF	0.741 ± 0.088	1.073 ± 0.020	0.83	5.93
	Average	0.758 ± 0.123	1.075 ± 0.028	0.48	4.37
MCEq SIBYLL-2.3c (with cov. matrix, no stopping muons)					
	H3a	0.813 ± 0.094	1.069 ± 0.020	0.90	6.14
	H4a	0.792 ± 0.093	1.069 ± 0.020	0.87	6.01
	GST	0.713 ± 0.089	1.074 ± 0.020	1.19	5.64
	GSF	0.738 ± 0.091	1.072 ± 0.020	0.83	5.76
MCEq SIBYLL-2.3c (no cov. matrix, with stopping muons)					
	H3a	0.705 ± 0.082	1.117 ± 0.006	0.74	6.12
	H4a	0.692 ± 0.081	1.117 ± 0.006	0.73	6.02
	GST	0.618 ± 0.078	1.118 ± 0.006	0.83	5.61
	GSF	0.647 ± 0.079	1.118 ± 0.006	0.70	5.79
	Average	0.669 ± 0.082	1.118 ± 0.006	0.69	5.78
MCEq DPMJET-III-19.1 (no cov. matrix, no stopping muons)					
	Average	0.510 ± 0.082	1.253 ± 0.031	0.36	4.41
MCEq DPMJET-III-19.1 (no cov. matrix, with stopping muons)					
	Average	0.610 ± 0.055	1.166 ± 0.007	0.78	7.77

F.5.3 Results Sorted by Hadronic Interaction Model Prediction

The tables below present the complete results sorted by the hadronic interaction model used in the flux prediction.

Table F.12: Fit results for CORSIKA SIBYLL 2.3d predictions without covariance matrix. The four individual unfoldings utilizing the H3a, H4a, GST and GSF primary composition models, as well as their average, as described in Paragraph 8.3.8, are compared to theoretical predictions weighted to all four primary models. Each section shows how one unfolded dataset fits to all four theoretical primary model predictions. The χ^2 values quantify the agreement between the measured flux and the model prediction of the total muon flux, with the prompt and conventional components as free parameters. The degree of freedom is nDoF = 8 for all fits. Significance in σ and p-values represent the detection significance for the prompt component, as explained in Section 8.3.10.

Unfolding	Prediction	Prompt Norm	Conv Norm	χ^2/nDoF	p-value	Significance/σ
H3a Unfolding (no cov. matrix, no stopping muons)						
	H3a	3.605 ± 0.366	0.956 ± 0.034	0.39	1.76e-12	6.96
	H4a	2.507 ± 0.296	0.991 ± 0.032	0.53	1.04e-09	5.99
	GST	1.823 ± 0.267	1.259 ± 0.035	1.11	6.90e-07	4.83
	GSF	0.619 ± 0.207	1.153 ± 0.032	1.33	1.70e-02	2.12
H4a Unfolding (no cov. matrix, no stopping muons)						
	H3a	3.564 ± 0.365	0.958 ± 0.034	0.37	2.40e-12	6.91
	H4a	2.471 ± 0.294	0.993 ± 0.032	0.52	1.43e-09	5.94
	GST	1.793 ± 0.265	1.261 ± 0.035	1.08	8.86e-07	4.78
	GSF	0.596 ± 0.205	1.155 ± 0.032	1.29	1.99e-02	2.06
GST Unfolding (no cov. matrix, no stopping muons)						
	H3a	3.287 ± 0.353	0.971 ± 0.034	0.58	2.31e-11	6.58
	H4a	2.203 ± 0.281	1.007 ± 0.032	0.78	1.49e-08	5.54
	GST	1.457 ± 0.246	1.281 ± 0.035	1.41	1.40e-05	4.19
	GSF	0.460 ± 0.194	1.159 ± 0.032	1.59	4.69e-02	1.68
GSF Unfolding (no cov. matrix, no stopping muons)						
	H3a	3.443 ± 0.355	0.964 ± 0.034	0.30	3.64e-12	6.85
	H4a	2.367 ± 0.286	0.998 ± 0.031	0.47	2.55e-09	5.84
	GST	1.714 ± 0.259	1.263 ± 0.035	1.00	1.44e-06	4.68
	GSF	0.523 ± 0.200	1.159 ± 0.032	1.18	3.20e-02	1.85
Average Unfolding (no cov. matrix, no stopping muons)						
	H3a	3.481 ± 0.365	0.963 ± 0.034	0.38	7.38e-12	6.75
	H4a	2.392 ± 0.293	0.998 ± 0.032	0.54	4.03e-09	5.77
	GST	1.700 ± 0.263	1.266 ± 0.036	1.10	2.41e-06	4.57
	GSF	0.548 ± 0.204	1.159 ± 0.033	1.29	2.88e-02	1.90

Table F.13: Fit results for CORSIKA SIBYLL 2.3d predictions with covariance matrix. The four individual unfoldings utilizing the H3a, H4a, GST and GSF primary composition models are compared to theoretical predictions weighted to all four primary models. Each section shows how one unfolded dataset fits to all four theoretical primary model predictions. The χ^2 values quantify the agreement between the measured flux and the model prediction of the total muon flux, with the prompt and conventional components as free parameters. The degree of freedom is nDoF = 8 for all fits. The average unfolding is not available with the covariance matrix. Significance in σ and p-values represent the detection significance for the prompt component, as explained in Section 8.3.10.

Unfolding	Prediction	Prompt Norm	Conv Norm	χ^2/nDoF	p-value	Significance/σ
H3a Unfolding (with cov. matrix, no stopping muons)						
	H3a	3.630 ± 0.271	0.952 ± 0.025	0.76	1.40e-21	9.47
	H4a	2.564 ± 0.222	0.986 ± 0.023	1.01	1.45e-16	8.18
	GST	1.915 ± 0.204	1.255 ± 0.026	2.04	1.76e-11	6.62
	GSF	0.628 ± 0.157	1.148 ± 0.024	2.43	2.32e-03	2.83
H4a Unfolding (with cov. matrix, no stopping muons)						
	H3a	3.579 ± 0.271	0.954 ± 0.025	0.72	5.06e-21	9.33
	H4a	2.517 ± 0.221	0.987 ± 0.023	0.96	4.45e-16	8.04
	GST	1.881 ± 0.204	1.254 ± 0.026	1.96	3.42e-11	6.52
	GSF	0.596 ± 0.156	1.148 ± 0.024	2.32	3.46e-03	2.70
GST Unfolding (with cov. matrix, no stopping muons)						
	H3a	3.323 ± 0.263	0.968 ± 0.025	1.11	1.88e-19	8.94
	H4a	2.273 ± 0.213	1.002 ± 0.023	1.46	1.98e-14	7.56
	GST	1.560 ± 0.191	1.277 ± 0.026	2.60	3.74e-09	5.78
	GSF	0.469 ± 0.149	1.155 ± 0.023	2.96	1.31e-02	2.22
GSF Unfolding (with cov. matrix, no stopping muons)						
	H3a	3.459 ± 0.265	0.960 ± 0.025	0.58	1.27e-20	9.24
	H4a	2.412 ± 0.216	0.992 ± 0.023	0.85	1.44e-15	7.90
	GST	1.798 ± 0.199	1.257 ± 0.026	1.81	8.92e-11	6.38
	GSF	0.520 ± 0.152	1.154 ± 0.024	2.10	7.91e-03	2.41

Table F.14: Fit results for MCEq SIBYLL 2.3c predictions without covariance matrix and with stopping muons to extend the low-energy region. The four individual unfoldings utilizing the H3a, H4a, GST and GSF primary composition models, as well as their average, as described in Paragraph 8.3.8, are compared to theoretical predictions weighted to all four primary models. Each section shows how one unfolded dataset fits to all four theoretical primary model predictions. The χ^2 values quantify the agreement between the measured flux and the model prediction of the total muon flux, with the prompt and conventional components as free parameters. The degree of freedom is nDoF = 18 for all fits. Significance in σ and p-values represent the detection significance for the prompt component, as explained in Section 8.3.10.

Unfolding	Prediction	Prompt Norm	Conv Norm	χ^2/nDoF	p-value	Significance/σ
H3a Unfolding (no cov. matrix, with stopping muons)						
	H3a	2.734 ± 0.176	0.935 ± 0.021	0.87	2.16e-28	10.99
	H4a	2.236 ± 0.150	0.927 ± 0.019	0.90	2.37e-26	10.56
	GST	1.675 ± 0.134	1.283 ± 0.024	1.39	3.71e-19	8.87
	GSF	0.811 ± 0.091	1.069 ± 0.020	0.88	1.79e-10	6.27
H4a Unfolding (no cov. matrix, with stopping muons)						
	H3a	2.705 ± 0.176	0.936 ± 0.021	0.85	5.81e-28	10.90
	H4a	2.208 ± 0.149	0.928 ± 0.020	0.89	6.50e-26	10.46
	GST	1.651 ± 0.133	1.284 ± 0.024	1.34	7.53e-19	8.79
	GSF	0.794 ± 0.091	1.070 ± 0.020	0.86	3.27e-10	6.18
GST Unfolding (no cov. matrix, with stopping muons)						
	H3a	2.536 ± 0.171	0.943 ± 0.021	1.09	4.27e-26	10.50
	H4a	2.046 ± 0.144	0.936 ± 0.020	1.19	5.56e-24	10.03
	GST	1.454 ± 0.125	1.295 ± 0.024	1.88	8.68e-17	8.24
	GSF	0.710 ± 0.087	1.073 ± 0.020	1.15	4.07e-09	5.77
GSF Unfolding (no cov. matrix, with stopping muons)						
	H3a	2.612 ± 0.171	0.940 ± 0.020	0.84	1.53e-27	10.81
	H4a	2.125 ± 0.145	0.932 ± 0.019	0.92	2.10e-25	10.35
	GST	1.585 ± 0.129	1.287 ± 0.024	1.21	2.19e-18	8.67
	GSF	0.741 ± 0.088	1.073 ± 0.020	0.83	1.47e-09	5.93
Average Unfolding (no cov. matrix, with stopping muons)						
	H3a	2.638 ± 0.238	0.942 ± 0.028	0.47	2.34e-15	7.84
	H4a	2.146 ± 0.202	0.934 ± 0.026	0.50	2.95e-14	7.51
	GST	1.588 ± 0.179	1.289 ± 0.033	0.74	1.61e-10	6.29
	GSF	0.758 ± 0.123	1.075 ± 0.028	0.48	6.21e-06	4.37

Table F.15: Fit results for MCEq SIBYLL 2.3c predictions with covariance matrix and without stopping muons. The four individual unfoldings utilizing the H3a, H4a, GST and GSF primary composition models are compared to theoretical predictions weighted to all four primary models. Each section shows how one unfolded dataset fits to all four theoretical primary model predictions. The χ^2 values quantify the agreement between the measured flux and the model prediction of the total muon flux, with the prompt and conventional components as free parameters. The degree of freedom is nDoF = 8 for all fits. The average unfolding is not available with the covariance matrix. Significance in σ and p-values represent the detection significance for the prompt component, as explained in Section 8.3.10.

Unfolding	Prediction	Prompt Norm	Conv Norm	χ^2/nDoF	p-value	Significance/σ
H3a Unfolding (with cov. matrix, no stopping muons)						
	H3a	2.740 ± 0.176	0.935 ± 0.020	0.89	2.06e-28	10.99
	H4a	2.257 ± 0.151	0.926 ± 0.019	0.91	2.32e-26	10.56
	GST	1.723 ± 0.138	1.280 ± 0.024	1.37	6.22e-19	8.81
	GSF	0.813 ± 0.094	1.069 ± 0.020	0.90	4.04e-10	6.14
H4a Unfolding (with cov. matrix, no stopping muons)						
	H3a	2.702 ± 0.176	0.936 ± 0.020	0.86	1.08e-27	10.84
	H4a	2.222 ± 0.151	0.927 ± 0.019	0.89	1.04e-25	10.41
	GST	1.696 ± 0.138	1.281 ± 0.024	1.33	8.70e-19	8.74
	GSF	0.792 ± 0.093	1.070 ± 0.020	0.88	6.01e-10	6.07
GST Unfolding (with cov. matrix, no stopping muons)						
	H3a	2.546 ± 0.171	0.942 ± 0.020	1.09	1.05e-25	10.53
	H4a	2.074 ± 0.146	0.935 ± 0.020	1.18	1.01e-23	10.07
	GST	1.511 ± 0.130	1.293 ± 0.024	1.85	8.21e-17	8.18
	GSF	0.713 ± 0.089	1.073 ± 0.020	1.14	5.64e-09	5.70
GSF Unfolding (with cov. matrix, no stopping muons)						
	H3a	2.611 ± 0.172	0.940 ± 0.020	0.85	1.07e-26	10.73
	H4a	2.139 ± 0.147	0.931 ± 0.019	0.93	1.03e-24	10.27
	GST	1.627 ± 0.134	1.286 ± 0.024	1.23	8.56e-18	8.58
	GSF	0.738 ± 0.091	1.072 ± 0.020	0.84	5.76e-09	5.81

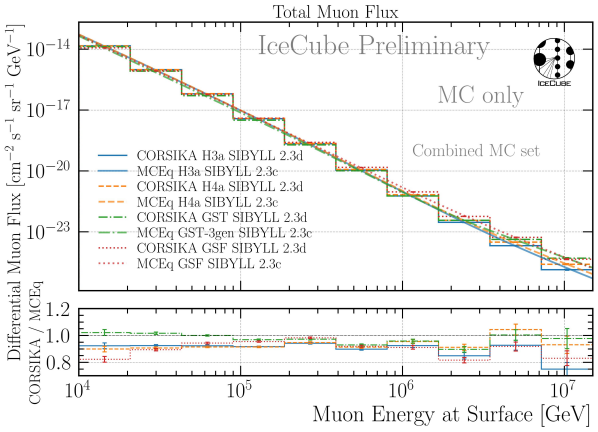
Table F.16: Fit results for MCEq SIBYLL 2.3c predictions without covariance matrix and with stopping muons to extend the low-energy region. The four individual unfoldings utilizing the H3a, H4a, GST and GSF primary composition models, as well as their average, as described in Paragraph 8.3.8, are compared to theoretical predictions weighted to all four primary models. Each section shows how one unfolded dataset fits to all four theoretical primary model predictions. The χ^2 values quantify the agreement between the measured flux and the model prediction of the total muon flux, with the prompt and conventional components as free parameters. The degree of freedom is nDoF = 18 for all fits. Significance in σ and p-values represent the detection significance for the prompt component, as explained in Section 8.3.10.

Unfolding	Prediction	Prompt Norm	Conv Norm	χ^2/nDoF	p-value	Significance/σ
H3a Unfolding (no cov. matrix, with stopping muons)						
	H3a	1.476 ± 0.152	1.058 ± 0.006	0.89	3.47e-12	6.86
	H4a	1.406 ± 0.132	1.035 ± 0.006	0.76	2.16e-14	7.55
	GST	1.233 ± 0.122	1.057 ± 0.006	0.60	5.31e-13	7.12
	GSF	0.705 ± 0.082	1.117 ± 0.006	0.74	4.81e-10	6.12
H4a Unfolding (no cov. matrix, with stopping muons)						
	H3a	1.458 ± 0.152	1.058 ± 0.006	0.86	5.99e-12	6.78
	H4a	1.387 ± 0.131	1.035 ± 0.006	0.74	4.30e-14	7.46
	GST	1.213 ± 0.122	1.058 ± 0.006	0.58	1.01e-12	7.03
	GSF	0.692 ± 0.081	1.117 ± 0.006	0.73	8.51e-10	6.02
GST Unfolding (no cov. matrix, with stopping muons)						
	H3a	1.334 ± 0.148	1.059 ± 0.006	0.96	8.99e-11	6.38
	H4a	1.267 ± 0.127	1.036 ± 0.006	0.87	9.29e-13	7.04
	GST	1.064 ± 0.115	1.058 ± 0.006	0.76	3.09e-11	6.54
	GSF	0.618 ± 0.078	1.118 ± 0.006	0.83	9.99e-09	5.61
GSF Unfolding (no cov. matrix, with stopping muons)						
	H3a	1.390 ± 0.149	1.059 ± 0.006	0.79	1.96e-11	6.61
	H4a	1.321 ± 0.128	1.036 ± 0.006	0.71	1.80e-13	7.27
	GST	1.151 ± 0.119	1.058 ± 0.006	0.54	3.84e-12	6.84
	GSF	0.647 ± 0.079	1.118 ± 0.006	0.70	3.44e-09	5.79
Average Unfolding (no cov. matrix, with stopping muons)						
	H3a	1.424 ± 0.154	1.059 ± 0.006	0.82	3.27e-11	6.53
	H4a	1.352 ± 0.133	1.036 ± 0.006	0.73	3.37e-13	7.18
	GST	1.170 ± 0.123	1.058 ± 0.006	0.58	7.58e-12	6.75
	GSF	0.669 ± 0.082	1.118 ± 0.006	0.69	3.84e-09	5.78

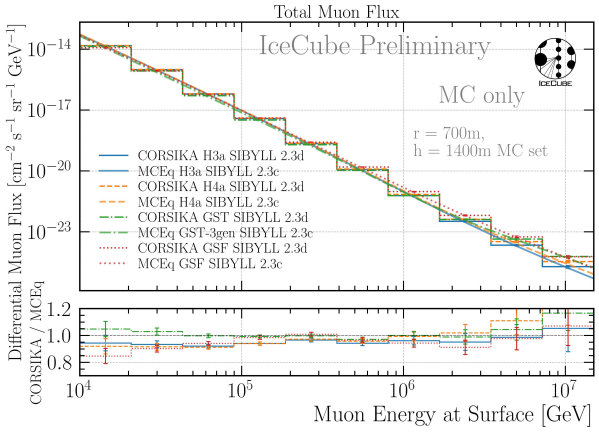
Table F.17: Fit results for MCEq DPMJet-III-19.1 predictions. Only the average unfolded spectrum, as described in Paragraph 8.3.8, is compared to theoretical predictions weighted to all four primary composition models H3a, H4a, GST and GSF. Fits are shown both without and with the preliminary stopping muon unfolding result to extend the low-energy region. The χ^2 values quantify the agreement between the measured flux and the model prediction of the total muon flux, with the prompt and conventional components as free parameters. The degree of freedom is nDoF = 8 for fits without stopping muons and nDoF = 18 including stopping muons. No covariance matrix is used. Significance in σ and p-values represent the detection significance for the prompt component, as explained in Section 8.3.10.

Unfolding	Prediction	Prompt Norm	Conv Norm	χ^2/nDoF	p-value	Significance/σ
Average Unfolding (no cov. matrix, no stopping muons)						
	H3a	1.551 ± 0.138	1.121 ± 0.031	0.31	8.98e-16	7.95
	H4a	1.326 ± 0.123	1.101 ± 0.030	0.35	1.25e-14	7.62
	GST	1.200 ± 0.129	1.485 ± 0.038	0.38	2.75e-11	6.56
	GSF	0.510 ± 0.082	1.253 ± 0.031	0.36	5.27e-06	4.41
Average Unfolding (no cov. matrix, with stopping muons)						
	H3a	1.179 ± 0.092	1.103 ± 0.006	1.16	9.32e-20	9.02
	H4a	1.108 ± 0.083	1.078 ± 0.006	1.51	1.89e-21	9.44
	GST	1.113 ± 0.088	1.099 ± 0.006	1.59	3.00e-19	8.89
	GSF	0.610 ± 0.055	1.166 ± 0.007	0.78	3.84e-15	7.77

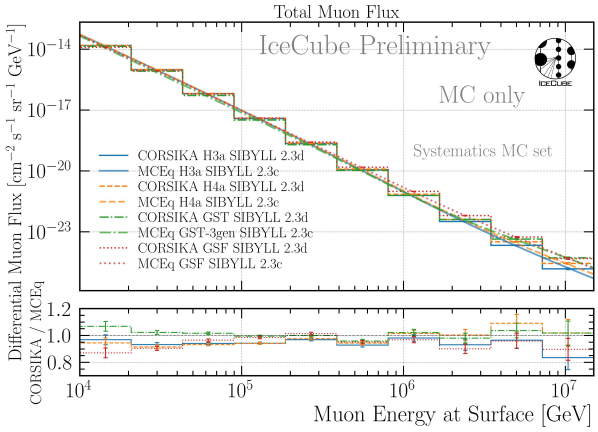
F.6 CORSIKA vs MCEq Comparison



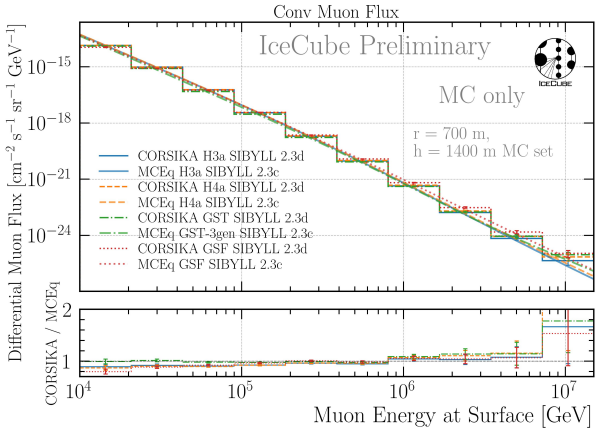
(a) Combined MC set total muon flux.



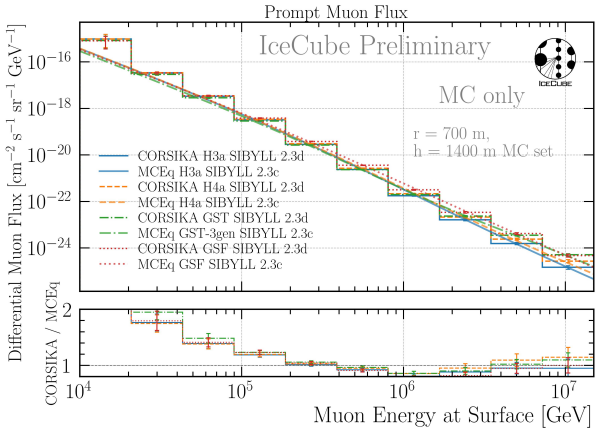
(b) MC set with $r = 700\,\mathrm{m}$ and $h = 1400\,\mathrm{m}$ total muon flux.



(c) Systematics MC set total muon flux.



(d) MC set with $r = 700\,\mathrm{m}$ and $h = 1400\,\mathrm{m}$ conventional muon flux.



(e) MC set with $r = 700\,\mathrm{m}$ and $h = 1400\,\mathrm{m}$ prompt muon flux.

Figure F.20: This figure presents the comparison between CORSIKA simulations and `MCEq` calculations for different MC datasets on *Final Level*. The top panel shows the total muon flux for the combined MC set, while the center left panel presents the total muon flux for the MC set with a cylinder radius of $r = 700\,\mathrm{m}$ and a height of $h = 1400\,\mathrm{m}$. The center right panel displays the total muon flux for the systematics MC set. The bottom left and right panels show the conventional and prompt muon fluxes, respectively, for the MC set with $r = 700\,\mathrm{m}$ and $h = 1400\,\mathrm{m}$. Only uncertainties arsing from the weights are considered, but not uncertainties from the effective area. Overall, there is an agreement between CORSIKA simulations and `MCEq` calculations across all datasets and muon components. However, some deviations and differences in the shapes of the spectra are visible, especially in the higher energy region for conventional, and the lower energy region for the prompt component.

F.7 Scaling Factors

Table F.18: Scaling factors $n_{\mathrm{s},\,i}$ calculated with Equation (7.2) are presented for the four primary weightings on *Final Level*. The factor is applied to MC to match a normalization offset to the data. Factors are calculated on the full sample for the different combinations of the MC datasets, as stated in Table 8.3. The combined dataset includes all datasets, while the systematics dataset includes only the datasets used to estimate ice systematic uncertainties (2277X, 2353X, 23111), and the $r = 700\,\mathrm{m}$ and $h = 1400\,\mathrm{m}$ dataset, which includes the extended history information to label a muon as conventional or prompt. Slightly different scaling factors are caused by different coverages and statistics in the energy range and the existance/non-existance of coincident events.

	Scaling factor $n_{\mathrm{s},\,i}$		
Primary weighting	Combined MC	Systematics MC	EHist MC
H3a	1.08	1.09	1.10
H4a	1.05	1.06	1.07
GST	1.27	1.29	1.30
GSF	1.05	1.06	1.07

F.8 Stopping Muon Event Display

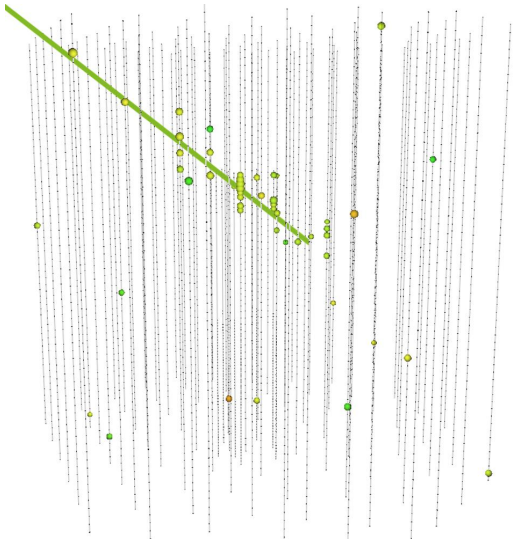


Figure F.21: Event display of a stopping muon event in the detector. The likely single muon looses all its energy and decays within the detector volume. Due to the low energy of the muon, only a few DOMs are triggered, which are shown in colored circles. The small size of the circles indicates a low amount of light detected. Figure taken from [160].

Bibliography

[1] IceCube Collaboration, *The IceCube Neutrino Observatory: instrumentation and online systems*, Journal of Instrumentation **12** (2017) P03012.

[2] J. Albrecht et al., *The Muon Puzzle in cosmic-ray induced air showers and its connection to the Large Hadron Collider*, Astrophysics and Space Science **367** (2022) 27.

[3] J. J. Thomson, *XL. Cathode Rays*, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science **44** (1897) 293.

[4] E. Fermi, *Tentativo di una Teoria Dei Raggi β*, Il Nuovo Cimento (1924-1942) **11** (1934) 1.

[5] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Physics Letters B **716** (2012) 1.

[6] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Physics Letters B **716** (2012) 30.

[7] S. Tomonaga, *On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields**, Progress of Theoretical Physics **1** (1946) 27.

[8] J. Schwinger, *On Quantum-Electrodynamics and the Magnetic Moment of the Electron*, Phys. Rev. **73** (4 1948) 416.

[9] R. P. Feynman, *Space-Time Approach to Quantum Electrodynamics*, Phys. Rev. **76** (6 1949) 769.

[10] A. Salam et al., *Weak and electromagnetic interactions*, Nuovo Cim. **11** (1959) 568.

[11] S. L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Physics **22** (1961) 579.

[12] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (21 1967) 1264.

[13] H. Fritzsch et al., *Advantages of the color octet gluon picture*, Physics Letters B **47** (1973) 365.

[14] D. J. Gross et al., *Ultraviolet Behavior of Non-Abelian Gauge Theories*, Phys. Rev. Lett. **30** (26 1973) 1343.

[15] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?* Phys. Rev. Lett. **30** (26 1973) 1346.

[16] F. Englert et al., *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (9 1964) 321.

[17] P. W. Higgs, *Broken symmetries, massless particles and gauge fields*, Physics Letters **12** (1964) 132.

[18] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508.

[19] IceCube Collaboration, *Measurement of Atmospheric Neutrino Oscillation Parameters Using Convolutional Neural Networks with 9.3 Years of Data in IceCube DeepCore*, Phys. Rev. Lett. **134** (9 2025) 091801.

[20] R. L. Workman et al., *Review of Particle Physics*, PTEP **2022** (2022) 083C01.

[21] G. Galilei, *Sidereus Nuncius*, Venice: Thomas Baglioni, 1610.

[22] A. Hewish et al., *Observation of a rapidly pulsating radio source*, Nature **217** (1968) 709.

[23] A. A. Penzias et al., *A Measurement of Excess Antenna Temperature at 4080 Mc/s*, The Astrophysical Journal **142** (1965) 419.

[24] C. T. Bolton, *Identification of Cygnus X-1 with HDE 226868*, Nature **235** (1972) 271.

[25] B. P. Abbott et al., *GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral*, Physical Review Letters **119** (2017) 161101.

[26] A. Goldstein et al., *An Ordinary Short Gamma-Ray Burst with Extraordinary Implications: Fermi-GBM Detection of GRB 170817A*, The Astrophysical Journal Letters **848** (2017) L14.

[27] D. A. Coulter et al., *Swope Supernova Survey 2017a (SSS17a), the Optical Counterpart to a Gravitational Wave Source*, Science **358** (2017) 1556.

[28] E. Troja et al., *The X-ray counterpart to the gravitational-wave event GW170817*, Nature **551** (2017) 71.

[29] G. Hallinan et al., *A Radio Counterpart to a Neutron Star Merger*, Science **358** (2017) 1579.

[30] The IceCube Collaboration and Fermi-LAT and MAGIC and AGILE and ASAS-SN and HAWC and H.E.S.S. and INTEGRAL and Kanata and Kiso and Kapteyn and Liverpool Telescope and Subaru and Swift/NuSTAR and VERITAS and VLA/17B-403 teams, *Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A*, Science **361** (2018) eaat1378.

[31] A. Knierim, *Cosmic Messengers*, version v3.0, 2024.

[32] T. K. Gaisser et al., *Cosmic Rays and Particle Physics: 2nd Edition*, Cambridge University Press, 2016.

[33] V. F. Hess, *Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten*, Phys. Z. **13** (1912) 1084.

[34] A. W. Strong et al., *Cosmic-ray propagation and interactions in the galaxy*, Annual Review of Nuclear and Particle Science **57** (2007) 285.

[35] K. Kotera et al., *The astrophysics of ultrahigh-energy cosmic rays*, Annual Review of Astronomy and Astrophysics **49** (2011) 119.

[36] K. Greisen, *End to the Cosmic-Ray Spectrum?* Physical Review Letters **16** (1966) 748.

[37] G. T. Zatsepin et al., *Upper Limit of the Spectrum of Cosmic Rays*, Journal of Experimental and Theoretical Physics Letters (JETP Letters) **4** (1966) 78.

[38] D. E. Nagle, *Delta: the first pion nucleon resonance - its discovery and applications*, tech. rep., Los Alamos National Laboratory, 1984.

[39] M. Aguilar et al., *The Alpha Magnetic Spectrometer (AMS) on the International Space Station: Part I – Results from the test flight on the space shuttle*, Physics Reports **366** (2002) 331.

[40] A. U. Abeysekara et al., *Sensitivity of the High Altitude Water Cherenkov Detector to Sources of Multi-TeV Gamma Rays*, Astroparticle Physics **50-52** (2013) 26.

[41] W. D. Apel et al., *The KASCADE-Grande experiment*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **620** (2010) 202.

[42] The Pierre Auger Collaboration, *Properties and Performance of the Prototype Instrument for the Pierre Auger Observatory*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **523** (2004) 50.

[43] T. K. Gaisser, *Spectrum of cosmic-ray nucleons, kaon production, and the atmospheric muon charge ratio*, Astroparticle Physics **35** (2012) 801.

[44] T. K. Gaisser et al., *Cosmic Ray Energy Spectrum from Measurements of Air Showers*, 2013.

[45] H. Dembinski et al., *Data-driven model of the cosmic-ray flux and mass composition from 10 GeV to* 10^{11} *GeV*, 2017.

[46] M. Benedikt et al., *Future Circular Collider Feasibility Study Report Volume 1*, tech. rep., CERN, 2025.

[47] W. Heitler, *The quantum theory of radiation*, vol. 5, International Series of Monographs on Physics, Oxford: Oxford University Press, 1936.

[48] J. Matthews, *A Heitler model of extensive air showers*, Astroparticle Physics **22** (2005) 387.

[49] L. Neste, *Prompt Atmospheric Muons: An Analysis Concept with IceCube*, MA thesis: TU Dortmund, 2023.

[50] M. Thunman et al., *Charm production and high energy atmospheric muon and neutrino fluxes*, Astroparticle Physics **5** (1996) 309.

[51] A. Fedynitch, *Cascade equations and hadronic interactions at high energies*, PhD thesis, 2015.

[52] F. Riehn et al., *Hadronic interaction model Sibyll 2.3d and extensive air showers*, Phys. Rev. D **102** (2020) 063002.

[53] A. Fedynitch et al., *Hadronic interaction model Sibyll-2.3c and inclusive lepton fluxes*, Physical Review D **100** (2019) 103018.

[54] S. Ostapchenko, *Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model*, Phys. Rev. D **83** (2011) 014018.

[55] T. Pierog et al., *EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider*, Phys. Rev. C **92** (2015) 034906.

[56] The IceCube Collaboration, *Characterization of the atmospheric muon flux in IceCube*, Astroparticle Physics **78** (2016) 1.

[57] IceCube Collaboration, *The design and performance of IceCube DeepCore*, Astroparticle Physics **35** (2012) 615.

[58] IceCube Collaboration, *Measurement of South Pole ice transparency with the IceCube LED calibration system*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **711** (2013) 73.

[59] R. C. Bay et al., *South Pole paleowind from automated synthesis of ice core records*, Journal of Geophysical Research: Atmospheres **115** (2010).

[60] IceCube Collaboration, *IceTop: The surface component of IceCube*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **700** (2013) 188.

[61] IceCube Collaboration, *Observation of high-energy neutrinos from the Galactic plane*, Science **380** (2023) 1338.

[62] IceCube Collaboration, *Detection of a particle shower at the Glashow resonance with IceCube*, Nature **591** (2021) 220.

[63] IceCube Collaboration, *Evidence for neutrino emission from the nearby active galaxy NGC 1068*, Science **378** (2022) 538.

[64] IceCube Collaboration, *Development of an analysis to probe the neutrino mass ordering with atmospheric neutrinos using three years of IceCube DeepCore data*, The European Physical Journal C **80** (2020) 9.

[65] IceCube Collaboration, *IceCube-Gen2: the window to the extreme Universe*, Journal of Physics G: Nuclear and Particle Physics **48** (2021) 060501.

[66] P. A. Cherenkov, *Visible luminescence of pure liquids under the influence of γ-radiation*, Dokl. Akad. Nauk SSSR **2** (1934) 451.

[67] P. A. Cherenkov, *Visible Radiation Produced by Electrons Moving in a Medium with Velocities Exceeding that of Light*, Phys. Rev. **52** (4 1937) 378.

[68] P. J. Mohr et al., *Codata Internationally Recommended 2022 Values of the Fundamental Physical Constants*, 2024.

[69] U. Katz et al., *High-energy neutrino astrophysics: Status and perspectives*, Progress in Particle and Nuclear Physics **67** (2012) 651.

[70] H. Bethe et al., *On the stopping of fast particles and on the creation of positive electrons*, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **146** (1934) 83.

[71] H. Bethe, *Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie*, Annalen der Physik **397** (1930) 325.

[72] H. J. Bhabha, *The Scattering of Positrons by Electrons with Exchange on Dirac's Theory of the Positron*, Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences **154** (1936) 195.

[73] G. Racah, *Sulla Nascita di Coppie per Urti di Particelle Elettrizzate*, Il Nuovo Cimento (1924-1942) **14** (1937) 93.

[74] R. P. Kokoulin et al., *Analysis of cross section of direct pair production by fast muons*, Acta Phys. Acad. Sci. Hung. 29: Suppl. 4, 277-84(1970). (1970).

[75] D. Clayton, *Principles of Stellar Evolution and Nucleosynthesis*, University of Chicago Press, 1983.

[76] IceCube Collaboration, *Observation of Seven Astrophysical Tau Neutrino Candidates with IceCube*, Phys. Rev. Lett. **132** (15 2024) 151001.

[77] IceCube Collaboration, *Search for Astrophysical Tau Neutrinos in Six Years of High-Energy Starting Events in IceCube*, PoS **ICRC2017** (2017) 974.

[78] IceCube Collaboration, *Energy reconstruction methods in the IceCube neutrino telescope*, Journal of Instrumentation **9** (2014) P03009.

[79] IceCube Collaboration, *A convolutional neural network based cascade reconstruction for the IceCube Neutrino Observatory*, Journal of Instrumentation **16** (2021) P07041.

[80] AMANDA Collaboration, *Muon track reconstruction and data selection techniques in AMANDA*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **524** (2004) 169.

[81] IceCube Collaboration, *The IceCube data acquisition system: Signal capture, digitization, and timestamping*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **601** (2009) 294.

[82] T. DeYoung, "IceTray: A software framework for IceCube," *14th International Conference on Computing in High-Energy and Nuclear Physics (CHEP 2005)*, Proceedings of CHEP 2005, 2005.

[83] G. C. Hill, *Experimental and theoretical aspects of high energy neutrino astrophysics*, PhD thesis: University of Adelaide, 1996.

[84] A. Gazizov et al., *ANIS: High energy neutrino generator for neutrino telescopes*, Computer Physics Communications **172** (2005) 203.

[85] IceCube Collaboration, *LeptonInjector and LeptonWeighter: A neutrino event generator and weighter for neutrino observatories*, Computer Physics Communications **266** (2021) 108018.

[86] IceCube Collaboration, *Measurement of the astrophysical diffuse neutrino flux in a combined fit of IceCube's high energy neutrino data*, PoS **ICRC2023** (2023) 1064.

[87] D. Heck et al., *CORSIKA: A Monte Carlo code to simulate extensive air showers*, Technical Report FZKA 6019, Forschungszentrum Karlsruhe, 1998.

[88] K. Bernlöhr et al., *The optical system of the H.E.S.S. imaging atmospheric Cherenkov telescopes*, Astroparticle Physics **20** (2003) 111.

[89] J. Aleksić et al., *The major upgrade of the MAGIC telescopes, Part I: The hardware improvements and the commissioning of the system*, Astroparticle Physics **72** (2016) 76.

[90] J. Holder et al., *The first VERITAS telescope*, Astroparticle Physics **25** (2006) 391.

[91] B. S. Acharya et al., *Introducing the CTA concept*, Astroparticle Physics **43** (2013) 3.

[92] T. Abu-Zayyad et al., *The surface detector array of the Telescope Array experiment*, Nuclear Instruments and Methods in Physics Research Section A **689** (2012) 87.

[93] Z. Cao, *The Large High Altitude Air Shower Observatory (LHAASO)*, Chinese Physics C **43** (2019) 045002.

[94] M. Ageron et al., *ANTARES: The first undersea neutrino telescope*, Nuclear Instruments and Methods in Physics Research Section A **656** (2011) 11.

[95] S. Adrian-Martinez et al., *Letter of intent for KM3NeT 2.0*, Journal of Physics G: Nuclear and Particle Physics **43** (2016) 084001.

[96] A. D. Avrorin et al., *The gigaton volume detector in Lake Baikal: Present and future*, EPJ Web of Conferences **121** (2015) 05003.

[97] S. Fukuda et al., *The Super-Kamiokande detector*, Nuclear Instruments and Methods in Physics Research Section A **501** (2003) 418.

[98] H. Fesefeldt, *Simulation of Hadronic Showers: Physics and Applications*, Nuclear Instruments and Methods in Physics Research Section A **292** (1990) 279.

[99] S. D. Ridder, *Sensitivity of IceCube Cosmic Ray Measurements to the Hadronic Interaction Models*, `http://hdl.handle.net/1854/LU-8609789`, `https://wiki.icecube.wisc.edu/index.php/SouthPole-Atmosphere`, PhD thesis: Ghent University, 2019.

[100] IceCube Collaboration, *Optical properties of deep glacial ice at the South Pole*, Journal of Geophysical Research: Atmospheres **111** (2006).

[101] J. Knapp et al., *Extensive Air Shower Simulation with CORSIKA: A User's Manual*, Visited on 05/06/2024, Kernforschungszentrum Karlsruhe, 1993.

[102] D. Heck et al., *The EHISTORY and MUPROD Options of the Air Shower Simulation Program CORSIKA*, tech. rep., Karlsruher Institut für Technologie (KIT), 2013.

[103] L. Neste, *Panama - A python toolkit for CORSIKA7*, 2023.

[104] L. Neste et al., *Prompt and conventional high-energy muon spectra from a full Monte Carlo simulation via CORSIKA7*, The European Physical Journal C **85** (2025) 759.

[105] K. R. The IceCube Collaboration (K. Meagher et al., *SimWeights: Pure python library for calculating the weights of Monte Carlo simulation for IceCube*, Version 0.1.1, 2023.

[106] E.-J. Ahn et al., *Cosmic ray interaction event generator SIBYLL 2.1*, Phys. Rev. D **80** (2009) 094003.

[107] J.-H. Koehne et al., *PROPOSAL: A tool for propagation of charged leptons*, Computer Physics Communications **184** (2013) 2070.

[108] J.-M. Alameddine et al., *Improvements in charged lepton and photon propagation for the software PROPOSAL*, Computer Physics Communications **302** (2024) 109243.

[109] P. Gutjahr et al., *Simulation of deflection uncertainties on directional reconstructions of muons using PROPOSAL*, The European Physical Journal C **82** (2022) 1143.

[110] C. Kopper, *clsim*, `https://github.com/claudiok/clsim`.

[111] IceCube Collaboration, *The IceCube Realtime Alert System*, Astroparticle Physics **92** (2017) 30.

[112] IceCube Collaboration, *Light propagation in the South Pole ice*, AIP Conference Proceedings **1630** (2014) 146.

[113] IceCube Collaboration, *Efficient propagation of systematic uncertainties from calibration to analysis with the SnowStorm method in IceCube*, Journal of Cosmology and Astroparticle Physics **2019** (2019) 048.

[114] IceCube Collaboration, *The IceCube Upgrade — Design and Science Goals*, PoS (ICRC2019) (2020), Presented at 36th International Cosmic Ray Conference (ICRC 2019).

[115] D. Chirkin, "Evidence of optical anisotropy of the South Pole ice," *33rd International Cosmic Ray Conference*, 2013 0580.

[116] IceCube Collaboration, *In situ estimation of ice crystal properties at the South Pole using LED calibration data from the IceCube Neutrino Observatory*, The Cryosphere **18** (2024) 75.

[117] IceCube Collaboration, *Measuring the optical properties of IceCube drill holes*, EPJ Web of Conferences **116** (2016).

[118] IceCube Collaboration, *The IceProd framework: Distributed data processing for the IceCube neutrino observatory*, Journal of Parallel and Distributed Computing **75** (2015) 198, IceCube Collaboration.

[119] R. J. Rossi, *Mathematical Statistics: An Introduction to Likelihood Based Inference*, Nashville, TN: John Wiley & Sons, 2018.

[120] D. Pandel, *Bestimmung von Wasser- und Detektorparametern und Rekonstruktion von Myonen bis 100 TeV mit dem Baikal-Neutrinoteleskop NT-72*, Ph.D. thesis: Humboldt-Universität zu Berlin, 1996.

[121] IceCube Collaboration, "Event Reconstruction in IceCube Based on Direct Event Re-simulation," *Proceedings of the 33rd International Cosmic Ray Conference (ICRC 2013)*, 2013 0581.

[122] IceCube Collaboration, *Improving Reconstruction of GeV-Scale Neutrinos in IceCube-DeepCore by Direct Event Simulation*, Physics in Canada **74** (2018) 125.

[123] S. Nowicki, "Improving Reconstruction of GeV-Scale Neutrinos in IceCube-DeepCore by Direct Event Simulation," *Proceedings of the XXVIII International Conference on Neutrino Physics and Astrophysics*, 2018 209.

[124] IceCube Collaboration, *Improving the Directional Reconstruction of PeV Hadronic Cascades in IceCube*, EPJ Web of Conferences **207** (2019) 05003.

[125] IceCube Collaboration, *An improved method for measuring muon energy using the truncated mean of dE/dx*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **703** (2013) 190.

[126] K. Hymon, *Measurements of Seasonal Variations of the Unfolded Atmospheric Neutrino Energy Spectrum with IceCube*, MA thesis: Ruhr-Universität Bochum, 2020.

[127] IceCube Collaboration, *Muon Track Reconstruction and Data Selection Techniques in AMANDA*, Nuclear Instruments and Methods in Physics Research Section A **524** (2004) 169.

[128] K. Schatto, *Stacked Searches for High-Energy Neutrinos from Blazars with IceCube*, Ph.D. thesis, defended June 2, 2014, PhD thesis: University of Mainz, 2014.

[129] Royal Swedish Academy of Sciences, *The Nobel Prize in Physics 2024: for foundational discoveries and inventions that enable machine learning with artificial neural networks*, `https://www.nobelprize.org/prizes/physics/2024/summary/`, 2024.

[130] M. Huennefeld, *icecube/dnn_reco: Version 1.0.1*, version v1.0.1, 2023.

[131] D. E. Rumelhart et al., *Learning representations by back-propagating errors*, Nature **323** (1986) 533.

[132] K. Hornik, *Multilayer Feedforward Networks Are Universal Approximators*, Neural Networks **2** (1989) 359.

[133] K. Fukushima, *Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position*, Biological Cybernetics **36** (1980) 193.

[134] V. Nair et al., "Rectified Linear Units Improve Restricted Boltzmann Machines," *Proceedings of the 27th International Conference on Machine Learning (ICML)*, Haifa, Israel: Omnipress, 2010.

[135] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, Gaussian likelihood and negative log-likelihood for regression; see Ch. 4, Cambridge, MA: MIT Press, 2012.

[136] D.-A. Clevert et al., "Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)," *Proceedings of the International Conference on Learning Representations (ICLR)*, 2016.

[137] O. I. Berngardt, *Improving Classification Neural Networks by using Absolute activation function (MNIST/LeNET-5 example)*, 2023.

[138] C. M. Bishop, *Pattern Recognition and Machine Learning*, Intro to squared-error loss and its link to linear regression; see Secs. 1.5 and 3.1, New York, NY: Springer, 2006.

[139] N. Srivastava et al., *Dropout: A Simple Way to Prevent Neural Networks from Overfitting*, Journal of Machine Learning Research **15** (2014) 1929.

[140] Y. LeCun et al., *Backpropagation Applied to Handwritten Zip Code Recognition*, Neural Computation **1** (1989) 541.

[141] Y. LeCun et al., *Gradient-Based Learning Applied to Document Recognition*, Proceedings of the IEEE **86** (1998) 2278.

[142] M. Hünnefeld, *Observation of high-energy neutrinos from the Milky Way*, PhD thesis: Technische Universität Dortmund, 2023.

[143] L. Flottau, *A deep learning based reconstruction of high-energetic muons in IceCube*, Bachelorthesis, 2023.

[144] B. Brandt, *A deep learning based angular reconstruction of atmospheric muons in IceCube*, Bachelorthesis, 2023.

[145] K. Pearson, *VII. Note on regression and inheritance in the case of two parents*, Proceedings of the Royal Society of London **58** (1895) 240.

[146] C. Spearman, *The proof and measurement of association between two things*, The American Journal of Psychology **15** (1904) 72.

[147] IceCube Collaboration, *Measurement of the mean number of muons with energies above 500 GeV in air showers detected with the IceCube Neutrino Observatory*, Phys. Rev. D **112** (8 2025) 082004.

[148] IceCube Collaboration, *An improved mapping of ice layer undulations for the IceCube Neutrino Observatory*, PoS **ICRC2023** (2023) 975.

[149] IceCube Collaboration, *State of the Ice Model in IceCube observatory*, PoS **ICRC2025** (2025) 1013.

[150] IceCube Collaboration, *A muon-track reconstruction exploiting stochastic losses for large-scale Cherenkov detectors*, Journal of Instrumentation **16** (2021) P08034.

[151] IceCube Collaboration, *Seasonal variations of the atmospheric muon neutrino spectrum measured with IceCube*, The European Physical Journal C **85** (2025) 1368.

[152] S. Verpoest et al., *Atmospheric muons and their variations with temperature*, Astroparticle Physics **161** (2024) 102985.

[153] I. Fredholm, *Sur une classe d'équations fonctionnelles*, Acta Mathematica **27** (1903) 365.

[154] A. N. Tikhonov, *On the stability of inverse problems*, Proceedings of the USSR Academy of Sciences **39** (1943) 195.

[155] V. Blobel, "Unfolding methods in high-energy physics experiments," *1984 CERN School of Computing*, CERN-1985-009, DESY-84-118, 1985.

[156] M. Börner, *Bestimmung des Energiespektrums von atmosphärischen Myonneutrinos mit 3 Jahren Daten des IceCube-Detektors*, PhD thesis: Technische Universität Dortmund, 2018.

[157] M. Börner et al., *funfolding: Poisson-likelihood unfolding framework*, GitHub repository, 2025.

[158] F. James et al., *Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations*, Computer Physics Communications **10** (1975) 343.

[159] H. Dembinski et al., *scikit-hep/iminuit*, 2020.

[160] IceCube Collaboration, "Unfolding the Atmospheric Muon Flux with IceCube: Investigating Stopping Muons and High-Energy Prompt Contributions," *Proceedings of 39th International Cosmic Ray Conference — PoS(ICRC2025)*, vol. 501, 2025 281.

[161] S. Roesler et al., "The Monte Carlo event generator DPMJET-III," *Advanced Monte Carlo for Radiation Physics, Particle Transport Simulation and Applications*, 2001 1033.

[162] IceCube Collaboration, *Measurement of the Diffuse Astrophysical Neutrino Spectrum above a TeV with All Flavor Starting Events in IceCube*, PoS **ICRC2025** (2025) 985.

[163] K. Pearson, *On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*, Philosophical Magazine Series 5 **50** (1900) 157.

[164] P. Virtanen et al., *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, Nature Methods **17** (2020) 261.

[165] H. Chernoff, *On the Distribution of the Likelihood Ratio*, The Annals of Mathematical Statistics **25** (1954) 573.

[166] M. Reininghaus et al., "CORSIKA 8 – Towards a modern framework for the simulation of extensive air showers," *EPJ Web of Conferences*, ed. by I. L.-Y. et al., vol. 210, 2019 02011.

[167] T. Huege et al., *The particle-shower simulation code CORSIKA 8*, (2023).

[168] R. Engel et al., *Towards a next generation of CORSIKA: a framework for the simulation of particle cascades in astroparticle physics*, Computing and Software for Big Science **3** (2019) 1.

[169] F. Riehn et al., *Sibyll*★, Astroparticle Physics **160** (2024) 102964.

[170] T. Pierog et al., "EPOS LHC-R: up-to-date hadronic model for EAS simulations," *PoS ICRC2023*, 2023 230.

[171] T. Pierog et al., "EPOS.LHC-R: a global approach to solve the muon puzzle," *PoS ICRC2025*, 2025 358.

[172] S. Ostapchenko, *QGSJET-III model of high energy hadronic interactions: II. Particle production and extensive air shower characteristics*, Phys. Rev. D **109** (2024) 094019.

[173] C. Bierlich et al., *A comprehensive guide to the physics and usage of PYTHIA 8.3*, SciPost Phys. Codeb. **2022** (2022) 8.

[174] T. Sjöstrand et al., *Hadron interactions for arbitrary energies and species, with applications to cosmic rays*, Eur. Phys. J. C **82** (2022) 21.

[175] H. Dembinski et al., *Global Spline Fit GSF-2025 - An update of the data-driven model of the cosmic-ray flux and its mass composition*, PoS **ICRC2025** (2025) 248.

[176] IceCube Collaboration, *Constraining the Prompt Atmospheric Neutrino Flux Combining IceCube's Cascade and Track Samples*, (2024).

[177] L. Kardum, *Unfolding the muon neutrino fux*, PhD thesis: TU Dortmund University, 2023.

[178] L. Flottau, *A forward folding based approach to measuring the prompt component of the atmospheric muon flux*, Masterthesis, 2025.

[179] C. A. Argüelles et al., *A binned likelihood for stochastic models*, Journal of High Energy Physics **2019** (2019) 30.

[180] OpenAI, *ChatGPT*, Large Language Model, 2025.

[181] Grammarly Inc., *Grammarly*, AI Writing Assistant, 2025.

[182] DeepL SE, *DeepL Translator*, AI Translator, 2025.

[183] Google, *Gemini*, Large Language Model, 2025.

List of Abbreviations

Abs absolute value

AGN active galactic nucleus

AMS-02 Alpha Magnetic Spectrometer

ANIS All Neutrino Interaction Simulation

CC charged current

CKM Cabibbo-Kobayashi-Maskawa

CMB cosmic microwave background

CORSIKA COsmic Ray SImulations for KAscade

CR cosmic ray

DNN deep neural network

DOM digital optical module

EAS Extensive Air Shower

EHISTORY Extended History

ELU exponential linear unit

EM electromagnetic

GRB gamma-ray burst

GSF Global Spline Fit

GST Gaisser-Stanev-Tilav

GZK Greisen-Zatsepin-Kuzmin

H3a Gaisser-Hillas 3

H4a Gaisser-Hillas

HAWC High Altitude Water Cherenkov Experiment

HGC Hadron Generation Counter

HIM hadronic interaction model

ICL IceCube Lab

KASCADE-Grande KArlsruhe Shower Core and Array DEtector-Grande

KDE kernel density estimator

LHC Large Hadron Collider

LIGO Laser Interferometer Gravitational-Wave Observatory

MC Monte Carlo Simulation

MCEq Matrix Cascade Equations

ML Machine Learning

MLE maximum likelihood estimation

MLP multilayer perceptron

MPE multi-photo-electron

MSE mean squared error

NC neutral current

NuGen Neutrino Generator

NuSim Neutrino Simulation

OF overflow

PANAMA PANdas And Multicore utils for corsikA 7

PDF probability density functions

PID Particle ID

PMT photomultiplier tube

PROPOSAL PRopagator with Optimal Precision and Optimized Speed for All Leptons

QCD Quantum Chromodynamic

ReLU Rectified Linear Units

SGD stochastic gradient descent

SM Standard Model of Particle Physics

SMT12 Simple Multiplicity Trigger 12

SMT8 Simple Multiplicity Trigger 8

SNR supernova remnant

UF underflow

UHECR ultra-high-energy cosmic ray

WHISP Working group for Hadronic Interactions and Shower Physics

Declaration on the Use of Generative AI

During the preparation of this thesis, the author used modern AI-assisted tools to improve language, wording, grammar, and readability. Specifically, the services ChatGPT [180], Grammarly [181], DeepL [182], and Gemini [183] were utilized. The author reviewed and edited the output and takes full responsibility for the content of this work.

Acknowledgements

I would like to express my deepest gratitude to everyone who has supported me throughout my PhD journey:

- My supervisor, Prof. Dr. Dr. Wolfgang Rhode, for his encouragement throughout my research. He was always available for discussions, provided valuable feedback, and guided me through my work. I am especially grateful for his extreme support during the entire time I have spent in his group, and for giving me the freedom to pursue my own research ideas. He gave me a lot of trust and responsibility, which I am very thankful for. This helped me not only to grow as a scientist, but also to gain confidence in my own abilities. And I am so thankful for all the opportunities to attend meetings, workshops and conferences, which helped me to build an international network in science. I very much enjoyed my time in your group. Thank you!

- Prof. Dr. Kevin Kröninger, for reviewing my thesis to ensure the quality of my work. I enjoyed my time during my Bachelor's studies at his chair, and I always appreciated his kindness and support.

- PD. Dr. Dominik Elsässer, who was always immediately available for all kinds of questions, no matter if it was day or night! His support was always more than 100 %, and he tried to make everything possible to help me reach my goals.

- The entire APP group in Dortmund for so many different reasons! Of course, I am grateful for the scientific discussions and collaborations, but I also want to thank you for the great atmosphere in the group. I always felt very welcome and supported, and I am happy to be part of this community.

- Spencer Klein, for his support and advice during my research stay at the University of California, Berkeley. I am very grateful for the opportunity to work with him and his group, and for the valuable insights he provided on my research. I felt welcome from the very first moment, and I am always happy to have had the chance to collaborate with him, and hopefully also in the future.

- Dennis and Anatoli, for help and support during the entire time of my PhD, and also for helping me to finish the work within the IceCube collaboration. I still enjoy remembering the "low hanging fruits"! And of course, I had a lot of fun with you at the conferences and meetings. But you have to work on your foosball table skills :D

- The entire IceCube collaboration, and especially the cosmic ray working group, for the great collaboration and all the discussions about my analysis that helped me to finish this work. And when you are 10 min over time, you are still below the average :D

- Christian Glaser and Steven Barwick, for very inspiring discussions that led to my decision to stay in science and start a postdoc position. Your way of thinking and Steve's love for physics is really inspiring!

- Gary and Sabrina, who gave me their trust and offered me a postdoc position, even though the PhD thesis and the analysis wasn't finished. My arrival in Adelaide was great and I am looking forward to the next chapter of my scientific career.

- Mirco and Ludwig, for so much. I love to remember the first time we met, and how I thought: oh my god! Mirco never talked at all, and Ludwig was such a geek. But after spending too much time together in the office, we not only became great working partners, but also good friends. We made so many great memories together, and even though we are living in three different time zones, we still keep in touch, and every time we meet, it feels like we just saw each other yesterday. This will be a friendship for life, and I am so grateful to have you both in my life!
- Lucas, Lene, Karo, Stefan and Jean-Marco for being always there for me, and joining me in the good and bad times. It was always so much fun to complain about the same things that we all had to deal with. Thank you for keeping my spirits up!
- All my friends: Juli, Nils, Philipp, Leon, Lars, Kai, Cyrus, Salvo, Michi, Lukas, Domenic, Timo, Chris, Cori, Becki, and Nathalie for their support and understanding during the last years. Every time we meet I have so much fun and I can just enjoy the moment. I am so grateful to have you in my life, and I am looking forward to many more years of friendship with you! And since all of you know my forgetfulness, I am sure that I forgot to mention some of you, but love you all!
- Lars aka *Reudiger, Rüdiger and dummski*, for being the best friend I could ask for. We have been through the entire studies from the very beginning together, and it is unbelievable what kind of friendship we have established over the years. I cannot mention everything I am grateful for, but I am so thankful to have you in my life, and I am looking forward to many more years of this *once-in-a-lifetime* friendship with you!...anscheinend muss das alles so!
- My family, for their unconditional love and support. I am so grateful to have you in my life. You supported me in every decision I made, and you always believed in me, without any doubt. Thank you so much! I love you!